INTERGALACTIC GAS: The Cosmic Importance of (Almost) Nothing

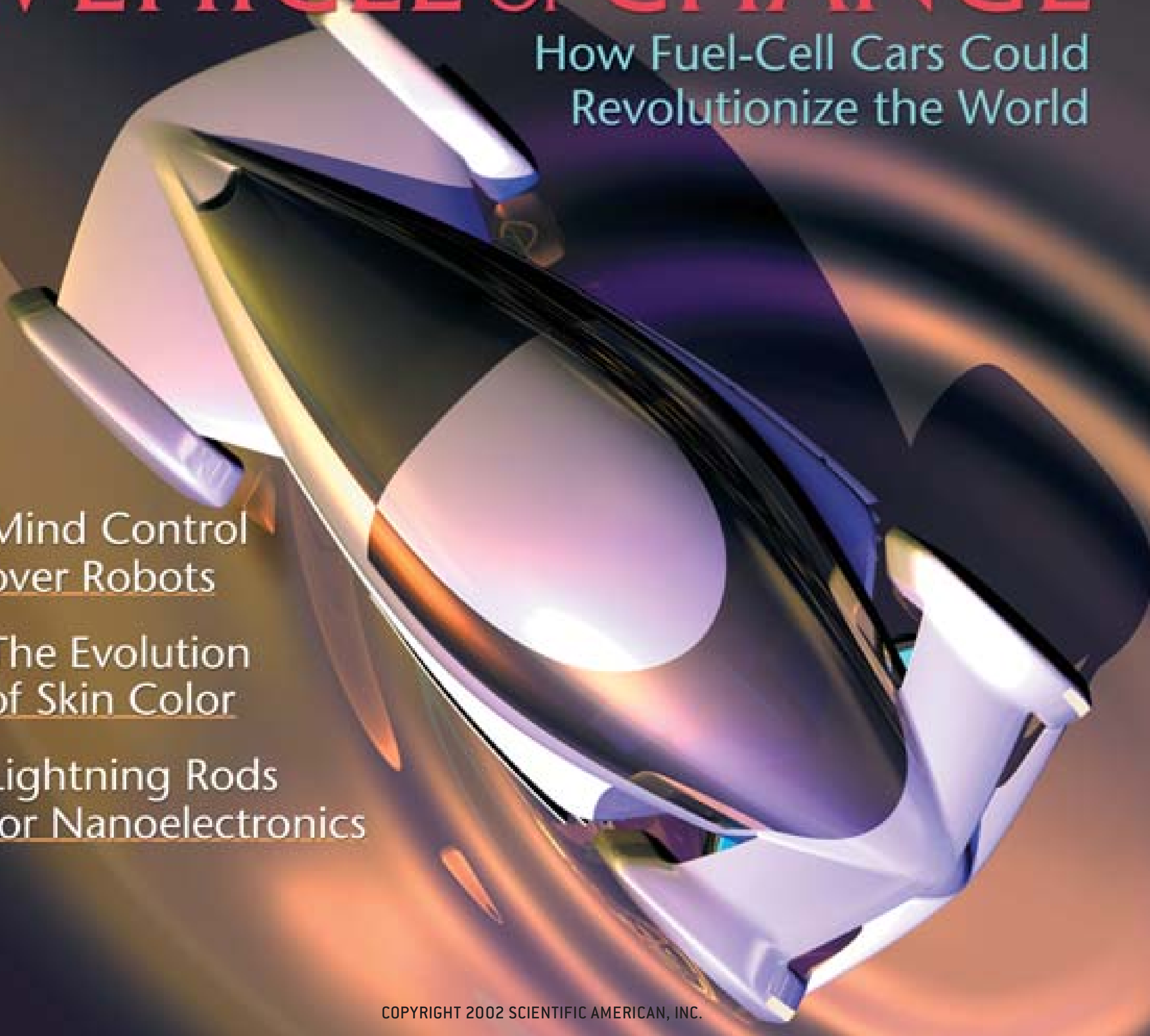
SCIENTIFIC AMERICAN

Terrorist Germs: An Early-Warning Defense System

OCTOBER 2002 $4.95
WWW.SCIAM.COM

VEHICLE OF CHANGE

How Fuel-Cell Cars Could Revolutionize the World



Mind Control over Robots

The Evolution of Skin Color

Lightning Rods for Nanoelectronics

october 2002

contents

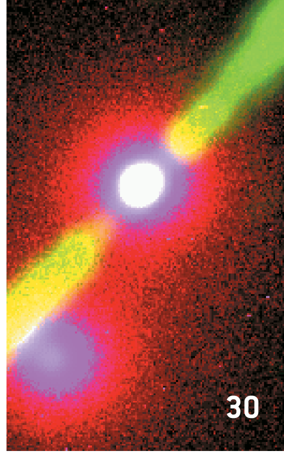
SCIENTIFIC AMERICAN Volume 287 Number 4

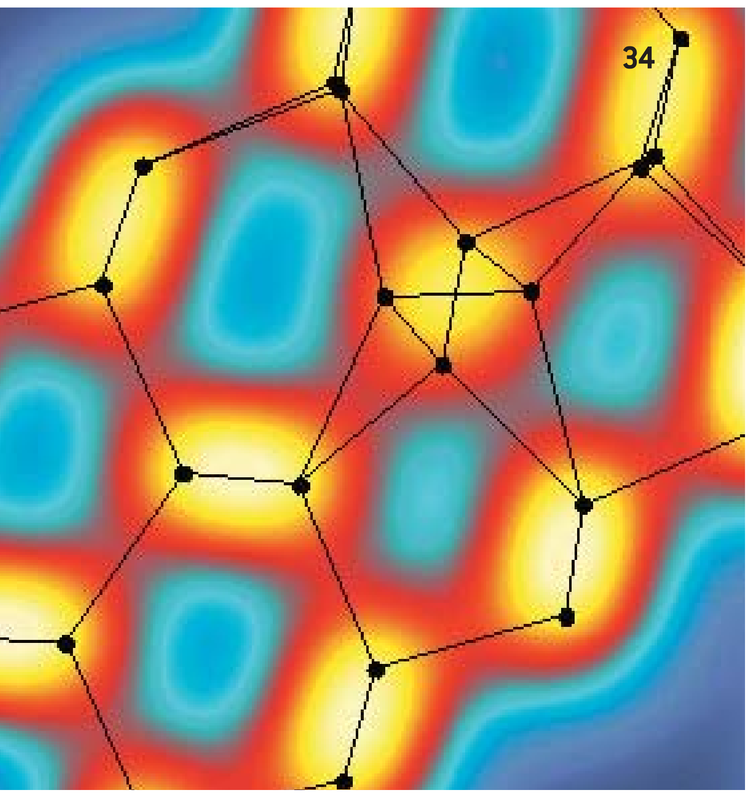
features

64 The future of transportation

SCIENTIFIC AMERICAN Volume 287 Number 4

departments





columns

Cover and page 3 images courtesy of General Motors

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian BN No. 127387652RT; QST No. Q1015332537. Subscription rates: one year $34.97, Canada $49, International $55. 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; (212) 451-8877; fax: (212) 355-0408 or send e-mail to sacust@sciam.com **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631.** Printed in U.S.A.

SA Perspectives

Greenwashing the Car

The automotive industry has never been known for taking the initiative in cleaning up the environment. Ever since the federal government forced auto manufacturers to lower exhaust pollution levels in the 1970s, industry lobbyists have waged a tough rearguard action on Capitol Hill against efforts to raise fuel economy standards. Meanwhile car companies have fiercely resisted the reclassification of their highly profitable small pickups, sport-utility and "crossover" vehicles from their current designation as moderately regulated light trucks for commercial use to what they often really are: gas-guzzling personal transport. Until not so long ago, many automakers denied even the possibility that carbon dioxide and other greenhouse gases might induce global warming. Following form, their representatives are fighting tooth and nail to block a recently passed California state law that restricts automotive carbon dioxide emissions.

EXHAUSTED: What hope for cleaner cars soon?

So what are we to make of carmakers' recent protestations that they want to be environmentally friendly? They are, after all, pouring large sums into the development of clean-diesel, hybrid and fuel-cell electric vehicles. And auto manufacturers have developed some promising fuel-saving technologies that they could roll out. But, perhaps most significantly, they are talking openly about making a revolutionary shift from today's oil-based economy to one founded on hydrogen. The entire industry now seems to agree that hydrogen fuel cells represent the only feasible long-term path toward addressing the environmental, economic and geopolitical issues associated with dependence on petroleum. The Bush administration, too, supports hydrogen fuel-cell development in its FreedomCAR public-private initiative.

The new reality is that auto manufacturers, and some global energy firms as well, now seem to see the hydrogen future as a potential moneymaker rather than the road to bankruptcy. Whenever the interests of business and the environment are aligned, real change for the better becomes possible.

In their article beginning on page 64, a trio of General Motors executives discusses their company's plans for vehicles powered by fuel cells rather than internal-combustion engines. In their vision, gas stations of the future would truly live up to their name by dispensing hydrogen gas. Reworking the car into a clean machine while driving the establishment of a nationwide hydrogen fuel distribution system costing hundreds of billions of dollars will certainly be a daunting task.

So two cheers for the fuel-cell-car pioneers. But this transformation will start to get serious only in a decade or so. Until then, industry lobbyists will apparently continue to battle against near-term measures to improve the environment. Skeptics note that the commitment to a far-off technology lets the auto industry earn environmental kudos without necessarily incurring the costs of producing high-mileage cars today. Environmentalists have a name for a strategy in which one flaunts green credentials while pushing to maintain the ability to pollute: "greenwashing."

The long, hard quest to build affordable, practical fuel-cell cars should not become an excuse to ignore what can and should be done more immediately. If we want car companies to design a greener future, then we need a system of incentives and market opportunities that steers them that way. In the meantime we must ensure that they make further reasonable efforts to clean up the trusty old internal-combustion engine.

THE EDITORS editors@sciam.com

GETTY IMAGES

How to Contact Us

EDITORIAL

For Letters to the Editors:
Letters to the Editors
Scientific American
415 Madison Ave.
New York, NY 10017-1111
or
editors@sciam.com
Please include your name and mailing address, and cite the article and the issue in which it appeared. Letters may be edited for length and clarity. We regret that we cannot answer all correspondence.

For general inquiries:
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-754-0550
fax: 212-755-1976
or
editors@sciam.com

SUBSCRIPTIONS

For new subscriptions, renewals, gifts, payments, and changes of address:
U.S. and Canada
800-333-1199
Outside North America
515-247-7631
or
www.sciam.com
or
Scientific American
Box 3187
Harlan, IA 51537

REPRINTS

To order reprints of articles:
Reprint Department
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-451-8877
fax: 212-355-0408
reprints@sciam.com

PERMISSIONS

For permission to copy or reuse material from *SA*:
permissions@sciam.com
or
212-451-8546 for procedures
or
Permissions Department
Scientific American
415 Madison Ave.
New York, NY 10017-1111
Please allow three to six weeks for processing.

ADVERTISING

www.sciam.com has electronic contact information for sales representatives of Scientific American in all regions of the U.S. and in other countries.

New York
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-451-8893
fax: 212-754-1138

Los Angeles
310-234-2699
fax: 310-234-2670

San Francisco
415-403-9030
fax: 415-403-9033

Midwest
Derr Media Group
847-615-1921
fax: 847-735-1457

Southeast/Southwest
MancheeMedia
972-662-2503
fax: 972-662-2577

Detroit
Karen Teegarden & Associates
248-642-1773
fax: 248-642-6138

Canada
Fenn Company, Inc.
905-833-6200
fax: 905-833-2116

U.K.
The Powers Turner Group
+44-207-592-8331
fax: +44-207-630-6999

France and Switzerland
PEM-PEMA
+33-1-4143-8300
fax: +33-1-4143-8330

Germany
Publicitas Germany GmbH
+49-69-71-91-49-0
fax: +49-69-71-91-49-30

Sweden
Andrew Karnig & Associates
+46-8-442-7050
fax: +49-8-442-7059

Belgium
Publicitas Media S.A.
+32-2-639-8445
fax: +32-2-639-8456

Middle East and India
Peter Smith Media & Marketing
+44-140-484-1321
fax: +44-140-484-1320

Japan
Pacific Business, Inc.
+813-3661-6138
fax: +813-3661-6139

Korea
Biscom, Inc.
+822-739-7840
fax: +822-732-3662

Hong Kong
Hutton Media Limited
+852-2528-9135
fax: +852-2528-9281

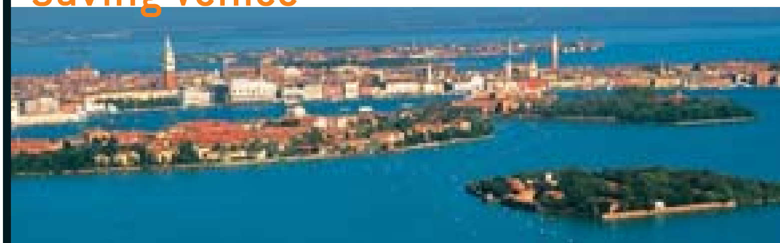
On the Web

WWW.SCIENTIFICAMERICAN.COM

FEATURED THIS MONTH

Visit **www.sciam.com/explore_directory.cfm** to find these recent additions to the site:

Saving Venice



© 1999 CONSORZIO VENEZIA NUOVA

By the end of the century, Venice could be a modern Atlantis. The picturesque Italian city is sinking as a result of geological plate shifting, which, along with now abandoned industrial practices that lowered land levels an entire foot in two decades, has left the famed St. Mark's Square hovering just two inches above the normal high-water mark. Project Moses, a controversial $3-billion government-funded plan to keep Venice from drowning, has finally received the green light from Italian officials. Not everyone approves of the floodgate scheme, however. Some scientists argue that it will harm local ecosystems. Furthermore, others contend, the gates won't be able to cope with the sea-level increases predicted by climate-change models.

NANOMACHINES FROM NATURE

Billions of years of evolution have left viruses well equipped to invade and multiply. But emptied of their infectious nucleic acids, the microorganisms can actually be put to good use, serving as highly modifiable and versatile additions to the nanoengineer's toolbox. Indeed, researchers are now using viral machinery to develop clever applications in medical imaging and drug delivery, as well as new approaches to building electronic devices.

ASK THE EXPERTS

What are the odds of a dead animal becoming fossilized?

Paleontologist Gregory M. Erickson of Florida State University explains.
www.sciam.com/askexpert_directory.cfm

SCIENTIFIC AMERICAN CAREERS

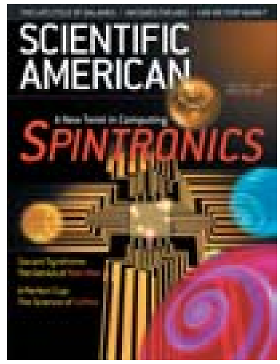
Looking to make a career move within the science and technology sectors?

Visit Scientific American Careers for positions in computers, sales and marketing, research and more.
www.sciam.com/careers

Search our database for hundreds of open positions. Better yet, let the right employers find you—POST YOUR RÉSUMÉ TODAY.
sciamjobs.sciam.com/texis/profile

Letters

EDITORS@SCIAM.COM



"AS A PERSON WHO ENJOYS Turkish coffee habitually, I was aghast to read in the otherwise excellent 'The Complexity of Coffee,' by Ernesto Illy [June 2002], that Turkish coffee is made in a special pot called an *ibrik*." Apparently that term is used only in the West, according to Selim Kusefoglu, chair of the chemistry department at the University of Bogazici in Istanbul. "An *ibrik* is used in a Turkish bath, another delightful custom, and is a metal container for holding water and should never be heated. Coffee, on the other hand, is made in a pot called a *cezve,* which has a straight, long handle and a side spout, a humble example of which, along with a few days' supply of Turkish coffee, is included with my letter. Illy's recipe is excellent, so please follow it. I hope you enjoy your Turkish coffee!" We found it to be a fine beverage choice for reading letters about the June 2002 issue, presented on the following pages.

MORE COFFEE TALK

Coffee is consumed especially by scientists, and Ernesto Illy is in a long tradition of researchers who turn their attention to the drink that literally stimulates them. One of the first and most eloquent was Benjamin Thompson, Count Rumford, who in 1812 wrote "On the Excellent Qualities of Coffee and the Art of Making It in the Highest Perfection." This essay is excerpted in *But the Crackling Is Superb,* an anthology by members of the Royal Society of Great Britain that is recommended reading for anyone who enjoys science with their eating and drinking.

Bruce Bayly
Tucson, Ariz.

THE MATH ON FALSE POSITIVES

"Lifting the Screen," by Alison McCook [News Scan], on screening for ovarian cancer, did not make the point clearly. The following should have been explicitly stated: despite the test's perfect sensitivity (all cases of ovarian cancer are detected) and its apparently high specificity of 95 percent (only 5 percent of women who do not have ovarian cancer will test positive), the specificity is still far too low considering that only one in 2,500 American women older than 35 have the disease. This is because for every 2,500 women tested, the one with cancer will test positive, and 5 percent of 2,500, or 125, women who do not have cancer will also test positive. That is, for every 126 women who test positive, only one will actually have cancer. Therefore, any individual positive test has less than a 1 percent chance of being correct.

Mark Herman
Shepherd, Mich.

BRING BACK DDT?

In a recent *Wall Street Journal* article, I was interested to read that "Malaria Strikes Growing Number of U.S. Travelers." I recalled the SA Perspectives "A Death Every 30 Seconds." Coincidentally or by design, in the same issue, in 50, 100 & 150 Years Ago, "Malaria, Italian-Style" notes the eradication of malaria in Italy with DDT and related insecticides. Although I am aware of the impact DDT had on wildlife and particularly on raptors, I think it's time to take it out of the closet and distribute it to these countries that are suffering such huge human and economic losses.

J. W. Heidacher
Hilton Head, S.C.

THOUGHTS ON AGING

I am puzzled by the contention in the essay "No Truth to the Fountain of Youth," by S. Jay Olshansky, Leonard Hayflick and Bruce A. Carnes, that there is no genetic component to aging. Why then do other sophisticated mammals have radically different life spans than humans do? My dog, for example, has an expected life span of 15 years with the best medical care that I can provide him. I will outlive him by a factor of five, even though we are both exposed to roughly the same environmental conditions.

James E. Lake
Tacoma, Wash.

I disagree with the assertion that "evolution is totally blind to the consequences of gene action (whether good, bad or indifferent) after reproduction is achieved." This may be true in the case of most earthly organisms, but in social mammals such as humans the course of aging of the elderly members of the community has a direct and significant impact on their descendants, whose lives they share on a daily basis. The elderly can enhance the group's chances of survival with the help of experience and information that they've gained in their own long lives. They can also decrease the group's chances by consuming too many of the available resources. I think it's likely that the aging members of a community of humans (and probably of chimpanzees, dogs, hyenas and others) considerably affect the reproductive success of their own direct descendants—and the continuation of the genes they gave them.

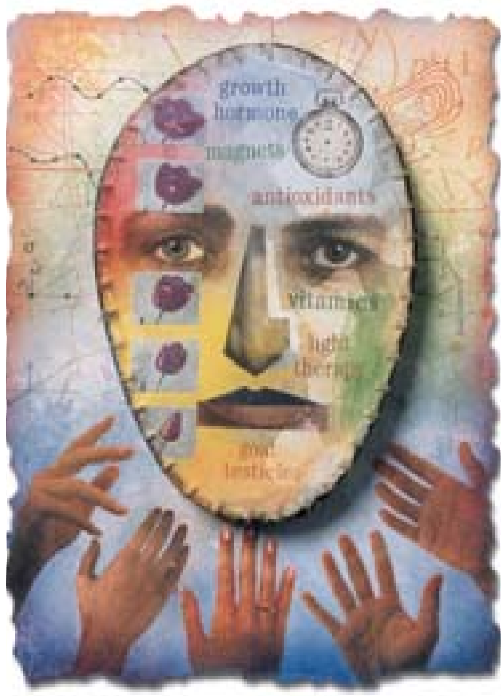
P. Rhiannon Griffith
Albuquerque, N.M.

The authors argue that genetic alterations to various model organisms—including fruit flies, whose average life span increased—did not affect the exponential increase in the risk of dying during adulthood. This is an important point, because the exponential increase in mortality is one of the widely accepted measures of aging in experimental research. In 1996 we and our colleague T. J. Nusbaum published an analysis of this parameter in genetically longer-lived fruit flies, finding that it was indeed altered in the way required by Olshansky et al. Presumably they will now be slightly more optimistic about the prospects for anti-aging medicine?

Michael R. Rose
Laurence D. Mueller
Department of Ecology and Evolutionary Biology
University of California, Irvine

The claim that "the primary goal of biomedical research and efforts to slow aging should not be the mere extension of life. It should be to prolong the duration of *healthy* life" must really warm the hearts of old people who have chronic illnesses but nonetheless have the temerity to find their lives well worth living … and prolonging.

Felicia Ackerman
Department of Philosophy
Brown University



ANTI-AGING REMEDIES: So far they're ineffective.

The authors warn against anti-aging fads, and their efforts are laudable. Nevertheless, is it not inevitable that in some future era our biological clock will be localized, characterized and turned off? Immortality! Many eagerly await that, but not I. A life without end would be a life of terminal ennui. Death is Tolkien's "gift of Iluvatar" that gives life its meaning.

Charles J. Savoca
Venice, Fla.

OLSHANSKY, HAYFLICK AND CARNES REPLY: Lake and Griffith fail to consider the critical distinction that must be made between the processes that cause aging and those that determine a species' longevity. The differences in the longevity of species are driven by the genes that determine growth and development, which influence longevity indirectly. That is why breeds of dogs larger than those of Lake, which also enjoy the same good care, will age and die well before 15 years. Once Lake and his dog reached sexual maturation, the molecular fidelity that both achieved during their genetically driven development began to succumb to random losses in the chemical energy necessary to maintain that fidelity. In an analogous fashion, our cars require a blueprint (the equivalent of genes in organisms) for their construction but do not require instructions on how to age.

As Griffith asserts, older members of social species can and do influence the survival of younger members. There is no evidence, however, that on an evolutionary timescale, assistance from older members leads to progressive increases in a species' longevity.

The point made by Rose and Mueller applies to actuarial aging (as measured by the rate of increase in the death rate by age); it has not been shown to apply to biological aging. As such, we are not "more optimistic about anti-aging medicine," because we do not think that humans come close to being the biological equivalent of big fruit flies.

Ackerman misunderstood our commitment to the health and welfare of the elderly. Our emphasis on quality of life, rather than length of life, is motivated by a deep concern for the toll that the nonfatal chronic conditions of aging take on mental and physical health as well as the economic consequences that are accompanying our rapidly expanding population of older people.

ERRATUM In "Divide and Vitrify," by Steven Ashley [News Scan], Mark A. Gilbertson is misidentified as director of the U.S. Department of Energy's Office of Environmental Management. His correct title is director of the Office of Basic and Applied Research in the Office of Environmental Management.

J. W. STEWART

50, 100 & 150 Years Ago
FROM SCIENTIFIC AMERICAN

Diphtheria Lethality ▪ Television's Ancient Ancestor ▪ A Diamond's Life

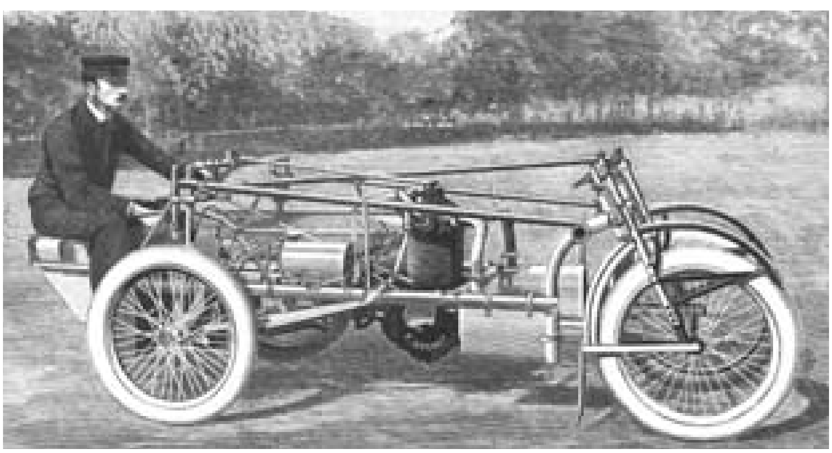
OCTOBER 1952

HOW DIPHTHERIA KILLS—"The substance secreted by the diphtheria bacillus is one of the most potent poisons known: one milligram of it is enough to kill 3½ tons of guinea pig. How does it work? Results from the diphtheria experiments with the Cecropia silkworm have been striking. The dormant pupa, which contains little cytochrome, will survive 70 micrograms of toxin for more than four weeks. Still more dramatic is the effect of toxin on the developing Cecropia adult. Although death may not come for days, the development of the insect is brought to a stop within a matter of hours. We assume that diphtheria toxin acts not by inhibiting any cytochrome component already formed, but by preventing the synthesis of new cytochrome."

OCTOBER 1902

(VERY) EARLY TELEVISION—"A Belgian engineer whose name is not known has devised a means to see electrically through long distances, just as we hear electrically by means of the telephone. At the transmitting station a rapidly rotating lens traverses, in a spiral pattern, forty times in each second, the surface of the body to which it is exposed. The lens is fitted with a screen so that only a small portion of its surface is exposed at any time. A selenium composition, the electric conductivity of which varies according to the intensity of the light to which it is exposed, is placed on the axis of rotation. At the receiving station is placed a conducting body and another lens, electrically synchronized with the first. The luminous image of the receiving body is projected in a spiral pattern on a white screen." [*Editors' note: This appears to have been a working version of the electromechanical "television" patented by German scientist Paul Gottlieb Nipkow in 1884.*]

RACING AUTOMOBILE—"The Truffault machine is constructed with the greatest simplicity. The machine was officially tested at Deauville in the 600-mile race, where it attained a speed of 51½ miles per hour and won the first place. The machine we illustrate is an experimental model in which the inventor has tried to ease as much as possible the terrible shocks and jars so familiar to all those who have taken long trips in these rapid and light vehicles. It is to be hoped that this experimental vehicle will, with some modifications, soon become an industrial one." [*Editors' note: J.M.M. Truffault designed and used one of the first shock absorbers.*]



TRUFFAULT RACER: The experimental model, 1902

OCTOBER 1852

WEAK STOMACHS—"The permanent opening made in the stomach of a soldier in Canada by a musket ball [*sic*], and described by Mr. William Beaumont, as well as experiments performed with animals, prove irrefragably that the process of digestion in animals which resemble man in their organization, is the same whether the action goes on in the stomach or in a vessel. It follows from this that it is very easy to obtain any quantity of the gastric juice, preferably from living animals. By this means, invalids and others, troubled with dyspepsia, may be supplied with the means of digestion."

RE-CUTTING THE KOH-I-NOOR—"This celebrated diamond, which created such a sensation in the Great Exhibition at Crystal Palace, was found to be very improperly cut, and did not exhibit half of its beauty. Consultation with the Queen, Prince Albert, and eminent scientific men were had, to see if it could be safely re-cut and improved. All the diamond cutting in the world, it seems, is done in Holland, by eminent and long practiced lapidaries, and the most famous of them, a person of the Jewish persuasion, was sent for, and consulted about the safety and certainty of cutting the famous 'mountain of light.' By late news from Europe we learn that the labor is now finished. It is now unsurpassed by any diamond above ground in shape, lustre, and beauty."

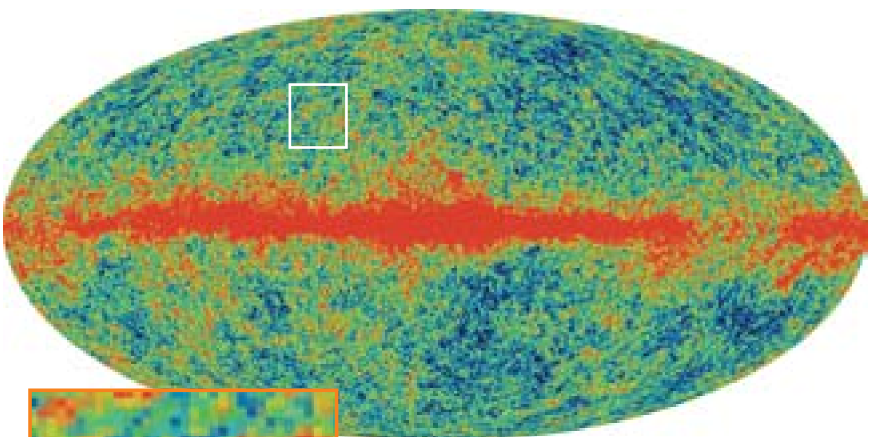
GARMENT WORKERS—"In Ulster, Ireland, and westwards, the embroidery trade is giving employment to a quarter of a million individuals. The females are almost invariably employed in their own homes under the eyes of their parents and friends, and they can thus obtain a livelihood without endangering their morals."

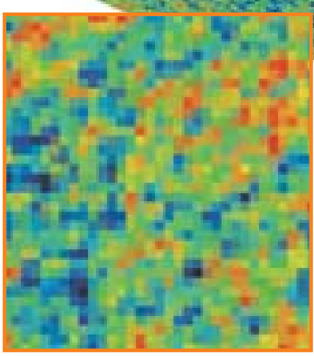
news SCAN

PHYSICS

A Pixelated Cosmos

HOW THE MICROWAVE BACKGROUND COULD HELP PROVE STRING THEORY **BY GEORGE MUSSER**





NOT A PRINTING ERROR: **If quantum gravity theories are right, the cosmic microwave background (*simulation shown*) might truly be a mosaic of pixels.**

NASA GODDARD SPACE FLIGHT CENTER

String theory has certainly had its detractors. It has been called an exercise in "recreational mathematical theology," a reprise of "the Dark Ages," a surrender to "the tyranny of belief," and a cryptophilosophical "ironic science." Any theory claiming to be an all-encompassing theory of everything would arouse people's contrarian instincts, but the rhetoric reflects a serious concern: How can a theory that deals in objects as small as 10^{-35} meter ever be tested, when particle accelerators lack the energy to probe anything smaller than 10^{-19} meter?

Over the past several years, though, cynicism has become harder to sustain. String theory and complementary efforts to produce a quantum theory of gravity have racked up conceptual successes. What is more, practitioners have brainstormed ways to test such theories, most recently by using the cosmic microwave background radiation. "Even though it's a long shot, the fact you can say the words 'string theory' and 'observation' in the same sentence is seductive," says Brian R. Greene of Columbia University, a leading string theorist.

Like other cosmological measurements, the newly proposed tests take advantage of the subtle unevenness of the microwave background. That unevenness is thought to originate during inflation, a burst of growth that the universe seems to have undergone early in its history. The energy field that drove inflation fluctuated in the way that all quantum fields do. Under ordinary circumstances, such fluctuations would have averaged out too quickly to be noticed, but cosmic expansion threw them off kilter, stretching them, weakening them and eventually locking them in place, like waves on a frozen pond.

String theory and related paradigms extend this picture by supposing that distances cannot be subdivided into chunks smaller than perhaps 10^{-35} meter across. Like a watercolor painting, in which brush strokes bleed together, space cannot accommodate an infinite amount of detail. If you could take an object and enlarge it enough, its boundaries would look blurry. And that is precisely what cosmic expansion does. If the universe grew by a factor of 10^{26} during inflation and an equivalent amount afterward, a fluctuation 10^{-35} meter across would now be a dozen light-years in size.

Greene and his colleagues Richard Easther and William H. Kinney, along with Gary Shiu of the University of Pennsylvania, have considered the circumstances under which this effect might be visible. Fluctuations as large as 10^{-32} meter might be blurred; inflation, at its most frenetic, would freeze those fluctuations once they had grown to 10^{-30} meter. During this 100-fold growth, the blurriness would become proportionally less conspicuous, leading the distribution of fluctuations to deviate by 1 percent from standard predictions. That might just show up in data from the Microwave Anisotropy Probe or the follow-up Planck satellite.

Two other groups—Achim Kempf of the University of Waterloo and Jens C. Niemeyer of the Max Planck Institute for Astrophysics in Garching, and Nemanja Kaloper, Matthew Kleban, Albion Lawrence and Stephen Shenker of Stanford University—argue that the effect is almost assuredly much smaller. But everyone agrees that we'll never know until we look. "This is an opportunity that should not be missed," Kaloper says.

Another idea, proposed by cosmologist Craig J. Hogan of the University of Washington, involves what is potentially a stronger and more distinctive phenomenon. It is based on one of the most profound concepts to have emerged from the nascent quantum gravity theories: the holographic principle, which restricts the amount of information a region of spacetime can contain. The amount depends not on its volume but, oddly, on the area of its boundary. Each quantum of area (10^{-35} meter on a side) can store one bit of information.

The principle even applies to the entire universe. During inflation, the freezing of fluctuations defined the effective boundary of space. At a distance of 10^{-30} meter, the boundary would have had an area of 10^{10} quanta, representing a gigabyte of data. That gigabyte would encode all the fluctuations we now see. If observers look at the microwave background closely enough, they might notice pixels or discrete colors, as though the sky were one great big computer screen. Although these numbers are guesswork, the most prominent pixels—those that encode the largest fluctuations—are the least dependent on uncertain parameters. Hogan estimates that they would always account for roughly 10 kilobytes, no more than a smallish computer image.

Even researchers who question the details agree with the basic point: Quantum gravity is no longer consigned to scribbles on a chalkboard. In fact, the fundamental nature of space and time might be written on the sky, and the entire initial state of the universe could be burned onto a single CD-ROM.

OTHER TESTS FOR QUANTUM GRAVITY

- **Cosmic particles:** Astronomers have detected cosmic- and gamma-ray particles with energies higher than any man-made accelerator could muster. Moreover, these particles travel huge distances, allowing any quantum gravity effects to accumulate.
- **Other oddball particles:** Neutral kaons involve such a fine balance of quantum effects that even a tiny push from quantum gravity might be observable.
- **Gravitational-wave observatories:** The next generation of instruments will gauge distances with such precision that they might be sensitive to the discrete nature of space.
- **Dark matter:** The missing mass of the universe is almost certainly a sign of exotic physics, either particles of a type consistent with string theory or corrections to existing laws of gravity.

ENVIRONMENT

From Flush to Farm

SEWAGE IS A GREAT FERTILIZER, BUT IS IT A HEALTH HAZARD? BY REBECCA RENNER

Ten years ago the U.S. stopped dumping sewage sludge in the ocean because of concerns about polluting the marine ecosystem. Since that time, the dregs from our drains have been going to farmland—as fertilizer. This practice has been contentious from the onset. Advocates enthuse about the success of sludge recycling. Opponents cite health complaints from those living nearby. But in terms of the science, "we are doing something on a big scale, and we don't know enough about it," says Thomas A. Burke, a public health professor at Johns Hopkins University.

Treated sludge, also known as biosolids, makes good fertilizer because it is high in organic content and plant nutrients. But sludge also harbors low levels of metals, organic pollutants and disease-causing microbes, so the U.S. Environmental Protection Agency has regulated its use under Part 503, a 1993 regulation of the Clean Water Act. The rule divides treated sludge into two classes. Class A sludge contains no detectable pathogens and can be used anywhere. Class B sludge, which accounts for the bulk of the fertilizer, is treated to reduce pathogen levels to below certain thresholds.

SEWAGE SLUDGE: WHAT A WASTE

There are good reasons to turn sludge into fertilizer: it saves money for farmers and water companies and conserves space in landfills. "Using sludge as fertilizer has got to be one of the most effective recycling programs ever instituted in this country," notes Greg Kester, who oversees Wisconsin's biosolids program.


LIMED SEWAGE SLUDGE dumped on a farm in DeSoto County, Florida, produces airborne dust that could affect those living nearby.

The EPA stipulates limited public access to sites where class B sludge has been applied.

Burke has spent the past two years chairing a National Research Council (NRC) committee to assess the scientific basis for the Part 503 rule and whether it adequately protects public health. The committee's findings, issued in July, are mixed. There is no documented scientific evidence that the rule has failed to protect public health. But then, no government agency has investigated, or even tried to track, health complaints. Moreover, there has been no risk assessment to justify the pathogen standards. Instead the EPA based the standards on historical observations of publicly restricted farm fields using anaerobically digested sludge. (Class B sludge breaks down both aerobically and anaerobically.)

These gaps don't necessarily mean that the program isn't working, says committee member Charles N. Haas, a microbiologist at Drexel University. "There is a long list of research to bring the biosolids rule up to the scientific standard of other EPA regulations. But the current rule works," he insists.

But the gaps indicated by the NRC report worry many activists and a smaller number of scientists who believe that exposure to sludge is making people ill. Cornell Waste Management Institute in Ithaca, N.Y., has compiled a database of more than 39 incidents in 15 states, affecting 328 people, as of August. The sick people who live near sludge-spreading operations complain of a common set of symptoms, according to Ellen Z. Harrison, the institute's director. Most frequent are respiratory and gastrointestinal symptoms, skin disorders and headaches.

Workers at wastewater treatment plants also experience gastrointestinal problems when they first start the job. This may lend some credence to the health complaints about sludge, says Joseph C. Cocalis, who recently retired from the National Institute for Occupational Safety and Health in Morgantown, W.Va. "Sludge workers have GI problems for about a month after they first start. Then they seem to acquire immunity," he explains.

According to the NRC report, the scientific issues are further complicated because the EPA doesn't enforce the sludge rule, a point acknowledged by the EPA. "Land application of sewage sludge is relatively safe compared with the other activities we regulate. It's a low priority for us," says Alan Hais of the EPA's Office of Water. The agency intends to publish a research plan in response to the NRC report. "We'll do the studies," Hais states. "If we find the risk is higher than we estimated, we'll change our approach to enforcement and compliance."

Rebecca Renner writes about environmental issues from Williamsport, Pa.

SICKNESS IN THE AIR?

The most vocal critic of using sludge as fertilizer is David L. Lewis, a microbiologist on leave from the EPA. He believes that people who live downwind of fields fertilized with sludge may face a two-pronged attack on their health. Chemicals such as ammonia that are used to treat the sludge irritate the respiratory system. This aggravation makes people more susceptible to airborne pathogens. It's just like hay fever season, he says, but worse: class B sludge "comes ready-made with the very bacteria and viruses that cause these infections."

UNIVERSITY OF GEORGIA/UNIVERSITY OF GEORGIA, SPECIAL

PUBLIC HEALTH

The Next Wave of AIDS

IGNORANCE AND DRUG RESISTANCE MAY WORSEN THE CRISIS BY LUIS MIGUEL ARIZA

The path to an AIDS-free world is looking especially bumpy. Various participants in the 14th International AIDS Conference, held in Barcelona this past July, suggest that, for several reasons, a global surge of cases could be in store.

