AUGUST 1994
$3.95

SCIENTIFIC AMERICAN

Red tides—a growing hazard.
The extreme ultraviolet universe.
SQUIDs for ultrafaint signals.

***The daily grind** of preparing flour left its mark on Neolithic bones.*

SCIENTIFIC AMERICAN

August 1994 Volume 271 Number 2

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Second-class postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. Reprints available: Write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111; fax: (212) 355-0408 or send E-mail to SCAinquiry@aol.com.

DEPARTMENTS

THE COVER painting evokes a daily task that left strong marks on the bones of Neolithic women. The task was grinding grain on a stone quern, shaped like a saddle so it could contain the grain and flour. Working for hours on her knees, a woman would push the rubbing stone forward to the far end of the quern and pull it back. In doing so, she put constant strain on the bones and joints of her back, arms, thighs, knees and toes. The work caused structural damage and arthritis (see "The Eloquent Bones of Abu Hureyra," by Theya Molleson, page 70).

THE ILLUSTRATIONS

Cover painting by Alfred T. Kamajian

Page	Source
27	Jane's Information Group
28	George Retseck
32–33	Center for EUV Astrophysics
34	Guilbert Gates/JSD
35	Center for EUV Astrophysics
36–38	Jared Schneidman/JSD (*drawings*)
40–41	Jeff W. Lichtman
42	Matthew H. Chestnut
43	Stephen J Smith and Michael E. Dailey (*top*), Jeff W. Lichtman (*bottom*)
44	Jared Schneidman/JSD (*drawings*), Jeff W. Lichtman and Susan Culican (*photos*)
45	Jeff W. Lichtman
46	Ian Worpole
47	David Scharf
48–49	Ian Worpole
50	Ian Worpole (*top*), Non Fan and John Clarke (*bottom*)
51	Ian Worpole
52	R. C. Black and F. C. Wellstood, University of Maryland
53	Christopher C. Gallen, Scripps Research Institute, and Eugene C. Hirschkoff, Biomagnetic Technologies
55	Dennis Kunkel/ Phototake, Inc.
56–57	Dimitry Schidlovsky
58–59	Guilbert Gates/JSD (*top*), Paul Travers, Birkbeck College (*middle*), Dimitry Schidlovsky (*bottom*)
60–61	Dimitry Schidlovsky
63	Susan Aviation, Inc.
64	Donald M. Anderson (*left* and *right*), David Wall, AMACO (*center*)
65	Greg Early, New England Aquarium
66	Jared Schneidman/JSD
67	H. Robert Guy, National Institutes of Health (*top*), M. Caruso, Woods Hole Oceanographic Institution (*bottom*)
68	Johnny Johnson after Gustaaf M. Hallegraeff, University of Tasmania
70	Roberto Osti
72–74	Roberto Osti (*top*)
75	Roberto Osti
76–77	Michael Hart/FPG
78	Science Photo Library/ Custom Medical Stock Photo (*left*), Gianne Carvalho/ Impact Visuals (*center*), Steve Winter/Black Star (*right*)
79	Catherine Leroy/SIPA (*left*), Mark Peterson/SABA (*center*), Carolina Kroon/ Impact Visuals (*right*)
80	Weybridge/SPL/Photo Researchers, Inc. (*left*), Mike Goldwater/Network/Matrix (*center*), Paul Harrison/Still Pictures (*right*)
81	Donna Ferrato, courtesy of Domestic Abuse Awareness Project (*left*), Mark Edwards/ Still Pictures (*center*), Judy Griesedieck/Black Star (*right*)
82	Mark Edwards/Still Pictures (*left*), Kevin Beebe/Custom Medical Stock Photo (*center*), Malcolm Linton/Black Star (*right*)
83	P. Motta/SPL/Photo Researchers, Inc. (*left*), Mark Edwards/Still Pictures (*right*)
90–92	Andrew Christie

SCIENTIFIC AMERICAN®

Established 1845

EDITOR: Jonathan Piel

BOARD OF EDITORS: Michelle Press, *Managing Editor;* John Rennie, *Associate Editor;* Timothy M. Beardsley; W. Wayt Gibbs; Marguerite Holloway; John Horgan, *Senior Writer;* Kristin Leutwyler; Philip Morrison, *Book Editor;* Madhusree Mukerjee; Corey S. Powell; Ricki L. Rusting; Gary Stix; Paul Wallich; Philip M. Yam

ART: Joan Starwood, *Art Director;* Edward Bell, *Art Director, Graphics Systems;* Jessie Nathans, *Associate Art Director;* Johnny Johnson, *Assistant Art Director, Graphics Systems;* Nisa Geller, *Photography Editor;* Lisa Burnett, *Production Editor*

COPY: Maria-Christina Keller, *Copy Chief;* Nancy L. Freireich; Molly K. Frances; Daniel C. Schlenoff

PRODUCTION: Richard Sasso, *Vice President, Production;* William Sherman, *Production Manager; Managers:* Carol Albert, *Print Production;* Janet Cermak, *Makeup & Quality Control;* Tanya DeSilva, *Prepress;* Carol Hansen, *Composition;* Madelyn Keyes, *Systems;* Ad Traffic: Carl Cherebin; Kelly Ann Mercado

CIRCULATION: Lorraine Leib Terlecki, *Associate Publisher/Circulation Director;* Katherine Robold, *Circulation Manager;* Joanne Guralnick, *Circulation Promotion Manager;* Rosa Davis, *Fulfillment Manager*

ADVERTISING: Kate Dobson, *Associate Publisher/Advertising Director.* OFFICES: NEW YORK: Meryle Lowenthal, *New York Advertising Manager;* Randy James, Elizabeth Ryan. CHICAGO: 333 N. Michigan Ave., Chicago, IL 60601; Patrick Bachler, *Advertising Manager.* DETROIT: 3000 Town Center, Suite 1435, Southfield, MI 48075; Edward A. Bartley, *Detroit Manager.* WEST COAST: 1554 S. Sepulveda Blvd., Suite 212, Los Angeles, CA 90025; Lisa K. Carden, *Advertising Manager;* Tonia Wendt. 235 Montgomery St., Suite 724, San Francisco, CA 94104; Lianne Bloomer. CANADA: Fenn Company, Inc. DALLAS: Griffith Group

MARKETING SERVICES: Laura Salant, *Marketing Director;* Diane Schube, *Promotion Manager;* Wendy Robinson, *Research Manager;* Nancy Mongelli, *Assistant Marketing Manager;* Ethel D. Little, *Advertising Coordinator*

INTERNATIONAL: EUROPE: Roy Edwards, *International Advertising Manager,* London; Vivienne Davidson, Linda Kaufman, Intermedia Ltd., Paris; Karin Ohff, Groupe Expansion, Frankfurt; Barth David Schwartz, *Director, Special Projects,* Amsterdam. SEOUL: Biscom, Inc. TOKYO: Nikkei International Ltd.; SINGAPORE: Hoo Siew Sai, Major Media Singapore Pte. Ltd.

ADMINISTRATION: John J. Moeling, Jr., *Publisher;* Marie M. Beaumonte, *General Manager*

SCIENTIFIC AMERICAN, INC.
415 Madison Avenue, New York, NY 10017-1111
(212) 754-0550

CHAIRMAN AND CHIEF EXECUTIVE OFFICER: John J. Hanley

CO-CHAIRMAN: Dr. Pierre Gerckens

CORPORATE OFFICERS: *President,* John J. Moeling, Jr.; *Chief Financial Officer,* R. Vincent Barger; *Vice Presidents,* Robert L. Biewen, Jonathan Piel

DIRECTOR, ELECTRONIC PUBLISHING: Martin Paul

CHAIRMAN EMERITUS: Gerard Piel

  PRINTED IN U.S.A.

LETTERS TO THE EDITORS

Stale Bread Mystery

Thank you for the delightful and informative article "Chemistry and Physics in the Kitchen," by Nicholas Kurti and Hervé This-Benckhard [SCIENTIFIC AMERICAN, April]. It's just the thing to amuse and console a lot of us physicists who are looking at other fields as we see our own evaporating (or in culinary terms, "reducing").

Perhaps the authors can help with a problem that has troubled me for years. The science of bread making has made progress in understanding how the gluten protein in flour is converted to give chewable bread with a tender crumb. We know that the sugars on the exterior caramelize to produce a golden-brown crust. We know that gluten gives the dough body and holds it together until baked. We know that the heat of baking alters the molecular bonds so the finished bread remains moist but no longer tough and elastic like the raw dough.

But what happens when slightly stale bread is freshened in a microwave oven? A conventional oven somehow partially restores the moist, tender constitution of fresh bread. A microwave oven, on the other hand, restores the moistness but also revives the undesirable toughness and elasticity of the gluten in the raw dough.

GERALD T. DAVIDSON
Menlo Park, Calif.

Kurti and This-Benckhard reply:

We inquired at the INRA Center in Nantes, where the laboratory of cereal technology is headed by Bernard Godon. Unfortunately, this effect has not yet been studied.

It is clear that in stale bread, water bound to the carbohydrates in a gel is slowly lost to either the air or the gluten network. When heated, the water bound to the gluten is taken up again by the carbohydrates, which partially gel. Heat enters the bread differently in the two types of ovens, however. The traditional oven creates a strong temperature gradient because the bread is a poor conductor of heat. The microwave oven heats the bread uniformly because the bread absorbs the energy directly. The microwaves can be absorbed by both the water and gluten molecules. Yet the behavior of the water can depend on whether it is bound to the carbohydrates or the gluten. These variables could affect the freshening of stale bread.

Prostate Cancer Screening

The impact of Marc B. Garnick's "The Dilemmas of Prostate Cancer" [SCIENTIFIC AMERICAN, April] went miles beyond the scope of most magazine articles. This one is literally a lifesaver. A friend sent the story to me from California. I was galvanized into being tested and discovered a cancer-causing polyp. I passed the article on to two friends, who had tests showing that both had prostatic malignancies. Because of your powerful story, we became some of the lucky ones: we can now do something about our problems.

Let's hope your article will impel researchers to get busy with serious study of this unglamorous disease.

SAMUEL A. HOUSTON
Houston, Tex.

Implicit in Garnick's endorsement of the screening recommendations of the American Cancer Society is a radical departure from the traditional medical ethic "first do no harm." Translated into a basic principle for the mass screening of asymptomatic individuals, that ethic means: do not recommend screening unless there is an effective proven treatment whose benefit outweighs the harm. As Garnick points out, the benefit/harm ratio of prostate-specific antigen (PSA) screening cannot be calculated at this time, because *there is no proven benefit.* We physicians must inform patients of that fact before asking them to consent to PSA testing.

DAVID L. HAHN
Madison, Wis.

All the scientific studies cited in the article recommended less aggressive treatment of mild disease. Yet Garnick favors aggressive treatment. Where are the data to substantiate his contention that the average patient in the U.S. benefits from early surgery for cancers detected by the PSA assay? Physicians in Europe use the PSA test less aggressively than those in the U.S.

Do you think the Food and Drug Administration would approve a new drug that rendered 70 to 80 percent of patients impotent, as early surgery does, based on the currently available data regarding its effectiveness?

MICHAEL D. SWEET
San Diego, Calif.

Garnick replies:

It may take years before the true value of screening becomes known. We are now witnessing, however, more men being diagnosed at a much less advanced stage of disease when their cancer is detected through PSA screening. It will probably require years of follow-up before the benefit of improved survival is realized through treating these patients at an earlier stage. Early diagnoses of breast and colon cancers have raised survival rates, but those benefits also sometimes did not appear until years later. On the basis of what is known today, some patients will decide early treatment is worthwhile; others will not. Many diseases that are vigorously treated in the U.S. do not receive the same attention in other countries.

Data suggest that prostate cancer, when detected and treated early, can be cured. If suffering and premature death can be avoided through early diagnosis and treatment, a physician will have behaved honorably. As recently stated in a national meeting on prostate cancer, the 70-year-old man dying of metastatic disease was probably at age 50 a man with a curable prostate cancer.

Letters selected for publication may be edited for length and clarity. Unsolicited manuscripts and correspondence will not be returned or acknowledged unless accompanied by a stamped, self-addressed envelope.

AMPLIFICATION

The biography box for "The Molecular Architects of Body Design" [February] neglected to mention that William McGinnis and Michael Levine collaborated on the homeobox discovery with Walter J. Gehring in his laboratory in Basel. The text should also have mentioned that the first "redesign" of the *Drosophila* body plan with an inducible promoter directing ectopic expression of *Antennapedia* was done by Gehring, Stephan Schneuwly and Roman Klemenz in 1987.

50 AND 100 YEARS AGO

AUGUST 1944

"Engineers for years have sought a practical method of gasoline injection for supplying fuel to the cylinders of gasoline engines. Such a method has now been perfected and is in production, according to Donald P. Hess, President of American Bosch Corporation. 'The gasoline, by this system, is delivered uniformly to every cylinder of the engine. The result is that all cylinders pull together in harmony, producing a smoother flow of power and quieter engine operation than has ever been possible with any other method,' Mr. Hess states."

"Cereals disguised as candy bars are the latest idea of the food industry, determined to make us eat cereals whether we want to or not."

"Sorting of mail electronically could be accomplished if a row or rows of black and white squares were used to designate the first main geographical subdivision in addresses. A second row would identify the postal substation and a third row the city postal carrier district. Envelopes could then be run through a scanning machine. As the letter whisked in front of the electric-eye, the machine would do the equivalent of reading the address in the coded squares and then automatically route the letter to the correct mail bag or container. This would be repeated for the second row and again for the third row when the letter arrived in the final postal sub-district. Thus it would have to be looked at only by the carrier."

"Magnesium threatens to take the place of celluloid as the most feared flammable material used in industry. The National Board of Fire Underwriters is preparing special data to show factories how to control this hazard. Absent from this will be the weird tales of factories which forbade their women operators to wear silk panties (if they could get any) lest sparks from friction set off the magnesium chips in their lathes."

SCIENTIFIC AMERICAN

AUGUST 1894

"June 30, 1894, was a gala day in London, the occasion being the opening of a new bridge over the Thames River located near the Tower. It is a heavy piece of work, occupying much more valuable space than was necessary. But it was considered by those who had the say that such a work, located, as it was, near the historical Tower of London, ought to be massive, and present a mediaeval architectural look. So they sank a pair of great piers in the narrow river, erected strong steel frames thereon to carry the cables and other parts, and then clothed the steel work with a shell of stone, the work, as a whole, being thus made to represent a structure of massive masonry."

"In writing of the last Royal Society conversazione, the *Lancet* mentioned an invention by Mr. C. T. Snedekor for heating by electricity a quilt or cushion. This quilt, which he named the thermogen, the *Lancet* has since had an opportunity of putting to practical trial, and has no hesitation in reporting upon it thoroughly favorably as an appliance that might be of great value in all hospitals or, for that matter, in all private houses where an electric main is handy."

"The citizens of Buffalo, N.Y., were treated to a remarkable mirage between 10 and 11 o'clock on the morning of August 16. It was the city of Toronto, with its harbor and small island to the south of the city. Toronto is fifty-six miles from Buffalo, but the church spires could be counted with the greatest ease. This mirage is what is known as a mirage of the third order. That is, the object looms up far above the real level and not inverted, as is the case with mirages of the first and second class, but appearing like a perfect landscape far away in the sky."

"As plainly shown in the illustration, a boat invented by Mr. H. B. Ogden, No. 204 Carroll Street, Brooklyn, N.Y., is propelled through the water in the same manner as one propels a bicycle on land. The boat is a long, easy running one, with the propelling machine dropped through its bottom into a very small brass boat or fin keel, large enough for the pedals. As shown in the sectional view at the top, the pedal cranks turn a gear which meshes into a worm of long pitch on the screw shaft; steering is effected by a rudder connected with the forward handle. These boats are designed to furnish a delightful means of recreation and healthful exercise, as well as serve useful purposes. Especial advantages are claimed for these boats for gunning service, as they are quiet, may be run fast, and the hands may be freed to use the gun at any time."

Ogden's marine velocipede, or bicycle boat

Star Gobbler

A black hole is identified in the core of the galaxy M87

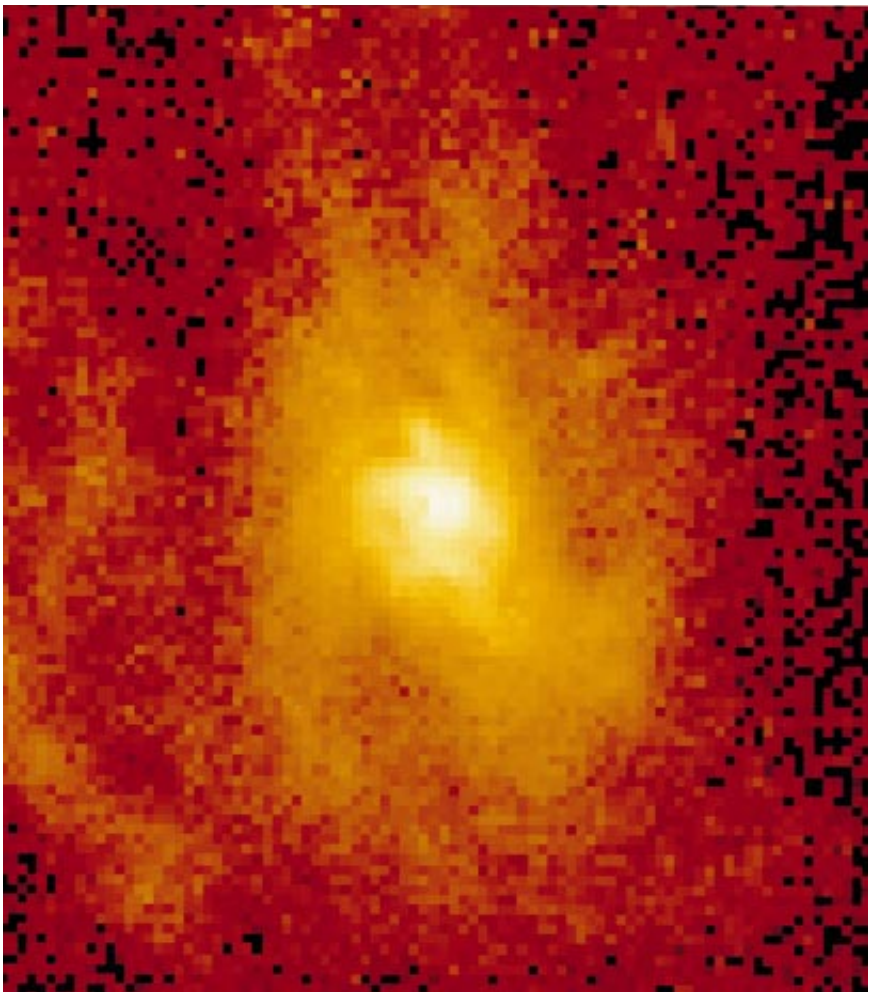
Scientists may not believe in monsters, but many astronomers believe—in the metaphoric sense—that ravenous beasts truly exist at the centers of some galaxies. These cosmic creatures are giant black holes, collapsed objects having millions or even billions of times the mass of the sun packed into a space no larger than our solar system. The gravitational field of such objects is so powerful that matter and even light that fall in cannot return to the outside universe.

For three decades, astronomers have eagerly sought signs that monster black holes were more than a figment of their imaginative theorizing. Now the *Hubble Space Telescope* has provided the strongest sign yet that these objects are indeed real. A team of astronomers led by Holland Ford of the Space Telescope Science Institute in Baltimore and Richard Harms of the Applied Research Corporation in Landover, Md., carried out the observations.

The scientists used *Hubble* to study the inner regions of M87, a huge elliptical galaxy located in the Virgo Cluster, some 50 million light-years from the earth. There they happened on a previously unknown disk of gas that, 60 light-years from its center, is whirling at a speed of 750 kilometers per second, some 25 times the velocity at which the earth orbits the sun.

From that exceedingly rapid motion, Harms and his colleagues estimate that the gas is orbiting a central mass possessing between two billion and three billion solar masses. The disk is oriented roughly perpendicular to the gas jets that shoot from the center of M87, exactly as astrophysical theory predicts. "All the evidence just fits together—it's kind of amazing!" Harms marvels.

"Many of us have believed in black holes for circumstantial evidence—this strengthens the evidence," says Martin Rees of the University of Cambridge, who traditionally takes a cautious view toward findings about black holes. Tod R. Lauer of the National Optical Astronomy Observatories, who has used *Hubble* before to probe the inner regions of M87, assumes a more definite stance. "I'd bet a good bottle of scotch, a good dinner and a trip to Hawaii" that the black hole is real, he says.



HOLLAND FORD AND RICHARD HARMS *NASA*

ROTATING DISK at the heart of the galaxy M87 was discovered using the* Hubble Space Telescope. *The hot gas probably orbits an unseen black hole at the center.

The new observations come 30 years after Edwin E. Salpeter, now at Cornell University, and the late Soviet astrophysicist Yakov B. Zel'dovich proposed that matter falling into black holes could power quasars and radio galaxies. As astronomers came to suspect that quasars merely represent an extremely active period in the early development of many galaxies, they realized that dormant black holes must remain in the cores of most large galaxies.

The long, radio-emitting jet of gas emanating from the center of M87 pegged the galaxy as a particularly likely place to find a massive black hole. In 1978 Peter Young of the California Institute of Technology conducted studies of stellar motion in the core of M87 that hinted at stars crowding around such an object. Images made by Lauer using *Hubble* before its recent optical fix strengthened the case. But the gas disk found by Ford and Harms and their co-workers presents a much more convincing argument. Rather than having to measure the motions of stars near the hole—a messy and inconclusive process—they could make a much simpler measurement of the rotation of what seems to be a single rotating disk. "Nature has given us a nice clean system here," Harms comments.

Alas, the search for black holes still fundamentally relies on indirect clues. Even the repaired *Hubble* cannot resolve the black hole itself; the hole

should measure about five billion kilometers in radius, 1/100,000th the size of the part of the disk seen by *Hubble*. But the small size and rapid motion of the disk effectively rule out just about any object except for a black hole. For instance, some devil's advocates have proposed that the concentrations of mass at the centers of some galaxies could be tightly bound clusters of faint, dense neutron stars or white dwarf stars; given the new observations of M87, "I don't think that's plausible anymore," Harms says.

Harms and his colleagues plan a follow-up *Hubble* session to determine velocities deeper in the disk, which should yield a nearly airtight case for the black hole. Astronomers can then ponder whether the seemingly exotic monster black holes are really a rather ordinary result of the way galaxies form. Rees, for instance, argues that massive black holes probably developed routinely during the process in which vast gas clouds gathered together into galaxies in the early universe, billions of years ago. "This has been fun, but I wouldn't mind seeing a second black hole," Harms laughs. "It's pretty hard to generalize from just a sample of one."

So will the black hole hunt never end? "The public doesn't understand what a human enterprise science is," Lauer muses. "It's like following Columbo on the chase. That's where the real excitement is." Black holes, well-camouflaged monsters that they are, will be keeping astronomers entertained for quite some time to come. —*Corey S. Powell*

Darling *Clementine?*

NASA-DOD tension may orphan the little probe that could

Clementine is a lightweight, low-cost, high-tech spacecraft that has produced the first comprehensive look at the moon since the termination of the Apollo missions more than 20 years ago. It is also a living (if limping) embodiment of the "better, faster, cheaper" mantra espoused by National Aeronautics and Space Administration head Daniel S. Goldin. So how is it possible that *Clementine* may be both the first and last of its breed?

Part of the answer lies in the craft's parentage. *Clementine* was built not by NASA but by the BMDO (Ballistic Missile Defense Organization—son of Star Wars) as a test bed for such antimissile technology as target acquisition and tracking equipment. At the same time, however, it was designed to produce scientific results useful to the civilian community.

Researchers involved with *Clementine* sound uniformly thrilled by the experience of working with the Department of Defense. Paul Spudis of the Lunar and Planetary Institute in Houston relates that planners at the BMDO "have bent over backward to accommodate every scientific request." Eugene Shoemaker of the U.S. Geologic Survey, who led the *Clementine* scientific team, also praises the efficient manner in which the spacecraft was built and managed. BMDO claims that it completed *Clementine* in two years at a cost of $75 million; both figures are a small fraction of those typical for NASA probes.

The outpouring of affection becomes even more apparent when *Clementine* scientists describe the mission's results. "The data from the moon are fantastically great," Spudis exults. *Clementine*'s most significant product is a digital map of the moon made at 11 separate wavelengths. Planetary scientists will be able to correlate the colors of the lunar surface seen on that map with studies of lunar samples returned by the Apollo missions. The product will be a vastly improved understanding of the distribution of rock types and, by extension, the geologic evolution of the moon.

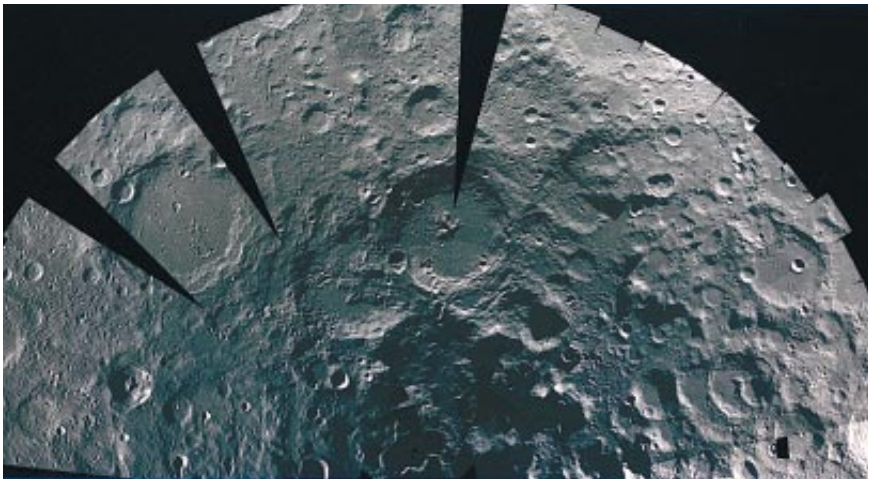
Clementine also conducted detailed studies of the moon's topography and gravitational field. David E. Smith of the NASA Goddard Space Flight Center reports that the range of elevations on the moon is much greater than scientists had realized. In particular, *Clementine* has revealed the surprising extent of the Aitken Basin near the south pole on the lunar farside. This basin, which averages 14 kilometers deep across a quarter of the moon's circumference, is one of the largest formations of its type in the solar system.

Cost and weight considerations lead to scientific trade-offs. For example, *Clementine* lacks a gamma-ray spectrometer, which could have searched for ice lining the shadowed craters at the moon's south pole. And the scientific part of the mission received a blow on May 7, when a software glitch sent *Clementine* into a spin. That accident scuttled the most exciting item on the spacecraft's agenda: a close encounter with the asteroid Geographos, one of the small rocky bodies whose orbits carry them perilously near the earth.

Stewart Nozette of the BMDO, who is the Clementine mission manager, claims that workers have identified the bug in the software and that fixes are in the works. Such mishaps are endemic among complicated robotic probes (recall the recent loss of the *Mars Observer* and the stuck antenna on *Galileo*). But *Clementine* has cost less than one tenth as much as those missions.

Will the Clementine concept overcome its political hurdles? NASA seems uncomfortable about embracing a project whose technology and can-do spirit come from the dark side. At the same time, the BMDO has distanced itself from the mission, leaving *Clementine* a bit of an orphan.

But a funding crunch looming in 1995 intensifies the long-simmering sense that NASA must radically change course if space science is to survive. Shoemaker judges *Clementine* to be "the wave of the future." Nozette acknowledges the "age-old rivalry between NASA and DOD" but sees an even deeper historical bond. "This is like old-style, 19th-century research," he reflects. "It's like Captain Cook taking the astronomers with him." —*Corey S. Powell*



NAVAL RESEARCH LABORATORY/DOD/NASA

PERMANENTLY SHADOWED CRATERS at the lunar south pole, seen in this mosaic view from* Clementine, *may contain hidden deposits of ice.

Radon's Risks

Is the EPA exaggerating the dangers of this ubiquitous gas?

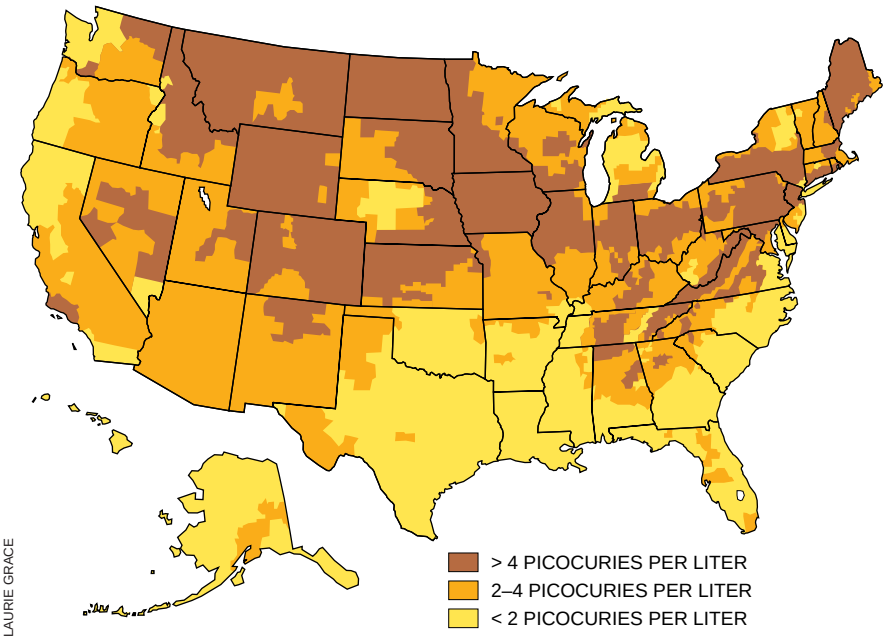
This very moment you are breathing radon, a naturally occurring gas generated by the decay of trace amounts of uranium found throughout the earth's crust. Should you be concerned? The Environmental Protection Agency thinks so. The agency has declared that five million or so of the nation's 80 million homes may have indoor radon levels that pose an unacceptably high risk of lung cancer to occupants.

The EPA has recommended that all homes be tested for radon and that they be structurally altered to reduce exposure should levels exceed a certain threshold established by the agency. Some scientists have challenged the EPA's recommendations, which could cost homeowners and landlords more than $50 billion if carried out. Critics claim that scientific data gathered to date do not support the EPA's estimates of the health risks from radon.

This issue can be traced to studies done decades ago showing that radon might be responsible for unusually high rates of cancer suffered by miners—particularly uranium miners. Whereas outdoor radon levels generally measure less than 0.5 picocurie per liter (pCi/L) of air, miners were often exposed to levels hundreds or even thousands of times higher. (A picocurie is a trillionth of a curie, which is the amount of radioactivity emitted by a gram of radium.)

Some 15 years ago tests revealed that radon seeping into homes and other buildings through fissures in foundations often accumulates to levels considerably higher than those measured outdoors. Only after the discovery in the mid-1980s of homes with levels as high as 1,000 pCi/L did the EPA take action. It based its policy on the controversial assumption that any amount of radiation exposure poses some risk and that the risk-exposure ratio is linear. That is, if long-term exposure to 100 pCi/L of radon in a mine increases the risk of lung cancer by 50 percent, then exposure to 10 pCi/L in a home increases the cancer risk by 5 percent, all other factors being equal.

The EPA now estimates that indoor radon causes between 7,000 and 30,000 of the 130,000 deaths from lung cancer a year in the U.S., making it second only to smoking as the leading cause of lung cancer. The agency contends that some 15 percent of these deaths could be avoided by reducing radon levels in the five million homes thought to have levels above 4 pCi/L.



LAURIE GRACE

***AVERAGE INDOOR RADON** levels of U.S. counties are estimated in this EPA map. Although the map's calculations are tentative, a bill before Congress requires EPA-approved radon-reduction measures in all new buildings in high-radon* **(brown)** *zones.*

Congress takes these claims seriously. A bill in the House of Representatives would require contractors in designated high-radon areas, which encompass roughly one third of the nation's counties [*see map above*], to follow new EPA guidelines for reducing radon. Such measures include installing pipes in the foundations of houses to route the gas outdoors. In addition, sellers of homes throughout the U.S. would have to provide buyers with EPA literature on radon and with the results of any previous radon tests. Every contract of sale would also warn buyers: "The U.S. Surgeon General has determined that prolonged exposure to radon can be a serious health hazard."

The EPA's position was bolstered this past January by a paper published in the *New England Journal of Medicine.* A team of Swedish workers compared 1,360 Swedish men and women who had cancer with a group of controls. The workers concluded that "residential exposure to radon is an important cause of lung cancer in the general population. The risks appear consistent with earlier estimates based on data in miners."

But other recent studies, while involving fewer subjects, have failed to corroborate this conclusion. A group led by Ernest G. Létourneau of the Radiation Protection Bureau of Health Canada measured radon levels in the homes of 738 lung cancer victims and an equal number of control subjects in Winnipeg, Manitoba. The average radon exposure of the cancer victims was slightly *less* than the exposure that the controls experienced.

An examination by a group from the University of Kansas School of Medicine of women living in 20 counties in Iowa corroborated previous evidence that radon may hasten the onset of lung cancer in smokers but does not pose a threat to nonsmokers. In *Health Physics,* the Kansas investigators reported a correlation between radon and risk of lung cancer in counties with high smoking rates. Counties with low rates of smoking showed an inverse relation between radon and cancer.

Finally, a study headed by Jay H. Lubin of the National Cancer Institute, published this year in *Cancer Causes and Control,* compared 966 women with lung cancer in Sweden, China and New Jersey with 1,158 controls. The workers found a slight but statistically insignificant correlation between radon and cancer. Asked if the studies done so far justify the EPA's 4 pCi/L threshold, Lubin declines to offer his personal opinion. But he says virtually all researchers would agree that levels above 20 pCi/L represent a genuine threat. That is the maximum amount of exposure to radiation now allowed by U.S. regulations.

Margo T. Oge, director of the EPA's Office of Radiation and Indoor Air, notes that over a dozen more radon studies are under way, and the EPA has asked the National Academy of Sciences to do a meta-analysis of available data. "We

obviously want to put forward an objective point of view," Oge says. Yet she insists that the EPA's 4 pCi/L "action level" is justified. The EPA's estimate of radon's risks, she asserts, stems from research on animals as well as epidemiological studies, and it is supported by the Centers for Disease Control, the surgeon general, the American Medical Association and other groups.

But these agencies fell in behind the EPA for political rather than scientific reasons, asserts Leonard A. Cole, a political scientist at Rutgers University. Cole is the author of *Element of Risk: The Politics of Radon,* a scathing critique of federal radon policy published last year. Cole suggests that the Reagan administration seized on the radon issue in the mid-1980s to counter its anti-environment image. The issue suited Republicans, he contends, because homeowners—rather than government or business—would bear the costs of fighting the threat. "Republican conservatives ran with this, and since then it's been picked up by Democrats," Cole says.

One of the most prominent critics of the EPA's handling of the radon issue is Anthony V. Nero, Jr., a pollution expert at Lawrence Berkeley Laboratory. Nero thinks all the data justify a policy that focuses on homes with levels of 20 pCi/L or above. By adopting such a stance, he argues, the EPA would reduce the number of homes targeted for remediation from over five million to perhaps 50,000 and thereby make it more likely that the job would be carried out.

Nero accuses the EPA of making "highly misleading" statements about the dangers of radon. A pamphlet equating radon levels of 4 pCi/L to smoking half a pack of cigarettes a day is "just wrong," Nero says, adding that the statistic applies only to those who already smoke one and a half packs a day. Although EPA officials state that they no longer distribute the pamphlet, Nero contends that such exaggerations continue to circulate in public and on the floor of Congress.

Nero also faults EPA officials such as Oge for comparing the EPA's recommended radon limit with its limit on radiation releases by nuclear power plants, which is some 80 times lower. It is completely appropriate, he points out, to set much stricter limits on industrial emissions than on a naturally occurring gas. Nero fears that by overstating its case, the EPA may trigger a backlash of skepticism and cause people to think, mistakenly, that *no* levels of radon pose a risk. The agency is "running backward very fast," he says, "instead of moving forward on the more pressing problem of very high levels." —*John Horgan*

Diversity Blues

Oceanic biodiversity wanes as scientists ponder solutions

The evidence is everywhere. Populations of fish and shellfish, of corals and mollusks, of lowly ocean worms, are plummeting. Toxic tides, coastal development and pollutant runoff are increasing in frequency and dimension as the human population expands. The oceans—near shore and in the abyssal deep—may be reaching a state of ecological crisis, but, for the public, what is out of sight is out of mind. "The oceans are in a lot more trouble than is commonly appreciated," rues Jane Lubchenco of Oregon State University. "There is great urgency."

To remedy this situation, marine scientists recently gathered in Irvine, Calif., to devise a national research strategy to protect and explore marine biodiversity. Although the variety of organisms found in the oceans is thought to rival or exceed that of terrestrial ecosystems, there is no large-scale conservation effort designed to protect these creatures. Indeed, there is no large-scale effort even to understand the diversity found in saltwater regions.

The National Research Council meeting attendees first set about establishing their ignorance: the system they study remains, in large part, a mystery. Several years ago, for instance, J. Frederick Grassle of Rutgers University reported that previous estimates of the number of organisms thriving on the deep-sea floor were probably too low. In analyzing sediment from an area off the coasts of Delaware and New Jersey, Grassle found 707 species of polychaetes, or worms, and 426 species of crustaceans. All these creatures were harvested in samples taken from boxes that measured only 30 centimeters per side and 10 centimeters in depth. Earlier studies had suggested a total of a mere 273 species of polychaetes.

As researchers at the Irvine meeting emphasized repeatedly, even the diversity of areas that have been exhaustively studied is not fully appreciated. New findings about star coral, or *Montastraea annularis,* offer a dramatic example. This organism "is sort of a lab rat of corals," explains Nancy Knowlton of the Smithsonian Tropical Research Institute in Panama. "It is an extremely intensively studied coral." Knowlton and her colleagues have discovered that this single species of coral is, in fact, three species in shallow waters. (There may be even more species in the star corals that inhabit deeper water.) These vari-

MICHAEL OSMOND *The Wildlife Collection*

***HUMPBACK WHALES** may provide crucial nutrients to ocean-bottom dwellers by sinking to the seafloor after they die. By severely limiting this food supply, the extensive hunting of whales may have already irreversibly altered the marine ecosystem.*

ous species have also been found to be adapted to different depths.

Knowing that diversity is out there, however, has not yet allowed marine researchers to make a stab at species numbers—something their peers on land have been able to do to galvanize public action. "We are not close to making an estimate," Knowlton acknowledges. "Even a seat-of-the-pants guess might be off by an order of magnitude."

Identifying threats to the oceans was less tricky. Although the usual suspects were in the lineup—including oil spills, the destruction of estuaries, toxic dumping and the introduction of nonindigenous species that outcompete the locals—conference attendees deemed fishing the greatest danger to marine biodiversity. "I was pretty surprised. The impacts of fishing have been at the top of my list for years," says Les Watling of the Darling Marine Center at the University of Maine. "But I thought there was not such a big awareness of that. The biggest problems are usually seen as pollutants or eutrophication." (Eutrophication is caused by excess nutrients from such chemicals as fertilizers and can lead to algal blooms.)

Nevertheless, reports about the global decline of fisheries keep coming in. As Carl Safina of the National Audubon Society outlined in a recent article in *Issues in Science and Technology,* catches of groupers and snappers fell by 80 percent during the 1980s, and the population of swordfish in the Atlantic Ocean has fallen by 50 percent since the 1970s.

In addition to the depletion of fish—which may have far-reaching but little understood ecological effects—fishing often wipes out habitat. By trawling on the seafloor, vessels disrupt bottom communities or coral reefs. Watling cites the destruction of sponges in the Gulf of Maine as one example. Last seen in 1987 on a videotape taken from a submarine, "the sponges are gone. They have been ground off the rocks," Watling states. These sponges may be important nursery habitats for species such as cod—of course, that possibility reveals another marine unknown. "The real problem is that we do not know anything about the first year of life in cod," Watling warns.

A crisis in taxonomy also worried the scientists. Every researcher had a complaint about years going by before he or she could get someone to identify an alga, about seminal papers misidentifying creatures, about graduate students receiving no training in taxonomy. Without good taxonomy, trying to identify and protect diversity becomes moot.

Beyond the challenge of identifying species correctly lies the challenge of understanding their interactions. If marine biology is going to help policymakers, it has to be at least somewhat predictive. Even if the effects of climatic change on a certain species are understood, for example, the implications for the entire ecosystem may be obscure. Unpublished studies by Lubchenco about increases in water temperature caused by a power plant in Diablo Cove, Calif., illustrate just this problem. "You could not have predicted the changes that occurred based on a knowledge of the individual species' sensitivity to water temperature," Lubchenco explains. "What is going on is greater than the individual response."

Getting the scientific community to voice concern about the threat to oceanic ecology was the first step, according to conference chairs Cheryl Ann Butman of the Woods Hole Oceanographic Institution and James T. Carlton of Williams College and Mystic Seaport. Designing a research program that will address the issue and receive funding from Congress is the next task at hand.

The most difficult hurdle may be catalyzing public awareness before the marine environment is altered beyond the point of no return. As Butman and Carlton describe, hunting whales may already have altered the oceans irrevocably. Because deep-sea organisms rely on food falling from the surface, large carcasses of whales may have been one of the major sources of nutrients for the bottom of the food chain. The sulfur-rich bones of whales may have provided stepping-stones for sulfur bacteria and other organisms as they moved from hydrothermal vent to vent. Fewer sinking cetaceans may have had important impacts on deep-sea processes.

"Unfortunately, the question is virtually impossible to answer now," Butman comments. "But it certainly would be irresponsible of us to put ourselves in a position like this again—that is, a position where we embark on a dramatic alteration of species diversity, which is what the whaling industry represents—without evaluating the ecological consequences." —*Marguerite Holloway*

The Riddle of [URE3]

The humble yeast cell hints at novel forms of heredity

A venerable biological mystery has taken a new twist. For several decades, researchers and clinicians have been intrigued by a family of fatal central nervous system disorders of humans and other mammals in which the brain degenerates. The diseases—among them kuru, Creutzfeld-Jakob disease and bovine spongiform encephalopathy ("mad cow disease")—are notable for the fact that they are not caused by ordinary infectious agents such as bacteria or viruses, whose genetic material consists of DNA or RNA. Work by Stanley B. Prusiner of the University of California at San Francisco and others strongly suggests that the agent, which is called a prion, consists of an aberrant form of a normal protein and includes no genetic material. When transmitted from another animal or produced spontaneously because of a prior mutation, it triggers the normal form to switch to the prion structure, thus initiating a runaway process that kills affected cells.

Prions have generally been considered a bizarre and isolated curiosity. Now Reed B. Wickner, a researcher at the National Institute for Diabetes, Kidney Disease and Digestive Disorders, has found evidence that prions have an analogue in yeast. Wickner's research, published in April in *Science,* focuses on a metabolic peculiarity that some mutations confer on yeast cells. The anomaly is the ability to feed on a chemical called ureidosuccinate. The mutations conferring this trait can be individually distinguished by the way they are passed on to offspring in experiments that cross cells of different types.

Most mutations that confer the ability to use ureidosuccinate have patterns of inheritance typical of mutations in genes on chromosomes. But one—[URE3]—is passed between individuals in ways that cannot be explained by what is known about how genes work. [URE3] is passed on to more offspring than a normal mutation should be when cells are crossed. It can be transmitted when cells exchange cytoplasm but not chromosomes. And a simple chemical treatment can reversibly "cure" [URE3], thus eliminating the cells' ability to use ureidosuccinate.

Somewhat similar strange patterns of inheritance can arise when mutations occur in DNA or RNA that replicates separately from the chromosomes. Yet this explanation does not apply to

[URE3], according to Wickner. The crucial clue to [URE3]'s nature, he says, is an observation that was originally made more than 20 years ago by François Lacroute of the Center of Molecular Genetics at Gif-sur-Yvette, France. Wickner has confirmed and extended that work. [URE3] can exist in a cell only if a protein called Ure2p, the product of a known gene, is present. If a cross is produced that lacks Ure2p, the [URE3] trait cannot appear in that cell. And a cell that lacks Ure2p has the ability to metabolize ureidosuccinate.

Wickner's explanation of this peculiar set of facts is that [URE3] is not really a mutation at all but rather the manifestation of cells that contain a variant form of the Ure2p protein. Normal Ure2p prevents uptake of ureidosuccinate, which is why cells lacking Ure2p can utilize the chemical. Wickner proposes that the variant form of Ure2p—which appears to the experimenter as the [URE3] trait—also fails to prevent metabolism of ureidosuccinate, which is why cells carrying the [URE3] trait can digest the chemical. In cells that initially contain some normal Ure2p and some of the abnormal form, the abnormal variant quickly converts all the cell's Ure2p into copies of itself, just as prion protein can convert its normal counterpart into more prion protein.

Prusiner notes that Wickner "has not done any experiments that prove it's a protein" that transmits the [URE3] trait. Even so, Prusiner is interested enough to have started studying the biochemistry of [URE3]. Wickner, for his part, is pressing ahead with attempts to prove that the [URE3] trait is indeed transmitted by a protein. Already he suspects that a second genetic system in yeast, [PSI], may follow the same pattern.

This latest turn in the prion story is unlikely to dethrone DNA and RNA as life's principal bearers of genetic information, Wickner acknowledges. Still, the apparent occurrence of protein-based inheritance in yeast raises the question of whether such mechanisms play a bigger role in life and death than has generally been believed.—*Tim Beardsley*

Anti-omniscience

An eclectic gang of thinkers pushes at knowledge's limits

The Danish physicist and poet Piet Hein once wrote: "Knowing what/ thou knowest not/is, in a sense,/ omniscience." The hope that science might achieve a kind of anti-omniscience drew together 20 thinkers, including mathematicians, physicists, biologists and economists, for a workshop on "the limits of scientific knowledge" held at the Santa Fe Institute.

"Can we prove there are limits to science?" asks Joseph F. Traub, a computer scientist at Columbia University, who is one of the meeting's organizers. Mathematics has had some success in delineating its own boundaries, Traub remarks. The most dramatic example was Kurt Gödel's demonstration in the 1930s that all moderately complex mathematical systems are "incomplete"; that is, they give rise to statements that can be neither proved nor disproved with the axioms of the system.

Gregory J. Chaitin, a mathematician at the IBM Thomas J. Watson Research Center, sees darker implications in Gödel's theorem. He notes that this insight has been followed by similar ones, notably Chaitin's own finding that mathematics is riddled with truths that have no logical, causal basis but are simply "random." As a result of these difficulties, he says, mathematics may become an increasingly empirical, experimental endeavor with less of a claim to absolute truth.

Other mathematicians find Chaitin's pessimism excessive. The hurdles identified by Gödel and others, declares Francisco A. Doria of the Federal University at Rio de Janeiro, can enrich mathematics. Doria suggests, for example, that at each point where an unprovable, or "undecidable," proposition obstructs them, mathematicians might simply make an arbitrary presumption about its truth or falsity to see whether fruitful results follow.

In fact, Gödel himself did not think his theorem posed any special barrier to knowledge, comments John Casti, a mathematician at the Santa Fe Institute and the workshop's other organizer. Casti believes mathematicians might avoid the Gödel problem by employing systems so simple that they do not suffer from incompleteness. He also expresses the hope that Gödel's theorem might turn out to be "a red herring" when it comes to natural science.

Others demur. Robert Rosen, a biophysicist at Dalhousie University in Halifax, contends that the "preternatural difficulty" biologists have had in arriving at a precise definition of life is related to the incompleteness concept. The incompleteness results are not just "intellectual curiosities," he insists. "I think it is in biology that you will see the true impact of these ideas."

Then there is the trap of the infinite regress. W. Brian Arthur, an economist who divides his time between Stanford University and Santa Fe, notes that in

trying to predict how the stock market will perform, an investor must guess how other investors will guess about how still others will invest—and so on. Economics is an intrinsically subjective—and hence unpredictable—enterprise, Arthur concludes.

Just because some aspects of a system are unpredictable, however, does not mean all aspects are, points out Lee Segel, a mathematician at the Weizmann Institute of Science in Israel. Although scientists cannot track the path of a single particle of air passing over a wing, they can calculate the pressure that the flow of air exerts on the wing, which amounts to much more useful information. "Before saying a problem will defeat us, we have to consider other approaches," Segel says.

Piet Hut, an astrophysicist at the Institute for Advanced Study in Princeton, N.J., offers a success story of this kind. He notes that one of the most difficult problems in astronomy, the *N*-body problem, involves predicting how three or more objects moving in one another's gravitational fields will behave over time. Hut and other investigators have sidestepped the issue by developing potent statistical methods for calculating the effect of the gravitational interactions of billions of stars within galaxies.

The history of computation also suggests that many perceived limits may be illusory, according to Rolf Landauer, a physicist at the IBM Watson center. For example, constraints once thought to be imposed on computation by the second law of thermodynamics or quantum mechanics have been shown to be spurious. The most immediate barriers to further advances in computation may be financial, Landauer says.

Even if our computers and mathematical tools continue to improve, cautions Roger N. Shepard, a psychologist at Stanford, we may not understand the world any better. If neuroscientists construct a computer powerful enough to simulate a human mind, they may simply substitute one mystery for another. "We may be headed toward a situation where knowledge is too complicated to understand," Shepard says.

The structure of the physical universe may represent the ultimate limit on human knowledge, according to Hut. Particle physicists may never be able to test theories that unify gravity and the other forces of nature because the predicted effects become apparent only at energies beyond the range of any conceivable experiment. Moreover, cosmologists can never know what, if anything, preceded the universe's birth.

One participant pleased by all this

rumination is Ralph E. Gomory, the former director of research for IBM who is now president of the Alfred P. Sloan Foundation in New York City, which sponsored the Santa Fe meeting. Gomory says he has long felt that the educational system places too little emphasis on what is unknown or even unknowable. To remedy the situation, the Sloan Foundation may initiate a program on the limits of knowledge.

Gomory also has a suggestion for mitigating science's task: make the world more artificial. Artificial systems, Gomory states, tend to be more predictable than natural ones. For example, to simplify weather forecasting, engineers might encase the earth in a transparent dome. Everyone sitting around the table stares at Gomory, whose expression remains deadpan. Then Traub remarks, "I think what Ralph is saying is that it's easier to create the future than to predict it." *—John Horgan*

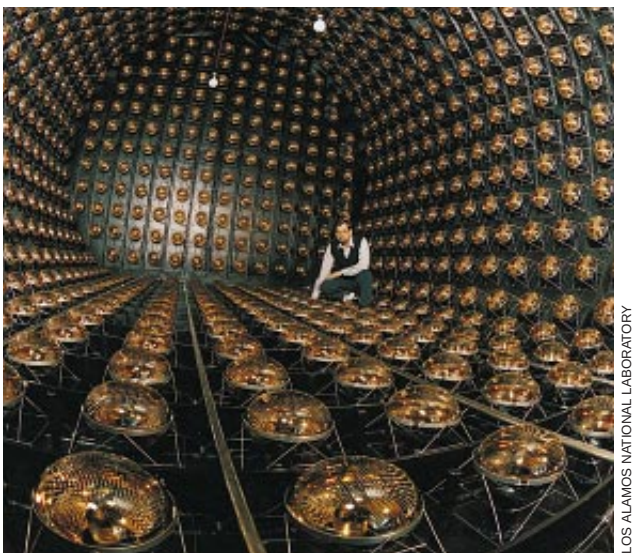
Missing Matter Found?

Ever since physicists discovered the massless neutrino—the "little neutral one"—they have wondered if this elusive particle might not in reality have some slight mass. Because neutrinos exist in great numbers in the universe, even a small mass could provide the "dark matter" that cosmologists believe makes up most of the substance of the cosmos. Having a mass, neutrinos might also be able to change into neutrinos of other types, by a process called an oscillation.

A particle detector at Los Alamos National Laboratory has captured eight events that could be the first direct sightings of neutrino oscillations. If verified, the observations will prove as well that neutrinos have mass. "It's too good to be true," said Baha A. Balantekin of the University of Wisconsin on viewing the data at a June conference.

Apparently the researchers agree. The Liquid Scintillator Neutrino Detector (LSND) experiment has the world's highest sensitivity to neutrino changes. But those involved are not making explicit claims. "We feel we have a high burden of proof," explains group leader D. Hywel White, "because if it's real, it's very important." Moreover, earlier reports of neutrino oscillations have themselves oscillated away.

Neutrinos come in three types: the electron neutrino, the muon neutrino and the tau neutrino. Why would the ability of neutrinos (or antineutrinos) to change from one kind to another indicate that the particles have mass? Mass determines the rate at which the wave function of a particle vibrates. If the waves of two neutrinos of different masses mingle, they beat against each other much like sound waves of different pitch. Then we sometimes see one neutrino, sometimes the other. If neutrinos had no mass, their waves would have the same frequency and would not be able to beat at all.



LOS ALAMOS NATIONAL LABORATORY

NEUTRINOS may have been caught in the act of changing by the phototubes lining this detector, shown without fluid.

Such fleeting events as neutrino oscillations are not easy to observe. In the LSND experiment, a beam of protons from an accelerator is shot into a water target. Many of the particles produced in the subsequent collisions are absorbed in the shielding around this "beam dump." But an occasional muon antineutrino escapes, flying for 30 meters to a detector filled with baby oil. On the way the particle might change into an electron antineutrino.

The electron antineutrino interacts with a proton in the oil, giving off a positron and a neutron. After some time, the neutron binds with another proton, yielding two photons with a characteristic total energy. The positron's bright track and the photons are observed by phototubes lining the oil tank. To avoid contamination from other particles that might have sneaked by the shielding, the experimenters look for positrons within a specific range of energy.