The U.S. Centers for Disease Control and Prevention presented evidence that the virus is spreading quickly among homosexual men, infecting nine for every one heterosexual male. Amazingly, three fourths of young gay men do not know they are infected; overall around 230,000 Americans are unaware that they are HIV-positive, regardless of their sexual orientation. (About 900,000 in the U.S. are HIV-positive.)

What's worrisome is that many of them admit to practicing unprotected sex. Risky behaviors are associated with violence, drug abuse and depression, but there might be a simpler reason for the virus rebound in young men: they are not scared of AIDS. "If you ask me if this is related to the increasing perception in the wealthier nations that AIDS is no longer invariably fatal, I would say yes," comments Jonathan Kagan, deputy director of the AIDS division of the National Institute of Allergy and Infectious Diseases (NIAID). There are 19 anti-HIV drugs approved, and some combinations lower the virus to undetectable levels in the blood, perhaps giving the false impression that AIDS can be cured. "To many, the sense of threat and urgency is diminished by this success," Kagan says.

REMEMBERING THOSE who have succumbed was part of the opening ceremony of the 14th International AIDS Conference in Barcelona.

The fact is that the disease is intrinsically incurable, meaning that therapies cannot eliminate the reservoirs of HIV in the body, explains Robert F. Siciliano of the Johns Hopkins University School of Medicine. Helper immune cells called CD4 cells are often killed immediately by the virus, but some enter a resting state after being infected. These memory cells and their progeny are designed to live for a long time, so the virus could persist in them for decades. Anthony S. Fauci, director of NIAID, unveiled data that show that the virus replicates inside these cells, even in patients with no detectable HIV in their blood.

The experts in Barcelona also expressed concern about the emergence of resistant strains. HIV can produce an astounding 10 million to 100 million variants daily in the human body. If the drugs cannot keep the viral replication below 50 particles per cubic millimeter, resistance is inevitable. A study of blood samples of 1,908 people infected with HIV showed that after two years of treatment, 78 percent developed resistance to one drug and more than half to combinations of medicines. About 100,000 in the U.S. could be infected with resistant strains of the virus.

Drug resistance will be a major issue in Africa. Bernhard Schwartländer, director of the HIV/AIDS program at the World Health Organization, announced that three million people could receive drug therapies by 2005. Less than 2 percent of Africans living with HIV or AIDS, about 50,000 people, are actually being treated, but resistant strains have already emerged. In 68 patients in Ivory Coast, resistance to one drug was found in 57 percent of the blood samples. In a Uganda study, some 19 percent of pregnant women showed resistance to nevirapine, a drug approved in 1996 to prevent mother-to-child transmission, two months after treatment. Nevertheless, the benefits of these drugs clearly outweigh the possibility of resistance, says Lynn Morris of the National Institute for Virology in Johannesburg: "My concern is that resistance can be used as an excuse not to treat."

Before the therapies arrive, WHO wants to establish a network of regional labs to know which resistant strains are circulating in Africa. "We need to invest resources to develop new drugs, so when resistance arises to first- or second-generation drugs, you can replace them with a drug" to which HIV is not resistant, Fauci remarks.

The pharmaceutical giant Roche described the results of clinical trials of one of these promising new drugs, the T-20 fusion inhibitor. T-20 is a synthetic peptide that blocks gp41, the protein that the virus uses to bind to the cell membrane. Twice-daily injections have produced good results with few side effects. But the drug has not yet been approved and could cost $11,000 annually—too much for Africa, where many live on less than $1 a day.

Two other large regions of the world are poised for an AIDS detonation. HIV cases in Russia have shot up by a factor of nearly 200 since 1995, from 1,047 to 197,497, according to the AIDS Foundation East-West, a nongovernmental organization focusing on international public health. Needle sharing among drug users is driving the increase. China could have 10 million new infections by the end of this decade, in part because 54 percent of the population do not know how the disease spreads. "These regions are like time bombs for the virus," Kagan laments. "Failure to control the epidemic in these countries will most likely result in millions more infections and deaths."

Luis Miguel Ariza is based in Madrid.

AFRICA AS ORPHANAGE

Peter Piot, executive director of UNAIDS, the United Nations program for the disease, called it "without doubt the most shocking report at this conference"—namely, that about 25 million children will lose one or both of their parents as a result of the AIDS epidemic in the next eight years. A joint effort by USAID, UNICEF and UNAIDS projects that in 2010, orphans will constitute 15 to 25 percent of the population younger than 15 years in 12 sub-Saharan African countries, mainly because of AIDS. Piot has appealed for a global response to help the children: "I've seen them taking the role of adults and working 14 hours a day. They saw how an entire generation of adults disappeared."

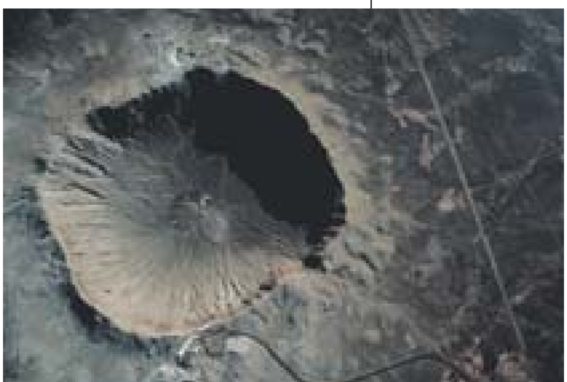
CESAR RANGEL *AP Photo*

NEAR-EARTH OBJECTS

Planetary Protection

X-RAY TESTS SHOW HOW TO DEFLECT AN INCOMING ASTEROID **BY STEVE NADIS**

John L. Remo has a modest goal: he'd like to save the planet. Unlike some delusional people who share his interest, Remo is a level-headed physicist, based at the Harvard-Smithsonian Center for Astrophysics, and his research might actually further that goal. Since the mid-1990s he and his colleagues at Sandia National Laboratories have conducted the first experiments aimed at seeing how momentum from high-intensity radiation bursts is transferred to meteorite fragments. With access to Sandia's Z machine, the world's most powerful x-ray generator, Remo and his team could guide efforts to divert an incoming asteroid or comet.



DEEP IMPACT: A meteorite only about 25 meters wide was still able to make a 1.2-kilometer-wide crater in Arizona.

A devastating collision with a near-earth object (NEO) may be only a matter of time. Consider asteroid 2002 MN: this past June the 100-meter-wide rock came within 120,000 kilometers of our planet. "That's almost too close for comfort," Remo says, especially considering that 2002 MN was discovered three days *after* its near miss. More unnerving were initial reports of asteroid NT7: this two-kilometer-wide rock swings by in 2019; if it were to collide, it would cause global havoc (the latest calculations indicate that it will miss).

Researchers have contemplated NEO mitigation or deflection for more than a decade, but discussions have been hampered by the lack of data. When Remo joined a deflection panel at Los Alamos National Laboratory in 1992, he emphasized the importance of understanding the material properties of NEOs to predict how they would react to an impulse.

Physicist Bruce A. Remington of Lawrence Livermore National Laboratory considers this kind of research long overdue. Despite years of debate, mitigation has remained an "abstract idea," Remington says. "Finally, people are getting real numbers that can help us figure out how much energy it would take to divert a menacing object." The problem, he adds, is too complicated to be calculated without seeing what happens experimentally.

For Remo, the crucial parameter is the "momentum coupling coefficient,"a gauge of the efficiency at which radiation striking an object is converted to kinetic energy. High-energy x-ray pulses produced by the Z machine impinge on the target material—six varieties of meteorites have been tested to date—boiling off the surface layer and creating a plasma jet that shoots backward. A momentum-conserving shock wave formed in its wake pushes the meteorite in the opposite direction. Remo, with Michael D. Furnish of Sandia, computed the velocity of these particles by measuring the Doppler shift of reflected laser light.

Because x-rays are a big component of nuclear blasts, the Z experiments are designed to simulate the detonation of a weapon near a threatening NEO to nudge it into a safe trajectory. Based on his computations of coupling coefficients, Remo believes that moderate-size nuclear explosives could do the job. A 25-kiloton device, for example, could move a one-kilometer-diameter object out of harm's way, assuming we had a few decades' advance notice. With longer lead times or smaller objects, nonnuclear options become more feasible.

There are, of course, serious challenges in scaling up results from centimeter-size shards to rocks hundreds of meters in diameter. Nevertheless, the coupling coefficients can be measured accurately in the lab, Furnish notes, because x-rays interact with matter on a microscopic scale—and that's true for giant asteroids and micrometeorites alike. Major uncertainty, however, stems from the question of whether asteroids are solid objects or loose assemblages of rocks. "If it's a rubble pile, you might move part of it the right way and other parts the wrong way," Furnish cautions.

Remo is devising experiments to investigate that possibility while also planning Z tests of different meteoritic and comet materials. Ultimately he hopes to turn the NEO peril into a straightforward physics and engineering problem. Rather than scaring people with forecasts of impending doom, Remo would like to tell them what can be done.

Steve Nadis is based in Cambridge, Mass.

OUT-OF-POCKET EXPENSES

In pursuing research on near-earth objects, John L. Remo has relied almost entirely on his own funds, underwritten by Quantum Resonance, a laser instrumentation company he runs in St. James, N.Y. "Using lasers and the Z machine to shock meteorites is so new, there weren't any programs to support it, because the work doesn't fit into any established research categories," explains Remo, a physicist affiliated with the Harvard-Smithsonian Center for Astrophysics. "If I had waited for funding, it would have taken years to start this work."

JONATHAN BLAIR *Corbis*

ROBOTICS

Human-Free Kick

AT ROBOCUP 2002, HUMANOIDS BATTLE IT OUT IN SOCCER **BY DENNIS NORMILE**

Foot-Prints, the Japanese striker, and Tao-Pie-Pie, the New Zealand goalkeeper, eyed each other as the ball was placed for the penalty kick. At the whistle, Foot-Prints sprang toward the ball, step by agonizingly slow step. Tao-Pie-Pie wobbled out to narrow the angle. Foot-Prints finally unleashed a nudge that shot the ball oh-so-slowly past Tao-Pie-Pie and just barely into the goal.

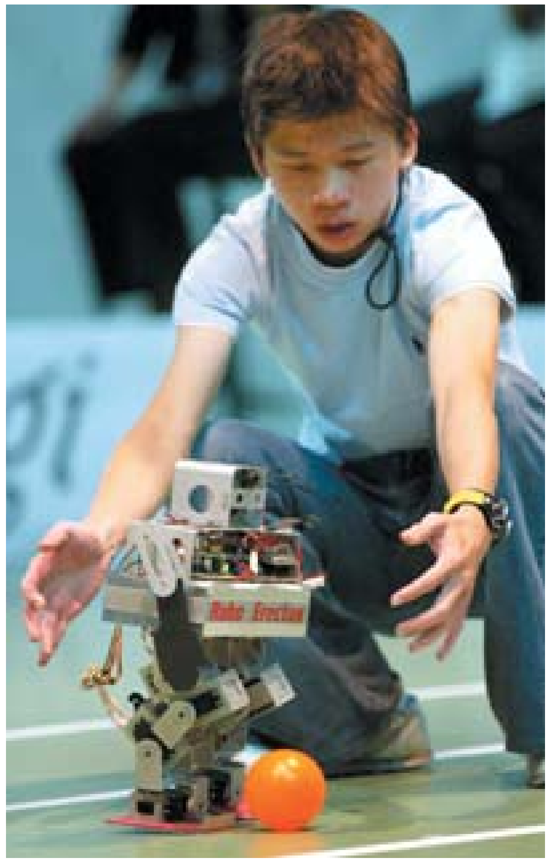
The crowd went wild, almost as if it were a World Cup match. Actually, it was RoboCup 2002. The annual robotic soccer tournament was held in Fukuoka, Japan, this past June as the World Cup was getting under way. The timing was no coincidence. "The goal of RoboCup is to develop a team of robots that can beat the human World Cup champions by 2050," says Hiroaki Kitano, a Sony artificial-intelligence specialist who is also president of the RoboCup Federation.

The notion of robots taking on Brazil would be laughable if roboticists around the world were not so enthusiastically answering the call. Kitano and his collaborators started RoboCup in 1997 with hopes that a grand challenge would spur advances in robotics and artificial intelligence. The first year only a couple dozen groups competed with wheeled robots and simulations, in which "players" were simulated on separate computers and a server governed the interactions. All the matches took place in one afternoon in the ballroom of a Nagoya hotel.

This year there were almost 200 teams. The matches were held in the Fukuoka Dome sports stadium, stretched over four days, and drew 127,000 spectators. More significantly, humanoid robots took to the RoboCup pitch for the first time. The two-leggers are not running yet, so they competed at walking speed—traversing a distance equal to five times their height, circling a pole and returning.

It was surprisingly entertaining, if not quite as exciting as the duel between Brazil's Ronaldo and Germany's Oliver Kahn. Some of the robots walked with the uncertainty of children taking their first steps; others, with the baby-step caution of the aged. Many of the robot makers hovered over their creations with outstretched hands, ready to catch a stumbler. (Human intervention netted a 30-second time penalty.) More soccerlike were the robot face-offs in penalty kicks. Because the robots' reactions take so much time, defense meant getting in front of the ball and hoping the kicks didn't go off at an angle.

The wheeled robots, which don't have to fritter precious computational power on balancing, can react in real time to moving balls. For pure efficiency, there would seem to be little reason to walk. So why bother with legs? That has been a perennial robotics question.



PENALTY SHOT is taken by Robo Erectus as one of its student builders from Singapore Polytechnic prepares to catch the bot should it stumble. This year's competition in Japan brought together 193 teams from 30 countries and regions.

SHIZUO KAMBAYASHI *AP Photo*

WINNERS ON THEIR FEET

Japan leads the way in humanoid robots, accounting for five of the 12 humanoid entries at RoboCup 2002. Rounding out the field were bots from Australia, Denmark, New Zealand, Sweden and Singapore. (The U.S. was not represented in that division, despite having dozens of teams in the wheeled and simulated robot competitions.) Japanese groups dominated the humanoid prizes: Foot-Prints, which was built by a pair of robot hobbyists, won the competition for humanoids 40 centimeters and smaller. Nagara, the creation of researchers at the Gifu (Prefecture) Industries Association, beat all comers in the 80-centimeter humanoid category.

"In the early 1980s there was a big debate in the U.S. over whether robots should look like humans or not," explains Christopher G. Atkeson, a roboticist at Carnegie Mellon University. At the time, researchers—and funding agencies—wanted robots for factory automation. Atkeson says the thinking was, "If the goal is to make VCRs, there is no need for robots to look human." He adds that the U.S. military wrote off the idea of robotic soldiers. "Work on humanoid robots in the U.S. has been slow to take off," he says.

In Japan, opting for legs or wheels has long depended simply on the application. Masato Hirose, who led the development of Asimo, Honda's quasi-autonomous walking humanoid, explains that Japanese researchers want a robot to assist humans with daily activities. "The merit of a humanoid is that it can go every place a human can go," he states, including up and down stairs and into confined spaces. He adds that robots that look human will also make interactions more natural.

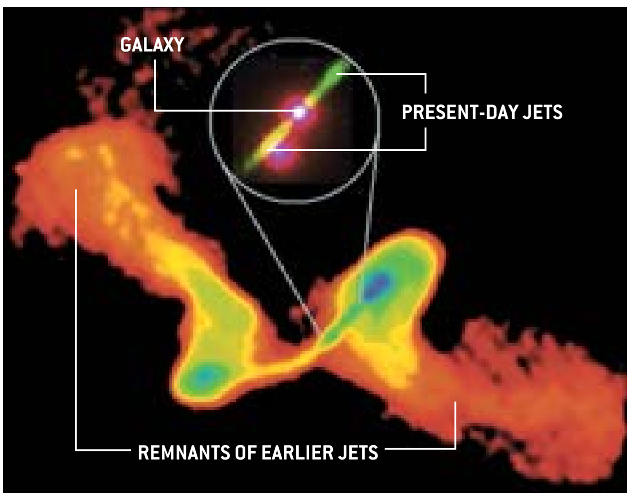
Kitano notes that humanoid success is not the final word. Japanese teams have not fared well in the wheeled and simulated robot competitions, where researchers from the U.S. and Europe have developed new approaches to robotic teamwork, among other techniques that rely on artificial intelligence. These advances could be quickly applied to humanoids, once their hardware is up to it. Beating the World Cup champs might require worldwide cooperation.

Dennis Normile is based in Tokyo.

ASTRONOMY

"X" Marks the Spot

SHIFTING RADIO JETS MAY SIGNAL THE COALESCENCE OF BLACK HOLES **BY JR MINKEL**



MERGER OF BLACK HOLES may have shifted the jets from radio galaxy NGC 326. The jets initially pointed to the 10 o'clock and 4 o'clock positions; they now point to 8 o'clock and 2 o'clock.

Researchers have assumed that supermassive black holes in the cores of galaxies come together when two galaxies collide, but they didn't have any evidence of the process. Now two astrophysicists, David R. Merritt of Rutgers University and Ronald D. Ekers of the Australia Telescope National Facility of CSIRO, argue that there are signs of such collisions—in the form of oddly shaped outflow jets from active galaxies. They propose that the direction of the jets, which are strong radio sources, shifts when a larger black hole absorbs a smaller one.

These jets, which result when matter spirals into black holes, are thought to align with a hole's spin axis. The researchers deduce that even a small black hole could cause its bigger partner to rotate when the two merge, thereby changing the outflows from an "I" to a distorted "X" shape. Given the number of galaxies displaying this characteristic and the 100-million-year lifetime of jets, they estimate that one merger occurs per year—useful information for those proposing gravitational-wave detectors. *Science* published the result online August 1.

JR Minkel is based in New York City.

BY THE NUMBERS

Quality of Life

IS THE U.S. THE BEST PLACE TO LIVE? BY RODGER DOYLE

Standard economic measures such as gross domestic product per capita and median family income were not designed to gauge the material quality of life. They don't, for instance, take into account inequality of income or damage to the environment. To get a better sense of how people experience everyday life, scholars have devised more sophisticated indices. One of the best examples comes from Lars Osberg of Dalhousie University in Nova Scotia and Andrew Sharpe of the Center for the Study of Living Standards in Ottawa.

They have measured economic well-being over time for 14 countries, using four classes of indicators: consumption (both private and governmental), wealth (which includes such diverse factors as housing and the social cost of environmental degradation), economic equality (measured by income distribution and degree of poverty), and security about future income (measured by, for example, risk of unemployment and illness). Their data for five of these countries show the U.S. with a somewhat less favorable trend since 1980 than that of Norway but better than that of the U.K. and Sweden [*see left chart*].

Some of the variations in the chart represent cyclical changes in business activity. The longer-term trends reflect a variety of factors. The favorable direction for Norway, for instance, results from higher consumption, wealth and security, whereas the poor performance of Sweden stems largely from increases in inequality and insecurity, combined with a mediocre increase in consumption.

The usefulness of the index is in raising questions such as, "Why is the U.S. at a lower level than Norway?" One component of the index suggests part of the answer [*see right chart*]. These may in turn point up other disparities. Why, for example, is financial security in the U.S. lower than in Canada in spite of a more robust U.S. economy? Data on the component of economic equality might prompt one to ask why it is falling in most of the 14 countries.

The Osberg-Sharpe indicators measure average quality of economic life and so tell us nothing about the poor or the rich. Economist Timothy M. Smeeding of Syracuse University and sociologist Lee Rainwater of Harvard University have explored this aspect by measuring the economic prospects of children whose family income is at the 10th, 50th and 90th percentiles of income distribution—in other words, poor, average and rich. Their data, which cover 13 industrial countries for the early and mid-1990s, show that the U.S. is the best place to be a rich child, but for the poor child the best place is Norway, which makes substantial cash payments to families. Poor children in the U.S. have worse prospects than their counterparts in all these countries but the U.K. The prospects of the average child, however, are better in the U.S. than in any of the other countries except Switzerland and Canada (the complete data for the 13 countries can be seen at www.sciam.com).

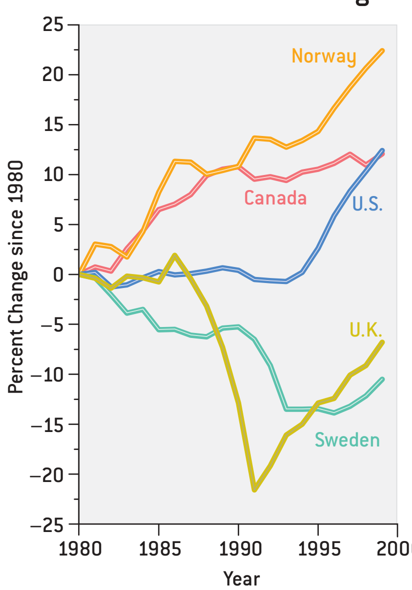
Next month: A look at subjective measures of well-being.

Rodger Doyle can be reached at rdoyle2@adelphia.net

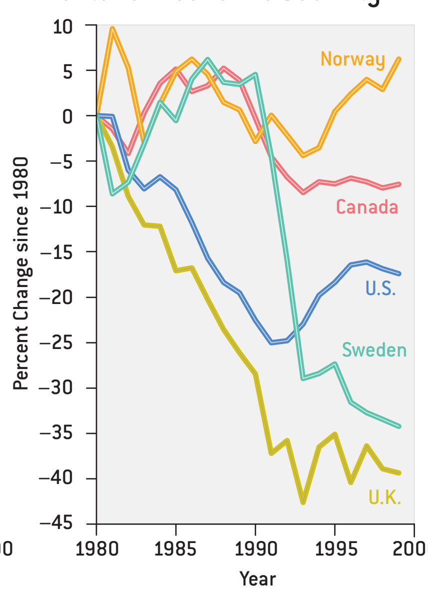
MORE TO EXPLORE

- **International Comparisons of Trends in Economic Well-Being.** Lars Osberg and Andrew Sharpe. Presented at the American Economic Association conference, January 2000. Available from the Center for the Study of Living Standards: www.csls.ca
- **Comparing Living Standards across Nations: Real Incomes at the Top, the Bottom, and the Middle.** Timothy M. Smeeding and Lee Rainwater. Revised edition at www.cpr.maxwell.syr.edu/faculty/smeeding
- See also the Webcast from the Jerome Levy Economics Institute conference on the quality of life, June 2001: www.levy.org/webcast/webcast.html

Index of Economic Well-Being



Indicator of Economic Security



SOURCE: Lars Osberg and Andrew Sharpe and the Center for the Study of Living Standards

RODGER DOYLE

BIOTECH

The Fat Just Melts Away

Mutant mice can now emulate those forever-slim folks who eat whatever they want. Rodents lacking a single fatty-acid-producing gene called *SCD-1* gorged on high-fat, sucrose-rich diets without packing on the pounds or sending their blood into a diabetes-inducing sugar rush. Instead they seemed to burn up the excess calories, judging by their oxygen consumption. The skin and eyes of the animals became dry as time went on, but those mice producing half the normal amount of the enzyme gained less weight than normal rodents and did not have obvious side effects, says lead investigator James M. Ntambi of the University of Wisconsin–Madison. This suggests that a tolerable drug to protect people from obesity and diabetes might be found, he explains. Human *SCD-1* is currently being analyzed. The study appeared in the August 12 online version of the *Proceedings of the National Academy of Sciences USA*.

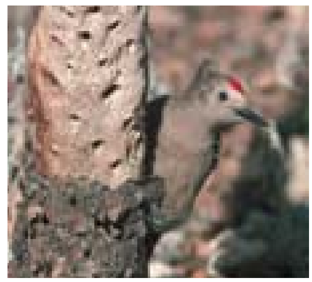
—JR Minkel

ECOLOGY

Sidling Up to the Rich

If you want to attract birds, think upper crust. Ann P. Kinzig and Paige S. Warren of Arizona State University found that the birds of Phoenix prefer the greenery of well-to-do neighborhoods over that in lower-income areas. Parks of the well-heeled contained an average of 28.2 species year-round, compared with 17.5 in depressed locales; middle-class parks fell in between, attracting 23.2 species. The researchers thought that the abundance and diversity of trees caused the disparities. Surprisingly, the vegetation factors did not correlate with the bird data—in fact, poor neighborhoods had a greater diversity of trees. It isn't the snazzy address that draws the feathers, of course. Socioeconomic status, Kinzig notes, is a surrogate for many possible reasons, such as landscaping or commercial activity, that may affect avian preferences. The results were presented at the August meeting of the Ecological Society of America.

—Philip Yam

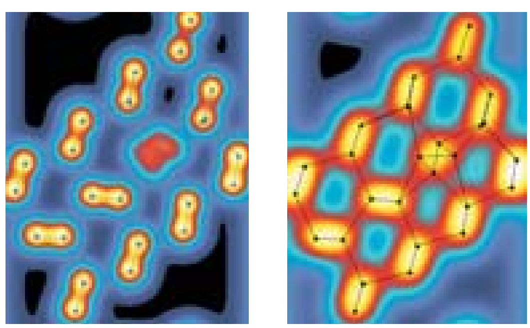


CITY BIRDS such as this Gila woodpecker prefer parks in affluent neighborhoods.

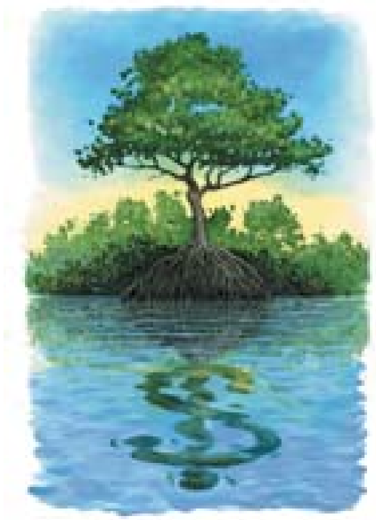
IMAGING

Aberration Negation

Like their optical counterparts, the magnetic lenses inside electron microscopes suffer from imperfections that blur images. Now researchers at the IBM Watson research center in Yorktown Heights, N.Y., and Nion R&D in Kirkland, Wash., have used multiple lenses and sophisticated software to counter the strongest of these flaws: spherical aberrations. The resulting electron beam is finer than a hydrogen atom and allows the first direct imaging of structures smaller than an angstrom (0.1 nanometer) by electron microscopy. As the dimensions of computer-chip elements shrink, scientists will need such resolution to view and fix atomic-level material defects, including missing or extra atoms. Previous studies had offered only imperfect glimpses of these defects. The August 8 *Nature* reports the advance. *—JR Minkel*



RESOLVED: Aberration correction reveals silicon atoms more distinctly (*left*) than does uncorrected imaging (*right*).



DATA POINTS: NATURE'S P/E RATIO

Bulldozing nature to create farms has short-term financial benefits, but it exacts an unprofitable long-term cost. Economically, it's better to keep nature as is: forests, swamps and reefs control flooding, absorb carbon dioxide and attract tourists. A new analysis by Andrew Balmford of the University of Cambridge and his colleagues quantifies some of the advantages.

Total economic value per hectare of
Forests: $2,570
Farms on former forests: $2,110

Mangroves: $60,400
Shrimp farms on former mangroves: $16,700

Wetlands: $8,800
Farms on former wetlands: $3,700

Percent of world's land that is reserves: 7.9
Percent needed to ensure future of wild nature: 15

Estimated annual cost to maintain: $20 billion to $28 billion
Estimated annual value of their goods and services: $4.4 trillion to $5.2 trillion

SOURCE: Science, August 9, 2002

TOM BRAKEFIELD *Corbis* (*top*); IBM (*bottom*); ILLUSTRATION BY MATT COLLINS

Phylis Morrison, 1927–2002


SETU SHAH

Phylis Morrison, a contributor to this magazine for decades and a dynamo of science education, passed away on July 9 after a recurrence of cancer. She and her co-author husband, the eminent physicist Philip Morrison, wrote the reviews of children's science books that were a holiday feature in *Scientific American* for many years. Between 1998 and 2001 their column, Wonders, offered a faithful glimpse into the extraordinary diversity of their interests, which ranged from travel, gardening and photography to sculpture.

Phylis refused to honor a creative distinction between art and science. She threw herself into the advancement of both, as is evident in her writing for the short film classic *Powers of Ten,* the television series *The Ring of Truth* and her numerous books. A more thorough appreciation of Phylis's accomplishments can be found online at www.sciam.com/request.cfm?source=1002issue_morrison

When the Morrisons were seeking a name for their Wonders column, they borrowed from a credo that they might have thought up if physicist Michael Faraday hadn't said it first: "Nothing is too wonderful to be true." Phylis almost was. *—The Editors*

COMPUTERS

No Strings Attached

Computer-music instruments usually rely on MIDI (musical instrument digital interface) devices to translate mechanical vibrations into data that software then turns into sound. Such instruments are surprisingly unresponsive to subtle motions of the musician's hands, and few offer haptic (tactile) feedback—crucial for accurate and expressive playing. Charles Nichols, formerly at Stanford University and now teaching composition and music technology at the University of Montana, has developed a computerized violin bow that provides the feel of traditional bows. The vBow, as Nichols calls it, is a fiberglass rod that rides in a channel connected to a violin-shaped base. Servomotors and cables closely mimic the violin's customary haptic feedback, and high-resolution sensors capture the fine gestures of fiddling. Before he can create more expressive computer music, however, Nichols has to develop the counterpart strings and body; he hopes to have a virtual violin completed in a few years. See www.charlesnichols.com *—Steven Ashley*


CHARLES NICHOLS

MERRIER MELODIES: Virtual violin bow offers the feel of the real thing.

BRIEF POINTS

- **Cows could become pharmaceutical factories: four have been genetically engineered to incorporate the human DNA that codes for disease-fighting immunoglobulins.**

  Nature Biotechnology, September 2002

- **On your marks: the Defense Advanced Research Projects Agency has offered $1 million to inventors who can build an autonomous machine that can travel from Los Angeles to Las Vegas, in a race set for 2004.**

  www.darpa.mil/grandchallenge/index.htm

- **Cameras can now "taste" apples visually by examining the light bouncing off the fruit. The degree of absorption and reflection depends on the apple's firmness and sugar content.**

  Agricultural Research magazine, August 2002

- **Something to crow about: a New Caledonian crow named Betty demonstrated high-level toolmaking skills when she figured out that bending the end of a wire is useful in hooking food.**

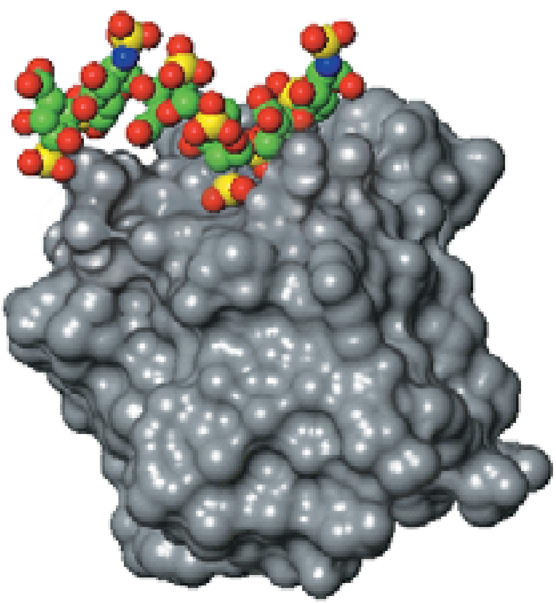
  Science, August 9, 2002; also at www.sciam.com/news_directory.cfm

Adding Sugar to Bioscience

A tennis game leads to a method for sequencing polysaccharides By MIKE MAY

A decade ago genes dominated the thinking of many molecular biologists, but not that of Ram Sasisekharan. He looked further ahead—past DNA, past proteins—to sugars. Polysaccharides cover proteins and the surface of all cells, but biologists at that time considered sugars inert, about as important to biological function as a plastic aspirin bottle is to fighting a headache. Still, Sasisekharan followed a hunch. "I instinctively felt there was something important there," he says. He was right. Cultivating these sweet opportunities, though, demanded teamwork.

In the late 1980s Sasisekharan started his doctoral work in Robert Langer's laboratory in the department of chemical engineering at the Massachusetts Institute of Technology. The reputation of Langer's lab as an invention factory would have intimidated the average student. Not Sasisekharan. He asked Langer for a project that nobody else wanted. So Langer set Sasisekharan loose on cloning heparinase, a task that had stymied previous postdoctoral and graduate students. This enzyme cuts up sugars in the heparin family, which surround the outside of cells. Such cutting can release growth factors stored in the extracellular matrix, the connective tissue–like coating on cells. Physicians use heparin to prevent blood clots after surgery and to treat clots that cause heart attacks.



FORM BEGETS FUNCTION: Sequencing of heparin (*colored molecule*) allowed researchers to determine its shape and hence how it binds to a growth factor (*gray molecule*).

MIMEON

From the start, Sasisekharan planned to follow a basic approach: find a compound's sequence, use that sequence to help unravel the compound's structure and, finally, determine how the substance works. To really figure out heparinase, Sasisekharan needed a teammate. Unexpectedly, he found one on a tennis court. While Sasisekharan volleyed with Ganesh Venkataraman, who was pursuing his Ph.D. in chemical engineering with M.I.T. professors T. Alan Hatton and Karen K. Gleason, these two doctoral students talked proteins. More to the point, Sasisekharan tried to recruit Venkataraman for help with heparinase, and he succeeded. "I had him work on making recombinant heparinase first," Sasisekharan says, "and at the same time convinced him to study sugars." Soon Sasisekharan had unraveled the sequence of amino acids that make up the protein heparinase.

All along, though, Sasisekharan wanted to go beyond enzymes to explore the biological role of sugars. His work on heparinase led naturally to an interest in the sugar that it cuts up—heparin. Sasisekharan first needed to determine the sequence of building blocks that make up this sugar. Then he hoped to use that sequence to find heparin's three-dimensional shape. He started preparing for this work years before, when he listened to his biophysicist father, Viswanathan, describe the importance of molecular shape and interactions in proteins and DNA. "I am highly influenced by my father's thinking," Sasisekharan notes. So the younger biologist knew, like his father before him, that he must unravel a molecule's shape to figure out its function.

Venkataraman began looking at heparin's structure. "As I started digging deeper," he relates, "I found there was very little information on the sequence." He and Sasisekharan needed a fast and accurate way to find the building blocks that make up heparin and other large polysaccharides. Heparin's sequence seems simple—just a repeating disaccharide, or two simple sugars linked together. Each of the simple sugars, though, can be modified in four places, which generates 16 possible versions. Consequently, the two simple sugars combined can come in 32 different "flavors." Human genes come with only four basic building blocks, and proteins use 20, so sugars looked considerably more complicated from the outset.

After completing their doctoral degrees, Sasisekharan and Venkataraman rejoined forces at M.I.T., as an associate professor and a research associate, respectively. They planned to develop a set of tools to sequence sugars. First they needed to name the different possible sugars. With 32 possible versions for a single disaccharide, even short chains skyrocket the permutations to a million or more. Venkataraman's engineering background pointed to numbers. "This was a problem that truly required a meeting of the minds," he asserts. "In hindsight, it was crucial that Ram used a biochemical approach and I used an engineering one." Sasisekharan saw the potential value of sequencing complex sugars, and Venkataraman devised a way to convert the complicated chemistry into a string of large numbers.

To do this, they utilized computers to keep track of the possibilities. A computer can compare and contrast sequences, showing where they are the same or different. In essence, the numbering system and computation let these scientists determine every possible sequence for a polysaccharide.

Getting a sequence for a sugar used to take a graduate thesis. Now it can be obtained in a day.

With a way to name the sugars, they needed to determine how to cut them up or break them into smaller pieces that would be easier to analyze. For that cutting, Sasisekharan developed a collection of other enzymes and chemicals to act as biological scissors, which accomplished the next step on the path to building a sugar-sequencing tool.

He and Venkataraman cut up heparin with scissors, determined how many pieces it made—which depends on the specific scissors and where it cuts the sugar being tested—and used a mass spectrometer to weigh each piece. Then their computers asked: Of all possible sequences, which ones would produce this number of pieces that weigh this much when exposed to this sugar-cutting tool? That eliminated some of the possibilities. The scientists then used a different pair of scissors and repeated the process.

Eventually only one possible sequence remained: the exact one for heparin. Sasisekharan and Venkataraman published that work in 1999 and patented the technique. "Getting the sequence for a polysaccharide used to take a graduate student thesis," Venkataraman says. "Now we can do that in one day" [see "Sweet Medicines," by Thomas Maeder; SCIENTIFIC AMERICAN, July].

After nearly a decade Sasisekharan and Venkataraman possessed the tool they needed to attack a crucial question: How does a sugar's structure affect its activity? Equally important, they knew that their research offered great benefits to medicine. For example, they could design a heparin with fewer side effects. To explore this commercial potential, Langer, Sasisekharan and Venkataraman founded a company called Mimeon. The company attracted a powerful chief executive officer, Alan Crane, who brought more than $2 billion in alliances to Millennium Pharmaceuticals as its vice president of corporate development.

In the near future, Mimeon will focus on heparin, already a market of more than $2 billion. Heparin was first collected from the livers of dogs. The version derived from pig intestines works well as an anticoagulant. But it must be given as an intravenous drug, and it reduces the platelets in the blood of certain patients, which can lead to dangerous bleeding. A smaller version, called low-molecular-weight heparin, has fewer side effects but also less potency to stop clotting. Nevertheless, scientists at Mimeon modified heparin's structure to make a better version—one with high anticoagulant properties and virtually no side effects. Clinical trials on multiple compounds may begin during 2003.

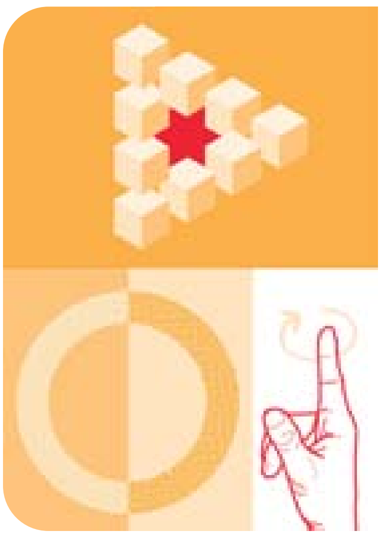
And it appears that much more lies ahead. Sasisekharan and Venkataraman have recently reported that making subtle changes to the sugars on cancer cells kills them. This could stimulate a series of specific tumor-fighting drugs. The next few years will show whether the two men can turn talks across a tennis net into a series of aces for biomedicine. SA

Mike May is based in Madison, Ind.

There's No Stopping Them

Perpetual motion is alive and well at the U.S. patent office By GRAHAM P. COLLINS

Robert Sheckley's mischievous 1954 short story "The Laxian Key" centers on a wonderful device called a Free Producer, an artifact of Meldgen Old Science that the hapless Arnold buys for next to nothing at Joe's Interstellar Junkyard. The machine "grabs energy out of the air, out of space ... anywhere. You don't have to plug it in, fuel or service [it]. It runs indefinitely." The Old Scientists of planet Meldge would've had a hard time getting a U.S. patent on their Free Producer: it sounds a lot like a perpetual-motion machine, which is verboten by U.S. Patent and Trademark Office policy, not to mention the laws of thermodynamics.



JOHN McFAUL

Or maybe not. Consider patent No. 6,362,718 for a "Motionless Electromagnetic Generator," granted in March of this year. The invention provides "a magnetic generator [in] which a need for an external power source during operation is eliminated." That is to say, once you start it up with a battery, it will keep on running and outputting power long after the battery is disconnected. Limitless power for next to nothing!

According to the patent, the motionless generator achieves its feat by being an "open, dissipative system, receiving, collecting and dissipating energy from its environment; in this case from the magnetic flux stored within the permanent magnet" that is a key element of the device. The patent cleverly points out that the invention therefore should not be considered a perpetual-motion machine, because it will stop when the magnet becomes demagnetized. (The patent neglects to add, however, that this demagnetization cannot actually be the source of the generator's mysteriously high output power.)

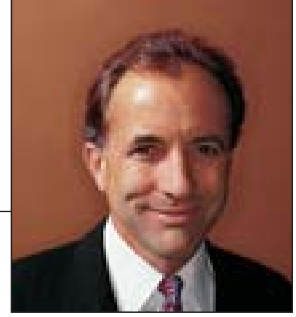
It's not the first time the impossible has been patented. The first English patent for a perpetual-motion machine was granted in 1635. We can excuse the examiner for not knowing about thermodynamics, which had not been invented yet. According to Eric Krieg, one of the founders of the Philadelphia Association of Critical Thinking, 600 patents for such devices had been granted by 1903 (see his history at www.phact.org/e/dennis4.html).

Patents such as the motionless generator illustrate that a patent is not a certification that a device will work. Examiners assess patent applications according to four criteria: novelty, usefulness, nonobviousness and enablement, the last of which means that the patent must disclose how to construct the patented device. A device that does not work as claimed should be rejected for failing usefulness and enablement, but initially the burden of disproof is on the examiner. Statements of fact in a patent application are presumed true unless a good reason for doubt is found. The device has only to be "more likely than not" to work.

The commissioner of patents may order a reexamination of a patent, and anybody can request one (for a fee of $2,800 to $8,500), but these reexaminations are rarely about bad science. Although thousands of patents are challenged in court for other reasons, no incentive exists for anyone to spend time and money debunking the science of an erroneous patent in court.

For these reasons, the endless stream of perpetual-motion machines and similar bogus devices will continue to yield occasional patents. And what of Arnold and his Meldgen Free Producer? Last seen, the machine was emitting a deluge of worthless gray powder, and no one, not even Joe the Interstellar Junkman, knows where he might get a Laxian Key—the only thing that can turn the machine off. It could almost be a metaphor. SA

Next month Staking Claims will continue its survey of perpetual motion at the U.S. patent office.



The Physicist and the Abalone Diver

The difference between the creators of two new theories of science reveals the social nature of the scientific process By MICHAEL SHERMER

Consider the following quotes, written by authors of recently self-published books purporting to revolutionize science:

"This book is the culmination of nearly twenty years of work that I have done to develop that new kind of science. I had never expected it would take anything like as long, but I have discovered vastly more than I ever thought possible, and in fact what I have done now touches almost every existing area of science, and quite a bit besides.... I have come to view [my discovery] as one of the more important single discoveries in the whole history of theoretical science."

"The development of this work has been a completely solitary effort during the past thirty years. As you will realize as you read through this book, these ideas had to be developed by an outsider. They are such a complete reversal of contemporary thinking that it would have been very difficult for any one part of this integrated theoretical system to be developed within the rigid structure of institutional science."

Both authors worked in isolation for years. Both produced remarkably self-consistent theories and make equally extravagant claims about overturning the foundations of physics in particular and science in general. Both shunned the traditional route of submitting their work to peer-reviewed scientific journals and instead chose to take their ideas straight to the public. And both texts are filled with self-produced diagrams and illustrations alleging to reveal the fundamental structures of nature.

There is one distinct difference between the two authors: one was featured in *Time*, *Newsweek* and *Wired,* and his book was reviewed in the *New York Times*. The other has been largely ignored, apart from a few exhibits at art museums. Their bios help to clarify these dissimilar receptions.

One of the authors earned his Ph.D. in physics at age 20 at the California Institute of Technology, where Richard Feynman called him "astonishing," and he was the youngest to ever win a prestigious MacArthur "genius award." He founded an institute for the study of complexity at a major university, then quit to start his own software company, where he produced a wildly successful computer program used by millions of scientists and engineers. The other author has been an abalone diver, gold miner, filmmaker, cave digger, repairman, inventor and owner-operator of a trailer park. Can you guess the names of the authors and which author penned which quote?

The first quote comes from Stephen Wolfram, the Caltech whiz and author of *A New Kind of Science,* in which the fundamental structure of the universe and everything in it is reduced to computational rules and algorithms that produce complexity in the form of cellular automata. The second comes from James Carter, the abalone diver and author of *The Other Theory of Physics,* proffering a "circlon" theory of the universe, wherein all matter is founded on hollow, ring-shaped tubes that link everything together.

JAMES CARTER'S THEORY bases the structure of the entire universe—from atoms to galaxies—on circlons, "hollow, ring-shaped mechanical particles that are held together within the nucleus by their physical shapes," as shown here in a helium atom.

Whether Wolfram is correct remains to be seen, but eventually we will find out because his ideas will be tested in the competitive marketplace of science. We may never know the veracity of Carter's ideas. Why? Because, like it or not, in science, as in most human intellectual endeavors, who is doing the saying matters as much as what is being said, at least in terms of getting an initial hearing.

Science is, in this sense, conservative and sometimes elitist. It has to be in order to survive in a surfeit of would-be revolutionaries. For every Stephen Wolfram there are 100 James Carters. There needs to be some screening process whereby truly revolutionary ideas are weeded out from ersatz ones.

Enter the skeptics. We are interested in the James Carters of the world because in the interstices between science and pseudoscience, the next great revolution may arise. Although most of these ideas will land on the junk heap, you never know until you look at them closely. SA

Michael Shermer is publisher of Skeptic *magazine (www.skeptic.com) and author of* In Darwin's Shadow.

ILLUSTRATION BY SARA CHEN; PHOTOGRAPH BY BRAD HINES

Salve for the Body and Mind

Palliative care is traditionally aimed at the terminally ill. But it should also treat sufferers of chronic disease, says Ann M. Berger of the National Institutes of Health By BOB KIRSCH

Massive textbooks, assorted journals, stuffed binders, miscellaneous folders and neatly framed family photos vie for shelf space in Ann M. Berger's office at the Warren Grant Magnuson Clinical Center of the National Institutes of Health. They would be fairly typical accoutrements of a physician's office—except for the 20 or so oversize, flouncy straw hats and the tea cart loaded with cups and saucers.

ANN M. BERGER: AIDING AND ABATING

- Senior editor of *Principles and Practice of Palliative Care and Supportive Oncology,* published this year. Plans to drop any reference to cancer in the title of future editions to reflect palliative care's broadened definition.
- Holds M.D. from the Medical College of Ohio and M.S.N. degree from the University of Pennsylvania; fellowships at the Yale University School of Medicine.
- Founder of dozens of palliative care programs.
- Pet peeve: "Doctors who don't see that palliative care goes along with curative care."

To Berger, these are medical tools, albeit ones for minds and moods. The hats and tea are brought out when doctors feel that a patient, or the family of a patient, needs a lift. Such a party is not given casually. It is a specific intervention, one that Berger instituted when she founded the palliative care service here.

Palliative care is the branch of medicine that addresses symptoms—both the physical, such as nausea and insomnia, and the psychological, such as worry and depression. As the 43-year-old Berger defines it, "palliative care is a combination of active and compassionate therapies that is primarily focused on the physical, psychological, social and spiritual suffering of the patient, family and caregiver." It begins, she says, at diagnosis and should be administered throughout the course of the disease.

Berger had an early, firsthand understanding of the need for palliative care. During the summer after her 14th birthday, she watched her grandfather, to whom she was close, die of bladder cancer. She witnessed the confused, ineffectual way her family absorbed the impact of the illness and learned how the effects of disease can ricochet among family members and back to the patient. She carried those lessons throughout her training as a nurse and on to medical school.

Berger arrived on the scene at just the right moment. By the 1980s the hospice care movement was gathering momentum, making great strides in addressing the multitudinous needs of dying patients. At the same time, some physicians specializing in the care of cancer cases were focusing on improving quality of life through better management of physical and psychological symptoms. Researchers were documenting that, for example, moderate to severe pain affects one in three patients being treated for cancer and between 60 and 90 percent of those with advanced disease. And yet pain management was being underutilized world-

PHOTOGRAPHS BY TOM WOLFF

wide, in part because of the hesitance of many physicians to prescribe morphine and related painkillers.

Berger realized that she wanted to work in this field when she was in the middle of a post–medical school fellowship at the Yale University School of Medicine. In 1992 she volunteered to found the palliative care service at Yale, and within three months she was invited to join the faculty. Since then she has founded palliative care services at 40 long-term-care facilities.

Berger and other advocates want palliative care to be more comprehensive than its traditional focus on terminally ill cancer patients. It should tackle symptoms that affect well-being—such as hair loss, drowsiness, anxiety, irritability and side effects of medication, including sexual dysfunction. Berger's perspective is that any patient with a chronic illness—especially those that are life-threatening, such as diabetes, emphysema, multiple sclerosis and cardiovascular disease—may require palliative care, which should be pursued in tandem with curative care. People with advanced heart disease can experience such severe shortness of breath that they can't even walk to their next-door neighbor's house; they may become as severely depressed and anxious as terminally ill patients.

At the clinical center, candidates for palliative care are first assessed to see how the sum of all the symptoms and suffering impinges on their quality of life. Once the assessment is complete, Berger and her team identify the health professionals who can best meet each patient's needs. They run the gamut from counselors, massage and music therapists, dietitians, social workers, clergy—even pet therapists and acupuncturists—to surgeons and neurologists who can conduct procedures that alleviate pain.

The array of problems that patients experience means that Berger may become involved in an astounding set of activities on any given day. Sometimes she prescribes a high-tech solution to a patient's complaint. Sometimes she is the friendly neighbor who listens and chats. Sometimes she is an educator, teaching physicians and nurses about palliative care. For instance, she recalls that when a patient had problems breathing and she recommended aerosolized morphine, the patient's pulmonologist—skilled in diagnosis and treatment but not up on symptom management—considered it a marginal therapy. She pointed out that a chapter in a textbook she had co-edited offered ample evidence supporting its use. Since that time, he has regularly prescribed aerosolized morphine.

But Berger also oversees something she considers equally important: palliative care for the doctors and other health professionals at the clinical center. "Being exposed to suffering on a daily basis can be emotionally taxing," she says. "What happens over time when many of your patients are seriously ill is you deal with crisis, crisis, crisis, crisis, and loss, loss, loss, loss." Few health care workers have an opportunity to contend with their feelings about the patients for whom treatment was not successful. "And providers who do not deal with their own feelings of loss are not going to be able to heal other people," she notes.

For this kind of integral care to be administered in hospitals is, in Berger's experience, impossible unless hospital leaders provide firm institutional support. Programs do not usually survive for very long without it, she observes.

Palliative medicine has not stirred much visible opposition among physicians, but it is challenging basic assumptions. According to Russell K. Portenoy, who chairs the palliative care department at Beth Israel Medical Center in New York City, patients typically struggle with questions about why they have become ill, and such questioning can threaten their essential belief structure. Yet those concerns, he says, "tend to be minimized by physicians whose training is not in the area of assessing and managing spiritual distress."

MEDICAL HUDDLE: Berger often meets with other health care professionals, because palliative care requires the skills of several specialists.

Berger would like every hospital to have at least one physician or nurse practitioner who is trained in and can focus on palliative care. She founded and leads an NIH working group that is trying to set up palliative medicine programs from coast to coast. But how can cash-strapped hospitals afford it? Berger says that most could start with modest ones. She points out that palliative care may reduce costs overall—by, for instance, helping patients to leave the hospital sooner or undergo fewer rehospitalizations. Clinical studies are under way to assess the economic feasibility of such programs.

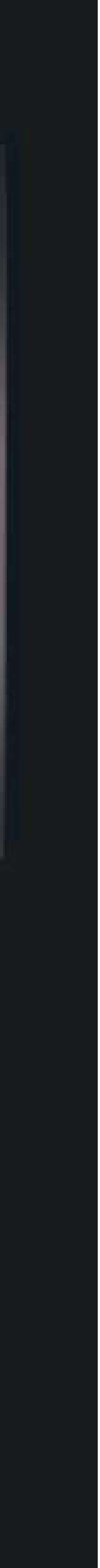
If Berger had any doubts about the course of her career, they were laid to rest three days after her NIH palliative care appointment in August 2000. Diagnosed with breast cancer, she endured the disease's course, although she has remained well since completing surgery. The experience reinforced her interest in understanding life-threatening illness from the other side of the stethoscope: "Once I became a patient, I saw that I wanted the cure but I also wanted myself—as a person—to be taken care of. Then it really jelled." SA

Bob Kirsch is a medical writer based in Ossining, N.Y.

Controlling Robots with the *Mind*

By Miguel A. L. Nicolelis and John K. Chapin

People with nerve or limb injuries may one day be able to command wheelchairs, prosthetics and even paralyzed arms and legs by "thinking them through" the motions



Belle, our tiny owl monkey, was seated in her special chair

inside a soundproof chamber at our Duke University laboratory. Her right hand grasped a joystick as she watched a horizontal series of lights on a display panel. She knew that if a light suddenly shone and she moved the joystick left or right to correspond to its position, a dispenser would send a drop of fruit juice into her mouth. She loved to play this game. And she was good at it.

Belle wore a cap glued to her head. Under it were four plastic connectors. The connectors fed arrays of microwires—each wire finer than the finest sewing thread—into different regions of Belle's motor cortex, the brain tissue that plans movements and sends instructions for enacting the plans to nerve cells in the spinal cord. Each of the 100 microwires lay beside a single motor neuron. When a neuron produced an electrical discharge—an "action potential"—the adjacent microwire would capture the current and send it up through a small wiring bundle that ran from Belle's cap to a box of electronics on a table next to the booth. The box, in turn, was linked to two computers, one next door and the other half a country away.

In a crowded room across the hall, members of our research team were getting anxious. After months of hard work, we were about to test the idea that we could reliably translate the raw electrical activity in a living being's brain—Belle's mere thoughts—into signals that could direct the actions of a robot. Unknown to Belle on this spring afternoon in 2000, we had assembled a multijointed robot arm in this room, away from her view, that she would control for the first time. As soon as Belle's brain sensed a lit spot on the panel, electronics in the box running two real-time mathematical models would rapidly analyze the tiny action potentials produced by her brain cells. Our lab computer would convert the electrical patterns into instructions that would direct the robot arm. Six hundred miles north, in Cambridge, Mass., a different computer would produce the same actions in another robot arm, built by Mandayam A. Srinivasan, head of the Laboratory for Human and Machine Haptics (the Touch Lab) at the Massachusetts Institute of Technology. At least, that was the plan.

If we had done everything correctly, the two robot arms would behave as Belle's arm did, at exactly the same time. We would have to translate her neuronal activity into robot commands in just 300 milliseconds—the natural delay between the time Belle's motor cortex planned how she should move her limb and the moment it sent the instructions to her muscles. If the brain of a living creature could accurately control two dissimilar robot arms—despite the signal noise and transmission delays inherent in our lab network and the error-prone Internet—perhaps it could someday control a mechanical device or actual limbs in ways that would be truly helpful to people.

Finally the moment came. We randomly switched on lights in front of Belle, and she immediately moved her joystick back and forth to correspond to them. Our robot arm moved similarly to Belle's real arm. So did Srinivasan's. Belle and the robots moved in synchrony, like dancers choreographed by the electrical impulses sparking in Belle's mind. Amid the loud celebration that erupted in Durham, N.C., and Cambridge, we could not help thinking that this was only the beginning of a promising journey.

In the two years since that day, our labs and several oth-

OWL MONKEY named Belle climbs on a robot arm she was able to control from a distant room purely by imagining her own arm moving through three-dimensional space.

JIM WALLACE *Duke University Photography* ers have advanced neuroscience, computer science, microelectronics and robotics to create ways for rats, monkeys and eventually humans to control mechanical and electronic machines purely by "thinking through," or imagining, the motions. Our immediate goal is to help a person who has been paralyzed by a neurological disorder or spinal cord injury, but whose motor cortex is spared, to operate a wheelchair or a robotic limb. Someday the research could also help such a patient regain control over a natural arm or leg, with the aid of wireless communication between implants in the brain and the limb. And it could lead to devices that restore or augment other motor, sensory or cognitive functions.

The big question is, of course, whether we can make a practical, reliable system. Doctors have no means by which to repair spinal cord breaks or damaged brains. In the distant future, neuroscientists may be able to regenerate injured neurons or program stem cells (those capable of differentiating into various cell types) to take their place. But in the near future, brain-machine interfaces (BMIs), or neuroprostheses, are a more viable option for restoring motor function. Success this summer with macaque monkeys that completed different tasks than those we asked of Belle has gotten us even closer to this goal.

From Theory to Practice

RECENT ADVANCES in brain-machine interfaces are grounded in part on discoveries made about 20 years ago. In the early 1980s Apostolos P. Georgopoulos of Johns Hopkins University recorded the electrical activity of single motor-cortex neurons in macaque monkeys. He found that the nerve cells typically reacted most strongly when a monkey moved its arm in a certain direction. Yet when the arm moved at an angle away from a cell's preferred direction, the neuron's activity didn't cease; it diminished in proportion to the cosine of that angle. The finding showed that motor neurons were broadly tuned to a range of motion and that the brain most likely relied on the collective activity of dispersed populations of single neurons to generate a motor command.

Overview/*Brain Interfaces*

- Rats and monkeys whose brains have been wired to a computer have successfully controlled levers and robot arms by imagining their own limb either pressing a bar or manipulating a joystick.
- These feats have been made possible by advances in microwires that can be implanted in the motor cortex and by the development of algorithms that translate the electrical activity of brain neurons into commands able to control mechanical devices.
- Human trials of sophisticated brain-machine interfaces are far off, but the technology could eventually help people who have lost an arm to control a robotic replacement with their mind or help patients with a spinal cord injury regain control of a paralyzed limb.

There were caveats, however. Georgopoulos had recorded the activity of single neurons one at a time and from only one motor area. This approach left unproved the underlying hypothesis that some kind of coding scheme emerges from the simultaneous activity of many neurons distributed across multiple cortical areas. Scientists knew that the frontal and parietal lobes—in the forward and rear parts of the brain, respectively—interacted to plan and generate motor commands. But technological bottlenecks prevented neurophysiologists from mak-

Belle's 600-Mile Reach

ON THE DAY BELLE first moved a multijointed robot arm with her thoughts, she wore a cap glued to her head. Beneath the cap, each of four plastic connectors fed an array of fine microwires into her cortex (*a*). As Belle saw lights shine suddenly and decided to move a joystick left or right to correspond to them, the microwires detected electrical signals produced by activated neurons in her cortex and relayed the signals to a "Harvey box" of electronics.

ing widespread recordings at once. Furthermore, most scientists believed that by cataloguing the properties of neurons one at a time, they could build a comprehensive map of how the brain works—as if charting the properties of individual trees could unveil the ecological structure of an entire forest!

Fortunately, not everyone agreed. When the two of us met 14 years ago at Hahnemann University, we discussed the challenge of simultaneously recording many single neurons. By 1993 technological breakthroughs we had made allowed us to record 48 neurons spread across five structures that form a rat's sensorimotor system—the brain regions that perceive and use sensory information to direct movements.

Crucial to our success back then—and since—were new electrode arrays containing Teflon-coated stainless-steel microwires that could be implanted in an animal's brain. Neurophysiologists had used standard electrodes that resemble rigid needles to record single neurons. These classic electrodes worked well but only for a few hours, because cellular compounds col-

The box collected, filtered and amplified the signals and relayed them to a server computer in a room next door. The signals received by the box can be displayed as a raster plot (*b*); each row represents the activity of a single neuron recorded over time, and each color bar indicates that the neuron was firing at a given moment.

The computer, in turn, predicted the trajectory that Belle's arm would take (*c*) and converted that information into commands for producing the same motion in a robot arm. Then the computer sent commands to a computer that operated a robot arm in a room across the hall. At the same time, it sent commands from our laboratory in Durham, N.C., to another robot in a laboratory hundreds of miles away. In response, both robot arms moved in synchrony with Belle's own limb. *—M.A.L.N. and J.K.C.*

BRYAN CHRISTIE DESIGN

lected around the electrodes' tips and eventually insulated them from the current. Furthermore, as the subject's brain moved slightly during normal activity, the stiff pins damaged neurons. The microwires we devised in our lab (later produced by NBLabs in Denison, Tex.) had blunter tips, about 50 microns in diameter, and were much more flexible. Cellular substances did not seal off the ends, and the flexibility greatly reduced neuron damage. These properties enabled us to produce recordings for months on end, and having tools for reliable recording allowed us to begin developing systems for translating brain signals into commands that could control a mechanical device.

With electrical engineer Harvey Wiggins, now president of Plexon in Dallas, and with Donald J. Woodward and Samuel A. Deadwyler of Wake Forest University School of Medicine, we devised a small "Harvey box" of custom electronics, like the one next to Belle's booth. It was the first hardware that could properly sample, filter and amplify neural signals from many electrodes. Special software allowed us to discriminate electrical activity from up to four single neurons per microwire by identifying unique features of each cell's electrical discharge.

A Rat's Brain Controls a Lever

IN OUR NEXT EXPERIMENTS at Hahnemann in the mid-1990s, we taught a rat in a cage to control a lever with its mind. First we trained it to press a bar with its forelimb. The bar was electronically connected to a lever outside the cage. When the rat pressed the bar, the outside lever tipped down to a chute and delivered a drop of water it could drink.

We fitted the rat's head with a small version of the brain-machine interface Belle would later use. Every time the rat commanded its forelimb to press the bar, we simultaneously recorded the action potentials produced by 46 neurons. We had programmed resistors in a so-called integrator, which weighted and processed data from the neurons to generate a single analog output that predicted very well the trajectory of the rat's forelimb. We linked this integrator to the robot lever's controller so that it could command the lever.

Once the rat had gotten used to pressing the bar for water, we disconnected the bar from the lever. The rat pressed the bar, but the lever remained still. Frustrated, it began to press the bar repeatedly, to no avail. But one time, the lever tipped and delivered the water. The rat didn't know it, but its 46 neurons had expressed the same firing pattern they had in earlier trials when the bar still worked. That pattern prompted the integrator to put the lever in motion.

After several hours the rat realized it no longer needed to press the bar. If it just looked at the bar and imagined its forelimb pressing it, its neurons could still express the firing pattern that our brain-machine interface would interpret as motor commands to move the lever. Over time, four of six rats succeeded in this task. They learned that they had to "think through" the motion of pressing the bar. This is not as mystical at it might sound; right now you can imagine reaching out to grasp an object near you—without doing so. In similar fashion, a person with an injured or severed limb might learn to control a robot arm joined to a shoulder.

A Monkey's Brain Controls a Robot Arm

WE WERE THRILLED with our rats' success. It inspired us to move forward, to try to reproduce in a robotic limb the three-dimensional arm movements made by monkeys—animals with brains far more similar to those of humans. As a first step, we had to devise technology for predicting how the monkeys intended to move their natural arms.

At this time, one of us (Nicolelis) moved to Duke and established a neurophysiology laboratory there. Together we built an interface to simultaneously monitor close to 100 neurons, distributed across the frontal and parietal lobes. We proceeded to try it with several owl monkeys. We chose owl monkeys because their motor cortical areas are located on the surface of their smooth brain, a configuration that minimizes the surgical difficulty of implanting microwire arrays. The microwire arrays allowed us to record the action potentials in each creature's brain for several months.

In our first experiments, we required owl monkeys, including Belle, to move a joystick left or right after seeing a light appear on the left or right side of a video screen. We later sat them in a chair facing an opaque barrier. When we lifted the barrier they saw a piece of fruit on a tray. The monkeys had to reach out and grab the fruit, bring it to their mouth and place their hand back down. We measured the position of each monkey's wrist by attaching fiber-optic sensors to it, which defined the wrist's trajectory.

Further analysis revealed that a simple linear summation of the electrical activity of cortical motor neurons predicted very well the position of an animal's hand a few hundred milliseconds ahead of time. This discovery was made by Johan Wessberg of Duke, now at the Gothenburg University in Sweden. The main trick was for the computer to continuously combine neuronal activity produced as far back in time as one second to best predict movements in real time.

As our scientific work proceeded, we acquired a more advanced Harvey box from Plexon. Using it and some custom, real-time algorithms, our computer sampled and integrated the action potentials every 50 to 100 milliseconds. Software translated the output into instructions that could direct the actions of a robot arm in three-dimensional space. Only then did we try to

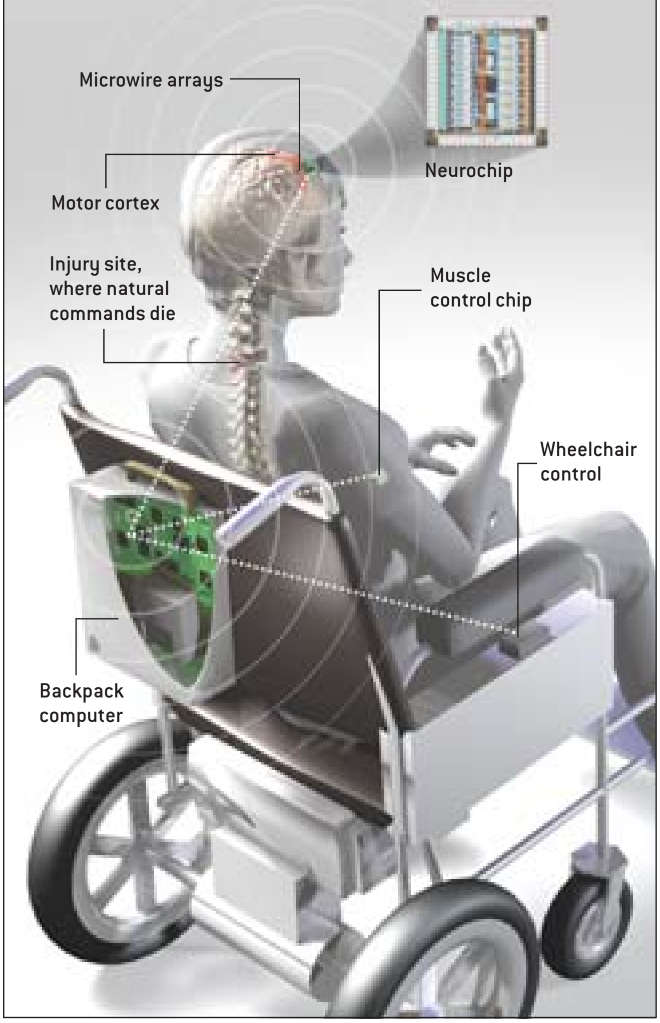
THE AUTHORS

MIGUEL A. L. NICOLELIS and *JOHN K. CHAPIN* have collaborated for more than a decade. Nicolelis, a native of Brazil, received his M.D. and Ph.D. in neurophysiology from the University of São Paulo. After postdoctoral work at Hahnemann University, he joined Duke University, where he now co-directs the Center for Neuroengineering and is professor of neurobiology, biomedical engineering, and psychological and brain sciences. Chapin received his Ph.D. in neurophysiology from the University of Rochester and has held faculty positions at the University of Texas and the MCP Hahnemann University School of Medicine (now Drexel University College of Medicine). He is currently professor of physiology and pharmacology at the State University of New York Downstate Medical Center.

A Vision of the Future

A BRAIN-MACHINE INTERFACE might someday help a patient whose limbs have been paralyzed by a spine injury. Tiny arrays of microwires implanted in multiple motor cortex areas of the brain would be wired to a neurochip in the skull. As the person imagined her paralyzed arm moving in a particular way, such as reaching out for food on a table, the chip would convert the thoughts into a train of radio-frequency signals and send them wirelessly to a small battery-operated "backpack" computer hanging from the chair.

The computer would convert the signals into motor commands and dispatch them, again wirelessly, to a different chip implanted in the person's arm. This second chip would stimulate nerves needed to move the arm muscles in the desired fashion. Alternatively, the backpack computer could control the wheelchair's motor and steering directly, as the person envisioned where she wanted the chair to roll. Or the computer could send signals to a robotic arm if a natural arm were missing or to a robot arm mounted on a chair. Patrick D. Wolf of Duke University has built a prototype neurochip and backpack, as envisioned here. —*M.A.L.N. and J.K.C.*



BRYAN CHRISTIE DESIGN

use a BMI to control a robotic device. As we watched our multijointed robot arm accurately mimic Belle's arm movements on that inspiring afternoon in 2000, it was difficult not to ponder the implausibility of it all. Only 50 to 100 neurons randomly sampled from tens of millions were doing the needed work.

Later mathematical analyses revealed that the accuracy of the robot movements was roughly proportional to the number of neurons recorded, but this linear relation began to taper off as the number increased. By sampling 100 neurons we could create robot hand trajectories that were about 70 percent similar to those the monkeys produced. Further analysis estimated that to achieve 95 percent accuracy in the prediction of one-dimensional hand movements, as few as 500 to 700 neurons would suffice, depending on which brain regions we sampled. We are now calculating the number of neurons that would be needed for highly accurate three-dimensional movements. We suspect the total will again be in the hundreds, not thousands.

These results suggest that within each cortical area, the "message" defining a given hand movement is widely disseminated. This decentralization is extremely beneficial to the animal: in case of injury, the animal can fall back on a huge reservoir of redundancy. For us researchers, it means that a BMI neuroprosthesis for severely paralyzed patients may require sampling smaller populations of neurons than was once anticipated.

We continued working with Belle and our other monkeys after Belle's successful experiment. We found that as the animals perfected their tasks, the properties of their neurons changed—over several days or even within a daily two-hour recording session. The contribution of individual neurons varied over time. To cope with this "motor learning," we added a simple routine that enabled our model to reassess periodically the contribution of each neuron. Brain cells that ceased to influence the predictions significantly were dropped from the model, and those that became better predictors were added. In essence, we designed a way to extract from the brain a neural output for hand trajectory. This coding, plus our ability to measure neurons reliably over time, allowed our BMI to represent Belle's intended movements accurately for several months. We could have continued, but we had the data we needed.

It is important to note that the gradual changing of neuronal electrical activity helps to give the brain its plasticity. The number of action potentials a neuron generates before a given movement changes as the animal undergoes more experiences. Yet the dynamic revision of neuronal properties does not represent an impediment for practical BMIs. The beauty of a distributed neural output is that it does not rely on a small group of neurons. If a BMI can maintain viable recordings from hundreds to thousands of single neurons for months to years and utilize models that can learn, it can handle evolving neurons, neuronal death and even degradation in electrode-recording capabilities.

Exploiting Sensory Feedback

BELLE PROVED THAT A BMI can work for a primate brain. But could we adapt the interface to more complex brains? In May 2001 we began studies with three macaque monkeys at

Duke. Their brains contain deep furrows and convolutions that resemble those of the human brain.

We employed the same BMI used for Belle, with one fundamental addition: now the monkeys could exploit visual feedback to judge for themselves how well the BMI could mimic their hand movements. We let the macaques move a joystick in random directions, driving a cursor across a computer screen. Suddenly a round target would appear somewhere on the screen. To receive a sip of fruit juice, the monkey had to position the cursor quickly inside the target—within 0.5 second—by rapidly manipulating the joystick.

The first macaque to master this task was Aurora, an elegant female who clearly enjoyed showing off that she could hit the target more than 90 percent of the time. For a year, our postdoctoral fellows Roy Crist and José Carmena recorded the activity of up to 92 neurons in five frontal and parietal areas of Aurora's cortex.

Once Aurora commanded the game, we started playing a trick on her. In about 30 percent of the trials we disabled the connection between the joystick and the cursor. To move the cursor quickly within the target, Aurora had to rely solely on her brain activity, processed by our BMI. After being puzzled, Aurora gradually altered her strategy. Although she continued to make hand movements, after a few days she learned she could control the cursor 100 percent of the time with her brain alone. In a few trials each day during the ensuing weeks Aurora didn't even bother to move her hand; she moved the cursor by just thinking about the trajectory it should take.

That was not all. Because Aurora could see her performance on the screen, the BMI made better and better predictions even though it was recording the same neurons. Although much more analysis is required to understand this result, one explanation is that the visual feedback helped Aurora to maximize the BMI's reaction to both brain and machine learning. If this proves true, visual or other sensory feedback could allow people to improve the performance of their own BMIs.

We observed another encouraging result. At this writing, it has been a year since we implanted the microwires in Aurora's brain, and we continue to record 60 to 70 neurons daily. This extended success indicates that even in a primate with a convoluted brain, our microwire arrays can provide long-term, high-quality, multichannel signals. Although this sample is down from the original 92 neurons, Aurora's performance with the BMI remains at the highest levels she has achieved.

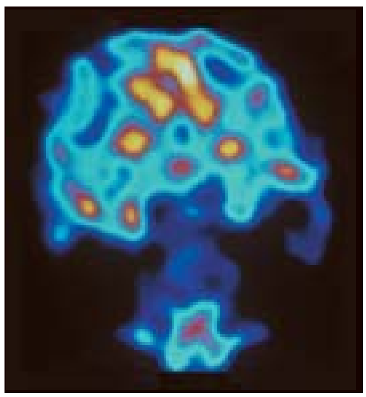
We will make Aurora's tasks more challenging. In May we began modifying the BMI to give her tactile feedback for new experiments that are now beginning. The BMI will control a nearby robot arm fitted with a gripper that simulates a grasping hand. Force sensors will indicate when the gripper encounters an object and how much force is required to hold it. Tactile feedback—is the object heavy or light, slick or sticky?—will be delivered to a patch on Aurora's skin embedded with small vibrators. Variations in the vibration frequencies should help Aurora figure out how much force the robot arm should apply to, say, pick up a piece of fruit, and to hold it as the robot brings it back to her. This experiment might give us the most concrete evidence yet that a person suffering from severe paralysis could regain basic arm movements through an implant in the brain that communicated over wires, or wirelessly, with signal generators embedded in a limb.

If visual and tactile sensations mimic the information that usually flows between Aurora's own arm and brain, long-term interaction with a BMI could possibly stimulate her brain to incorporate the robot into its representations of her body—schema known to exist in most brain regions. In other words, Aurora's brain might represent this artificial device as another part of her body. Neuronal tissue in her brain might even dedicate itself to operating the robot arm and interpreting its feedback.

Stopping Seizures

RECENT EXPERIMENTS SUGGEST that brain-machine interfaces could one day help prevent brain seizures in people who suffer from severe chronic epilepsy, which causes dozens of seizures a day. The condition ruins a patient's quality of life and can lead to permanent brain damage. To make matters worse, patients usually become unresponsive to traditional drug therapy.

A BMI for seizure control would function somewhat like a heart pacemaker. It would continuously monitor the brain's electrical activity for patterns that indicate an imminent attack. If the BMI sensed such a pattern, it would deliver an electrical stimulus to the brain or a peripheral nerve that would quench the rising storm or trigger the release of antiepileptic medication.



PET SCAN taken during an epileptic seizure highlights regions of excessive brain activity in yellow.

At Duke we demonstrated the feasibility of this concept in collaboration with Erika E. Fanselow, now at Brown University, and Ashlan P. Reid, now at the University of Pennsylvania. We implanted a BMI with arrays of microwires in rats given PTZ, a drug that induces repetitive mild epilepsy. When a seizure starts, cortical neurons begin firing together in highly synchronized bursts. When the "brain pacemaker" detected this pattern, it triggered the electrical stimulation of the large trigeminal cranial nerve. The brief stimulus disrupted the epileptic activity quickly and efficiently, without damaging the nerve, and reduced the occurrence and duration of seizures. *—M.A.L.N. and J.K.C.*

TIM BEDDON *SPL/Photo Researchers, Inc.*

To test whether this hypothesis has merit, we plan to conduct experiments like those done with Aurora, except that an animal's arm will be temporarily anesthetized, thereby removing any natural feedback information. We predict that after a transition period, the primate will be able to interact with the BMI just fine. If the animal's brain does meld the robot arm into its body representations, it is reasonable to expect that a paraplegic's brain would do the same, rededicating neurons that once served a natural limb to the operation of an artificial one.

Each advance shows how plastic the brain is. Yet there will always be limits. It is unlikely, for example, that a stroke victim could gain full control over a robot limb. Stroke damage is usually widespread and involves so much of the brain's white matter—the fibers that allow brain regions to communicate—that the destruction overwhelms the brain's plastic capabilities. This is why stroke victims who lose control of uninjured limbs rarely regain it.

Reality Check

GOOD NEWS NOTWITHSTANDING, we researchers must be very cautious about offering false hope to people with serious disabilities. We must still overcome many hurdles before BMIs can be considered safe, reliable and efficient therapeutic options. We have to demonstrate in clinical trials that a proposed BMI will offer much greater well-being while posing no risk of added neurological damage.

Surgical implantation of electrode arrays will always be of medical concern, for instance. Investigators need to evaluate whether highly dense microwire arrays can provide viable recordings without causing tissue damage or infection in humans. Progress toward dense arrays is already under way. Duke electronics technician Gary Lehew has designed ways to increase significantly the number of microwires mounted in an array that is light and easy to implant. We can now implant multiple arrays, each of which has up to 160 microwires and measures five by eight millimeters, smaller than a pinky fingernail. We recently implanted 704 microwires across eight cortical areas in a macaque and recorded 318 neurons simultaneously.