As of June, the experiment had run for a month and a half. The small number of electron antineutrinos observed suggests that muon antineutrinos convert only slightly to the former; their mixing strength is about 1 percent. The experimenters do not quote a mass difference. But the 30 meters over which the muon antineutrino can change its type means that the apparatus is sensitive to mass differences of a little less than an electron volt. This mass difference implies a large neutrino mass.

Although the result could help solve the dark matter problem, it contributes nothing to the solution of another major puzzle that physicists are currently enjoying: the solar neutrino problem. The number of electron neutrinos coming from the sun is less than half the number that theory predicts. The deficit might be explained by presuming that the particles change to muon neutrinos and therefore escape detection. But if neutrinos change type over a distance of 30 meters, as in this experiment, the oscillations would average out over the 92 million miles that separates the sun from the earth. No deficit would be observed. The new findings may, however, illuminate the lack of muon neutrinos in cosmic rays reaching the earth's surface, a mystery dubbed the "atmospheric" neutrino problem.

Early in August the LSND experiment will start running again and will take data for three more months. Despite the precarious state of finances at Los Alamos National Laboratory, White is hopeful that the experiment can continue for at least another year. Already, one observer remarked, the neutrino oscillations "look at least as real as the top quark," evidence for which was announced in March at Fermi National Accelerator Laboratory in Batavia, Ill. Both groups are currently walking a thin line between presenting suggestive data and making a claim. The team at Los Alamos should be able to verify—or otherwise—their nonclaim much sooner. *—Madhusree Mukerjee*

Darwin's Current Bulldog

In 1859 Darwin published his theory of common descent through natural selection. I don't think there has ever been a set of theories so heavily attacked or that has had so many alternative theories to face," exults Ernst Mayr. "Look at it now. It stands there, not a dent in it."

The Alexander Agassiz Professor of Zoology, Emeritus, of Harvard University might just as well be speaking about himself. Mayr is the unscathed survivor of a lifetime's battles over evolution. If Charles Darwin's work is intact, no small part of the credit belongs to Mayr, who has probably done as much as anyone to advance evolutionary theory and to entrench it at the core of all biological thought.

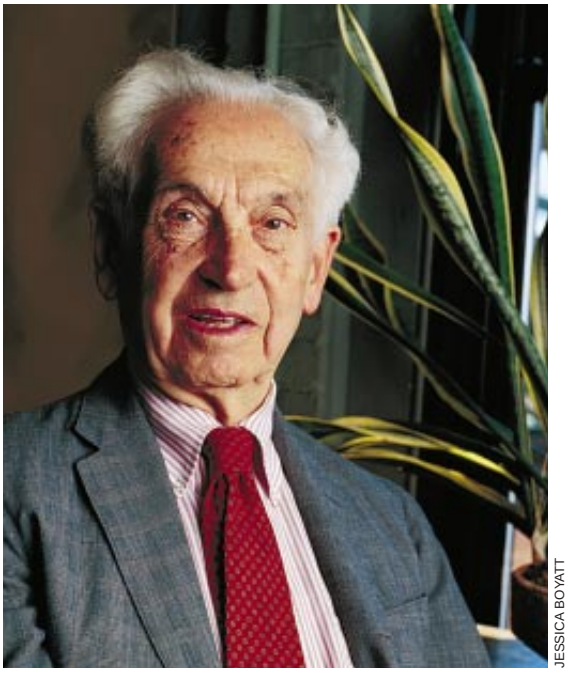
"Certainly, Ernst Mayr is a leading biologist of the mid- to late 20th century," remarks science historian John C. Greene of the University of Connecticut, who organized a symposium celebrating Mayr last year. "He's one of the founders of modern neo-Darwinism and has restored natural selection to a central place in the theory of evolution."

In his passion for evolution, Mayr is reminiscent of the 19th-century scientist Thomas H. Huxley, remembered as "Darwin's bulldog" for his championing of the naturalist's views. Is Mayr Darwin's new bulldog? "Yes, very much more so than Huxley, in a way," Mayr says without hesitation. "Huxley did not believe in natural selection."

At the age of 90, Mayr remains a dynamo. Every morning he visits his cool green office under the eaves of Harvard's Museum of Comparative Zoology, a neatly dressed, gray-haired figure walking without benefit of cane. His vitality strips at least a decade off his appearance. "I admired myself a couple of days ago," he chuckles. "I am in between with getting household help, and I noticed that the kitchen floor was very dirty. So I got a bucket, and I washed it."

He is the author of 20 books (so far)—nine of which appeared after his 65th birthday. Among them are *The Growth of Biological Thought,* a monumental overview of his field's development, and *One Long Argument,* a popular account of Darwin's ideas. The number of Mayr's published papers stands somewhere around 650 (and counting). In recent years, he has also commented on overpopulation, the loss of biodiversity and the search for extraterrestrial intelligence.

Mayr has been a leader in ornithology, systematics, evolutionary biology and both the history and philosophy of biology. "And I have received world-class distinctions in all of them," he adds. They include the National Medal of Science, the Balzan Prize (the highest honor in evolutionary biology) and the Sarton Medal (the highest honor in the history of science). At the International Ornithological Congress in 1986, he was declared the preeminent bird researcher of our time. He holds 11 honorary degrees and membership in 45 academic societies around the world.



JESSICA BOYATT

ERNST MAYR is evolution's leading defender.

About the only great prize he hasn't won is the Nobel, which recognizes practical research but not Mayr's specialty, conceptual advances. "I believe that in biological science, the concepts are the crucial thing," he observes. The lack of a Nobel does not seem to bother him. "I have always said that if there were a Nobel Prize for Biology—which there isn't, because Nobel was an engineer and too ignorant about biology!—if there were such an award, Darwin could never have received it for evolution through natural selection, because that was a concept, not a discovery!"

Part of Mayr's personal charm is that he can somehow make such blunt statements without sounding arrogant or malicious. Even his scientific adversaries, whom he attacks without quarter, seem to forgive him. "Oh, I'm fierce, because I don't give in, you know?" Mayr says. "But with all my opponents—there are maybe one or two exceptions—I am on very good terms."

Mayr was born in Germany in 1904. Every weekend he and his parents hiked, watched birds, looked at the season's flowers or collected fossils in nearby quarries. "All my high school days, as soon as I was done with my homework, I would be out with my bicycle in a park or someplace, bird-watching. That was the foundation for my whole career."

In 1923 Mayr spotted a rare species of duck that had not been seen in central Europe for 75 years. The publication of that discovery allowed him to meet Germany's leading ornithologist, who encouraged Mayr's interests. Although Mayr had been training as a medical student, he eventually chose to be a naturalist instead. In 1926 he graduated from the University of Berlin with a doctorate in zoology and became an assistant curator for the university's museum.

His career change had less to do with purely scientific zeal than with a boyish love of adventure. "I was told, 'If you become a naturalist, you can go on expeditions,' and that's really what I wanted," Mayr admits. Between 1928 and 1930 Mayr worked on ornithological expeditions in New Guinea and the Solomon Islands. "In those days, New Guinea was very wild country," he recalls. "You traveled into the interior for one day's walking, and you came to villages where no white man had ever been before."

While in New Guinea, Mayr noticed

that the natives recognized the same species distinctions in the local birds that Western naturalists did. That fact convinced Mayr that species are real biological units and not arbitrary taxonomic inventions. In addition, like Darwin in the Galápagos Islands, Mayr found several populations of birds that appeared to have become species because they were geographically isolated.

Because of connections with U.S. scientists that he had made in the Solomon Islands, Mayr landed a job in New York City in 1932 as a curator of the bird collection at the American Museum of Natural History. He began publishing voluminously and formulating ideas about evolution.

Strange though it now sounds, the young Mayr was not a follower of Darwin. Rather, like most other naturalists of the day, he subscribed to Jean-Baptiste Lamarck's theory about the inheritance of acquired characteristics. "And that was very logical," Mayr explains. Mendelian genetics did not seem to support the possibility of gradual adaptive changes, which Darwinism required. Geneticists favored the idea that species evolved suddenly through massive mutations. "But we naturalists realized that species developed gradually. The only evolutionary theory that was gradual was Lamarckism, and so to oppose the mutationists, we all became Lamarckians." During the 1930s, however, geneticists started recognizing the evidence for small mutations, and Mayr's reservations about the occurrence of natural selection melted away.

Mayr's work made him one of the architects of the modern evolutionary synthesis, a select group that included such giants as Theodosius Dobzhansky, George Gaylord Simpson, G. L. Stebbins, and Bernhard Rensch. During the 1930s and 1940s, these biologists wedded Darwin's ideas about natural selection, population genetics and the field studies of naturalists into a cohesive explanation for evolution.

Mayr's first major contribution to the synthesis pulled together adaptation and speciation. Previously, according to Mayr, geneticists had concerned themselves with how natural selection might adapt organisms to their environment; naturalists had separately pondered how and why species arose. No one had united the problems. "I was going to fill that niche," Mayr says. "Dobzhansky had already done it to some extent in 1937 but only in a very tentative, preliminary way. My 1942 book really filled it pretty completely."

In that book, *Systematics and the Origin of Species,* Mayr introduced the biological species concept, which defined a species as a set of interbreeding populations that is reproductively isolated by behavior and physiology from other groups. He also argued that new species could evolve only through allopatry—that is, through the geographical separation of a population from the rest of its kind. In 1954 he distinguished a second form of geographical speciation, peripatry, in which the founding population of the isolated group is very small: this circumstance has genetic ramifications that can sometimes accelerate species divergence.

On both species and speciation, Mayr maintains, he has been shown to be correct many times over. Nevertheless, critics have been legion, and Mayr has spent much of the past five decades rebutting them. All these opponents, he believes, misunderstand and distort what he, the other synthesists and Darwin said, then try to knock down those straw men. "All the attacks that I find are based on ignorance," he declares.

"Hybrids are mistakes," the species expert says. "To me, this all seems so obvious and so simple."

For example, some naysayers claim that fertile hybrid organisms, such as wolf-coyote crosses, are living contradictions of the biological species concept. "This argument is so misleading that it pains me to have to refute it again and again and again," Mayr sighs. The designation "species," he expounds, belongs to an entire population, not to individuals. The group survives even if individuals deviate from the group's normal behavior. "Hybrids are mistakes," he insists. "But the isolating mechanisms are good enough to prevent the merging of the species." Mayr shakes his head in dismay. "To me, this all seems so obvious and so simple."

Mayr has also strenuously argued against what he sees as the unrealistic simplifications of geneticists. Deriding their efforts as "beanbag genetics," he accuses them of trying to reduce evolution to a mere change in gene frequencies, without considering how populations of organisms live. The eminent geneticists R. A. Fisher and J.B.S. Haldane, for example, had concluded that large populations should evolve faster than small ones because they had a larger reservoir of genetic diversity. Yet Mayr observed just the opposite. "The larger a species is, the slower it evolves. It becomes evolutionarily inert," Mayr states. "My conclusions were not based on mathematics but on the simple matter of observation. And of course, the observation won out, and the mathematics turned out to be all wrong!"

For a time, the controversial theory of punctuated equilibrium was being trumpeted as a challenge to Darwinism. During the 1970s, Niles Eldredge of Columbia University and Stephen Jay Gould of Harvard pointed out that in the fossil record, many species remain unchanged for millions of years and then undergo rapid evolution in the virtual blink of an eye. But Mayr dismisses the idea that this observation refutes the synthesis as "total rot," "a lead balloon" and "a red herring." Not only are long periods of evolutionary stasis compatible with natural selection, Mayr says, but "the whole theory was already indicated in my 1954 paper."

Mayr believes the existing framework of the evolutionary synthesis is essentially unshakable, although he does not consider it complete. "The synthesis up to now has been rather coarse. There is room for more fine-grained analysis," he reflects. The greatest advance, he believes, will come from integrating the synthesis with the information emerging about genes and their interactions, particularly during development.

His unflagging productivity makes it clear that Mayr fully intends to stay in the thick of further work on evolution—and as many other subjects as he can. "Right now I have probably about 15 papers in press," he estimates. One of them tackles the theory of philosopher Thomas S. Kuhn about the nature of scientific revolutions. "My paper shows that Kuhn's idea absolutely doesn't fit any scientific revolution in biology," Mayr crows. A second article denounces the search for extraterrestrial intelligence, which Mayr regards as a colossally foolish waste of money. "This is another case of physicists talking about things they know nothing about," he says with amusement.

Those papers are only a warm-up for Mayr's bigger undertakings, however. He is currently writing another book with the modest tentative title of *This Is Biology: The Science of the Living World.* "It's a sort of life history of the science of biology," Mayr explains, then adds unnecessarily, "It's a very ambitious project." And when he is done with that, he continues, "I plan to write a very simple book on evolution for the layperson."

Is that all? "Oh, I have several more projects in mind for after that," he laughs. "I consider every publication a stepping-stone to the next. Activity is what keeps me going." —*John Rennie*

SCIENTIFIC AMERICAN

Third World Submarines

The proliferation of submarines may be a threat to established navies and regional stability, but to arms manufacturers it is a market opportunity

by Daniel J. Revelle and Lora Lumpe

During the spring of 1993, Iran put the first of its new Russian-built Kilo-class submarines through sea trials in the Persian Gulf. Its presence raises the specter of an Iranian attempt to close the Strait of Hormuz, the narrow waterway through which a fourth of the world's oil now passes.

Throughout the cold war, the U.S. Navy's highest priority mission was to engage Soviet nuclear-powered submarines in a global game of hide-and-seek. As that threat has faded, conflicting priorities have emerged. On one hand, the U.S. Navy is concerned about the threat that growing Third World naval forces pose to its ability to operate in coastal waters around the world. On the other hand, concern about the fate of the cold war industrial base is creating pressures for the U.S. to join former allies and enemies in supplying advanced diesel-powered attack submarines to developing countries.

More than 20 developing countries currently operate over 150 diesel attack submarines. North Korea has 25 such vessels, India 18, Turkey 15, Greece 10, Egypt 8, Libya 6 and Pakistan 6. Many of these boats are obsolescent, poorly maintained or operated by ill-trained crews. Others, however, could be a match for many vessels in the navies of the industrial world.

DANIEL J. REVELLE and LORA LUMPE worked together in the Arms Sales Monitoring Project at the Federation of American Scientists (FAS) in Washington, D.C. Revelle received a degree in physics from Carleton College in Northfield, Minn., and is currently a graduate student in aerospace engineering at the University of Colorado at Boulder. Lumpe directs the FAS's Arms Sales Monitoring Project and edits a bimonthly newsletter on weapons exports.

Third World nations have purchased their most advanced vessels from Russia and western European countries, both of which have a submarine manufacturing base far in excess of their own needs. Hans Saeger, sales director for the German submarine builder HDW, has estimated that NATO countries have the capacity to build 19 vessels a year, although NATO members generally purchase only two or three. The incentive to employ the remaining capacity is strong.

Germany in particular is a major exporter of submarines. Its sales are of exceptional concern because they frequently involve the transfer not only of vessels but also of production equipment and know-how for building submarines. Such "coproduction" deals promote sales, but they also lead to an increase in the number of nations competing to sell submarines, thus making proliferation even more difficult to contain. Germany has made coproduction agreements with South Korea, India and Argentina—the last has been licensed to produce two additional submarines for reexport.

Russia looks to weapon sales as a source of desperately needed hard currency. The Russian navy stated several years ago that it intended to continue producing two diesel submarines a year, keeping one for itself and selling the other for ready cash. Soviet customers have included Libya, North Korea, India and Algeria. More recently Iran purchased two of the Kilo boats with the option to buy a third.

Other nations are in the business, too. France has supplied its Daphne and more modern Agosta models to Pakistan. China has sold somewhat outdated Romeo-class submarines to North Korea and Egypt. Sweden is marketing submarines to Malaysia and is looking for other sales in South Asia. The Netherlands is considering the sale of 10 submarines to Taiwan in what is expected to be the last big sale of the century. Britain, meanwhile, is selling off four new Upholder-class diesel boats that its fleet no longer has the money to support, even offering to lease them complete with mercenary crews.

Although the U.S. Navy has purchased only nuclear-powered attack submarines since the 1960s, the U.S. government recently gave approval for domestic production of diesel vessels. In a 1992 report to Congress, the navy argued: "Construction of diesel submarines for export in U.S. shipyards would not support the U.S. submarine shipbuilding base and could encourage future development and operation of diesel submarines to the detriment of our own forces." Nevertheless, in April 1994 the State Department gave Ingalls shipyard in Pascagoula, Miss., the go-ahead to produce HDW's Type 209 under a license from the German firm. Egypt wants to buy two of these boats but cannot afford to purchase them directly from Germany. The vessels built by Ingalls will be bought using U.S. military aid, which may be spent only on weapons of American manufacture.

UPHOLDER-CLASS SUBMARINE (*shown here*) is one of four that the British Royal Navy built during the 1980s but can no longer afford to maintain. Britain is now offering to sell the diesel vessels or to lease them out, complete with crews.

Once this new production line is in place, economic considerations will probably generate pressure to make further sales to developing countries. Taiwan and Saudi Arabia are the next likely customers for U.S.-made Type 209 vessels.

As shrinking military budgets add to economic woes, arms manufacturers are aggressively seeking to expand their markets. Submarine merchants have targeted nations bordering on the Gulf of Oman, the Mediterranean, the Arabian Sea and northern Indian Ocean, the South China Sea, and Pacific waters near the north Asian coast. If successful, their sales campaign could pose serious risks to international stability.

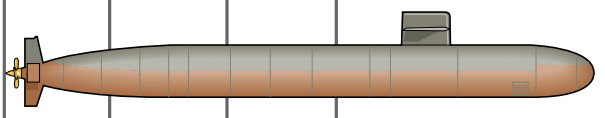
Even a handful of modern, well-maintained diesel submarines could have made a significant difference in the Persian Gulf War. If Saddam Hussein had bought six modern vessels "and positioned three of them on either side of the Strait of Hormuz, that would have complicated matters," according to U.S. vice admiral James Williams. "One diesel sub can make a great difference to how you drive your ships," he asserts.

During the Falklands/Malvinas war, a single Argentine Type 209 managed to elude 15 British frigates and destroyers and the antisubmarine aircraft of two carriers. The *San Luis* maneuvered into torpedo range of the British fleet and launched three torpedoes, although all three shots were unsuccessful. Early in the conflict a British submarine sank the Argentine cruiser *General Belgrano* with two straight-running torpedoes of a design that dated to World War II.

Both the U.S. and British navies are developing active antitorpedo weapons

Attack Submarines for Sale

Diesel-powered attack submarines now being sold to developing nations are smaller and slower than are the superpowers' nuclear versions (such as the U.S. Los Angeles–class vessel pictured immediately below). Nevertheless, they pose a significant threat to shipping and to naval forces that might wish to intervene in regional conflicts.



	LENGTH (METERS)	MAXIMUM SPEED (KNOTS)	DIVING DEPTH (METERS)	ARMAMENT
LOS ANGELES *U.S.*	110	30	450	4 Torpedo tubes 16 Missile tubes 18 Torpedoes 4 Subroc missiles 12 Submarine-launched cruise missiles 6 Harpoon antiship missiles
AGOSTA *FRANCE*	68	20.5	300	4 Torpedo tubes 20 Torpedoes or Exocet missiles
KILO *RUSSIA*	73	25	N/A	6 Torpedo tubes 12 Torpedoes or 24 mines
TYPE 209 (SSK-1500) *GERMANY*	64	22.5	N/A	8 Torpedo tubes 14 Torpedoes Strap-on mine-laying pods
UPHOLDER *U.K.*	70	20	> 250	6 Torpedo tubes 18 Torpedoes or Harpoon antiship missiles
VÄSTERGÖTLAND *SWEDEN*	49	20	> 300	10 Torpedo tubes 18 Torpedoes
ZEELEEUW *NETHERLANDS*	68	21	300	4 Torpedo tubes 20 Torpedoes or Harpoon antiship missiles

for the turn of the century, but at present evasion and electronic countermeasures are the only way to avoid a torpedo already in the water. Courtesy of the industrial nations, most Third World navies now have advanced torpedoes that can home in on a ship and explode just underneath its keel for maximum damage.

Some also possess submarine-launched antiship missiles. The U.S. has sold the Harpoon missile to Israel, Pakistan and others, and the French are marketing a submarine-launched version of the Exocet missile.

The deadliness of submarine-launched weaponry makes early detection and destruction of attacking submarines a crucial factor in antisubmarine warfare (referred to as ASW). Submarines in general are obviously much more difficult to detect than are surface ships or aircraft. Diesel attack submarines can be very quiet. When moving slowly, they can rely for days on battery power, eliminating engine noise or any need to surface or snorkel for air.

Diesel submarines have a relatively short range, and so they tend to inhabit littoral waters rather than the mid-ocean depths. Indeed, most developing countries have only a few vessels deployed defensively near their own coastlines, leading some analysts to deride them as mere "intelligent minefields." Nevertheless, the task of tracking and destroying these submarines can be complex and fraught with pitfalls.

The "shallow" areas that usually harbor diesel submarines may be as deep as 300 meters, giving a vessel plenty of space to hide. At the same time, the bottom is close enough that false sonar echoes can mask a boat's location, much as "ground clutter" can hide low-flying aircraft from radar. Ships, oil rigs and sea life can add noise in coastal waters, further complicating the ASW operator's job. Magnetic anomaly detectors, used to find submarines in the open ocean, can be especially confounded by the clutter of a shallow seafloor and the "magnetic garbage" that litters the coastal plain.

To detect submarines and determine their location, ASW operators must catalogue other sound sources in the region where submarines might travel and map thermal, depth and salinity profiles and bottom conditions that can affect the path of acoustic emissions and sonar returns [see "The Amateur Scientist," page 90]. The U.S. Navy has only begun to turn its attention to this problem for waters such as the Persian Gulf, which was free of submarines un-

til 1992. At that time, Iran acquired its first Kilo boat, and the U.S. assigned two Los Angeles–class nuclear-powered attack submarines to patrol and map the area.

Although diesel submarines have many advantages when deployed under appropriate conditions, they are not without weaknesses. Their engines make more noise than do nuclear reactors and cannot drive a submarine as fast. When running at high speed under electric power, a submarine can deplete its batteries in a few hours. Even at slower speeds it must still approach the surface to take in air every four to 10 days, depending on the submarine's capabilities and the captain's willingness to risk running out of power to avoid detection. Consequently, ASW forces can prevail by blanketing an area with vessels and aircraft. Admiral Henry Mauz, U.S. Atlantic commander in chief, explains, "If you don't let him snorkel, you hold him down. Pretty soon he can't work—it's too hot, too steamy, too much carbon dioxide and monoxide."

The newest submarine designs aim to reduce these liabilities. The Kilo and Type 209, for example, emit much less noise when snorkeling than do their predecessors. Moreover, Swedish, German, Italian, Russian and South Korean shipyards are developing air-independent propulsion (AIP) systems, which eliminate the need for frequent snorkeling and may enable a vessel to remain at depth for up to a month. Sweden has tested and incorporated into its next-generation design an AIP system using a Stirling engine, an external combustion engine that does not burn fuel explosively and is thus much quieter than a standard gasoline or diesel engine. Other designs may use liquid oxygen and high-efficiency combustion systems, or chemical fuel cells with up to five times the net energy density of lead-acid batteries.

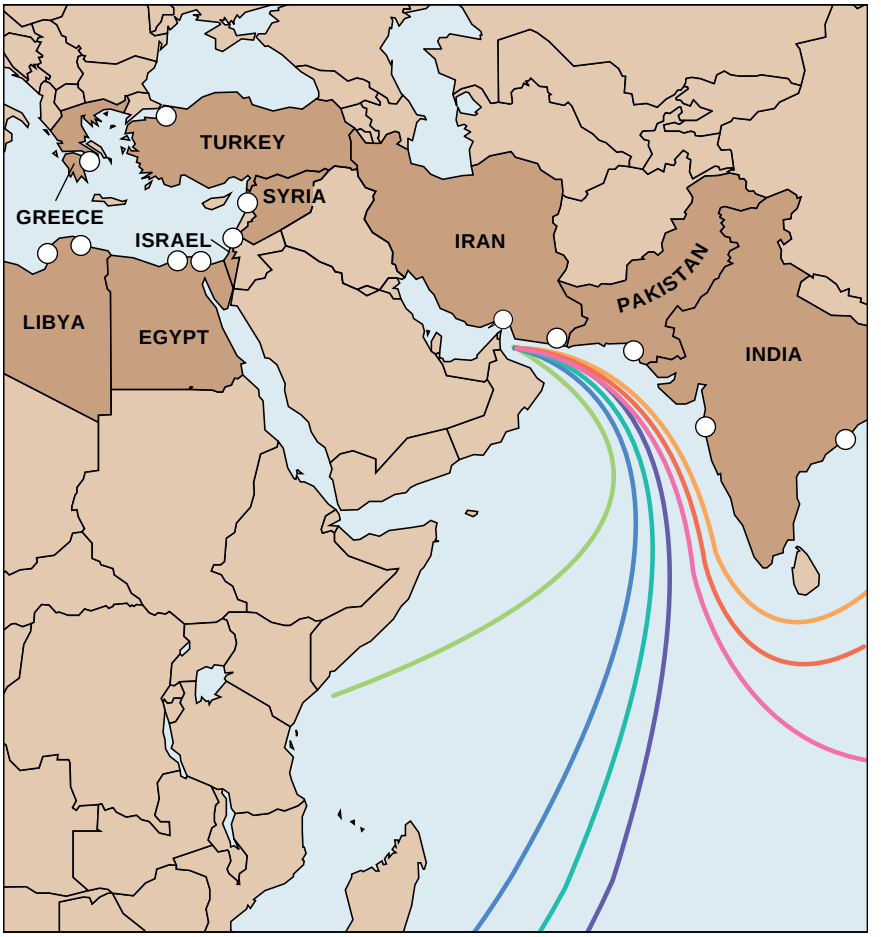
Most submarine fleets fielded by Third World countries do not currently present an insuperable threat to naval operations. U.S. Navy representatives point out that "only a relatively small proportion of the ocean is less than 1,000 feet deep, and most of that is less than 30 miles from shore. Controlling the deeper water," they contend, "guarantees battle group operation safety and 'bottles up' potential threats in restricted shallow water areas, where they are more susceptible to mines and other forces, while ensuring the sea lanes of communication remain open."

The new Kilos, to be based in southern Iran, are regarded by one U.S. intelligence official as so easy for U.S. aircraft to find and destroy that eliminating them would be little more than a "live fire exercise." Less capable submarines do not necessarily pose a serious danger even in large numbers. North Korea's fleet, for example, consists of antiquated Chinese-built Romeo-class vessels, a type the Soviet Union stopped selling in 1960. Libya's submarine crews have a reputation for being poorly trained, and their boats are so shoddily maintained that only one or two out of six may be operable—not one has routinely gone to sea since 1985.

Faced with this mixed situation, the U.S. Navy has taken two contradictory positions. In its posture statement the service pledges to "ensure we maintain the ASW edge necessary to prevail in combat along the littoral," thus implicitly acknowledging that its current ASW forces are adequate to meet existing and near-term threats. At the same time, officials are justifying a new nuclear attack submarine program and several new helicopter, sonar, radar, torpedo and ship defense projects based in large part on the peril that could arise from diesel submarines in shallow water.

Indeed, the dangers that submarine fleets of the developing world present to U.S. forces will increase if nations continue to export more advanced and stealthy diesel submarines and weapon systems. Are there ways to limit the spread of the submarines?

It is difficult to convince exporters that halting the sale of submarines to



PERSIAN GULF has been the site of submarine operations since 1992, when Iran received its first submarine from Russia and built a base at Bandar Abbas. The U.S. then assigned two Los Angeles–class nuclear-powered attack submarines to patrol and map the area. Roughly a quarter of the world's oil passes this single maritime choke point.

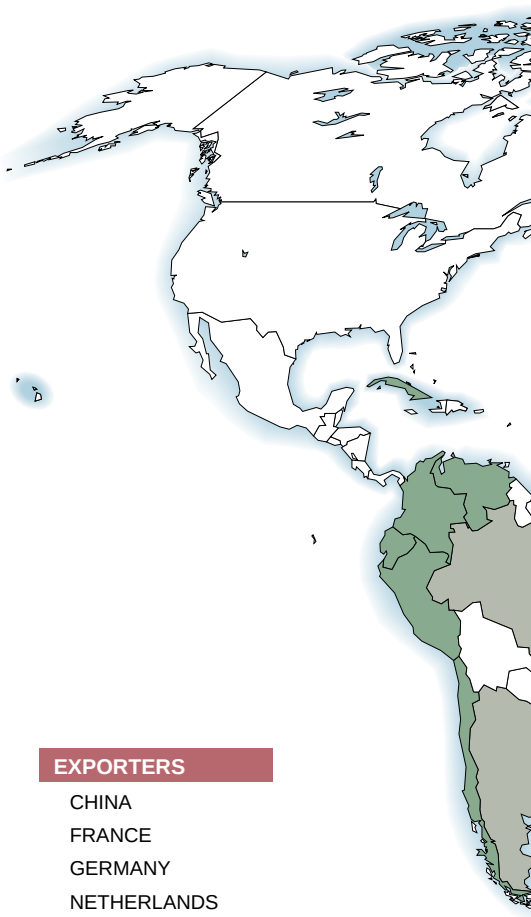
the Third World would be in their best interests, but the idea of forgoing potential sales is not unprecedented. In 1987, when Western countries became sufficiently alarmed about ballistic missile proliferation, they managed to put aside their financial interests to limit the sale of missiles and related technology. The Missile Technology Control Regime (MTCR) bars the transfer of missiles, equipment or know-how that could lead to widespread proliferation.

Missiles were an object of special concern because they could penetrate enemy defenses and were highly suitable for surprise attack—destabilizing characteristics also shared by submarines. Attack submarines in the hands of rogue states raise the specter of terrorism against commercial shipping and could also wreak havoc against major-power forces attempting to operate in littoral waters. As with the MTCR, the best way to stop the spread of submarines to potentially hostile regimes is to control the export of these weapons worldwide. Routine sales of ballistic missile capabilities are no longer considered a legitimate commercial opportunity for nations to exploit. The same can be done for submarines.

IMPORTERS	HAVE	PLAN
ALGERIA	2	–
CHILE	4	–
COLOMBIA	2	–
CUBA	3	–
ECUADOR	2	–
EGYPT	8	2–6
GREECE	10	–
INDONESIA	2	–
IRAN	2	–
ISRAEL	3	2
LIBYA	6	–
MALAYSIA	–	?
PAKISTAN	6	3
PERU	9	–
PHILIPPINES	–	?
SAUDI ARABIA	–	?
SINGAPORE	–	?
SOUTH AFRICA	3	–
SYRIA	3	–
TAIWAN	4	4
VENEZUELA	2	–

CO-PRODUCERS	HAVE	PLAN
ARGENTINA	4	4
BRAZIL	4	3
CHINA	45	–
INDIA	18	6
NORTH KOREA	25	–
SOUTH KOREA	4	8
TURKEY	15	7

EXPORTERS
CHINA
FRANCE
GERMANY
NETHERLANDS
RUSSIA
SWEDEN
U.K.



PRIMARY SOURCE: International Institute for Strategic Studies

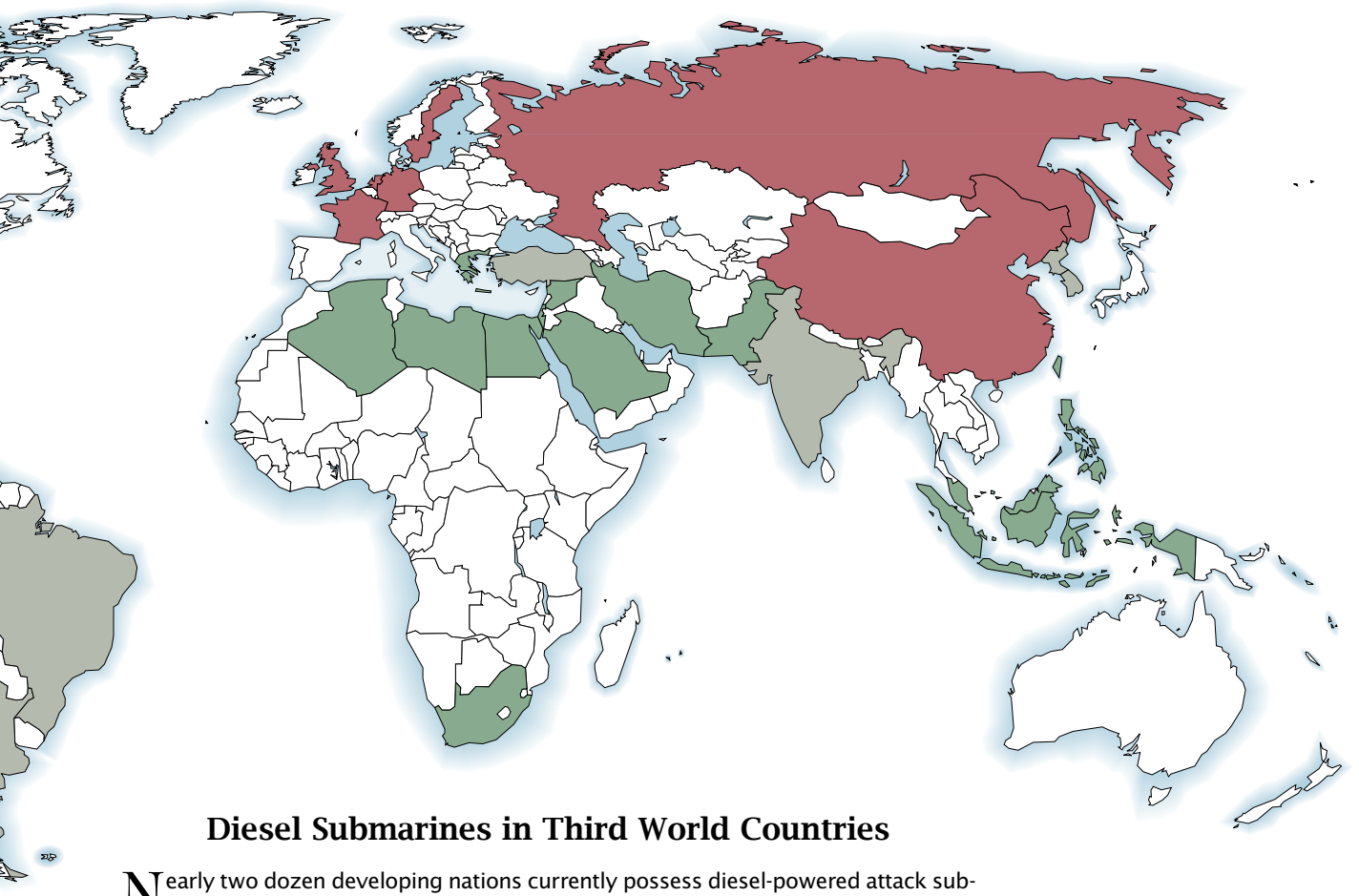
The market may not be such a large one for the developed countries to give up. Modern submarines cost too much for most countries—Pakistan, for example, would pay $233 million for each of three Agosta 90 models it is seeking to purchase from France. But China is competing with France for the Pakistani sale. Both countries are offering generous financing packages that reduce the profitability of the deal. In today's buyers' market, cash-paying customers are few. In the U.S. deal with Egypt, the revenues that Ingalls shipyard would receive are U.S. taxpayer dollars, already required to be spent on U.S. goods and services.

Many submarine sales involve agreements to license the designs and technology for building the boats. Thus, the purchaser may become independent and may even compete with the original seller for future orders. Brazil, Argentina, South Korea and India, all former submarine purchasers, have produced some of their own vessels. It was precisely such proliferation of production capabilities that spurred formation of the MTCR. The developed countries may similarly wish to act before losing control of the world trade in submarines, along with the market itself, to Third World submarine producers.

Submarine exports are sometimes justified on the basis of the need to preserve the defense industrial base, but the capabilities that are preserved may not be all that useful for a modern nation's own defense. Germany has sold Type 209 submarines for nearly 20 years, but there is not a single Type 209 in the German navy. Of greater aid in maintaining a submarine industrial base in Germany and Sweden are current domestic construction orders for submarines with air-independent propulsion systems, which will provide work through the late 1990s. For the U.S., production of diesel vessels in Mississippi would not help maintain nuclear submarine production in Virginia and Connecticut, although it would help keep Ingalls afloat. Instead it would create a production line whose output the U.S. Navy is interested neither in purchasing nor in seeing proliferated around the globe.

A good step toward eventual control of submarine exports might be to restrict the sale of advanced submarine-launched weapons, such as modern torpedoes and antiship cruise missiles. These weapons, a single one of which can sink a large surface vessel, are particularly destabilizing. Furthermore, the U.S. could set an example by stopping the export of Harpoon missiles. These antiship weapons allow a submarine to attack a target such as an aircraft carrier from as far away as 90 miles, well beyond the reach of its inner defenses.

Missile and torpedo sales valued in the hundreds of thousands of dollars may be easier for governments to resist than submarine sales worth hundreds of millions. Whereas even the most basic torpedo can sink a ship, more modern weapons, which are faster, stealthier, longer range and better guided and which can defeat modern countermeasures, could place naval forces in imminent peril. By limiting sales of undersea ordnance to the most basic types, exporters would limit the threat from



Diesel Submarines in Third World Countries

Nearly two dozen developing nations currently possess diesel-powered attack submarines. Many of these countries are seeking to expand or modernize their fleets, and a handful of additional nations intend to join the submarine club. Meanwhile a growing set of exporters (including some former and current submarine buyers) is competing for the developing nations' business. The U.S., which has not made diesel submarines for about 30 years, is about to reenter the export market.

existing boats. An agreement restricting coproduction or sale of submarine production technology would be another logical move toward cessation of submarine exports in general.

Countries that purchase submarines would be expected to object to restrictions on their availability. An outright ban on sales would affect neighbors and enemies equally, however. An effective international agreement could prevent naval arms races before they begin.

Given the long lifetime of submarines and other advanced weapons, exporting them even to countries that are now staunch allies is a risky business. Iran had six German Type 209 submarines on order at the time of its fundamentalist revolution. Had those weapons been delivered, Iran would likely have used them to great effect against Kuwaiti and Iraqi oil shipments during the Iran-Iraq war and could have turned them against the U.S. fleet when it intervened to protect those deliveries.

Although Third World submarines do not pose an overwhelming threat at present, continued sales of modern submarines and munitions have led to real and serious proliferation risks. Submarine-producing countries need to look beyond short-term commercial interests to long-term security necessities and organize a regime whereby the sale of advanced submarines is slowed or halted entirely.

FURTHER READING

THERE IS A SUB THREAT. Rear Admiral James Fitzgerald, U.S.N., and John Benedict in *Proceedings of the U.S. Naval Institute*, Vol. 116, No. 8, Issue 1050, pages 57–63; August 1990.

...FROM THE SEA: PREPARING THE NAVAL SERVICE FOR THE 21ST CENTURY. U.S. Department of the Navy, September 1991.

THE SUBMARINE. Special section in *Navy International*, Vol. 97, Nos. 11/12, pages 311–330; November/December 1992.

THIRD WORLD SUBMARINES AND ASW IMPLICATIONS. John R. Benedict, Jr., in *ASW Log* (now called *Airborne Log*), pages 5–8; Spring 1992.

ATTACK SUBMARINES IN THE POST–COLD WAR ERA: THE ISSUES FACING POLICYMAKERS. Center for Strategic and International Studies, June 1993.

NAVY SEAWOLF AND CENTURION ATTACK SUBMARINE PROGRAMS: ISSUES FOR CONGRESS. Ronald O'Rourke. Congressional Research Service Issue Brief, April 7, 1994.

THE SUBMARINE REVIEW. Published quarterly by the Naval Submarine League, Annandale, Va.

Extreme Ultraviolet Astronomy

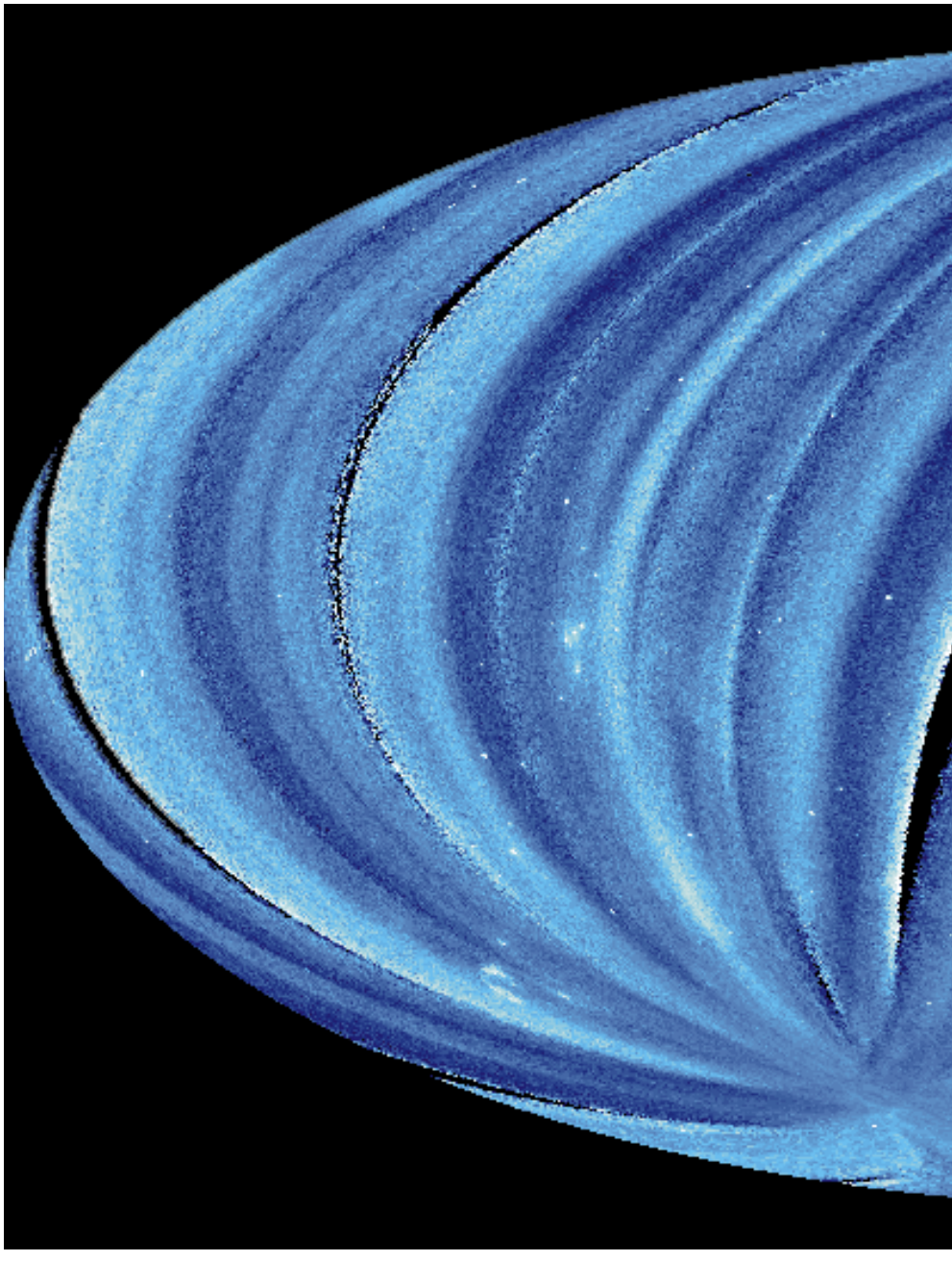
Observations at these wavelengths, once thought impossible, are extending knowledge of the cosmos

by Stuart Bowyer

At noon on June 7, 1992, tense with anticipation, I watched a Delta II rocket lift off its pad at Cape Canaveral, Fla., carrying the National Aeronautics and Space Administration's *Extreme Ultraviolet Explorer* (*EUVE*) satellite. About an hour later the launch vehicle placed *EUVE* into an orbit 550 kilometers above the earth.

The satellite's performance has surpassed expectations. Soaring over the atmosphere, which prevents extreme ultraviolet radiation from reaching earthbound telescopes, *EUVE* has detected a wide variety of astronomical objects. Among them are white dwarfs, coronally active stars, neutron stars and planetary objects in our solar system, all radiating in this high-frequency band. *EUVE* has even seen 10 sources of extreme ultraviolet radiation beyond the Milky Way galaxy. This observation was all the more satisfying because of the long-standing prediction that interstellar gas would absorb all EUV radiation coming from even nearby stars, let alone that from extragalactic objects.

The first satellite dedicated to ex-



STUART BOWYER received his Ph.D. in physics from the Catholic University of America in 1965 and, soon after, joined the faculty of its department of space sciences. In 1967 he became a professor of astronomy at the University of California, Berkeley. There he created a research group involved in extreme ultraviolet and far ultraviolet astronomy and related topics in high-energy astrophysics. In 1989 he founded the Center for Extreme Ultraviolet Astrophysics at Berkeley.

Among other honors, Bowyer has received the National Aeronautics and Space Administration's highest award, the Distinguished Public Service Medal, for his work in developing the field of extreme ultraviolet astronomy.

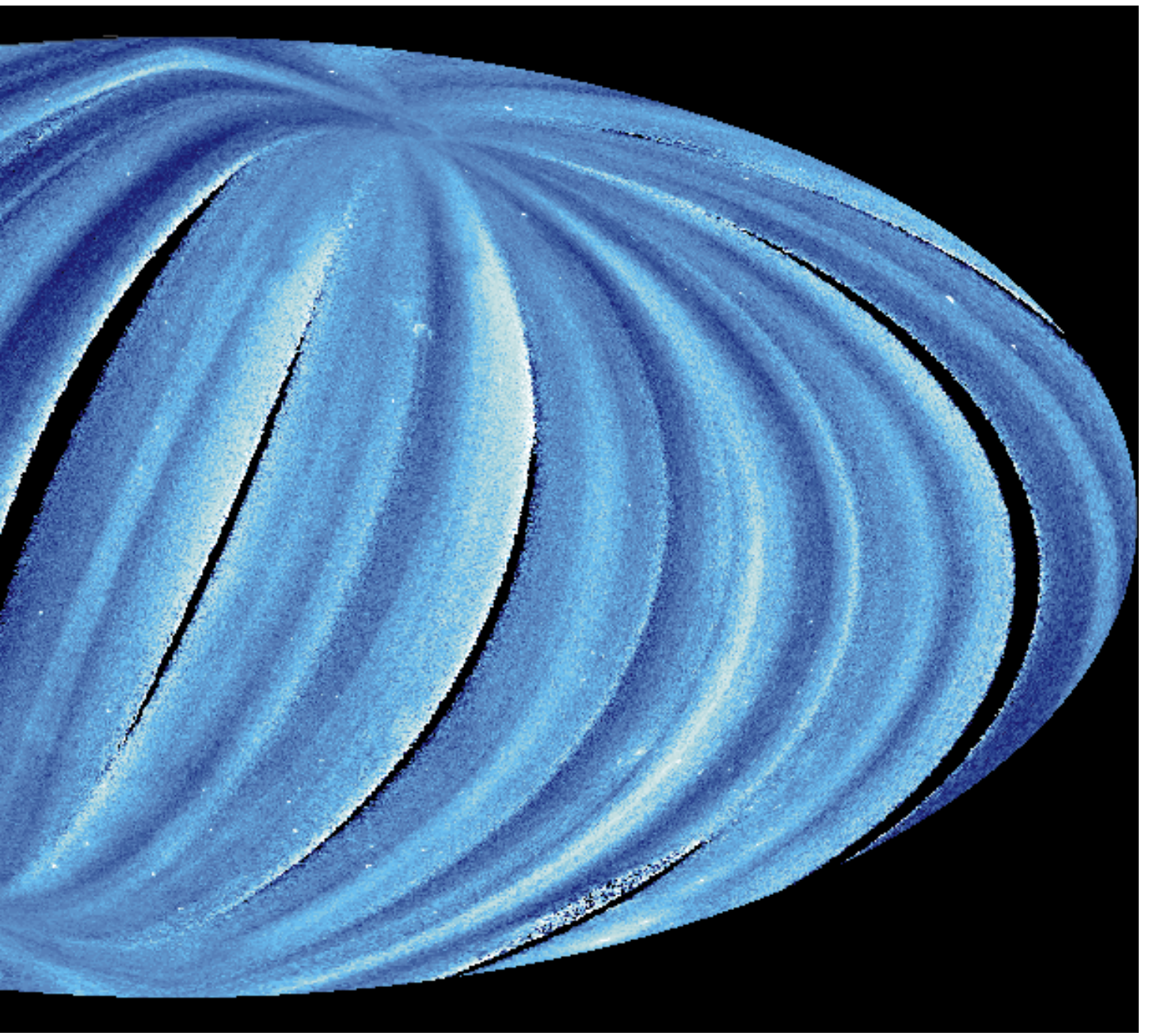
treme ultraviolet astronomy, *EUVE* has in its two years of flight already collected crucial information on a range of astronomical objects. The observations are forcing us to revise our models of hot young stars and white dwarfs, as well as yielding new information on stellar coronae, the interstellar medium and planets in the solar system.

For me and the students and postdoctoral fellows who have worked with me in developing extreme ultraviolet astronomy, *EUVE* and its discoveries represent the culmination of a vision stretching back more than two decades. Seeing our dreams come true has been all the sweeter in light of the prediction that extreme ultraviolet astronomy was a science that was doomed to failure.

During the 1960s and early 1970s, astronomers believed that extreme ultraviolet radiation—having wavelengths between roughly 100 and 1,000 angstroms—would be completely absorbed by the interstellar medium. Thus, such light, if emanating from any star other than the sun, could not reach the earth's vicinity. This calculation was based on an estimate of the average density of gas in interstellar space: one hydrogen atom per cubic centimeter, with lesser amounts of helium and other elements. If this material were uniformly distributed throughout the galaxy, EUV astronomy would indeed be impossible.

There was also a technical hitch: instruments to detect and analyze EUV radiation—in the laboratory, let alone

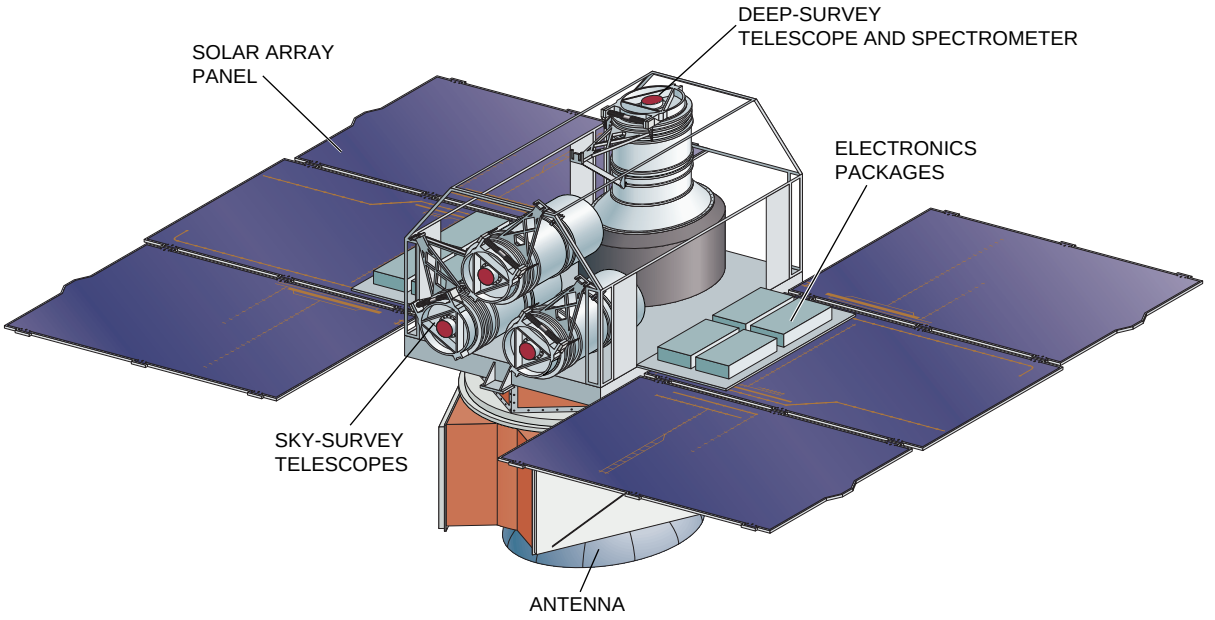
NIGHT SKY glows in extreme ultraviolet radiation of short wavelengths (about 100 angstroms). Six months of observations by the *Extreme Ultraviolet Explorer* (*EUVE*) satellite were compiled in this map. Each stripe corresponds to a continuous phase of data taking; the dark stripes, representing periods during which the survey was suspended for calibration, have since been filled in. The circular feature to the lower left is the Vela supernova remnant. Orion's belt can be seen to the left of center.



The *EUVE* Satellite

Launched in June 1992, the *EUVE* satellite is dedicated to extreme ultraviolet astronomy. The satellite is shown here stripped of its outer insulating covering. Three "sky survey" telescopes look out perpendicular to the axis of rotation of the satellite (*vertical*). The axis lies in the plane of the ecliptic—that in which the earth circles the sun—and points toward and away from the sun. As the satellite rotates, the sky-survey telescopes scan the sky perpendicular to the spin axis; in six months the entire sky is mapped.

The fourth, "deep survey" telescope looks along the rotation axis and scans a small strip of sky in the direction away from the sun. This portion of the sky is observed for substantially longer exposure times, revealing fainter sources of EUV radiation. Half the radiation it collected goes to three spectrometers used by NASA's guest observers since the sky survey was completed. *EUVE*'s observations will continue through 1995; the satellite itself should remain in orbit until at least 1999.



any devices capable of being launched in a rocket or satellite—were nonexistent or in a primitive stage of development. Together these factors conspired to make the extreme ultraviolet band the last frontier in observational astronomy.