In addition, considerable miniaturization of electronics and batteries must occur. We have begun collaborating with José Carlos Príncipe of the University of Florida to craft implantable microelectronics that will embed in hardware the neuronal pattern recognition we now do with software, thereby eventually freeing the BMI from a computer. These microchips will thus have to send wireless control data to robotic actuators. Working with Patrick D. Wolf's lab at Duke, we have built the first wireless "neurochip" and beta-tested it with Aurora. Seeing streams of neural activity flash on a laptop many meters away from Aurora—broadcast via the first wireless connection between a primate's brain and a computer—was a delight.

More and more scientists are embracing the vision that BMIs can help people in need. In the past year, several traditional neurological laboratories have begun to pursue neuroprosthetic devices. Preliminary results from Arizona State University, Brown University and the California Institute of Technology have recently appeared. Some of the studies provide independent confirmation of the rat and monkey studies we have done. Researchers at Arizona State basically reproduced our 3-D approach in owl monkeys and showed that it can work in rhesus monkeys too. Scientists at Brown enabled a rhesus macaque monkey to move a cursor around a computer screen. Both groups recorded 10 to 20 neurons or so per animal. Their success further demonstrates that this new field is progressing nicely.

The most useful BMIs will exploit hundreds to a few thousand single neurons distributed over multiple motor regions in the frontal and parietal lobes. Those that record only a small number of neurons (say, 30 or fewer) from a single cortical area would never provide clinical help, because they would lack the excess capacity required to adapt to neuronal loss or changes in neuronal responsiveness. The other extreme—recording millions of neurons using large electrodes—would most likely not work either, because it might be too invasive.

Noninvasive methods, though promising for some therapies, will probably be of limited use for controlling prostheses with thoughts. Scalp recording, called electroencephalography (EEG), is a noninvasive technique that can drive a different kind of brain-machine interface, however. Niels Birbaumer of the University of Tübingen in Germany has successfully used EEG recordings and a computer interface to help patients paralyzed by severe neurological disorders learn how to modulate their EEG activity to select letters on a computer screen, so they can write messages. The process is time-consuming but offers the only way for these people to communicate with the world. Yet EEG signals cannot be used directly for limb prostheses, because they depict the average electrical activity of broad populations of neurons; it is difficult to extract from them the fine variations needed to encode precise arm and hand movements.

Despite the remaining hurdles, we have plenty of reasons to be optimistic. Although it may be a decade before we witness the operation of the first human neuroprosthesis, all the amazing possibilities crossed our minds that afternoon in Durham as we watched the activity of Belle's neurons flashing on a computer monitor. We will always remember our sense of awe as we eavesdropped on the processes by which the primate brain generates a thought. Belle's thought to receive her juice was a simple one, but a thought it was, and it commanded the outside world to achieve her very real goal. SA

MORE TO EXPLORE

Real-Time Control of a Robot Arm Using Simultaneously Recorded Neurons in the Motor Cortex. J. K. Chapin, K. A. Moxon, R. S. Markowitz and M.A.L. Nicolelis in *Nature Neurosciences,* Vol. 2, pages 664–670; July 1999.

Real-Time Prediction of Hand Trajectory by Ensembles of Cortical Neurons in Primates. J. Wessberg, C. R. Stambaugh, J. D. Kralik, P. D. Beck, J. K. Chapin, J. Kim, S. J. Biggs, M. A. Srinivasan and M.A.L. Nicolelis in *Nature,* Vol. 408, pages 361–365; November 16, 2000.

Actions from Thoughts. M.A.L. Nicolelis in *Nature,* Vol. 409, pages 403–407; January 18, 2001.

Advances in Neural Population Coding. Edited by M.A.L. Nicolelis. Progress in Brain Research, Vol. 130. Elsevier, 2001.

Neural Prostheses for Restoration of Sensory and Motor Function. Edited by J. K. Chapin and K. A. Moxon. CRC Press, 2001.

The Emptiest Places

Space comes in degrees of emptiness, but even in the wasteland between galaxies it is not a complete void

By Evan Scannapieco, Patrick Petitjean and Tom Broadhurst

LIKE DEWDROPS ON A SPIDER'S WEB, galaxies collect on the filaments of material that stretch across the vast reaches of intergalactic space. Much of the history of the universe may have been determined by the give-and-take between galaxies and intergalactic gas. This artist's conception is based on computer simulations of the gas.

A trip from the earth to beyond the Milky Way is a journey to the emptiest places imaginable. Leaving the cozy confines of our solar system, we find ourselves in the interstellar regions of the galaxy. Here light from the nearest stars takes years to reach us, and the density of gas averages about one atom per cubic centimeter.

But we are headed to a place that is much more desolate. As we continue outward into the farthest reaches of the galactic disk, the stars are separated by dozens, then hundreds, of light-years, and the interstellar gas thins out 100-fold. Finally, passing into the vast inky blackness beyond the galaxy, we come upon a gas so tenuous that it scarcely seems worthy of the name, with an average density of only 10^{-5} atom per cubic centimeter.

In terms of density, the voyage from interplanetary to intergalactic space is more drastic than going from water into air. You might be forgiven for expecting that the end point of the trip, the deepest recesses of space, would give new meaning to the word "boring." Even astronomers used to think little of intergalactic space. Why bother with a thin gruel of atoms when the universe abounds in richly textured planets, luxurious galaxies and ravenous black holes?

But that attitude has been shifting. Far from an austere backwater, the intergalactic medium (IGM) is turning out to be the central staging area for cosmic evolution. The IGM predates galaxies. At early times, all matter took the form of a hot and all-pervading gas. Through the expansion of the universe, the gas cooled and condensed into the myriad galaxies found today. Anything left behind became ever more diffuse.

This much has been clear for several decades. Astronomers long assumed, however, that the details of the intergalactic gas were unimportant and that gravity alone called the shots in galaxy formation. According to the prevailing view, once the IGM had cooled from its hot, ionized state to a colder mixture of neutral hydrogen and helium, it offered no effective resistance to gravity. Places that had an unusually high density pulled in material at the expense of sparser regions—a process that continues unimpeded to this day. In this picture, the densities, positions and sizes of galaxies and larger structures depend only on the random primordial distribution of mass. Even if the medium had some internal complexity, a possibility that struck most researchers as unlikely, it exerted no effect on the truly interesting parts of the cosmos.

Yet the more astronomers began to uncover the properties of the gas, the more their observations came into conflict with this simple theory. They discovered that the IGM has an intricate history, including several important transitions intimately related to the formation of structure. And they found that this most delicate of materials is drawn out into a vast network of gaseous sheets and filaments, draped between the galaxies like a spider web.

These investigations began to gather momentum, and the past two years have seen an explosion of research activity. It is not easy, though, to study something that can barely be seen. Like detectives, astronomers are gathering indirect clues and carefully piecing them together to reveal the story of the gas between the galaxies.

Overview/*Intergalactic Medium*

- Near-Earth space, where astronauts roam, is nearly a vacuum by terrestrial standards, but the space between galaxies is even emptier, with a millionth the density. Astronomers once doubted that anything interesting could happen in such an incomprehensibly tenuous gas.
- Yet a steady accumulation of observations shows that this gas, known as the intergalactic medium, has undergone at least three dramatic transitions, with profound effects on the formation of galaxies and other cosmic structures.

Seeing the Forest for the Lines

THESE CLUES COME FROM four types of observational evidence: the cosmic microwave background radiation, quasar spectra, x-rays from galaxy clusters and magnetic field measurements. The microwave background provides a snapshot of the IGM at the moment it changed from ionized to neutral, approximately 300,000 years after the big bang, when the gas temperature had fallen to a few thousand kelvins. Patterns in this radiation are the starting point for all models of the IGM.

The second type of evidence involves quasars. Thought to be powered by young supermassive black holes, these extremely bright objects act as lighthouses that illuminate narrow stretches of intergalactic space. Material between us and a quasar absorbs light of specific wavelengths, leaving a telltale imprint on the quasar spectrum. Interpreting such spectra requires a degree of care. They contain lines at wavelengths that do not appear to correspond to any known substance. This discrepancy is thought to be a product of the expansion of the universe, which, by stretching the light waves, causes the spectral lines to move from their usual positions to longer wavelengths—a pro-

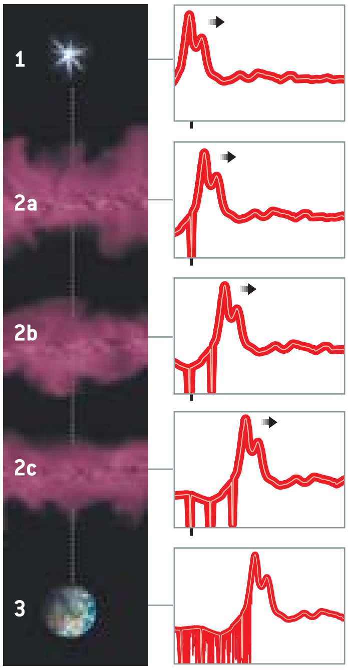
MARK A. GARLICK (*preceding pages*)

THE FOREST PRIMEVAL

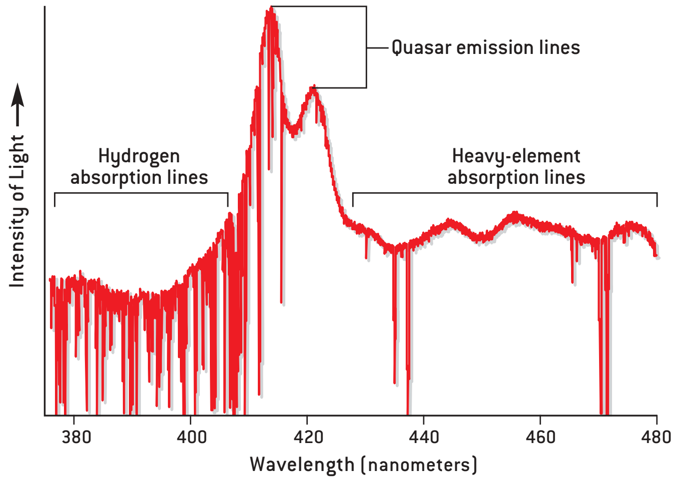
WISPY THOUGH IT MAY BE, intergalactic gas betrays its presence by distorting the light of other objects—particularly quasars, the brightest objects in the known universe. Acting like rose-tinted glasses, gas clouds block light at certain wavelengths but let the rest through. The process shows up as a series of absorption lines (*sharp dips*) in quasar spectra. A typical spectrum has so many such lines that metaphorically minded astronomers refer to them as the Lyman-alpha forest. *—E.S., P.P. and T.B.*

1 The light begins its journey with a fairly smooth spectrum (*red curve*). It peaks at a wavelength of 122 nanometers, referred to as Lyman alpha.

2 As the light travels toward the earth, two effects change the spectrum: cosmic expansion shifts it to longer wavelengths, and each hydrogen cloud takes a bite out of the spectrum (*sharp dip*). Each bite leaves a new spectral line at 122 nanometers, which later gets shifted along with the rest of the spectrum.



3 By the time the light arrives at the earth, the spectrum has been thoroughly chewed up, with hundreds of hydrogen lines and even the occasional line of a heavier element. In this spectrum of quasar HE 1122–1628, the original peak has been shifted from 122 to 414 nanometers—an indication of the quasar's distance.



cess known as redshifting. The farther away the subject is, the more the universe has expanded since the light began its journey and thus the greater the redshift.

The first quasar spectra were observed in the mid-1960s [see "The Absorption Lines of Quasi-Stellar Objects," by E. Margaret Burbidge and C. Roger Lynds; SCIENTIFIC AMERICAN, December 1970], but it was not until the late 1970s that detectors reached the sensitivity required to yield high-quality spectra. Alec Boksenberg, then at University College London, and Wallace L.W. Sargent of the California Institute of Technology realized that each spectrum had hundreds of absorption lines. On a graph it looks like a dense thicket of lines—hence the name that astronomers give to this phenomenon, the Lyman-alpha forest. The term "Lyman alpha" indicates that the lines appear to be produced by neutral hydrogen gas. That they occur in such profusion indicates that the space between us and the quasar is filled with hundreds of gas clouds, each at a different distance and therefore a different redshift [*see illustration above*].

Ironically, although neutral hydrogen neatly accounts for the lines, it can constitute only a small fraction of the clouds. Ionized hydrogen and helium must make up the bulk. This is because neutral gas naturally absorbs radiation at a range of wavelengths, as the random thermal motion of atoms leads to additional shifts in the spectrum. The mathematically ideal lines broaden into bands of noticeable width. In 1965 James E. Gunn and Bruce A. Peterson, both then at Caltech, showed that if a sufficient fraction of the IGM is neutral—more than one part in a million—the broadening would cause absorption lines of different clouds to overlap. Instead of a forest, astronomers would observe a continuous trough.

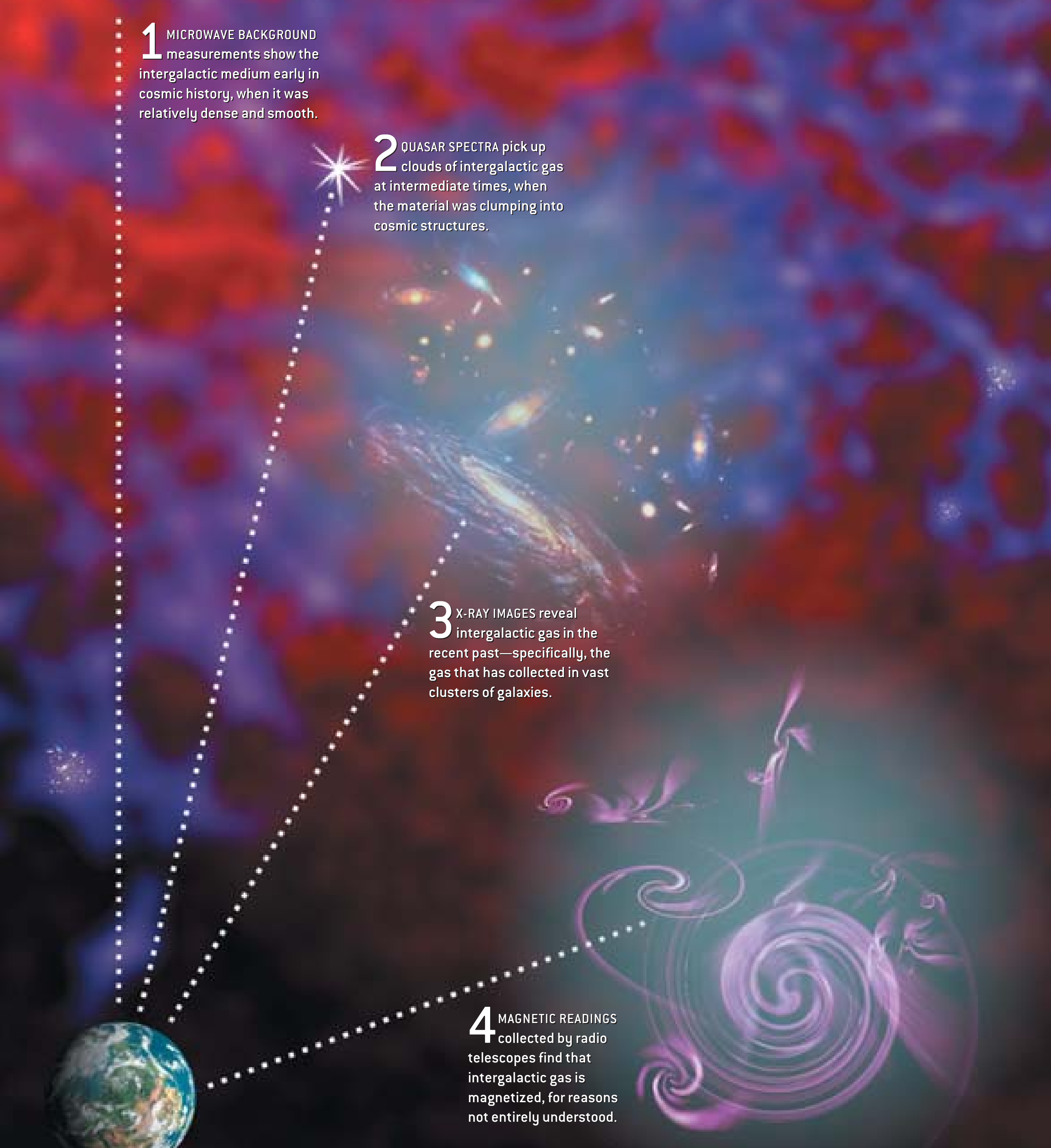
Thus, the simple presence of the Lyman-alpha forest proves that the cool, predominantly neutral IGM necessary for purely gravity-driven galaxy formation was relatively short-lived. Something must have reionized the gas before most quasars formed. An exciting recent discovery concerns one of the most distant and ancient quasars known, named SDSSpJ103027.10+052455.0, which was detected by the Sloan Digital Sky Survey, the most detailed effort yet made to map the sky. Last year Robert H. Becker of the University of California at Davis and his colleagues found an extended range of Lyman-alpha absorption in the spectrum of this object—perhaps the first observation of a trough as predicted by Gunn and Peterson. It may be a glimpse of the period when reionization was still under way.

Not only do quasar spectra tell us about the density and ionization of the IGM, they hint at how the material is distributed in space. In essence, each forestlike spectrum is a core sample through the universe. By comparing core samples with one another and with computer simulations of structure formation, astronomers have sought to reconstruct the full three-dimensional arrangement of matter. Gravitational lensing, whereby

DON DIXON (*illustration*); ALICIA CALLE (*left graphs*); LAURIE GRACE (*right graph*)

FOUR WAYS TO SEE THE UNSEEABLE

INTERGALACTIC GAS is nearly invisible, so astronomers cannot study it directly. Instead they must act as cosmic detectives, reconstructing the history of the gas from four main types of indirect clues.



the gravity of an intervening body bends the quasar light, can help in this process. The bending produces two core samples closer to each other than dumb chance would otherwise allow. In this way, Michael Rauch of the Carnegie Observatories in Pasadena, Calif., and Sargent and Thomas A. Barlow of Caltech measured gas motions within the IGM. They found that although most of the medium is quiescent, the densest patches have been stirred repeatedly by energetic events occurring every 100 million years or so.

In the past seven years, absorption-line studies have detected not only neutral hydrogen but also a smattering of heavier elements. Ionized carbon, with its characteristic "doublet" of twin absorption lines at wavelengths close to Lyman alpha, was the first of these elements to be observed, and others, notably magnesium and oxygen, have followed. In galaxies, atoms of these substances often clump into large molecules—that is, dust particles—that act to redden the light passing through them. No such reddening occurs in the Lyman-alpha clouds, indicating that the heavy elements there remain as individual atoms with a density of about one for every million hydrogen atoms. Although this is not a lot, it is enough to indicate that the IGM is not merely leftover material from galaxy formation. Elements synthesized by stars have somehow made their way out of galaxies and into the space between them.

Seeds of Construction

WHEREAS THE QUASAR SPECTRA probe small, tenuous clouds typically located at enormous distances from the Milky Way (and therefore seen as they were at a much earlier period in cosmic history), the third type of observation concerns itself with the opposite: massive, dense pockets of gas in the comparatively nearby universe. This gas resides in the largest gravitationally bound structures, the massive galaxy clusters. The name "galaxy clusters" is somewhat of a misnomer; these bodies are mostly hot plasma, with some galaxies thrown in like seeds in a watermelon. The ionized gas—nothing more than a compressed form of the intergalactic medium—has been heated to several million kelvins and shines brightly in x-rays. The Chandra X-ray Observatory and X-ray Multi-Mirror Mission have greatly improved our ability to study this gas.

In the conventional view of structure formation, the cluster gas was heated purely by gravitational collapse. If so, its temperature should be related to its mass and density and therefore to its luminosity; specifically, the luminosity should be proportional to the square of the temperature. Yet observations show that luminosity is proportional to temperature to the 3.5th power. Again, it seems that the IGM was the site of some kind of unexpected activity.

The fourth and final type of empirical finding concerns one of the most uncertain, yet potentially crucial, properties of the IGM: its magnetic structure. As electrons move through magnetized regions, they emit light at radio wavelengths. This emission is polarized in the same direction as the magnetic field. Unfortunately for observers, the low density of intergalactic gas makes the signal extremely weak. In 1989 Kwang-Tae Kim and Philipp P. Kronberg, both then at the University of Toronto, and their colleagues found a diffuse bridge of magnetized material that connects two clusters of galaxies, but such measurements have not extended into deeper reaches of space [see "Magnetic Anomalies," by George Musser; News and Analysis, SCIENTIFIC AMERICAN, August 2000]. For the most part, astronomers have relied on clues from large galaxies and clusters. Most spiral galaxies have magnetic fields that are sufficiently strong to affect the galaxies' formation and spin. Their ordered structure implies that a "seed" magnetic field predated the galaxy and strengthened as it took shape. On larger scales, radio studies have found diffuse magnetized gas in several nearby galaxy clusters. A clear implication is that the IGM as a whole is magnetized.

Take a Break

AS INCOMPLETE AS these four types of evidence are, they indicate that the IGM has undergone at least three dramatic changes over the course of cosmic time. The first transition, from ionized to neutral, is the best understood. Known as recombination, it was the event responsible for releasing the microwave background radiation.

The second transition, from neutral back to ionized, is murkier. This reionization may have been caused by quasars, by the stars in early galaxies or even perhaps by a hitherto undetected population of massive stars uniformly distributed through space [see "The First Stars in the Universe," by Richard B. Larson and Volker Bromm; SCIENTIFIC AMERICAN, December 2001]. Although the event seems to have had little effect on the formation of massive galaxies, it may have generated enough thermal pressure to impede the formation of smaller galaxies, complicating the simple picture of purely gravitational structure formation.

To determine which of the many possible sources of reionization played a role, astronomers have studied each in turn. The results are still inconclusive. The best observations of the stellar contribution involve the so-called Lyman-break galaxies, which take their name from a sharp cutoff in their spectra that occurs as neutral hydrogen within the galaxies absorbs starlight. For sufficiently distant galaxies, the break is redshifted from its usual position in the ultraviolet part of the spectrum to the visible part. By searching for a visible-light break, astronomers can identify distant galaxies without having to resort to tricky line-

THE AUTHORS

EVAN SCANNAPIECO, PATRICK PETITJEAN and *TOM BROADHURST* bring both theory and observation to the study of intergalactic space. Scannapieco and Broadhurst did the first theoretical analysis of the effect that galaxy outflows have on the formation of other galaxies. Scannapieco and Petitjean are working together on the clustering of heavy elements observed in quasar spectra. Scannapieco, who also dabbles in cosmology, works under the auspices of the National Science Foundation at the Arcetri Astrophysical Observatory in Florence and the Institute of Astrophysics in Paris. Petitjean is deputy director of the institute and a leader of the European research network on the intergalactic medium. Broadhurst, a visiting professor at the Hebrew University in Jerusalem, is the discoverer of some of the most distant known galaxies.

DON DIXON (*opposite page*)

by-line redshift measurements. This technique, originally developed by Charles C. Steidel of Caltech and his colleagues, has enabled observers to build up sizable catalogues of distant galaxies—whose starlight may have helped reionize intergalactic gas. Unfortunately, the technique suffers from a selection effect: it tends to pick out only the brightest galaxies. Therefore, it does not capture the full stellar contribution to reionization.

Another method is to examine the abundance and distribution of heavy elements. If these elements are observed everywhere, the first objects were probably massive stars smoothly distributed in space. Quasars or dwarf galaxies would scatter the elements more unevenly. So far, however, measurements are too imprecise to provide much guidance. For now, the best scientists can do is to place limits on the spatial distribution of gas. They do so by combining quasar spectra with numerical simulations of structure formation. By adjusting the parameters of the simulation until the spectra match, modelers have drawn our picture of the cosmic web into clearer focus.

Blowout

THE THIRD IGM transformation, the one that accounts for the observed relation of luminosity and temperature in galaxy clusters, remains even more mysterious. The most convincing account dates to work by Nicholas Kaiser, then at Toronto, in 1991. He speculated that cluster gas was preheated to several million kelvins long before gravitational collapse began. This preheating would have reduced the density of the cluster gas, with the largest effect on the smaller clusters, in which gravity is weaker. The decrease in density would have led to lower luminosities and would have accentuated the dependence on temperature, which is related to cluster mass.

The most natural drivers of this preheating were supernova explosions. A rapid succession of supernovae blasts material out of galaxies, injecting not only energy but also heavy elements into the IGM. X-ray satellites have shown that the gas in galactic clusters is indeed enriched in these elements. Furthermore, the degree of enrichment is roughly the same no matter how young or old the clusters are, suggesting that the enrichment occurred early in the clusters' lives. Supernovae would naturally account for this abruptness, as the first wave of stars to form in a galaxy will explode within just a few million years.

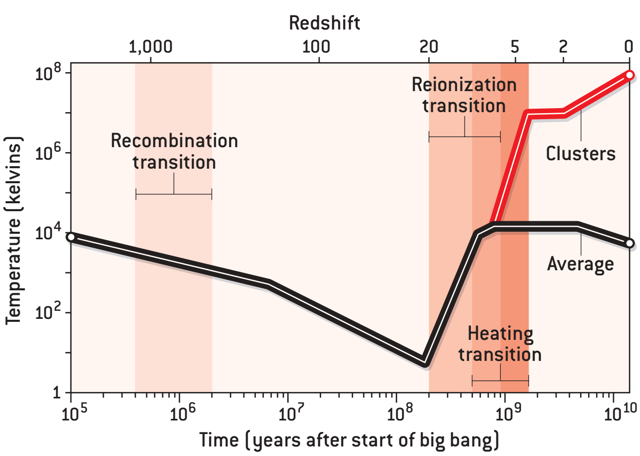
The strongest evidence for the supernova mechanism involves direct observations of distant starbursting dwarf galaxies, which, lacking strong gravity, should be more susceptible to disruption by exploding stars. Max Pettini of the University of Cambridge, Steidel and Alice E. Shapley of Caltech and their collaborators combined galaxy spectra taken in both visible and infrared light. The visible-light spectra contained two sets of lines, one from hydrogen as it emitted light, the other from heavy elements as they absorbed the light of background objects. The infrared spectra contained one set of lines, which were emitted by gaseous nebulae within the galaxy.

Pettini and his colleagues found that these three sets of lines were redshifted by different amounts: the heavy elements by less than the galaxy, the hydrogen by more. In other words, relative to the center of the galaxy, the heavy elements are moving toward us at about 300 kilometers a second, whereas the hydrogen is moving away from us at the same velocity.

This pattern is strange and unexpected. The simplest interpretation is that material is streaming out of the galaxy—a cosmic wind blowing into space. This outflow contains both heavy elements and hydrogen, but in some regions the heavy elements are easier to see, and in other regions the hydrogen is easier to see. For the heavy elements to be visible, they must lie between us and the bulk of the galaxy; otherwise there would be no light for them to absorb. Thus, they must be moving away from the center of the galaxy. The reasoning is reversed for the hydrogen. For it to be visible it must also be moving away from the center but lie on the far side of the galaxy. That way its emitted light is redshifted beyond the wavelength at which intervening matter could block it [*see illustration on opposite page*].

This pattern has been seen in all distant dwarf galaxies for which it is detectable, a fact that suggests that these outflows were once commonplace in the universe. Astronomers have seen gargantuan plumes of material from nearer galaxies as well. One particularly striking case is the dwarf galaxy NGC 1569, which was recently observed by Crystal Martin of the University of California at Santa Barbara and her colleagues. The team found that huge quantities of oxygen and other heavy elements are escaping from the galaxy in bubbles of multimillion-kelvin gas.

The winds stirred the densest regions of the IGM, magnetized vast regions of space and may even have suppressed the formation of small galaxies. The transformation wrought by outflows was much more severe than the earlier reionization. Whereas reionization kept galaxies smaller than a few hundred million solar masses from forming, outflows may have squelched galaxies 10 times larger. This process could resolve one of the most puzzling discrepancies of cosmology: simulations of structure formation predict many more small galaxies than actually exist [see "The Life Cycle of Galaxies," by Guinevere Kauffmann and Frank van den Bosch; SCIENTIFIC AMERICAN, June].

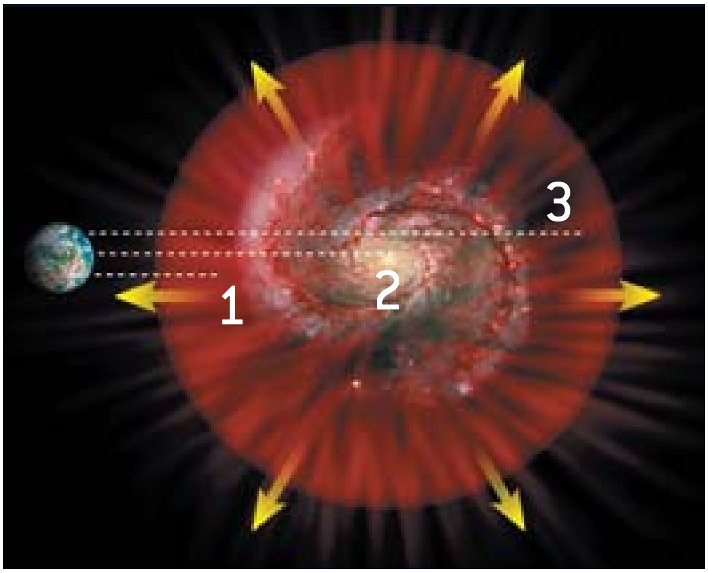


THERMAL HISTORY of the intergalactic medium reveals three important transitions. Evidently the medium has both affected and been affected by the formation of cosmic structures, such as galaxy clusters. Observations indicate that the transitions occur at particular redshifts, which translate (with some uncertainty) into specific times.

SARA CHEN

THE SHAPE OF THE WIND

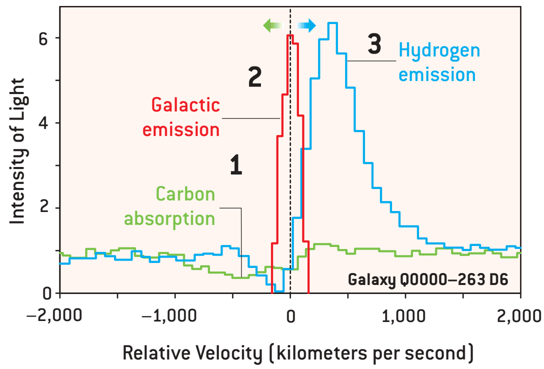
THE STRONGEST EVIDENCE that galaxies are blowing material into intergalactic space comes from studies of galaxy spectra. The wind shows up as a distinctive trio of spectral lines, representing the front (*1*), center (*2*) and rear (*3*) of the galaxy. Two are peaks, indicating emission; the other is a dip, indicating absorption. *—E.S., P.P. and T.B.*



1 As galactic light travels toward Earth, it passes through heavy elements such as carbon, which absorb certain wavelengths. This material is moving toward us, shifting the absorption to negative velocities with respect to the galaxy.

2 Nebulae within the galaxy itself emit infrared light, serving as a reference point.

3 Hydrogen on the far side of the galaxy emits light. The motion of the hydrogen away from us shifts the emission to positive velocities. This shift has the side effect of allowing the emission to pass through the galaxy without getting reabsorbed.



Exerting Self-Control

THUS, EACH GENERATION of objects alters the IGM, which in turn determines the properties of the next generation. The sources that reionized the universe generated enough thermal pressure to regulate their own formation, and the winds from starbursting galaxies may have been strong enough to nip other such galaxies in the bud. Analogous feedback processes occur within galaxies themselves, in which supernovae and ultraviolet starlight act on the interstellar gas out of which the next generation of stars forms [see "The Gas between the Stars," by Ronald J. Reynolds; SCIENTIFIC AMERICAN, January]. In fact, the concept of feedback is becoming a unifying theme in astronomy, seemingly repeating itself on all scales.

The four main types of observation are continually being complemented by new results and innovative methods. Observations of the microwave background radiation, for example, are now sensitive enough to detect the slight blurring caused by IGM material. Patchiness during reionization should scatter some of the microwave photons, and hot areas of the IGM, such as galaxy clusters, should further distort the radiation. The latter contribution, known as the Sunyaev-Zeldovich effect, has already been studied in individual clusters, and its broad-scale effect was tentatively detected by the Cosmic Background Imager experiment this past summer.

In an ingenious variation on quasar absorption-line studies, Kenneth R. Sembach of the Space Telescope Science Institute, Blair Savage and Bart Wakker of the University of Wisconsin–Madison and their colleagues have used the Far Ultraviolet Spectroscopic Explorer satellite to study the IGM in the immediate neighborhood of the Milky Way. Their observations indicate that nearby gas clouds are distributed in an uneven manner reminiscent of the cosmic web between galaxy clusters and yet are moving through a million-kelvin medium like the gas within clusters. Thus, the Local Group of galaxies may be surrounded by an extended corona of gas, whose properties are suggestive of both the most diffuse and the densest regions of the IGM. Similar hot regions could make up a hitherto unknown component of the IGM and account for a large fraction of its mass.

New studies such as these make it clear that the story of the intergalactic medium is only beginning to be told and that surprises await us. Like an intrepid intergalactic traveler, we have worked our way from the familiar solar neighborhood into the depths of the most desolate places imaginable. Our eyes are still adjusting to the unexpected and intricate beauty of the cosmic web that stretches across the emptiest places. SA

MORE TO EXPLORE

The Lyman-Alpha Forest at Low-Redshift: Tracing the Dark Matter Filaments. Patrick Petitjean, Jan P. Mücket and Ronald E. Kates in *Astronomy & Astrophysics,* Vol. 295, No. 1, pages L9–L12; March 1995. Available online at **arXiv.org/abs/astro-ph/9502100**

The Intergalactic Medium. Piero Madau in *Encyclopedia of Astronomy and Astrophysics.* Edited by Paul Murdin. Institute of Physics Publishing, 2001. **astro-ph/0005106**

The Role of Heating and Enrichment in Galaxy Formation. Evan Scannapieco and Tom Broadhurst in *Astrophysical Journal,* Vol. 549, No. 1, pages 28–45; March 2001. **astro-ph/0003104**

DON DIXON (*illustration*); SARA CHEN (*graph*)

VEHICLE OF CHANGE

Hydrogen fuel-cell cars could be the catalyst for a cleaner tomorrow

BY LAWRENCE D. BURNS, J. BYRON McCORMICK AND CHRISTOPHER E. BORRONI-BIRD

RUNNING BOARD: General Motors's Hy-wire prototype hydrogen fuel-cell vehicle sits up on a lift, stripped down to its operational essentials—a "skateboard" chassis filled with a fuel-cell stack, electric motor and programmable electronic controls.

When Karl Benz rolled his Patent Motorcar out of the barn in 1886, he literally set the wheels of change in motion. The advent of the automobile led to dramatic alterations in people's way of life as well as the global economy—transformations that no one expected at the time. The ever increasing availability of economical personal transportation remade the world into a more accessible place while spawning a complex industrial infrastructure that shaped modern society.

Now another revolution could be sparked by automotive technology: one fueled by hydrogen rather than petroleum. Fuel cells—which cleave hydrogen atoms into protons and electrons that drive electric motors while emitting nothing worse than water vapor—could make the automobile much more environmentally friendly. Not only could cars become cleaner, they could also become safer, more comfortable, more personalized—and even perhaps less expensive. Further, these fuel-cell vehicles could be instrumental in motivating a shift toward a "greener" energy economy based on hydrogen. As that occurs, energy use and production could change significantly. Thus, hydrogen fuel-cell cars and trucks could help ensure a future in which personal mobility—the freedom to travel independently—could be sustained indefinitely, without compromising the environment or depleting the earth's natural resources.

A confluence of factors makes the big change seem increasingly likely. For one, the petroleum-fueled internal-combustion engine (ICE), as highly refined, reliable and economical as it is, is finally reaching its limits. Despite steady improvements, today's ICE vehicles are only 20 to 25 percent efficient in converting the energy content of fuels into drive-wheel power. And although the U.S. auto industry has cut exhaust emissions substantially since the unregulated 1960s—hydrocarbons dropped by 99 percent, carbon monoxide by 96 percent and nitrogen oxides by 95 percent—the continued production of carbon dioxide causes concern because of its potential to change the planet's climate.

Even with the application of new technologies, the efficiency of the petroleum-fueled ICE is expected to plateau around 30 percent—and whatever happens, it will still discharge carbon dioxide. In comparison, the hydrogen fuel-cell vehicle is nearly twice as efficient, so it will require just half the fuel energy. Of even more significance, fuel cells emit only water and heat as byproducts. Finally, hydrogen gas can be extracted from various fuels and energy sources, such as natural gas, ethanol, water (via electrolysis using electricity) and, eventually, renewable energy systems. Realizing this potential, an impressive roster of automotive companies are making a sustained effort to develop fuel-cell vehicles, including DaimlerChrysler, Ford, General Motors, Honda, PSA Peugeot-Citroën, Renault-Nissan and Toyota.

Overview/*Hydrogen Fuel Cells*

- Fuel cells convert hydrogen gas into electricity cleanly, making possible nonpolluting vehicles powered by electric drive motors. When combined with compact drive-by-wire electronic steering, brakes and throttle controls, fuel-cell technology allows engineers to split a vehicle into a rolling chassis and a (potentially interchangeable) body with an expansive interior.
- The prospect of clean hydrogen fuel-cell vehicles could also augur an altered energy economy and a sustainable environment without compromising personal mobility.
- The chicken-and-egg problem: large numbers of fuel-cell vehicles require adequate fuel availability to support them, but the required infrastructure is hard to build unless there are significant numbers of fuel-cell vehicles on the roadways.

It's an Automotive World

IT IS IMPORTANT to find a better solution to the problems posed by personal transportation because the environmental impact of vehicles is apt to wax as use booms. In 1960 fewer than 4 percent of the world's population possessed vehicles. Twenty years later 9 percent were owners, and currently the share has reached 12 percent. Based on present growth rates, as many as 15 percent of the people living on the planet could have a vehicle by 2020. And because the world's population may climb from six billion today to nearly 7.5 billion two decades hence, the total number of vehicles could increase from about 700 million to more than 1.1 billion. This projected expansion will be spurred by the burgeoning of the middle class in the developing world, which translates into rising per capita income. Higher income correlates almost directly with automobile ownership.