It was a frontier that appealed to me. Immediately after completing my Ph.D. in physics with a thesis in x-ray astronomy, I developed some crude instrumentation to detect extreme ultraviolet radiation. Could such technology find a place in astronomy? Not being formally trained as an astronomer, I was not overly impressed by the standard pessimistic picture. But it certainly was cause for concern.

Soon after I came to the University of California at Berkeley in 1967, George B. Field, a leading expert on the interstellar medium, proposed that interstellar matter might be distributed quite unevenly. Its density in many directions might be only one tenth of the average. In that case, extreme ultraviolet radiation would penetrate—in those directions—10 times farther than was normally assumed. Therefore, a volume of space 1,000 times greater than was commonly believed to exist would be observable by EUV light. That would amount to a 1,000-fold increase in the potential number of such sources! I said to myself, "If George Field with a snap of his theoretical fingers can extend the range of EUV observations by 1,000 times, who knows what may be possible in real life?" I decided that EUV astronomy was something I should pursue after all.

Since 1968 I have led a research group at Berkeley with a special focus on EUV astronomy. To this group I have been fortunate to attract a series of talented, adventurous graduate students and postdoctoral fellows.

During the early years, it was difficult to convince astronomers, and NASA in particular, to support efforts to overcome the technical difficulties facing extreme ultraviolet astronomy. But NASA did provide me with a modest grant. So I focused on using sounding rockets—which stayed above the atmosphere for only about five critical minutes before falling back to the earth—as a means of testing and evaluating new technology for extreme ultraviolet astronomy.

In the mid-1970s I led an effort to fly a telescope, primitive by today's standards, that would make extreme ultraviolet observations during the U.S.-Soviet *Apollo-Soyuz* spaceflight. NASA selected our proposal, and in 1975 *Apollo-Soyuz* carried our instrument above the atmosphere. We found four sources of EUV radiation. Two turned out to be hot white dwarf stars and one a star with an active corona. The fourth was a cataclysmic variable star, a binary system that occasionally brightens by a factor of five to 100, named SS Cygni. Because the white dwarfs and SS Cygni are 100 to 200 light-years away, the *Apollo-Soyuz* experiment demonstrated that EUV radiation can, at least in some directions, pass through the interstellar medium for astronomical distances.

The success of *Apollo-Soyuz* gave us a big push forward. In response to a NASA announcement for an explorer satellite program, we submitted a proposal to develop instruments that could map the entire sky in the extreme ultraviolet band. NASA chose our proposal, and so began the years of toil that

would culminate in the *Extreme Ultraviolet Explorer* satellite.

Not all the obstacles were technological. Financial restrictions kept initial funding at a low level. And, notwithstanding the success of EUV observations on *Apollo-Soyuz,* a report issued by the National Academy of Sciences in 1979 suggested that NASA cancel *EUVE.* The recommendation was based on the grounds that less than a dozen EUV sources were likely to be detected. To its credit, NASA continued its support for our work.

The technical difficulties were indeed considerable. Focusing, detecting and analyzing high-energy EUV radiation require instruments that are quite different from those used with visible light. For example, the best optical telescopes use mirrors that gather and focus light by reflecting it at large angles. If the wavelength of the radiation is less than about 500 angstroms, however, it will be absorbed by this type of mirror.

To tackle this problem, x-ray astronomers had pioneered the development of "grazing incidence" telescopes. With these instruments, the radiation strikes the mirror at an angle almost parallel to the surface and is reflected at a similar small angle [*see illustration on next page*]. The trick can be used to focus EUV radiation, but this type of mirror is exceedingly difficult to fabricate. The mirror surfaces must be shaped and polished with painstaking precision, but few of the comparatively inexpensive techniques used for making optical telescopes can be applied to grazing-incidence telescopes.

The cost of figuring and polishing glass surfaces for an EUV grazing-incidence instrument was prohibitive for a field that had yet to prove itself. At Berkeley I involved three graduate students—Webster C. Cash, Roger F. Malina and David S. Finley—in an effort to develop metal telescopes for EUV astronomy. We crafted these from carefully shaped aluminum, coated with nickel (to give a good polishing surface) and then with a thin layer of gold for maximum reflectivity.

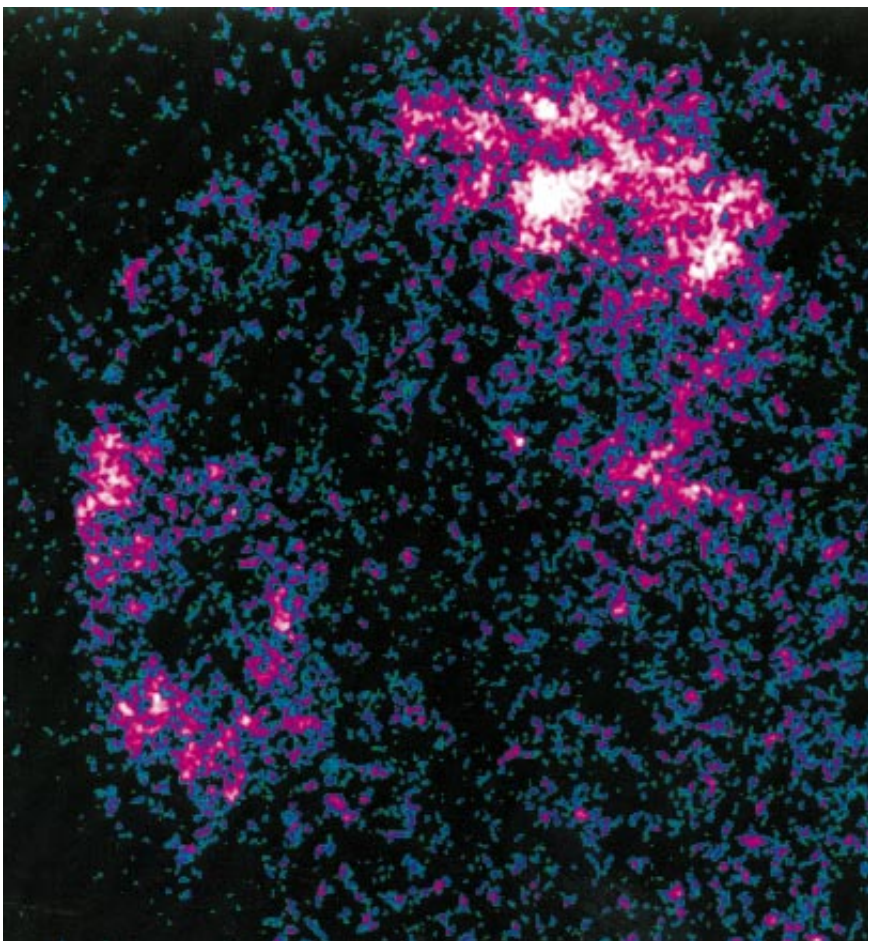
We were fortunate to enlist the help of Lawrence Livermore National Laboratory in this project. Using diamond-turning lathes originally developed to make nuclear weapons, we were able to make mirrors having characteristics that were far better than similar mirrors fabricated elsewhere. In the end, our mirrors were about as good as the glass one flown in the *Einstein* x-ray satellite in 1978, which cost nearly 30 times more.

Developing the grazing-incidence mirrors for EUVE was only part of the battle. We also had to invent and develop detectors for the EUV radiation collected by our telescopes. Several extreme ultraviolet detectors were designed in my laboratory, all based on a common principle. Incoming photons jar loose an electron, which accelerates down a capillary tube—one of an adjoining honeycomb of such tubes—knocking out additional electrons. These in turn eject still more electrons, producing a cascade of several million electrons from a single initial electron.

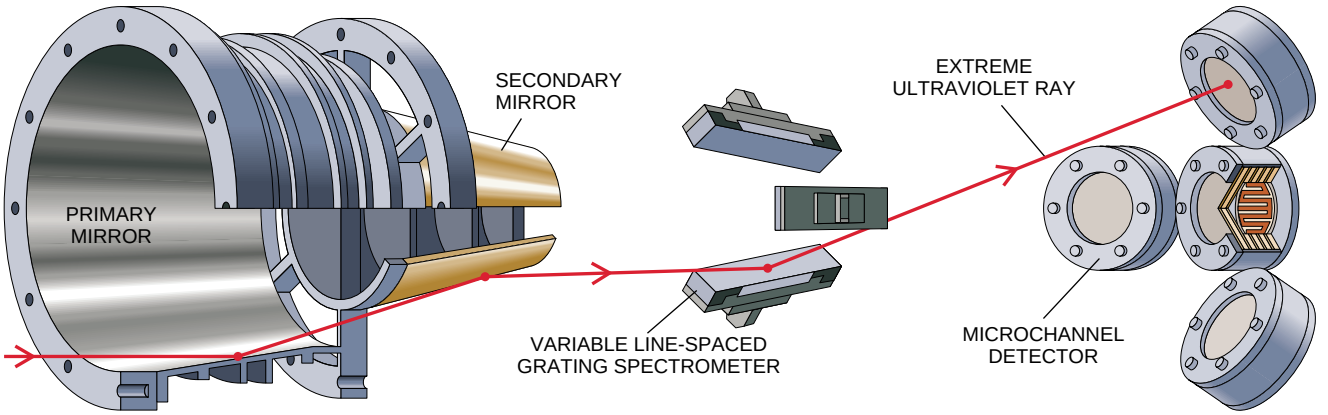
This principle has been employed in many military and civilian night-vision systems. But applying it to extreme ultraviolet astronomy was no easy task. Because EUV sources, like most astronomical objects, are amazingly faint, we had to create an instrument of tremendous sensitivity, one capable of detecting a single photon. In spite of this sensitivity, the detector had to produce an extremely low level of spurious, random signals. Furthermore, we needed a way to convert the cascade of electrons into an electronic signal that would faithfully replicate the image produced by the telescope.

Michael Lampton, my first postdoctoral fellow and now my colleague and collaborator, invented several schemes for reproducing images of the sky from the signals from the detector. The detector we used for *EUVE* generates a picture defined by a matrix of 2,048 by 2,048 pixels. The detector has very low internal noise and is nearly insensitive to the longer-wavelength ultraviolet radiation. [see "The Microchannel Image Intensifier," SCIENTIFIC AMERICAN; November 1981].

But the telescopes and the detector were only part of the story. In the long run, I knew that spectroscopy—the science of analyzing the amounts of radiation of different wavelengths—would be critical to the growth of EUV astronomy, just as it had been in all other fields of astronomy. Hence, I engaged four graduate students—Malina, Patrick Jelinsky, Michael Hettrick and Mark V.



VELA SUPERNOVA REMNANT is one of the brightest extreme ultraviolet sources in the sky. It is, however, very hard to see in visible light. The bright blotches show gas thrown off when the star exploded. Now the gas forms an expanding shell, radiating profusely in the extreme ultraviolet and x-ray bands.



EXTREME ULTRAVIOLET ASTRONOMY requires telescopes, spectrometers and detectors that are very different from those used in optical astronomy. Because this short-wavelength radiation is absorbed by mirrors designed for optical telescopes, "grazing incidence" mirrors had to be employed. In those shown here the radiation strikes the mirrors at a very small angle to the surface. The drawing shows a ray of EUV radiation being focused by mirrors (with the covers removed) in the deep-survey telescope. The radiation then diffracts off a variable line-spaced grating, which disperses it into component wavelengths. Finally, the ray is observed by a microchannel detector, which can "see" even a single photon of radiation.

Hurwitz—in an attempt to fashion EUV spectrometers.

The heart of the device that was ultimately developed is a radically different ruled grating in which the spacing between adjacent lines increases slowly from one end to the other. This "variable line-spaced grating" spectrometer disperses the light in a way that provides unique advantages. In essence, it has high efficiency, is compact and easily adapts to different telescopes.

Notwithstanding our successes in the laboratory, all our instruments were originally designed for use on sounding rockets. So they had to be reconfigured to meet NASA's "spaceworthiness" specifications in order to fly in *EUVE*. Malina was subsequently appointed by NASA to lead a team of scientists and engineers to accomplish these goals and to make sure that the equipment would function properly in orbit.

In the meantime the observations made by *Apollo-Soyuz* had sparked interest in extreme ultraviolet astronomy elsewhere. A group of British astrophysicists decided to build an instrument to survey the sky in the shortest-wavelength part of the EUV band. Their Wide-Field Camera, using extreme ultraviolet technology first developed at Berkeley, was designed for launching piggyback on the German satellite *ROSAT*. (The primary mission of the satellite—named after the discoverer of x-rays, Wilhelm Röntgen—is the observation of x-ray sources.)

The Wide-Field Camera was formally approved in 1980, four years after NASA approved *EUVE,* but fate ordained that it precede *EUVE* into orbit by almost two years. The camera found 350 sources of short-wavelength extreme ultraviolet radiation. In addition to this catalogue of sources, some astrophysical insights were derived from its data.

We had initially planned to operate the post-launch phase of the mission within one of the organizational structures already available at Berkeley. But by the late 1980s it had become clear that the project was too large in scope for the facilities and staff of the available structures. I was able to persuade the university to found the Center for Extreme Ultraviolet Astrophysics (CEA). Scientists and engineers at the CEA monitor the health of *EUVE*'s instruments, control their operation and analyze data from the first phase of observations. They also provide support for NASA's "guest observers"—astronomers from all over the world who use the spectrometer for specific observations.

Four telescopes form the complex of instruments on *EUVE* [*see box on page 34*]. Three of these, the "sky survey" telescopes, point in the same direction. We had originally made each one to observe a different band of wavelengths in the EUV, using special filters developed in my laboratory. Later we devised a fourth filter and a clever packaging scheme that would allow us to explore the EUV sky in four bands instead of three. We therefore reconfigured the survey telescopes so that all four bands would be observed.

The direction in which the survey telescopes look out is perpendicular to the axis of rotation of the *EUVE* satellite. This axis lies in the plane of the ecliptic—that of the earth's orbit around the sun—and is pointed toward and away from the sun. As the satellite spins, the telescopes scan a strip of the sky; the strip shifts daily as the earth travels in its orbit around the sun. The entire sky is mapped in six months.

The fourth, "deep survey" telescope is aligned parallel to the axis of rotation of *EUVE*. Thus, it looks away from the sun. Within six months the telescope scans a small strip of sky in the plane of the ecliptic. The prolonged exposure allows more sensitivity than does the main survey and reveals fainter sources. Half of the incoming radiation is used for the deep survey; the other half is divided equally among the three spectrometers. Because a single grating can efficiently reflect wavelengths differing by only a factor of two or three, three spectrometers are needed to cover the entire EUV band.

In January 1993 *EUVE* finished the first surveys of the sky covering the entire range of extreme ultraviolet radiation. To incorporate the enormous amounts of data returned by the satellite, we had to address a common problem. As well as extreme ultraviolet photons, our detectors are sensitive to cosmic rays and charged particles captured in the earth's magnetic field. These particles cause a background, or noise, in which our faint astronomical signals are embedded.

Through considerable effort, we were able to create reliable algorithms to distinguish true sources from false ones. Using these programs, we have now carried out our initial processing of the all-sky data and found more than 400 distinct EUV sources. In the next processing, due to be completed next month, we will be introducing even more sophisticated techniques. We expect to find up to 1,000 sources.

Our first results show that many of the extreme ultraviolet sources detected are stars with active coronae and hot white dwarfs. The remainder of the

current harvest constitutes a diverse collection of astronomical objects: cataclysmic variable stars, hot young stars, extragalactic sources, supernova remnants and neutron stars.

In January 1993 *EUVE* entered its second phase of operation: studying particular sources intensively with the spectrometers. Astronomers who have convinced NASA of the value of their research proposals are participating in this second phase. These guest observers each look at their chosen object for some 10 to 100 hours during the "dark time," in which the satellite is shielded by the earth from the sun's radiation. Notably, *EUVE* is able to direct its axis toward a location in the sky, while moving in orbit, to an accuracy better than 10 arc seconds, the angular size of a dime seen from 700 meters away.

A group of astronomers led by Arieh Königl of the University of Chicago made one of the most exciting observations using *EUVE*. The discovery concerns an extragalactic object, PKS 2155-304, so called because it was first seen with the Parkes radio telescope in Australia. PKS 2155-304 is an elliptical galaxy, which emits an extremely bright jet of hot material that we seem to be observing head-on. Radiation from the jet completely dominates all other emissions from the galaxy.

Such "BL Lac objects" are known to radiate light of almost all wavelengths, from x-rays through radio waves. A unique feature of their radiation is that it is virtually featureless and thus provides few clues as to the physical conditions in these objects. *EUVE*'s spectroscopy established that PKS 2155-304 is detectable to wavelengths as long as 120 angstroms. The observation proved that EUV radiation, which was not supposed to be able to penetrate the interstellar medium, could travel intergalactic distances. More important, a number of absorption features were found in the EUV spectrum. A detailed study of these features will reveal profound insights into the physical conditions of such objects. In particular, we may learn how matter falling into the black hole at the galaxy's center is converted into the relativistic jet that we observe.

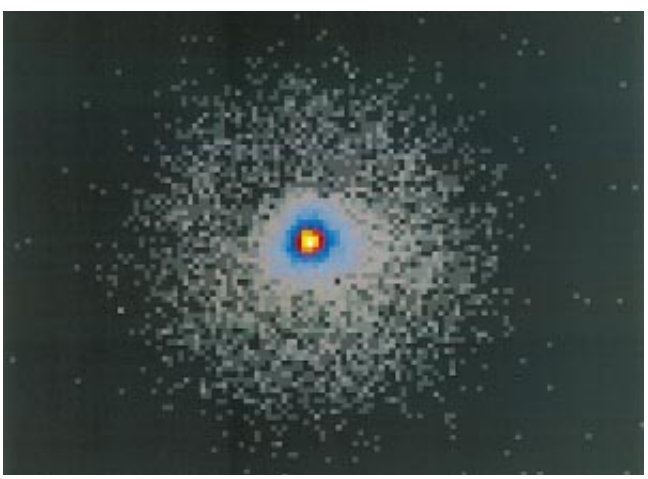
Most sources detected by *EUVE* are stars within the Milky Way that exhibit active coronae. In "normal" stars like our sun, the corona is an extended envelope of rarefied gas. Energy transported from the underlying denser, cooler layers heat the corona in some unknown manner to one or two million degrees Celsius. Between the star's apparent surface—called a photosphere—and its corona, there lies a transition region through which the temperature rises abruptly. The radiation from the transition region consists primarily of extreme ultraviolet rays and x-rays. The gas in the transition region is very thin, however, and the total energy radiated is only a millionth of that of the photosphere.

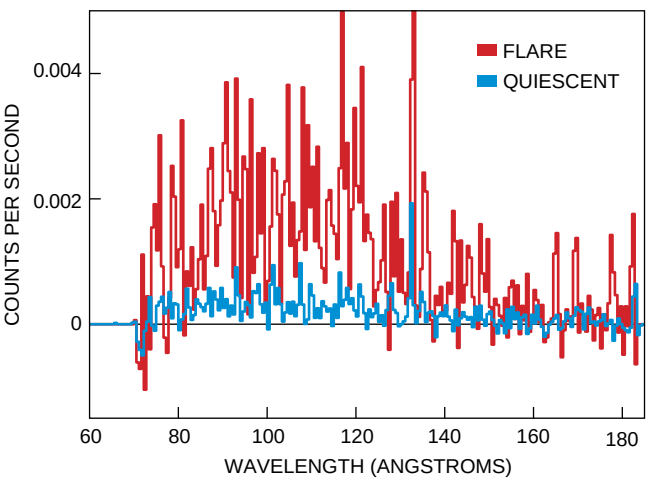
In some stars the corona radiates a lot more energy than usual. In addition, the radiation typically comes from plasma at greater temperatures than in a normal corona. Such coronally active stars reveal much about the corona and transition region even of sunlike stars.

Andrea K. Dupree of the Harvard-Smithsonian Center for Astrophysics and her colleagues have used *EUVE* to study the extreme ultraviolet spectrum of Capella. This system of two coronally active yellow giant stars is 45 light-years away. These astronomers have found strong emission lines produced by ions of iron—atoms of iron that have lost anywhere between 14 and 23 of their 26 electrons. The emission lines demonstrate for the first time that the system contains ionized gas (plasma) at temperatures ranging from a few tens of thousands of degrees up to many millions of degrees. Surprisingly, there is a paucity of plasma at close to one million degrees. In this respect, the Capella system is very different from the sun, which has a (seemingly well understood) excess of material at one million degrees C. What could possibly produce a lack of material at this temperature? So far we do not know.

Dupree and her colleagues have observed a large amount of plasma at a temperature close to six million degrees C in the Capella system; we also do not know what produces this hot material. Plasma confined by magnetic fields may undergo fusion near one or both of the stars, releasing energy. Or the matter could be heated through some interaction between the two stars. By observing variations in the EUV radiation from Capella, we may eventually be able to determine the underlying mechanisms operative in this stellar system.

Flares, or sudden outbursts of extreme ultraviolet and x-ray radiation, often occur on the sun and other stars. The outbursts on coronally active stars make those on the sun look quite modest. Less than six weeks after *EUVE* was launched, we observed two large flares on the star AU Microscopii, about 30 light-years away. In the first of these flares, AU Mic increased its extreme ultraviolet output 20-fold during a few minutes and then sank back to its usual level of emission over the next few





AU MICROSCOPII, a coronally active star, is imaged by the short-wavelength band of the sky-survey telescope (*left*). The spectrum of the star is shown (*right*) in its quiet phase and during a flare. The flare consisted of a sharp peak in EUV radiation that lasted for two hours, followed by a decaying tail that lasted for more than a day.

hours. Analyzing the radiation from AU Mic with the spectrometers, we noted substantial differences in its quiet and flare phases [*see illustration on preceding page*]. Through additional observations we hope to improve our understanding of these violent, transitory phenomena.

A major surprise from *EUVE*'s all-sky survey was the discovery that Epsilon Canis Majoris, a young, extremely massive hot star in the Big Dog constellation, is by far the brightest EUV source in the sky, even though it is more than 600 light-years away. John Vallerga and his co-workers at the CEA found this star in the longest-wavelength band of the all-sky survey. In this band, absorption by interstellar matter is the most severe. The observation implies that in the direction of this star, there are fewer than 0.002 hydrogen atom per cubic centimeter, about a factor of 1,000 below the average for our galaxy.

The EUV spectrum of Epsilon Canis Majoris defies all that we know about the atmospheres of hot young stars. Joseph P. Cassinelli of the University of Wisconsin led a group that analyzed the spectrum of this star, in the hope of understanding the winds that x-ray studies have indicated are emanating from these objects. Emission from these winds was detected and is being analyzed. Much more unexpected was that emission was also detected from the star's photosphere.

Dozens of studies carried out on Epsilon Canis Majoris in the far ultraviolet, the visible and the infrared bands of radiation have all confirmed current models for this class of stars. But the EUV flux from the star's photosphere exceeds the predictions of these "well understood" models by a factor of 30. Although several astrophysicists have speculated about causes of this excess, their explanations are widely divergent. Further work will certainly be needed before we clarify this anomaly.

Epsilon Canis Majoris has also elucidated the ionization of the interstellar medium. Such ionization occurs when an interstellar atom absorbs a photon, liberating one of the atom's electrons. The EUV radiation from this star is so intense that—in its quadrant of the sky—the star is the dominant cause of the ionization of the interstellar medium. Detailed studies of the character of the interstellar medium, incorporating this new result, are now being carried out.

When extreme ultraviolet light causes ionization, it is absorbed in the process. This absorption is evidenced in the radiation from an astronomical source by dark lines or an absorption edge in the spectrum. By studying these effects, we can learn about the temperature of the interstellar medium and the density and degree of ionization of each element in it. Hydrogen, helium atoms and singly ionized helium ions are the primary absorbers of extreme ultraviolet radiation. My collaborators and I at the CEA have studied this absorption in the extreme ultraviolet spectrum of a hot white dwarf, GD 246, which lies about 200 light-years away. Along the line of sight to this star the hydrogen atoms have an average density of about 0.04 per cubic centimeter, and about 25 percent of the helium is ionized.

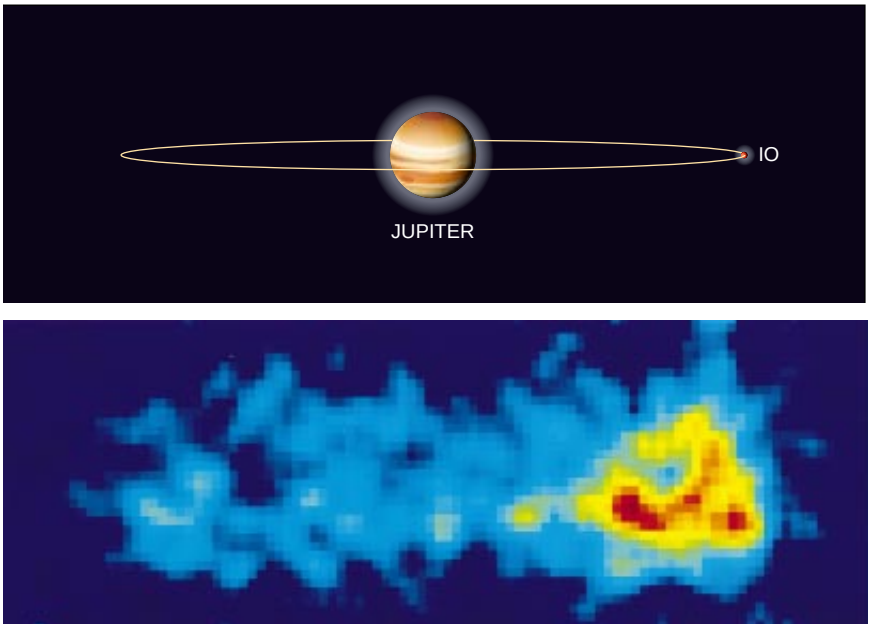
The continuous spectra produced by white dwarfs are ideal for carrying out these studies. We have now made similar observations of about a dozen white dwarfs viewed in different directions. By analyzing these results, we will be able to deduce details about the overall ionization state of the interstellar medium. This information is crucial in developing an understanding of how this material has evolved. Ultimately it should tell us how concentrations of the interstellar gas arise and eventually form new stars.

Hot white dwarf stars, the second most numerous class of EUV sources, have in fact furnished their own set of surprises. The extreme ultraviolet spectra of many white dwarfs have proved far weaker than theorists had expected, forcing us to revise our models of their atmosphere. It is worth taking a moment to see why white dwarfs were expected to dominate the sky in the extreme ultraviolet band—and why they do not.

A white dwarf packs a mass roughly equal to the sun's within a volume equal to the earth's and therefore produces tremendous gravitational forces at its surface, typically 100,000 times the force of gravity on the earth. For 50 years, theorists had concluded that such a large gravitational field would make the denser material in the white dwarf sink downward, separating the atmosphere into layers of varying composition. All the heavy elements would be at the center, leaving the white dwarf's outer layers made up of nearly pure hydrogen or, in the absence of hydrogen, nearly pure helium.

In theory, it was straightforward to demonstrate that any hot white dwarf whose outer layers consisted of pure hydrogen or helium must emit enormous amounts of EUV. But the observations made by *Apollo-Soyuz,* by *EXOSAT* (another x-ray satellite with some capacity to observe in the extreme ultraviolet) and now by *EUVE* revealed very modest amounts of such radiation.

Stephane Vennes of the CEA had earlier postulated why white dwarfs might suffer a deficit in extreme ultraviolet radiation. *EUVE*'s spectroscopy proved him to be correct. Extreme ultraviolet light emitted from lower regions of the white dwarf's atmosphere pushes iron



IO TORUS, a ring of gas containing oxygen and sulfur ions that surrounds Jupiter, was first seen in a *Voyager* flyby. The gas is emitted by Jupiter's volcanic moon Io (*top right*). The *EUVE* picture (*bottom*) shows that one side of the ring is much brighter than the other. Because the gas is probably heated by Jupiter's magnetic field, the observation provides clues to the field's structure.

DEEP-SURVEY TELESCOPE and spectrometers are being prepared for integration with the *EUVE* satellite. The telescope takes long exposures of the sky in the direction of the earth's shadow. The spectrometers, which receive light from sections of the same mirror, analyze the radiation from individual sources into component wavelengths.

ions into the upper atmosphere: the iron absorbs the photons, whose upward momentum produces an upwardly directed radiation pressure. The ions are levitated to the surface, where they substantially reduce the amount of EUV radiation emitted from the white dwarf.

Data from the deep survey telescope have provided information about the diffuse high-energy astronomical background, radiation believed to be generated by the hot material in the interstellar medium. The origin, stability and characteristics of this hot material are not well understood; they are not even well defined. A team at the CEA found a shadow in the diffuse background. The shadow is cast by an isolated cirrus cloud that, though very tenuous, completely absorbs EUV radiation from more distant locations.

All the extreme ultraviolet radiation we detect in the direction of the cloud emanates from the sharply demarcated region between the earth and the cloud. We enlisted the help of Jens Knude of the Copenhagen University Observatory, who determined the distance of the cloud to be 120 light-years. With this information we were able to determine directly the pressure of the hot interstellar medium. We found this pressure to be surprisingly high compared with earlier (indirect) estimates of 700 to 6,000 kelvins per cubic centimeter for this parameter; it turned out to be 19,000. (In terms of human experience, this pressure is in fact exceedingly low: about a millionth of a trillionth of atmospheric pressure at sea level.)

Even in the well-explored realm of planetary physics, *EUVE* has begun to make important contributions. A team of astronomers led by H. Warren Moos and Doyle T. Hall of Johns Hopkins University has obtained striking EUV images of the plasma torus around Jupiter [*see illustration on opposite page*]. The *Voyager* flybys in 1979 revealed that the torus is made mostly of oxygen and sulfur ions initially emitted by Jupiter's volcanic moon Io. The EUV images show that one side of the Io torus is brighter than the other, which means that the gas on that side is in a hotter and denser environment. Because this compression and heating are very likely the result of the motion of the gas within Jupiter's magnetic field, observations of the Io torus provide valuable clues about the interior structure of Jupiter's magnetosphere.

EUVE's phase of guest-observer spectroscopy will continue through 1995. Although we cannot know what discoveries are in store, those that have been made so far are certainly encouraging. Further down the line, progress in extreme ultraviolet astronomy will depend in large part on NASA's backing for the endeavor. The *EUVE* satellite itself will remain in orbit until at least 1999. If an extended mission for the satellite is authorized, we will be able to enter the next millennium with our first orbiting extreme ultraviolet observatory still sending streams of new results back to the earth.

FURTHER READING

VARIABLE LINE-SPACE GRATINGS: NEW DESIGNS FOR USE IN GRAZING INCIDENCE SPECTROMETERS. Michael C. Hettrick and Stuart Bowyer in *Applied Optics,* Vol. 22, No. 24, pages 3921–3924; December 15, 1983.

EXTREME ULTRAVIOLET EXPLORER MISSION. Special issue of *Journal of the British Interplanetary Association,* Vol. 46, No. 9; September 1993.

ASTRONOMY AND THE EXTREME ULTRAVIOLET EXPLORER SATELLITE. Stuart Bowyer in *Science,* Vol. 263, pages 55–59; January 7, 1994.

Confocal Microscopy

For producing sharp two- or three-dimensional images with light, this microscopic technique is unsurpassed. It can also be applied for seeing deep inside the tissues of living specimens

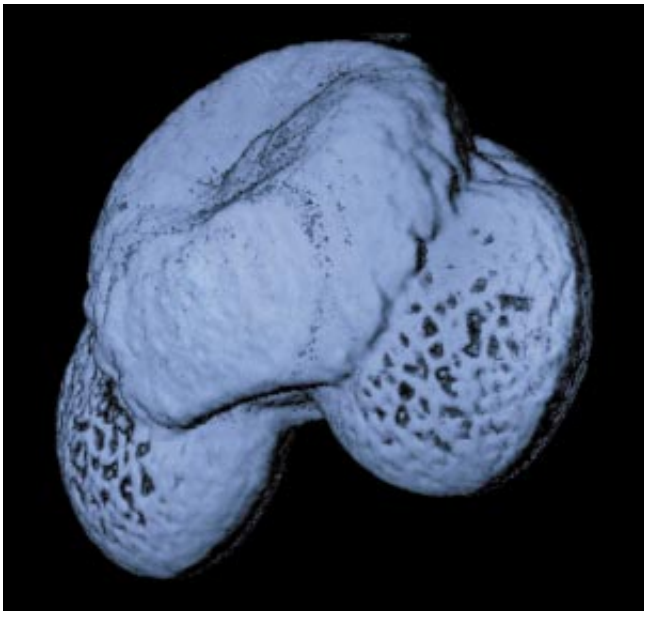
by Jeff W. Lichtman

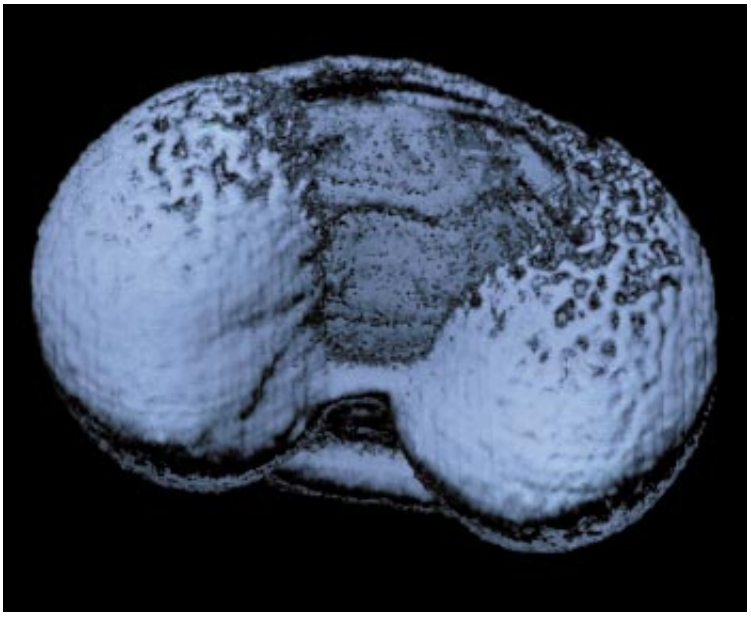
Marvin Minsky is famous as the father of artificial intelligence, but he was also the author of another signal achievement. In the 1950s, as a postdoctoral fellow at Harvard University, he built a revolutionary light microscope that enabled him to view successively deeper layers in a specimen with astonishing clarity, without first having to undertake the laborious task of cutting the specimen into thin sections. Minsky's invention did not earn wide acclaim at the time. In fact, when he patented his "double-focussing stage-scanning microscope" in 1961, few people understood what it could do. During the 17-year life of the patent, he received no royalties, and no instruments of similar design were manufactured. Unappreciated for his foray into optics, Minsky moved on to other challenges, leaving his prototype to rust in a corner of his basement.

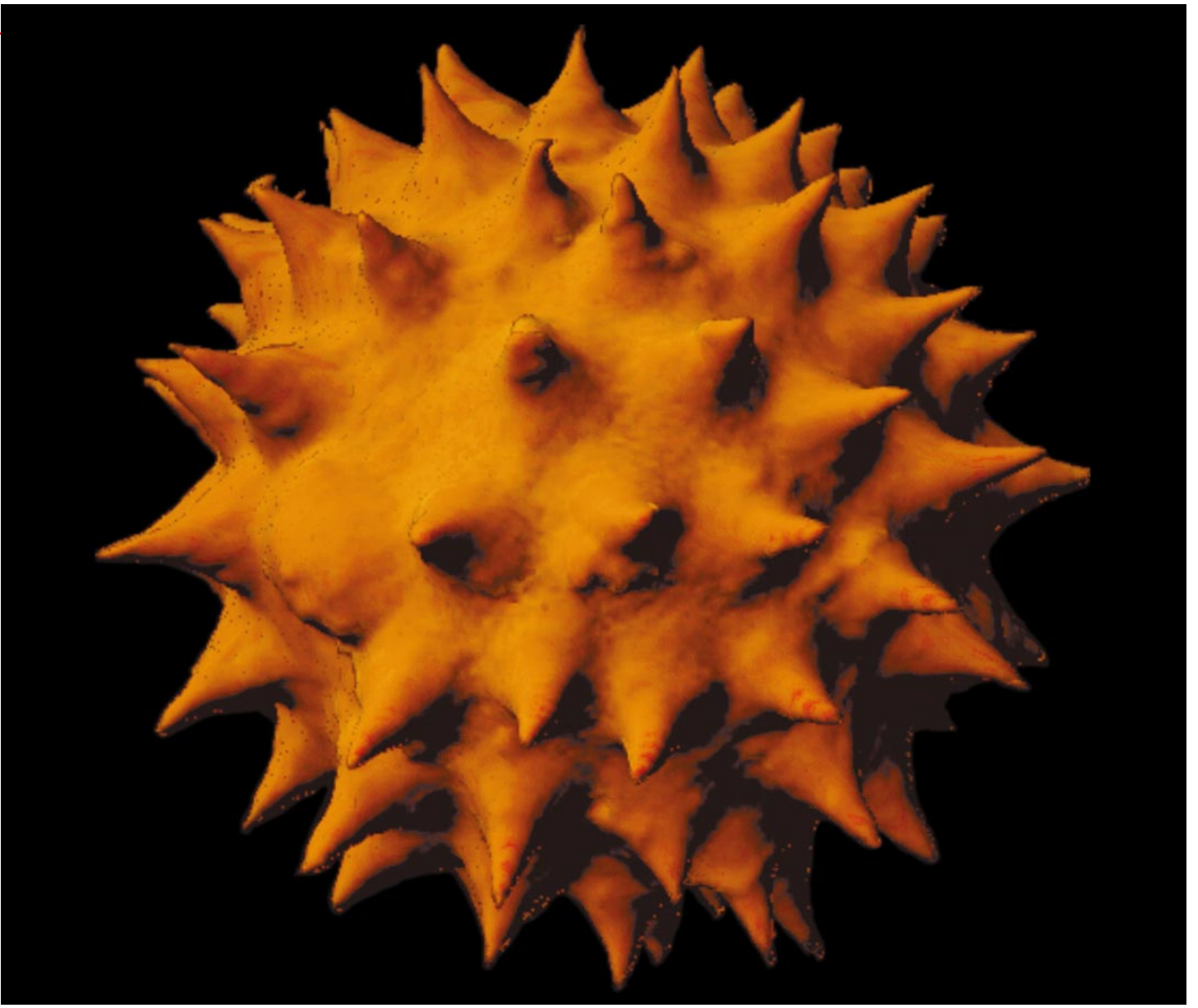
Thirty years later his approach—otherwise known as confocal microscopy—has caught on with a vengeance. Indeed, the technology is proving to be one of the most exciting advances in optical microscopy in this century. The extent to which current interest was sparked by rediscovery of Minsky's early work or by independent reinvention of his concept by others is not completely clear. Nevertheless, the happy result is that scores of different kinds of confocal microscopes are now available—in forms that range from rudimentary to baroque. Whether researchers need to image the ultrastructure of potato chips or computer chips, the diseased eye or the

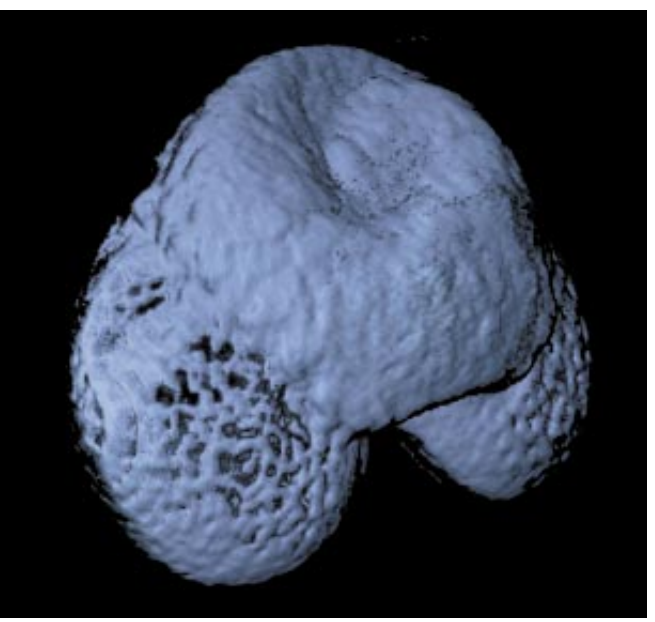
JEFF W. LICHTMAN, who earned an A.B. from Bowdoin College and an M.D. and Ph.D. from Washington University, is professor of neurobiology at the Washington University School of Medicine. He spends most of his time studying long-term changes in the structure and function of synapses. Lichtman is also an inventor of microscopic equipment; he holds several patents for confocal microscopes and their components.

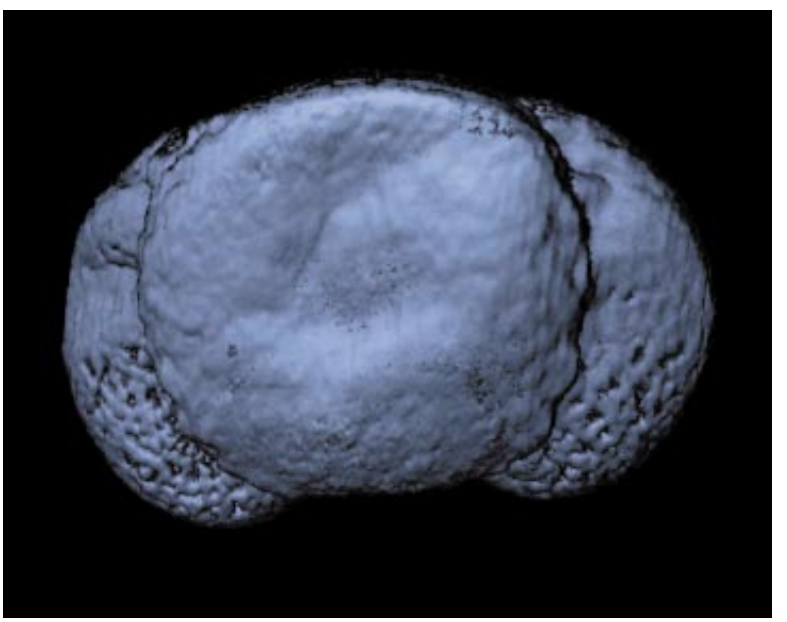
PORTRAITS OF POLLEN GRAINS from a sunflower (*top*) and a pine (*bottom*) were made by imaging successive planes in each fluorescently stained grain with a confocal microscope. A computer digitized the images, or optical sections, and combined them. Such digital reconstructions can be viewed in any orientation; the pine pollen is shown (*left to right*) from one side, from the opposite side, rotated 72 degrees relative to the first position, and from above.

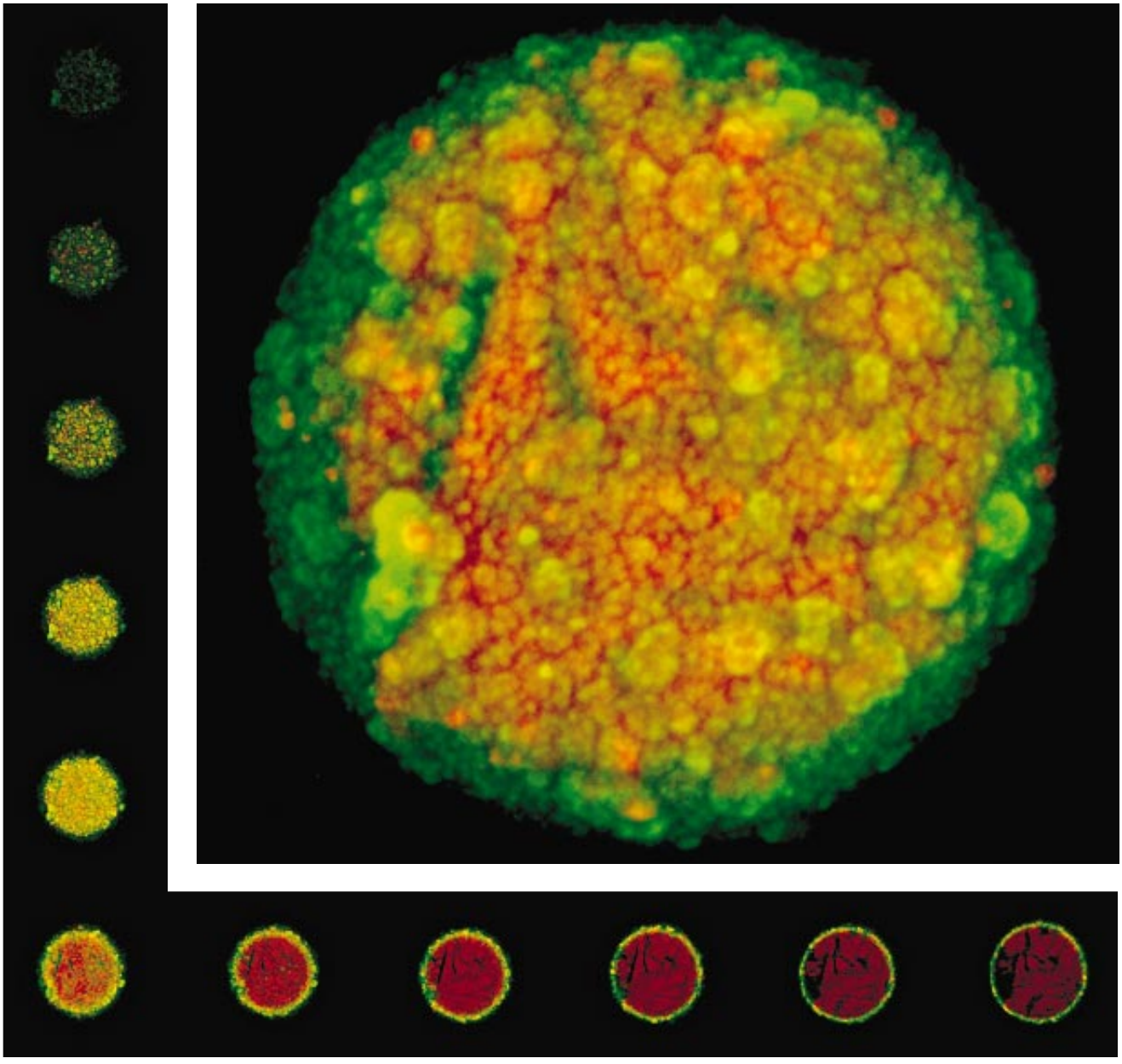












FLUID-FILLED POLYMER MICROCAPSULE (*large sphere*), about 0.1 millimeter in diameter, was rendered from a stack of optical sections that included the smaller spheres shown here. Matthew H. Chestnut of the Procter & Gamble Company made the images so as to compare the structural integrity of this capsule with that of others having a different composition. He distinguished the shell (*green*) from the fluid inside (*red*) by labeling those components with different dyes. Detailed analysis of many views revealed no obvious breaks in the shell but indicated some leakage was occurring.

developing brain, confocal microscopy is allowing them to see their subjects quite literally in a new light.

Minsky, who has been at the Massachusetts Institute of Technology for many years, developed the technique in the course of pondering how the human brain works. He reasoned that if the connections between all neurons in the brain could be mapped, the resulting circuit diagram should uncover clues to the brain's operation. Unfortunately, anyone who tries to apply conventional optical microscopy to identify the tiny interconnections between nerve cells in a slab of cerebral tissue immediately encounters a serious technical obstacle.

In standard microscopes the magnifying lens or system of lenses—commonly called the objective—both illuminates and provides a view of a specimen. As the objective focuses light on planes underneath the surface of brain tissue (or any thick, translucent material), the image rapidly becomes incomprehensible. Trying to view neural elements deep in such tissue is a lot like trying to see an object below the surface of a muddy pond by shining a flashlight into the water; the light is reflected by so many small particles that distinguishing the object from its environment is impossible.

To attain a perfectly sharp representation of a single plane in a specimen, one would ideally collect light that was reflected directly—and only—from the plane of interest. Material above and below that plane also returns light, however, giving rise to blurring, a bane of light microscopy. At the same time, a troublesome phenomenon called scatter can reduce contrast. Scatter arises when light hits minute particles and caroms off them into other particles before reaching a detecting surface. Signals produced by such randomly deflect-

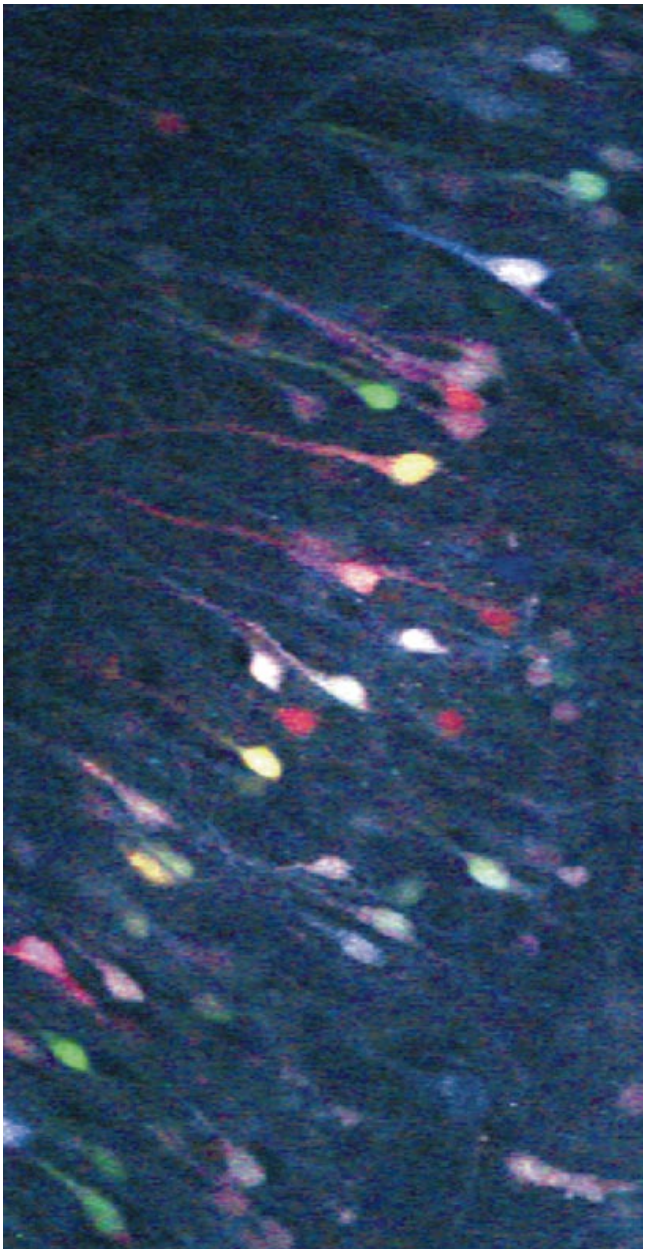
ed light convey no meaningful information; they create a diffuse glow that can swamp out light coming back from the plane of interest.

Minsky succeeded in minimizing blurring and enhancing contrast by making just a few modifications to the standard microscope. First, he prevented much of the scatter from occurring. He passed the illuminating light through an objective that focused the rays into an hourglass-shaped beam [*see illustration on next page*]. And he trained the "waist" of this beam—a sharp, intense point—on a single spot at some selected depth in the specimen. This procedure ensured that the spot would be the most intensely illuminated point in the specimen and would thus reflect the most light. Equally important, by focusing on one spot, Minsky guaranteed that no light would hit and be scattered by much of the remaining material in the specimen. In conventional microscopy, in contrast, the entire specimen would be lit and free to deflect illuminating rays.

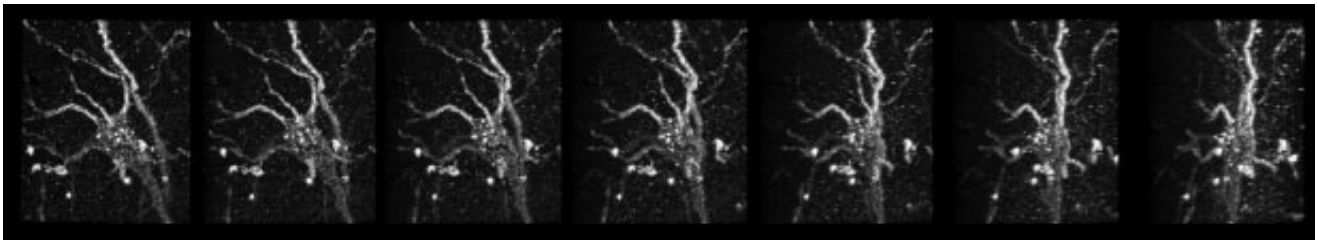
The strategy of focusing on a small area limited the total amount of scatter. But it did not prevent light from being returned and scattered by illuminated tissue lying above and below the spot of interest (within the hourglass shape of the beam). By means of a second crucial adjustment, Minsky was also able to keep much of this extraneous light from reaching the detecting surface. He knew that the objective focused light returning from the brightly illuminated spot onto a plane far above the specimen. He placed a mask containing a pinhole aperture in that plane, positioning the aperture so that the return light passed through the aperture to the detecting surface. The effect was dramatic: all of the signal from the brightly illuminated spot in the specimen reached the detector through the pinhole aperture; at the same time, the mask filtered out most of the light emanating from tissue outside the spot. In consequence, a nearly perfect image of the spot was formed, essentially undisturbed by scatter and by blurring from light in out-of-focus areas.

The obvious problem with Minsky's first two steps was that they yielded a sharp image of only a minute dot. To produce an equally impressive representation of an entire plane, the young inventor added a final feature: scanning. He moved the specimen bit by bit in a raster pattern across successive rows in a plane, so that eventually each point at some given depth visited the sharply focused illuminating beam and, in sequence, sent a clear signal through the pinhole aperture to the detector. He maneuvered the specimen with two electromagnetic tuning forks. One moved it across each row, and the other moved it from one row to the next in the plane.

In order to see the entire image of a plane, Minsky had light that passed through the pinhole strike a photomultiplier detector. This detector, in turn, generated a flow of electrical signals that yielded an image on a military-surplus, long-persistence radar screen. By lowering or raising the objective lens and repeating the scanning process, he could view another plane of the specimen on the rather large screen. The



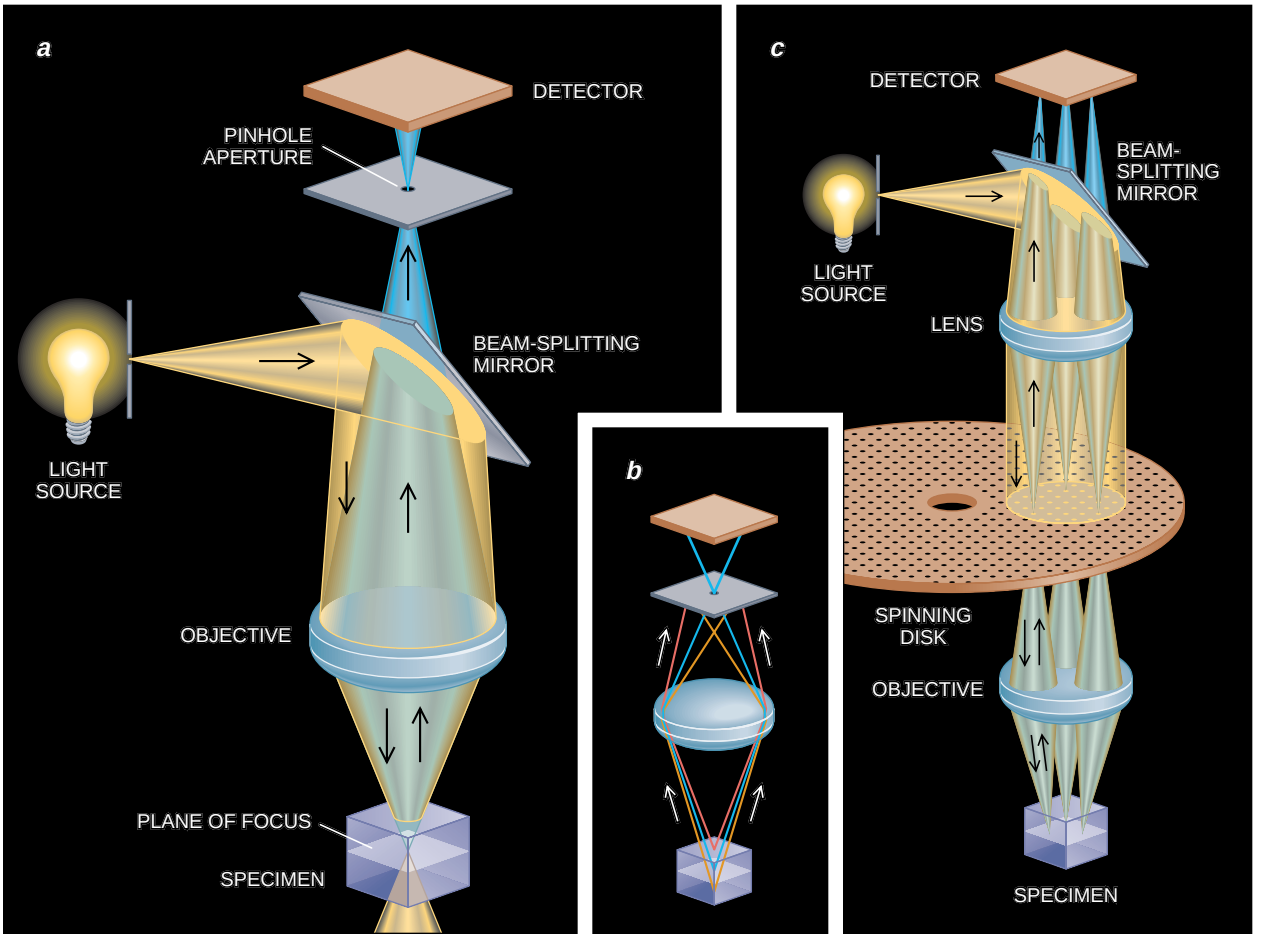
ACTIVE NEURONS (*colored bodies*) are highlighted in this slice of rodent brain tissue that was kept alive artificially. The picture is a computer-generated compilation of three confocal images made 12 seconds apart by Michael E. Dailey and Stephen J Smith of Stanford University. Each time point was coded by a single color—first red, then green and then blue. The image thus reveals that neurons fired at different times and that some of them were active during two of the shots (such as the yellow cell) or during all three (*white*).

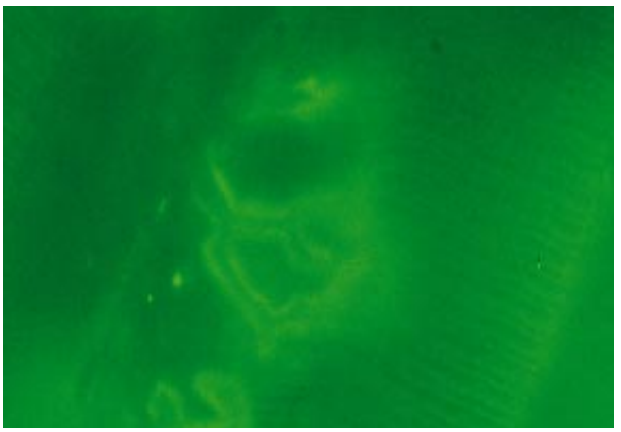


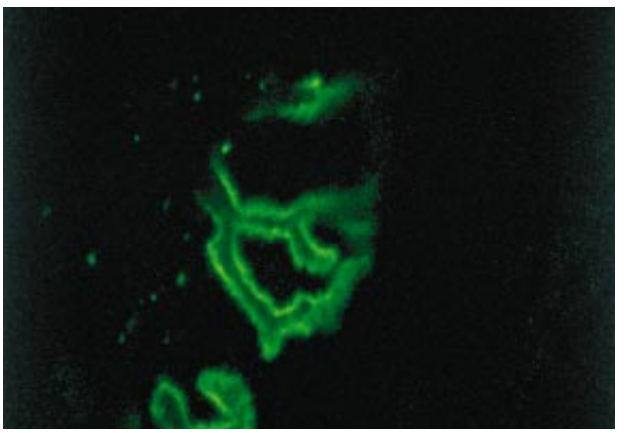
THREE-DIMENSIONAL RECONSTRUCTION of a neuron stars in this filmstrip; in each successive frame the structure is rotated some 10 degrees around the vertical axis. The strip yields a movie of the rotating cell. To view the neuron in three dimensions, cross your eyes as you look at a pair of images, focusing each eye on a different frame.

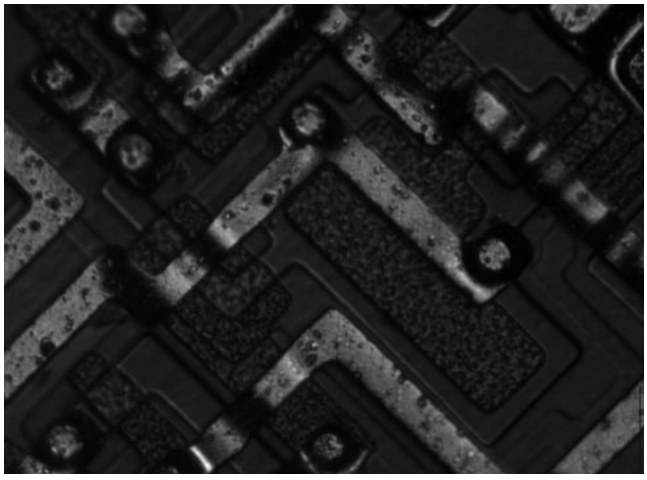
How Confocal Microscopy Works

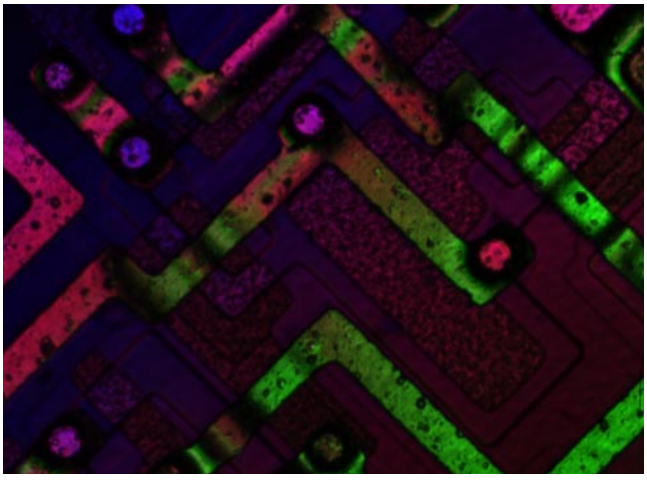
Confocal microscopes achieve high resolution of a selected plane in a specimen by means of three basic steps. First, light (*yellow in a*) is focused by an objective lens into an hourglass-shaped beam so that the bright "waist" of the beam strikes one spot at some chosen depth in a specimen. Next, light reflected from that spot (*blue*) is focused to a point and allowed to pass in its entirety through a pinhole aperture in a mask positioned in front of a detecting device. Meanwhile the opaque regions around the pinhole block out most of the rays that would tend to obscure the resulting image—namely, those reflected by illuminated parts of the specimen lying above (*red in b*) and below (*orange*) the plane of interest. Finally, the light is moved rapidly from point to point in the specimen until the entire plane has been scanned. The unusual clarity provided by the technique is evident in the micrographs at bottom, which were produced by a conventional (*left*) and a confocal (*right*) microscope. Both images depict the same mouse muscle, fluorescently labeled to highlight sites that are contacted by a motor neuron. To speed the scanning process, some instruments incorporate a disk containing hundreds of pinhole apertures through which light is sent and collected (*c*); the disk rotates to ensure that every spot in a plane will be visited.











TEXTURED SURFACE OF A COMPUTER CHIP appears in a standard light micrograph (*left*) and a composite confocal image (*right*). In the latter, scans made at three depths have been superimposed. The deepest layer is green; the topmost layer, red. The confocal image provides height information that cannot be gleaned from the conventional micrograph.

choice of a big screen may have been a tactical error. When Minsky asked friends and colleagues at Harvard to look at his invention, the observers often had difficulty interpreting what they were seeing. As Minsky later deduced, the display was just too spread out.

"I demonstrated the confocal microscope to many visitors, but they never seemed very much impressed with what they saw on that radar screen," says Minsky in a 1988 memoir. "Only later did I realize that it is not enough for an instrument merely to *have* a high resolving power; one must also make the image *look* sharp. Perhaps the human brain requires a certain degree of foveal compression in order to engage its foremost visual abilities. In any case, I should have used film—or at least installed a smaller screen!" He did neither, and perhaps for this reason confocal microscopy languished for decades.

In spite of the early disinterest, investigators and manufacturers have since devised many ways to combine the essential features of confocal microscopy—illumination of only a small region of a specimen, passing of the return light through an aperture aligned with the illuminated region, and scanning of the specimen. Few versions move the specimen anymore; in most devices the light beam travels. To increase the speed of image acquisition, some microscopes move the light beam with mirrors that pivot, forcing light that strikes them to flow swiftly across a specimen in a raster-scan pattern. These mirrors make it possible to reconstruct an image in less than a second. Such instruments require brighter light sources than were available to Minsky; after all, they have to produce a detectable signal from each spot almost instantly. Lasers, which are very intense and are easily focused to a fine point, are widely exploited for this purpose.

Another time-saving strategy deploys multiple spots of light to scan different regions of the specimen simultaneously, much as parallel computers carry out different operations at one time. Some of these devices incorporate spinning disks that contain many apertures through which both the illuminating and the return light pass. Other machines exploit slit-shaped apertures that shorten scanning time by illuminating lines rather than just spots. Fast-scanning strategies have made it possible to view complete planes of a specimen in real time, often directly through an eyepiece.

Most modern confocal microscopes benefit enormously from another revolutionary advance: development of computers that perform digital imaging processing. As a confocal microscope scans successive planes in a specimen, it produces a stack of images, each of which is an optical section; such sections are analogous to images of fine slices that have been physically cut from a specimen. Image-processing programs record not only the brightness of every spot in every section but also the spot's position in the specimen—its location in a single plane (its x and y coordinates) as well as its depth (its z coordinate). The locales defined by the triple coordinates are called voxels. They are the three-dimensional equivalent of pixels in a two-dimensional image.

Image-processing programs can compile voxels into three-dimensional reconstructions of microscopic objects. They can also manipulate voxels with ease, allowing reconstructed images to be rotated around an axis and viewed from any vantage. The advent of such technology has enabled scientists to readily make observations that would often have been highly expensive and time-consuming to achieve by other means. For instance, brain researchers have found computer-linked confocal microscopes to be extremely helpful for uncovering the detailed structure of the nervous system, and they are beginning to view living brain tissue with these devices.

From very modest beginnings, confocal microscopy has become an ultrasophisticated melding of lasers, optics, electromechanical scanning and computerized image processing. The result has given microscopists the ability to see into objects and to create fully dimensional images almost at will. Minsky's dream of mapping brain circuits with microscopy seems to be occurring after all.

FURTHER READING

AN EVALUATION OF CONFOCAL VERSUS CONVENTIONAL IMAGING OF BIOLOGICAL STRUCTURES BY FLUORESCENCE LIGHT MICROSCOPY. J. G. White, W. B. Amos and M. Fordham in *Journal of Cell Biology,* Vol. 105, No. 1, pages 41–48; July 1987.

MEMOIR ON INVENTING THE CONFOCAL SCANNING MICROSCOPE. M. Minsky in *Scanning,* Vol. 10, No. 4, pages 128–138; July/August 1988.

HIGH-RESOLUTION IMAGING OF SYNAPTIC STRUCTURE WITH A SIMPLE CONFOCAL MICROSCOPE. J. W. Lichtman, W. J. Sunderland and R. S. Wilkinson in *New Biologist,* Vol. 1, No. 1, pages 75–82; October 1989.

CONFOCAL MICROSCOPY. Edited by Tony Wilson. Academic Press, 1990.

SQUIDs

(for superconducting quantum interference devices) are the most sensitive detectors of magnetic fields. Their applications range from diagnosis of brain tumors to tests of relativity

by John Clarke

An array of detectors close to the head of a patient suffering from focal epilepsy picks up tiny magnetic-field fluctuations, pinpointing the location of the lesion in the brain responsible for the disorder. A five-ton aluminum bar suspended in a vacuum chamber at a temperature near absolute zero awaits the minuscule disturbance that would signify the arrival of a gravity wave from a supernova. A lonely instrument in Baja California records subtle variations in the magnetic field, helping geophysicists to locate a source of geothermal energy below the surface.

Each of these disparate measurements is made possible by an instrument called the SQUID, short for superconducting quantum interference device. The SQUID, which picks up changes in magnetic field, is the most sensitive detector of any kind available to scientists. Only inherent quantum effects set its limits. It has become by far the most widely used small-scale superconducting device. Although it is not a particularly new invention—it celebrated its 30th birthday last year—it has recently undergone a revolution in its accessibility. The advent of the high-temperature superconductors in the late 1980s has enabled SQUIDs to operate in liquid nitrogen, at a "warm" 77 kelvins (–196 degrees Celsius). As such, newer SQUIDs will be simpler to use and more widely applicable than those built from conventional superconductors, which function only at temperatures near absolute zero.

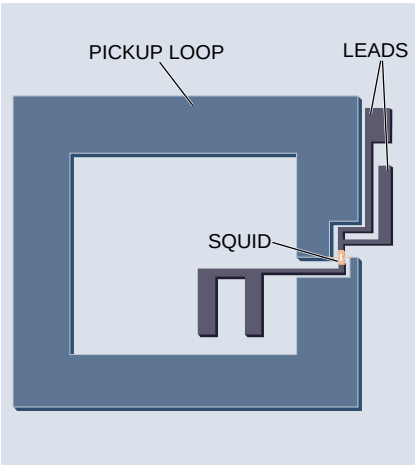
JOHN CLARKE is a professor at the University of California, Berkeley, and a faculty senior scientist at Lawrence Berkeley Laboratory. He earned his B.A. degree in 1964 and his M.A. and Ph.D. degrees in 1968, all from the University of Cambridge. Clarke is a fellow of the Royal Society, the American Physical Society and the American Association for the Advancement of Science. He has spent most of his career working on superconductivity, particularly the development of SQUIDs and their applications to a variety of topics.

The SQUID derives its phenomenal properties from a combination of several quantum-mechanical effects. The resistanceless flow of electric current is the most apparent. The Dutch physicist Heike Kamerlingh Onnes earned a Nobel Prize for his 1911 discovery that mercury became a superconductor when cooled by liquid helium to 4.2 kelvins. Subsequently, many metals—for example, tin, lead and niobium—and a vast number of alloys were shown to lose all resistance to current when cooled to low temperatures.

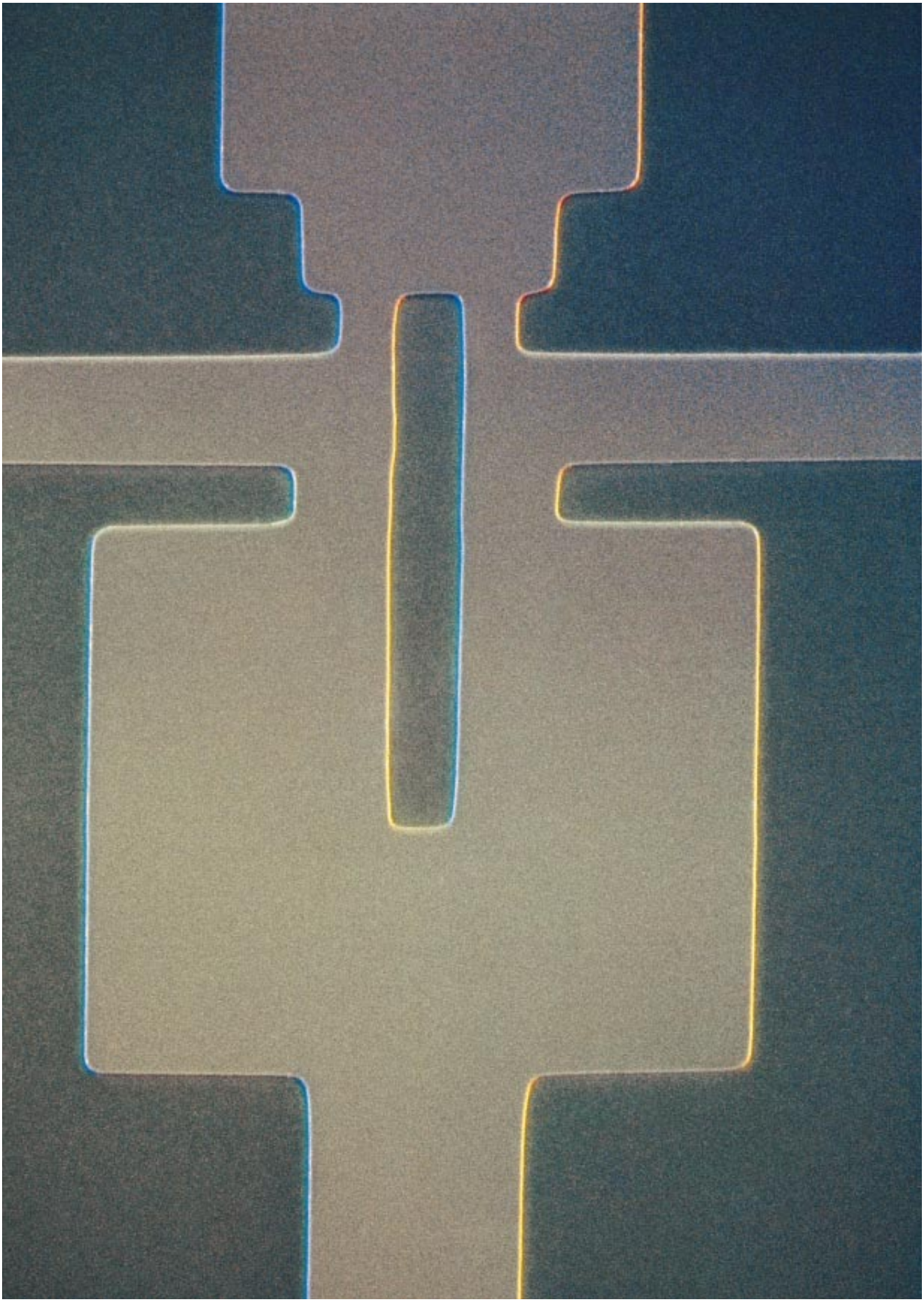
An explanation of why materials became superconducting was a long time coming. It waited until 1957, when John Bardeen, Leon N. Cooper and J. Robert Schrieffer published their seminal paper, reporting theoretical work that also earned a Nobel Prize. The central feature of their idea, called the BCS theory, is the Cooper pair: two electrons of opposite spin and momentum are bound together so that they have zero net spin and momentum. The attractive force behind this pairing is a subtle interaction between the negative charge of electrons and the positive charge of ion cores in the superconducting material. These ion cores are simply atoms that have lost one or more of their outermost electrons, which become free to conduct electricity. The ion cores are pulled in toward an electron as it moves through the lattice of a solid, creating a region of enhanced positive charge. This region attracts another, nearby electron. The effect is analogous to two baseballs on a water bed: if the indentations caused by the baseballs overlap, the baseballs become attracted to each other. The two electrons are weakly bound together, with an energy typically of one millielectron volt.

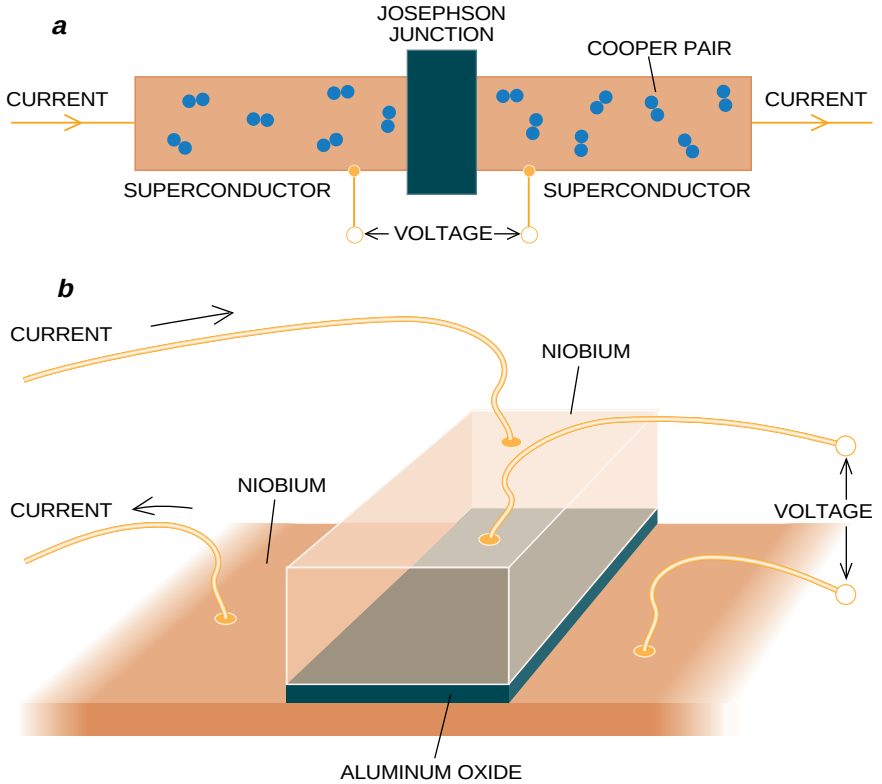
How do paired electrons move without resistance, whereas single electrons do not? In ordinary conductors, impurities, defects and, especially, lattice vibrations called phonons deflect the movement of single electrons. Such scattering of electrons endows the substance with resistance. The energy binding the electrons in a Cooper pair, though low, is high enough to prevent the pair from being separated by scattering. Hence, Cooper pairs propagate through the material without resistance. Deep cooling is essential because it quiets the lattice vibrations. At higher temperatures, the thermal energies become large enough to disrupt the Cooper pair.

A remarkable fact about a superconductor concerns its wave function. A wave function is a mathematical tool physicists use to represent particles in quantum systems. Like any wave, this function has both amplitude and phase—a simple example is a sine wave. It gives the probability for a given particle to be in a particular place at a particular time. What is curious about a su-



SQUID MAGNETOMETER is etched in a single layer of a high-temperature superconductor. The SQUID itself (*photograph*) is about 30 microns wide. The two grain boundaries that serve as Josephson junctions (*not visible*) lie just above the horizontal strip that runs across the image. The strip is coupled to a pickup loop, about eight millimeters wide (*diagram*).





JOSEPHSON JUNCTION consists of an insulating barrier separating two superconductors (*a*). Cooper pairs of electrons quantum-mechanically tunnel through the barrier. In a practical realization of the junction (*b*), an aluminum oxide film, which acts as the barrier, separates two layers of niobium. This "trilayer," grown on a silicon substrate in a vacuum chamber, is subsequently patterned to form individual Josephson junctions of typically a few microns across.

perconductor is that a single wave function can describe the entire collection of Cooper pairs. When no current flows, all the pairs have the same phase—that is, they are said to be phase coherent.

A third piece of Nobel-winning work at the heart of the SQUID comes from Brian D. Josephson, who predicted the effect that now bears his name. As a research student in 1962 at the University of Cambridge, Josephson considered two superconductors separated by a layer of an insulating material, which acts as a barrier to the flow of current [*see illustration above*]. The quantum-mechanical wave functions associated with the Cooper pairs leak into this "forbidden" region from each side. Provided the barrier is not too thick, the two wave functions will overlap. If this overlap is sufficiently large, the phases of the two wave functions "lock together." Under these conditions, Cooper pairs can "tunnel" through the barrier without breaking up. The junction hence acts as a weak superconductor. The critical current—the maximum supercurrent that can flow through the junction—depends on the size of the junction, the superconducting material and the temperature.

This phenomenon is described as the direct-current (dc) Josephson effect. Experiments conducted a few months later verified it. Philip W. Anderson and John M. Rowell, then at Bell Telephone Laboratories, made the observations. An alternating-current (ac) effect exists as well. Here a voltage maintained across the junction causes the amplitude of the supercurrent to oscillate in time.

Beyond their role in SQUIDs, Josephson junctions have many other applications. Because they can switch rapidly from the superconducting state to the resistive state—in just one or two picoseconds—they appear in experimental ultrafast digital components, including logic circuits, shift registers and analog-to-digital converters. Standards laboratories also use the Josephson junction to maintain the reference for the volt. Irradiating a junction with microwaves of a given frequency induces voltage steps. These steps occur at voltages that are precisely some integer multiple of that frequency.

Besides zero resistance and the Josephson effect, the SQUID exploits a third quantum-mechanical phenomenon: flux quantization. We are accustomed to thinking of quantization as something that happens on an atomic scale—for example, the occupation by electrons of discrete energy levels as they move around the nucleus. An analogous effect occurs in superconducting rings on a macroscopic scale. Suppose a current flows around the ring. The current produces a magnetic field threading through the ring. The product of the magnetic field and the area enclosed by the ring—the magnetic flux—cannot take on an arbitrary value. It must equal an integral number of a quantity called the flux quantum. A flux quantum is extremely small: a red blood corpuscle, about seven microns in diameter, in the earth's magnetic field (about 0.00005 tesla) embraces roughly one flux quantum.

A dc SQUID is rather simple. It consists of two Josephson junctions formed into a superconducting ring [*see box on opposite page*]. Applying current to the SQUID (biasing it) sends Cooper pairs of electrons tunneling through the junctions. A magnetic field applied to the ring, however, alters the flow. Specifically, it changes the quantum-mechanical phase difference across each of the two junctions. These phase changes, in turn, affect the critical current of the SQUID. A progressive increase or decrease in the magnetic field causes the critical current to oscillate between a maximum value and a minimum one. The maximum occurs when the flux administered to the SQUID equals an integral number of flux quanta through the ring; the minimum value corresponds to a half-integral number of quanta. (The flux applied to the SQUID can assume any value, unlike the flux contained within a closed superconducting ring, which must be an integral number.) In practice, we do not measure the current but rather the voltage across the SQUID, which also swings back and forth under a steadily changing magnetic field.

This quantum interference effect provides us with a digital magnetometer. Each "digit" represents one flux quantum. In fact, conventional electronics can detect voltages corresponding to changes in magnetic flux of much less than one flux quantum. The SQUID in essence is a flux-to-voltage transducer, converting a tiny change in magnetic flux into a voltage.

In my early days as a research student at Cambridge, my supervisor, Brian Pippard, proposed that I use a SQUID to make a highly sensitive voltmeter. In those days, procedures for making Josephson junctions were in their infancy and not practicable for manufacturing instruments. One day early in 1965, over

the traditional afternoon tea at the Cavendish Laboratory, I was discussing this problem with Paul C. Wraight, a fellow student. He suggested that a molten blob of solder (an alloy of lead and tin that becomes superconducting in liquid helium) deposited onto a niobium wire might just conceivably make a Josephson junction. His rationale was that niobium has a native oxide layer that might behave as a suitable tunnel barrier.

We rushed back to the laboratory, begged a few inches of niobium wire from a colleague, melted a blob of solder onto it, attached some leads and lowered it into liquid helium. As we hoped, Josephson tunneling! The fact that Wraight's idea worked the first time was important. If it had not, we would never have bothered to try again. Because of its appearance, we christened the device the SLUG. Later I was able to make a voltmeter that could measure 10 femtovolts (10^{-14} volt), an improvement over conventional semiconductor voltmeters by a factor of 100,000.

Needless to say, the technology of SQUID sensors has evolved beyond recognition in the intervening years. Most modern dc SQUIDs follow a design proposed by Mark B. Ketchen and Jeffrey M. Jaycox of the IBM Thomas J. Watson Research Center. They consist of multiple layers of thin films deposited on silicon wafers. The photolithographic and etching techniques of the semiconductor industry pattern these films. These methods can produce as many as 400 SQUIDs on a four-inch wafer. The wafer is then diced into individual chips, each bearing one SQUID. The SQUID itself consists of a square washer of niobium that has two Josephson tunnel junctions. The barriers consist of aluminum oxide, an electrical insulator, grown on top of one of the niobium layers [*see illustration on next page*].

Just how sensitive is such a SQUID? A convenient criterion is the energy associated with the smallest change in magnetic flux the device can detect in one second: typically about 10^{-32} joule. This incredibly tiny amount is roughly equal to the mechanical energy required to raise a single electron one millimeter in the earth's gravitational field. In fact, the best SQUIDs ever manufactured are 100 times more sensitive than that. They approach the limit set by Heisenberg's uncertainty principle, which sets fundamental boundaries on the accuracy of measurements.

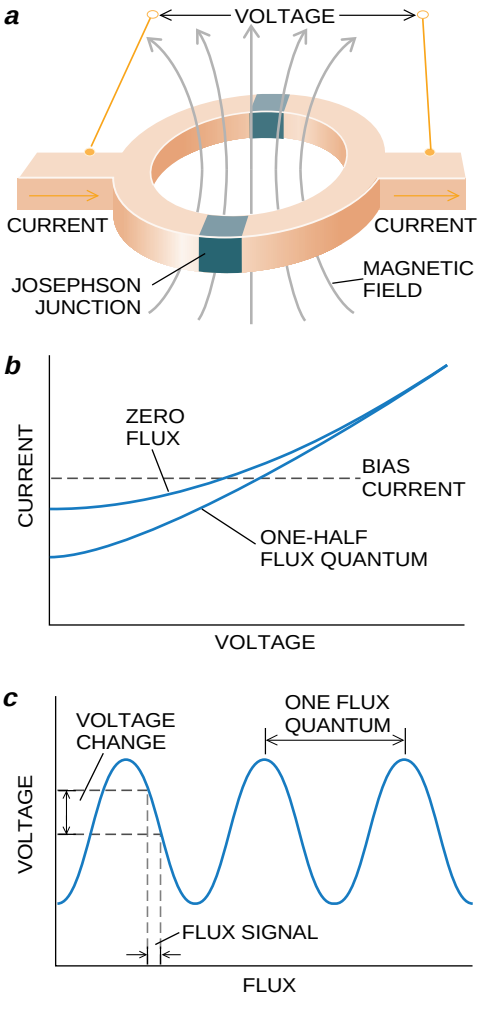
I should also mention that SQUIDs based on alternating current exist. This kind of instrument is known as the radio-frequency (rf) SQUID, because it is biased with a flux oscillating in the megahertz range. The device consists of a single Josephson junction in a superconducting loop, which is coupled to an inductor connected across a capacitor. This design forms a so-called resonant circuit, which is driven by an rf current. The amplitude of the rf voltage across this circuit oscillates in response to a magnetic flux.

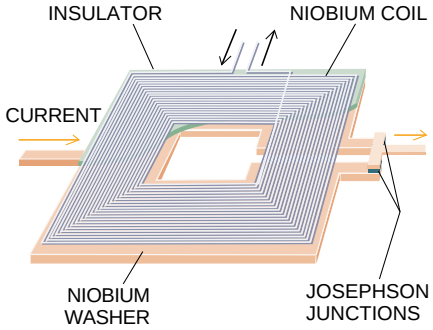
Manufacturers sold rf SQUIDs long before they did dc SQUIDs, even though the direct-current variety is generally more sensitive. Presumably, rf SQUIDs were easier to make, since each one requires only a single junction. Now advanced thin-film technology allows reliable production of large numbers of junctions. As a result, most SQUIDs sold today are of the direct-current type. Nevertheless, rf SQUIDs have not disappeared from the shelves, as many investigators find them quite adequate.

To take advantage of the extraordinary sensitivity of the SQUID, the devices are almost always coupled to an input circuit. For magnetometers, this circuit enhances the SQUID's sensitivity to magnetic fields, often by 100-fold. This so-called flux transformer simply consists of a loop of superconducting material coupled to a SQUID [*see illustration on page 51*]. The flux transformer boosts the field sensitivity because the loop encloses a much larger area than can a SQUID. An external magnetic field causes a persistent supercurrent to circulate in the loop. This current in-

The Direct-Current SQUID

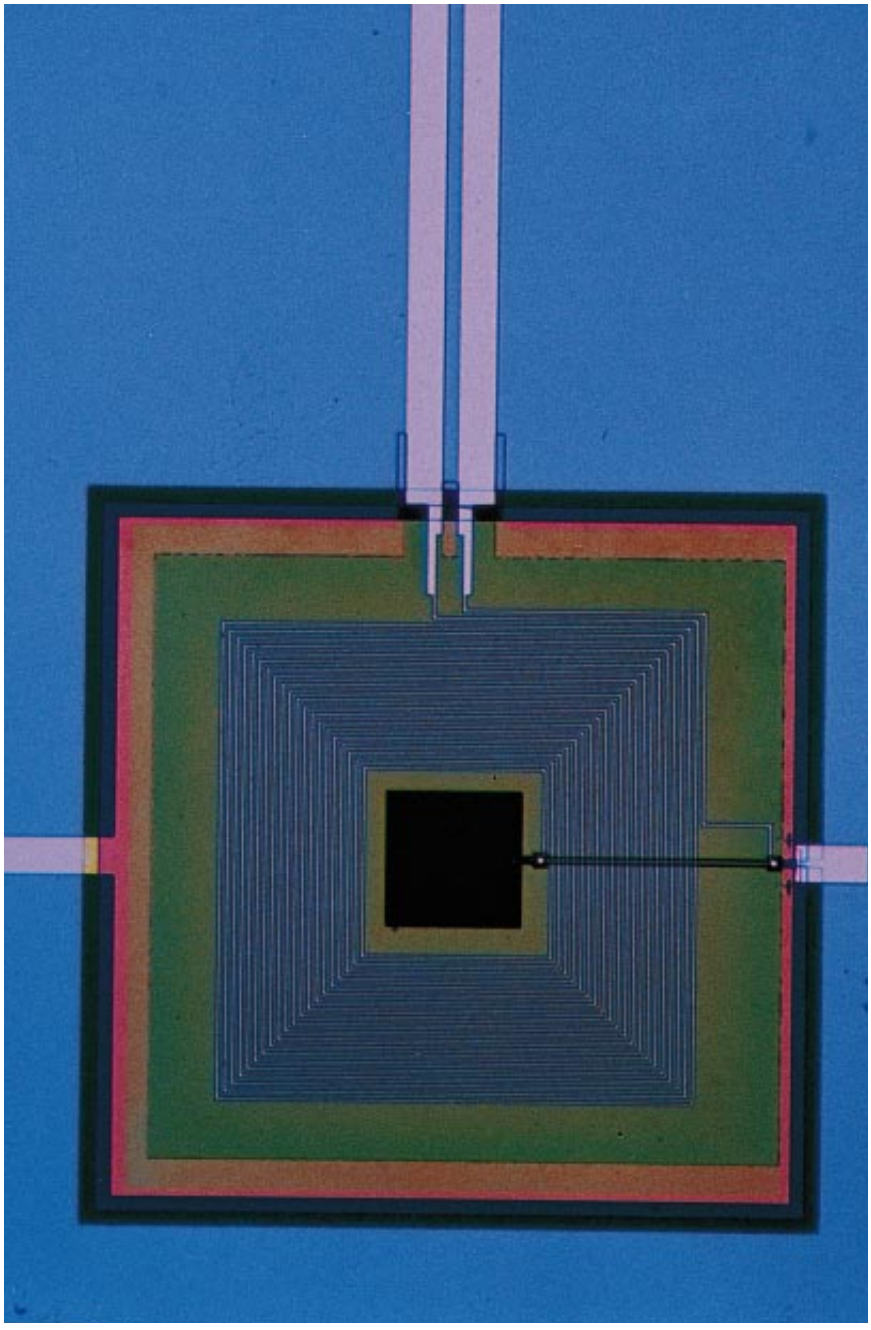
The direct-current (dc) SQUID consists of two Josephson junctions arranged on a superconducting ring (*a*). A current applied to the SQUID, called a bias current, divides between the junctions and, if greater than the critical current, produces a voltage across the SQUID. Plotting this current against the voltage yields characteristic curves (*b*). Steadily increasing the magnetic flux threading through the ring (for instance, by bringing in a small magnet) causes the critical current to decrease and then increase successively. The critical current is a maximum for zero flux (or an integer number of flux quanta) and a minimum for a half-integer number of flux quanta. The period of these oscillations is the flux quantum (*c*). This effect closely resembles the double-slit experiment in optics: when coherent light (such as that from a laser) passes through two parallel slits, the emerging beams "interfere" with each other to produce a series of light and dark fringes. In a superconductor a single wave function describes all the Cooper pairs. The wave functions at the two Josephson junctions interfere with each other to produce the current and voltage swings. In practice, we can detect changes that are smaller than the flux quantum. A tiny flux signal produces a corresponding voltage swing across the SQUID, which conventional electronics can measure.





duces a flux in the SQUID. With a flux transformer, a SQUID can reach femtotesla (10^{-15} tesla) resolution. One femtotesla corresponds to one part in 10^{11} of the earth's magnetic field.

Another variation of this SQUID magnetometer is the SQUID gradiometer. It measures the difference between magnetic field values at two different places, that is, the gradient. The gradiometer relies on two flux-transforming loops wound in opposite directions. An alternative approach employs two SQUID magnetometers separated by a short distance; electronic circuitry then subtracts one output from the other to determine how the field changes over this distance. This method is particularly appealing when applied to large arrays of magnetometers. In one variant, Roger H. Koch and his colleagues at the IBM Watson center developed the "three-SQUID gradiometer." In this device the output of one magnetometer channel cancels the ambient magnetic noise at two others, thereby creating a quiet environment for the gradiometer.

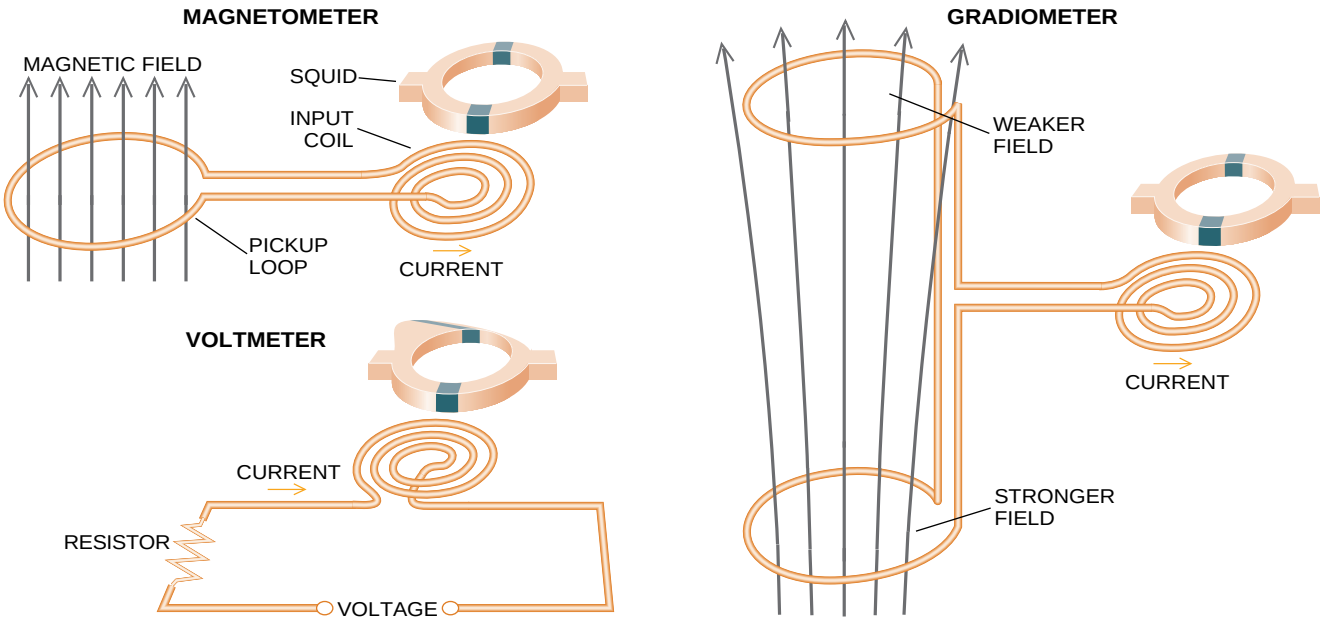


DIRECT-CURRENT SQUID made of conventional superconducting material typically operates in liquid helium. In its most widely used configuration, it consists of several layers of thin film (*diagram*). The body of the SQUID is a square washer of niobium. A Josephson junction is on each side of the slit at the right edge of the washer. A 20-turn niobium spiral coil is deposited on an insulating layer; the two leads at the top of the photograph provide external connections to this coil.

Measuring magnetic-field gradients is especially useful in medical diagnosis: the electric currents in the human body provide a rich source of time-varying magnetic signals. The tiny magnetic signals vary from a few femtoteslas from the brain to 50,000 femtoteslas from the heart. Until the development of SQUIDs, these signals were too weak to be studied. In addition, fluctuations in the earth's magnetic field and magnetic noise produced by the motion of elevators and automobiles and, particularly, by the 60-hertz hum from the electric grid system overwhelm the body's magnetic signals. The SQUID gradiometer attenuates background noise because the sources are usually far from the patient and tend to be nearly uniform. Hence, they evoke only a weak response in the SQUID gradiometer, which is sensitive to nonuniform fields.

In practice, an array of SQUID sensors maps the spatial variation in the magnetic fields produced by the body. From this contour map, a computer can reconstruct the region inside the body that produced the signals. This procedure is entirely noninvasive. During the past 20 years, the number of SQUIDs in the array has grown from about seven to more than 100; recently the Superconducting Sensor Laboratory in Inzai Chiba, Japan, announced a prototype system consisting of 250 channels.

Such instruments provide the physician with crucial information about various illnesses. For example, in patients who have focal epilepsy, a relatively localized electrical discharge in the brain triggers a seizure. Mapping magnetic-field spikes by a SQUID array can pinpoint the source of the discharge. When superposed on an image produced by magnetic resonance imaging, this source may correlate with some abnormality, such as scar tissue. Under favorable circumstances, the surgeon can excise it or destroy it with a "gamma knife"—collimated gamma rays. In a different approach, one evokes a magnetic response by means of a specific stimulus. For example, in San Diego, Eugene C. Hirsch-



SQUID-BASED INSTRUMENTS usually need auxiliary components. A magnetometer has a "flux transformer," which consists of a pickup loop connected to the SQUID's input coil. When a magnetic field is applied, a persistent current develops in the loop. The current flows to the input coil, producing a flux in the SQUID. In a gradiometer, two pickup loops are wound in opposite directions to measure the magnetic field at different places simultaneously. A flux develops in the SQUID only if the field is not the same at those points. In a SQUID-based voltmeter, a voltage at the input terminals produces a current equal to this voltage divided by the value of the resistor connected to the SQUID's input coil.

koff of Biomagnetic Technologies and Christopher C. Gallen of the Scripps Research Institute use a 74-channel SQUID system to map the response of the cortex surrounding a brain tumor to tactile stimulation.

Another medical application of widespread interest concerns the heart. Cardiac arrhythmia—erratic heartbeat—comes from spurious electrical pathways connecting the atria and the ventricles, short-circuiting the normal cardiac signals. In severe cases, the arrhythmia can prove fatal. To treat this ailment (with an electrical discharge from a catheter), one must localize this pathway, which can sometimes entail a prolonged search with one or more catheters. Several groups, including Gerhard Stroink and his colleagues at Dalhousie University in Halifax, and researchers at Siemens in Erlangen and at the University of Erlangen-Nürnberg have shown that SQUID imaging can localize the site of the electrical discharge. In this way, SQUIDs can substantially reduce the time required to find the anomaly.

Despite the impressive results of biomagnetic measurements, the high cost of multichannel machines has kept them from achieving general acceptance. Yet the technology has the potential to reduce health care costs dramatically. Locating the epileptic focus with SQUIDs may take three hours, whereas the alternative method of implanting electrodes on the surface of the brain may last as long as one week. The savings could easily reach $50,000. Similarly, the removal of a brain tumor without significant loss of neural function prevents trauma for the patient as well as saving the enormous expense of rehabilitation. The insurance industry has begun to accept the cost-savings possibilities of biomagnetic measurements; eight companies have reimbursed patients for the presurgical screening of brain tumors at Scripps. Blanket approval for the SQUID-based procedure may emerge by the end of this year.

SQUIDs also play an essential role in countless nonmedical applications, both in fundamental science and in routine measurements. A SQUID recently set an upper limit on the mass of the photon (if it has one at all: conventional theory dictates that it does not). The data indicate that the photon's mass must be less than about 10^{-46} gram. This limit is the strictest yet set in a cryogenic laboratory experiment.

Another fundamental application of a SQUID lies in the attempt to detect gravity waves. Relativity predicts that such waves—displacements in space and time—ought to be abundant in the universe. Collapsing stars, black holes and other movements of dense celestial bodies should ripple space in all directions. Researchers try to pick up these shifts with a giant metal bar. Weighing typically five tons and cooled by liquid helium, the bar would be sent into minute longitudinal oscillations by an incident gravity wave. A displacement in the end of the bar is detected as a flux change in a SQUID at a resolution of 10^{-18} meter—roughly 0.001 the diameter of an atomic nucleus. Indeed, several gravity-wave antennae are deployed worldwide. To date, the antennae have not recorded gravity waves. But the next generation coming on-line, with perhaps two orders of magnitude greater sensitivity, will most likely do so.

Probably the most widespread SQUID-based system is a sophisticated "susceptometer." Manufactured by Quantum Design in San Diego, the instrument enables scientists to measure the magnetic properties of samples from a few kelvins to well above room temperature. Hundreds of such machines figure in the workplaces of physicists, chemists, materials scientists and biologists.

Although SQUIDs have proved their utility for many years, interest in them has recently exploded, thanks to J. Georg Bednorz and K. Alexander Müller of the IBM Research Laboratory in Zurich. These researchers discovered the so-called high-temperature superconductors (which promptly earned them the fourth Nobel Prize in the field). Other workers quickly pushed the transition temperature to over 100 kelvins (–173 degrees C).

The popular media heralded this breakthrough as the greatest scientific

revolution since the lightbulb (or, perhaps, merely the transistor). The hype promised an almost instant multibillion-dollar business. Of course, as with all scientific innovations, progress takes time. Today the only devices commercially available that rely on this breakthrough are SQUIDs.

Unlike most conventional superconductors, the high-temperature materials are ceramics organized in complex layers. For devices such as the SQUID at least, the most common material is an oxide of yttrium, barium and copper. It bears the formula $YBa_2Cu_3O_{7-x}$, where x equals approximately 0.15. The substance is known as YBCO (often pronounced "ibco"). Its transition temperature is about 90 kelvins, so that it becomes a superconductor at the temperature of liquid nitrogen.

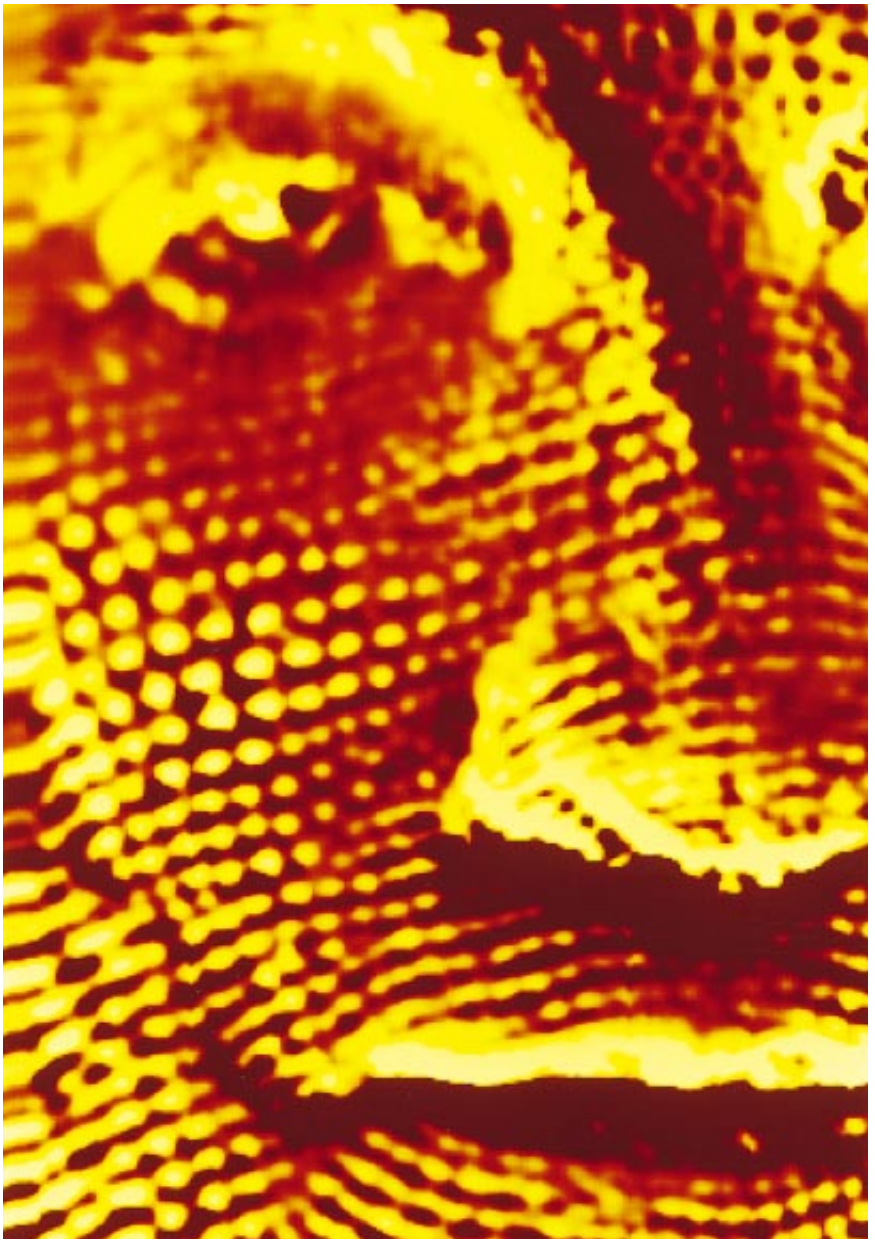
Because they are ceramics, the new superconductors are brittle and difficult to work. Workers cannot readily bend the wires into shape as they can with the relatively ductile low-temperature superconductors. Researchers have devised various schemes to make high-quality thin films out of them. Of those approaches, deposition by a high-power pulsed excimer laser has proved very useful, not least because the process is relatively rapid. The laser pulses reside in the ultraviolet part of the light spectrum, with a typical wavelength of 248 nanometers. They strike a rotating YBCO target mounted in a chamber containing oxygen. Each pulse vaporizes a small amount of YBCO, forming a beautiful rose-colored plume. The material collects on a nearby substrate kept at about 800 degrees C. The film grows epitaxially—in other words, its crystal structure mimics that of the substrate—with the correct chemical makeup.