GENERAL MOTORS (*preceding pages*)

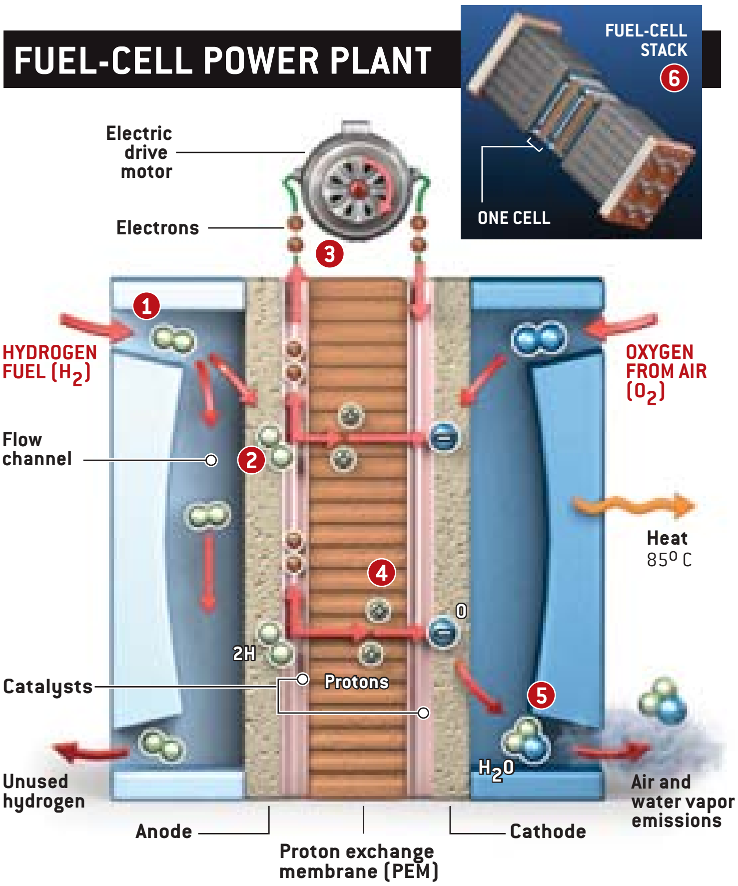
Three quarters of all automobiles are now concentrated in the U.S., Europe and Japan. We expect, however, that more than 60 percent of the increase in new vehicle sales during the next 10 years will occur in eight emerging markets: China, Brazil, India, Korea, Russia, Mexico, Poland and Thailand. The challenge will be to create compelling, affordable and profitable vehicles that are safe, effective and environmentally sustainable.

Rethinking Automotive Propulsion

TO UNDERSTAND WHY this technology could be so revolutionary, consider the operation of a fuel-cell vehicle, which at base is a vehicle with an electric traction drive. Instead of an electrochemical battery, though, the motor gets power from a fuel-cell unit [*see illustration below*]. Electricity is produced when electrons are stripped from hydrogen fuel traveling through a membrane in the cell. The resulting current runs the electric motor, which turns the wheels. The hydrogen protons then combine with oxygen and electrons to form water. When using pure hydrogen, a fuel-cell car is a zero-emission vehicle.

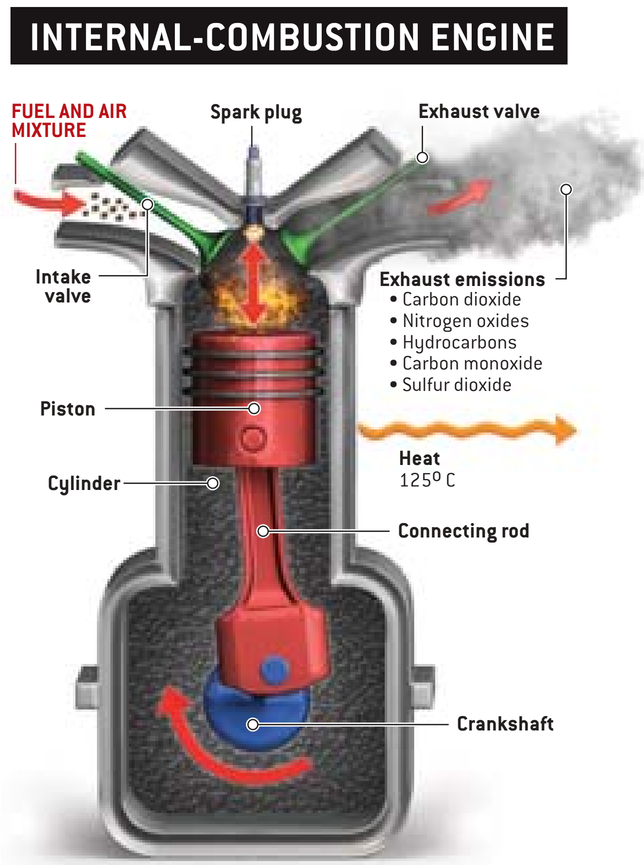
Although it takes energy to extract hydrogen from substances, by either reforming hydrocarbon molecules with catalysts or splitting water with electricity, the fuel cell's high efficiency more than compensates for the energy required to accomplish these processes, as we will show later. Of course, this energy has to come from somewhere. Some generation sources, such as natural gas-, oil- and coal-burning power facilities, produce carbon dioxide and other greenhouse gases. Others, including nuclear plants, do not. An optimal goal would be to produce electricity from renewable sources such as biomass, hydroelectric, solar, wind or geothermal energy.

By adopting hydrogen as an automotive fuel, the transportation industry could begin the transition from near-total reliance on petroleum to a mix of fuel sources. Today 98 percent of the energy used to power automobiles is derived from



UP TO 55% EFFICIENCY

ELECTROCHEMISTRY VS. COMBUSTION: A proton exchange membrane (PEM) fuel cell comprises two thin, porous electrodes, an anode and a cathode, separated by a polymer membrane electrolyte that passes only protons. Catalysts coat one side of each electrode. After hydrogen enters (*1*), the anode catalyst splits it into electrons and protons (*2*). The electrons travel off to power a drive motor (*3*), while the protons migrate through the membrane (*4*) to the cathode. Its catalyst combines the protons with returning electrons and oxygen from the air to form water (*5*). Cells can be stacked to provide higher voltages (*6*).



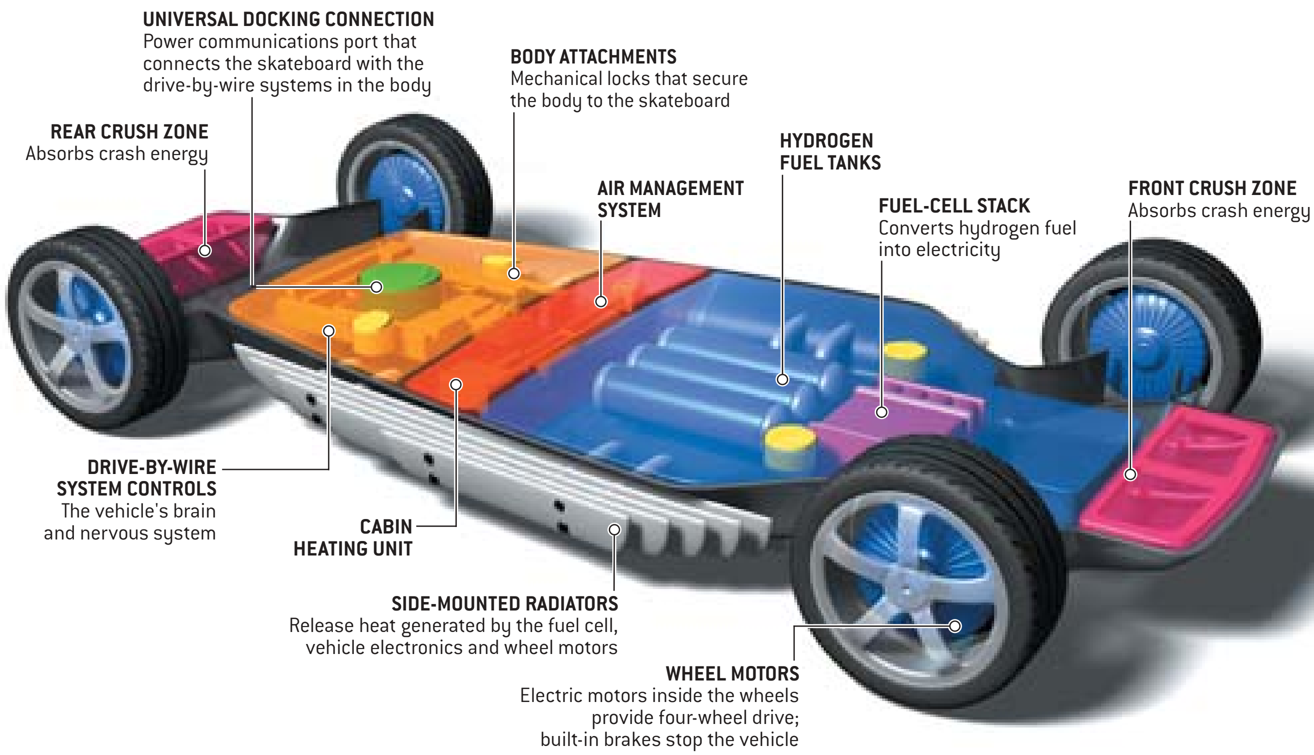
UP TO 30% EFFICIENCY

MOST U.S. CARS use four-stroke internal-combustion engines. The piston, which travels up and down when the crankshaft rotates, starts at the top of the cylinder. The intake valve opens and the piston drops, allowing the fuel/air mixture to enter the cylinder. The piston moves back up, compressing the gasoline and air. The spark plug fires, igniting the fuel droplets. The compressed charge explodes, driving the piston down. The exhaust valve opens, allowing the combustion products to exit the cylinder.

JOE ZEFF

AUTONOMY'S "SKATEBOARD" CHASSIS

SHOEHORNING functional automotive systems into the flat, skateboard-like chassis is the key to General Motors's AUTOnomy concept for a future hydrogen fuel-cell vehicle. That and the use of compact electronic drive-by-wire technology for steering, braking and throttling permits designers much greater freedom in configuring the upper bodies. It means no more bulky engine compartment, awkward center cabin hump or conventional steering wheel to work around. The novel approach also allows bodies to be interchangeable. Owners could have new, personalized bodies "plugged in" to their used chassis at the dealership, or do it themselves—turning, say, a family sedan into a minivan or a luxury car.



JOE ZEFF

petroleum. As a result, roughly two thirds of the oil imported into the U.S. is devoted to transportation. By supplementing fossil fuels, the U.S. can reduce dependence on foreign oil and foster development of local, more environmentally friendly energy sources. This effort will also introduce competition into energy pricing—which could lower and stabilize fuel and energy costs in the long term.

Revamping Vehicle Design

ANOTHER KEY to producing a truly revolutionary vehicle is the integration of the fuel cell with drive-by-wire technology, replacing previous, predominantly mechanical systems for steering, braking, throttling and other functions with electronically controlled units. This frees up space because electronic systems tend to be less bulky than mechanical ones. By-wire system performance can be programmed using software. In addition, with no conventional drivetrain to limit structural and styling choices, automakers will be free to create dramatically different designs to satisfy customer needs.

Replacing conventional ICEs with fuel cells enables the use of a flat chassis, which gives designers great freedom to create unique body styles. Drive-by-wire technology similarly liberates the interior design because the driving controls can be radically altered and can be operated from different seating positions. Recognizing this design opportunity, General Motors came up with a concept called AUTOnomy, which the company introduced early this year. A drivable prototype, Hy-wire (for *hy*drogen by-*wire*), debuted at the Paris Motor Show in late September.

The AUTOnomy concept and the Hy-wire prototype were created, literally, from the wheels up. The foundation for both is a thin, skateboard-like chassis containing the fuel cell, electric drive motor, hydrogen storage tanks, electronic controls and heat exchangers, as well as braking and steering systems [*see illustration above*]. There is no internal-combustion engine, transmission, drivetrain, axles or mechanical linkages.

In a fully developed AUTOnomy-type vehicle, drive-by-wire technology would require only one simple electrical connection and a set of mechanical links to unite chassis and body. The body could plug into the chassis much like a laptop connects to a docking station. The single-electrical-port concept creates a quick and easy way to link all the body systems—controls, power and heating—to the skateboard. This simple separation of body and chassis can help keep the vehicle body lightweight and uncomplicated. It also makes the body easily replaceable.

The hydrogen fuel-cell vehicle is nearly twice as efficient as an internal-combustion engine, so it will require only half the fuel energy.

GENERAL MOTORS

In principle, simply by having the dealer or car owner "pop on" an interchangeable body module, the vehicle could be a luxury car today, a family sedan next week or a minivan next year.

Much like a computer, vehicle systems would be upgradable through software. As a result, service personnel could download programs as desired to improve vehicle performance or to tailor particular ride and handling characteristics to suit a particular vehicle brand, body style or customer preference.

With drive-by-wire electronic controls, the driver needs no steering wheel, gear shifter or foot pedals. GM's Hy-wire prototype is equipped with a steering guide control called X-Drive that easily moves from side to side across the width of the car to accommodate left- and right-hand driving positions. The X-Drive operates something like a motorcycle's handgrips: the driver accelerates by twisting the handgrips and brakes by squeezing them. Steering involves a turning action similar to today's steering wheel. The driver also has the option to brake and accelerate with either the right or left hand, with braking taking priority in the case of mixed signals. Motorists start the vehicle by pushing a single power button and then select one of three settings: neutral, drive or reverse.

X-Drive also eliminates the conventional instrument panel and steering column, which frees up the vehicle interior and allows novel placement of seats and storage areas. For example, because there is no engine compartment, the driver and all passengers have more visibility and much greater legroom than in a conventional vehicle of the same length.

By lowering the vehicle's center of gravity and eliminating the rigid engine block in front of the passengers, an AUTOnomy-like skateboard chassis can improve ride, handling and stability characteristics beyond what is possible with conventional vehicle architecture.

Reorganizing the Automobile Business

THE SIMPLIFIED DESIGN of an advanced fuel-cell vehicle, as suggested by the AUTOnomy concept and the Hy-wire prototype, could have a profound effect on vehicle manufacturing, perhaps setting the stage for a reinvention of the automobile business. Today's auto industry is capital-intensive, with modest profit margins. Even as car companies are aggressively managing the costs of vehicle development and manufacturing, excess production capacity in the global industry is driving down vehicle prices. At the same time, the regulatory standards–driven content of cars and trucks continues to grow, pushing up costs. Taken together, lower prices and higher costs are threatening profit margins.

A concept such as that of AUTOnomy, however, could significantly change the current business model. It could conceivably lower vehicle development costs because, with modules able to be produced independently, design changes to the body and chassis modules could be made more easily and cheaply. As with today's truck platform derivatives, it will be possible to design the chassis only once to accommodate various body styles. These derivatives could easily have different front ends, interior layouts and chassis tuning. With perhaps only three chassis needed—compact, midsize and large—production volumes could be much larger than those now, bringing greater economies of scale.

Having far fewer components and part types will further reduce costs. The fuel-cell stack, for example, is created from a series of identical individual cells, each comprising a flat cathode sheet and similar anode component separated by a polymer-electrolyte membrane. Depending on the power requirements of a particular vehicle (or other device, such as a stationary electricity generator), the number of cells in the stack can be scaled up or down.

Although automotive fuel-cell technology is far from economical at present (thousands of dollars per kilowatt for a handbuilt prototype), costs have begun to decline dramatically. For instance, the 10-fold increase in the power density of fuel-cell stacks achieved over the past five years has been accompanied by a 10-fold decrease in their cost. And whereas fuel cells currently require precious metals for catalysts and expensive polymer membranes, scientists are making progress in finding ways to minimize the use of catalysts and make membrane materials cheaper.

The AUTOnomy concept also makes it possible to decouple body and chassis manufacture. A global manufacturer could

THE AUTHORS

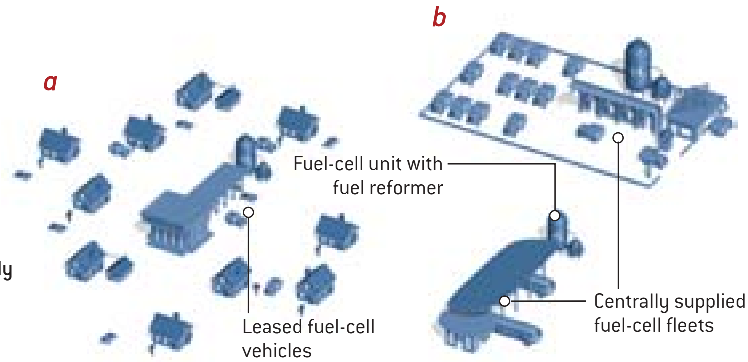
LAWRENCE D. BURNS, J. BYRON McCORMICK and *CHRISTOPHER E. BORRONI-BIRD* play leading roles in the fuel-cell development efforts of General Motors. Burns is vice president of GM Research & Development and Planning. He oversees the company's advanced technology and innovation programs and is responsible for the company's product portfolio, capacity and business plans. Burns is a member of the Automotive Strategy Board, GM's highest-level management team. McCormick is executive director of GM's Fuel Cell Activities. He has been involved in fuel-cell research throughout his career, initiating and then heading the Fuel Cells for Transportation program at Los Alamos National Laboratories before joining GM in 1986. Borroni-Bird joined GM in June 2000 as director of design and technology fusion, a group that applies emerging technology to improve vehicle design. He is also director of the AUTOnomy program, which includes the Hy-wire prototype vehicle. Previously Borroni-Bird managed Chrysler's Jeep Commander fuel-cell vehicle program.

STEPS TOWARD A HYDROGEN SOCIETY

WITHIN A FEW YEARS

a Small numbers of prototype vehicles are tested by leasing them to residents living near a hydrogen fueling station.

b Transit and business fleets that return to the garage each day, such as buses, mail trucks and delivery vans, start to be supplied by centrally located hydrogen stations.

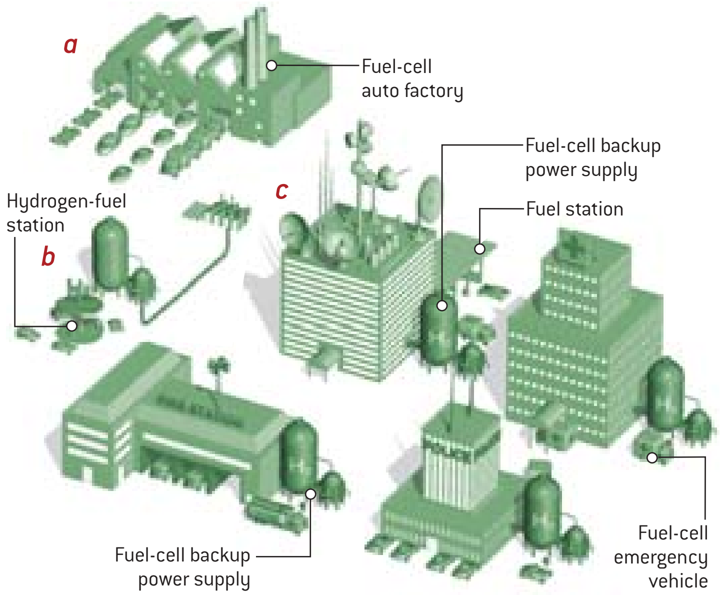


WITHIN A DECADE

a Car plants manufacture fuel-cell-powered "skateboards" and a few different "snap-on" body types.

b Hydrogen fueling stations with on-site natural gas reformers (chemical cracking units) are installed to provide hydrogen to early production vehicles.

c Stationary power generators, which reform natural gas into hydrogen and feed into the fuel cell, are installed in enterprises that require high-reliability "premium" power for data communications, continuous manufacturing, or emergency medicine. For example, ambulances and emergency vehicles refuel at the hospital fuel-cell unit.



IN MORE THAN A DECADE

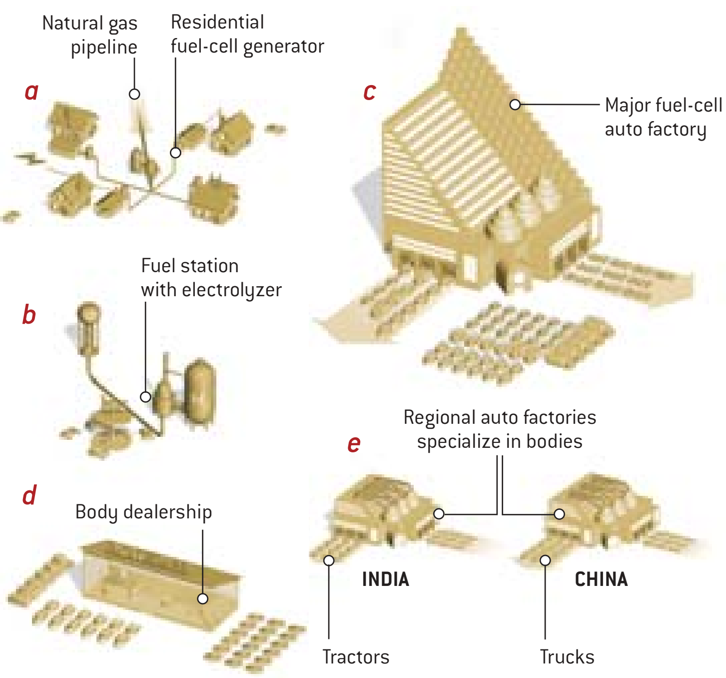
a Stationary reformer/fuel-cell units sited at more types of businesses and, eventually, homes sell extra power to the electricity grid in what's called a distributed generation system. These installations begin to provide hydrogen locally to employees.

b More hydrogen stations that use electrolyzers come online.

c Huge assembly plants put out three sizes of fuel-cell skateboard chassis (compact, midsize, large).

d Dealers sell new bodies in various styles for drivers' used skateboards.

e Other plants in different regions build diverse bodies for their local markets (for example, in India and China, tractors and trucks).



ENERGY SOURCES

Current fossil-fuel, nuclear and hydroelectric generation will be increasingly augmented by cleaner and renewable technologies.

SLIM FILMS

Hydrogen fuel-cell cars and trucks could help ensure a future in which personal mobility could be sustained indefinitely.

build and ship the chassis (an ideal scenario, given its thin profile), and local firms could build the bodies and assemble the complete vehicles. The chassis could be very economical because it would be mass-produced.

In high-end markets, this kind of arrangement might mean that new chassis might debut every three or four years—when software upgrades could no longer match performance desires—but that customers could purchase a new body module annually or lease one even more frequently. In addition, if chassis hardware is developed appropriately, then new hardware and software upgrades could become practical. Alternatively, consumers who wish to keep their vehicle body but want a higher-performance chassis could buy one. In less affluent markets, the chassis would comprise durable hardware and could be financed for much longer periods, perhaps decades.

Storing Hydrogen

THIS IS NOT TO SAY that all the technical barriers to engineering practical fuel-cell vehicles have been surmounted. Many obstacles have yet to be overcome before they will achieve the convenience and performance that customers have come to expect from their ICE automobiles. One of the biggest hurdles is the development of safe and effective onboard hydrogen storage technology that would provide sufficient driving range—about 300 miles. Any acceptable storage technology must be durable enough to run for at least 150,000 miles. It must function in temperatures from –40 to 113 degrees Fahrenheit (–40 to 45 degrees Celsius). And the refueling process must be simple and take less than five minutes to complete. There are various approaches to storing hydrogen, including liquid, compressed gas and solid-state methods. All are promising, yet all present challenges.

Compressed-gas tanks are the most likely to be used early on, but high compression remains a perceived safety issue. Currently these systems carry about 5,000 pounds per square inch (psi) of hydrogen (350 bars), but the goal is 10,000 psi (700 bars) to extend vehicle range. For safety purposes, the tank must have an impact burst strength of at least twice the pressure of the fuel. Vessels are currently made from materials that are either very expensive, such as carbon fiber, or very heavy. They are also relatively large, making it difficult to fit them in a vehicle.

Hydrogen can also be stored in liquid form, but a substantial amount of energy is needed to chill it to the extremely low temperatures required (–423 degrees F or –253 degrees C). Further, as much as 3 to 4 percent of the hydrogen will still "boil off" every day. Although most of this boil-off will be used by the vehicle, it would be a concern for cars parked for several days between trips.

A longer-term solution is to transport hydrogen using a solid-state approach. One promising alternative is metal-hydride storage. In this method, the hydrogen is held in the interstices of pressed metallic alloy powder, much like a sponge absorbs water. This technique has many encouraging aspects, including straightforward construction, a high degree of safety and promising storage capacities. But temperatures in the range of 150 to 300 degrees C are needed to extract the hydrogen from the metal hydride. To avoid an energy penalty, the hydrogen must be released at a temperature nearer to 80 degrees C. Although research is still in the early stages, solid-state storage is tantalizing.

Reworking the Infrastructure

MOMENTOUS AS THE CHANGES to the automotive business might be, they could be overshadowed by the potential influence of AUTOnomy-type vehicles on the world's energy supply system. Viewed from where we are today, fuel cells and a hydrogen fueling infrastructure are a chicken-and-egg problem. We cannot have large numbers of fuel-cell vehicles without adequate fuel availability to support them, but we will not be able to create the required infrastructure unless there are significant numbers of fuel-cell vehicles on the roadways. Given that the creation of a potentially costly hydrogen generation/distribution network in the U.S. is a prerequisite to commercializing fuel-cell cars and trucks, strong advocacy from local and national leaders in the public and private sectors is crucial. Key issues that must be addressed include subsidy funding, incentives for developing refueling stations, creation of uniform standards, and general education about the topic. The FreedomCAR initiative announced by the U.S. Department of Energy earlier this year, a public-private partnership to promote the development of fuel-cell power and hydrogen as a primary fuel for cars and trucks, is a step in the right direction. Government

Major Automotive Fuel-Cell Developers

DaimlerChrysler AG	*Stuttgart-Mohringen, Germany*
Ford Motor Co.	*Dearborn, Mich.*
General Motors Corp.	*Detroit, Mich.*
Honda Motor Company Ltd.	*Tokyo*
PSA Peugeot Citroën	*Paris*
Renault-Nissan Alliance	*Paris/Tokyo*
Toyota Motor Corp.	*Toyota City, Japan*

support for the research and pilot demonstrations required to prove the feasibility of the infrastructure will be needed.

To be sure, industry also must do its part to enable the difficult transition to the hydrogen economy. GM is now developing a bridge strategy that should move things along. We are working on bringing to market interim hydrogen-based fuel-cell products that will earn revenues to help offset the hundreds of millions of dollars that the company is investing on fuel-cell technology, while providing real-world operating experience.

It is likely that fuel-cell generators will be marketed for use in businesses and, eventually, homes, before fuel-cell vehicles are widely available. These applications are much less complex than automobiles, which have very demanding performance requirements. GM has developed prototype stationary fuel-cell generators that run on hydrogen extracted from fossil fuels.

Within the next few years, GM plans to unveil a range of stationary fuel-cell generators that are aimed at the "premium power," or high-reliability "guaranteed power," energy market segment. This $10-billion-a-year business encompasses energy consumers that cannot afford to be without electricity, including digital-data centers, hospitals, factories using continuous industrial processes, and telecommunications companies. These generators would enable cost reduction through the ability to cut power usage during peak periods as well as provide revenue through net metering (selling power back to the grid). One of our initial products will be a 75-kilowatt unit incorporating a reformer that extracts hydrogen for the fuel-cell stack from natural gas, methane or gasoline. No breakthrough technical developments are needed to build these stationary power products. When operational, these decentralized power systems can also be used to refuel vehicles with hydrogen.

Once safe and reliable hydrogen storage methods are available, off-board fuel processing at the filling station becomes a viable avenue for generating the hydrogen needed for transportation. An advantage of fuel processing, of course, is that most of the infrastructure required to implement it already exists. The current petroleum-based fuel distribution network could be retrofitted by installing fuel reformers or electrolyzers right at the corner gas station, allowing local operators to generate hydrogen on the spot and pump it for their customers. With this approach, there would be no need to build new long-distance pipelines or dismantle the present automotive servicing infrastructure. As we begin the transition from petroleum to hydrogen, this might well be the optimal way to proceed.

An even more radical scheme would be to refuel at home or at work using the distribution network that currently provides natural gas to individual homes and businesses. Natural gas pipelines are as common in many areas as gasoline stations, making this infrastructure an ideal conduit for hydrogen. The natural gas could be reformed into hydrogen and then stored onboard the vehicle. Alternatively, electricity from the utility grid could produce the hydrogen. Electricity purchased during off-peak hours, such as when your car is housed overnight in a garage, might eventually be an affordable way to refuel in some locales.

As vehicle power-generation systems become more sophisticated, we see the role of the automobile within the global power grid changing. Vehicles could at some point become a new power-generation source, supplying electricity to homes and work sites. Most vehicles sit idle about 90 percent of the time, so imagine the exponential growth in power availability if the electrical grid could be supplemented by the generating capacity of cars and trucks in every driveway or parking garage. Consider, for example, that if only one out of every 25 vehicles in

HY-WIRE OPENS UP THE INTERIOR

FLEXIBILITY IN DESIGN and consumer choice is the key to GM's strategy of stuffing all the car's operational systems into the skateboard chassis. Body designers are now free to explore passenger compartment configurations unhindered by traditional limits such as the dashboard and central hump. The body designs could also be interchangeable, allowing a single chassis to feature a wardrobe of alternative tops.

GENERAL MOTORS

Hydrogen can be generated from natural gas at a cost that is comparable with conventional fuel costs.

California today were a fuel-cell vehicle, their combined generating capacity would exceed that of the state's utility grid.

Obviously there are multiple options to choose from in creating a hydrogen distribution network. Although the scenarios we have painted are plausible, one of the most important factors in determining what the infrastructure will eventually look like is cost. Energy companies around the world are studying the economics of hydrogen. In recent testimony before the U.S. House of Representatives Committee on Science's Energy Subcommittee, James P. Uihlein of BP stated that hydrogen can be generated from natural gas at a cost that is comparable with conventional fuel costs. In fact, he went on to note, at the refinery gate, hydrogen's cost-per-mile-driven is actually substantially less than conventional fuel because of the outstanding efficiency of the fuel-cell engine. Hydrogen's current high cost, Uihlein said, can be attributed to the expense of transporting and dispensing it.

Hydrogen Matters

DEPENDING ON THE FEEDSTOCK and the production and distribution methods used, the cost of a kilogram of hydrogen can be four to six times as high as the cost of a gallon of gasoline or diesel fuel. (A kilogram of hydrogen is the energy equivalent of a gallon of petroleum-based fuel.) Yet because an optimized fuel-cell vehicle is likely to be at least twice as efficient as an ICE vehicle, it will go twice as far on that kilogram of fuel. Therefore, hydrogen should become commercially viable if its retail price per kilogram is double that of a gallon of gasoline. As improvements in hydrogen storage, fuel processing and electrolysis technologies are achieved and as demand for hydrogen increases, the cost of hydrogen should move nearer to the required price range. In fact, recent studies indicate that with today's technology we are within a factor of 1.3 of where we would like to be in terms of price.

Even though we are in the early stages of exploring solutions, we believe that when the infrastructure is required, it could develop rapidly, despite the enormous challenges involved. That was the case a century ago, when the gasoline automobile was proving its usefulness to customers, and the infrastructure to support it grew quickly. Entrepreneurs are always ready to seize new opportunities. The world is already beginning to develop the technologies needed for hydrogen production and distribution. Nevertheless, the size and scope of this particular infrastructure are huge, and there are significant technical obstacles ahead.

As discussions about how to create the required distribution network continue, it is interesting to note that hydrogen infrastructures are currently installed in several locations, most notably along the U.S. Gulf Coast and in Europe around Rotterdam, the Netherlands. Hydrogen is produced by the oil and chemical industries (it is used for sulfur removal in the petroleum-refining process), so it flows today through hundreds of miles of pipeline in a number of countries. The existing infrastructure annually produces approximately 540 billion cubic meters of hydrogen, primarily reformed from natural gas. On an energy-equivalent basis, this equates to roughly 140 million tons of petroleum a year, which is almost 10 percent of the present transportation demand. Even though the infrastructure is dedicated to other uses, the fact that it is already in place demonstrates that a great deal of expertise on generating and transporting hydrogen is available.

Like any advance that has the potential to change the dominant technology completely, the implementation of fuel cells and the transition to a hydrogen-based energy infrastructure will take time. Although a precise timetable is hard to predict, given our current technological momentum and business realities, we aim to have compelling and affordable fuel-cell vehicles on the road by the end of this decade. We then anticipate a significant increase in the penetration of fuel-cell vehicles between 2010 and 2020 as automakers begin to create the installed base necessary to support high-volume production. Many of these companies, including GM, have invested hundreds of millions of dollars in fuel-cell research and development, and the sooner they can anticipate a return on these investments, the better.

Because it takes about 20 years to turn over the entire vehicle fleet, it will take at least that long to reap the full extent of the environmental and energy benefits that hydrogen fuel-cell vehicles can provide. But the AUTOnomy concept brings that future nearer—and makes it clearer. Instead of the historical evolution of the automobile, we now see the development of revolutionary technologies that fundamentally reinvent the automobile and its role in our world. SA

MORE TO EXPLORE

Prepared Statement of James P. Uihlein to the U.S. House of Representatives Committee on Science, Subcommittee on Energy. Field Hearing on Fuel Cells: The Key to Energy Independence? June 24, 2002.

Designing AUTOnomy. Christopher E. Borroni-Bird. Available at **www.sciam.com/explore_directory.cfm**

GENERAL MOTORS

Skin

Deep

Throughout the world, human skin color has evolved to be dark enough to prevent sunlight from destroying the nutrient folate but light enough to foster the production of vitamin D

By Nina G. Jablonski and George Chaplin

Among primates, only humans have

a mostly naked skin that comes in different colors. Geographers and anthropologists have long recognized that the distribution of skin colors among indigenous populations is not random: darker peoples tend to be found nearer the equator, lighter ones closer to the poles. For years, the prevailing theory has been that darker skins evolved to protect against skin cancer. But a series of discoveries has led us to construct a new framework for understanding the evolutionary basis of variations in human skin color. Recent epidemiological and physiological evidence suggests to us that the worldwide pattern of human skin color is the product of natural selection acting to regulate the effects of the sun's ultraviolet (UV) radiation on key nutrients crucial to reproductive success.

From Hirsute to Hairless

THE EVOLUTION OF SKIN PIGMENTATION is linked with that of hairlessness, and to comprehend both these stories, we need to page back in human history. Human beings have been evolving as an independent lineage of apes since at least seven million years ago, when our immediate ancestors diverged from those of our closest relatives, chimpanzees. Because chimpanzees have changed less over time than humans have, they can provide an idea of what human anatomy and physiology must have been like. Chimpanzees' skin is light in color and is covered by hair over most of their bodies. Young animals have pink faces, hands, and feet and become freckled or dark in these areas only as they are exposed to sun with age. The earliest humans almost certainly had a light skin covered with hair. Presumably hair loss occurred first, then skin color changed. But that leads to the question, When did we lose our hair?

The skeletons of ancient humans—such as the well-known skeleton of Lucy, which dates to about 3.2 million years ago—give us a good idea of the build and the way of life of our ancestors. The daily activities of Lucy and other hominids that lived before about three million years ago appear to have been similar to those of primates living on the open savannas of Africa today. They probably spent much of their day foraging for food over three to four miles before retiring to the safety of trees to sleep.

By 1.6 million years ago, however, we see evidence that this pattern had begun to change dramatically. The famous skeleton of Turkana Boy—which belonged to the species *Homo ergaster*—is that of a long-legged, striding biped that probably walked long distances. These more active early humans faced the problem of staying cool and protecting their brains from overheating. Peter Wheeler of John Moores University in Liverpool, England, has shown that this was accomplished through an increase in the number of sweat glands on the surface of the body and a reduction in the covering of body hair. Once rid of most of their hair, early members of the genus *Homo* then encountered the challenge of protecting their skin from the damaging effects of sunlight, especially UV rays.

Built-in Sunscreen

IN CHIMPANZEES, the skin on the hairless parts of the body contains cells called melanocytes that are capable of synthesizing the dark-brown pigment melanin in response to exposure to UV radiation. When humans became mostly hairless, the ability of the skin to produce melanin assumed new importance. Melanin is nature's sunscreen: it is a large organic molecule that serves the dual purpose of physically and chemically filtering the harmful effects of UV radiation; it absorbs UV rays, causing them to lose energy, and it neutralizes harmful chemicals called free radicals that form in the skin after damage by UV radiation.

Anthropologists and biologists have generally reasoned that high concentrations of melanin arose in the skin of peoples in tropical areas because it protected them against skin cancer. James E. Cleaver of the University of California at San Francisco, for instance, has shown that people with the disease xeroderma pigmentosum, in which melanocytes are destroyed by exposure to the sun, suffer from significantly higher than normal rates of squamous and basal cell carcinomas, which are usually easily treated. Malignant melanomas are more frequently fatal, but they are rare (representing 4 percent of skin cancer diagnoses) and tend to strike only light-skinned people. But all skin cancers typically arise later in life, in most cases after the first reproductive years, so they could not have exerted enough evolutionary pressure for skin protection alone to account for darker skin colors. Accordingly, we began to ask what role melanin might play in human evolution.

Overview/*Skin Color Evolution*

- After losing their hair as an adaptation for keeping cool, early hominids gained pigmented skins. Scientists initially thought that such pigmentation arose to protect against skin-cancer-causing ultraviolet (UV) radiation.
- Skin cancers tend to arise after reproductive age, however. An alternative theory suggests that dark skin might have evolved primarily to protect against the breakdown of folate, a nutrient essential for fertility and for fetal development.
- Skin that is too dark blocks the sunlight necessary for catalyzing the production of vitamin D, which is crucial for maternal and fetal bones. Accordingly, humans have evolved to be light enough to make sufficient vitamin D yet dark enough to protect their stores of folate.
- As a result of recent human migrations, many people now live in areas that receive more (or less) UV radiation than is appropriate for their skin color.

The Folate Connection

IN 1991 ONE OF US (Jablonski) ran across what turned out to be a critical paper published in 1978 by Richard F. Branda and John W. Eaton, now at the University of Vermont and the University of Louisville, respectively. These investigators showed that light-skinned people who had been exposed to simulated strong sunlight had abnormally low levels of the essential B vitamin folate in their blood. The scientists also observed that subjecting human blood serum to the same conditions resulted in a 50-percent loss of folate content within one hour.

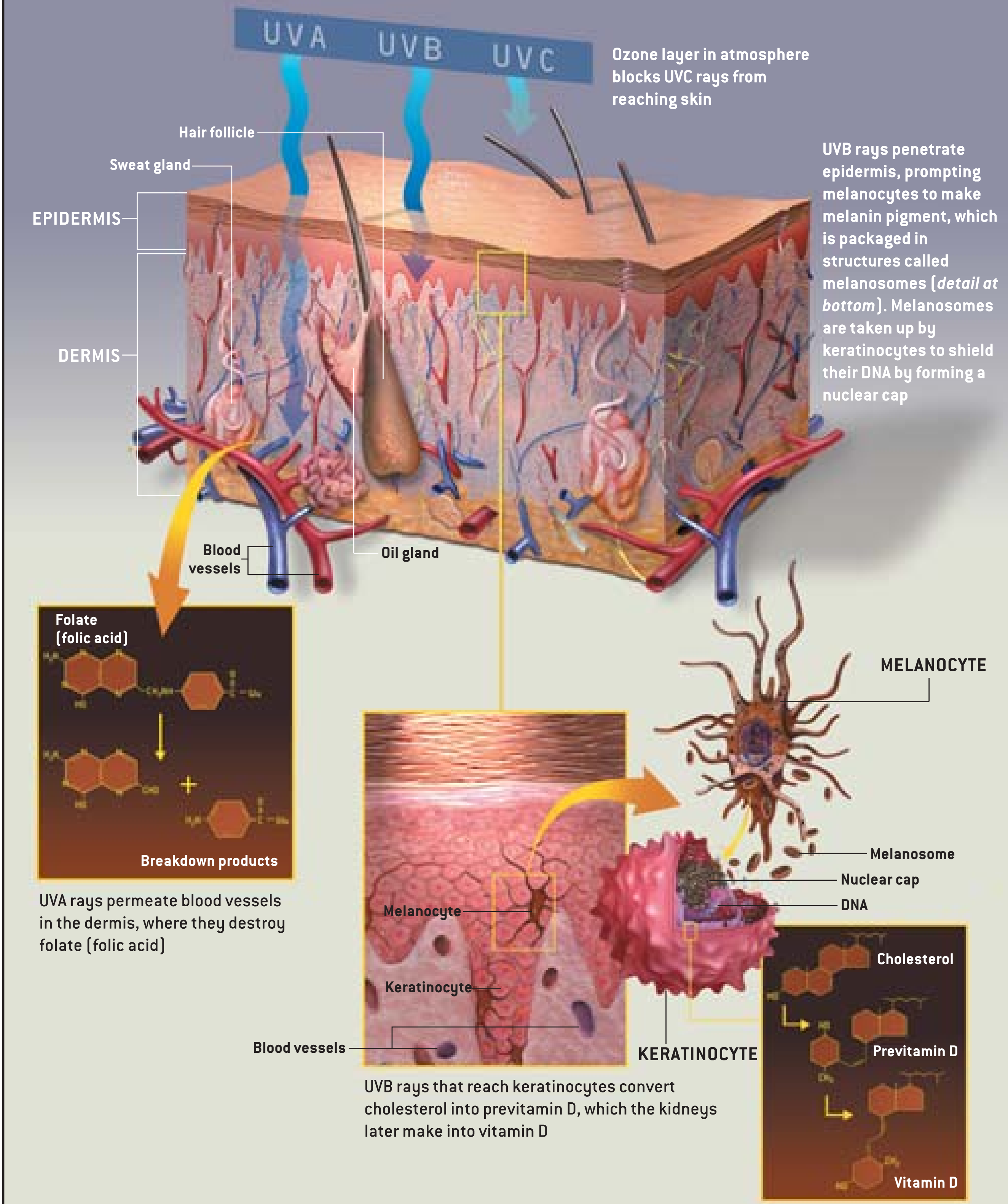
The significance of these findings to reproduction—and hence evolution—became clear when we learned of research being conducted on a major class of birth defects by our colleagues at the University of Western Australia. There Fiona J. Stanley and Carol Bower had established by the late 1980s that folate deficiency in pregnant women is related to an increased risk of neural tube defects such as spina bifida, in which the arches of the spinal vertebrae fail to close around the spinal cord. Many research groups throughout the world have since confirmed this correlation, and efforts to supplement foods with folate and to educate women about the importance of the nutrient have become widespread.

We discovered soon afterward that folate is important not only in preventing neural tube defects but also in a host of other processes. Because folate is essential for the synthesis of DNA in dividing cells, anything that involves rapid cell proliferation, such as spermatogenesis (the production of sperm cells), requires folate. Male rats and mice with chemically induced folate deficiency have impaired spermatogenesis and are infertile. Although no comparable studies of humans have been conducted, Wai Yee Wong and his colleagues at the University Medical Center of Nijmegen in the Netherlands have recently reported that folic acid treatment can boost the sperm counts of men with fertility problems.

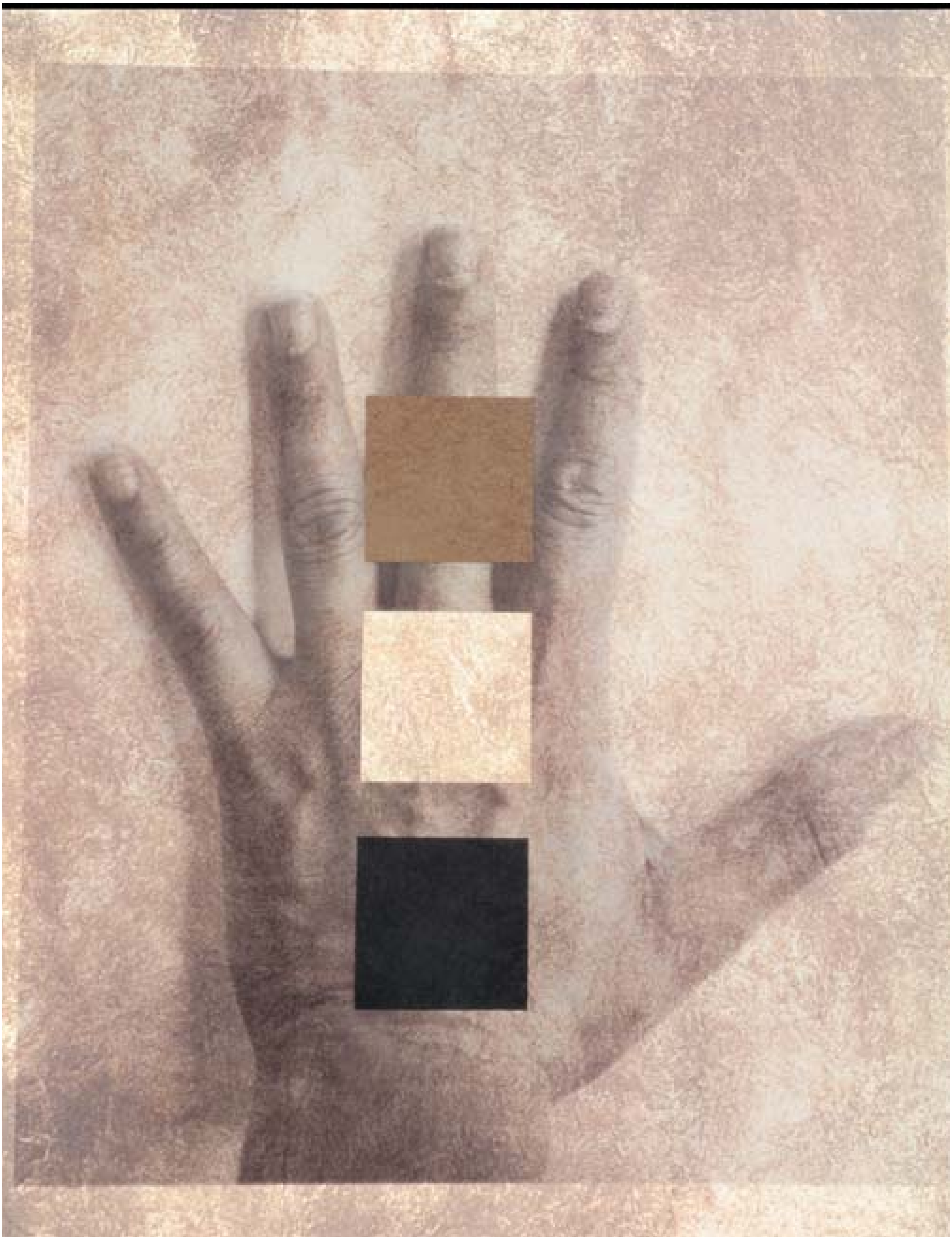
IRAIDA ICAZA (*preceding pages*)

SKIN IN THE SUN

THE ULTRAVIOLET (UV) RAYS of the sun are a mixed blessing: they spur the production of vitamin D but destroy folate and can cause cancer by damaging DNA. Melanin pigment produced by melanocytes protects against DNA damage and folate breakdown. But keratinocytes must get enough UV rays to make vitamin D. *—N.G.J. and G.C.*



KEITH KASNOT



IRAIDA ICAZA

Such observations led us to hypothesize that dark skin evolved to protect the body's folate stores from destruction. Our idea was supported by a report published in 1996 by Argentine pediatrician Pablo Lapunzina, who found that three young and otherwise healthy women whom he had attended gave birth to infants with neural tube defects after using sun beds to tan themselves in the early weeks of pregnancy. Our evidence about the breakdown of folate by UV radiation thus supplements what is already known about the harmful (skin-cancer-causing) effects of UV radiation on DNA.

Human Skin on the Move

THE EARLIEST MEMBERS of *Homo sapiens,* or modern humans, evolved in Africa between 120,000 and 100,000 years ago and had darkly pigmented skin adapted to the conditions of UV radiation and heat that existed near the equator. As modern humans began to venture out of the tropics, however, they encountered environments in which they received significantly less UV radiation during the year. Under these conditions their high concentrations of natural sunscreen probably proved detrimental. Dark skin contains so much melanin that very little UV radiation, and specifically very little of the shorter-wavelength UVB radiation, can penetrate the skin. Although most of the effects of UVB are harmful, the rays perform one indispensable function: initiating the formation of vitamin D in the skin. Dark-skinned people living in the tropics generally receive sufficient UV radiation during the year for UVB to penetrate the skin and allow them to make vitamin D. Outside the tropics this is not the case. The solution, across evolutionary time, has been for migrants to northern latitudes to lose skin pigmentation.

The connection between the evolution of lightly pigmented skin and vitamin D synthesis was elaborated by W. Farnsworth Loomis of Brandeis University in 1967. He established the importance of vitamin D to reproductive success because of its role in enabling calcium absorption by the intestines, which in turn makes possible the normal development of the skeleton and the maintenance of a healthy immune system. Research led by Michael Holick of the Boston University School of Medicine has, over the past 20 years, further cemented the significance of vitamin D in development and immunity. His team also showed that not all sunlight contains enough UVB to stimulate vitamin D production. In Boston, for instance, which is located at about 42 degrees north latitude, human skin cells begin to produce vitamin D only after mid-March. In the wintertime there isn't enough UVB to do the job. We realized that this was another piece of evidence essential to the skin color story.

During the course of our research in the early 1990s, we searched in vain to find sources of data on actual UV radiation levels at the earth's surface. We were rewarded in 1996, when we contacted Elizabeth Weatherhead of the Cooperative Institute for Research in Environmental Sciences at the University of Colorado at Boulder. She shared with us a database of measurements of UV radiation at the earth's surface taken by NASA's Total Ozone Mapping Spectrophotometer satellite between 1978 and 1993. We were then able to model the distribution of UV radiation on the earth and relate the satellite data to the amount of UVB necessary to produce vitamin D.

We found that the earth's surface could be divided into three vitamin D zones: one comprising the tropics, one the subtropics and temperate regions, and the last the circumpolar regions north and south of about 45 degrees latitude. In the first, the dosage of UVB throughout the year is high enough that humans have ample opportunity to synthesize vitamin D all year. In the second, at least one month during the year has insufficient UVB radiation, and in the third area not enough UVB arrives on average during the entire year to prompt vitamin D synthesis. This distribution could explain why indigenous peoples in the tropics generally have dark skin, whereas people in the subtropics and temperate regions are lighter-skinned but have the ability to tan, and those who live in regions near the poles tend to be very light skinned and burn easily.

One of the most interesting aspects of this investigation was the examination of groups that did not precisely fit the predicted skin-color pattern. An example is the Inuit people of Alaska and northern Canada. The Inuit exhibit skin color that is somewhat darker than would be predicted given the UV levels at their latitude. This is probably caused by two factors. The first is that they are relatively recent inhabitants of these climes, having migrated to North America only roughly 5,000 years ago. The second is that the traditional diet of the Inuit is extremely high in foods containing vitamin D, especially fish and marine mammals. This vitamin D–rich diet offsets the problem that they would otherwise have with vitamin D synthesis in their skin at northern latitudes and permits them to remain more darkly pigmented.

Our analysis of the potential to synthesize vitamin D allowed us to understand another trait related to human skin color: women in all populations are generally lighter-skinned than men. (Our data show that women tend to be between 3 and 4 percent lighter than men.) Scientists have often speculated on the reasons, and most have argued that the phenomenon stems from sexual selection—the preference of men for women of lighter color. We contend that although this is probably part of the story, it is not the original reason for the sexual difference. Females have significantly greater needs for calcium throughout their reproductive lives, especially during pregnancy and lactation, and must be able to make the most of the calcium contained in food. We propose, therefore, that women

THE AUTHORS

NINA G. JABLONSKI and *GEORGE CHAPLIN* work at the California Academy of Sciences in San Francisco, where Jablonski is Irvine Chair and curator of anthropology and Chaplin is a research associate in the department of anthropology. Jablonski's research centers on the evolutionary adaptations of monkeys, apes and humans. She is particularly interested in how primates have responded to changes over time in the global environment. Chaplin is a private geographic information systems consultant who specializes in describing and analyzing geographic trends in biodiversity. In 2001 he was awarded the Student of the Year prize by the Association of Geographic Information in London for his master's thesis on the environmental correlates of skin color.

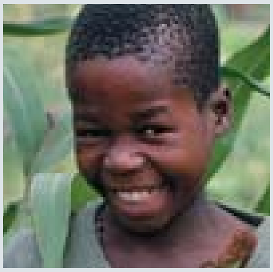
SKIN AND MIGRATION

THE SKIN OF PEOPLES who have inhabited particular areas for millennia has adapted to allow vitamin D production while protecting folate stores. The skin tones of more recent immigrants will take thousands of years to catch up, putting light-skinned individuals at risk for skin cancer and dark-skinned people in danger of vitamin D deficiency. *—N.G.J. and G.C.*

LONG-TERM RESIDENT	RECENT IMMIGRANT

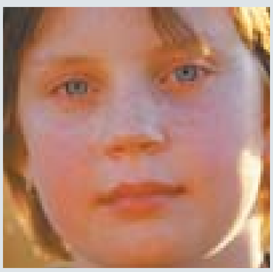
SOUTHERN AFRICA: ~20–30° S


Khoisan (Hottentot)


Zulu: arrived about 1,000 years ago

AUSTRALIA: ~10–35° S


Aborigine


European: ~300 years ago

BANKS OF RED SEA: ~15–30° N


Sudanese


Arab: ~2,000 years ago

INDIA: ~10–30° N


Bengali


Tamil: ~100 years ago

tend to be lighter-skinned than men to allow slightly more UVB rays to penetrate their skin and thereby increase their ability to produce vitamin D. In areas of the world that receive a large amount of UV radiation, women are indeed at the knife's edge of natural selection, needing to maximize the photoprotective function of their skin on the one hand and the ability to synthesize vitamin D on the other.

Where Culture and Biology Meet

AS MODERN HUMANS MOVED throughout the Old World about 100,000 years ago, their skin adapted to the environmental conditions that prevailed in different regions. The skin color of the indigenous people of Africa has had the longest time to adapt because anatomically modern humans first evolved there. The skin-color changes that modern humans underwent as they moved from one continent to another—first Asia, then Austro-Melanesia, then Europe and, finally, the Americas—can be reconstructed to some extent. It is important to remember, however, that those humans had clothing and shelter to help protect them from the elements. In some places, they also had the ability to harvest foods that were extraordinarily rich in vitamin D, as in the case of the Inuit. These two factors had profound effects on the tempo and degree of skin-color evolution in human populations.

Africa is an environmentally heterogeneous continent. A number of the earliest movements of contemporary humans outside equatorial Africa were into southern Africa. The descendants of some of these early colonizers, the Khoisan (previously known as Hottentots), are still found in southern Africa and have significantly lighter skin than indigenous equatorial Africans do—a clear adaptation to the lower levels of UV radiation that prevail at the southern extremity of the continent.

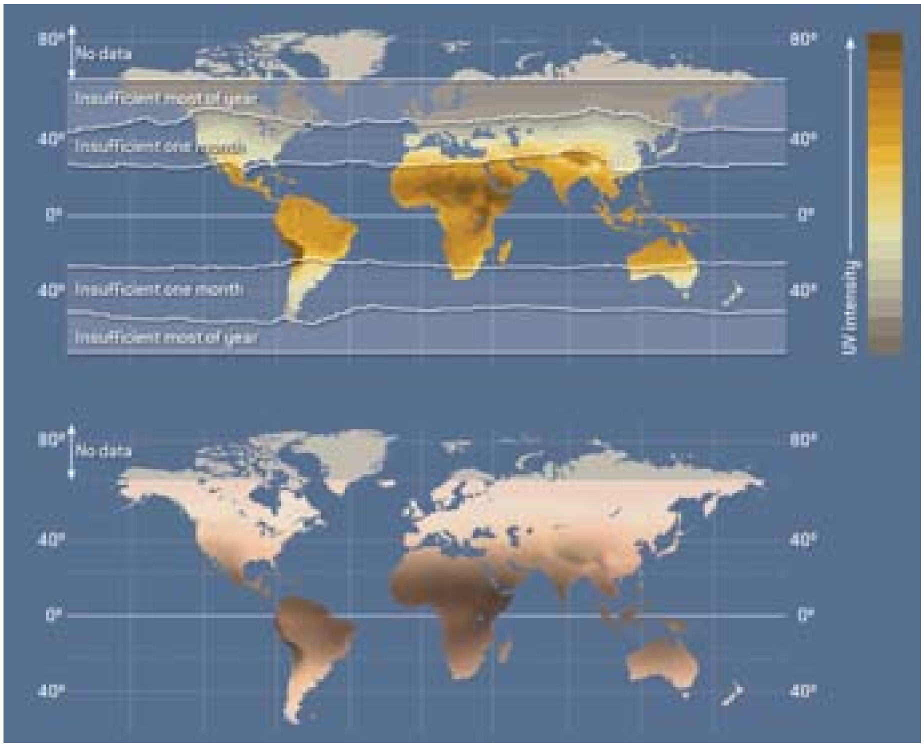
Interestingly, however, human skin color in southern Africa is not uniform. Populations of Bantu-language speakers who live in southern Africa today are far darker than the Khoisan. We know from the history of this region that Bantu speakers migrated into this region recently—probably within the past 1,000 years—from parts of West Africa near the equator. The skin-color difference between the Khoisan and Bantu speakers such as the Zulu indicates that the length of time that a group has inhabited a particular region is important in understanding why they have the color they do.

Cultural behaviors have probably also strongly influenced the evolution of skin color in recent human history. This effect can be seen in the indigenous peoples who live on the eastern and western banks of the Red Sea. The tribes on the western side, which speak so-called Nilo-Hamitic languages, are thought to have inhabited this region for as long as 6,000 years. These individuals are distinguished by very darkly pigmented skin and long, thin bodies with long limbs, which are excellent biological adaptations for dissipating heat and intense UV radiation. In contrast, modern agricultural and pastoral groups on the eastern bank of the Red Sea, on the Arabian Peninsula, have lived there for only about 2,000 years. These earliest Arab people, of European origin, have adapted to very similar environmental condi-

PETER JOHNSON *Corbis* (*Khoisan*); BARBARA BANNISTER *Gallo Images/Corbis* (*Zulu*); PENNY TWEEDIE *Corbis* (*Aborigine*); DAVID McLAIN *Aurora* (*European*); ERIC WHEATER *Lonely Planet Images* (*Sudanese*); WAYNE EASTEP *Getty Images* (*Arab*); ROGER WOOD *Corbis* (*Bengali*); JEREMY HORNER *Corbis* (*Tamil*)

WHO MAKES ENOUGH VITAMIN D?

POPULATIONS THAT LIVE in the tropics receive enough ultraviolet (UV) light from the sun (*top map, brown and orange*) to synthesize vitamin D all year long. But those that live at northern or southern latitudes do not. In the temperate zones (*light-shaded band*), people lack sufficient UV light to make vitamin D one month of the year; those nearer the poles (*dark-shaded band*) do not get enough UV light most months for vitamin D synthesis. The bottom map shows predicted skin colors for humans based on UV light levels. In the Old World, the skin color of indigenous peoples closely matches predictions. In the New World, however, the skin color of long-term residents is generally lighter than expected—probably because of their recent migration and factors such as diet. —*N.G.J. and G.C.*



tions by almost exclusively cultural means—wearing heavy protective clothing and devising portable shade in the form of tents. (Without such clothing, one would have expected their skin to have begun to darken.) Generally speaking, the more recently a group has migrated into an area, the more extensive its cultural, as opposed to biological, adaptations to the area will be.

Perils of Recent Migrations

DESPITE GREAT IMPROVEMENTS in overall human health in the past century, some diseases have appeared or reemerged in populations that had previously been little affected by them. One of these is skin cancer, especially basal and squamous cell carcinomas, among light-skinned peoples. Another is rickets, brought about by severe vitamin D deficiency, in dark-skinned peoples. Why are we seeing these conditions?

As people move from an area with one pattern of UV radiation to another region, biological and cultural adaptations have not been able to keep pace. The light-skinned people of northern European origin who bask in the sun of Florida or northern Australia increasingly pay the price in the form of premature aging of the skin and skin cancers, not to mention the unknown cost in human life of folate depletion. Conversely, a number of dark-skinned people of southern Asian and African origin now living in the northern U.K., northern Europe or the northeastern U.S. suffer from a lack of UV radiation and vitamin D, an insidious problem that manifests itself in high rates of rickets and other diseases related to vitamin D deficiency.

The ability of skin color to adapt over long periods to the various environments to which humans have moved reflects the importance of skin color to our survival. But its unstable nature also makes it one of the least useful characteristics in determining the evolutionary relations between human groups. Early Western scientists used skin color improperly to delineate human races, but the beauty of science is that it can and does correct itself. Our current knowledge of the evolution of human skin indicates that variations in skin color, like most of our physical attributes, can be explained by adaptation to the environment through natural selection. We look ahead to the day when the vestiges of old scientific mistakes will be erased and replaced by a better understanding of human origins and diversity. Our variation in skin color should be celebrated as one of the most visible manifestations of our evolution as a species. SA

MORE TO EXPLORE

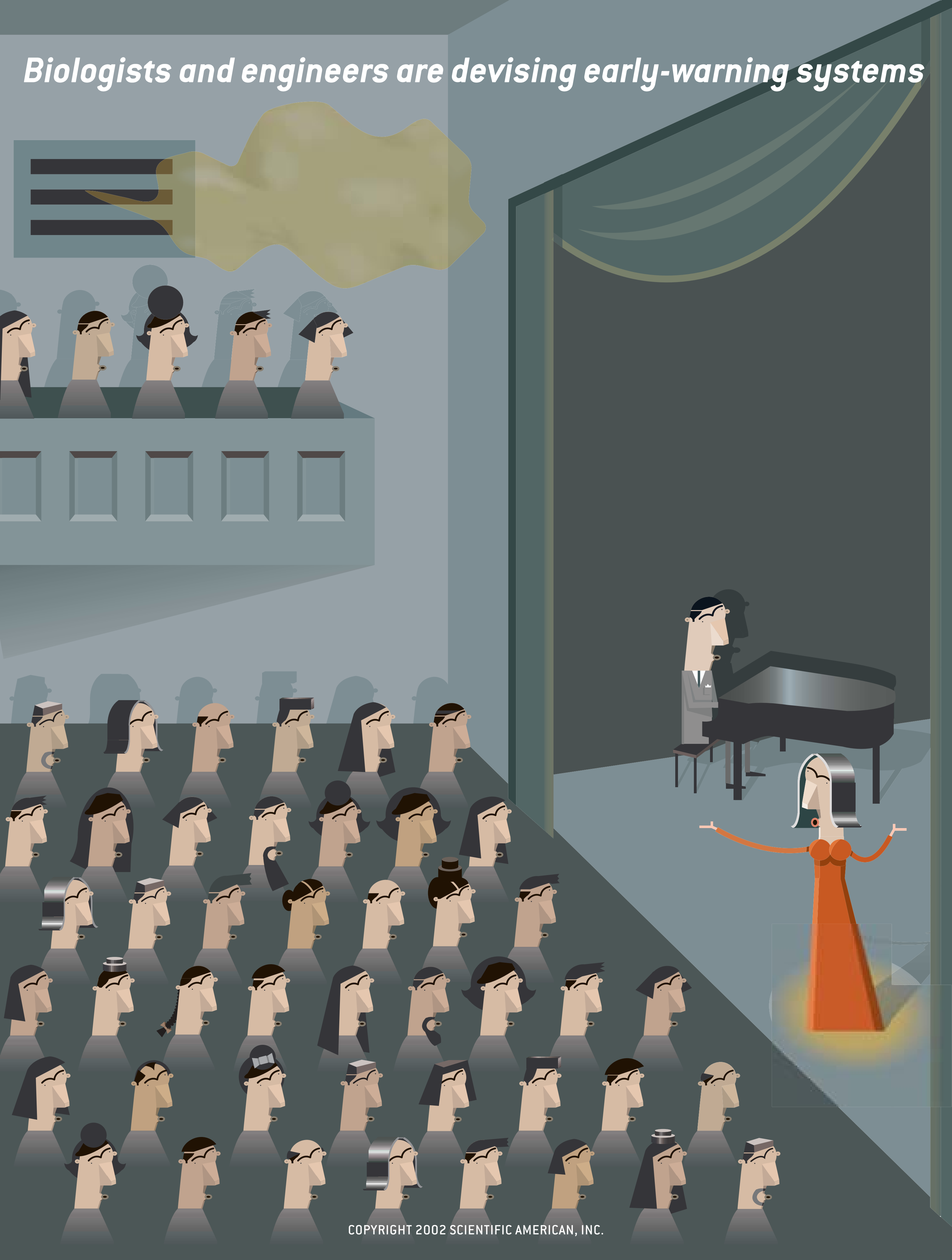
The Evolution of Human Skin Coloration. Nina G. Jablonski and George Chaplin in *Journal of Human Evolution,* Vol. 39, No. 1, pages 57–106; July 1, 2000. An abstract of the article is available online at **www.idealibrary.com/links/doi/10.1006/jhev.2000.0403**

Why Skin Comes in Colors. Blake Edgar in *California Wild,* Vol. 53, No. 1, pages 6–7; Winter 2000. The article is also available at **www.calacademy.org/calwild/winter2000/html/horizons.html**

The Biology of Skin Color: Black and White. Gina Kirchweger in *Discover,* Vol. 22, No. 2, pages 32–33; February 2001. The article is also available at **www.discover.com/feb_01/featbiology.html**

SARA CHEN

Biologists and engineers are devising early-warning systems



that can detect a bioterrorist attack in time to blunt its effects

TECHNOLOGY AGAINST Terror

By Rocco Casagrande

In May 2000 high-ranking members of the U.S. government watched as a cloud of bacteria wafted through the Denver Center for the Performing Arts, a complex of seven theaters that seats a total of 7,000. One week later thousands of people were dead or dying from the plague, the state borders of Colorado were closed, food and medical supplies began to run short, and medical care was all but shut down as doctors and nurses fell ill and antibiotics were used up. Luckily, this scenario wasn't real; it was a computerized exercise to simulate the effects of a biological attack against a target in the U.S. Part of a test named TopOff, it served as a wake-up call to civic leaders that they can't wait for sick people to start showing up at emergency rooms if they hope to wage an effective defense against biological weapons. Scientists are now devising a range of early-warning systems to alert government officials to an attack as it is happening. These technologies include DNA- and antibody-based biochips as well as "electronic noses" that can sniff out deadly microbes.

Are We Under Attack?

BIOLOGICAL WARFARE is insidious. Airborne clouds of bacterial or viral agents are nearly invisible and odorless; people who inhale the agents would not know they had been attacked until they fell ill days later. By that time, it might be too late to treat those victims or to protect others from infection. Although most biological agents are not very contagious, in many instances the unknowingly infected could pass on the disease.

Fortunately, the incubation period of biological agents provides a window of time in which public health officials could quarantine and treat victims and vaccinate others. Prior to the onset of symptoms, many diseases caused by biological agents are treatable with antibiotics; after symptoms appear, some victims will be beyond treatment.

Early detection is particularly important because many of the diseases caused by biological warfare agents trigger initial symptoms, such as fever and nausea, that could easily be mistaken for the flu. Medical students are generally taught the phrase "When you hear hoofbeats, think horses—not zebras," as a way to remind them to rule out common disorders before considering more exotic diagnoses. Although this dictum saves time and effort in everyday situations, it could lead doctors to initially overlook a biological attack. For this reason, some biological detectors are colloquially referred to as zebra chips, or Z chips, because they can tip off physicians that a metaphorical zebra is on the loose.

Biological warfare can be waged by contaminating food or water supplies or via disease-carrying insects such as mosquitoes, but these methods are unlikely to affect thousands of vic-

FRED RIX

tims during a single attack. Biological weapons reach the level of weapons of mass destruction—with a potential for human casualties rivaling that of nuclear weapons—only when they are disseminated through the air as a breathable aerosol of particles about one millionth of a meter in size. These tiny droplets can float through the air for long distances and become lodged deep within the lungs to cause dangerous systemic infections.

Airborne biological agents are tough to detect, however, because of their variety: they can come in the form of bacteria, viruses or nonliving toxins produced by microbes. Biological agents can be deadly, even when extremely dilute. A healthy person breathes in roughly six liters of air a minute, and certain pathogens can cause disease when as few as 10 organisms are inhaled. To protect people who are present for short periods in a contaminated area, a device would have to pick up two individual pathogens per liter of air—an extremely daunting task.

The first successful biological detectors merely spotted clouds of small particles. Some of these machines—such as the U.S. Army's XM2, which was deployed during the Gulf War—sample the air around them and are linked to machines that count particles of the appropriate size for bioweapons. If that particle count exceeds a certain threshold, an alarm sounds to notify troops to evacuate the area. Other particle detectors employ lidar, a system much like radar that emits a laser beam and then detects the light that bounces back from objects in its path. In dry conditions, lidar-based devices function from a distance of 50 kilometers, but they cannot distinguish between mists of biological agents and clouds of fine dust or smoke.

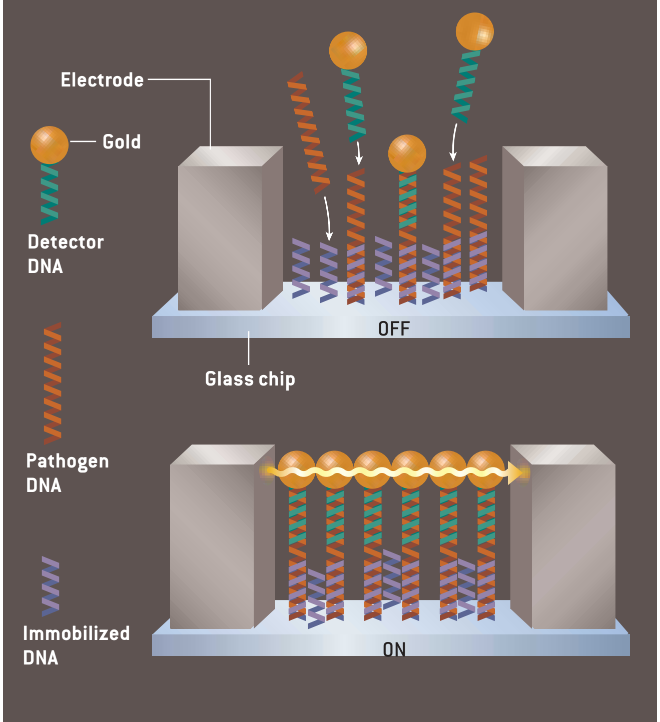
Newer lidar systems take advantage of certain molecules in almost all living cells that fluoresce when excited by ultraviolet (UV) light. These UV-lidar devices excite particles with a UV beam and watch for fluorescence emitted from the cloud. But even UV-lidar cannot discern pathogens from clouds of harmless microorganisms or pollen or from plumes of mold spores. Despite their shortcomings, particle detectors are useful for guiding troops away from areas that might pose a danger from a biological aerosol. They might also be deployed to indicate when more sensitive detectors should begin to analyze samples [*see box on next page*].

Overview/Biodetectors

- Biowarfare agents are colorless and odorless and could take days to cause symptoms. Accordingly, a society might not know it was under attack until it was too late to respond.
- Biologists and engineers are developing detectors consisting of chips that detect pathogens using antibodies or DNA. They are also coming up with devices that "sniff out" odors emitted by microbes or the additives used to make them into weapons.
- Public officials must decide how best to deploy the new bioweapons detectors; it would be impractical, for example, to put them on every street corner.

A Needle in a Haystack

SOME OF THE NEWEST biological weapons detectors can distinguish pathogens from benign microorganisms or other particles based on differences in their genetic makeup. Because DNA is located within microbes, the cells must first be broken open so their DNA can be extracted. Some devices, such as the GeneXpert system made by Cepheid in Sunnyvale, Calif., have built-in cell disrupters; others require a technician to isolate the DNA.



DNA-BASED BIOCHIP detects DNA from a pathogen by making a bridge of gold particles linking two electrodes. The bridge completes an electrical circuit, raising an alarm.

One of the first DNA chips, which was developed at Northwestern University, relies on the complementary nature of the two strands that constitute the DNA double helix. The DNA helix is like a twisted ladder in which each rung is composed of two subunits called bases. The ladder splits down the middle when genes are turned on or when a cell is copying its genes before dividing. Four bases make up the rungs of the helix: adenine (A), thymine (T), cytosine (C) and guanine (G). A always binds to T, and C always pairs with G. Knowing the sequence of bases from one strand—ATCGCC, for example—one can predict the complementary sequence of the other strand, in this case TAGCGG.

The sensing element of the Northwestern University system contains single strands of DNA that are complementary to a short sequence of DNA that is specific to a given pathogen. The strands are immobilized on a glass chip between two electrodes. When DNA from that particular pathogen enters the system, it sticks to, or hybridizes with, one end of the immobilized DNA. To detect this hybridization, a technician adds pieces of DNA

FRED RIX

Deploying the Defense

Deciding when—and where—to use biological warfare detectors may be the hardest part

A BIOLOGICAL ATTACK could occur anytime, anywhere. Madmen bent on killing as many people as possible could just as easily release a cloud of pathogens at a rural state fair as they could unleash a biological agent in an urban subway train during rush hour. (In the former case, however, they would have to pick an overcast day: bright sunlight kills most microbes.)

Since September 11, 2001, the fear of a terrorist attack has pervaded the thinking of events planners from New York City to Punxsutawney, Pa., which beefed up security this past Groundhog Day to foil possible attacks. Although the diversity and abundance of potential targets will make it impossible or impractical to protect all of them completely, properly deploying biological detectors could reduce the likelihood that the worst attacks would succeed.

Currently, biological agent detectors are too expensive and require too much maintenance to be placed on every street corner. Common sense dictates, however, that certain events or locations deserve tighter security because of their importance or the large number of potential victims there. The Capitol building and the Pentagon are at the heart of U.S. democracy and power and therefore deserve around-the-clock biological surveillance. Eventually technology may progress to the point where biological agent detectors will be reliable, cheap and self-sufficient enough to guard the municipal buildings of every major city.

Unfortunately, no biological agent detector available now can both distinguish harmful organisms from benign ones and monitor its surroundings constantly for pathogens. Some devices cannot collect samples automatically and require a human operator. Others can take samples mechanically, by sampling air or water, but are able to do so only when directed by an operator, who must acquire the samples at intervals to allow adequate time for analysis. Operators could take samples at set times—such as every hour or as soon as the previous sample was analyzed—but a cloud of biological agents could pass over an area or be dispersed in a matter of minutes. Taking samples at the wrong time could miss an attack.

U.S. ARMY'S Long Range Biological Standoff Detection System operates on board a Black Hawk helicopter. Such lidar-based particle detectors can be coupled to biochips that can distinguish particles of harmless microbes from those containing bioweapons.

LOS ALAMOS NATIONAL LABORATORY

Some detectors are linked to lidar systems or particle counters and collect samples only when a cloud of particles of the right size is present. Similar systems could be used to test the water that flows in the water mains that supply sensitive buildings: if the particle count in a water main spiked, the device would divert a sample of water for further analysis.

A system that could monitor its environment continuously, analyze samples rapidly and operate at low cost would be ideal, but such a system has yet to be perfected. In the meantime, epidemiologists and computer scientists have collaborated to create a database for monitoring the symptoms of patients who visit emergency rooms to detect the earliest signs of a biological attack. The protocol relies on handheld devices that physicians use to upload symptom information to a database that can pick up patterns of illness—such as an unusual number of flu-like illnesses outside flu season—that are consistent with the leading edge of a biological attack.

One such system—known as the Lightweight Epidemiology Advanced Detection and Emergency Response System (LEADERS)—has been used since 2000 to mine hospital databases in areas near major events such as political party conventions, the Super Bowl and the World Series. Mindful of the fact that most people do not go to the hospital when they think they have the flu, programmers are adding metrics into the LEADERS database such as sales of over-the-counter medications, sick-day tallies and tollbooth receipts (sick people are less likely to drive). Ideally, these databases would log patient information from all over the nation constantly so that attacks in the country could be detected early, no matter where, when or how they occur [see "The Vigilance Defense," on page 88]. *—R.C.*

that have gold particles tethered to them and that are complementary to the other end of the target DNA sequence. Where the gold-bearing DNA sequences bind, they complete an electrical circuit between the electrodes and raise an alarm.

Several other DNA-based detectors rely on the fact that specific sequences of DNA can be amplified through a process called the polymerase chain reaction (PCR). In this technique, scientists heat DNA so that the bonds between the two bases that make up each rung break and the two strands of DNA separate. Then they cool the solution and add two short pieces of DNA, called primers, that are designed to hybridize specifically to either end of the DNA sequence they are attempting to detect. Enzymes latch on to the primers and extend them, copying the initial two strands of DNA into four. Scientists can double the number of copies of the target DNA sequence each time they repeat the cycle until there is enough to detect.