Besides forming them into thin films, SQUID makers also need to form Josephson junctions on the wafer. Investigators have invented many clever processes to form them in high-temperature compounds. Duane Dimos and his co-workers at the IBM Watson center have developed an especially successful method. One starts with a crystal, usually of strontium titanate, that has been cut and fused together to produce a deliberate misorientation of the crystal axes along a line. This dislocation is called a grain boundary. When the YBCO film is grown epitaxially on the substrate, it replicates the abrupt change in crystal orientation. The grain boundary greatly diminishes YBCO's supercurrent-carrying capability, hence behaving as a Josephson junction. Another successful way of making junctions calls for an epitaxial sandwich; workers interpose a thin slice of normal material between two superconductors.

As with their low-temperature counterparts, high-temperature SQUIDs rely on a flux transformer to improve sensitivity to magnetic fields. A simple means of accomplishing this task is to form the flux transformer in the same YBCO layer as the SQUID. A second flux transformer can further enhance the sensitivity. With the latter of these designs, Dieter Koelle and his co-workers at the University of California at Berkeley, Lawrence Berkeley Laboratory and Conductus, Inc., in Sunnyvale, Calif., achieved a noise level of about 30 femtoteslas. Michael Mück and his colleagues at the Jülich Research Center in Germany achieved 24 femtoteslas using a similar transformer coupled to an rf SQUID.

It should be noted that SQUIDs operating in liquid nitrogen will never achieve as good a resolution as can their counterparts working in liquid helium. So why are the high-temperature devices creating such a stir? The popular view was expressed many years ago by Nobelist Ivar Giaever, now at the Rensselaer Polytechnic Institute: liquid helium is as expensive as Scotch, whereas liquid nitrogen is as cheap as milk.

Although this statement is more or less true, the heart of the matter is that liquid nitrogen vaporizes much more slowly than does liquid helium. Instead



SQUID IMAGE of George Washington was made by scanning a high-temperature SQUID over a one-dollar bill. As it scans, the SQUID detects the variations in the magnetic field produced by the ink particles on the bill.

of having to refill a dewar of liquid helium every few days, one replenishes the supply of liquid nitrogen every few weeks. Furthermore, liquid helium is available only in major metropolitan areas; in more remote regions of the world it is an exotic substance that cannot readily be procured every few days. Liquid-nitrogen cooling thus offers a superconducting technology that would otherwise be denied.

The growing promise of high-temperature SQUIDs is especially apparent in geophysics research, which is often conducted in inaccessible locations. One such application is in the study of magnetotellurics ("magnetism of the earth"). Its practitioners measure the electrical resistivity of the ground below the surface to infer the underlying structure. In essence, very low frequency (0.001 to 100 hertz) electromagnetic waves from the upper atmosphere propagate down to the earth's surface. The solar wind blowing on the charged layers in the magnetosphere or ionosphere generates these waves, which are reflected by the ground. But a component of these waves decays into it. By measuring the tiny fluctuating magnetic fields (using magnetometers) and electric fields (using buried electrodes), the geophysicist can map out the resistivity of the earth's crust at depths of as much as several tens of kilometers. From this information the researcher infers the subsurface hydrology and porosity, seeking valuable clues on the whereabouts of oil or geothermal energy sources. In complementary techniques the geophysicist supplies magnetic pulses and measures the response. Pulsed methods are widely applied down exploratory boreholes to search for oil and may be able to locate buried hazardous wastes.

The ease granted by liquid nitrogen will likely encourage the proliferation of SQUID systems in medicine. Already, several research groups have used high-temperature SQUIDs to obtain magnetocardiograms and even magnetoencephalograms. An intriguing extension of the technology may aid fetal cardiology. Heart rate variability is one way physicians assess the health of the fetus, and electrocardiograms can be taken until the seventh or eighth month of pregnancy. Beyond that period, the signal declines because the fetus becomes electrically insulated from the mother. A magnetocardiogram, however, remains unaffected. Moreover, the superior spatial resolution of magnetic measurements makes distinguishing the fetal signal from the maternal one easier.

Workers have begun to obtain magnetic images by sweeping SQUIDs over an object. Such a "scanning SQUID microscope" can resolve objects down to a few microns wide. This technique has applications in studying not only magnetic materials but also metals and electronic circuits, which produce magnetic fields when a current flows. Scientists are also exploring the use of SQUIDs in nondestructive testing—say, in looking for corrosion of aluminum sheets riveted together in aircraft. The SQUID measures the influence of the aircraft skin on an applied oscillating magnetic field; a change in electrical conductivity reveals the defects.

At this juncture, the most sophisticated high-temperature SQUID commercially available is "iMAG," a version of the single-layer YBCO magnetometer. Made by Conductus, a complete system with electronics, dewar and probe sells for about $10,000. Highly versatile, this instrument is suitable for laboratory experiments as well as for geophysics and magnetocardiology studies.

To be sure, such SQUIDs will by no means replace their low-temperature cousins in all situations. Gravity-wave detectors and other applications with the most demanding requirements will continue to call for liquid helium. The liquid nitrogen–based devices will, however, open up a range of possibilities that were not previously realistic, bringing this wonderfully sensitive technology out of the laboratory and research hospital and into the marketplace.

FOCAL EPILEPSY can arise from a localized neural defect that produces magnetic signals. An array of SQUIDs measures the signals, and the inferred position of the epileptic focus (*yellow areas*) is superposed on a magnetic resonance image.

FURTHER READING

SQUID MAGNETOMETERS FOR LOW-FREQUENCY APPLICATIONS. T. Ryhänen, H. Seppä, R. Ilmoniemi and J. Knuutila in *Journal of Low Temperature Physics,* Vol. 76, Nos. 5–6, pages 287–386; September 1989.

PRINCIPLES AND APPLICATIONS OF SUPERCONDUCTING QUANTUM INTERFERENCE DEVICES. Edited by A. Barone. World Scientific Publishing, 1992.

CARDIOMAGNETIC IMAGING. Gerhard Stroink in *Frontiers in Cardiovascular Imaging.* Edited by B. L. Zaret, L. Kaufman, A. S. Berson and R. A. Dunn. Raven Press, 1993.

MAPPING THE BRAIN WITH MSI. Christopher C. Gallen and Floyd E. Bloom in *Current Biology,* Vol. 3, No. 8, pages 522–524; August 1993.

SQUIDS: THEORY AND PRACTICE. John Clarke in *The New Superconducting Electronics.* Edited by H. Weinstock and R. W. Ralston. Kluwer Academic Publishers, 1993.

How Cells Process Antigens

Cells alert the immune system to the presence of infections by displaying molecular complexes made from bits of their own proteins and those of invading organisms

by Victor H. Engelhard

All multicellular organisms have some relatively primitive defense system that distinguishes foreign pathogens from the body of the host and eliminates them. In addition, higher vertebrates have evolved a more advanced immune system that can discriminate among pathogens and make selective responses to each. The advantage of such specific immunity is that the immune system can rapidly adapt to those pathogens that are most often encountered in the local environment.

At the biomolecular level, the immune system's surveillance of the vertebrate body depends on a search for antigens—immunologic target molecules—that signal the presence of an invader. Antigens are not merely pieces of a pathogen. Rather they are often molecules constructed by the host's cells from bits of the pathogen's proteins and from cellular proteins called major histocompatibility complex (MHC) molecules. The processing and assembly of antigens are the key to the flexibility, the specificity and the thoroughness of all immune responses.

The creation of antigens and their presentation on cell surfaces, where the immune system can inspect them, constitute a complex phenomenon. Researchers already understand many of the steps in some detail. Interestingly enough, the processing of antigens is intimately tied to the mechanisms that synthesize and recycle all proteins inside cells and transport them between intracellular compartments. Greater understanding of antigen processing will therefore clarify what is occurring at a molecular level inside both diseased and healthy cells. The benefits of this work could eventually take the form of better treatments for a spectrum of illnesses, from infections to cancer.

VICTOR H. ENGELHARD is professor of microbiology at the Beirne Carter Center for Immunological Research of the University of Virginia Health Sciences Center at Charlottesville. In 1973 he graduated from Rice University with a baccalaureate in biochemistry, then continued his studies in that field at the University of Illinois, where he received master's and doctoral degrees. During postdoctoral work at Harvard University, he concentrated on the immunologic problem of antigen processing. In 1993 he and his colleagues discovered the alternate pathway for class I antigen processing.

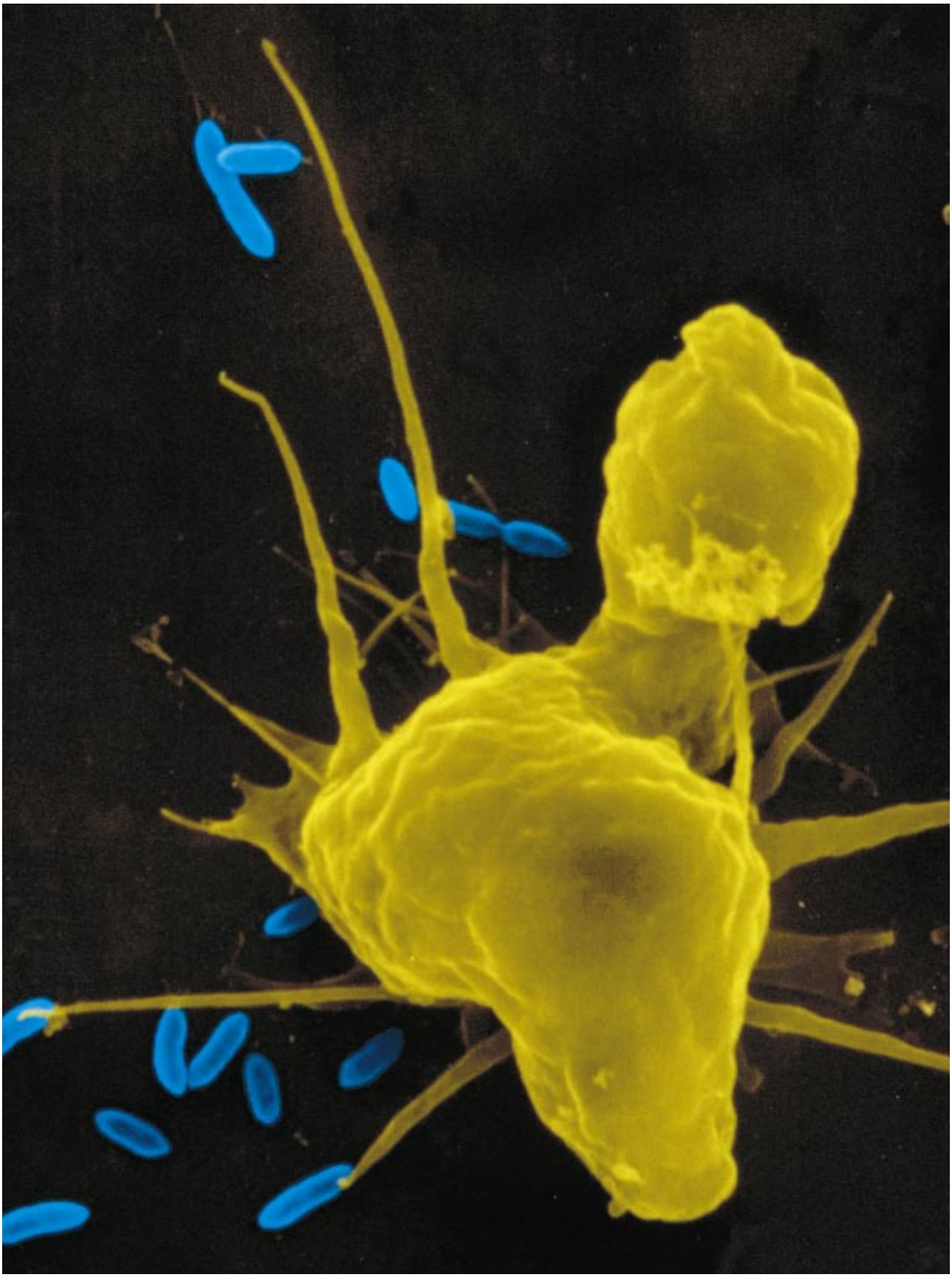
Before continuing with this discussion of antigen processing, it may be helpful to review briefly some aspects of how the immune system works. To make specific responses, the immune system employs a large population of white blood cells called lymphocytes. Lymphocytes have surface receptors that bind with high affinity to antigens. Each lymphocyte expresses receptors with slightly different structures; consequently, each lymphocyte is specific for just one type of antigen. Immunologists estimate that in a typical human being, the lymphocyte population expresses more than 10^8 distinct antigen receptors. This repertoire enables the immune system to respond with exquisite specificity to virtually any foreign antigen it encounters.

The immune system tailors the quality of its response to the nature of the pathogen and the way in which it invades the body. Many bacteria and large parasites such as worms establish infections in the extracellular spaces of the body, such as the bloodstream or intestinal lumen. To control these organisms, the immune system deploys soluble antigen receptors called antibodies, which are made by *B* lymphocytes. Antibodies bind directly to a parasite and provide a focus for the destructive action of other immunologic molecules and cells.

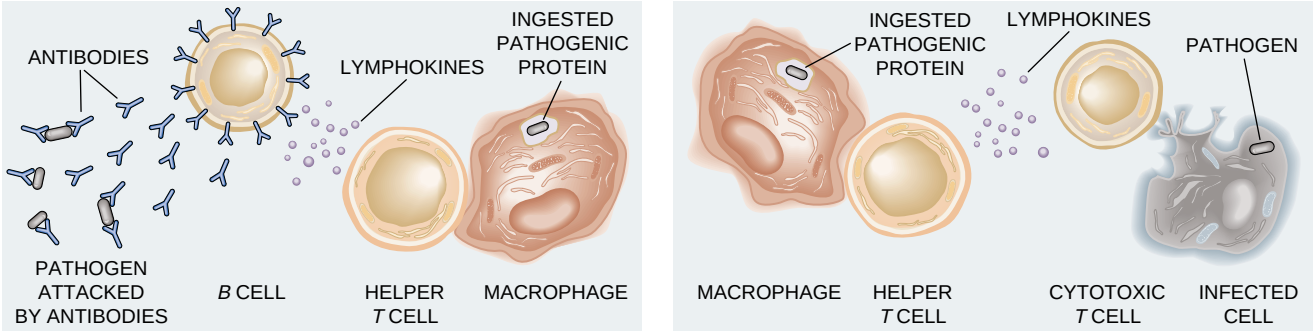
Viruses and many other bacteria and protozoan parasites, such as those that cause malaria, sleeping sickness and leishmaniasis, are not so easily thwarted. They establish their infections inside the host's cells, where antibodies cannot reach them. To rebuff these organisms, another arm of the immune system comes into play. The host's cells carry MHC molecules on their surface. In infected cells, these MHC molecules bind to and display small peptides, or fragments of proteins, that come from the parasite. The complexes of parasite peptides and host MHC molecules form the antigens that can be recognized by antigen receptors on cytotoxic (killer) *T* lymphocytes. In this way, the *T* lymphocytes can identify and kill infected cells selectively, sparing healthy cells. One function of MHC-peptide complexes, therefore, is as a signal that a cell is infected.

MHC-peptide complexes are also important in the regulation of immune responses. Some specialized cells, such as the white cells called macrophages, roam the body, ingesting extracellular materials they find, degrading them to produce peptides and presenting the peptides as antigens. These antigen-presenting cells travel from sites of infection to the lymph nodes, where they recruit lymphocytes for the immune response; in effect, the antigen-presenting cells are like messengers from the front lines of battle. When helper *T* lymphocytes recognize an MHC-peptide complex on these antigen-presenting cells, they secrete hormonelike molecules (lymphokines) that promote the differentiation of immune system cells.

Thus, the recognition of an MHC-peptide complex on the surface of a cell is a critical event in the initiation of all immune responses and, particularly, in the effective elimination of intracellular parasites. During the past 20 years, immunologists around the world have sought to discover how the complex between an MHC molecule and a peptide forms. These studies have led to an understanding of how the structure of MHC molecules enables them to bind to many different peptides from the extraordinary variety of infectious agents that an organism encounters.



MACROPHAGE INGESTS BACTERIA as part of the immune response to infection. Inside this white blood cell, bacterial proteins are degraded into peptides and presented as antigens by specialized molecules on the cell's surface. The ways in which cells turn their own and foreign proteins into antigens have gradually been revealed in recent decades.



RECOGNITION OF ANTIGENS regulates immune responses. When helper *T* cells recognize an antigen complex on a macrophage or other specialized presenting cell, the *T* cells release chemical signals called lymphokines. Those signals can direct *B* cells to release antibodies against extracellular bacteria (*left*) or cytotoxic *T* lymphocytes to attack cells infected with viruses (*right*). Antigen complexes also allow the cytotoxic *T* cells to identify their targets.

MHC molecules were originally identified by investigators working on tissue transplantation. In the 1930s George D. Snell of Jackson Laboratories in Bar Harbor, Me., and Peter A. Gorer of the Lister Institute of Preventive Medicine in Middlesex, England, described a locus, or genetic position, on chromosome 17 of mice that was the primary determinant of whether tissues transplanted from one strain of mouse to another would be accepted or rejected. They designated this locus H-2. (The H stands for "histocompatibility," a term meaning "tissue acceptance.") A similar genetic locus in humans was defined in the 1950s by Jean Dausset of the University of Paris and others.

Further work in many laboratories showed that H-2 contained many genes for transplantation antigens—proteins that are expressed on the surface of cells and can be recognized by the immune system. The name "major histocompatibility complex" was coined to reflect the importance of this closely linked group of genes in graft rejection and acceptance. (Historically, the human versions of these molecules have often been called human leukocyte antigens, or HLAs, but MHC is now accepted as a general label.)

Structural studies showed that the transplantation antigens encoded by the MHC genes are of two fundamental types, designated class I and class II molecules. Each of these MHC classes is highly diverse: mouse and human populations contain more than 100 forms of the molecules, although individuals typically express only between three and six of each class.

The physiological function of MHC molecules became apparent only in the late 1960s. Working independently, Baruj Benacerraf of New York University and Hugh O. McDevitt (initially working in Israel and later at Harvard University) found that some individual guinea pigs and mice could make antibodies against certain simple protein antigens but that other individuals could not. Using many inbred strains of mice, McDevitt showed that "responsiveness" and "unresponsiveness" were genetically determined traits that depended on the particular type of class II MHC molecules expressed by the mice.

Similarly, in 1974 Rolf Zinkernagel and Peter Doherty of the John Curtin School of Medical Research in Canberra discovered that some inbred strains of mice died if they were infected intracerebrally with the virus for lymphocytic choriomeningitis, whereas others lived. In response to the virus, the casualties were producing cytotoxic *T* lymphocytes that attacked the infected nervous system. (Ironically, *T* lymphocytes usually protect the body from harm, but in this case they were engaging in a deadly autoimmune reaction.) As the researchers showed, the ability to produce those *T* lymphocytes was linked to the expression of a particular set of class I MHC molecules by the mice.

Zinkernagel and Doherty then made a critical advance by demonstrating that *T* lymphocytes isolated from one mouse could recognize virally infected cells from a second mouse—but only if both mice expressed the same class I molecules. In short, the immune response could occur only if both the right antigen and the right MHC molecule were present. This dual requirement for a foreign antigen and an appropriate MHC molecule was termed MHC-restricted recognition of antigen.

Several groups, most notably those of Alan S. Rosenthal of the National Institutes of Health and David H. Katz of Harvard, demonstrated that MHC-restricted antigen recognition also explained McDevitt's observations of immune responsiveness. *B* cells did not produce antibodies against McDevitt's simple protein antigens unless they were stimulated by the helper *T* lymphocytes. Those *T* lymphocytes recognized only the antigen-presenting cells that had been exposed to the protein antigen and that offered the appropriate class II MHC molecules.

For the next 10 years, many groups of researchers attempted to understand how *T* cells recognized both the antigen and the MHC molecules. Separate studies conducted in the laboratories of Emil R. Unanue, first at Harvard and later at Washington University, and of Howard M. Grey of the National Jewish Center for Immunology and Respiratory Medicine in Denver provided a seminal breakthrough. They discovered that to stimulate an immune response, extracellular proteins must first be endocytosed (ingested) and broken into peptides by an antigen-presenting cell. These peptides then bind to class II MHC molecules and appear on the cell surface as a complex that helper *T* cells can recognize. This sequence of events—the ingestion of antigens, their fragmentation into peptides and their binding to MHC molecules—was called antigen processing.

Class I MHC molecules are also involved in antigen processing. As Alain R. M. Townsend of John Radcliffe Hospital in Oxford, England, learned, cytotoxic *T* lymphocytes identify virally infected cells by looking for viral peptides presented by a class I MHC molecule. Further work done in the laboratories of Thomas J. Braciale of Washington University and Michael J. Bevan of the Scripps Research Institute established that all the peptides naturally presented by class I MHC molecules are derived from proteins in a cell's cytoplasm.

All these results and others indicate that the two kinds of MHC molecules sample antigens that are processed in different intracellular compartments. The peptides associated with class I

MHC molecules have invariably been found to originate from a cell's own proteins. The source proteins for class II–associated peptides are sometimes found to be in the medium in which a cell grows. More frequently, however, those peptides come from proteins located on the outer membrane.

An important implication of those observations is that most of the MHC molecules on a cell are presenting peptides from normal cellular proteins and not ones from pathogens. Even when a cell has ingested a foreign antigen or been infected, the number of MHC molecules presenting those alien peptides is only a small fraction of the total.

The ability of MHC molecules to bind to specific peptides and to participate in antigen processing is a consequence of their structure and synthesis. Both classes of MHC molecules are composed of two protein subunits. The class I molecules each consist of one heavy protein chain and a much smaller light chain called ß2m. The two chains in the class II molecule are of roughly the same size and are smaller than the class I heavy chain.

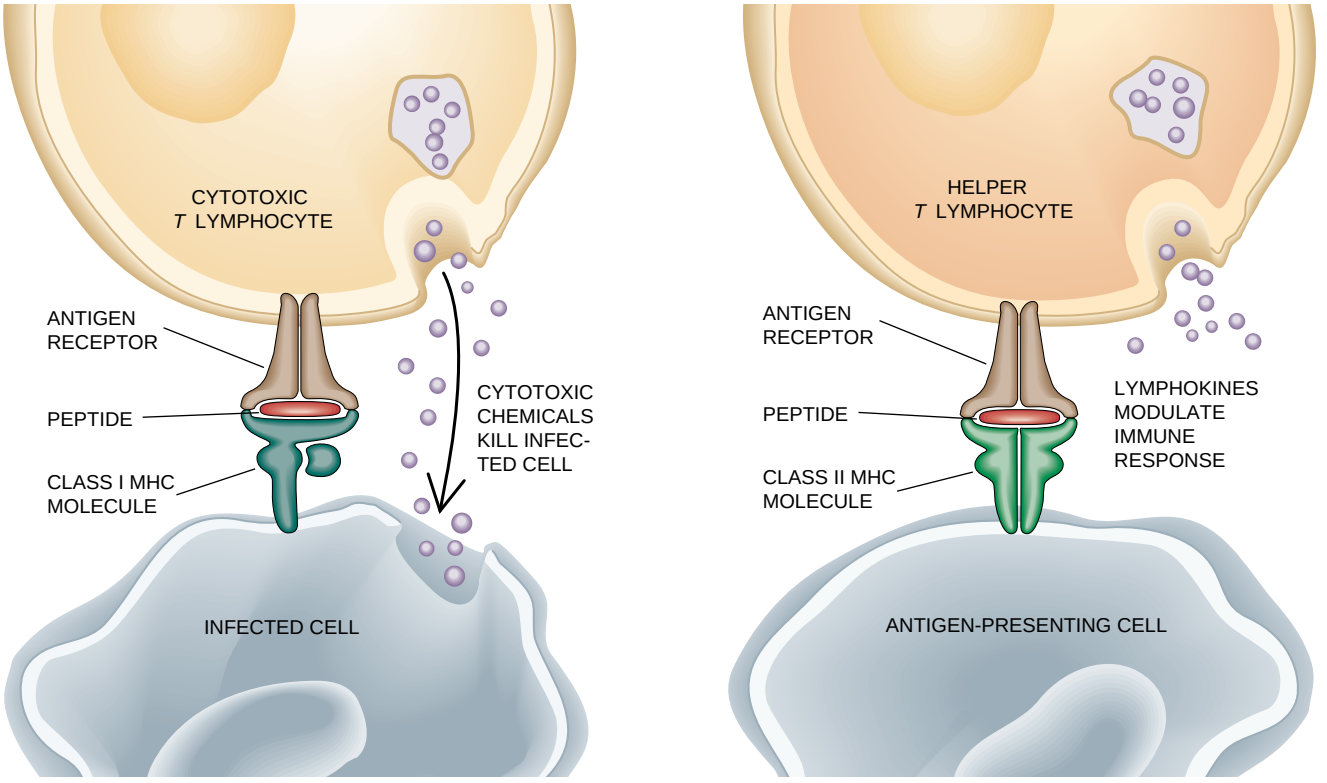
Notwithstanding those differences, x-ray crystallographic analyses by Don C. Wiley of Harvard and his colleagues have revealed that class I and II MHC molecules have strikingly similar structures. Both types feature a deep cleft in their top surface into which peptides bind. The structure of the cleft itself is complex and contains several pockets that can interact with different parts of a peptide. Differences in the shapes and properties of these pockets endow the various forms of MHC molecules with their selective affinity for certain peptides.

Many molecular immunologists and protein chemists are still trying to understand more precisely what determines those affinities. Crystallographic studies have given some clues about the binding between peptides and the MHC molecules. Taking an alternative approach, immunochemists have also looked for features common to all the peptides that bind to one form of MHC molecule, although the structural complexity of peptides makes that work difficult.

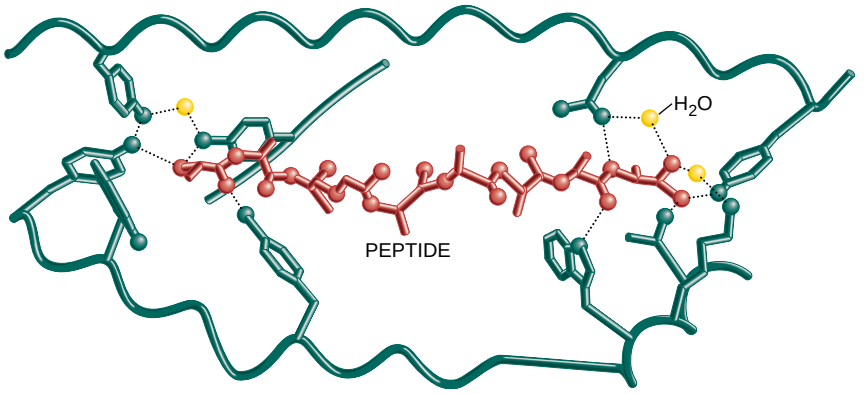
Tandem mass spectrometry has greatly advanced the effort. In this technique, peptides are extracted from MHC molecules in acid, purified and then run through a mass spectrometer, which can be used to determine the amino acid sequences constituting each of the peptides. Robert A. Henderson, Eric L. Huczko and Ye Chen in my laboratory and Andrea L. Cox, Hanspeter Michel, Wanda M. Bodnar, Theresa A. Davis and Jeffrey Shabanowitz in the laboratory of Donald F. Hunt of the University of Virginia are among the practitioners of this technique. They have used it to analyze the structure of peptides associated with several human class I MHC molecules. Those studies have confirmed that MHC molecules can bind to an extraordinarily diverse set of peptides. A human cell has between a half million and a million class I molecules of a single form; we estimate those molecules present more than 10,000 different peptides and perhaps as many as 100,000.

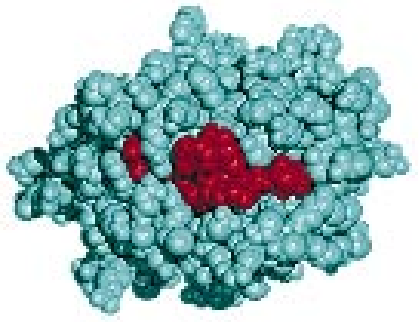
We and others have found that most of the peptides associated with a particular class I form share some simple structural motifs, or features, that support binding. Peptides that bind to class I MHC molecules are usually eight or nine amino acid residues long. That length seems to be optimal for allowing the two ends of a peptide—the amino and carboxyl terminals—to fit into pockets at opposite ends of the MHC binding cleft.

Amino acids at certain positions in the peptides are also highly conserved. For example, most of the peptides that

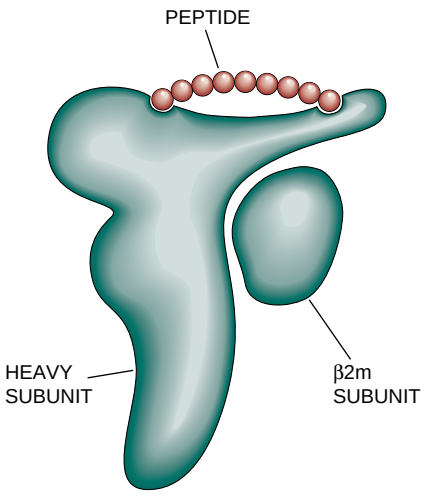


ANTIGEN COMPLEXES of major histocompatibility complex (MHC) molecules and peptides can be recognized by *T* lymphocytes. Class I MHC molecules, found on all nucleated body cells, can present peptides from viruses. *T* cells respond to such complexes by killing the infected cell (*left*). Class II MHC molecules, found only on antigen-presenting cells, present peptides scavenged from extracellular proteins. These complexes trigger the release of lymphokines (*right*).





CLASS I MHC molecules consist of a heavy subunit and a lighter ß2m subunit (*bottom*). As seen from above, a cleft at the top of the molecule holds a peptide, which is typically about nine amino residues long (*middle*). The ends of the peptide are held in place by bonds (*dotted lines*) within pockets at the edges of the cleft (*top*).



bind to a human class I molecule called HLA-A2.1 have the amino acid leucine in the second position from their amino terminus; at the carboxyl end of the molecule, the last amino acid is always uncharged and hydrophobic. Conversely, the peptides that bind to the human class I molecule HLA-B27 have the amino acid arginine in the second position and end with a residue that is positively charged and hydrophilic.

Such information, as well as other structural data, has painted a fairly simple picture of how peptides bind to class I MHC molecules. The two ends of the peptide and two or three additional amino acid residues fit into well-separated pockets in the MHC binding cleft. Those connections provide most of the binding energy between the molecules. The rest of the peptide chain extends above the surface of the cleft and is not strongly constrained by interactions with the MHC molecule, which means that the latter can accommodate a wide variety of peptide structures. Many peptides derived from the proteins of pathogens have structural motifs that would allow them to bind to an MHC molecule under the right circumstances. Indeed, using this motif information, investigators have in a few cases been able to predict which peptide an MHC molecule on an infected cell would present.

The peptide binding site of class II molecules is similar to that of class I molecules, but the differences are significant. Most important, the binding cleft of class II molecules lacks pockets for specifically grabbing the ends of a peptide. Instead most of the binding takes place in the middle of the cleft and, therefore, more toward the middle of the bound peptide. As a consequence, the peptides associated with class II molecules exhibit greater variation in length and are on average considerably longer than those associated with class I. Many peptides associated with any given form of class II molecule make up a nested set: that is, they share a core sequence of amino acids (which bind within the MHC cleft) and vary only in the lengths of their amino and carboxyl terminal ends.

The binding cleft of class II molecules contains more centrally located pockets than those of class I molecules, but so far much less is known about their importance in constraining which amino acid features a peptide must have. As a result, structural motifs that would make it possible to predict reliably which peptides would bind to a class II MHC molecule are now only beginning to be defined.

Association with peptides is a normal step in the biosynthesis and assembly of both class I and II MHC molecules. But just as the sources of the peptides for each class differ, so, too, do the mechanisms for their assembly.

After the heavy-chain and ß2m light-chain subunits that make up the class I molecules are synthesized, they join together inside the membranous organelle called the endoplasmic reticulum. If ß2m is not present, the heavy chain cannot fold into its proper conformation. It therefore cannot move through another organelle, the Golgi apparatus, and be routed to its final destination on the cell surface. As recent research has now made clear, the heavy chain–ß2m complex must also bind to a peptide while inside the endoplasmic reticulum to make this trip.

Townsend and Klaus Kärre of the Karolinska Institute in Stockholm demonstrated that point most elegantly. They identified mutant cells that had only 5 percent of the standard number of class I MHC molecules on their surface even though they were synthesizing heavy chains and ß2m light chains in the normal amounts. The chains turned out to be unfolded and trapped in the endoplasmic reticulum. When Townsend and Kärre added the appropriate peptides to these cells, the chains folded correctly, and the cells expressed more normal levels of class I molecules. Thus, peptides stabilize the interactions of the heavy and light chains, acting in many ways as a third subunit of the class I molecule.

The identification of mutant cells, like those studied by Townsend and Kärre, in which peptides fail to associate with class I MHC molecules, spurred several laboratories to investigate the defect. In late 1990 four groups simultaneously identified two genes in the MHC that encode transport proteins. These proteins are members of an extensive family of similar molecules that help to transport small molecules across cell membranes in diverse organisms. Researchers hypothesized that the new MHC-associated proteins carried peptides from the cytoplasm into the endoplasmic reticulum. They aptly named the molecule made by the genes TAP (for *T*ransporter associated with *A*ntigen *P*rocessing).

Follow-up studies have revealed that

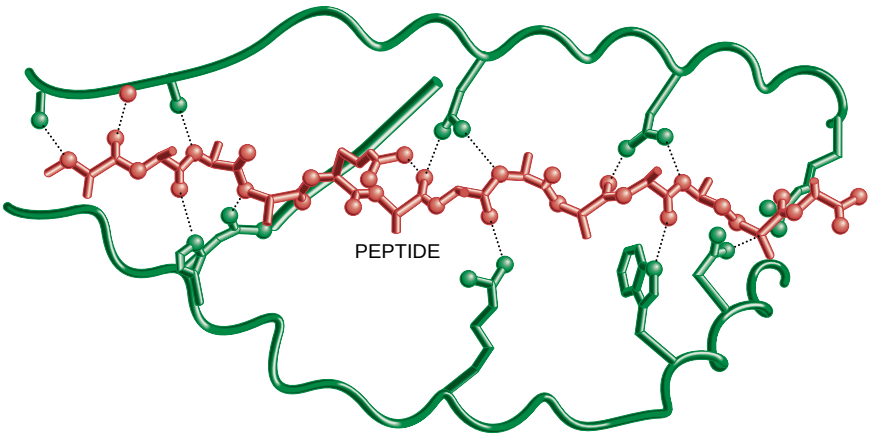
all the known lines of mutant cells that exhibit faulty antigen processing of this type have defects in one or the other of these TAP genes. A variety of other experiments have also supported the hypothesis, including a direct demonstration that membrane vesicles containing TAP catalyzed the transport of small peptides.

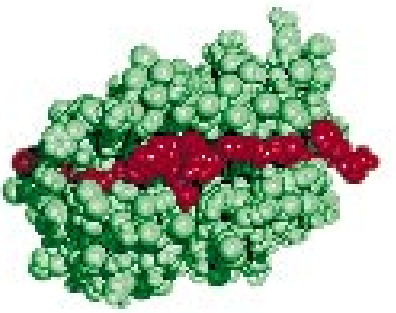
How are the peptides transported by TAP made? A definite answer is still unknown, but a strong circumstantial case can be made for the involvement of an enzymatic complex called the proteasome. Proteasomes are large cylindrical structures found in many compartments of a cell. They are an amalgamation of several different proteases, or protein-splitting enzymes, and they appear to be a cell's principal mechanism for degrading proteins that have either outlived their usefulness, been damaged or been folded incorrectly.

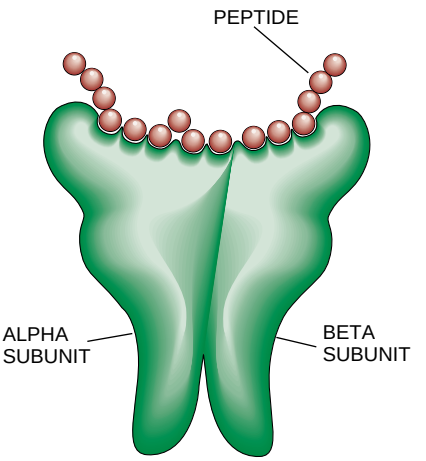
John J. Monaco of the Medical College of Virginia has performed some of the most intriguing studies of proteasomes. He demonstrated that two subunits sometimes found in proteasomes are encoded by genes in the MHC that are immediately adjacent to those for TAP. Normally, only about 10 percent of the proteasomes in cells contain these subunits. But if a cell is exposed to gamma-interferon, a lymphokine released during immune responses, the cell's expression of these subunits increases, and they are associated with more proteasomes. (The cell's expression of MHC molecules and TAP also increases.)

Kenneth L. Rock and Alfred L. Goldberg of Harvard have recently shown that the inclusion of these subunits in a proteasome causes it to produce peptides that end in basic or hydrophobic amino acids—exactly the types to which most class I molecules bind. Whether the two subunits also alter the length of the produced peptides to the optimum size for class I molecules is not known. Nevertheless, it is tempting to believe that proteins made in the cytoplasm are degraded by proteasomes and transported by TAP into the endoplasmic reticulum, where they can bind to class I MHC molecules.

One curious observation, however, is that many mutant cells that lack one or both components of TAP still express fairly high levels of some forms of class I MHC molecules on their surface. Henderson in my laboratory and Michel in Hunt's laboratory at the University of Virginia both looked at why that is the case. They discovered that on these cells, the peptides associated with the class I molecules all seem to come from the signal sequences of cellular proteins. Signal sequences are features commonly found at the amino terminal ends of newly synthesized proteins that are heading to the cell surface or other internal cellular compartments. When the organelles called ribosomes synthesize such proteins, these sequences ensure that the ribosomes will attach to the endoplasmic reticulum before the protein is completed. In effect, the signal sequences help to direct the new proteins down the path toward their final destinations.







CLASS II MHC molecules have alpha and beta subunits that are nearly the same size (*bottom*). These molecules also hold peptides in a cleft (*middle*). These peptides, however, are gripped primarily by bonds near the center of the cleft (*top*). The peptides bound by class II molecules are therefore generally longer and of more variable length than those associated with class I.

As the protein extrudes into the endoplasmic reticulum, an enzyme clips the signal sequence from its front end. Once freed in this way, the signal sequences in the endoplasmic reticulum provide a ready source of peptides that could associate with class I MHC molecules, thereby circumventing the antigen-processing defects in the mutant cell. Researchers have learned that at least two peptides recognized by *T* lymphocytes are derived from signal sequences, which indicates that this alternate antigen-processing pathway may be quite important.

Given that both the class I and II MHC molecules assemble inside the endoplasmic reticulum, it is surprising that they do not bind to the same peptides. Part of the reason may be that the peptides transported into the endoplasmic reticulum by TAP do not have structural features that would enable them to bind to class II molecules stably.

Perhaps a more cogent reason, however, is that immediately after their synthesis, the class II subunits associate with a third molecule, called the invariant chain, or Ii. The invariant chain prevents peptides from binding to class II molecules, either by directly interfering with peptide binding or by keeping the class II molecules in a partially unfolded state. It also redirects class II MHC molecules along a path to the cell surface that class I molecules and most other surface membrane proteins do not follow: through the Golgi apparatus and into endosomes.

Endosomes are membrane vesicles formed by invaginations of the cell-surface membrane. They therefore often contain surface proteins and their associated ligands (the molecules to which

they bind). As endosomes pass through the cell, their interior becomes acidic, and they accumulate proteases that degrade many of the encapsulated surface proteins and ligands. Ultimately, endosomes recycle to the surface membrane, fuse with it and return their contents to the surface.

Peter Cresswell of Duke University found that when class II–Ii complexes move into endosomes, the movement of the vesicles toward the surface stops for up to six hours. During this time, endosomal proteases digest the invariant chain, which frees the class II molecules to bind to other peptides in the vesicle. Many of those peptides are of course derived from extracellular sources. Finally, the class II–peptide complexes move out to the cell surface.

Another interesting point about the formation of class II–peptide complexes was learned by looking at certain mutant cells created by Elizabeth D. Mullins and Donald A. Pious of the University of Washington. On the surface of these cells, the class II molecules have an oddly floppy, easily denatured form. The appearance and behavior of these molecules resemble that of newly synthesized class II molecules in the endoplasmic reticulum. One might naturally presume that these floppy surface molecules also lack a stabilizing peptide. Yet direct isolation of the molecules from the mutants has shown that is not the case: the floppy molecules are bound to a set of peptides derived from one small region of the invariant chain. These peptides are called CLIPs (class II–associated invariant chain peptides).

The mutation in these cells seems to interfere with the ability of class II MHC molecules to bind to any peptides other than those from the invariant chain. In independent experiments, Mullins and Pious recently showed that the defect in these cells affects a newly identified molecule, called DM, that is structurally related to but distinct from conventional class II molecules. Before their work, the function of DM was completely unknown.

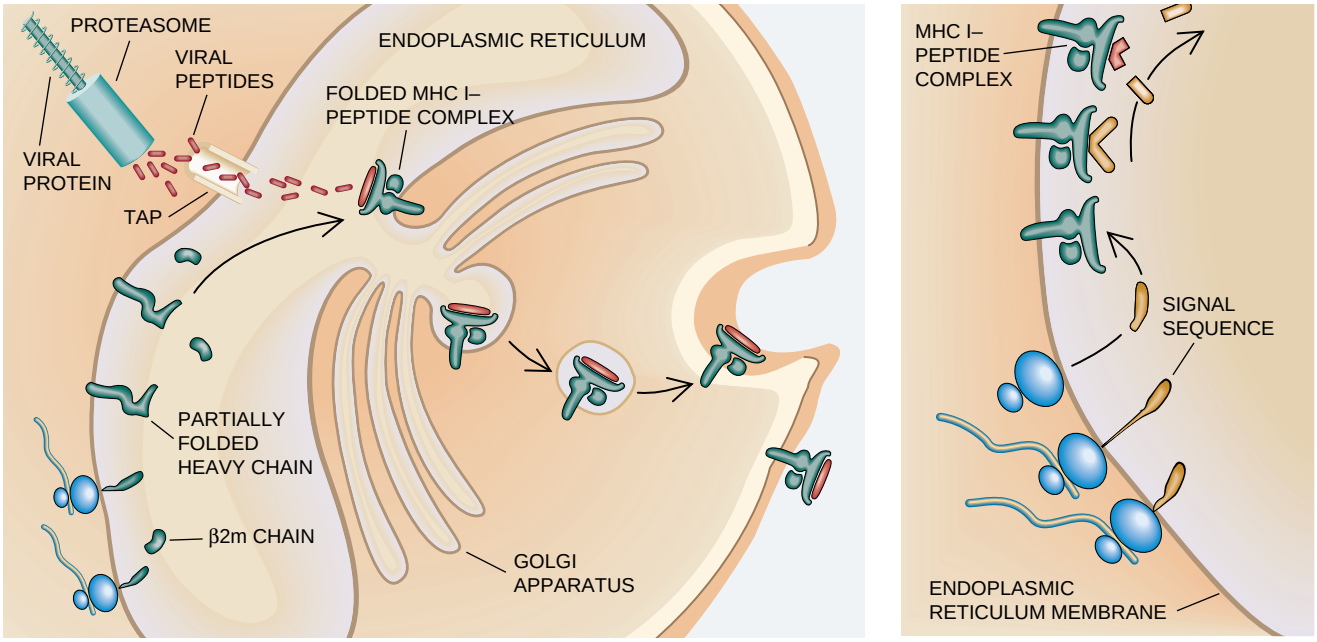
The exact roles of the CLIPs and DM in the normal processing pathway for class II molecules are not yet understood. One attractive hypothesis is that CLIP is the part of the invariant chain that physically occupies the peptide binding cleft on the class II molecule or that at least alters the structure of the class II molecule to prevent other peptides from binding. After the invariant chain is degraded in the endosome, CLIP still remains associated with the class II molecule until it is actively removed by DM.

As we have seen, the antigen-processing mechanisms create a representative sample of peptides from the plethora of proteins that a cell makes and ingests. The display of those peptides on MHC molecules in turn allows the immune system to identify and destroy cells that are havens for infectious agents or are otherwise abnormal.

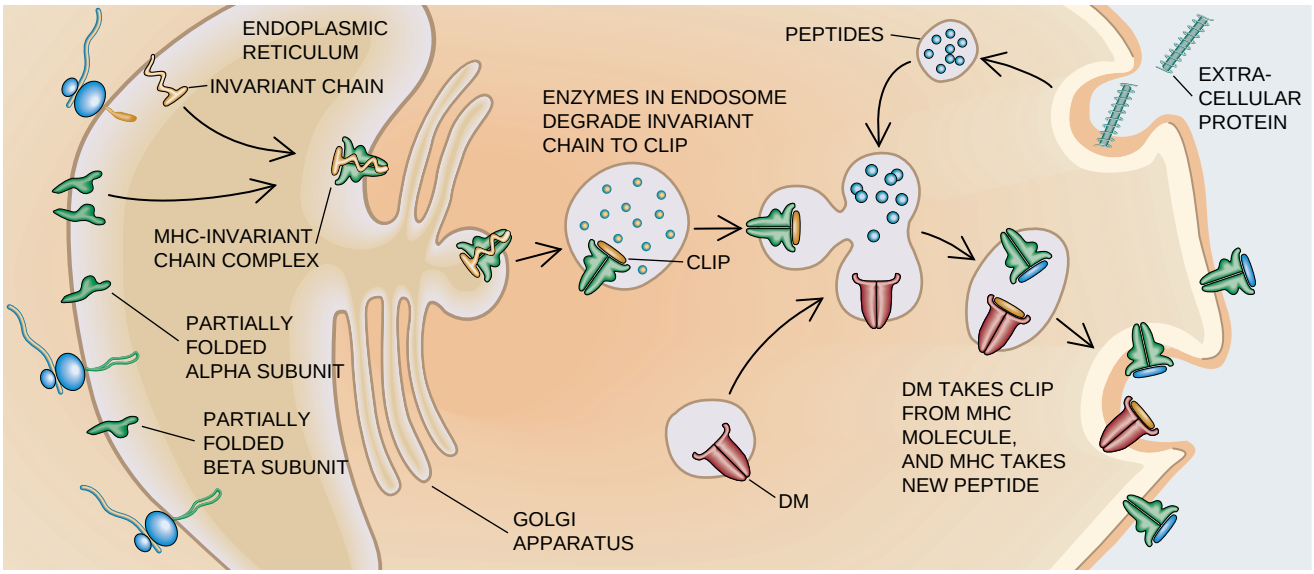
MHC molecules must be able to present many foreign peptides and do so in such a way that the complex looks different from one formed with a similar host peptide. That requirement probably explains why individuals express several forms of class I and II MHC molecules and why the population contains hundreds of forms. Some of those forms are more likely than others to bind to peptides from specific pathogens.

Real examples of the significance of MHC variety are beginning to appear in the literature. A few years ago Adrian Hill of the University of Oxford and his colleagues found evidence that in humans, susceptibility to malaria varies with the expression of certain class I MHC molecules. Those forms of class I molecules that seem to confer the greatest resistance are especially common among people living where malaria is widespread, as natural selection would dictate.

As might be expected, however, a few pathogens have learned to fight back against the antigen-processing system that routinely frustrates them. Several kinds of viruses, for example, can suppress the expression of MHC molecules



CLASS I ANTIGEN PATHWAY (*left*) typically begins with intracellular proteins, such as those from viral infections, being reduced to peptides by proteasomes. These peptides are then carried into the endoplasmic reticulum by a transporter protein, TAP. Inside the endoplasmic reticulum, the peptides and the partially folded class I subunits associate into a folded MHC-peptide complex. This complex is routed through the Golgi complex and carried to the cell surface. In a variation on this pathway discovered by the author and his colleagues (*right*), the peptides can originate with the signal sequences that are clipped from the beginnings of cellular proteins being assembled in the endoplasmic reticulum.



CLASS II ANTIGEN PATHWAY begins with the binding of the MHC subunits to an invariant chain molecule. After this complex passes through the Golgi complex, the invariant chain is reduced to a smaller peptide, CLIP. It is thought that inside endosome vesicles, a molecule called DM steals the CLIP from the MHC complex, which frees the MHC molecule to bind with peptides derived from extracellular proteins. This final class II–peptide complex then travels to the cell surface.

during the early stages of an infection. Many kinds of adenoviruses make a molecule that binds to newly synthesized class I molecules in the endoplasmic reticulum and prevents their expression on the cell surface. Other adenoviruses make a molecule that interferes with the expression of the class I gene. Interference with the surface expression of class I MHC molecules has now also been described for cytomegaloviruses and herpes simplex virus, although the mechanisms they employ are not yet known. Notwithstanding these examples, in general, antigen processing enables the immune system to control infections extremely well.

It may also be relevant to controlling cancer. Because many tumors express mutated proteins, the immune system may look for peptides derived from such proteins as indicators that a cell has been transformed and is on its way to creating a tumor. Many tumors have been identified in which expression of class I MHC molecules is reduced. Experimental manipulations that increase MHC expression often make these tumors more controllable by the immune system. Some recent evidence also suggests that a few types of tumor cells may lower their expression of TAP, presumably to avoid recognition by *T* lymphocytes.

The nature of the peptides that tumor-specific *T* lymphocytes can recognize is just now being unraveled. Thierry Boon and his colleagues at the Ludwig Institute for Cancer Research in Brussels have looked in some detail at *T* lymphocytes that recognize human melanoma cells [see "Teaching the Immune System to Fight Cancer," by Thierry Boon; SCIENTIFIC AMERICAN, March 1993]. They have shown that one target of the *T* cells appears to be a set of peptides from a protein called MAGE-1, which is expressed in a variety of tumors but is almost undetectable in normal tissue. Boon, Steven A. Rosenberg of the National Cancer Institute and my colleagues and I at the University of Virginia have also identified antigenic peptides that arise from three proteins that both normal melanocytes and melanoma tumor cells express.

These results suggest that the effectiveness of antitumor immunity may be limited in part by the availability of unusual target peptides in tumor cells. One approach to augmenting antitumor immunity will certainly involve the further identification of class I–peptide complexes on tumor cells that *T* cells can recognize, as well as the development of strategies to increase their immunogenicity.

Perversely enough, there seem to be instances in which the power of the antigen-processing system works against the health of the body. Over the past decade, researchers have made the thought-provoking discovery that the expression of certain class II MHC molecules is linked to many autoimmune diseases, such as juvenile-onset diabetes and rheumatoid arthritis, in which the immune system tragically attacks the host's tissues. These MHC molecules are presumably presenting peptides from the host and precipitating these immune attacks. The immune system has mechanisms that normally prevent such presentations or halt the destructive responses they would elicit. Why the mechanisms fail in these cases is still a puzzle.

How the presentation of peptides from host and foreign sources is linked to the development of these conditions is among the most intriguing questions facing immunologists today. As the details become clearer, immunologists may eventually be able to design therapies that can selectively manipulate antigen processing. We may someday be able to block the presentation of antigens that exacerbate autoimmune diseases or enhance the processing of antigens that would reveal infections or tumors. For the moment, however, the mysteries at hand are more than enough to keep immunologists well occupied.

FURTHER READING

THE BIOCHEMISTRY AND CELL BIOLOGY OF ANTIGEN PROCESSING AND PRESENTATION. Ronald N. Germain and David H. Margulies in *Annual Review of Immunology,* Vol. 11, pages 403–450; 1993.

NATURALLY PROCESSED PEPTIDES. Edited by Alessandro Sette. Karger, 1993.

ANTIGENIC PEPTIDE BINDING BY CLASS I AND CLASS II HISTOCOMPATIBILITY PROTEINS. Lawrence J. Stern and Don C. Wiley in *Structure,* Vol. 2, No. 4, pages 245–251; April 15, 1994.

STRUCTURE OF PEPTIDES ASSOCIATED WITH CLASS I AND CLASS II MHC MOLECULES. Victor H. Engelhard in *Annual Review of Immunology,* Vol. 12, pages 181–207; 1994.

Red Tides

Many experts believe these blooms of toxic algae have recently become more prevalent, posing a greater threat to human and marine health

by Donald M. Anderson

Late in 1987 scientists faced a baffling series of marine catastrophes. First, 14 humpback whales died in Cape Cod Bay, Mass., during a five-week period. This die-off, equivalent to 50 years of "natural" mortality, was not a stranding, in which healthy whales beach themselves. Instead the cetaceans died at sea—some rapidly—and then washed ashore. Postmortem examinations showed that the whales had been well immediately before their deaths and that many of them had abundant blubber and fish in their stomachs, evidence of recent feeding. Alarmed and saddened, the public and press blamed pollution or a chemical spill for the mysterious deaths.

Two more mass poisonings occurred that month, but the victims in these new cases were humans. Fishermen and beachgoers along the North Carolina coast started complaining of respiratory problems and eye irritation. Within days, residents and visitors who had eaten local shellfish experienced diarrhea, dizziness and other symptoms suggesting neurotoxic poisoning. The illnesses bewildered epidemiologists and even prompted public conjecture that a nearby sunken submarine was leaking poison gas.

Concurrently, hospitals in Canada began admitting patients suffering from disorientation, vomiting, diarrhea and abdominal cramps. All had eaten mussels from Prince Edward Island. Although Canadian authorities had dealt with shellfish poisoning outbreaks for decades, these symptoms were unfamiliar and disturbing: some patients exhibited permanent short-term memory loss. They could remember addresses but could not recall their most recent meal, for example. The officials quickly restricted the sale and distribution of mussels but eventually reported three deaths and 105 cases of acute poisoning in humans.

We now know that these seemingly unrelated events were all caused, either directly or indirectly, by toxic, single-celled algae called phytoplankton—vast blooms of which are commonly referred to as red tides. Although red tides have been recorded throughout history, the incidents mentioned above were entirely unexpected. As we shall see, they illustrate several major issues that have begun to challenge the scientific and regulatory communities.