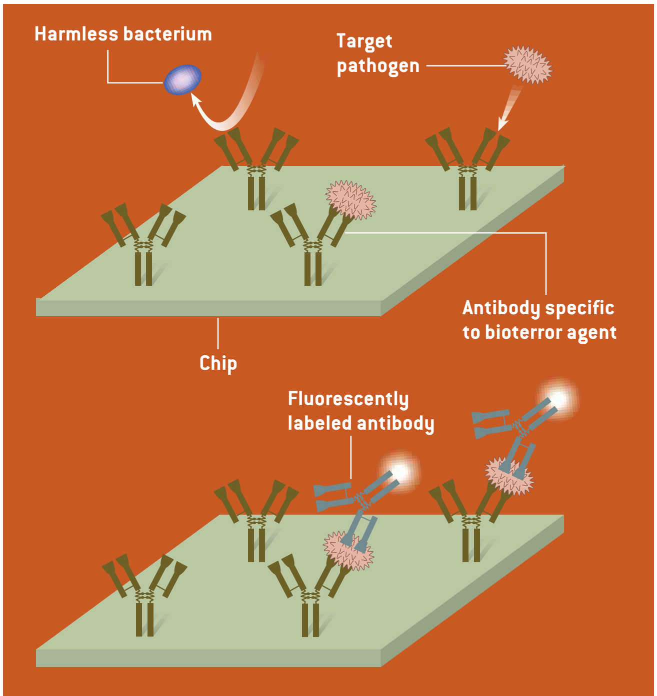
By incorporating fluorescent molecules in the newly synthesized pieces of DNA, researchers can monitor the amplification process as it progresses. Also available are machines called rapid thermal cyclers that complete each round of heating and cooling in less than a minute, allowing 30 doublings of very scarce DNA sequences to be accomplished within half an hour.

But the circuit- and PCR-based systems must be preloaded with reagents that are specific to particular pathogens, which means that users must guess correctly beforehand which pathogens a potential terrorist might choose. To circumvent this shortcoming, scientists at Ibis Therapeutics in Carlsbad, Calif., and Science Applications International Corporation in San Diego have developed a system called Triangulation Identification Genetic Evaluation of biological Risks (TIGER). Like many other DNA chips, TIGER amplifies target DNA using PCR. It is different, though, in that it employs primers that hybridize to a segment of DNA involved in controlling protein synthesis that is part of the basic machinery of all cells. TIGER can still be exquisitely sensitive because the sections between the primers are so highly variable that almost every microorganism has a unique sequence. Technicians can then analyze those amplified sequences using a mass spectrometer and compare them against a database of patterns from all known microorganisms to identify the agent.

DNA-based devices have their limitations, however. They cannot detect toxins, which have no DNA, and their half-hour reaction time makes them too slow to be used to alert people of an oncoming cloud of biowarfare pathogens in time for them to evacuate.

Canary in a Coal Mine

CHIPS BASED ON ANTIBODIES—Y-shaped molecules produced by the immune system that bind to specific target molecules on invaders—can surmount these hurdles. Because antibodies can detect molecules on the surfaces of microbes, no additional time is needed to break open the target cells. They can also ferret out toxins as well as whole microorganisms.



ANTIBODY-BASED CHIPS specifically bind pathogens, which are detected using other antibodies with fluorescent tails.

Antibodies are the heart of a biowarfare detection system designed by the U.S. Naval Research Laboratory called Raptor. The system is a so-called sandwich assay: target pathogens stick to antibodies on the chip and are detected when they are sandwiched between another layer of antibodies labeled with fluorescent dyes. Raptor can pick up on several pathogens at once because it can accommodate spots of several types of antibodies, each specific to a different bioweapon. The Origen system from Igen in Gaithersburg, Md., also detects pathogens using a sandwich assay, but instead of fluorescent dyes it relies on a compound that emits a burst of light when exposed to an electric field. The light is brighter than normal fluorescence, allowing the analysis of samples that contain only a few pathogens. Furthermore, one of the antibodies is tethered to a surface that allows the target pathogens to be concentrated before detection.

At Surface Logix, we have collaborated with Radiation Monitoring Devices in Watertown, Mass., to develop technology that can be used to pick up pathogens on a continuous basis. The machine, which can be connected to an air sampler,

FRED RIX

THE AUTHOR

ROCCO CASAGRANDE is a scientist at the biotechnology company Surface Logix in Brighton, Mass., where he designs and tests biological agent detectors. He received his Ph.D. from the Massachusetts Institute of Technology in 2001. Casagrande has consulted for the U.S. Defense Threat Reduction Agency and is a member of the Controlling Dangerous Pathogens Group at the University of Maryland. He is also on the roster of inspectors of the United Nations Monitoring, Verification and Inspection Commission, which is charged with ensuring Iraq's biological and chemical weapons disarmament.

mixes any particles present in the sampled air with a solution of microscopic magnetic beads. Each bead is coated with fluorescently labeled antibodies that bind to a particular microbe.

The sample stream containing the beads flows down a microscopic channel about the width of a human hair, where it meets a clean stream free of microorganisms. The clean stream and the sample stream flow side by side without mixing until they hit a fork in the channel. A magnet placed just before the fork pulls the magnetic beads—and any pathogens bound to them—into the clean stream. That stream then flows into a detector that registers the presence of pathogens by their fluorescence.

A major advantage of our system is that it removes target pathogens from the thousands of harmless organisms present in a given sample. Smoke and other environmental contaminants in the sample do not affect detection, because the microbeads are pulled into the clean stream before the fluorescence-detection step. In addition, the machine accepts samples from the environment constantly and analyzes them in real time.

Other systems use antibodies to capture passing pathogens onto vibrating devices such as quartz crystals, thin membranes or microscopic cantilevers. As these devices capture pathogens, they become weighed down, which slows their vibrations. This change is detectable by electronics.

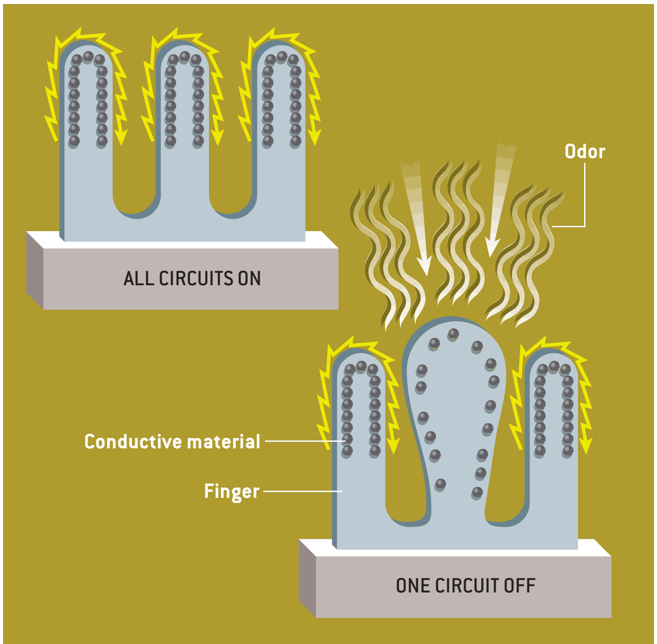
Sniffing Out Invaders

THE DEVICES MENTIONED in this article are either available now or expected to be within the next few years. But newer, more powerful technologies are being developed all the time.

So-called electronic noses—which are currently used to spot explosives and chemical weapons—are being adapted to sniff out biological bombs. One such device, the Cyranose, made by Cyrano Sciences in Pasadena, Calif., contains an array of pegs made of slightly different polymers. Each polymer peg has a specific capacity for absorbing a particular chemical, which causes it to swell. The pegs contain flecks of material that conduct electricity. When the pegs are not swollen, the flecks are close enough to one another to conduct electricity; as they swell, the flecks become separated, breaking the circuit and yielding a positive signal. The pattern of broken circuits is different for each odor. Researchers are devising noses that can pick up metabolites given off by dangerous bacteria or by chemicals such as stabilizers that are often part of biological weapons. The hope is to find a pattern unique to a biological agent.

In a highly innovative approach, BCR Diagnostics in Jamestown, R.I., uses dormant forms of bacteria called spores to tip off the presence of bioweapons. When bacteria enter the detector, their normal metabolism cleaves an inactive nutrient into an active one, allowing the detector's spores to germinate. Because the spores have been genetically modified to emit light as they undergo this transformation, the detector registers the presence of the pathogen. Unfortunately, this device detects the presence of all bacteria, whether harmful or not. But it could be adapted by specially designing inactive nutrients that would be converted to active ones only by the types of bacteria most likely to be used in bioweapons.

A clever terrorist could fool even the best biodetector, however, by genetically engineering an otherwise harmless organism to produce deadly toxins. The ideal detector would respond to the presence of a biological weapon exactly as its target would, but more quickly. To this end, the Defense Advanced Research Projects Agency is currently supporting biodetector research involving living cells from humans, animals or plants. The idea behind this kind of detector is that a human pathogen must be harmful to at least one type of human cell; measuring the extent of cell death in the detector would indicate the presence of a harmful organism in the environment.



ELECTRONIC NOSE has polymer fingers that swell when they sense a particular odor. The swelling separates bits of conductive material lodged within the fingers and breaks an electrical circuit, triggering an alarm.

Although biological weapons are horrifying, no country or terrorist group has yet wielded them to kill thousands of people. Biodetectors can in principle help protect populations against such an unlikely event, but they can also fill other roles. Pathogen detectors can be deployed to identify contaminated food or to diagnose infectious diseases in a doctor's office. Devices that measure cellular responses can also be used evaluate the susceptibility of cancer cells to various drugs, allowing scientists to identify potential therapeutics more rapidly. In this way, it might be possible to turn shields—instead of swords—into plowshares. SA

MORE TO EXPLORE

America's Achilles' Heel: Nuclear, Biological and Chemical Terrorism and Covert Attack. Richard A. Falkenrath, Robert D. Newman and Bradley A. Thayer. MIT Press, 1998.

Biological Warfare: Modern Offense and Defense. Edited by Raymond A. Zilinskas. Lynne Rienner Publishers, Boulder, Colo., 1999.

Information on the TopOff exercise and Dark Winter is at **www.house.gov/reform/ns/web_resources/briefing_memo_july_23.htm**

FRED RIX

The VIGILANCE Defense

Proven systems and

By Stephen S. Morse

Last fall, when physicians diagnosed anthrax in a Florida man named Robert Stevens, they initially suspected that they were seeing a rare, natural case of the disease. The infection of a co-worker of Stevens, however, put the country on alert. Five people eventually died, but the quick recognition that the earliest anthrax infections resulted from a purposeful release of bacteria surely saved many lives: infected people were diagnosed and treated sooner, and officials isolated anthrax-tainted mail.

Federal and local governments have long considered bioterrorism a law enforcement issue. But both the anthrax episode and recent simulations of simultaneous smallpox releases in Atlanta, Philadelphia and Oklahoma City and of a plague attack on Denver highlight what perhaps should be obvious: bioterrorism is first and foremost a public health concern. For now and the foreseeable future, the first indication of a bioterror attack—as in a natural outbreak of a dangerous infectious disease—will be sick people showing up at emergency rooms rather than a high-tech device sounding an alarm that a pathogen has been released in a public place. The smallpox exercise in particular showed that the health care system is unprepared for the large numbers of people who would become sick in a worst-case scenario, not to mention the demands for attention from the "worried well," who would merely fear that they might be ill.

Regardless of whether a disease outbreak results from bioterrorism or natural causes, preparation can make the difference between a quickly contained threat and a widespread health crisis. And those who will make the most difference are the frontline defenders of the public's health: doctors, nurses and others who care for the first patients; laboratory researchers charged with determining the cause of the patients' illnesses; and officials in public health agencies charged with protecting whole populations. Disease control depends on the ability of these caregivers to recognize that something unusual and possibly dangerous is going on and to mobilize the individuals and resources that can best limit the disease's spread. It also requires having action plans in place that clearly define the roles and responsibilities of the nation's public health defense teams.

An effective early-warning system requires three elements, one of which I call "clinical recognition." The system needs physicians or other health workers who can recognize an unusual disease, who will order tests needed for a definitive diagnosis when symptoms seem atypical, and who will report such symptoms and troubling lab results to local or state departments of public health. In the first Florida case, doctors initially suspected meningitis, but thorough testing revealed the cause of the meningitis to be anthrax—half of inhalation anthrax victims also get meningitis—which Stevens's physicians dutifully reported to public health officials.

Hoping that an individual physician will connect one patient's symptoms with an outbreak, however, is unrealistic. A better approach is to train all clinicians to practice "syndromic surveillance": to actively look for and report certain syndromes, such as flu-like illnesses or rashes, that are common with potential bioterror pathogens. In most cases, lab tests will determine the cause to be a natural, familiar disease. Syndromic surveillance for flu-like illnesses should detect—no surprise—the flu, which would be valuable in itself, as pandemic flu strains are a major concern. But potential terror agents, such as pulmonary anthrax and plague acquired through inhalation, also begin with flu-like symptoms.

In addition, municipalities and states can monitor other, more subtle signs of an outbreak. Sudden increases in drug purchases at pharmacies, in workers or students calling in sick and in specific symptoms appearing at hospitals can all be tracked and can warn public health officials of the need for further investigation.

In 1993 my colleagues and I started the Program for Monitoring Emerging Diseases (ProMED), designed to promote global infectious disease surveillance and especially to look for outbreaks of novel diseases. The beginnings of a way to accomplish that goal came the next year when Jack Woodall, then of the New York State Department of Health, and I developed ProMED-mail. This open e-mail system, now administered by the International Society for Infectious Diseases, allows interested parties anywhere in the world to report clinical disease observations. E-mail is edited and evaluated for scientific validity before being sent to the entire subscriber list of 25,000 individuals, including disease experts, who can post comments and take personal action. No formal response system exists, however, although the World Health Organization does monitor ProMED-mail and other sources and has a system for notifying local representatives and recommending further

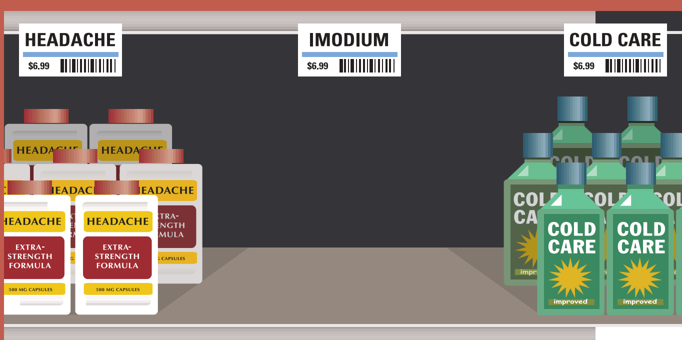
THE AUTHOR

STEPHEN S. MORSE is a virologist who in 1988 coined the term "emerging viruses" and who has led the movement to recognize the threat of emerging infectious diseases. He is director of the Center for Public Health Preparedness at the Mailman School of Public Health of Columbia University and is on the faculty of Columbia's epidemiology department. He was previously at the U.S. Department of Defense's Defense Advanced Research Projects Agency, where he co-directed the pathogen countermeasures program and directed the advanced diagnostics program.

investigation. More comprehensive worldwide surveillance, akin to worldwide weather observation, would only improve our capacity to notice any storms on the horizon and to respond in a timely fashion.

Even one case of a highly unusual disease, such as inhalation anthrax or Ebola, should trigger further epidemiological investigation (including a search for how the disease was acquired), the second key feature of the early-warning system. Unfortunately, despite the commitment and talent of people at the Centers for Disease Control and Prevention (CDC) and local health departments, the current national system is inadequate: public health is still a patchwork, with great variations in capacity from place to place.

Laboratory capacity is the third component. Labs are crucial for identifying specific disease agents in the early stages of an investigation and for determining whether a sample of material from a patient or a place contains that agent. The techniques of molecular epidemiology, which can identify individual strains of a pathogen based on molecular variations, are also invaluable for forensics. Molecular analysis could determine, for example, if a single substrain of anthrax was involved in all the 2001 infections, suggesting that the same person or group was involved.



SOLD-OUT ITEMS in pharmacies can be a sign of widespread illness. After public health officials in Milwaukee in 1993 became aware that the antidiarrheal agent Imodium had disappeared from pharmacy shelves, they discovered that the parasite *Cryptosporidium*, in tainted water, had infected some 400,000 people.

FRED RIX

Improved disease surveillance and response capability is a necessary first step. But further improvements are needed: the public health system is increasingly challenged in a variety of emergencies, from hurricanes to heat waves, and would play a major role in the unfortunate event of chemical or radiological terrorism, in addition to bioterrorism. Public health is no less a part of the national infrastructure than are bridges and highways, and a weak link anywhere is cause for concern.

That said, there is some good news to report. Since 1999 the CDC has been developing the Laboratory Response Network for bioterrorism, to complement the regular public health laboratory system. This network includes clinical microbiology labs around the country, which perform basic diagnoses and can culture and identify some pathogens. More specialized labs exist nationally in a system of "upward referral." A low-level lab would send unresolved or suspicious samples up the ladder to those better equipped to examine unusual materials. The ladder's highest rungs include laboratories at the CDC and at the U.S. Army Medical Research Institute of Infectious Diseases.

Ever since September 11 and the ensuing anthrax attacks that rocked the U.S., our government has been addressing the need to improve preparedness for bioterrorism. The establishment of a new Office of Public Health Preparedness in the U.S. Department of Health and Human Services in October 2001 was followed in January by supplemental federal appropriations of $2.9 billion in funding for bioterrorism preparedness—probably the largest single infusion of money into public health, after decades of chronic underfunding. The bulk of this money will help states improve their public health infrastructures. Among the items included is technology to enhance disease surveillance, as well as trained personnel—often the resource in shortest supply—to run that technology, carry out diagnoses and safely handle hazardous material. Another priority is expansion of the number of labs able to make fast, reliable identifications of the most important threats, such as anthrax, smallpox, plague, botulism toxin and tularemia.

In addition, the CDC will receive $116 million, most of which is earmarked for laboratory modernization. The agency will also improve its electronic Health Alert Network, which keeps state and local health departments in touch with one another and clinicians regarding active disease threats. The National Pharmaceutical Stockpile will expand: the number of "Push Packages"—50-ton medical supply kits (syringes, bandages, respiratory masks, antibiotics, etc.) ready to be delivered to disaster sites—will increase from the current eight to 12. And the funding will buy a supply of smallpox vaccine sufficient to protect the entire population if needed. A smallpox outbreak is highly improbable, but the consequences would be so significant that preparation is nonetheless necessary.

Once a disease is identified, the next step is to limit its spread and to treat those exposed. But this, too, is dependent on the quality of surveillance and communications. Where are clusters of clinical cases? What medications are needed where? How can cleanup crews best detoxify a given area, and which areas have priority? What physical evidence can be gathered about the disease agent that might lead to those responsible for perpetrating the attack? What, if any, quarantine measures should be considered? Sharing data is a must, both among local, state and federal health departments and between the public health community and the law enforcement community.

The public health infrastructure is not merely an essential component of biodefense—it may well be the *only* component in the earliest phases of a response to a bioterror attack. Outbreaks of infectious disease occur frequently, and global conditions suggest that the emergence and spread of infectious diseases will increase in the near future. Some, such as a new influenza pandemic, may spread rapidly and with devastating force. A fortunate aspect of the investment in biological defense is that it is sure to pay great dividends: whether a disease emerges naturally or is released purposefully, public health preparedness can save lives either way. SA

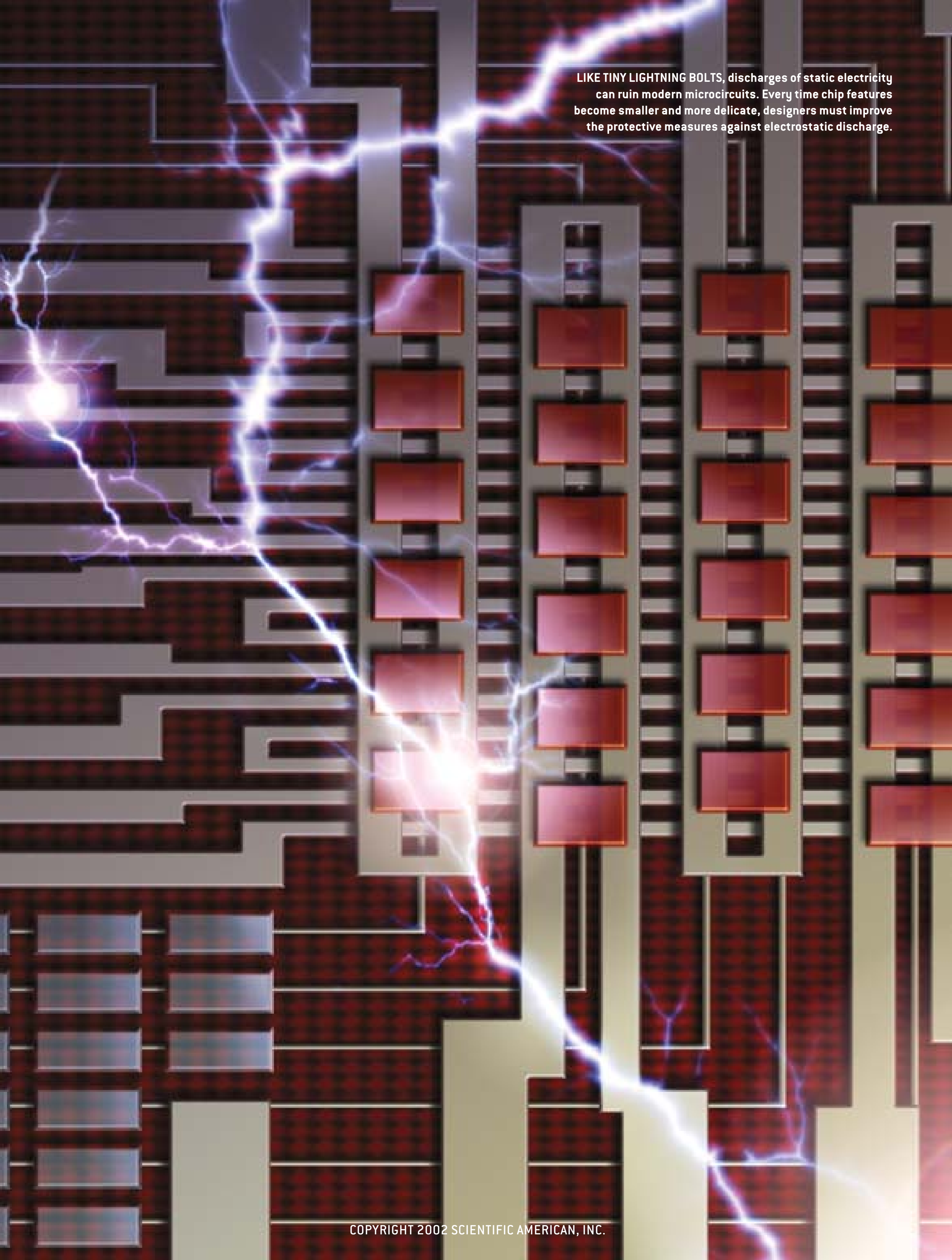
Lightning Rods for Nanoelectronics

Electrostatic discharges threaten to halt further shrinking and acceleration of electronic devices in the future

BY STEVEN H. VOLDMAN

We're all familiar with electrostatic charge: shuffle across a shag carpet in sneakers, touch a piece of metal, and zap. The slight prick we feel—caused when the electric charge built up by the shuffling suddenly leaps to another object—is nothing compared with what modern electronic equipment experiences.

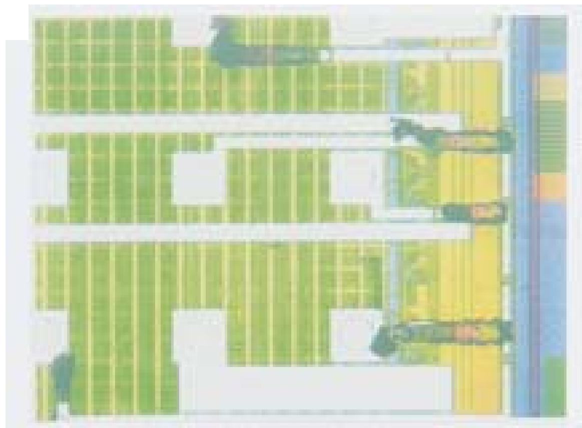
On a dry winter day, walking on a new carpet can generate a whopping 35,000-volt discharge. We are not harmed by this high voltage, because the amount of charge that flows is puny. Still, it is large enough to destroy sensitive micro-

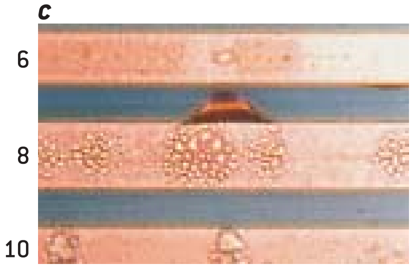


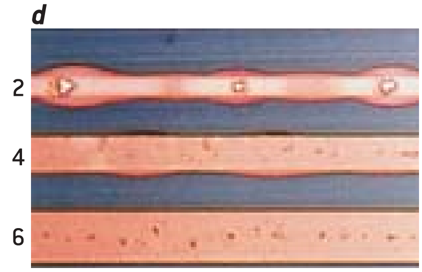
LIKE TINY LIGHTNING BOLTS, discharges of static electricity can ruin modern microcircuits. Every time chip features become smaller and more delicate, designers must improve the protective measures against electrostatic discharge.

DAMAGE TO INTERCONNECTS

INTERCONNECTS BETWEEN COMPONENTS and layers of chips can be damaged by excessive heating when a pulse of current from an electrostatic discharge passes through them. Aluminum interconnects are sandwiched between thin layers of titanium, which adhere better to the adjoining insulator. When the wire heats up, insulator on each side cracks and molten aluminum oozes out, as shown here for a silicon-on-insulator microprocessor that was subjected to a 7,500-volt discharge (*a*). More modern copper wiring has a higher melting point. The copper is in a trough of tantalum. The layer of insulator above the copper cracks when the wire is heated (*b*). When there is no insulator on top, the copper blisters or vaporizes (*c*). At extremely high temperatures (3,017 degrees Celsius and up), the tantalum sidewalls melt, allowing copper to extrude (*d*). (Numbers indicate wire widths in microns.)







electronic components. Researchers have come up with clever ways to prevent such damage. But as circuits get smaller, they become more sensitive to electrostatic discharge (ESD) and the old tricks no longer work. Can we continue to find new ways to prevent electrostatic damage and thereby maintain the pace of innovation?

People who like to tinker with their computers know that when they open up their machines, they should "ground" themselves—perhaps by touching the metal radiator panel or attaching a wire from their fingers to a metal fixture. This grounding diverts any built-up charge into another object. Microprocessors and other chips have built-in protection circuits, but for tomorrow's equipment such precautions will be of even greater concern. ESD is an issue not only for finished products but also during their manufacture, from wafer fabrication to packaging to the assembly of complete systems. Each step has its own electrostatic hazards.

In general, electrostatics poses the greatest threat during manufacturing and handling to install the devices and is less of a concern once the components are safely ensconced inside machines such as the computer on your desk. Some protection methods rely explicitly on this assumption [*see box on opposite page*]. The hazards start at the earliest stages of manufacture: even photolithographic masks, whose function is entirely mechanical and not electrical, are at risk [*see box on page 95*]. The chief danger to microelectronics is damage to the active elements caused by heating and by electrical breakdown of insulating layers. Magnetic disk drive heads, however, face their own unique problems, including magnetic aspects of discharges and considerations of aerodynamics [*see box on page 97*].

ESD protection devices have been incorporated into microchips since the 1960s and have evolved over the decades according to technological necessity and corporate strategy. The main goal with each new generation of microelectronics is to perpetuate Moore's Law by building elements such as transistors that are smaller and faster. Someday in the not too distant future the industry will hit a wall that blocks further progress. We will

Overview/*Electrostatic Protection*

- Processes as simple as a person walking across carpet can produce high voltages of electrostatic charge. Modern microelectronics is extremely sensitive and can be ruined by the pulse of electricity of an electrostatic discharge (ESD) that can occur from mere handling of a chip.
- Discharges melt metal and silicon and can punch holes into insulating layers. Modern ESD protection includes the use of sturdier materials and a variety of extra on-chip circuitry to divert discharges away from active elements.
- Devices become more sensitive to discharges with each generation of smaller circuitry, making ESD protection an ongoing challenge. Failure to meet that challenge would halt the progress of miniaturization and higher performance.

SLIM FILMS (*preceding pages*); STEVEN H. VOLDMAN (*above*) reach the point where we cannot design and build a smaller, faster transistor. But there is a second wall, because even if we can build the next transistor, it will be useless if we have no way to protect it adequately against ESD. No one knows which wall is coming first.

Thermal Runaway

WHAT CAUSES ELECTRONICS to fail when ESD occurs? The main culprit is heat generated by the electric current of the discharge, which can be enough to melt the material. Internal temperatures from ESD events can exceed 1,500 degrees Celsius (2,700 degrees Fahrenheit), above the melting points of aluminum, copper and silicon. Damage occurs even without melting. The properties of diodes and transistors are determined by the doping of the semiconductor: carefully introduced impurity atoms, or dopants, produce regions having specific electronic properties. Excessive heating can allow dopants to migrate, ruining the precise pattern of regions that is essential for the device to function properly.

Processes known as electro-current constriction and thermal runaway make matters worse by concentrating the heating in a hot spot: when one location of a semiconductor heats up significantly, its resistance falls, so that more of the current flows through the hottest place, heating it even more [*see illustration on next page*]. Geometry and design symmetry play a key role in distributing current evenly through a device to forestall the onset of thermal runaway. The material's thermal conductivity, heat capacity and melting temperature are all important in determining its ability to store the heat or diffuse it evenly.

Just as important as the sophisticated transistors of modern devices are the electrical connections between different elements. These include interconnect wires along the surface of the chip's semiconductor layers and "vias" that connect vertically from one layer to another. All of these are reduced in size along with the rest of the device in the effort to improve the speed and computational power of high-performance semiconductor chips. For many years, aluminum was the metal of choice for interconnects, but aluminum melts at only 660 degrees C. Beginning around 1997 (after 10 years of research), the microelectronics industry migrated to copper interconnects, primarily because of copper's superior electrical conductivity, enabling smaller and faster circuits. An additional benefit is copper's higher melting point, 1,083 degrees C, which gives interconnects higher tolerance to heating.

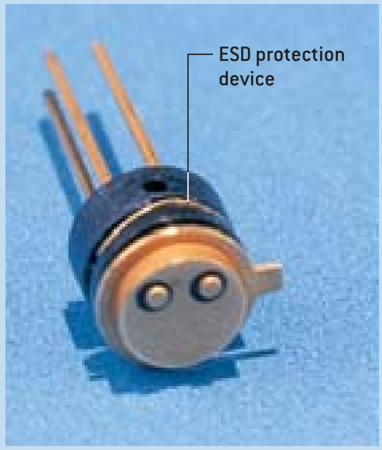
In contrast to the transition to copper, a generational change in insulating materials has had a minor negative side effect on sensitivity to ESD. These "low-k" materials form the insulating regions between metal lines in devices being rolled out in the marketplace. The materials' low dielectric constant (k) reduces the capacitance between the lines, which in turn reduces cross talk (interference between the lines) and increases the travel speed of high-frequency signals and short pulses. Unfortunately, low-k materials have lower thermal conductivity than silicon dioxide (the traditional insulator, or dielectric), so they are not as effective at dissipating energy from electrostatic events. This has to be compensated for by careful electrical design, wider interconnects or other techniques to reduce heating. Still, the net effect of introducing copper and low-k materials together is to improve sturdiness against ESD. Their introduction helped to bring about the transition to one-gigahertz (GHz) applications.

We now turn to the transistors, the workhorses of microchips. The primary

Making Sturdier Lasers and LEDs

Semiconductor lasers and light-emitting diodes (LEDs) are becoming ubiquitous, from supermarkets to large displays. Although these components are made mostly from the same materials as microprocessors, integrating electrostatic discharge (ESD) protection devices right on the semiconductor wafers can be prohibitively expensive. For displays, the cost per unit area is of tremendous importance, and the LEDs are made small and closely packed, leaving limited room for ESD protection circuits. Fragility to ESD is regarded as an acceptable trade-off. Researchers are therefore focusing on developing ways to make the materials used for LEDs intrinsically more robust against ESD.

David V. Cronin of Polaroid invented a mechanical solution to protect individual laser diodes when they are being handled: When the diode is not in its socket, conductive metal springs short the electrodes to the diode's metal casing (*see photograph at right*). Any ESD on the electrodes will flow to the casing instead of to the diode's semiconductor. When the laser diode is inserted into its socket, the metal spring disengages. *—S.H.V.*



BETH PHILLIPS

THE AUTHOR

STEVEN H. VOLDMAN is an electrostatic discharge (ESD) engineer/scientist on the silicon-germanium development team at IBM in Burlington, Vt. In his academic years he was a member of the Massachusetts Institute of Technology Magnetic Mirror Fusion Research Group and the M.I.T. High Voltage Research Laboratory. His research includes electrostatic issues in CMOS, silicon-on-insulator, silicon-germanium and magnetic disk drive technologies. He was responsible for the design and development of ESD circuits for advanced microprocessors and other semiconductor chips. The recipient of 100 U.S. patents, he was an IBM Top Inventor in 1998, 2000 and 2001. For fun, he plays ice hockey and the shofar.

BASIC EFFECTS OF HIGH VOLTAGES IN SEMICONDUCTORS

Dielectric breakdown

a Thin insulating layers are vulnerable to dielectric breakdown: A sufficiently strong electric field creates chains of molecular defects that cross from one side of the insulator to the other, similar to lightning cutting a path through air. The defects remain as permanent "pinholes."

b Because the electric field is voltage (V) per unit distance, a thicker layer experiences a weaker electric field and is less vulnerable than a thinner one.

Avalanche multiplication

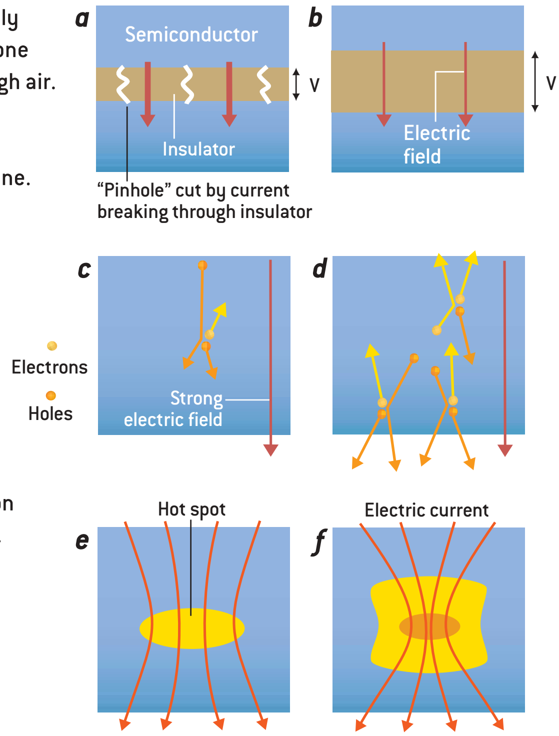
c Avalanche multiplication occurs when an excessive electric field acts in a semiconductor. The field accelerates electrons and holes to sufficient energy to knock other electrons and holes free, which add to the current.

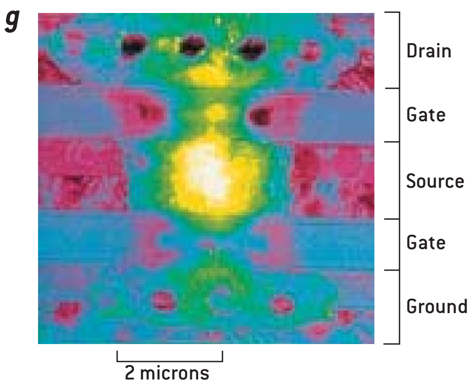
d The field accelerates these additional current carriers, which knock even more carriers free in a growing cascade that produces a very large current flow.

Electro-current constriction

e Resistance in a semiconductor falls with rising temperature, so if a location heats up significantly (*yellow*), more current flows through the hot region.

f The narrowing current exacerbates the heating (*orange*), which further concentrates the current.





MOSFET breakdown

g In this atomic force microscopy image of a MOSFET damaged by electrostatic discharge, electro-current constriction is apparent. Colors indicate the height of the semiconductor. The large green and yellow shape is the result of melted silicon welling up.

digital technology today is the MOSFET device, named after the metal-oxide semiconductor field-effect transistors that they contain. The basic MOSFET structure consists of two doped regions called the source and the drain, separated by a region called the channel [*see illustration on page* 96]. An electrode called the gate sits above the channel, separated from it by a thin layer of silicon dioxide dielectric. The voltage applied to the gate controls how current flows in the channel between source and drain.

Such devices have entered the nanostructure age in recent generations. In August, for example, Intel announced plans to manufacture chips with gates 50 nanometers long and gate oxides 1.2 nanometers thick—a mere five atomic layers. The thinner the dielectric, the lower the voltage needed to cause breakdown. Dielectric breakdown is caused not by heating but by the electrical carriers (electrons or holes) breaking molecular bonds and cutting a path through the insulator like a tiny bolt of lightning [*see illustration above*]. The defects formed by oxide failure are known as pinholes. With very thin oxide layers, mere handling of microelectronic chips can produce pinholes in the gates.

The source and drain of a MOSFET are also sensitive, and ESD on those regions leads to MOSFET thermal breakdown. When the high voltage of the discharge arrives at the drain, say, it increases the electric field there. This strong field accelerates the current-carrying electrons, making them energetic enough to knock other electrons free. These secondary electrons (and corresponding holes) increase the current flow even more and are themselves accelerated enough to knock more electrons free, and so on. Called avalanche multiplication, this process causes current to flow from the transistor into the nearby substrate, which puts the transistor into an unstable "negative resistance" state, further exacerbating the situation. As the current increases, heating leads to the thermal runaway, or thermal breakdown, described earlier.

The chief technique used to protect delicate transistors is to include ESD protection circuits on the chip, to divert cur-

ALICIA CALLE (*illustrations*); STEVEN H. VOLDMAN (*micrograph*) rents from discharges away from the transistors toward the ground or power electrodes. Other circuits can then transfer the current from one power rail to another until it finds the ground potential.

In the 1960s and 1970s these ESD devices were generally silicon-controlled rectifiers, but 1980s and 1990s technology often used MOSFETs themselves—transistors to protect transistors. The ESD transistors are designed to be much sturdier than the transistors they protect, which is possible because they do not have to meet the same high-performance standards as the active transistors. Adding such circuits does, however, affect the microchip's performance, and optimizing the design to minimize this impairment for each new microchip is a huge chore.