Indeed, there is a conviction among many experts that the scale and complexity of this natural phenomenon are expanding. They note that the number of toxic blooms, the economic losses from them, the types of resources affected and the kinds of toxins and toxic species have all increased. Is this expansion real? Is it a global epidemic, as some claim? Is it related to human activities, such as rising coastal pollution? Or is it a result of increased scientific awareness and improved surveillance or analytical capabilities? To address these issues, we must understand the physiological, toxicological and ecological mechanisms underlying the growth and proliferation of red tide algae and the manner in which they cause harm.

Certain blooms of algae are termed red tides when the tiny pigmented plants grow in such abundance that they change the color of the seawater to red, brown or even green. The name is misleading, however, because many toxic events are called red tides even when the waters show no discoloration. Likewise, an accumulation of nontoxic, harmless algae can change the color of ocean water. The picture is even more complicated: some phytoplankton neither discolor the water nor produce toxins but kill marine animals in other ways. Many diverse phenomena thus fall under the "red tide" rubric.

Of the thousands of living phytoplankton species that make up the base of the marine food web, only a few dozen are known to be toxic. Most are dinoflagellates, prymnesiophytes or chloromonads. A bloom develops when these single-celled algae photosynthesize and multiply, converting dissolved nutrients and sunlight into plant biomass. The dominant mode of reproduction is simple asexual fission—one cell grows larger, then divides into two cells, the two split into four, and so on. Barring a shortage of nutrients or light, or heavy grazing by tiny zooplankton that consume the algae, the population's size can increase rapidly. In some cases, a milliliter of seawater can contain tens or hundreds of thousands of algal cells. Spread over large areas, the phenomenon can be both visually spectacular and catastrophic.

Some species switch to sexual reproduction when nutrients are scarce. They form thick-walled, dormant cells, called cysts, that settle on the seafloor and can survive there for years. When favorable growth conditions return, cysts germinate and reinoculate the water with swimming cells that can then bloom. Although not all red tide species form cysts, many do, and this transformation explains important aspects of their ecology and biogeography. The timing and location of a bloom can depend on

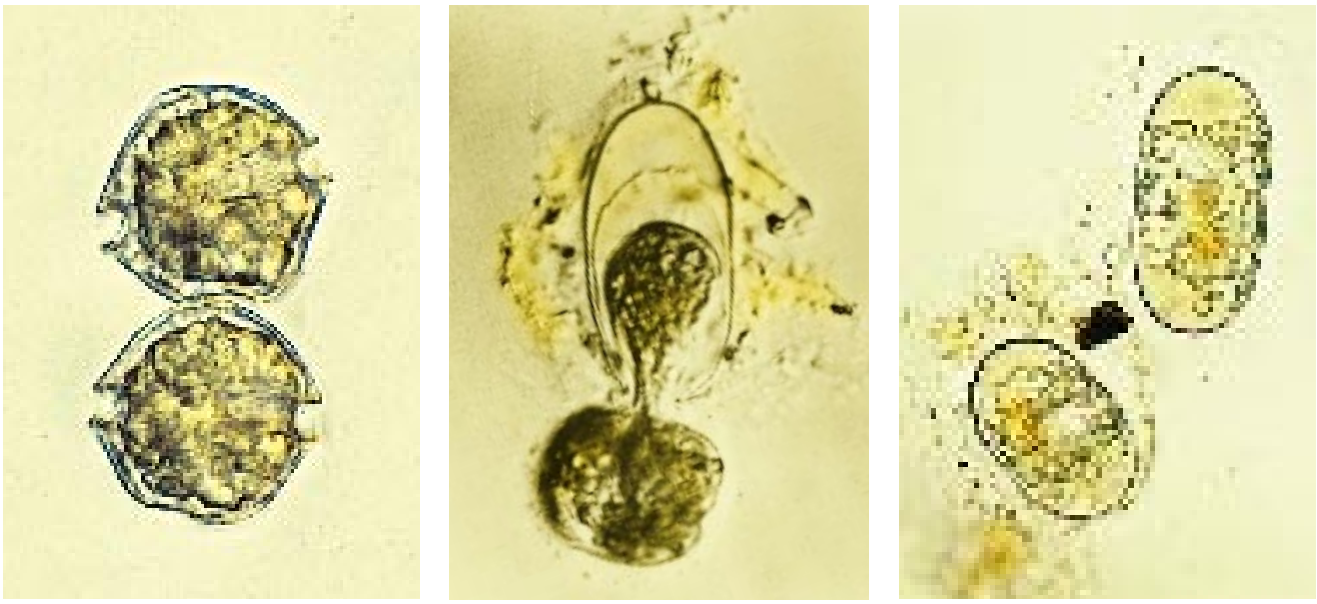
DONALD M. ANDERSON is a Senior Scientist at the Woods Hole Oceanographic Institution. In 1977 he earned a doctorate in aquatic sciences from the department of civil engineering at the Massachusetts Institute of Technology. Anderson studies the physiological and genetic regulation of toxicity in dinoflagellates, their bloom dynamics and the global biogeography of toxic *Alexandrium* species. He also participates in various international programs for cooperative research and training on red tides, marine biotoxins and harmful algae.

RED TIDES appear when pigmented algae proliferate and form blooms. Even nontoxic species, including *Noctiluca,* shown blooming in a Japanese harbor, can kill marine animals by depleting the oxygen in shallow waters.

when the cysts germinate and where they were deposited, respectively. Cyst production facilitates species dispersal as well; blooms carried into new waters by currents or other means can deposit "seed" populations to colonize previously unaffected areas.

A dramatic example of natural dispersal occurred in 1972, when a massive red tide reaching from Maine to Massachusetts followed a September hurricane. The shellfish toxicity detected then for the first time has recurred in that region virtually every year now for two decades. The cyst stage has provided a very effective strategy for the survival and dispersal of many other red tide species as well.

How do algal blooms cause harm? One of the most serious impacts on human life occurs when clams, mussels,



LIFE CYCLE of many toxic species enables them to survive for years under adverse conditions. When nutrients are scarce, the algae form thick-walled, dormant cysts (*left*). The cysts often travel into new waters. When favorable growth conditions return, the cysts germinate (*center*) and colonize previously unaffected areas with harmful organisms (*right*).

oysters or scallops ingest the algae as food and retain the toxins in their tissues. Typically the shellfish themselves are only marginally affected, but a single clam can sometimes accumulate enough toxin to kill a human being. These shellfish poisoning syndromes have been described as paralytic, diarrhetic and neurotoxic, shortened to PSP, DSP and NSP. The 1987 Canadian outbreak in which some patients suffered memory loss was appropriately characterized as amnesic shellfish poisoning, or ASP. The North Carolina episode was NSP.

A related problem, ciguatera fish poisoning, or CFP, causes more human illness than any other kind of toxicity originating in seafood. It occurs predominantly in tropical and subtropical islands, where from 10,000 to 50,000 individuals may be affected annually. Dinoflagellates that live attached to seaweeds produce the ciguatera toxins. Herbivorous fishes eat the seaweeds and the attached dinoflagellates as well. Because ciguatera toxin is soluble in fat, it is stored in the fishes' tissues and travels through the food web to carnivores. The most dangerous fish to eat are thus the largest and oldest, often considered the most desirable as well.

Symptoms do vary among the different syndromes but are generally neurological or gastrointestinal, or both. DSP causes diarrhea, nausea and vomiting, whereas PSP symptoms include tingling and numbness of the mouth, lips and fingers, accompanied by general muscular weakness. Acute doses inhibit respiration, and death results from respiratory paralysis. NSP triggers diarrhea, vomiting and abdominal pain, followed by muscular aches, dizziness, anxiety, sweating and peripheral tingling. Ciguatera induces an intoxication syndrome nearly identical to NSP.

Illnesses and deaths from algal-derived shellfish poisons vary in number from year to year and from country to country. Environmental fluctuations profoundly influence the growth and accumulation of algae and thus their toxicity as well. Furthermore, countries differ in their ability to monitor shellfish and detect biotoxins before they reach the market. Developed countries typically operate monitoring programs that permit the timely closure of contaminated resources. Illnesses and deaths are thus rare, unless a new toxin appears (as in the ASP crisis in Canada) or an outbreak occurs in an area with no history of the problem (as in North Carolina). Developing countries, especially those having long coastlines or poor populations who rely primarily on the sea for food, are more likely to incur a higher incidence of sickness and death from algal blooms.

Phytoplankton can also kill marine animals directly. In the Gulf of Mexico, the dinoflagellate *Gymnodinium breve* frequently causes devastating fish kills. As the wild fish swim through *G. breve* blooms, the fragile algae rupture, releasing neurotoxins onto the gills of the fish. Within a short time, the animals asphyxiate. Tons of dead fish sometimes cover the beaches along Florida's Gulf Coast, causing several millions of dollars to be lost in tourism and other recreation-based businesses.

Farmed fish are especially vulnerable because the caged animals cannot avoid the blooms. Each year, farmed salmon, yellowtail and other economically important species fall victim to a variety of algal species. Blooms can wipe out entire fish farms within hours, killing fingerlings and large fish alike. Algal blooms thus pose a large threat to fish farms and their insurance providers. In Norway an extensive program is under way to minimize these impacts. Fish farmers make weekly observations of algal concentrations and water clarity. Other parameters are transmitted to shore from instruments on moored buoys. The Norwegian Ministry of Environment then combines this information with a five-day weather forecast to generate an "algal forecast" for fish farmers and authorities. Fish cages in peril are then towed to clear water.

Unfortunately, not much more can be done. The ways in which algae kill fish are poorly understood. Some phytoplankton species produce polyunsaturated fatty acids and galactolipids that destroy blood cells. Such an effect would explain the ruptured gills, hypoxia and edema in dying fish. Other algal species produce these hemolytic compounds and neurotoxins as well. The combination can significantly reduce a fish's heart rate, resulting in reduced blood flow and a deadly decrease in oxygen.

Moreover, nontoxic phytoplankton can kill fish. The diatom genus *Chaetoceros* has been linked to dying salmon in the Puget Sound area of Washington State, yet no toxin has ever been identi-

fied in this group. Instead species such as *C. convolutus* sport long, barbed spines that lodge between gill tissues and trigger the release of massive amounts of mucus. Continuous irritation exhausts the supply of mucus and mucous cells, causing lamellar degeneration and death from reduced oxygen exchange. These barbed spines probably did not evolve specifically to kill fish, since only caged fish succumb to the blooms. The problems faced by fish farmers are more likely the unfortunate result of an evolutionary strategy by certain *Chaetoceros* species to avoid predation or to stay afloat.

Algal toxins also cause mortalities as they move through the marine food web. Some years ago tons of herring died in the Bay of Fundy after consuming small planktonic snails that had eaten the PSP-producing dinoflagellate *Alexandrium.* From the human health standpoint, it is fortunate that herring, cod, salmon and other commercial fish are sensitive to these toxins and, unlike shellfish, die before toxins reach dangerous levels in their flesh. Some toxin, however, accumulates in the liver and other organs of certain fish, and so animals such as other fish, marine mammals and birds that consume whole fish, including the viscera, are at risk.

We now can reconstruct the events that killed the whales in 1987. A few weeks of intense investigations that year by marine pathologist Joseph R. Geraci of the Ontario Veterinary College, myself and many others revealed that the PSP toxins most likely caused these deaths. The dinoflagellate *Alexandrium tamarense* produced the toxins, which reached the whales via their food web. We analyzed mackerel that the whales had been eating and found saxitoxin, not in their flesh but concentrated in the liver and kidney. Presumably the mackerel ate zooplankton and small fish that had previously dined on *Alexandrium.*

The humpbacks were starting their southward migration and were feeding heavily. Assuming that they consumed 4 percent of their body weight daily, we calculated that they received a saxitoxin dosage of 3.2 micrograms per kilogram of body weight. But was this a fatal dose? Unfortunately, in 1987 we had no data that directly addressed how much toxin would kill a whale. We knew the minimum lethal dose of saxitoxin for humans is seven to 16 micrograms per kilogram of body weight, but that was two to five times more than what the whales had probably ingested.

Our calculations were initially disheartening, but as we thought about it we realized that whales might be more sensitive to the toxins than are humans. First, whales would have received continual doses of toxin as they fed, whereas human mortality statistics are based on single feedings. Second, during a dive, the mammalian diving reflex channels blood and oxygen predominantly to the heart and brain. The same mechanism sometimes protects young children who fall through thin ice and survive drowning, despite being underwater for half an hour or longer. For humans, cold water induces the reflex, but for whales, it is activated during every dive.

Each dive then would expose the most sensitive organs to the toxin, which would bypass the liver and kidney, where it could be metabolized and excreted. Finally, saxitoxin need not have killed the whales directly. Even a slightly incapacitated animal might have difficulty orienting to the water surface or breathing correctly. The whales may actually have drowned following a sublethal exposure to saxitoxin. The exact cause will never be known, but the evidence strongly suggests that these magnificent creatures died from a natural toxin originating in microscopic algae.

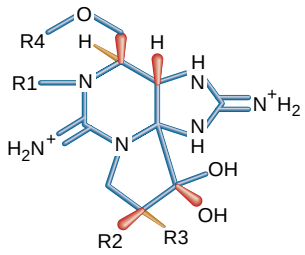
Other examples of toxins traveling up the food web appear nearly every year. In 1991 sick or dying brown pelicans and cormorants were found near Monterey Bay, Calif. Wildlife experts could find no signs that pesticides, heavy metals or other pollutants were involved. The veterinarian in charge of the study telephoned Jeffrey Wright of the National Research Council laboratory in Halifax, Nova Scotia. Wright had directed the Canadian Mussel Toxin Crisis Team that identified the poison responsible for the mysterious ASP episode in 1987. His team had isolated a toxin from the Prince Edward Island mussels, called domoic acid, and traced it to its source—a diatom, *Pseudonitzschia pungens,* that had been considered harmless. Four years later members of the same Canadian team quickly ascertained that the sick and dying birds in California had eaten anchovies that contained domoic acid, again from *Pseudonitzschia* (but a different species).

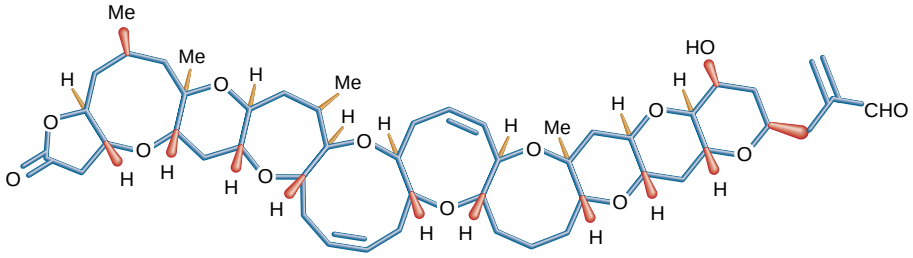
The toxins responsible for these syndromes are not single chemical entities but are families of compounds having similar chemical structures and effects. For example, the saxitoxins that cause PSP are a family of at least 18 different compounds with widely differing potencies. Most algal toxins cause human illness by disrupting electrical conduction, uncoupling communication between nerve and muscle, and impeding critical physiological processes. To do so, they bind to specific membrane receptors, leading to

HUMPBACK WHALES, a total of 14, died suddenly from exposure to a bloom of *Alexandrium tamarense* in Cape Cod Bay, Mass., in 1987. Researchers later learned that the whales had eaten mackerel whose organs contained high concentrations of saxitoxin, a neurotoxin produced by the algae.

Algal Toxins

The structure of red tide toxins varies considerably. Saxitoxin compounds (*one example shown at left*) sport different combinations of H^+, OH^- and SO_3^- on the R1 to R4 sites, but all members of this family block the sodium channel and thus prevent communication between neurons and muscles. H. Robert Guy of the National Institutes of Health has proposed a structural model of this interaction (*right*). The carbon backbone of the sodium channel is colored gray, the carboxyls are red, nitrogen is blue, and hydrogen is white. Saxitoxin binds in the narrowest region of this channel. The brevetoxins (*one example shown at center*) that cause NSP are much larger molecules that also affect the sodium channel.





changes in the intracellular concentration of ions such as sodium or calcium.

The saxitoxins bind to sodium channels and block the flux of sodium in and out of nerve and muscle cells. Brevetoxins, the family of nine compounds responsible for NSP, bind to a different site on the sodium channel but cause the opposite effect from saxitoxin. Domoic acid disrupts normal neurochemical transmission in the brain. It binds to kainate receptors in the central nervous system, causing a sustained depolarization of the neurons and eventually cell degeneration and death. Memory loss in ASP victims apparently results from lesions in the hippocampus, where kainate receptors abound.

Why do algal species produce toxins? Some argue that toxins evolved as a defense mechanism against zooplankton and other grazers. Indeed, some zooplankton can become slowly incapacitated while feeding, as though they are being gradually paralyzed or otherwise impaired. (In one study, a tintinnid ciliate could swim only backward, away from its intended prey, after exposure to toxic dinoflagellates.) Sometimes grazing animals spit out the toxic algae as though they had an unpleasant taste. These responses would all reduce grazing and thus facilitate bloom formation.

All the same, nontoxic phytoplankton also form blooms, and so it is unlikely that toxins serve solely as self-defense. Scientists are looking within the algae for biochemical pathways that require the toxins, but the search thus far has been fruitless. The toxins are not proteins, and all are synthesized in a series of chemical steps requiring multiple genes. Investigators have proposed biosynthetic pathways, but they have not isolated chemical intermediates or enzymes used only in toxin production. It has thus been difficult to apply the powerful tools of molecular biology to these organisms, other than to study their genes or to develop detection tools.

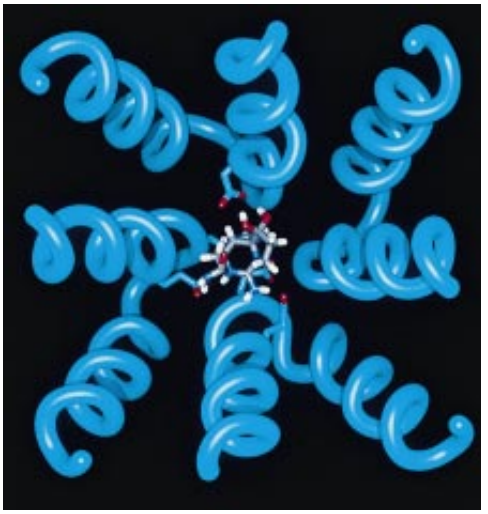
We do have some tantalizing clues about toxin metabolism. For example, certain dinoflagellate strains produce different amounts of toxin and different sets of toxin derivatives when we vary their growth conditions. Metabolism of the toxins is a dynamic process, but we still do not know whether they have a specific biochemical role. As with the spiny diatoms that kill fish, the illnesses and mortalities caused by algal "toxins" may be the result of the accidental chemical affinity of those metabolites for receptor sites on ion channels in higher animals.

The potential role of bacteria or bacterial genes in phytoplankton toxin production is an area of active research. We wonder how a genetically diverse array of organisms, including phytoplankton, seaweeds, bacteria and cyanobacteria, could all have evolved the genes needed to produce saxitoxin [see "The Toxins of Cyanobacteria," by Wayne W. Carmichael; SCIENTIFIC AMERICAN, January]. Several years ago Masaaki Kodama of Kitasato University in Japan isolated intracellular bacteria from antibiotic-treated *A. tamarense* cultures and showed that the bacteria produced saxitoxin. This finding supported an old and long-ignored hypothesis that toxins might originate from bacteria living inside or on the dinoflagellate cell.

Despite considerable study, the jury is still out. Many scientists now accept that some bacteria produce saxitoxins, but they point out that dense bacterial cultures produce extremely small quantities. It is also not clear that those bacteria can be found inside dinoflagellates. That intracellular bacteria produce all of the toxin found in a dinoflagellate cell therefore seems unlikely, but perhaps some synergism occurs between a small number of symbionts and the host dinoflagellate that is lost when the bacteria are isolated in culture. Alternatively, a bacterial gene or plasmid might be involved.

Given the diverse array of algae that produce toxins or cause problems in a variety of oceanographic systems, attempts to generalize the dynamics of harmful algal blooms are doomed to fail. Many harmful species, however, share some mechanisms. Red tides often occur when heating or freshwater runoff creates a stratified surface layer above colder, nutrient-rich waters. Fast-growing algae quickly strip away nutrients in the upper layer, leaving nitrogen and phosphorus only below the interface of the layers, called the pycnocline. Nonmotile phytoplankton cannot easily get to this layer, whereas motile algae, including dinoflagellates, can thrive. Many swim at speeds in excess of 10 meters per day, and some undergo daily vertical migration: they reside in surface waters by day to harvest sunlight like sunbathers, then swim down to the pycnocline to take up nutrients at night. As a result, blooms can suddenly appear in surface waters that are devoid of nutrients and would seem incapable of supporting such prolific growth.

A similar sleight-of-hand can occur horizontally, though over much larger distances. The NSP outbreak in North Carolina illustrates how ocean currents can transport major toxic species from one area to another. Patricia A. Tester,



a biologist at the National Oceanic and Atmospheric Administration's National Marine Fisheries Service laboratory in Beaufort, examined plankton from local waters under a microscope soon after the initial reports of human illnesses. She saw cells resembling the dinoflagellate *G. breve,* the cause of recurrent NSP along Florida's western coast. Experts quickly confirmed her tentative identification, and for the first time in state history, authorities closed shellfish beds because of algal toxins, resulting in a loss of $20 million.

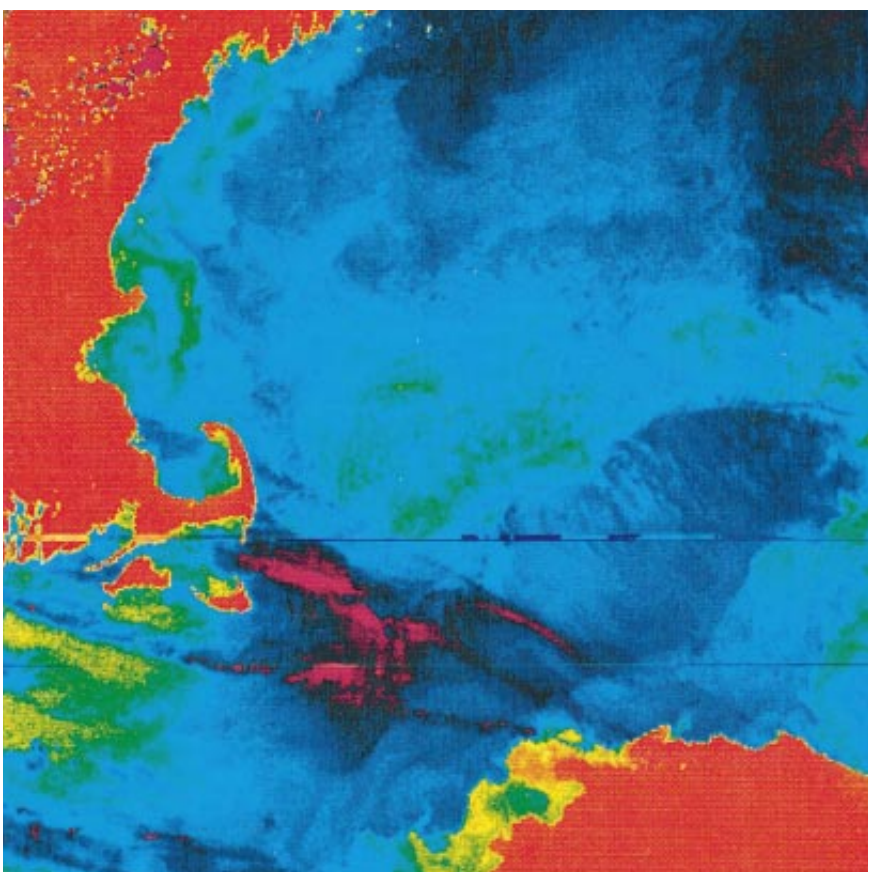
Tester and her co-workers have since used satellite images of sea-surface temperatures to argue that the *G. breve* population in North Carolina originated off the southwestern coast of Florida, nearly 1,000 kilometers away. That bloom traveled from the Gulf of Mexico up the southeastern coast of the U.S., transported by several current systems culminating in the Gulf Stream. After 30 days of transport, a filament of water separated from the Gulf Stream and moved onto North Carolina's narrow continental shelf, carrying *G. breve* cells with it. The warm water mass remained in nearshore waters, identifiable in satellite images for three weeks. Fortunately, *G. breve* does not have a known cyst stage, so it could not establish a seedbed and colonize this new region.

This incident, taken together with many others like it throughout the world, speaks of an unsettling trend. Problems from harmful red tides have grown worse over the past two decades. The causes, however, are multiple, and only some relate to pollution or other human activities. For example, the global expansion in aquaculture means that more areas are monitored closely, and more fisheries' products that can be killed or take up toxins are in the water. Likewise, our discovery of toxins in algal species formerly considered nontoxic reflects the maturation of this field of science, now profiting from more investigators, better analytical techniques and chemical instrumentation, and more efficient communication among workers.

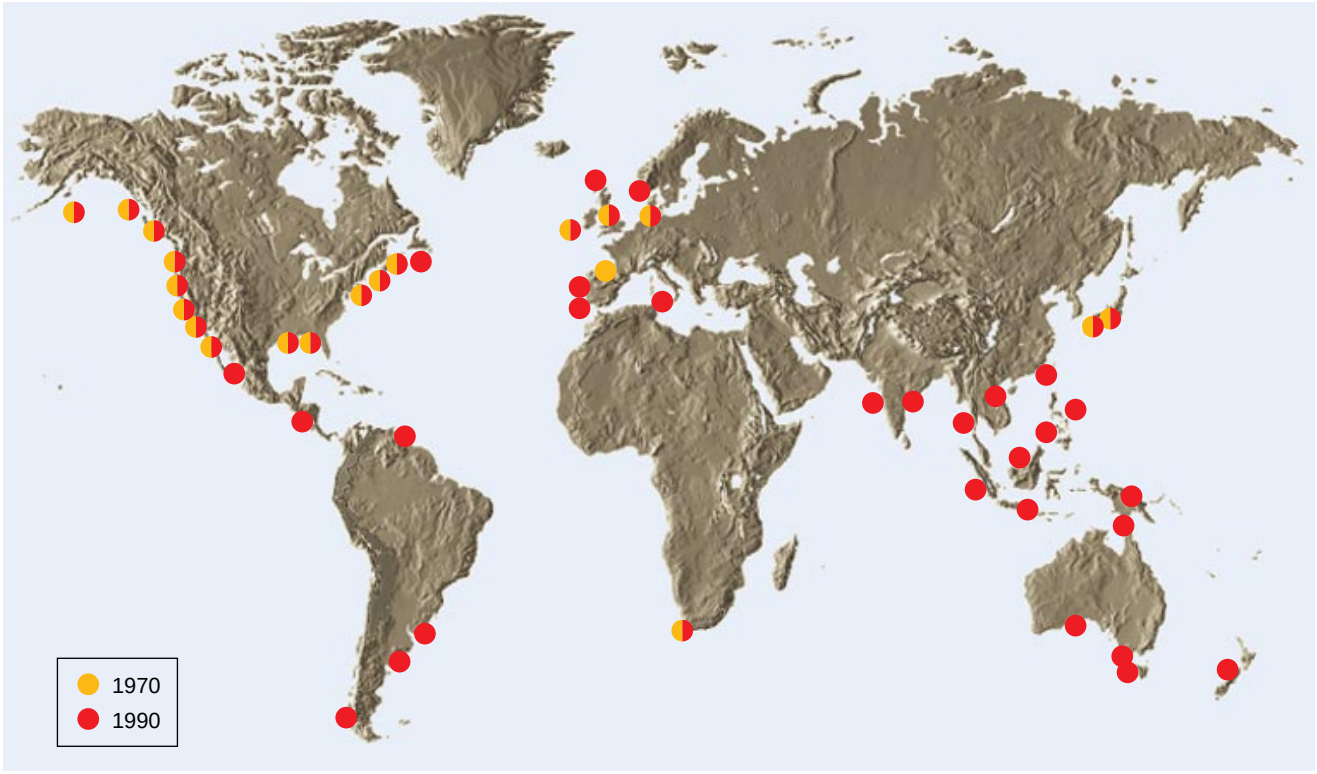
Long-term studies at the local or regional level do show that red tides (in the most general sense of the term) are increasing as coastal pollution worsens. Between 1976 and 1986, as the population around Tolo Harbor in Hong Kong grew sixfold, red tides increased eightfold. Pollution presumably provided more nutrients to the algae. A similar pattern emerged in the Inland Sea of Japan, where visible red tides proliferated steadily from 44 per year in 1965 to more than 300 a decade later. Japanese authorities instituted rigorous effluent controls in the mid-1970s, and a 50 percent reduction in the number of red tides ensued.

These examples have been criticized, since both could be biased by changes in the numbers of observers through time, and both are tabulations of water discolorations from blooms, not just toxic or harmful episodes. Still, the data demonstrate what should be an obvious relationship: coastal waters receiving industrial, agricultural and domestic waste, frequently rich in plant nutrients, should experience a general increase in algal growth. These nutrients can enhance toxic or harmful episodes in several ways. Most simply, all phytoplankton species, toxic and nontoxic, benefit, but we notice the enrichment of toxic ones more. Fertilize your lawn, and you get more grass—and more dandelions.

Some scientists propose instead that pollution selectively stimulates harmful species. Theodore J. Smayda of the University of Rhode Island brings the nutrient ratio hypothesis, an old concept in the scientific literature, to bear on toxic bloom phenomena. He argues that human activities have altered the relative availability of specific nutrients in coastal waters in ways that favor toxic forms. For example, diatoms, most of which are harmless, require silicon in their cell walls, whereas other phytoplankton do not. Because silicon is not



WARM WATER PLUME (*green*), revealed in a satellite infrared image of sea-surface temperature, traveled from southern Maine into Massachusetts Bay in 1987, carrying *A. tamarense*. Remote-sensing techniques help investigators track blooms traveling within discrete water masses.



OUTBREAKS of paralytic shellfish poisoning affected more than twice as many areas in 1990 as they did in 1970. Some experts believe coastal pollution and shipping practices have contributed to the expansion.

abundant in sewage, but nitrogen and phosphorus are, the ratio of nitrogen to silicon or of phosphorus to silicon in coastal waters has increased over the past several decades. Diatom growth ceases when silicon supplies are depleted, but other phytoplankton classes, which often include more toxic species, can proliferate using "excess" nitrogen and phosphorus. This idea is controversial but not unfounded. A 23-year time series from the German coast documents a fourfold rise in the nitrogen-silicon and phosphorus-silicon ratios, accompanied by a striking change in the composition of the phytoplankton community: diatoms decreased, whereas flagellates increased more than 10-fold.

Another concern is the long-distance transport of algal species in cargo vessels. We have long recognized that ships carry marine organisms in their ballast water, but evidence is emerging that toxic algae have also been hitchhiking across the oceans. Gustaaf M. Hallegraeff of the University of Tasmania has frequently donned a miner's helmet and ventured into the bowels of massive cargo ships to sample sediments accumulated in ballast tanks. He found more than 300 million toxic dinoflagellate cysts in one vessel alone. Hallegraeff argues that one PSP-producing dinoflagellate species first appeared in Tasmanian waters during the past two decades, concurrent with the development of a local wood-chip industry. Empty vessels that begin a journey in a foreign harbor pump water and sediment into their tanks for ballast; when wood chips are loaded in Tasmania, the tanks are discharged. Cysts easily survive the transit cruise and colonize the new site. Australia has now issued strict guidelines for discharging ballast water in the country's ports. Unfortunately, most other nations do not have such restrictions.

The past decade may be remembered as the time that humankind's effect on the global environment caught the public eye in a powerful and ominous fashion. For some, signs of our neglect come with forecasts of global warming, deforestation or decreases in biodiversity. For me and my colleagues, this interval brought a bewildering expansion in the complexity and scale of the red tide phenomenon. The signs are clear that pollution has enhanced the abundance of algae, including harmful and toxic forms. This effect is obvious in Hong Kong and the Inland Sea of Japan and is perhaps real but less evident in regions where coastal pollution is more gradual and unobtrusive. But we cannot blame all new outbreaks and new problems on pollution. There are many other factors that contribute to the proliferation of toxic species; some involve human activities, and some do not. Nevertheless, we may well be witnessing a sign that should not be ignored. As a growing world population demands more and more of fisheries' resources, we must respect our coastal waters and minimize those activities that stimulate the spectacular and destructive outbreaks called red tides.

FURTHER READING

PRIMARY PRODUCTION AND THE GLOBAL EPIDEMIC OF PHYTOPLANKTON BLOOMS IN THE SEA: A LINKAGE? Theodore J. Smayda in *Novel Phytoplankton Blooms: Causes and Impacts of Recurrent Brown Tide and Other Unusual Blooms.* Edited by E. M. Cosper, V. M. Bricelj and E. J. Carpenter. Springer-Verlag, 1989.

MARINE BIOTOXINS AT THE TOP OF THE FOOD CHAIN. Donald M. Anderson and Alan W. White in *Oceanus,* Vol. 35, No. 3, pages 55–61; Fall 1992.

DOMOIC ACID AND AMNESIC SHELLFISH POISONING: A REVIEW. Ewen C. D. Todd in *Journal of Food Protection,* Vol. 56, No. 1, pages 69–83; January 1993.

A REVIEW OF HARMFUL ALGAL BLOOMS AND THEIR APPARENT GLOBAL INCREASE. Gustaaf M. Hallegraeff in *Phycologia,* Vol. 32, No. 2, pages 79–99; March 1993.

MARINE TOXINS. Takeshi Yasumoto and Michio Murata in *Chemical Reviews,* Vol. 93, No. 5, pages 1897–1909; July/August 1993.

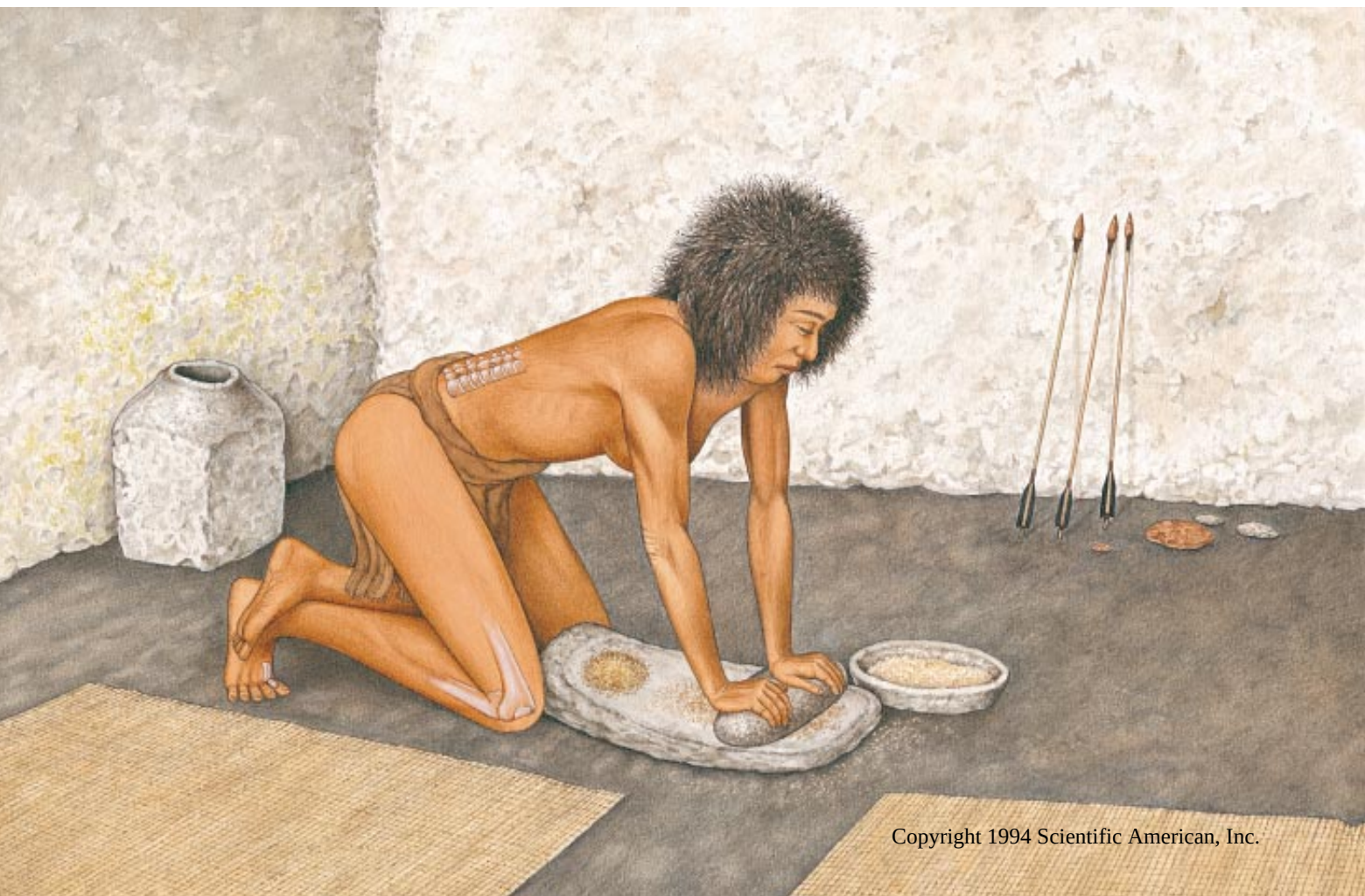
The Eloquent Bones of Abu Hureyra

The daily grind in an early Near Eastern agricultural community left revealing marks on the skeletons of the inhabitants

by Theya Molleson

Reconstructing how people lived in ancient times is like detective work: clues are scarce. Inferences must be made from spotty evidence such as bones, durable artifacts and the ruins of habitations. In my work as a paleontologist at the Natural History Museum in London, I knew that a collection of early Neolithic human bones had been brought to England from excavations at Abu Hureyra, in what is now northern Syria. The archaeological work was done in 1972 and l973—shortly before the site was due to be flooded by the reservoir behind the new Tabqa dam—by Andrew M. T. Moore, then at the University of Oxford [see "A Pre-Neolithic Farmers' Village on the Euphrates," by Andrew M. T. Moore; SCIENTIFIC AMERICAN, August 1979]. The skeletal remains of about 162 individuals—75 children and 87 adults, of whom 44 were female, 27 male and 16 of undetermined sex—have been identified from seven trenches dug at Abu Hureyra. The deposits span about 3,000 years.

It seemed to me and my colleagues that the bones might reveal details of the daily life of the Abu Hureyra people and therefore that of other Neolithic groups whose members had made the transition from hunting and gathering to an agricultural economy. The marks of life experience—some wrought by disease, some by work—can be imprinted on the bones and teeth of the skeleton. Close study has indeed yielded a fund of information that might not otherwise have been discovered, particular-



ly about the women of the community.

Abu Hureyra was inhabited in two different times. The first one was from roughly 11,500 to 10,000 years ago, just preceding the development of agriculture. The pre-Neolithic people of this settlement gathered a wide range of wild seeds, including lentils, einkorn, rye, barley, hackberries and pistachios. They also hunted the gazelles that migrated toward the Euphrates in the spring. The second settlement followed an unexplained hiatus of 200 years. The early Neolithic people of the later settlement cultivated a range of domestic cereals: emmer, einkorn, oats, barley, chickpeas and lentils. All these plants required preparation before they could be eaten. The preparation took much labor and time.

The record of this effort can be read in the bones of the people of Abu Hureyra. One of the first skeletal traits we noticed were signs of extra and sometimes excessive strains caused by the carrying of loads, most likely game, grain and building materials. The evidence was most conspicuous among the young. If adolescents are required to labor in this way, one can expect changes in the shape of the upper vertebrae. That is what we found. It is also probable that the loads were carried on the head: the hook-shaped parts of the vertebrae in the neck are enlarged, indicating that the bones developed a buttressing support. Otherwise, the neck might have wobbled under the weight of a heavy burden. In some individuals, we found degenerative changes in the neck vertebrae that may have arisen from injuries sustained by bearing weight.

These cases were not common. In fact, the general health of the people appears to have been good, except for bone deformities that turned up repeatedly: collapsed vertebrae (always the last dorsal one) and grossly arthritic big toes. These malformations were associated with evidence of muscular arms and legs. Clearly, the bones bespoke a demanding physical activity that was also injurious.

For a time, we actually entertained the idea that the people of Abu Hureyra had engaged in some sport or athletics, but crippled ballerinas seemed unlikely to have appeared during the Neolithic. We remained mystified until a colleague who was vacationing in Egypt noticed that the kneeling suppliants depicted on temple walls always had their toes curled forward. This observation suggested that some activity that involved kneeling had produced the pathology that we observed among the residents of Abu Hureyra.

During the excavations, Moore had found saddle querns in the rooms of the houses, abandoned after they had last been used. (A quern is a primitive stone mill for grinding grain by hand; a saddle quern is so named because it resembles a saddle in shape.) I was convinced that the kneeling action consisted of long hours spent grinding cereal grains on the saddle quern. Gordon Hillman of the University of London, who had worked on the plant remains from the site, was not so sure. He pointed out that removing the outer husk of the seeds by pounding them with a pestle in a mortar—another chore done while kneeling—would have been an essential step in preparing the grains. Probably both tasks were involved in creating the vertebral deformities, but it is unlikely that mortar-and-pestle work caused the toe deformities: the laborer could have changed positions while pounding but not while grinding.

So it was the preparation of grain for eating that was the most demanding and labor-intensive activity of the settlement, as it still is in many places. The grain had to be pounded every day because the seeds would not keep once they were dehusked. The dehusking with mortar and pestle and the subsequent grinding in a saddle quern would have taken many hours. What we had found on the bones, then, were the telltale signs of long hours spent at such labor. Also evident were marks of injuries, perhaps caused by using the saddle quern with too much enthusiasm or haste.

Querns and rubbing stones found at Abu Hureyra suggest how such wear and tear came about. The querns were set directly on the ground rather than mounted on a plinth or other raised structure, a practice followed in later times (debris surrounding the querns supports the conclusion that each was found where it had been used). Thus, the individual using the quern would have had to kneel.

Picture the operation. The grinder puts the grain on the quern and holds

THEYA MOLLESON works in the paleontology department of the Natural History Museum in London, where she does research on the effects of the environment on the human skeleton, both in life and after burial. She also lectures on human osteology at Birkbeck College of the University of London. Molleson studied geology at the University of London and social anthropology at the University of Edinburgh.

GRINDING GRAIN on a saddle quern, a daily task for Abu Hureyra females, put strain on several of the joints. On her knees, the woman repeatedly pushed the rubbing stone forward and then pulled back to her starting position. The activity, taking up several hours a day, affected particularly the bones shown above: the big toe, the spine and the leg. The toe is hyperflexed and damaged; the spine shows bony growths of the vertebrae; the leg, pictured with the femur (thigh bone) at the top and the tibia (shin bone) below it, has a buttress along the shaft of the femur and bony growths at the knee.

the rubbing stone with both hands. On her knees (yes, it was women's work, as we shall see), with toes bent forward, she pushes the stone toward the far end of the quern, ending the stroke with her upper body almost parallel to the ground so her arms are at or near the level of her head. On reaching the far end of the quern, she jerks back to her starting position. I call this part of the grinding action the recoil. The movement that raises the arms as the grinder pushes forward employs the deltoid muscles of the shoulder. During this stroke, the arms also turn inward, a motion accomplished by the biceps muscles.

It is precisely the places where the deltoid muscles attach to the humerus (the long bone of the upper arm) and the biceps muscles to the radius (one of the two forearm bones) that are markedly developed in these individuals. The overdevelopment of the muscles was symmetrical, affecting both arms equally. On the forearm of these individuals, the radial tuberosity—the bulged area of the radius where the biceps muscle attaches—is particularly noticeable.

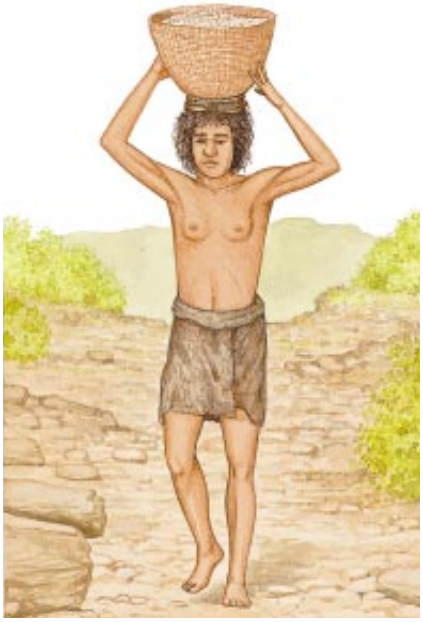
Kneeling for many hours strains the toes and knees, whereas grinding puts additional pressure on the hips and, especially, the lower back. The characteristic injuries we found on the last dorsal vertebra were disk damage and crushing. Such injuries could occur if the grinder overshot the far end of the saddle quern during the forward push or recoiled to the starting position too quickly or vigorously.

During grinding, the body pivots alternately around the knee and hip joints. The movement subjects the femurs (thigh bones) to considerable bending stresses. These bones thus develop a distinct buttress along the back to counteract the bending moments imposed from the hip and the knee as the weight of the body swings back and forth across the saddle quern. The knee also takes a lot of pressure because it serves as the pivot for the movement. Thus, the joint surfaces enlarge. All these effects appear on a set of bones we studied. The femurs were curved and buttressed. The knees show bony extensions on their articular surfaces.

The feet are also subjected to heavy pressure as one grinds grain on a quern. The toes are curled forward to provide leverage, which is supplied in large part by the big toes. In the remains from Abu Hureyra, the first metatarsal joints of the toes are enlarged and often injured. There are also signs of cartilage damage: smooth, polished surfaces at the metatarsal joint indicate that bone had rubbed on bone. In some individuals, a gross osteoarthritis had developed. In one case, the right big toe is much more severely affected than the left. Although an infective origin for this condition cannot be ruled out, perhaps the grinder was in the habit of resting one foot on the other to relieve the pain. Just such a position is shown in a model, illustrated in J. H. Breasted's *Egyptian Servant Statues,* of an Egyptian woman using a quern.

The changes to the arm, thigh and toe bones that we observed affected the overall bone morphology. This result would come about only if the stresses had been applied to the bones for long

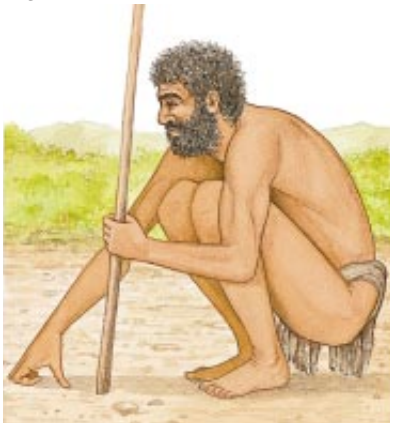
LOAD BEARING

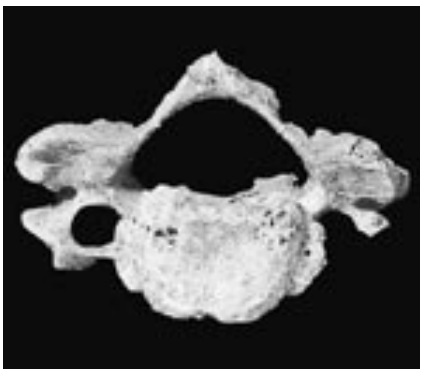


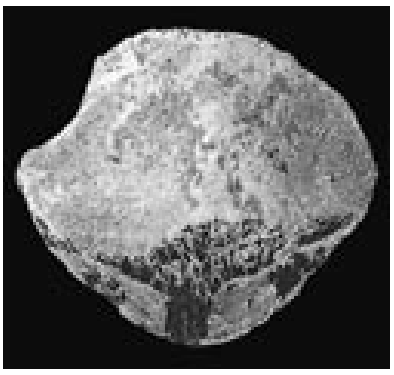
BONE ABNORMALITIES appeared among the people of Abu Hureyra as a result of the activities depicted here. Carrying loads on the head deformed the bones of the upper spine; the pitting on the vertebra indicates disk damage. Pounding grain in a mortar and pestle and operating a quern strongly developed the arm muscles, as reflected by the bulging in the two humerus (upper arm) bones (*top of photograph*),

USING MORTAR AND PESTLE

SQUATTING AT REST







hours daily while the individual was still growing. Travelers have observed such activity quite recently. Michael Aster writes in *A Desert Dies:* "Life in the [Saharan] oasis seemed to grind on at its own pace. For the women this was literally true, for they spent much of their time grinding grain on their hand mills.... I often watched Hawa as she placed a few grains at a time on the stone-base and let them trickle down as she ground them, sweeping the flour into a bowl every few minutes. After an hour or so her little daughter, aged about nine, would take over and begin grinding furiously. It might take several hours to produce enough flour for one meal."

We wanted to know whether members of both sexes ground grain at Abu Hureyra. Finding the answer proved difficult. The skeletons were so fragmented that we had to devise a way of determining the sex of an individual from the specific bones that showed the changes we believed resulted from using a saddle quern. Measurements of the first metatarsal bone of the foot demonstrated that it was generally larger in males, and by this means we could see that most of the bones showing the saddle-quern effects were from females.

We concluded that the grain was usually prepared by the women and girls in the household. A rather loose division of roles thus appears among these early Neolithic people. The inhabitants of Abu Hureyra must soon have discovered that the most efficient way to operate was to divide up the work of supplying food. We can assume that the men hunted and, with the advent of agriculture, cultivated food plants. The women of the household took on the job of grain preparation—a laborious task, or rather a series of tasks, that occupied many hours a day and could lead to back, knee and toe injuries. These are the repetitive stress injuries of the Neolithic. There is no need to assume that this division of roles implies any inequality between the sexes or between roles—that comes later.

The women were not the only ones to suffer. The coarsely ground grain had an appalling effect on everyone's teeth. One precaution necessary with all grain products except sifted flour is careful sorting to remove hard kernels and small stones. The number of fractured teeth among the early Neolithic people of Abu Hureyra bears witness to a failure to do this sorting effectively—and probably to an absence of sieves. For the same reason, awns or glumes from the outer covering of the grains remained in the flour and occasionally became lodged between the teeth, causing gum infection. On the other hand, caries (tooth decay) was rare. Apparently the flour was not sufficiently refined or cooked (if it was cooked) to provide the right environment for the bacteria that cause cavities.

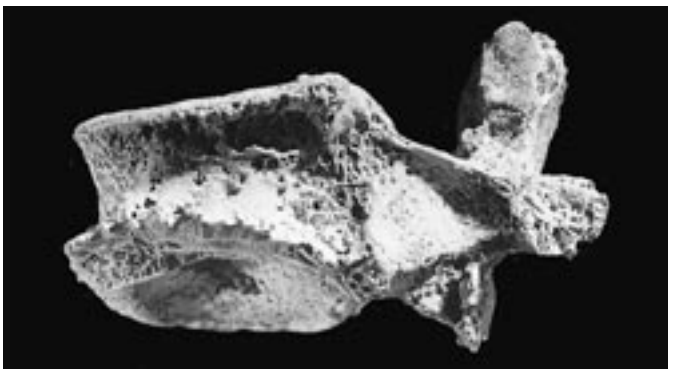
Fracturing was only one problem. The grains, even after being pounded and ground, yielded a hard meal that was exceedingly abrasive. Apart from the damage caused by rock powder from the grindstone, the flour itself rapidly wore down the teeth. Many people lost teeth at an early age. Moreover, scanning electron micrographs of teeth from Abu Hureyra show pits comparable in size to those that date stones and other hard objects make on the teeth of nonhuman primates.

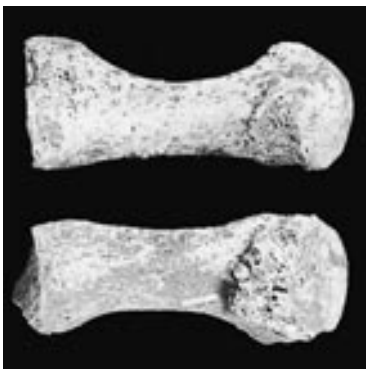
Something had to be done about the horrendous wear on the teeth. The archaeologists at Abu Hureyra had occasionally noticed the imprints of woven mats in plaster from later levels of the settlement. This finding was evidence that the people had by then mastered the skills of weaving. The invention of the sieve—an application of the principles of weaving—would have meant that grain could be sifted from grit and coarse chaff. Women in the Near East today can operate a sieve so deftly that they produce three piles on it: stones, chaff and grains. They then flip the stones into the palm of the hand. The result is fewer fractured teeth. We have no direct evidence of sieves at Abu Hureyra, but tooth wear is notably less severe in the later times.

Some way also had to be found to contain the harvested grain in order to bring it in from the fields.

where the deltoid muscles attach, and in the two radius (forearm) bones (*bottom*), where the biceps muscles attach. Squatting to rest put strain on the knee, resulting in this notched patella (kneecap). Using a quern damaged the last dorsal vertebra; wedging and pitting indicate crushing and disk damage. Also affected were the bones of the big toe; here there is wear near the right end of the upper toe bone and severe osteoarthritis near the right end of the lower bone.

USING QUERN





Baskets may have been the solution. We noted strange grooves on the front teeth of individuals from the later levels at Abu Hureyra. In making a basket, three canes have to be maneuvered at once. Because the hands are occupied holding the first staves of the basket, the teeth are used to control the working canes. Clark S. Larsen of the University of North Carolina at Chapel Hill has illustrated how a modern Paiute Indian woman holds the canes between her teeth. The habit of weaving in this way forms grooves on the surface of the front teeth. The grooves are almost identical to those we have observed on the teeth from Abu Hureyra.

The skeletal evidence for weaving and basket making is rare among the bones we studied, presumably because the skills for those crafts were confined to a few people. Those individuals are all from one part of the settlement, which suggests a craft area. Such specialization would be a natural outcome of any division of roles. Role specialization allows the development of expertise, speed and improved technology. If an expert is relieved of the need to produce her own food, she can manufacture more than enough sieves or baskets to supply the community. Any surplus can be used in trade.

From a different part of the settlement came evidence for another group of craftswomen. We noticed that several jaws found there have enormously enlarged joint surfaces, together with extremely uneven wear on the teeth. To display this pattern of wear, the teeth must have been subjected to immense crushing forces that abraded the lower teeth on the outside and the upper teeth on the inside. In some cases, the wear extends right down to the root.

Tetsuya Kamegai of Iwate Medical University in Japan has found similar changes among Maori people who chew plant stems to make fiber string. Some years ago J. D. Jennings of the University of Utah described the marks on quids chewed by worn teeth. The quids, made by people of the same epoch as the Abu Hureyra community, are found by the thousands at Danger Cave in Utah. The cave yielded pieces of cord made of chewed bulrush stems and mats bound with the cord. I believe mats were being made in a similar way at Abu Hureyra, a view also supported by the impressions of matting found during the excavation.

Some 7,300 years ago the new technology of pottery making brought great changes to the community. Pottery vessels provided a container in which grains could be soaked and cooked. That made the cereals so much softer that wear on the teeth was significantly reduced, as can be seen in scanning electron micrographs.

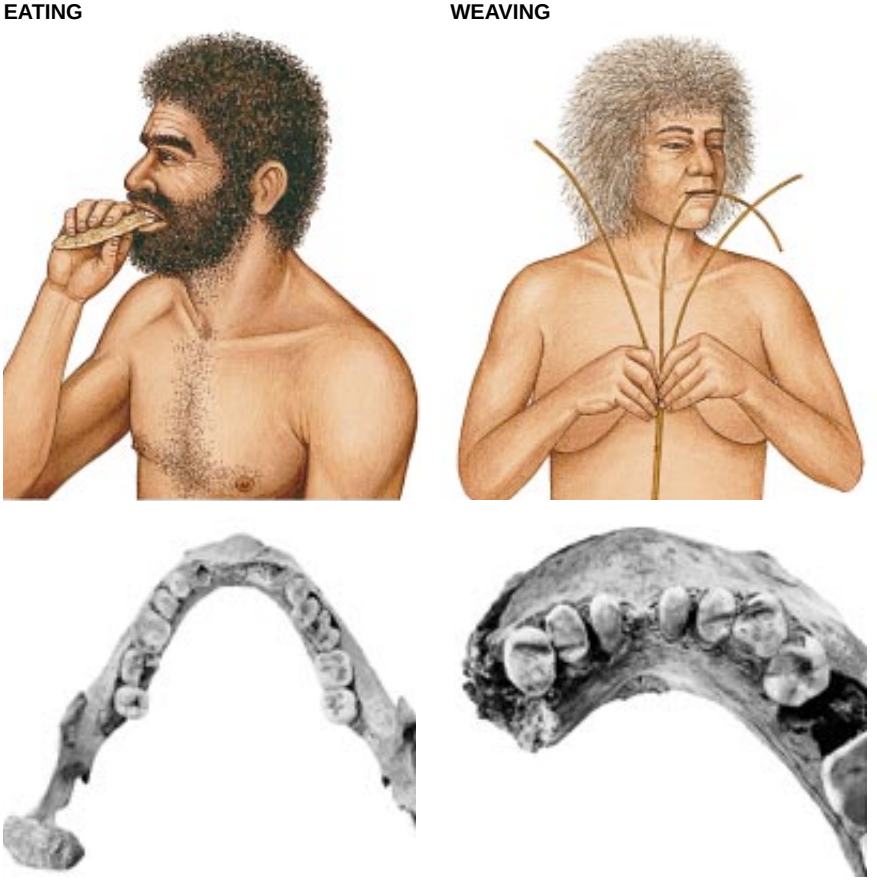
Cooked cereal is also tastier and easier to digest. Cooking releases the carbohydrates from the grain and makes them easy for the digestive system to absorb. One result was porridge, which soon had a dramatic effect on the community's population structure. A single consequence is evident in the unmended fracture of a woman's jaw; it is unlikely that she could have survived if a nutritious gruel or porridge were not available. Much more significant is that once porridge was available, women could give it to infants in place of breast milk. The mothers, too, consumed a diet quite rich in carbohydrates. The result of early weaning and better nourishment was to increase fertility substantially by reducing the interval between births.

This effect can be seen in the much larger proportion of infant skeletons recovered from the pottery levels compared with their percentage in the earlier strata. The proportion is so high as to suggest that infants were at increased risk of dying from disease, presumably because the rising population density gave more opportunity for pathogens to spread from one person to another. Some of the children have a thickening and pitting of the eye sockets, known as cribra orbitalia, that probably was the result of anemia following long-term infection by parasites.

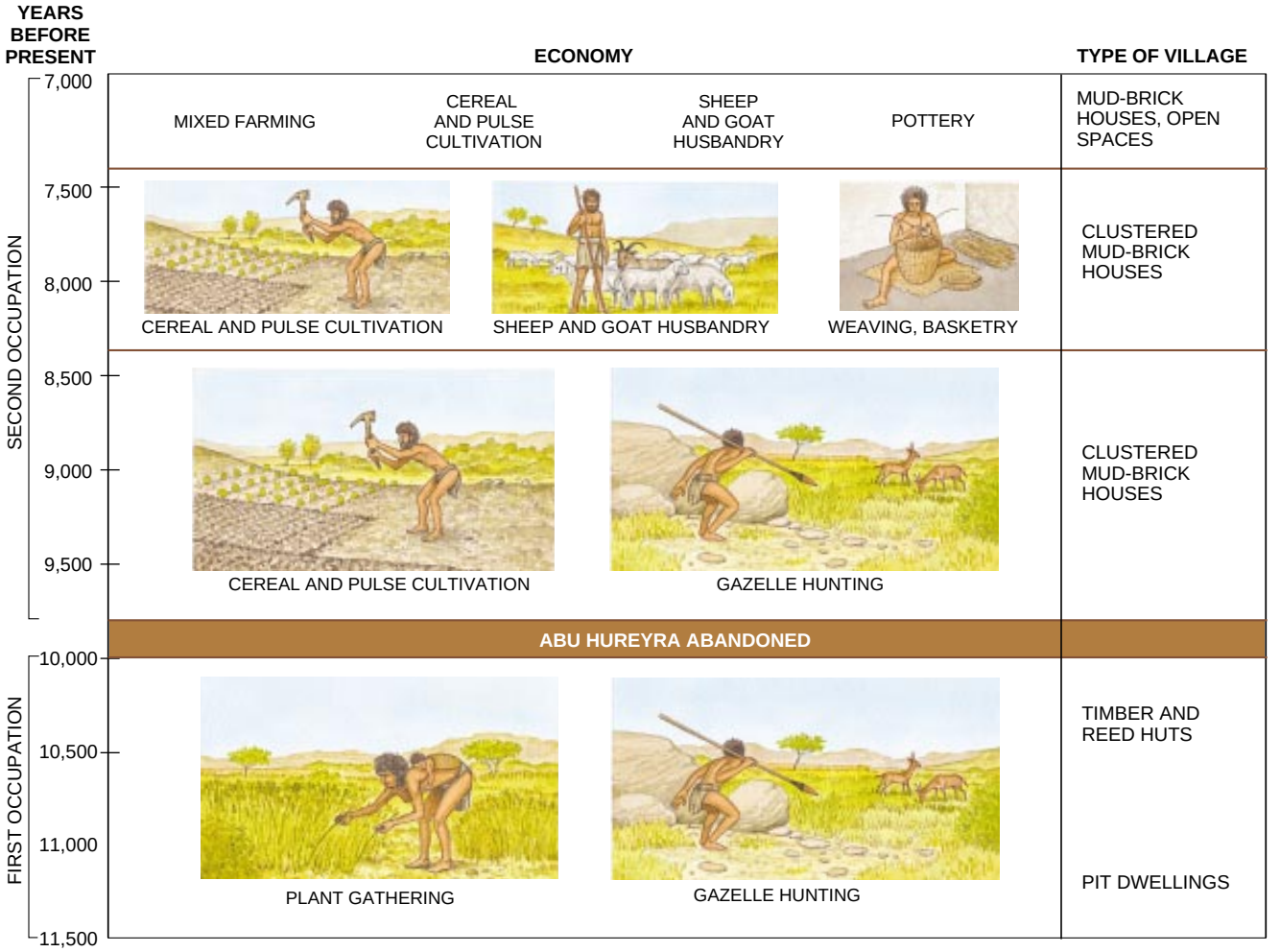
It is from the pottery levels that we find evidence of dental caries. The change in food preparation, with greater emphasis on cooked cereals made into bread and porridge, created sticky foods that adhere to the teeth and provide a medium for the growth of the bacteria that cause caries.

Abu Hureyra was abandoned about 7,000 years ago, as were many other Neolithic sites in the Near East. One cannot say why; disease, famine and climatic change are all possibilities. Abu Hureyra, although it was a structured society, remained egalitarian to the end—at least in terms of burial practices. But during the Neolithic, roles probably became more defined and more circumscribed.

The incorporation of role in the so-



TOOTH WEAR was severe among the early Neolithic people of Abu Hureyra. The coarse flour produced by grinding on a quern abraded teeth. Pulling canes through the teeth while making baskets resulted in deep grooves.



ABU HUREYRA'S CHRONOLOGY extended through two different occupations of the site over some 4,500 years. The first occupants were pre-Neolithic people who lived primitively and did not farm. Early Neolithic people of the second occupation gradually came to the cultivation of crops, the domestication of animals and such crafts as pottery and basket making.

cial fabric is reflected in the burial practices. The dead were buried under the floors of the houses or in pits in the yards outside. Many more women than men are buried in the rooms. This was their domain, where they had lived and worked. The women, it seems, had specific parts of space bounded by the limits of the house; their territory was a frame for their activities. John Gold of Oxford Brookes University sees this territoriality as a fundamental expression of social organization. The role boundaries established in life were maintained after death. The skeletal changes indicating how women spent their days—grinding, spinning, making baskets and mats—reflect a commitment in terms of time and economics that constitutes role specialization.

The very division of roles may have encouraged people not immediately immersed in preparing food to develop crafts. Crop cultivation, in fact, created its own challenges. Water had to be conveyed to seedlings for irrigation; animals had to be kept from destroying the crops; and harvested grain had to be transported. These problems provoked the exploration of technologies to solve them. Vessels, fences and baskets were devised, and certain people became expert in making them.

At Abu Hureyra, we see a progression of changes that can be understood in the light of such innovations. The improvements brought problems that called for further innovations. There was a constant progress toward a better life—a striving that continues to this day. Abu Hureyra represents the first step on the path toward civilization. But signs of wealth, class, elite institutions and scholarship have not been found in this settlement. We must look for them elsewhere.

FURTHER READING

THE EXCAVATION OF TELL ABU HUREYRA IN SYRIA: A PRELIMINARY REPORT. A.M.T. Moore in *Proceedings of the Prehistoric Society,* Vol. 41, pages 50–77; December 1975.

DENTAL MODIFICATIONS AND TOOL USE IN THE WESTERN GREAT BASIN. Clark Spencer Larsen in *American Journal of Physical Anthropology,* Vol. 67, No. 4, pages 393–402; August 1985.

SEED PREPARATION IN THE MESOLITHIC: THE OSTEOLOGICAL EVIDENCE. T. Molleson in *Antiquity,* Vol. 63, No. 239, pages 356–362; June 1989.

DENTAL EVIDENCE FOR DIETARY CHANGE AT ABU HUREYRA. Theya Molleson and Karen Jones in *Journal of Archaeological Science,* Vol. 18, No. 5, pages 525–539; September 1991.

DIETARY CHANGE AND THE EFFECTS OF FOOD PREPARATION ON MICROWEAR PATTERNS IN THE LATE NEOLITHIC OF ABU HUREYRA, NORTHERN SYRIA. T. Molleson, K. Jones and S. Jones in *Journal of Human Evolution,* Vol. 24, No. 6, pages 455–468; June 1993.

A Global View

by Marguerite Holloway, *staff writer*

Improving women's health means overhauling attitudes toward sex and addressing hidden epidemics, such as domestic violence

Women around the world share a common health problem, whether they are missing from U.S. clinical drug trials or are nowhere to be seen in a malaria clinic in Thailand. Social customs, health policies and a largely male medical community have tended to treat women as wives or wombs and little else. Consequently, as chronicled in report after report, aspects of women's health—including nutrition and aging, responses to illnesses such as AIDS, and diseases that plague only women—have received minimal attention or funding. Gaps pockmark the formal knowledge of women's health.

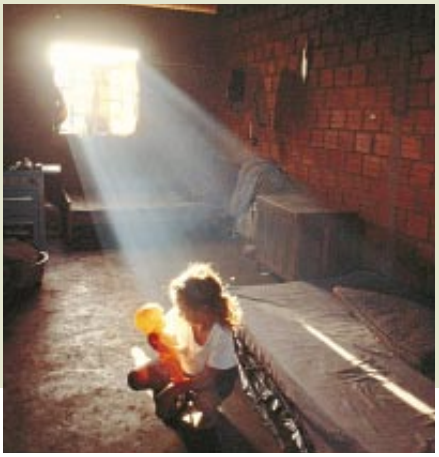
Female politicians, petitioners, patients and practitioners are beginning to cut away this legacy of prejudice. By widening the aperture to view all components of health, they have called for an approach to care that is based on the entire life cycle—one that would confront with equal rigor malnutrition in young girls, cervical cancer in middle-aged women and osteoporosis in older women. This strategy has exposed what could prove to be some of the most intractable issues endangering women's well-being. They are the least acknowledged threats: domestic violence, the damaging effects of illegal abortion, sexually transmitted diseases and genital mutilation.

Somewhat paradoxically, these issues ultimately have less to do with women's health in and of itself than they have to do with society as a whole. By seeking to redress an imbalance, physicians, women's health advocates and family planners have focused on areas that have been ignored not just for women but also for men. Upgrading women's health entails, in large part, reexamining sexuality and cultural mores; it entails improving male and adolescent sexual health; it entails involving men. Put simply, securing women's health means social transformation.

The challenge of domestic violence indicates just how extensive a change is needed. Violence—namely infanticide, sexual abuse, rape, battering and bride burning—is perhaps the most universally ignored but pervasive problem facing women. Because few women report domestic violence, experts say estimates of incidence are unreliable: between 20 and more than 50 percent of women throughout the world may be abused. Regional data indicate an epidemic. One survey in Papua New Guinea found that 56 percent of married women in cities reported being battered, and 18 percent of wives had gone to a hospital because they had been beaten; in the countryside, 67 percent of wives had suffered from domestic violence. In Bombay, one of four deaths in women between the ages of 15 and

***FOUR GENERATIONS OF WOMEN** together embody the concept of a life cycle approach to health care. Medical policies often treat women as mothers or wives, ignoring other aspects of their well-being. Women and health care workers are demanding and achieving a fuller approach to care, one that is applicable to all stages of life.*

SONOGRAM of a fetus can reveal sex. Some couples in China and India use this information to abort females.



POVERTY can be more detrimental to girls, including this Brazilian child, than to boys. Girls are often less valued.

MALNUTRITION contributes to a death rate of 61.2 per 1,000 Haitian girls versus 47.8 per 1,000 boys.

24 is caused by "accidental" burning (a means of murdering a wife in order to get a higher dowry through another marriage). In the U.S. between 22 and 35 percent of visits to emergency rooms are for injuries caused by domestic violence. Battering may be the leading cause of injury to American women.

"People always ask me where it is worse—in India, where they burn women, or in the U.S.," says Lori Heise of the Pacific Institute for Women's Health in Washington, D.C. "I do not think these kinds of comparisons are constructive. You see the systematic abuse of women in many societies being tolerated." One incident that Heise often uses to illustrate this tolerance took place in Kenya in 1991 when boys at the St. Kizito boarding school stormed the girls' dormitory, raping 71 and killing 19. The principal was quoted as saying, "The boys never meant any harm to the girls. They just wanted to rape."

Such sanctioning of abuse—whether overt or implicit—is common everywhere. It is this attitude, rooted in social relationships between men and women, that health care workers are recognizing as a major threat to women's health. It is this attitude that has prevented women from successfully escaping their batterers, physicians from recognizing the extent of the medical problem and men from changing their behavior. "Violence is socially constructed," Heise says. "There is enough evidence to show that whatever disposition men may have toward aggressive behavior, it can be overcome."

The acknowledgment that women's health is predicated on societal transformation—and that the medical system generally ignores problems such as domestic violence—has arisen in large part from the women's health movement in the U.S. and from the field of international family planning.

Although American women have been advocating for better and more gender-tailored health care since the 1960s, not until a few years ago were some of these concerns incorporated into public policy. In 1990 the General Accounting Office published a report that detailed the absence of women and minorities from clinical trials of medication. Because many women—some of them pregnant—were taking these drugs, the absence of data on hormonal interactions or differences in metabolism suggested that the health of women and minorities was being ignored and thereby put at risk. In fact, although researchers are just beginning to look into these topics, sex-related responses to some antidepressants and anticonvulsants as well as racial variations in reactions to drugs for hypertension have been discovered.

A Legacy of Inattention

At the same time that women's absence from clinical trials was being decried, it became clear that health subjects specific to women, breast cancer and menopause among them, were also not being investigated with the same energy as were some male concerns. In addition, many studies on illnesses that affect both sexes—specifically, heart disease and AIDS—had not scrutinized the etiology or expression of the disorders in women.

In the case of AIDS, this inattention has been well documented. It is not that women with AIDS or those infected with HIV had received no attention before women's health became politically correct. A host of studies have examined HIV transmission from female prostitutes to their male clients, as well as the epidemiology of AIDS among this group of women. Other investigators have looked at prenatal transmission, studying how and when and with what degree of prevalence mothers infect their fetus or infant.

The consequences of viewing women solely in relation to men or to children have become tragically obvious. The proportion of women infected with HIV has increased more rapidly than has the proportion of men carrying the virus, in the U.S. and elsewhere. The World Health Organization (WHO) estimates that by the year 2000 between 30 and 40 million people will have been infected, half or more than half of them women. AIDS is already the leading cause of death for African-American women between the ages of 20 and 40 in New York State and New Jersey; it is also the principal killer of women in the same age group in sub-Saharan Africa.

Until 1992, despite the magnitude of this part of the epidemic, the list of conditions set by the U.S. Centers for Disease Control and Prevention used to define AIDS did not include gynecological symptoms. Because the search for alternative means of protection has only just begun, women remain reliant on men to use condoms. The recently marketed female condom may confer the same degree of protection that the male condom does—it still, however, requires male consent. Only recently have laboratories begun to investigate microbicides that women can administer directly and, if need be, privately. Researchers also hope to find a form of AIDS prevention that could permit pregnancy.

In the U.S. the medical community and the government have taken steps to ensure that women will no longer be marginalized. Because the onus for ignoring women's health appears to lie

GENITAL MUTILATION is practiced in some African countries, such as Egypt, where this procedure took place. About two million girls are mutilated annually.

TEENAGE MOTHER in Iowa is one of the one million U.S. girls who become pregnant each year.

HIV-INFECTED WOMAN is a member of the fastest-growing segment of the population carrying the virus—women.

largely with the traditionally male-heavy medical establishment, the profession is under review. Horror stories from female residents about their male colleagues lopping off a cadaver's breasts and discarding them as unimportant have set in motion curriculum debates; educators are weighing the merits of creating a specialty in women's health.

At the federal level, an Office for Research on Women's Health was established at the National Institutes of Health in 1990. Under the auspices of this office, a multimillion-dollar study of the effects of nutrition on the health of older women is scheduled to start next year. Clinical trials must also now include fair representation of women and minorities. This requirement is, however, under attack by some scientists on the grounds that meeting quotas could, in certain cases, undermine both the feasibility of the study and the validity of the findings.

What works in the U.S. may not in other cultural and political settings. As Beverly Winikoff of the Population Council in New York City points out, there has been no movement calling for the inclusion of women in clinical trials in developing countries. "In fact, it has been the opposite: Do not experiment on women," Winikoff notes. A legacy of being used as subjects in unethical birth-control experiments has made women skittish. In one such trial, Mexican-American women were told they were being administered a birth-control pill, when in fact many were being given a placebo. Those who became pregnant and requested an abortion were not permitted to have the procedure.

Shifts toward the focus on a life cycle perspective in the international community have arisen instead in large part from women's dissatisfaction with the scope and quality of family-planning services. Many programs established in the 1960s to curb population growth by providing contraception or sterilization evolved into the principal channels through which to treat women's health. As in the U.S., however, the view of women embodied in these services primarily derived from their role as reproducers. At first, the goal was to ensure that they had access to birth control so as to curb population growth. Later, in the early 1980s, some of these programs were expanded in an effort to lower infant mortality.

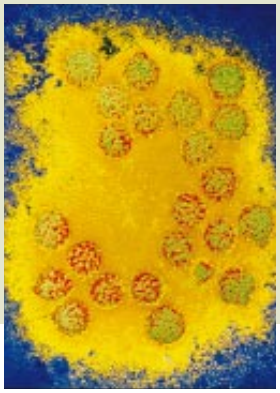
Over time it became obvious that services designed to achieve population-control goals did not necessarily consider women's reproductive health needs. Some programs were perceived as coercive or were incomplete in following up any adverse side effects caused by contraceptives. "It also became quite clear that it was only the child's health that mattered," notes Adrienne Germain, vice president of the International Women's Health Coalition in New York City. "The mother counted only as a means of ensuring it."

There were, and continue to be, grievous ramifications of this view—ramifications that reveal the paucity of medical information on women's health. Many practitioners, projecting industrial nations' experience onto Southern Hemisphere settings, assumed that what was good for the mother was good for the child, explains Deborah Maine, an epidemiologist at Columbia University. For instance, improving the nutritional intake of a fetus seems, in a cultureless context, a good idea. But in an impoverished region of Latin America, many girls are stunted as a result of being inadequately fed when they were younger, and their pelvic bones and birth canal remain small. Thus, without access to modern health services, a large baby can mean death for the mother.

Despite such mistakes, the global family-planning campaign to improve children's health and to lower fertility has been successful. Between 1960 and today, infant mortality fell 43 percent, at the same time the proportion of married women in developing countries using some form of contraception increased from 9 to 51 percent. Still, maternal mortality remains high. Some 500,000 women die every year as a result of complications during pregnancy or delivery; an estimated 50 million suffer from related complications. "I have been in a district hospital in Bangladesh where they will do surgery for accidents or fights among men, but not cesarean sections" that are needed, Maine says. "It is not a high priority. Well, that can kill you." And it clearly does. The difference between maternal mortality rates in developing countries as opposed to rates in industrial countries continues to be significant: an average of 420 deaths per 100,000 births versus an average of 26 per 100,000.

Unsafe abortions account for between 25 and 40 percent of this mortality. As many as 60 million abortions are performed annually, at least 50 percent of them clandestinely in the 100 or so countries where the procedure is illegal or severely restricted. Many of them are sought by teenagers and young unmarried women, who have little access to contraception because family-planning programs were designed for married women. In many countries, providing contraceptives to teenagers is illegal.

"Most of the attention on women's health is paid in the first few months of life when there is attention on infant mortality, then women disappear until



PAPILLOMAVIRUS is associated with cervical cancer, which kills 200,000 yearly.

ANEMIA PATIENTS in Ethiopia wait for treatment. Some 450 million women in developing countries are anemic.

PRIMARY HEALTH CARE in western Kenya is typical for many rural areas. Providing fuller services for women will entail expanding the entire delivery system.

we are concerned about their reproductive life, and then they disappear again," concludes Maureen Law of the International Development Research Center in Ottawa and chair of the Institute of Medicine report on women and nutrition in sub-Saharan Africa that will be released this fall.

Disappearing Women

Indeed, as health care advocates and researchers point out, women and girls simply disappear, and not just from the minds of policymakers. A 1992 study in Bombay found that of 8,000 abortions performed after parents had determined the sex of the fetus, only one fetus was male. Beyond selective abortion and infanticide of girls, boys often receive preferential treatment. Development experts, notably Amartya Sen of Harvard University, have determined that some 100 million women are missing worldwide—the result of many sociocultural factors, including poor health care, nutrition and poverty.

Given equal care, the ratio of female-to-male survival rates has been found to be as high as 1.15. In places such as Pakistan, however, the mortality rate of girls is 1.5 times higher than that of boys—the result of better nutrition for boys. Throughout the world about 450 million women are stunted because of malnutrition, as opposed to 400 million men. Monica Das Gupta of Harvard found that boys in India were 50 times more likely to be treated for malnutrition than were their sisters, even though the condition was four to five times more prevalent in the girls. In Punjab, according to the United Nations, only 14 percent of low-income boys weigh less than 70 percent of what is considered normal for their age, compared with 50 percent of the girls.

The inability of family-planning services to deal with malnutrition, maternal mortality and other aspects of women's health has led many specialists to push for more realistic, and higher quality, care. Instead of "providing family planning and then trying to show that it is good for women's health needs, you have to start with the needs and work from there," Maine comments. Rather than focusing on fertility, women and doctors argue that women's health is best served if treatment is comprehensive and takes into account the patient's social situation. Tackling maternal mortality in Guatemala, for instance, will make little difference if girls are not adequately nourished from birth.

By the same token, dispensing medication for malaria means nothing if women are not showing up at the clinic. Work by Carol Vlassoff of WHO suggests that some tropical diseases, such as schistosomiasis and malaria, may not be adequately understood or treated in women. Although the incidence appears to be the same for both sexes, studies have found that men are six times more likely to request medication or care. Vlassoff notes that because women or girls are frequently not highly valued and often have no political or economic power, the health care system makes little effort to reach them. In many African and Middle Eastern countries, women must obtain their husband's permission to visit health services. (In contrast, American women appear to visit their physicians more frequently than men do. Researchers attribute this phenomenon to gynecological and obstetric care, women's role as the family health provider and the socialization of men, which deters them from seeking treatment.)

Thus, if a woman's social context—including her poverty, family structure or lack of political power—is not taken into account, it is argued, health needs can never be fully met. Countless studies have demonstrated correlations between improving women's educational opportunities and financial clout and improving their health and that of their children. For instance, a strong link has been found between literacy and infant mortality. One study found that 40 of 1,000 children born to literate mothers died, as opposed to 120 of 1,000 born to illiterate mothers. The U.N. estimates that 597 million women are unable to read, compared with 352 million men.

The idea of medicine as a culturally tailored continuum of care forms the core of the life cycle approach. "A lot of interventions for women's health are best accomplished early," says Anne Tinker, senior health specialist at the World Bank and editor of a report, *Women's Health and Nutrition: Making a Difference.* By considering the nutritional intake of a young girl, health practitioners can ensure safer pregnancy and delivery later. Encouraging young women to delay pregnancy until they have finished school can lead to better employment opportunities and health.

The implementation of this philosophy can be seen in the changing emphasis of international bodies such as WHO, the World Bank and the U.S. Agency for International Development. It can also be seen in the explosion of women's health groups around the world. Representatives from such groups strongly influenced the guiding document drafted for the U.N.'s International Conference on Population and Development

DOMESTIC VIOLENCE is a leading cause of injury to women. Up to 35 percent of women's visits to U.S. emergency rooms are to seek treatment for battering.

CARE BY A MIDWIFE is often the only prenatal service provided to women in Ghana.

FAMILY-PLANNING CLINIC in India offers contraceptives. Globally, some 100 million women desire these services.

that will be held this September in Cairo. The theme of this conference, which takes place once a decade, reflects changes apparent in family planning. In 1974 the key to reducing fertility was articulated as the delivery of contraceptives; in 1994 the key is ensuring women's health and economic well-being.

Victory still lies a long way off. Many experts worry that expanding family-planning services to include treatment for, say, sexually transmitted diseases (STDs) may undermine the quality of care any one program could provide. "Conceptually, there is a lot of resistance. Services do not want to be identified with STDs. And they are politically afraid of abortion," Winikoff says. "Services want to be integrative but without actually changing." Nevertheless, health care experts recognize the fact that women have so little free time, so little access and often travel such great distances to reach clinics that they need one-stop shopping.

Although the Cairo conference will probably set aside money for model projects, a few examples already provide grounds for optimism that integration can work, assuming high-quality services shaped in consultation with female patients. The Bangladesh Women's Health Coalition is often cited as a successful example of integration. Women can get treatment for primary health needs and gynecological problems as well as care for their children. They also have access to contraception and abortion. The coalition has reported an annual increase in use by women in rural areas of 15 percent, an increase attributed to integrating services.

The life cycle concept has cast the net wide, encouraging the consideration of all facets of a woman's experience and how they relate to her health. But the net keeps getting snagged on one matter: sexuality. Sex has to be dealt with apolitically and with more thorough clinical care before STDs can be controlled, safe abortions can be ensured and adolescent pregnancy can be reduced. Such an approach collides with social convention, political sensitivity and religious stricture in many cultures around the world—the Vatican's efforts to oppose the Cairo conference are but one example. The conflict has become so acute that simple facts about behavior cannot be ascertained.

Even in the U.S., where the gathering of demographic information is a well-established industry, virtually no national surveys of sexual practices have been made since Alfred Kinsey's work in the 1950s. In 1987 the National Institute of Child Health and Human Development (NICHD) requested proposals for a national survey of sexual practices and attitudes among adults in the age of AIDS and STDs (meaning, in this case, infections such as *Chlamydia,* gonorrhea, syphilis and human papillomavirus). Yet as Edward O. Laumann of the University of Chicago and several colleagues discovered when they submitted a proposal and it was accepted, the invitation actually meant very little. The plan for the study was attacked by Senator Jesse Helms of North Carolina, and the NICHD backpedaled. Seven years later the call for proposals still stands—and, amazingly enough, it does not even include the word "sex." As Laumann describes in a recent issue of *Family Planning Perspectives,* "only such terms as the 'proximate determinants of fertility' or 'fertility-related behavior' are used."

The effects of this attitude on health are obvious. WHO estimates that about 250 million new cases of STDs occur every year, mostly in people between the ages of 20 and 24. One in four Americans will contract an STD at some time, a proportion that is among the highest in any developed country, according to a report by Patricia Donovan of the Alan Guttmacher Institute in New York City. Because many of these diseases are asymptomatic in women, they often go untreated until they cause such effects as infertility, pelvic inflammatory disease or death. Strains of the human papillomavirus can cause cervical cancer, which kills more than 200,000 women annually. Many STDs cause HIV to be transmitted more readily. Although men also suffer in great numbers, many clinics and programs have been established with them in mind. Furthermore, there is often less stigma associated with STDs in men than there is in women.

Sex and the Single Teen

The other victims of the need for more data on sexual behavior that could shape realistic prevention programs are adolescents. "People do not like to think of young people as being sexually active, and they do not want to think about the consequences," says Judith Senderowitz, founder and former director of the Center for Population Options. "And we have a new situation, a bio-social gap." As the age at marriage has increased and the age of menarche has decreased as a result of improvements in nutrition, Senderowitz explains, there is a window of about 10 years during which young people are unmarried and sexually active but often without access to sex education or birth control. According to WHO, more than half of the people infected with HIV

NURSING MOTHER AND HER CHILD work on a tree farm in Bhutan. Many women have no choice but to combine rigorous child care with physical labor.

BREAST CANCER, which afflicts one in nine women in the U.S., is frequently diagnosed through the use of mammography.

POVERTY AND OLD AGE can be linked for women, as they are for this Russian citizen.

worldwide are younger than 25 years.

The benefits of being open about sexuality are clear in Sweden. "There is a consensus in Sweden that sex is natural—it is part of life," Senderowitz notes. Since 1942 sex education has been a part of many school programs; since 1956 it has been mandated. As the rate of abortions rose 43 percent in American girls under the age of 20 between 1975 and 1981, the rate fell 30 percent in Sweden. Its rate of STD infection is one of the lowest in the world.

In contrast, in the U.S., teenagers now make up one fourth of those infected with STDs. Sex education in the U.S. varies from state to state; some programs do not even review contraception. "Many of the existing sex education programs provide too little information too late," asserts Shanti R. Conly, a researcher at Population Action International and co-author with Stephanie L. Koontz of a recent report on adolescent health, *Youth at Risk: Meeting the Sexual Health Needs of Adolescents.* In the developing world the situation is much worse because sex education is virtually nonexistent. In one case cited by Conly and Koontz, one fourth of the teenagers surveyed in Sri Lanka thought a woman could become pregnant if she wore clothing that had been worn by a man. "It is very important for national governments to recognize this as a public health issue and to get over the moral and social taboos. I mean, this is a crisis," Conly urges.

Looking at sexual behavior in young women requires consideration of the subject in young men. Bruce Armstrong of the Young Men's Clinic at Columbia University's School of Public Health points out: "I think there is a growing awareness that not to include men—given what we know and given the biology of reproduction—is to really miss the big picture." Armstrong, who set up his program in the mid-1980s, goes on to note that "in many relationships men have a lot of influence over what behavior gets performed."

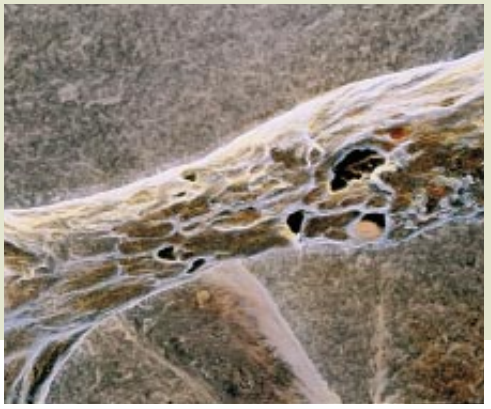
Male Responsibility

Similar insights have catalyzed a cautious but growing male responsibility movement. "No one wants to shift the burden to men, because that also shifts the power," Winikoff says. "But it is good to put on the table." Because the original family-planning programs focused on women's fertility, few sought to include men. Many women were forced to hide their use of contraceptives. One family-planning expert cites the example of women in Zimbabwe who have health workers hide birth-control pills under rocks in the fields where they farm. They do not want their husbands to know. "I guess we came up against the difficulty of increasing contraceptive use and facing up to the fact that women do not make these decisions on their own," says Susheela Singh, associate director of research at the Alan Guttmacher Institute. "There is something of a philosophical shift taking place in the women's health field. The thinking is that women should not be taking on so much of the burden."

Currently roughly 75 percent of sterilizations take place in women, even though tubal ligation can be risky and expensive, when compared with vasectomies. Yet vasectomy is increasingly being promoted as contraceptive services at least try to involve men in the decision-making process. As Jill Gay of the Pan American Health Organization says, "Men are not getting the right treatment. Men are excluded from reproductive health care. There are no workshops for men, none on male sexuality, none on male responsibility." One program in Egypt made such an effort, and within a year contraceptive use doubled. Other programs have sought to educate men about the importance of nutrition during pregnancy. Efforts in Thai refugee camps brought about health improvements in mothers and infants.

Men are also being included in the Demographic and Health Surveys, which track changes in international fertility rates. Such data gathering has led to the observation that in many places one man can be responsible for creating more children in his lifetime than one woman may be. As Aaron Sachs, a researcher at the Worldwatch Institute, reviews in a recent report, *Men, Sex, and Parenthood in an Overpopulating World,* men in Cameroon who have children father an average of 8.1—often by different partners—by the time they reach age 50, whereas women who have children average 4.8.

Even sperm are beginning to shoulder some of the responsibility. The egg is no longer solely liable for the health or impairment of the fetus. This view had, in some cases, put the fetus before the mother, jeopardizing women's right to choose where to work. (This topic came before the U.S. Supreme Court in the case of Johnson Controls. The defendant corporation had banned women from lucrative work because of concern that exposure to lead might damage a potential fetus. In 1991 the court ruled against Johnson Controls.) As more investigators concentrate on the male role in fetal health, it is becoming clear that male gametes also appear responsible for genetic defects. Furthermore, they may be just as susceptible to occupational and lifestyle hazards as



OSTEOPORITIC BONE is less dense and more friable than healthy bone. In the U.S. more than 24 million women suffer from this condition.

LIFE CYCLE approach to health care would ensure that good nutrition in girls, such as this infant from Sri Lanka, would lead to better health when they are grandmothers.

ova are. Benzene, lead and some pesticides are among the compounds that may impair the viability of sperm and may then, in turn, cause stillbirths or developmental problems.

By turning attention to men as well, women are showing that the protection of health is multifaceted and that neither sex bears all the responsibility. This shift in gender relations is reflected in another transformation that is under way in the field of women's health: a change in cultural relations. There is a growing awareness that edicts or policies from developed countries are often doomed to fail when imposed on developing countries.

Female genital mutilation offers one particularly powerful example. Every year about two million girls in at least 26 African countries are clitoridectomized or infibulated—the process in which the clitoris as well as the labia majora and minora are destroyed—often without anesthesia. In addition to mutilation, the procedure can lead to infection, septicemia and death. The cultural origins of the practice are varied, but some themes recur from place to place. Genital mutilation is seen as a rite of passage that ensures marriage. Many women desire it because they have been socialized to view their genitals as unattractive or because they think it is religiously mandated (which it is not) or because their mothers and role models have undergone the ritual.

Although Nahid Toubia, a physician from Sudan who was formerly at the Population Council, says the comparison is problematic, a quick glance at the tradition of male circumcision can be illustrative. "A lot of people in the U.S. do not even want to talk about it, because it is linked to their own beliefs. Which just shows you how people in Africa behave," Toubia notes. "It is part of their belief system, so they cannot fathom living life without doing it." Some 85 percent of men in the U.S. are circumcised—although the American Academy of Pediatrics does not recommend or discourage the procedure. One group of physicians suggests that circumcision confers health benefits; others argue that good hygiene eliminates this advantage and that removing the foreskin reduces sexual pleasure.

Health surveys are just beginning to incorporate questions on genital mutilation. As is true in so many other areas of women's health, there are few rigorous data to go on. Nevertheless, Toubia says, between 85 and 114 million women alive today have undergone the operation. (Female genital mutilation was practiced in Europe and the U.S. as a remedy for hysteria, lesbianism and masturbation, Toubia reports.)

Not only has the recognition of the scope of such mutilation been slow, but confronting it has revealed the limitations of well-intentioned efforts to improve women's health. Although organizations of African women have been trying to combat the tradition since the 1960s, Western politicians and women's groups have recently alighted on female genital mutilation as the most hellish practice around. Last October, Congress drafted legislation, H.R. 3247, to make it illegal in the U.S., as it is in France and the U.K. Because of increased immigration, there have been anecdotal reports of clitoridectomy and infibulation in the U.S. and in Europe.

According to many women in the health field, and according to many Americans of African origin who are working on the matter, punishment and imperialistic reactions are exactly the wrong approach. "It is a gruesome and extreme form of what women all over the world suffer: male-dominated societal control over their bodies. This one is so direct and seems easy to deal with, which is why people have focused on it," Toubia notes. But "we are talking about a perceptual shift."

Improving women's health appears to depend on just such a change in perception, which is why the hardest work for women's health advocates and physicians may be yet to come. Ensuring women's health involves more than drafting legislation to guarantee the inclusion of women in clinical trials or to allocate funds to study and treat any sex-specific aspects of a tropical disease. Instead the future of women's well-being seems to center on the world's approach to sex and to cultural differences. Without the involvement of men and without a revolution of openness in considerations of sexuality, society cannot expect to conquer the ultimate threats to the health of everyone—male and female. "This is not about vaccinating people," Toubia remarks. "It is about tackling sociocultural ideas."

FURTHER READING

THE HEALTH OF WOMEN: A GLOBAL PERSPECTIVE. Edited by Marge Koblinsky, Judith Timyan and Jill Gay. Westview Press, 1993.

POPULATION POLICIES RECONSIDERED: HEALTH, EMPOWERMENT, AND RIGHTS. Edited by Gita Sen, Adrienne Germain and Lincoln C. Chen. Harvard School of Public Health, 1994.

WOMEN AND HEALTH RESEARCH: ETHICAL AND LEGAL ISSUES OF INCLUDING WOMEN IN CLINICAL STUDIES. Edited by Anna C. Mastroianni, Ruth Faden and Daniel Federman. National Academy Press, 1994.

Welfare Plastic

Electronic benefits may become a big business

Citicorp has entered a business traditionally served by the corner check-cashing outlet. Recipients of public assistance who used to redeem a welfare check at a window in a local storefront may soon be using an electronic network supplied by the nation's largest bank holding company.

Citicorp and other financial institutions are starting to serve these unlikely markets because of state and federal efforts to deliver government entitlement programs on the network to the millions of Americans who do not even hold a savings passbook. The Clinton administration has embraced electronic delivery of benefits as a goal of Vice President Al Gore's National Performance Review, which outlined a blueprint for streamlining government.

In late May, Gore announced a five-year, nationwide program to furnish electronically a broad range of federal benefits, from food stamps to welfare to military pensions. "The plan is elegantly simple with a dramatic impact," Gore commented at a news conference. "It says we will create one card, like other ATM cards, to deliver all federal and state government benefits."

Gore could be talking real money. The administration has estimated that $112 billion of federal benefits could be made accessible annually by swiping a magnetic card through a retailer's point-of-sale terminal or by tapping into an automatic teller machine. That amount is equal to more than 60 percent of the transactions that VISA tallies each year. An interagency government task force set up to administer this program has estimated that it could yield savings of $195 million a year. Savings would be achieved by reducing the number of paper forms to be processed and using the efficiencies of commercial financial networks.

So-called electronic benefits transfer (EBT) could also turn into a substantial market for financial service providers. Through the government, banks and other institutions could find at little cost a large volume of customers they do not now serve. "There's not a lot of marketing involved; nobody has to run a 30-second commercial," says Brian Claire, senior director of Citibank EBT Services in Chicago, which has contracts to provide electronic benefits in Texas, South Carolina and New York State. The federal government is contemplating spending $83 million as its share in establishing a nationwide EBT program.

KATHERINE LAMBERT

CHECKING IT OUT: the state of Maryland has instituted a program for electronic delivery of food stamps* (shown here), *welfare and child support.

The banks see another advantage: the technology may help spur the advent of the noncash economy. The point-of-sale terminals used by retailers for food-stamp accounts might also be adapted for debiting the bank accounts of customers not receiving benefits. That could be a step toward weaning consumers from greenbacks and from paper checks and invoices.

Today recipients tap benefits accounts electronically in about 10 states, although plans to implement a system are being readied in more than half of the 50 states. Only in Maryland does the system operate statewide. Since April of 1993 Maryland's Department of Human Resources has distributed an average of $57 million each month in benefits, comprising food stamps, welfare and a child-support program, to 170,000 households. The "Independence Cards" distributed to these homes provide access within Maryland to 1,800 automatic teller machines and 3,300 food stores with point-of-sale terminals.

A card, swiped through the retailer terminal in the checkout line, serves as a key that unlocks a benefits holder's account. The recipient then punches into the terminal a four-digit identification code. If sufficient funds have been credited to a welfare or child-support account, a cardholder can request money from the cashier at some stores. Or a food-stamp account can be debited to purchase groceries. The card can also be used to withdraw money from a welfare or child-support account over a bank's automatic teller network.

Maryland has already achieved a measure of success. Abt Associates, a Cambridge, Mass., social sciences consulting firm, estimated that EBT achieved a modest savings over the existing paper system: issuing paper welfare checks

turns out to be less expensive than having recipients use automatic teller machines. That deficit is more than offset, however, by electronic food stamps. It costs more to issue and distribute food-stamp booklets than to have clients use supermarket point-of-sale terminals. Economies were realized in other areas, mostly for reduced fraud and error. In total, the estimated annual savings from electronic benefits amounted to $454,000, an almost 4 percent reduction in costs to run the undertaking. "The EBT program we evaluated was a success," says John A. Kirlin, of Abt Associates.

Maryland might have lowered program expenditures still further if retailers had to purchase their own point-of-sale terminals, if banks discounted fees because of the volume of transactions generated or if benefits recipients had to pay their own transaction fees.

The state also discovered the idiosyncrasies inherent in maintaining its benefits caseload on the network. Although the cards are designed to prevent food stamps or welfare checks from being stolen, the state ran into a few instances in which its own workers allocated themselves cards and cash balances. In addition, the cards can be easily lost or the magnetic validation stripe may wear off when a card is stored in a back-pocket wallet; the agency now issues as many as 6,000 replacement cards a month.

Training people who cannot read or who have no dealings with a bank is as much a challenge as preventing card losses. Deluxe Data Systems, which supplies the electronic networking services, had to recruit interpreters who spoke a Southeast Asian dialect that had no word for automatic teller machine. The loose translation for ATM amounts to "the big box with the little window." "This takes a type of technology that a lot of us have taken for granted and applies it to a group that has been passed by," says Tom McLaughlin, Deluxe Data's vice president of government services. Also, many food-stamp recipients prefer standing in line with a magnetic card rather than a stamp booklet. "It really has lifted people's self-esteem," says Dale E. Brown, project manager for benefits transfer for Maryland's Department of Human Resources.

Maryland has been heartened enough by the program results that it has decided to place point-of-sale terminals for paying bills in a few utility company and state housing department offices. It has even begun to test letting food-stamp clients pay for fruit and vegetables at a farmer's market through an electronic linkup by cellular telephone.

The Maryland experience suggests that EBT may help spur a gradual movement away from cash. Shortly after the point-of-sale terminals were installed, some 450 Maryland stores contracted with area banks to allow electronic payments for nonbenefits transactions, allowing customers to pay with a bank card or a credit card. "EBT puts an infrastructure in place for debit cards," McLaughlin says.

With the advent of a nationwide program, EBT will go to the 31 million benefits clients who receive food stamps or who do not maintain a bank account into which cash benefits are deposited directly. States administer some but not all federal benefits programs. But the goal of the Clinton administration to provide EBT will require that a cardholder be able to get funds across a state line. The federal government is now organizing a regional test of EBT in nine southern states.

As benefits go electronic, privacy will surely become an issue. Checkout scan-

From Swords to Mouse Ears

The parents of Mickey Mouse and Donald Duck are thinking of joining forces with the folks who brought you Fat Man and Little Boy. The vast entertainment empire founded by Walter Elias Disney has shown a keen interest in how it might put to good use a range of technologies developed at the immense Los Alamos and Sandia weapons complexes in New Mexico.

The unlikely collaboration might benefit both bomb maker and theme park owner. The two laboratories could perhaps achieve a high-profile partnership that would burnish their image as they inch toward a post–cold war identity. Meanwhile instrumentation, sensors, microchips and other gizmos from the innards of the New Mexican research compounds could help make the Disney parks run more efficiently or perhaps even lend high-tech cachet for revivifying aging displays at Epcot Center.

So far a contract to work together closely has yet to be signed, and details of the talks are sparse. In a reversal of roles, laboratory officials eagerly want to discuss an agreement that would trumpet the commercial relevance of their labors, but a reticent Disney organization has tried to keep specifics cloaked in Manhattan Project–like secrecy.

Yet this spring a group of Los Alamos researchers visited Disney World in Orlando, Fla., to demonstrate a testing technique—resonant spectroscopy—that might probe the structural integrity of park rides. It works by using a remote-sensing device that detects the vibrations, measured as the resonant frequencies, that are generated in steel and concrete by the motions of, say, a ride vehicle. A crack or other flaw causes an expected pattern of resonances to be altered.

Disney officials have pondered the potential of other laboratory offerings: specialized microchips that can be used as electronic timing devices for fireworks; advanced materials and fuels that increase the life or efficiency of rides; computerized control systems that add lifelike movement to a mannequin of Pluto or Goofy; tickets encoded with information about an individual patron; and laser remote sensing to help optimize the amount of moisture in the air of the hydroponics exhibit at Epcot Center.

As with other ventures between national laboratories and private industry, the final terms of an agreement may prove nettlesome. Despite efforts at streamlining these types of contracts, the Department of Energy, which funds the national laboratories, continues to wrestle with how to give private industry the exclusive access it wants to technology originally paid for from public coffers.

If Sandia and Los Alamos can strike a bargain with hard-nosed Disneyites, a new era of deal making between government researchers and entertainment and consumer products companies may open. Laboratory chemists under contract to McDonald's might apply their skills to make a low-fat hamburger—or maybe even a high-fat one—that tastes like real meat. Domino's Pizza could equip its delivery trucks with radar or infrared sensors to warn of impending collisions with other vehicles. Disney could take advantage of Los Alamos's supercomputing expertise to remake *Fantasia* in three or perhaps more dimensions.

For physicists ensconced in the high Southwest, these options may lack the thrill of setting off a multimegaton underground nuclear test. But, heck, they beat working at McDonald's. *—Gary Stix*

ners now identify eligible food-stamp purchases; nonfood items are prohibited under the program. But the computers that disburse food credits could potentially keep a record of a food-stamp recipient's shopping list. "You would try not to collect that data when setting up a privacy-safe system," says Marc Rotenberg, director of the Electronic Privacy Information Center.

Another simmering controversy is whether benefits recipients should be afforded the same protection against theft or misuse of their cards as are bank automatic-teller cardholders. A ruling by the Federal Reserve in February held that Regulation E, which makes a cardholder liable for no more than $50 if notification is given within two days of when the card is lost or stolen, should be extended to EBT users within three years. The delay in implementation of the regulation is intended to allow the federal and state governments to resolve the issue of who pays when a card is fraudulently used. Neither state governments or federal agencies nor banks or retailers want to foot these costs. States have predicted that the annual liability across the nation could total $500 to $800 million. "Regulation E gives recipients far more rights than they have today," McLaughlin says.

Welfare-rights groups, however, have lobbied hard for these protections, saying that without the regulation in place, a two-tiered system would emerge that would discriminate between rich and poor. "Our concerns are that recipients need to be protected from losses that occur through no fault of their own when going to the new system," says Barbara Leyser, a senior policy analyst at the Center on Social Welfare Policy and Law, an advocacy group for those receiving public assistance.

The states have labeled the liability question one that could become an obstacle in making progress toward a system that delivers benefits electronically. Distributing federal money to those in need will inevitably provoke debate whether a check is on the network or in the mail. *—Gary Stix*

Antigreen Greenies

Activists flag opposition to the royal and ancient game

Because fairways, greens and roughs can consume 150 acres and frequently require copious amounts of fertilizer and pesticide, the game of golf is an environmentally dubious symbol of membership in the world's business community. In the U.S. alone there are more than 14,500 courses, occupying an area bigger than the state of Delaware. An estimated 25,000 links dot the face of the globe.

But the royal and ancient sport, as it is known, has provoked an equally global backlash from environmentalists. They believe that space given over to doglegs and bunkers serves only to rob farmland, destroy national parks, waste precious water and increase the use of toxic pesticides and herbicides needed to keep greens green.

A worldwide antigolf campaign began last year when a group of linksophobes from Indonesia, Malaysia, India, Hawaii, Hong Kong, the Philippines, Japan and Thailand established the Global Anti-Golf Movement (GAGM), which is contracted to an acronym with the forbidding pronunciation of "gag 'em." This grassroots movement has broadcast its presence by declaring a World No Golf Day—and subsequently a World No Golf Year. GAGM has also militated for a moratorium on new golf course construction and called for a ban on the introduction of the sport into the Olympics.

The United Nations Security Council is not contemplating a special session on golf. But tensions between course developers and their opponents have at times degenerated into violence. In the newly developing countries of Southeast Asia, the game has become a symbol of the lifestyle of emerging elites that threatens the livelihoods of traditional farmers.

Par-3 holes, in fact, can displace farms. In Cijayanti, Indonesia, 30 miles south of Jakarta, farmers burned a construction encampment of a company, Light Instrumenindo, after the developer's security guards reportedly attacked a local villager. The company offered to pay a mere 60 rupiah (five cents) per square meter for the farmland there. One Indonesian activist, Dedi Ekadibrata, was sentenced this spring to 18 months in prison for provoking the protest and arson.

A Thai documentary, *Green Menace: The Untold Story of Golf,* chronicled how some golf courses have illegally drained large quantities of water from reservoirs. Once built, the courses provide scant employment, according to Thai activist Anita Pleumarom. The female caddies who do work there, she says, are often called on for extracurricular favors. "It's clear that they're not only carrying golf bags but are available for other services as well," Pleumarom says.

The trophy for the most bizarre golf-related spectacle goes to Vietnam. A hundred Vietnamese soldiers, metal detectors in hand, were reported to have combed the countryside in Song Be Province for unexploded bombs, a prelude to clearing this once bloody battleground for a golf resort.

The golf establishment, whose political epicenter lies in the U.S., has gone on the offensive by showcasing projects that minimize the disruption of land sculpted into fairway and green. The American Society of Golf Course Architects has cited a course designed

COURTESY OF THE TOURISM AUTHORITY OF THAILAND

THAILAND's more than 160 golf courses have come under fire from activists who have criticized the game because of environmental excesses.

by one of its members in Gaylord, Mich., called the Natural, that has only 32.2 acres, or 17 percent of its terrain, maintained as manicured turf; the rest has been left untouched. In spirit, the Natural constitutes homage to earlier practitioners of the sport. They would cut holes with a penknife and stick a feather in them. Rabbits or sheep were coopted as greenskeepers.

This strategy may not prevail in golf's new empire. Playing on an untended course may not carry over easily to the Thai or Indonesian countryside. The desire in those places to imitate the unearthly verdure of the chemical-fed greens and fairways of Augusta or Palm Springs will require healthy doses of pesticides and fertilizer. Komodo dragons, moreover, have never shown an appetite for imported American turf grasses. *—Gary Stix*

Rotaxanes

These organic molecules may serve as nanoswitches

In their quest to build viable, nanoscale devices, researchers have turned most often to the tools of solid-state physics, using methods such as lithography and vapor deposition to fashion ever shrinking switches, levers and ratchets from silicon. Now organic chemistry may be making a contribution to this field as well, in the form of a class of molecules called rotaxanes.

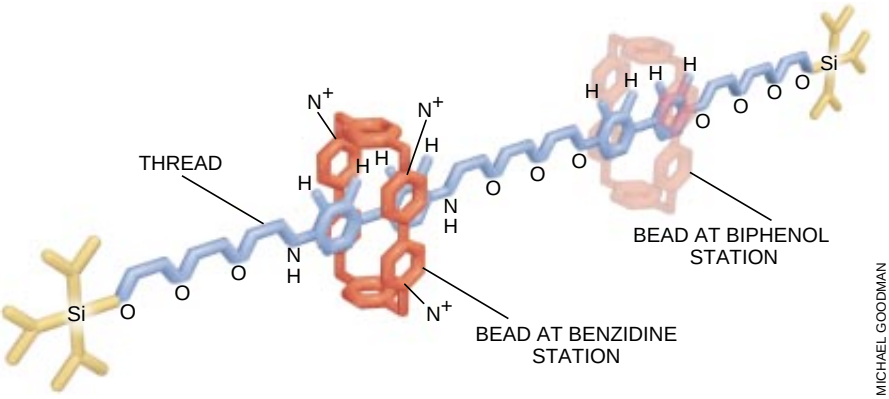
In general, these structures resemble a bead on a thread. At room temperature the bead, an electron-accepting cyclophane, shuttles rapidly along the thread, stopping at certain sections that can donate electrons to it. Chemists can use light, chemicals or electrical charges to control the bead's exact location. In this way, the interplay between the bead and the thread provides the basis for an ideal molecular switch.

Anthony Harriman and Andrew C. Benniston, both at the Center for Fast Kinetics Research at the University of Texas at Austin, have created what could be a rotaxane switch thrown by light. The compound comprises a cyclophane bead strung on a thread having a central benzoxy station and a ferrocene knot at each end. When the workers illuminate the system using a laser, photonic energy prompts an electron-transfer reaction that leaves a positive charge at one ferrocene knot. A coulomb repulsion between this end point and the positively charged bead then pushes the bead to the far side of the thread.

Unfortunately, after a few picoseconds, the original charge distribution is restored, shutting the putative switch off. "This is absolutely useless," Harriman jokes. "But after the first reaction, you can initiate a second electron transfer that incorporates the ferrocene stopper," he notes. The rate of this reaction depends on the distance between the charges. Harriman therefore hopes to devise a less fickle switch by developing a system in which either a membrane or a series of smaller steps increases the distance between charges.

A simpler, more direct approach may be to move the bead using electrochemical and chemical persuasion. Angel E. Kaifer, with Richard A. Bissell and Emilio Córdova of the University of Miami and J. Fraser Stoddart of the University of Birmingham, reported in *Nature* another rotaxane that contains a cyclophane bead, a benzidine station, a biphenol station and two silicon knots. At 229 kelvins, for 84 percent of the molecules, the bead encircles the benzidine group. When the workers place a positive charge on this group, by protination using an acid or electrochemical oxidation, the like-charged bead then hovers exclusively around the biphenol station. "This rotaxane system is unique in that we have two handles by which we can control it," Kaifer says. Unlike Harriman and Benniston's system, the bead will remain at the biphenol station, say, in the "on" position, until switched off by a reverse reaction.

Nevertheless, because the electron transfer requires several milliseconds, the chemical switch is slower than its light-activated counterpart. In a light-triggered system the reaction takes but a few picoseconds. "You can also use several lasers of different color in sequence to excite different components in the molecule," Harriman adds. For now, neither scheme clearly holds the upper hand. "This is really very preliminary research," Kaifer says. "These are all relatively primitive molecules, and we should expect to see much better versions in the future."



ROTAXANE provides the basis for a molecular switch. Chemists control the position of the bead along the thread using photochemical or electrochemical methods.

To concoct rotaxanes, chemists rely on template-directed synthesis, a technique Stoddart pioneered. The strategy mimics the mechanism by which intricate molecules assemble themselves in nature. Scientists follow the way in which a product's ingredients will align themselves naturally (according to weak electrostatic interactions between them) as a template of sorts. Using this pattern, they can lay the pieces out so that the proper bonds will eventually snap into place, like a self-pitched tent.

More sophisticated rotaxanes continue to appear, but chemists have yet to incorporate them successfully into working widgets. "We still need to determine how we are going to hardwire these molecules to the outside world," Kaifer says. He suggests that perhaps the threads can be attached to an electrode surface, in which case an applied potential could bounce the bead back and forth. These movements might be harnessed to drive a molecular, electrical engine. "It's very easy to say that this is the objective," Harriman states, "but such feats are several generations of molecules away." *—Kristin Leutwyler*

More Bits from Pits

Will optical disks be the dominant storage medium?

Recent innovations are thrusting the optical disk into contention as the primary storage medium of the information age. The advances may enable optical disks to supplant videotapes, hard drives in computers and other magnetic storage devices.

Optical disks store data in the form of minute, oblong pits that cause fluc-

tuations in the brightness of a laser beam reflected off the disk's surface. These fluctuations metamorphose into a stream of ones and zeros and thus into information that the user receives as text, image, sound—or a combination of the three.

In May, Sony revealed an innovation that greatly increases the capacity of a disk. In conventional optical disks, information is stored by varying the length of each pit and the distances between the pits. The Sony workers placed the centers of the pits at regular intervals with respect to one another while varying the distance from the front and rear edge of each pit to its center. This scheme doubles the number of bits that each pit holds.

Rewritable optical disks are challenging the supremacy of the rewritable hard drives and floppy disks found in most computers. Called phase-change optical disks, they are coated with a thin, crystalline film. When a laser beam strikes the film, it forms glasslike, or amorphous, spots that change the film's reflectivity. A low-intensity laser beam can restore the crystalline structure.

Toshiba has developed a 3.5-inch, double-sided disk that holds 606 megabytes. Late last year Matsushita rolled out 5.25-inch phase-change disks having a capacity of 1.5 gigabytes. This May the company announced a method—similar to Sony's edge-modulation technique—that could quadruple the capacity of a phase-change disk to six billion bytes. A single such disk could store a million pages of text or several feature-length movies.

At the IBM Almaden Research Center, workers have multiplied the capacity of disks by stacking them on top of one another. Hal J. Rosen of IBM describes the innovation as making compact discs a three-dimensional rather than merely a two-dimensional storage technology.

Unlike conventional compact discs, which are coated with opaque metal films, the IBM optical disks are virtually transparent, allowing the laser to penetrate them and scan the surface of disks below. At the same time, the surfaces are reflective enough so that when the laser focuses on them, it can read the pits. A movable lens adjusts the focal point of the laser, allowing it to read the surface of any disk in the stack.

So far the IBM researchers have managed to place as many as six disks in a single stack. They say there are no technical obstacles to creating sandwiches of 10 disks, which would be only slightly thicker than a standard compact disc. The technique could also be applied to rewritable disks, Rosen says.

All these developments, combined with shorter-wavelength semiconductor lasers and superior video-compression techniques, may soon enable consumers to record full-length videos, for example, on small optical disks rather than big, boxy tapes. On the other hand, new products must have a clear margin of superiority to displace existing technologies—particularly if these products are not compatible with the old. Even in the era of compact discs, there are no doubt a few diehards who have held on to their record players because they have no other way of listening to their vintage collection of Sinatra. Although bankers and scientists might not suffer from such sentimental attachments, they might also be reluctant to invest in technologies that render obsolete their prized collections of financial data or records of emissions from pulsars.

—John Horgan and Robert Patton

Illegitimi non Carborundum

Although blessed with an extraordinary strength-to-weight ratio and good tolerance of heat, carbon fiber has not yet moved out of its boutique niche in the manufacturing world. The fiber is found in tennis and squash racquets, in missile nose cones, in rocket-engine nozzles and in structural components of spacecraft such as the *Hubble Space Telescope.* The material also appears in military aircraft. But in down-to-earth civil engineering—where carbon fiber's properties should also give it a powerful advantage over traditional materials—the wonder fiber remains largely unknown.

The primary reason is cost: materials for a cable made out of carbon fiber are about 20 times more expensive than is a steel cable capable of carrying the same weight, according to John M. Walton of Bridon Ropes, a British company in Yorkshire that is a major supplier for suspension bridges. Bridon has decided to challenge these economics. The firm is supplying cable for what Walton believes will be the first bridge to use supports made of carbon fiber. The cables consist of 61 rods, each containing hundreds of fibers thinner than a human hair, twisted together in a matrix of epoxy resin.

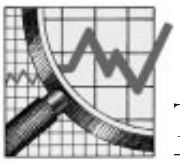
The price disadvantage is less daunting in finished cables than it is in the raw material, Walton says, because carbon-fiber cables are lighter to work and to transport. The economic pendulum swings even farther in favor of carbon when maintenance costs over the product's life cycle are factored in. The cables Bridon will use in the bridge perform better in fatigue tests than does steel, and they need no painting. Tests at the University of Cambridge show the rods that make up the cables become even stronger when they are wet.

Furthermore, the properties of carbon fiber mean it can be fashioned into engineering achievements that would otherwise be impossible. For example, carbon fiber's fivefold weight advantage over steel means that the material could be used to build a suspension bridge whose span can extend three times farther than one supported by the conventional material—almost 10 miles, as opposed to less than three miles.

But the first carbon fiber–supported bridge will not, alas, vault the Straits of Gibraltar. Instead it will reach 38 meters over the A50 highway at Meir in Britain's West Midlands, allowing pedestrians to cross in safety. Ironically, the fiber cable has to be encased in stainless steel to protect it from social rather than engineering forces (the knives of vandals: though strong, carbon-fiber cable can be cut quite easily). Walton hopes the experience gained with the bridge, which was designed by the engineering company Scott Wilson Kirkpatrick and commissioned by the Department of Transport, will lead to bigger things.

In the U.S., engineers expect to dispense with the preliminaries. A traffic-bearing bridge using carbon-fiber cables is being planned by a team at the University of California at San Diego to cross Interstate 5, which runs the length of the West Coast. The university, together with the Center for Composite Materials at the University of Delaware, recently received federal funding to find ways of rehabilitating battered and worn-out bridges and highways.

Carbon-fiber cables might also be used for anchoring deep-sea floating oil rigs. At present, production platforms are impractical in more than 4,500 feet of water, because the amount of buoyancy needed to offset the weight of the mooring cables becomes prohibitive. Carbon cables would allow platforms to be anchored in waters 9,000 feet deep, according to one estimate. Should oil companies pursue that option, carbon fiber will help them increase their access to an even more eagerly sought form of the element. *—Tim Beardsley*



THE ANALYTICAL ECONOMIST

A Workaholic Economy

For the first century or so of the industrial revolution, increased productivity led to decreases in working hours. Employees who had been putting in 12-hour days, six days a week, found their time on the job shrinking to 10 hours daily, then, finally, to eight hours five days a week. Only a generation ago social planners worried about what people would do with all this newfound free time. In the U.S., at least, it seems they need not have bothered.

Although the output per hour of work has more than doubled since 1945, leisure seems reserved largely for the unemployed and underemployed. Those who work full-time spend as much time on the job as they did at the end of World War II. In fact, working hours have increased noticeably since 1970—perhaps because real wages have stagnated since that year. Bookstores now abound with manuals describing how to manage time and cope with stress.

There are several reasons for lost leisure. Since 1979, companies have responded to improvements in the business climate by having employees work overtime rather than by hiring extra personnel, says economist Juliet B. Schor of Harvard University. Indeed, the current economic recovery has gained a certain amount of notoriety for its "jobless" nature: increased production has been almost entirely decoupled from employment. Some firms are even downsizing as their profits climb. "All things being equal, we'd be better off spreading around the work," observes labor economist Ronald G. Ehrenberg of Cornell University.

Yet a host of factors pushes employers to hire fewer workers for more hours and, at the same time, compels workers to spend more time on the job. Most of those incentives involve what Ehrenberg calls the structure of compensation: quirks in the way salaries and benefits are organized that make it more profitable to ask 40 employees to labor an extra hour each than to hire one more worker to do the same 40-hour job.

Professional and managerial employees supply the most obvious lesson along these lines. Once people are on salary, their cost to a firm is the same whether they spend 35 hours a week in the office or 70. Diminishing returns may eventually set in as overworked employees lose efficiency or leave for more arable pastures. But in the short run, the employer's incentive is clear.

Even hourly employees receive benefits—such as pension contributions and medical insurance—that are not tied to the number of hours they work. Therefore, it is more profitable for employers to work their existing employees harder. Back in 1930s, when laws that mandated time-and-a-half for overtime were first enacted, notes Lotte Bailyn of the Massachusetts Institute of Technology, benefit packages were much smaller, and increased wages spurred employers to hire new workers. Now, however, benefits have increased to the point where employers find overtime more profitable. Bailyn says many have even structured wages so that employees consider extra payments part of their basic pay. In addition, many employers are cutting back on benefits, while salaries are remaining the same.

Output per hour of work has doubled since 1945—and leisure time has dwindled.

Payroll taxes imposed by the government tip the scales even further away from hiring additional workers, Ehrenberg explains. Unemployment insurance, workers' compensation and Social Security taxes are ostensibly levied according to each employee's earnings. But wage caps—above which additional pay results in no additional tax—are set far too low, he says. Unemployment insurance contributions, for example, take into account only the first $12,000 of each worker's pay.

For all that employees complain about long hours, they, too, have reasons not to trade money for leisure. "People who work reduced hours pay a huge penalty" in career terms, Schor maintains. "It's taken as a negative signal" about their commitment to the firm. Bailyn adds that many corporate managers find it difficult to measure the contribution of their underlings to a firm's well-being, so they use the number of hours worked as a proxy for output. "Employees know this," she says, and they adjust their behavior accordingly.

Although "the image of the good worker is the one whose life belongs to the company," Bailyn says, "it doesn't fit the facts." She cites both quantitative and qualitative studies that show increased productivity for part-time workers: they make better use of the time they have, and they are less likely to succumb to fatigue in stressful jobs. Companies that employ more workers for less time also gain from the resulting redundancy, she asserts: "The extra people can cover the contingencies that you know are going to happen," such as when crises take people away from the workplace. Positive experiences with reduced hours have begun to change the more-is-better culture at some companies, Schor reports.

Larger firms, in particular, appear to be more willing to experiment with flexible working arrangements. Nevertheless, significant changes lie ahead if work schedules are to return to the downward trend they followed for more than a century before World War II. Financial barriers as well as cultural ones must go, Ehrenberg says. He recommends lifting the wage caps for unemployment insurance and other government levies—with a concomitant lowering of rates. He also suggests that new deductions, such as funds for national health insurance, be based on hourly payrolls. The cost of hiring new workers will thus no longer be artificially inflated.

It may take even more than changes in the financial and cultural structures of employment for workers successfully to trade increased productivity and money for leisure time, Schor contends. She says the U.S. market for goods has become skewed by the assumption of full-time, two-career households. Automobile makers no longer manufacture cheap models, and developers do not build the tiny bungalows that served the first postwar generation of home buyers. Not even the humblest household object is made without a microprocessor. As Schor notes, the situation is a curious inversion of the "appropriate technology" vision that designers have had for developing countries: U.S. goods are appropriate only for high incomes and long hours.

—Paul Wallich

Scanning Underwater Surfaces

The principles of tracking submarines and other objects underwater have, perhaps surprisingly, remained largely intact since early this century. One simple means is to lower a hydrophone into the water to pick up sound. Indeed, this column described how to build such a device years ago [see "The Amateur Scientist," March 1964 and August 1970]. A more active approach sends sound waves that echo off a target. Objects of various shapes, sizes and composition reflect sound waves differently. This acoustic signature enables a listener to determine whether an enemy submarine or some other object is approaching [see "Third World Submarines," by Daniel J. Revelle and Lora Lumpe, page 26]. Although the military relies on sophisticated electronic equipment to decipher the echoes, a simple, low-cost transducer can illustrate the essential principles. Easy to build, such transducers are also much cheaper than commercial fish finders, which operate in a similar way.

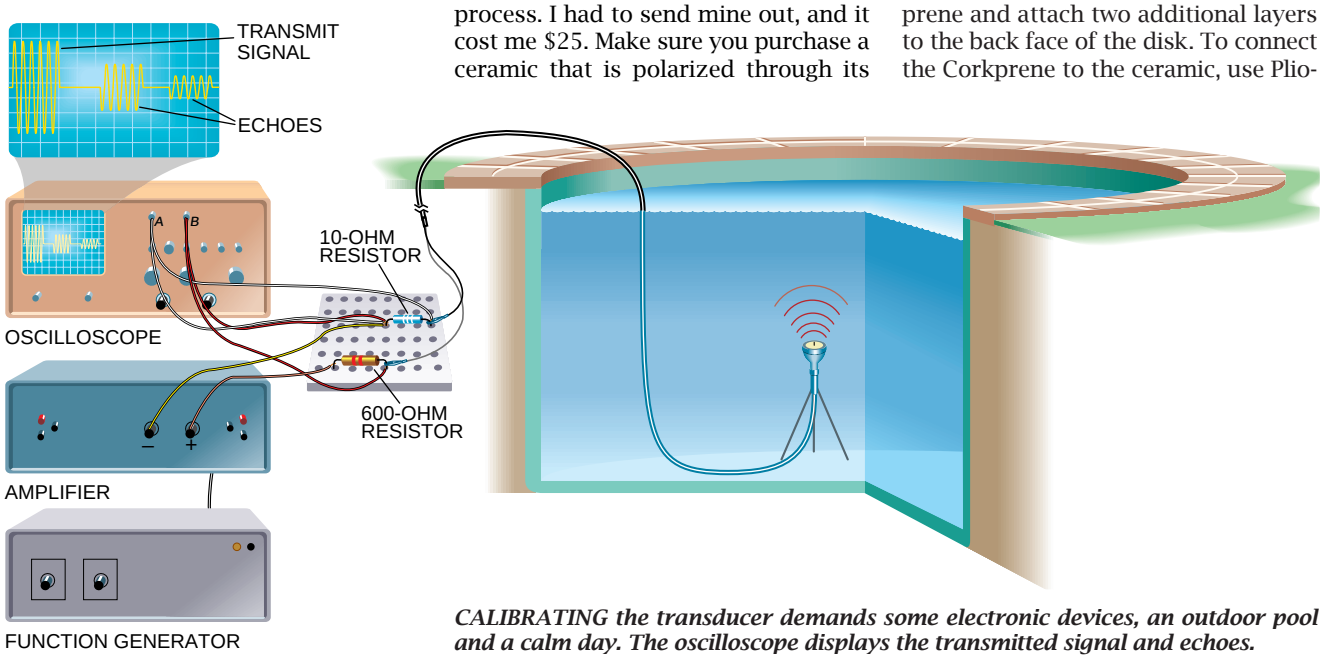
KYLE M. BECKER, who earned his degree in mechanical engineering from Boston University, is in the graduate program of acoustics at Pennsylvania State University. His thesis explores how differences in the surface along the ocean bottom affect acoustic signals.

The basis for the transducer design is a piece of piezoelectric ceramic. When squeezed, this substance develops an electrical voltage across its ends. Conversely, applying a charge to the faces makes the material flex. An alternating voltage causes the ceramic to expand and contract in succession, yielding an acoustic wave. Acoustic waves likewise cause an alternating charge to develop across the faces of the ceramic. Measuring the characteristics of the voltage indicates the kind of object from which the sound waves are reflecting.

The ceramic used for this construction is Navy type-I lead zirconate titanate or Vernitron type PZT-4. It can be obtained in several forms from various manufacturers. I got mine from EDO Acoustics (Ceramics Group, 2645 South 300 West, Salt Lake City, UT 84115; telephone: 801-486-2115). You will need a ring of the material that is half an inch thick. A ring breaks up unwanted vibration patterns that exist radially through the disk and that compromise its performance. The disk should have a two-inch outside diameter and a $^{3}/_{8}$-inch inside diameter. It is probably best to order the piece as a ring, as drilling holes through ceramic is not a trivial process. I had to send mine out, and it cost me $25. Make sure you purchase a ceramic that is polarized through its thickness and has both faces silvered.

Building the transducer means connecting a couple of wire leads to the ceramic disk and then mounting it in a suitable housing. Cast-iron pipe fittings, available from hardware or plumbing supply stores for about $5, make a good container. I used a pipe two and a half inches long and half an inch in diameter and a pipe converter. The converter is just a piece of tubing wider on one end than it is on the other. The tail end of the pipe converter should be half an inch wide so that the other pipe can screw into it. The mouth should be just wide enough to accept the ceramic. I used one whose mouth end was two inches wide, but I had to machine the threads out to give myself room to work. File smooth the top surface of the pipe converter and connect the pipe to the tail of the converter [*see illustration on opposite page*].

Next, prepare the ceramic for mounting. Begin by soldering a short lead to each face of the disk. To further suppress unwanted vibrations in the ceramic, you will need some Corkprene, a pliable, corklike material (made by Armstrong World Industries in Lancaster, Pa.; grade number DC-100). I received a three-foot-square, $^{1}/_{16}$-inch-thick sheet courtesy of Expanco Cork Company (P.O. Box 384, 1139 Phoenixville Pike, West Chester, PA 19380; telephone: 610-436-8300). Wrap the perimeter of the ceramic with the Corkprene and attach two additional layers to the back face of the disk. To connect the Corkprene to the ceramic, use Plio-



***CALIBRATING** the transducer demands some electronic devices, an outdoor pool and a calm day. The oscilloscope displays the transmitted signal and echoes.*

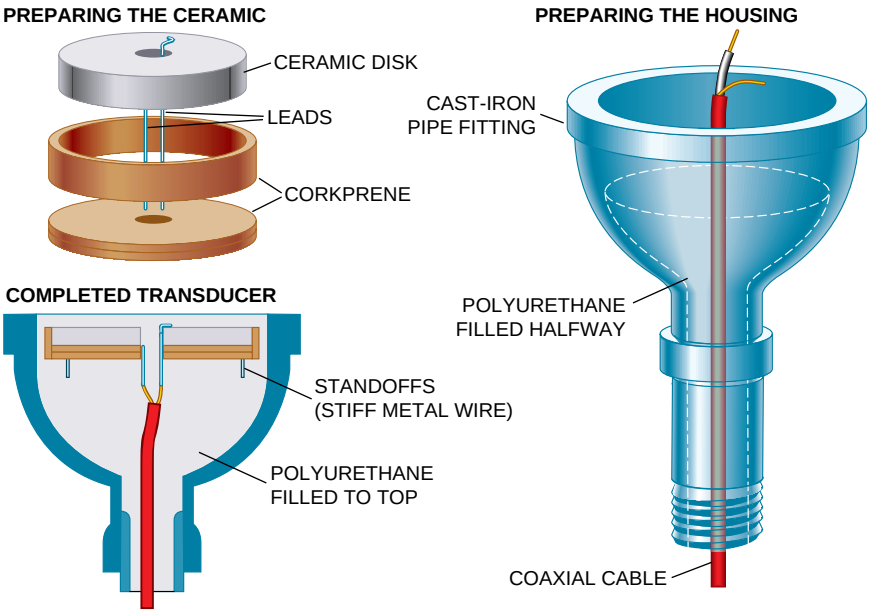
bond, an adhesive available at hardware stores. With the leads and cork in place, the ceramic is ready for mounting.

Prepare the housing by threading some coaxial cable up through the narrow end of the pipe assembly. The cable should extend at least 20 feet. Pull it through so that about an inch sticks out of the top of the housing. Secure the cable in place by wrapping the narrow end of the housing with heavy tape. A tight seal is important.

The next step is to fill the bottom half of the housing with a polyurethane potting compound. I used Conap EN-4 polyurethane (Conap, 1405 Buffalo Street, Olean, NY 14760; telephone: 716-372-9650). Follow the package instructions in preparing the compound. Pour enough to fill the mouth end of the pipe converter halfway, but be sure to leave enough room for the ceramic and Corkprene. After the polyurethane sets (about 24 hours), trim the coaxial cable as closely as possible to the polyurethane. Cut the cable so that the center conductor and shield are visible (be sure they do not touch each other).

Insert three or four pieces of short, stiff wire into the dried polyurethane. Cut pieces from sewing needles will work. These pieces hold up the ceramic disk and allow polyurethane to flow under it. Place them so that the ceramic will sit level to the front of the housing, about $^1/_8$ inch below the lip. Solder one of the leads to the center conductor of the coaxial cable and the other to the cable's metal shield. Place the ceramic on the stiff wire pieces and pour in polyurethane all the way up to the brim of the pipe converter. (The polyurethane will not affect the function of the ceramic.) For best results, pour slowly and try to prevent any air bubbles from forming, especially on the front face of the transducer. Allow it to set for 48 hours.

To make full use of the transducer, it is necessary to calibrate it. First, determine the frequency at which the transducer vibrates most readily. This value, called the resonance frequency, can easily be calculated from the dimensions of the ceramic. For a thin disk in the arrangement described, the resonance frequency in kilohertz is equal to 80 divided by the thickness of the ceramic in inches. For a half-inch piece of ceramic, resonance should occur at 160 kilohertz. A more accurate determination demands the self-reciprocity calibration technique. This method, presented formally in *Principles of Underwater Sound,* by Robert J. Urick, requires three pieces of electronic equipment. The first instrument is a signal generator, which must be capable of running in a pulsed mode at frequencies between 100 and 200 kilohertz; I used a Hewlett-Packard 3314A function generator. The second piece of equipment is a signal amplifier, which must be able to handle high frequencies without distortion. Finally, you will need a two-channel analogue oscilloscope. Although certainly not household items, such devices can usually be found in the physics or engineering laboratories of most colleges and universities. You might be able to arrange to



***UNDERWATER TRANSDUCER** is a piezoelectric ceramic housed in pipe fittings. Leads connected to the opposite faces of the ceramic are soldered to the coaxial cable. Polyurethane is poured into the housing in two stages for waterproofing.*

Calculating the Impedance

The impedance (Z) is the alternating current analogue of resistance in direct-current circuits. It is found by taking the ratio of transmit voltage (V) to transmit current (I)—that is, $Z = V/I$. It is best to use absolute magnitudes, so that you can work with real numbers.

The next step is to calculate the theoretical "spherical reciprocity parameter" (J_s), or $J_s = 2r/\rho f$. In this equation, r is the transducer's depth below the surface, the Greek letter rho (ρ) is the density of the water (about 1,000 kilograms per cubic meter for fresh water), and f is the driving frequency. Because the transmit response is determined relative to one meter, it is necessary to set r equal to 1.

Now J_s is equal to the receive sensitivity, M_s (in volts per pascal, a unit of pressure), divided by the transmit sensitivity, S_s (in pascals per ampere at one meter)—that is, $J_s = M_s / S_s$. To determine M_s, use the relation $M_s = \sqrt{(2r\,|E_r|\,J_s/I)}$. In this equation, E_r is the actual voltage across the transducer. You will need to correct for the 10-ohm resistor: $|E_r| = |V_r|\,[Z/(10+Z)]$. V_r is the return voltage, which appears as the second pulse envelope on the oscilloscope.

Once you find M_s, you can solve for S_s using the reciprocity parameter (J_s). Finally, convert the units from pascals to micropascals and from amperes to volts (use the transducer impedance values). Then you can determine the sensitivity of the transducer in decibels:

$$M_s\,(\text{dB}) = 20 \log (M_s),$$
$$S_s\,(\text{dB}) = 20 \log (S_s).$$

The value for M_s is relative to one volt per micropascal, and the value for S_s is relative to one micropascal per volt at one meter.

Determining the Scattering Strength

The scattering strength, S, is given by $S = 10 \log (s)$, where s is the scattering coefficient determined by

$$s = \frac{E[|E_r|^2] r^4}{|V|^2 M_s^2 S_s^2 A.}$$

$E[|E_r|^2]$ is the expected, or mean, value of each received echo voltage squared, r is the distance from the surface, and A is the area covered by the transducer sound. The area is given by

$$A = \frac{\pi r^2 \sec \theta \sin^2 \beta/2}{(\cos^2 \theta - \sin^2 \beta/2) \sqrt{1 - \tan^2 \theta \tan^2 \beta/2}}$$

where θ is the angle of incidence and β the width of the beam.

use the equipment at a local institution.

A calm water surface is also crucial. An indoor swimming pool, or even an outdoor pool on a calm day, suffices nicely. A bathtub or a 55-gallon drum is probably too small to yield accurate results. The transducer must be held in a mount that can be rotated. At the laboratory, we use a sophisticated computer-controlled system, but an expendable camera tripod and a clamp should do the job.

Calibrating the transducer requires measuring the voltages across the faces of the device and the current transmitted through it. Run the input signal from the signal generator to the amplifier. Connect the transducer to the amplifier in series with a 600-ohm resistor preceding it and a 10-ohm resistor directly after it [*see illustration on page 90*]. The voltages transmitted and received can then be measured directly across the transducer and displayed simultaneously on the oscilloscope. The voltage drop across the 10-ohm resistor gives the transmit "current." To determine the actual current, display the signal on the other channel of the oscilloscope and divide the measured voltage by 10 ohms.

After setting up the circuit, place the transducer in the water so that it points up at about eight to 12 inches below the surface. Set the signal generator and amplifier to provide pulsed sine waves. The transmit voltage should be around 20 volts peak to peak (that is, 10 volts above and below zero). On the signal generator, set the frequency at 160 kilohertz (or whatever number you determined to be the working frequency of your transducer) and the number of pulses at 35. The oscilloscope should show a distinct transmit signal of 10 volts strong and 35 cycles long, followed by a weaker "echo" of the same duration as the transmit signal. This echo is the first reflection from the sound scattered by the water-air interface. In fact, there should be a train of echoes corresponding to second and third and even fourth reflections. These reflections are strongest when the transducer points straight up. To take measurements, apply as much gain as you can to the device without distorting the transmit voltage on the oscilloscope. Then record the levels of the transmitted voltage, the received voltage (the level of the first reflected pulse) and the transmitted current.

These measurements will enable you to determine the transmit signal and receive sensitivities of the transducer. These characteristics depend on the transducer's impedance. The method for calculating the impedance appears in the box on the preceding page. You must carry out the procedure for a range of frequencies centered on the theoretical working frequency. The results will allow you to pinpoint the exact frequencies at which your transducer is most sensitive.

To determine your transducer's directivity, or beam pattern, set the frequency back to where it is most sensitive. Mine was 156 kilohertz. After making sure the transducer is pointing straight up, take a measurement. Rotate the transducer in one direction by one or two degrees before conducting another measurement. Continue this procedure until the first echo no longer appears on the oscilloscope. Reset the transducer so it is aimed straight up again and repeat the entire procedure, but rotate the transducer in the other direction.

Once you have your data, you must "normalize" your results. That is, divide all your measurements by the result obtained at zero degrees, or looking straight up. Take the logarithm of these normalized quantities and multiply by 20; this procedure converts the readings into decibels. Plot these values on the y axis against the angles for which the readings were taken on the x axis. (Note that the decibel values are negative, so that zero should be at the top.) You should get some kind of bell-shaped curve, with the maximum occurring at the zero-degree mark. Look for the points on the curve that are six decibels below zero. The expanse of the curve at this point indicates the effective beam width.

After you determine the operating characteristics of your transducer, you can use it for many purposes. You can determine the position of a large object by looking for its echoes. Multiply the time between transmit and receive pulses by the speed of sound in water (about 1,500 meters per second) to calculate its distance. If the target is moving, you will need to repeat the procedure several times to track its distance from the transducer.

Because most amateurs will probably not need to locate submarines and torpedoes, it is perhaps more practical to observe how different surfaces scatter sound waves. To start, I would suggest scattering signals off the bottom of a swimming pool. Many aboveground pools have a sandy bottom covered by a pool liner, which is acoustically transparent. Another good surface would be an in-ground pool, which often has a rough concrete bottom.

Take at least 100 separate measurements at a time, scanning the surface with the transducer positioned at a known distance and angle from the bottom. You must then manipulate the data statistically [*see box above*]. Each kind of surface will produce a characteristic curve. Urick's book offers several examples. With this information and your transducer, you should be able to deduce the kind of bottom below bodies of water down to at least 30 meters deep.

For more information on scattering curves, write to: The Amateur Scientist, Underwater Transducer, Scientific American, 415 Madison Avenue, New York, NY 10017-1111.

FURTHER READING

FUNDAMENTALS OF ACOUSTICS. L. E. Kinsler, A. R. Frey, A. B. Coppens and J. V. Sanders. John Wiley & Sons, 1982.

PRINCIPLES OF UNDERWATER SOUND. R. J. Urick. McGraw-Hill Publishing, 1983.

UNDERWATER ACOUSTIC SYSTEM ANALYSIS. W. S. Burdic. Prentice Hall, 1984.

BOOK REVIEWS by Philip Morrison

The Smallest Biochemists

POWER UNSEEN: HOW MICROBES RULE THE WORLD, by Bernard Dixon. W. H. Freeman and Company, 1994 ($22.95).

At least you learn how microbes astonish the world. Item: On plate IV appears the bacterium *Haloarcula,* square as a picture frame, four microns on each edge, shown in a sharp electron micrograph image 10 centimeters square. There is a little fuzz and some bending, but the cell is square all right, a lively surprise. Tony Walsby found it around 1980 with its kin by the tens of millions in any cubic centimeter of solar-heated brine on a seaside salt flat in the Sinai Desert. The name it bears means "salt box." The flimsy, clear, thin squares float and swim to and fro very well, using twin reversible flagellae. In 1986 Koki Horikoshi and his team discovered a triangular counterpart in a Japanese salt farm.

Bernard Dixon is an English microbiologist turned celebrated science journalist. With his flair for news and a contagious pleasure in the ancient lineage of the single-celled, he has assembled this delightful book of short stories about them and naturally also about those who study, harbor, employ and fear them. He has compiled 75 such three-page tales as that of the squares and triangles, easy to consume as salted peanuts and even more nourishing. This is science in practice, biodiversity itself packaged along with a critical look at the nature of evidence, all written very small, very simple, prose as edifying as it is entertaining.

Item (one of the older ones): Chaim Weizmann arrived in Manchester in 1904 with "more or less the proverbial half-a-crown in his pocket." A young Jew denied advancement in his homeland of Belarus, he hoped to make his way in liberal Britain on the strength of his fine chemical training in Zurich. Superb lecturer, beloved of the students, he had become a "very remarkable" professor of chemistry when World War I began. When the *Guardian* editor interviewed the busy minister of munitions, David Lloyd George, he was surprised to find him preoccupied, not by policy or even labor but by a shortage of the solvent acetone, needed to make propellant for the 12-inch guns of the new Dreadnoughts. Acetone was made then by the antique and expensive process of distillation of hardwood. Offered the volunteer services of Manchester's most brilliant chemist, the politician was at once charmed by Weizmann, who came back within weeks with an utterly original technical fix. He had isolated from the microflora of maize a small bacterium, *Clostridium acetobutylicum,* able in huge vats to ferment acetone and butyl alcohol (also worthwhile) from cheap cereal starch.

That was the first sizable extension of the traditional fermentation industries and marked the beginning of modern biotechnology. Weizmann sought only one reward: British help in repatriating the Jewish people back to Zion. As prime minister, Lloyd George set that process in motion for good and bad in 1917. Chemist Weizmann would become the first president of beset Israel in 1948. *C. acetobutylicum* synthesis, however, is now as obsolete as dreadnoughts; these days acetone is made very cheaply from petrochemical feedstock using mineral catalysts.

Item: The colon bacterium, the slime molds, the yeasts, the bread mold fungus are all microbes essential to modern microbiological research, as well as a long list of microorganisms useful in synthesis and diagnosis. But few viruses are friends to science; they are foes, at most eventual agents of their own medical defeat. There are new exceptions. One herpesvirus (and some strains of rabies and pseudorabies, too) has become a subtle and important tool for neurobiology. A key task is to plot the labyrinthine circuits of the neural networks. Two neuroanatomists published the way in 1990.

For years, the connections have been traced by adding enzymes that living fibers can take up; chemical activity then serves as the marker. Fluorescent dyes have sometimes worked well, and even neurotoxins like that of tetanus. But such labels hardly travel across the gaps at the neuron-to-neuron contacts. The tracers diffuse out into the gaps, and labeling weakens along the chain. The trick is simple and striking: live neuroviruses multiply in living nerve cells, so they will amplify as they go instead of weakening. Detection is not hard if one uses specific antibodies; the

MICROBIOLOGICAL PESTICIDE: a microscopic fungus penetrates the threadlike hyphae of another microbe, which causes disease in lettuce.

scheme works both in glassware and in live lab animals. Perhaps we will one day praise neuroviruses for their exquisite mapping of the circuits of thought. Dr. Dixon arrays 70 more little microbial narratives for a reader's dipping pleasure. Try them soon.

Not by Bread Alone

WINE SCIENCE: PRINCIPLES AND APPLICATIONS, by Ron S. Jackson. Academic Press, 1994 ($99.95).

No fruit crop is as large as the 60 million tons of grapes grown worldwide, outweighing all the year's oranges or bananas. Two thirds of that crop is made into wine. France and Italy each grow about a quarter of the globe's wine grapes (and produce somewhat more of its wines), exchanging places in a minuet they dance to weather and economies. A phenomenon of the North Temperate Zone, though grown nowadays on both sides of the equator, vineyards share the climate of the natural habitat of the predominant wine grape species, *Vitis vinifera*. One familiar plant of the same genus is Boston Ivy; many wild grapes of other *Vitis* species grow in northern temperate lands. Wine is certainly nothing new; the oldest residues we know were found (by infrared absorption spectroscopy) near Anatolia and the southern Caucasus and date to about 4000 B.C.

Wine is by now so well integrated an indulgence—a "seraphic beverage," the author calls it—that it improves public health (in spite of the red-wine headaches caused by a certain identified enzyme). Health benefits derive in part because wine displaces less angelic concentrations of alcohol, in part because it confers specific cardiac benefits, so recently established that they do not enter this comprehensive, engaging and otherwise up-to-date treatise. Not for the coffee table, this volume has plenty of clear technical text, with many diagrams, graphs and maps. General readers will indeed have no easy time over the 50 pages of organic chemicals found in wines or with the precise botany of the flowers of the vine, but the book touches on and clarifies so many disciplines that it can draw in all the scientific and engineering readers who know the fragrant and glowing glass.

Table wine is no matter of grapes alone but rests on the properties and interactions of half a dozen distinct organisms: the vine itself, a specific yeast (the chief fermenting industries each domesticate strains of a single yeast species, selected for wine maker, brewer, distiller or baker), one bacterial species, oak trees (especially the white oak and the cork oak) and the all-too-human consumers. It is a happy accident of history that the very beginnings of Western agriculture took place just where the vine grows near or even on the oaks, at the southern edge of the once glaciated part of Europe and western Asia. For the particular yeast species now indispensable in wine making seems not to be among the many yeasts adapted to grow naturally on grape skin but rather is one found wild on the bark and sap of oak trees.

A new vineyard harbors few or none of the cells of wine maker's yeast; however, once wine is made there, the strain is always present. The bacterium friend of the wine—many other bacteria are hostile—is the one able to ferment malic and tartaric acids into ethyl alcohol. The white oak provides the wood best suited for the cooper's work, the wooden tankage and casks wherein wine is traditionally made and stored. These wooden vessels are now giving way to the stainless steel of chemical engineering. But the very flavors and body of red wine owe much to the chemistry of white oak wood, the tannins and lignin degradation products that the wine slowly extracts from the oaken staves. The unique resinous flavors of Greek wine come from added exudates of the Aleppo pine. Of course, it is the thick bark of the cork oak of Portugal that has provided the stopper of choice ever since glass bottles first became cheap.

Several hundred compounds, many volatile, can be identified in wine by gas chromatography and mass spectroscopy, down to the parts per trillion level. Those that have "impact" include the macroconstituents: water, alcohol, the grape sugars fructose and glucose, the two well-known organic acids, and ethyl alcohol. Fermentation brings as well a string of longer straight-chain alcohols at the parts per 1,000 level; their aromas and those of the esters they form with the organic acids do much for wine's appeal. But the long catalogue of trace constituents gives little understanding of desirable fragrance and flavor, although it has ordered the pathologies of wine. One or a few intricate compounds—for instance, the molecule *N*-(*N*-hydroxy-*N*-methyl-gamma-aminobutyryl) glycin—may account for the "foxiness" of New World grapes (they are quite another species, *V. lambrusca*), but the tasting panels always find complex multiple factors in the subtleties of good wine.

The literature here so well surveyed and the technology it records are rich. There are 15,000 named cultivars of the vine; some 300 clones can be distinguished within one famous variety, Pinot noir. Clonal selection is the breeders' art of our day, based on vegetative techniques that use cuttings and even buds rather than flower hybridization. The results have raised yields and prevented infection. So far tissue culture has been disappointing; plantlet yield from single cells remains far too low.

You can read of much inorganic engineering, too, of the remarkable elastic response of compressed cork, of well-measured microclimates and new ways to array the grape leaf canopy to take full advantage of flatland sunshine. Here, too, you will find how hand-selected grapes derive their honeyed flavor from high sugar content and from a furanone and an ester with very long structural names, the fruit all "nobly" rotting under fermentation by exactly the right fungus. (Vive the vulnerable variety Sémillon.) You can admire—or not—the ingenuity of the new "bag in a box" packaging for bulk wines, as well as the fermentation of uncrushed grapes held in an atmosphere of carbon dioxide that can win quick fragrance for wines like Beaujolais nouveau.

Professor Jackson is a botanist at Brandon University in wintry Manitoba, nowhere near the lands of oak and vine. His devoted engagement to wine making began when he worked in grape country, both in Ontario and in the Finger Lakes of New York State. (There is evidence that his Francophone wife, Suzanne Ouellet, was a further benign influence.) No mere academic, but a consultant to official Manitoba, his authority demonstrates that expertise can flow like wine from producer to consumer, to be widely and pleasurably shared.

Robin Hood Redux

GLOBAL ECOLOGY: ENVIRONMENTAL CHANGE AND SOCIAL FLEXIBILITY, by Vaclav Smil. Routledge, 1994 ($59.95; paperbound, $17.95).

Under a rather too general title, this direct, appealing book amounts to a personal statement by a cogent analyst, well supported, as usual for him, by a penetrating and compact account of the facts.

Of course, we do not know enough; each farmer in the field must act lacking enough data to optimize his work before the uncertainties of water, weather, soil, prices. We need more knowledge than we can have. Many eloquent voices therefore cry "Jeremiah." Professor Smil names half a dozen names among his references; in 1968 one population ex-

pert grimly foresaw "the battle to feed all humanity is over.... nothing can prevent a substantial increase in the world death rate." In fact, in both the largest lands, China and India, the expectancy of life has improved since then by 10 or 15 years. "The great doomsayer" (Paul R. Ehrlich) offered no correction, instead publishing with a co-author yet another gloomy book, replacing "bomb" by "explosion" in its title.

"Pushing preconceived ideas disguised as objective evaluations" wins bold headlines but is manifestly one of the most frequent failings of environmental science. No more substantial to Smil are the claims of those who see only a cornucopia ahead, a view as suasive as it is naive. Science has been successful overall, he agrees, although it is certainly incomplete, and perhaps it is incapable of completeness. Yet its "qualified utilitarian success" does not mean that it can be sole guide to the future. Even a satisfactory science is "no substitute for morality, and to believe that the ethics of limits and sharing has no place in dealing with our environmental dilemmas would be to forfeit any hopes for real successes in solving them."

His presentation opens with the realities as he sees them, goes on to outline promising strategies for remedial actions in the face of uncertainty and closes with a sober account of "cascading complications" ahead. For it is not marginal adjustments that will work but rather a slow, "profound socio-economic transformation which will demand...new ways...genuine sacrifices"—especially, reader, from the prosperous. Who else?

What abides are the existential necessities. How many will sit at the world table? The population curves, long dubbed exponential, are here to the year 2000. Already we know from the observed rates of change that the term is in error. The author very properly notes that the highest percentage rate of population increase was past by the 1970s and has fallen steadily ever since. (Even more striking is one United Nations forecast that the absolute decadal increment in human head count will have peaked by that millennial year, to fall thereafter, likely a benchmark for all human history.) How will we feed the next five billion persons as our family doubles? Most of their bowls must be filled from increased yield of field and paddy, little from new-sown croplands. That is nothing unusual; four fifths of the crop gains of the 20th century were won by intensification. Doubling is doable, given more fertilizer, more irrigation water, more skill and, behind those, more subsidizing energy, da capo.

Minerals are not in short supply; plant macronutrient solids (the elements nitrogen, phosphorus, potassium, sulfur) are plentiful, and all special needs are wide open to substitution and economic recycling. The Bronze Age is over, even the Iron Age declines, and the Silica Age has lots of sand to mine. For a couple of generations, non–fuel resource demands will certainly grow as the populous lands consume more, but we expect no major problems of supply or of impact. Soil erosion is widespread and adds a tone of anxiety to our expectation of higher yields, but its quantitative toll worldwide has not yet risen out of the noise, rather like global warming itself.

What rich and poor alike must do is seek all-around efficiency, let frugality grow, organize a market-wise payment of more of the real price of natural resources and services. Hardly anyone grumbles about efficiency. The exceptions are consumers who dine well on meat and seafood, and big-engine fanciers, all sure to be hard hit by realistic taxes and prices, and whole countries short on capital.

On a large scale, coal will return, to burn in hot fluidized beds to feed gas turbines. At farmside scale, much more will change. Even little gypsum blocks buried in the fields, their resistance measurable with Radio Shack technology, can monitor and much economize water use. The old rotation of grain fields to bear legumes and green manure will return renewed and modernized. But grand novelties—plentiful methane, ocean waves and surface heat, wind, safe fission or hot fusion—should not be planned for but taken as welcome bonuses. More likely arrivals are photovoltaics at large scale, gene-engineered, self-nitrifying cereal crops, and enzymatic fermentation of cellulose wastes into alcohol fuel. We might get lucky!

So far the good-enough news. Now for the bad. Who will vote for frugality? Not the 100 million Sichuan farm folk already chilly and half-hungry during one entire season of the year and not the Brazilian *favelas.* Will the middle class of North America vote against the economic growth they are promised every day? Meanwhile terawatts (1 TW = 1 billion kilowatts as heat) mount up; the power used for transportation alone among our 5 percent of humankind in the U.S. runs at a steady 1 TW, whereas the poorest 80 percent of humanity disposes of only 5 TW for every purpose under heaven. Industrializing, populous China is likely to supplant the U.S. as the biggest carbon dioxide producer by the year 2020, for good human reasons.

Two large changes lie ahead; we cannot very well assess their scale or their arrival time. The poor are urbanizing fast, making their way from subsistence farming to an industrial base. They surely aim at continued economic growth. The well-off face reduction, not growth, in their current high consumption of "energy, water, food, feed, or wood." The richer need to begin transition to a more enduring economic system and to a healthier life at low growth rate, aimed at the "delinking of social status from material consumption." That the quality of life can remain quite high under lower but better standards of consumption is effectively argued here through an eye-opening study of affluence as a function of energy use in France. There is time left for a buffer, the use of our present surplus capacity—much of it released from unneeded military investment—to improve the lot of others, including the many who still dwell poor within wealthy lands.

History has contradictory precedents to illumine these deep tensions. Once sugar from the West Indies was Europe's largest single import. Long ago the industry to make sugar from domestic beets began in Europe. During the late 1980s, the sugar makers of Europe and the U.S. were paid four to five times the world price, whereas a good-size low-wage labor force grew sugar in the tropics at double the yield with much less input, an advantage built into tropical sugarcane and its all-year growth. Now world sugar is overproduced at low price. Rich countries have kept efficient producers poorer ever since the Napoleonic Wars, in order to subsidize a small number of farmers at the expense of the environment and the consumers—a defense against naval blockade that has hardened into a long-vested interest.

Less than 10 years back, Robert Jay Lifton wrote in fear of large-scale nuclear conflict: "We need to be neither optimistic nor pessimistic about the human future; rather, we must hope. Hope means...possibility and includes desire, anticipation, and vitality.... It is a necessity and...virtue...that must itself be nurtured, shared, mutually enhanced." Along a course that few expected, that compelling fear has much receded. Lifton's hope proved justifiable, and hope anticipates further action. Our author's view denies any "single-vision salvation," any "sweeping normative solutions," neither small paradises nor safe massive power. We ought to live neither in confidence nor in despair but instead with shared hope and a steady sense for possibilities old and new.

Sex, Death and Kefir

The certainty of death was absent at the origin of life. Unlike humans and other mammals, many organisms do not age and die. The process of programmed, inevitable death evolved only after our symbiotic microbial ancestors, some two billion years ago, became sexual individuals.

Any organism can die because of circumstances beyond its control: the ambience grows too hot, a predator attacks, poison gas permeates. But programmed death happens independently of environmental action: cornstalks topple at the end of the season, or a healthy elephant succumbs at the end of a century. Monthly, in menstruating women, the dead cells of the uterine lining flow through the vagina. Each autumn in the deciduous trees and shrubs of the North Temperate Zone, rows of cells at the base of the leaf stem die.

Unlike animals and plants that grow from embryos and die on schedule, all bacteria and most other microorganisms remain eternally young. These other organisms are protoctists and fungi. Protoctists constitute a diverse group that includes our animal ancestors, as well as seaweeds, ciliates, slime molds, foraminifera, diatoms and many others. Like the fungi (yeasts, molds and mushrooms), protoctists are symbiotic aggregates of nucleated cells that reproduce by cell division. Protoctist and fungal individuals can grow and reproduce without any sexual partners.

But in some protoctists—those that became the ancestors of the fungi, plants and animals—our kind of sex, which involves mating and cell fusion by fertilization, first appeared. I propose that it did so as an accidental consequence of a desperate strategy for survival. Sex began when unfavorable seasonal changes in the environment caused our protoctist predecessors to engage in attempts at cannibalism that were only partially successful.

The result was a monster bearing the cells and genes of at least two individuals (as does the fertilized egg today). The return of more favorable environmental conditions selected for survival those monsters able to regain their simpler, normal identity. To do so, each had to slough off half or more of the "extra" cell remains. Death and the genes that caused death evolved. "Death genes" have now been isolated and their operation studied. Lawrence M. Schwartz, here at the University of Massachusetts, for example, can predict the demise of cells in a laboratory culture to within a few hours when he introduces DNA containing death genes.

Those microbial ancestors that fused survived, whereas those that evaded sexual liaisons died. Cell fusion that guaranteed survival triggered development of simpler, normal individuals again. Those very few individuals that indulged each winter or dry season in body fusion and its relief by death of extra genes and cells survived to become our sexual ancestors. Cannibalistic fusion and its thwarting by programmed death became inextricably linked to seasonal survival and to individuality. Embryo development required sex cell fusion, cell movement, cell interaction and programmed cell death. It still does.

Surprisingly, kefir, a nutritious drink popular in the Caucasus Mountains of southern Russia and Georgia, illustrates how symbiogenesis—the appearance of new species by symbiosis—works and how, in evolution, symbiosis preceded sex. Kefir refers both to the effervescent dairy drink and to the individual curds that ferment milk to make the drink. These curds, like our eventually sexy protoctist ancestors, evolved symbiogenetically.

Legend says the prophet Muhammad gave the original kefir curds to the Orthodox Christian peoples near Mount Elbrus with strict orders never to give them away. Nevertheless, secrets of preparation of "Muhammad pellets" have been shared. Kefir, which looks like large-curd cottage cheese, grows by division. It ferments milk sugars and proteins, making the yogurtlike drink. When the active metabolism that assures individuality ceases, kefir curds dissolve and die without aging.

After the curds die, kefir individuals become an arbitrary mix of fermenting microbes rather than the specific combination of bacteria and yeast that forms each curd. Like our protoctist ancestors that evolved from symbioses among bacteria, kefir individuals arose from the physical association of 30 different kinds of microbes. These yeasts and bacteria remain together in precise relationships as each divides, maintaining the integrity of the individual curd.

Symbiogenesis led to complex individuals that *can* die (like kefir and most protoctists) before sexuality led to organisms that *had* to die (such as elephants and us). A kefir individual, like any other, requires behavioral and metabolic reaffirmation: component microbial cells that grow too fast or that do not help to make the curd are forced by the others to die. During the course of brewing the beverage, people inadvertently bred for kefir individuality.

Kefir microbes are integrated into kefir curds just as the former symbiotic bacteria became components of our cells. The kefir curd is a new individual more complex than its components. Kefir can no more be made by the "right mix" of chemicals or microbes than can elephants.

Like kefir, we and all other organisms made of nucleated cells, from amoebae to whales, are not only individuals, we are aggregates. For example, plant cells come from ancestors with indigestion. Ancestral translucent swimming cells that acquired green photosynthetic cyanobacteria began as green monsters. From many partly digested cyanobacteria in a hungry protoctist, a new individual, the green alga cell, and ultimately the plant, evolved.

Kefir is a sparkling demonstration that integration processes by which our cells evolved still occur. Kefir also helps us see how the origin of complex new individuals preceded the evolution of programmed death of the individual. Kefir instructs us, by its existence, about how the tastes and choices of one species (ours) influence the evolution of others, the 30 intertwined microbes that became the kefir curd. Although the kefir curd is a complex individual, a product of interacting aggregates of bacteria and fungi, it does not reproduce by sex. Rather the kefir curd, which has no sex life, enlarges by direct growth, division and death of its components. When mistreated by adverse environments, it disintegrates and dies. And, like any live individual, it never returns to life as that same individual.

LYNN MARGULIS, distinguished university professor, is in the biology department at the University of Massachusetts, Amherst.