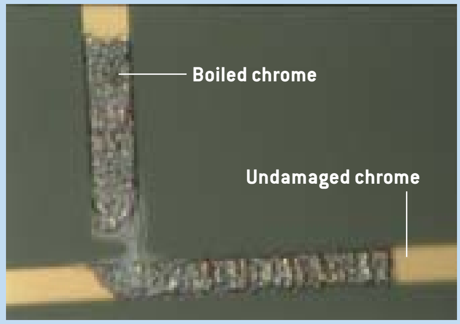
At present the favorite ESD device is the silicon *p-n* diode, which can handle larger discharges than a transistor ESD device of comparable size can: the transistor is limited because it must send the current through its narrow channel and its gate must resist dielectric breakdown. The widely used CMOS (complementary metal-oxide silicon) technology has a natural *p-n* diode structure running from the MOSFET to the chip substrate, which can be adapted to function as the ESD device. Additional *p-n* diodes running down to the substrate are easy to incorporate, although optimization remains a problem.

Since about 1995, when circuit line widths shrank to 250 nanometers and less, "smart" circuits known as ESD power clamps have been used to discharge the ESD current through the final stage, from the power rail to the ground. For example, some power clamps use a simple frequency-dependent filter to discriminate an ESD pulse from normal signals. Others detect the excess voltage of the discharge. Once the device senses the pulse, a signal powered by the pulse turns on robust transistor circuits to discharge the current safely to the ground plane. Many years ago transistors did not respond fast enough to cope with the highest-frequency part of an ESD pulse, which can be as high as a gigahertz. Modern high-speed transistors no longer have that difficulty.

Today all microprocessors and many other devices, such as radio-frequency chips, use ESD protection circuits and power clamps. With each new generation of smaller circuitry, the task of designing these devices becomes harder because they, too, must be smaller while providing better protection and without disturbing the ever higher performance of the transistors they are protecting. The jury is still out on whether these techniques will suffice for future semiconductors.

When Photomasks Spark

Photolithographic masks define which materials are laid down where in a microelectronics device, and their functioning at finer size scales is vital to the ability to continue miniaturizing technology. It may come as a surprise that electrostatics poses a problem to these photomasks even though they are purely mechanical and not electrical in function.



ELECTROSTATIC DAMAGE to a photomask halted production at a major European semiconductor manufacturer. The bridge defect is chrome oxide.

Recent studies by Julian Montoya of Intel and Arnold Steinman of Ion Systems in Berkeley, Calif., show that shapes on masks can become charged and subsequently discharge to adjacent shapes, causing damage to the masks (*see micrograph above*). When the spacing between two lines is wide, it takes a higher voltage to jump the gap. Such large discharges release enough energy to cause highly visible damage. When lines are closer together, however, discharges occur at lower voltages, producing less visible damage, which can be overlooked by inspections of the photomasks but is still destructive enough to spoil the devices made with the mask.

Practical techniques to improve the reliability of these photomasks include increasing the humidity and ionizing the air (to help dissipate charges) when the masks are assembled and monitoring factory tools for buildup of charge. The masks themselves could be made of dissipative materials that bleed off the charge slowly before it can reach a dangerous level. Much more remains to be understood about the discharge effects and the mask damage that they cause. *—S.H.V.*

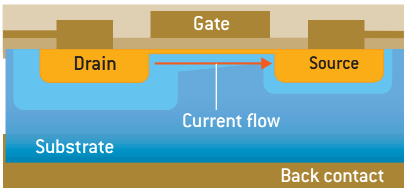
Silicon-on-Insulator

IN A MICROELECTRONICS technology called silicon-on-insulator, the ESD protection methods used in traditional MOSFETs face an obstacle. This technology promises higher speeds by reducing transistor capacitance (recall that capacitance slows down signals in the metal interconnects as well). In silicon-on-insulator devices the active circuit elements are in a thin silicon film that is separated from the rest of the chip substrate by a layer of insulator. The insulating film means that there is no natural diode structure available to divert ESD away from the sensitive electronics and down through the substrate. In the absence of vertical diodes, the only choice is to build lateral structures in the thin silicon film. In 1994, when research work began, some observers doubted that suitable ESD protection could be built for silicon-on-insulator technology. But by 2000 lateral *p-n* diodes running from the input signal lines to power rails were providing excellent ESD protection in IBM's mainstream silicon-on-insulator microprocessors. This technique will continue to suffice for future ultrathin devices, in which the silicon layer may be as thin as 20 nanometers.

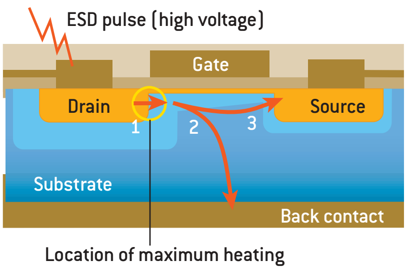
CMOS systems are well suited to digital applications, such as microprocessors, in part because they draw very little current in the 0, or off, state, and they conduct efficiently in the 1, or on, state. They

JAMES WILEY *KLA-Tencor Corporation*

THERMAL BREAKDOWN OF A MOSFET

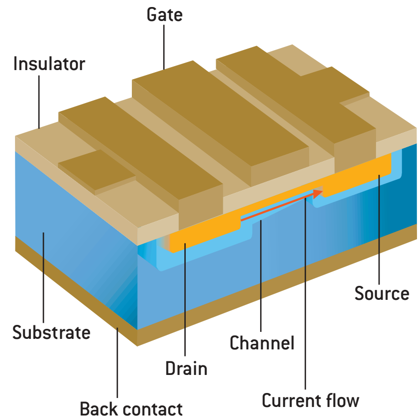


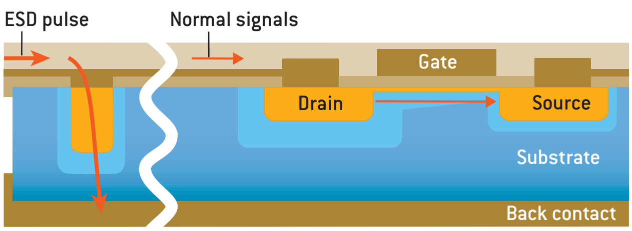
In normal operation of a MOSFET, current flows through a thin channel between the drain and the source. Voltage applied to the gate turns this current on and off.



When the very high voltage of an electrostatic discharge (ESD) pulse arrives at the drain, the current leaving the drain becomes extremely large (*1*) because of avalanche multiplication (*see illustration on page 94*). This current flows down through the substrate to the back contact (*2*), which changes the voltage in the substrate until the transistor "snaps back" and the current again flows to the source (*3*). Generally the maximum heating occurs at the junction of the drain and the channel.

Structure of a MOSFET





ESD protection circuits must divert such a pulse before it reaches the transistors on a chip, but at other times they should not disturb the normal flow of data signals or the states of the active transistors.

are not so well suited to applications such as the high-speed communications that are increasingly a part of modern life, including cellular telephones, laptop computers and personal digital assistants connected to the Internet, whether by wires or wirelessly. This market will continue to grow rapidly in the next decade, and application speeds are increasing from 10 to 100 GHz, requiring faster, smaller circuits that will be inherently more sensitive to ESD. (A technology called radio-frequency CMOS works between 1 and 10 GHz but is not yet used beyond 10 GHz.)

The technology that MOSFETs replaced, the bipolar transistor, is making a comeback for these types of applications through the use of new materials, in particular silicon germanium (SiGe) and gallium arsenide (GaAs). Bipolar transistors differ from field-effect transistors in that current flowing between two regions, called the emitter and the collector, is controlled by a small current entering an intervening region, called the base. A property of a semiconductor called the band gap determines how much it is like a conductor or an insulator. By engineering the band gap, one can build transistors that run 100 times as fast as in pure silicon. Transistor manufacturers control the band gap directly in the base of a SiGe transistor by adjusting the percentage of germanium. A transistor whose regions have different band gaps in this way is called a heterojunction bipolar transistor. (Herbert Kroemer was awarded the 2000 Nobel Prize in Physics for his early work on developing heterojunction transistors.) SiGe transistors are used in high-speed oscilloscopes, cell phones, Global Positioning System (GPS) devices and high-speed communications.

One can combine the advantages of SiGe with those of CMOS by building SiGe transistors on top of a layer of standard silicon CMOS technology. ESD protection of these devices is very similar to that used for CMOS: diodes built from bipolar transistors in the SiGe divert current from the SiGe transistors, and the usual diodes and MOSFETs do the same for the CMOS components.

These techniques have worked well all the way out to 100 GHz, but what will happen beyond that speed? (Already IBM has demonstrated 200-GHz transistors in the laboratory and is manufacturing 120-GHz technology.) At speeds of a few hundred gigahertz, we will probably have to start from scratch with a completely different form of ESD protection from that used now.

Another heterojunction technology uses gallium arsenide, which has better electrical characteristics than silicon. Physicists employ extremely high quality GaAs to study the fractional quantum Hall effect and other esoteric phenomena. Commercially, GaAs shows up in the power amplifiers of mobile phones and the optical interconnects that join elec-

ALICIA CALLE

tronic circuits to fiber optics. It is widely utilized for space applications such as satellites and interplanetary probes.

Standard CMOS devices are difficult to make with GaAs. In silicon, insulating layers can be made out of silicon dioxide, but there is no comparable native oxide in GaAs. This lack of an oxide also stymies the use of the type of ESD protection that is used in silicon, silicon-on-insulator and SiGe technologies. Consequently, GaAs devices are comparatively sensitive to ESD: few would survive contact by a person carrying as little as 1,000 volts, let alone the several thousands of volts easily generated just by walking on carpet.

Spark Gaps

ESD PROTECTION is of great concern for the space applications, in which charges build up on surfaces from sources such as the Van Allen belts near the earth and particles streaming from the sun. High-energy electrons can penetrate the interior and cause charge buildup on circuit boards deep inside. Developing adequate ESD protection for GaAs will be a challenge but is essential if future missions are to succeed.

Nearly a decade ago Karlheinz Bock, then at the University of Darmstadt in Germany, demonstrated a new type of protection for GaAs chips called field-emission devices or spark gaps. These are conical shapes etched into the GaAs, with an air gap between the cone's tip and another part of the device. A voltage from a discharge produces a very high electric field at the pointed tip, and sparking across the gap discharges the voltage. Field-emission devices have a number of advantages over diode- and transistor-based ESD protection: first, they have low capacitance and therefore minimal impact on the normal functioning of the device. More important, they can discharge high currents, and they can do so repeatedly. Field-emission devices will provide the level of ESD protection needed for GaAs power amplifiers and space applications once they make the transition out of the laboratory and into practical devices.

In years to come, traditional methods of ESD protection for semiconductors may not be acceptable with smaller, faster devices. Technologies other than GaAs might move to solutions such as spark gaps. Alternatively, designers might use new materials to make intrinsically sturdier transistors and rely on off-chip devices to prevent ESD pulses from reaching the nanocircuitry.

In 600 B.C., when Thales of Miletus was exploring electrostatic charge, little did he dream that 26 centuries later, electric charge would be influencing the direction and reliability of technology and that we would still be struggling with electrostatics. Understanding how semiconductors, magnetic recording heads, photomasks and other nanostructures charge and discharge will continue to be a focus of research and development. Inventions and patents of ESD circuits will burgeon as technologies are scaled to smaller sizes and as new disciplines realize that electrostatics matter. SA

Magneticians and Electrostatics

The magnetic recording industry uses magnetoresistor elements to read information stored on computer disks. The read head is a titanium carbide wafer with a thin-film stripe of magnetoresistive material. As the head sweeps over the disk, variations in the magnetic field from the disk alter the resistance of the stripe, which translates the signal into voltage. As each new design of disk drive increases the density of information on the disk, the magnetoresistor stripe has to be made correspondingly narrower.

Electrostatic discharges pose a variety of threats in addition to melting the magnetoresistor stripe. The current from a discharge can produce a magnetic field that can alter the data recorded on the disk. In disk drives, the disks and heads move at high velocities in extremely close proximity: the system has been compared to flying a fighter jet at 100 feet. A minute blister or nodule produced by ESD can spoil the aerodynamics of this flight and create havoc.

Devices to protect against ESD were not built into older magnetoresistor heads. Those heads were bigger and less sensitive, and there was no silicon wafer on which to build conventional protections. Protection would have been expensive. Magneticians, focused on the magnetics of drives, have only recently begun to address the electrostatic issues.

Low-resistance fuses have been built adjacent to magnetoresistor heads to serve as an alternative current path and to avoid failures during manufacturing. Mechanical solutions short the wires leading to the head so that failure does not occur during assembly of disk drives. (This is similar to the springs used on laser diodes mentioned in the box on page 93.) New head designs put the magnetic elements on silicon wafers, allowing integrated protection like that used for microprocessors.

Yet these read heads remain among the most sensitive of devices to ESD. Discharge voltages as low as about 35 volts can cause damage in read heads used today, which rely on an effect called giant magnetoresistance. The next generation of read heads, known as tunneling magnetoresistor devices, rely on quantum tunneling between magnetic films and are even more fragile—discharges as low as 10 volts would be a problem. That level of sensitivity makes manufacturing them very difficult. Researchers do not know if they can overcome these obstacles. *—S.H.V.*

MORE TO EXPLORE

Basic ESD and I/O Design. Sanjay Dabral and Timothy J. Maloney. John Wiley & Sons, 1998.

Investigating a New Generation of ESD-Induced Reticle Defects. James Wiley and Arnold Steinman in *Micro,* Vol. 17, No. 4, pages 35–40; April 1999. Available online at **www.micromagazine.com/archive/99/04/wiley.html**

ESD in Silicon Integrated Circuits. Second edition. Ajith Amerasekera and Charvaka Duvvury. John Wiley & Sons, 2002.

The ESD Association offers an introduction to ESD at **www.esda.org/aboutesd.html**

FLAT DISPLAYS

Vying for Eyes

Plasma televisions are the new status symbol. Even though they retail for $5,000 and up, they're selling like hotcakes. The displays are so thin, flat and light that they can be hung on a wall like a painting, transforming the TV set from a bulky, space-consuming appliance to a way-cool objet d'art.

The leading plasma displays measure 42, 50 and 60 inches diagonally yet are only three to five inches deep. Their resolution and brightness rival those of the best conventional cathode-ray tubes (CRTs). Although the obvious downside is cost, there is a less discussed fault: because a high voltage, from 160 to 200 volts, is required to excite the gas sandwiched between two glass plates, the units draw a lot of power and thus give off a lot of heat. The screens can also suffer burn-in—image artifacts etched into the pixels—in 10 years or so, according to Peter H. Putman, a display expert and president of Roam Consulting in Doylestown, Pa., if static images are persistently displayed at high contrast.

Plasma TVs have made quite a splash, but they do have a rival: liquid-crystal displays. LCDs are thin and light, run cooler, and cost somewhat less. But the picture decomposes if viewers watch from extreme angles, and manufacturing problems at large screen sizes will probably limit LCDs to 36 inches, says Richard M. Lewis, senior vice president of technology at Zenith. That means plasma displays could corner the large-screen market, because CRTs may top out at 40 inches; Lewis notes that CRTs become too deep and hard to manufacture at larger sizes because the electron gun that sweeps across the glass must be positioned farther and farther back.

Then again, the future may lie in organic light-emitting diodes—thin polymer films that glow to create an image. Putman says that manufacturers such as DuPont and Samsung have made OLEDs that are just an eighth of an inch thick, and some can bend like a plastic sheet. Screen sizes have reached only a few inches, but the displays are starting to appear in cellular telephones and handheld games. "It's too soon to tell which technology might prevail," Putman says. "But the CRT may soon be dead." *—Mark Fischetti*

PLASMA DISPLAY applies voltage to transparent electrodes, which discharge into a noble gas, such as neon or xenon, creating a plasma. The plasma conducts electricity briefly and emits a burst of ultraviolet rays, which stimulate phosphors to emit red, green or blue light back out the front panel toward the viewer. The polarity is reversed to clear the cell of any charge and prepare it for the next cycle. Each set of adjacent red, green and blue chambers constitutes one pixel; a screen may have up to a million pixels.

DANIELS & DANIELS

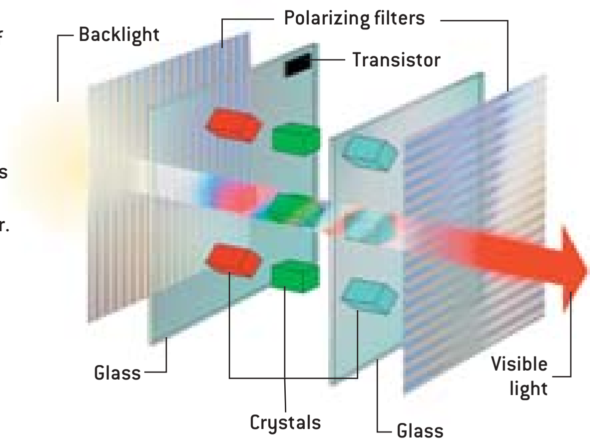
DID YOU KNOW . . .

➤ R NOT D: American engineers have conducted much of the pioneering research leading to ultrathin displays, but overseas companies have often introduced the first products. According to Peter H. Putman, the University of Illinois devised the plasma concept in 1964, with Owens-Illinois, IBM and Plasmaco conducting subsequent research, yet Asian companies such as Fujitsu, Sony and Samsung drove the retail rise in the late 1990s. RCA created the earliest prototype LCD panels in the 1960s only to see Sharp Electronics include the first commercial screens on calculators in the early 1970s.

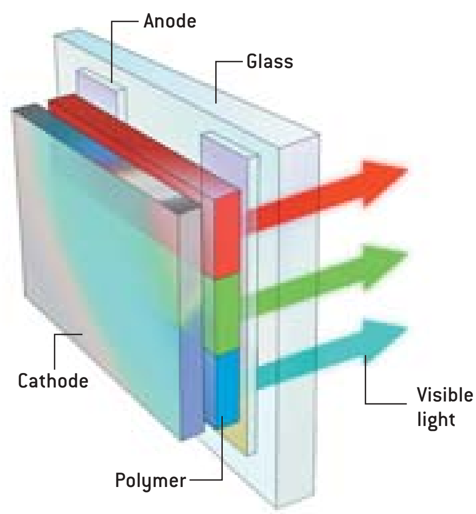
➤ SIZE MATTERS: Zenith's 40-inch-diagonal plasma TV, including its cabinet, is 3.1 inches deep and weighs 72.5 pounds. The company's conventional 36-inch CRT television is 23.9 inches deep and weighs 151 pounds—and the picture is 20 percent smaller. Price difference: about $3,000.

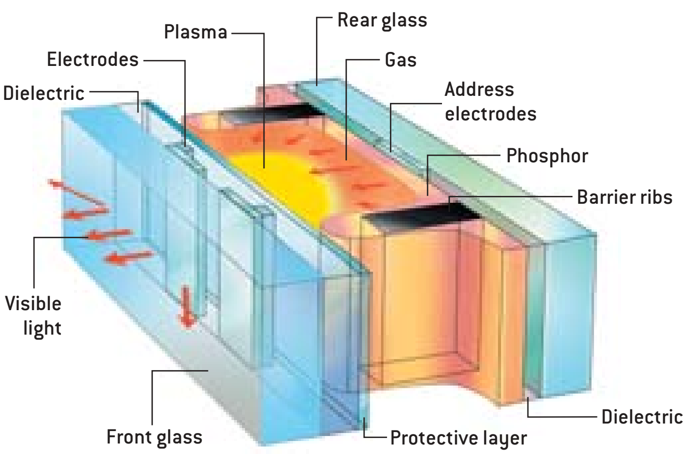
➤ RATIONALIZATION: Few consumers could justify the exorbitant price of the first retail plasma displays, but *New York Times* culture writer Rick Marin found a way in March 2001 after he saw a 42-inch Sony unit that had just been "reduced" to $7,999. "My apartment is probably worth $500 a square foot," he wrote from his Manhattan abode. "With a flat TV on the wall, I could get rid of my [old] TV and reclaim 16 square feet [of floor space]. That's $8,000 worth of real estate. I'd actually be saving $1."

LIQUID-CRYSTAL DISPLAY backlight shines through a matrix of tiny pixels, each filled with liquid crystals. A thin-film transistor applies a voltage to each pixel, aligning the crystals to various degrees that allow different amounts of red, green or blue light to pass through the front glass to the viewer.



ORGANIC LIGHT-EMITTING DIODES are thin films. An anode metal, an organic polymer and a cathode metal are deposited on a flexible glass substrate to form a pixel. Current applied to the anode and cathode sends electrons and holes into the polymer, where they collide and cause the polymer to glow red, green or blue toward the viewer.





This month's topic was suggested by reader Steve Dunivin. Have an idea for a future column? Send it to workingknowledge@sciam.com

Computers for the Third World

THE SIMPUTER IS A HANDHELD DEVICE DESIGNED FOR RURAL VILLAGERS BY FIONA HARVEY

It doesn't look like much. A drab, gray piece of plastic, about five inches long and three inches wide. A black-and-white screen, three inches by two inches, showing a few simple snippets of text. And yet this nondescript little computer may hold the key to bringing information technology to Third World countries.

The device is known as the Simputer. I recently got a chance to evaluate one of the preproduction models that have been put together by the Simputer Trust, a nonprofit organization based in Bangalore, India. This year Encore Software, a Bangalore company that licensed the technology from the trust (not to be confused with the California software company of the same name), plans to sell thousands of the handheld devices, capping an effort that began in 1998.

Simputer stands for "simple, inexpensive, multilingual computer." It was designed to meet the needs of rural villagers in countries such as India, Malaysia, Nigeria and Indonesia. Many of these potential users are illiterate and have never even seen a computer before. Loaded with some elementary software, the Simputer will sell for about $250 (or $300 for a model with a color screen). That's a sizable chunk of the yearly per capita income in many developing nations. But the Simputer's proponents argue that a single device could enable an entire village to access the Internet, perform transactions, keep track of agricultural prices and educate its children. Says Shreyas Patel, a consultant to Encore who has been setting up pilot tests of the Simputer in East Africa: "This will bring computing power to isolated communities. It can have an enormous impact."

FARM WORKERS IN BOLARE, a village located southwest of Bangalore, India, try out a literacy tutorial program on the Simputer. Because the device can convert text to speech, it can help teach villagers how to read the local language, Kannada.

The Simputer was conceived by a team of computer scientists at the Indian Institute of Science in Bangalore. To make the machine cheap enough to sell in poor regions, the developers kept the hardware requirements to a minimum. The Simputer's microprocessor is an Intel StrongARM chip, which is known for its low power consumption. The device will have as much as 64 megabytes of random-access memory and 32 megabytes of flash memory, as well as a modem that can connect to a telephone line. And the computer runs on the Linux operating system, which is available free of charge.

Like the Palm, the Visor and other personal digital assistants (PDAs), the Simputer has a touch-sensitive screen. You use a stylus to tap on icons and to input information. The device doesn't have a keyboard or handwriting-recognition software, but in certain applications the user can select letters or numbers from a software-generated keyboard that pops up on the screen. In addition, the Simputer has a program called Tapatap that displays a three-by-three grid; you can input a letter or number by tapping on the

PHOTOGRAPHS COURTESY OF ENCORE SOFTWARE LTD.

squares of the grid in a particular sequence. Although this method is easier than hunting and pecking on a software keyboard, it is still somewhat laborious, so the Simputer's applications have been carefully designed to minimize the need for tapping in text.

But how will illiterate people be able to use the Simputer if they can't read the directions on the screen? There are two answers. One is the simplicity of the device's interface: because each display page shows only a few possible commands, even illiterate users should be able to learn by trial and error the purpose of the icons and buttons on each page. The second answer is software that can turn text into speech. The Simputer holds a database of phonemes—basic linguistic sounds—and from these it can generate an audio representation of any word as long as it is spelled phonetically and in characters from the Roman alphabet. It will work for various Indian languages, including Hindi, Kannada and Tamil, allowing the Simputer to read the text aloud on its tiny built-in speakers. The Simputer Trust says the software will be made available in other languages as well, depending on where the device is used.

I was unable to test this function on my preproduction model, which lacked the text-to-speech program. I can confirm, however, that the Simputer is remarkably easy to use. Its screen is free of the annoying graphical clutter that most of us are accustomed to seeing on our PCs. Below the screen are seven small buttons, one an on-off switch and the rest for use with certain embedded applications such as the Tapatap program. I found that I did not need to bother with the buttons very often, because the design of the software made it easier to use the stylus.

The Simputer also has a slot for "smart" cards, a feature that its makers see as crucial. Because the device lacks a hard drive, smart cards will act as the device's portable storage units. In this way, many people will be able to share a single Simputer without having to share their private information with one another.

PREPRODUCTION MODEL of the Simputer has a black-and-white screen and built-in speakers. A Bangalore company named Encore Software plans to sell the device for about $250.

The cards will cost between $1 and $3 apiece and will hold four to eight kilobytes of data—not very much by commercial standards but enough to carry some basic information for each user. "We envisage that a village might club together to buy one," notes Shashank Garg, vice president for product development at Encore. A farmer in India, for example, could use the Simputer to find out the latest prices for cotton, allowing him to strike a better deal when selling his crop. The next day one of his neighbors could use the same device to examine government property records, eliminating the need to make a difficult journey to the city.

The Simputer Trust believes the range of applications will prove compelling. But the device does have some drawbacks. It's slow, taking about 15 seconds to boot up and often needing several seconds to digest the information that the user inputs. And the Simputer sometimes crashes when it is left idle for a while, making it necessary to reboot the machine. Also, powering the device may be a daunting task in areas that do not have a reliable electricity supply. Although the Simputer can run on three AAA batteries, it can operate for only a few hours before draining them. And in the developing world, even batteries are expensive and hard to come by.

Fortunately, Simputer users may be able to draw on muscle power instead. A decade ago English inventor Trevor G. Baylis created the Freeplay radio, which is powered by turning a crank that winds up a spring inside the machine. As the spring unwinds, it turns a shaft that runs a small electric generator. Freeplay Energy Group, the company that now sells these radios, recently produced a similar charger that can power a mobile phone. In a demonstration this year some energetic hand-cranking yielded enough energy to run an Apple laptop for a few minutes. With a few adaptations, devices such as these could charge up the Simputer.

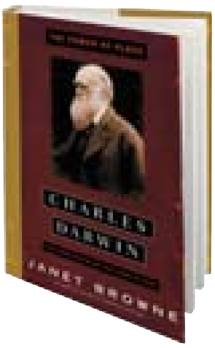
But the Simputer may not be the best tool for bringing information to the world's poorest nations. Because most people in developing countries have no access to fixed telephone lines, many mobile-phone operators are setting up networks in those areas. Mobile phones are cheaper than the Simputer, and the most advanced models can send text messages and access the Internet. Communities choosing between the devices may find a mobile phone more immediately attractive for keeping in touch with the outside world and conducting business.

Perhaps the greatest obstacle for the Simputer, though, is cost. Will people in developing countries be able to justify the expenditure of $250 on a device that may be helpful but is not essential? When so many communities in the Third World still lack clean drinking water and adequate medical facilities, are computers really a priority? SA

Fiona Harvey is a journalist based in London who writes about all aspects of technology for the Financial Times.

Putting Darwin in His Place

USING HIS QUIET COUNTRY ESTATE AS HEADQUARTERS, THE GREAT NATURALIST WAS A RECLUSIVE REVOLUTIONARY **BY RICHARD MILNER**



CHARLES DARWIN: THE POWER OF PLACE
by Janet Browne
Alfred A. Knopf, 2002
($37.50)

After years of immersion in Charles Darwin's 14,000 letters at the Cambridge Library, Janet Browne—an editor of the Darwin correspondence project—has published the second half of her sprawling, magnificent biography. Integrating the best of current scholarship with her own discoveries, Browne's account is state of the art. That said, as Stephen Jay Gould once opined, "too many Darwins dwelled within this enormously complex man" to allow any one biography to be "definitive." But if the most any author can do is to find his or her own Darwin, Browne's finely detailed portrait ranks among the best ever attempted.

Voyaging, her first volume, published in 1995, followed the young naturalist through his adventures on board HMS *Beagle,* his return home and his marriage to his cousin Emma Wedgwood. While still on the ship, in 1835, Charles had written his sister that he'd had enough seafaring. "I am convinced it is a most ridiculous thing to go round the world," he wrote, "when by staying quietly, the world will go round with you." That seemingly casual comment turned out to be his plan for life. After purchasing a comfortable home in the rural outskirts of London, the Darwins settled in to raise seven children, assisted by a staff of maids, cooks and gardeners.

Little by little, Darwin transformed Down House into a self-contained biological field station, a stationary *Beagle* trawling for facts. It became headquarters for fomenting, in Darwin's words, "a considerable revolution in natural history," the site of a coordinated family effort, the focus of a worldwide correspondence and the hub of a social network that reached deep into London's scientific community. Browne's latest volume picks up the story in the late 1850s, with events leading up to publication of *On the Origin of Species.*

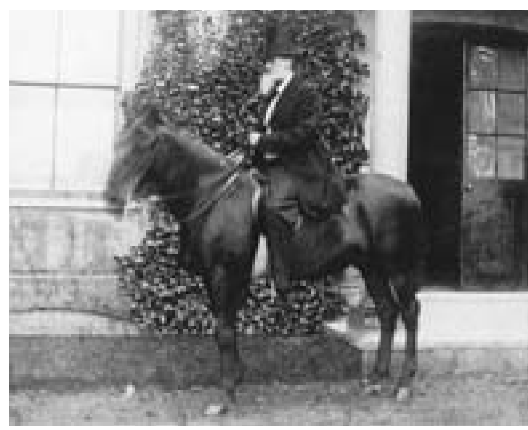
We've become familiar with the modest, reticent, semi-invalided Victorian gentleman who ceaselessly badgered correspondents for seeds, skins and field observations when he wasn't pottering among his pigeons, earthworms and carnivorous plants. But Browne offers surprising details about another Darwin—"a remarkable tactician … a canny and dedicated publicist." Darwin well understood that posterity does not necessarily remember the discoverer of an idea but the individual who first convinces the world of its truth.

Browne shows a Darwin obsessed over details of publishing, timing, choosing the right translators for foreign editions and closely monitoring book sales. While his lieutenants Thomas Huxley and Joseph Hooker promoted evolution within scientific circles, Darwin kept an eye on how he was faring with the general public. He clipped, catalogued and indexed hundreds of offprints, about 350 reviews and 1,600 articles, as well as satires, parodies and *Punch* caricatures, with which he filled hefty scrapbooks—a seemingly incongruous activity for a self-effacing sage. He remarked, "I am really quite sick of myself."

Browne also provides new and welcome details about Darwin as a pigeon breeder, botanist and crusader against spiritualist con men who drew members of his family into séances. His interest in varieties of domestic fruits, she notes, led him to raise 57 kinds of gooseberries!

As for Darwin's mysterious malady, Browne oddly dismisses any attempt to diagnose his illness "at such a distance" as "surely impossible" and considers Chagas' disease "or any latent tropical infection picked up during the *Beagle* voyage" as "unlikely," without offering her reasons. While in Argentina, Darwin was bitten by the Benchuca bug ("the great black bug of the Pampas"), which is now known to often carry the Chagas' blood

DARWIN on Tommy, outside Down House. The horse threw him in 1869, ending his riding days.



DARWIN COLLECTION, COURTESY OF THE SYNDICS OF CAMBRIDGE UNIVERSITY LIBRARY parasite in its feces. Darwin's chronicling of the invasive contact, along with his symptoms—40 years of flatulence, vomiting, weakness and fatigue—are entirely consistent with a diagnosis of chronic (arrested) Chagas' disease. His doctors were baffled, because the connections between the stomach pathology, the insect bite and the insidious trypanosome were not known to medicine until almost three decades after Darwin's death. In addition, historian and psychiatrist Ralph Colp, Jr., has correlated some of Darwin's worst attacks with psychosomatic stress.

Colp demonstrated years ago that Darwin suffered many of the most severe bouts of illness whenever his work or his theory was attacked in print. In 1863, for instance, just after his geological mentor Charles Lyell published an infuriatingly tepid endorsement of evolution in his *Antiquity of Man,* Darwin's disappointment brought on 10 days of vomiting, faintness and stomach distress. (He then canceled a long-planned visit by Lyell to Down House.) Similarly, in 1871, anatomist St. George Mivart's scathing attack on *The Descent of Man* triggered two months of "giddiness" and inability to work—part of a striking and recurrent pattern that Browne inexplicably ignores. That lapse aside, however, this marvelous book is indispensable for anyone seeking to understand the man (and the crucial role played by his family and friends) as he revolutionized Western thought about human nature and origins.

Browne has appropriately titled her concluding volume *The Power of Place.* Even present-day visitors to Down House are often struck by how the house and grounds still seem to bear the impress of Darwin's powerful personality. She invites the reader to enjoy extended and enlightening visits with the Darwins and their friends. The best part is that whenever you want to drop in, they are always home. SA

Richard Milner wrote The Encyclopedia of Evolution *and the one-man musical* Charles Darwin: Live & in Concert.

THE EDITORS **RECOMMEND**

THE LUNAR MEN: FIVE FRIENDS WHOSE CURIOSITY CHANGED THE WORLD

by Jenny Uglow. Farrar, Straus and Giroux, New York, 2002 ($30)

"We were united," one of them said, "by a common love of *science,* which we thought sufficient to bring together persons of all distinctions." That was Joseph Priestley. A number of other men, many of whom were or would become famous, belonged to the Lunar Society of Birmingham in England during its lifetime in the latter decades of the 18th century. "They are a small, informal bunch who simply try to meet at each other's houses on the Monday nearest the full moon, to have light to ride home (hence the name) and, like other clubs, they drink and laugh and argue into the night. But the Lunar men are different—together they nudge their whole society and culture over the threshold of the modern, tilting it irrevocably away from old patterns of life towards the world we know today." Uglow, biographer of (among others) William Hogarth and George Eliot, focuses on the five members who formed "the core." Besides Priestley, they were Erasmus Darwin, Matthew Boulton, James Watt and Josiah Wedgwood. Through her portraits of them and their work, she evokes vividly the state of science and technology on the eve of the industrial revolution. This "small group of friends," Uglow writes, "really was at the leading edge of almost every movement of its time in science, in industry and in the arts—even in agriculture."

WILD ORPHANS

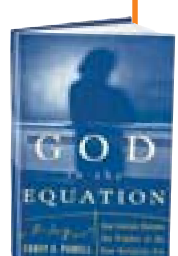
Photographs and text by Gerry Ellis. Welcome Books, New York, 2002 ($24.95)

This is the inaugural volume in a series of books featuring photographs and stories from the various field sites of GLOBIO's Wild Orphans project. Photojournalist Ellis founded GLOBIO to help children learn about biodiversity by illustrating the relation between an orphaned wild animal and the state of its ecosystem. (Orphaned wildlife are key indicators of the pressures exerted on biodiversity by poaching, habitat loss, toxins, overhunting and wars.) In this book, Ellis documents the story of eight baby elephants brought to the David Sheldrick Wildlife Trust in Kenya during the drought-plagued summer of 1999. He chronicles their lives over two years, beginning with haunting photographs of the rescues of the abandoned baby elephants, through their upbringing by dedicated 24-hour surrogate "mothers" (all African men of almost superhuman patience and generosity), to their release back into the wild. It's an amazing project and a stunning book. More photos and information are available at www.sciam.com

GOD IN THE EQUATION: HOW EINSTEIN BECAME THE PROPHET OF THE NEW RELIGIOUS ERA

by Corey S. Powell, Free Press (Simon & Schuster), New York, 2002 ($24)

Science writer Powell casts science as the new religion, with Einstein as god. "Sci/religion," as he calls it, "offers a positive and immensely appealing alternative way to look at the world, a religion of rational hope." Even if you disagree with Powell's premise, his book is a delight to read—lively, well-informed, personable. And as a bonus, it provides an unusually graceful account of the history of cosmology.



All the books reviewed are available for purchase through www.sciam.com

Prime Spies

BY DENNIS E. SHASHA

Number theory, once thought an esoteric discipline concerned with curious properties of prime numbers, turns out to form the underpinnings of modern cryptography. The Rivest-Shamir-Adleman public-key cryptography algorithm (RSA, for short) used in most e-commerce transactions relies on the (unproved but widely believed) difficulty of factoring a number that is the product of two primes.

Multiplying two large primes is an example of a so-called one-way function: the multiplication takes only a few microseconds, and the time is proportional to the length of the binary representations of the numbers; in contrast, discovering the two primes given the product is slow, taking hours for a 512-bit product and continuing a slow exponential climb (in the length of the product) after that. For 2,048-bit numbers, factoring is considered impractical, as far as is publicly known. Fast factoring algorithms, if they exist, would have untold applications for industrial and even military espionage.

This brings us to a puzzle first posed by John McCarthy (inventor of the programming language Lisp and theoretical Artificial Intelligence) and solved by Michael Rabin (the playful inventor of so many important computer algorithms) in the 1950s. The puzzle goes like this: You have a bunch of spies ready to go into enemy territory. When they return to cross the frontier into your country, you want to avoid getting them shot, while at the same time preventing enemy spies from entering. So each must present a password to the guards, which the guards will verify. Whereas you trust your spies, and your guards are loyal, you believe the guards may loosen their tongues in bars at night. What information should the guards receive, and how should the spies present their passwords, so that only your spies get through and nobody else, even if the guards go out for a couple? Consider the discussion about primes to be a hint. SA

Dennis E. Shasha's third book of puzzles, Dr. Ecco's Cyberpuzzles, *has just been published.*

Answer to Last Month's Puzzle

If you spread your investments over 10 companies, giving each $1.43 million, the chance that at least seven will yield 10-fold returns is more than 95 percent. If that happens, the total return would be $100.1 million. This strategy leaves $2.7 million in reserve for future investments. For a full explanation, visit www.sciam.com

Web Solution

For a peek at the answer to this month's problem, visit www.sciam.com

SARA CHEN

The 2,000-Year-Old Menace

IS IT POSSIBLE THAT FUNNY IS THE ROOT OF ALL EVIL? **BY STEVE MIRSKY**

U.S. Attorney General John Ashcroft announced in July that the Bush administration would go ahead with the creation of a program called the Terrorism Information and Prevention System, or TIPS. The system would encourage people whose jobs put them in contact with a wide range of fellow Americans to report any "suspicious" activity. Some lawmakers worried about TIPS, expressing concern that it might turn us into a nation of snitches. Never mind that. These are desperate times, and desperate times call for desperation. Therefore, it is with a heavy heart that I am duty-bound to alert America to the clear and present danger posed by Mel Brooks.

I first became suspicious of Mr. Brooks when the Enron scandal broke. My suspicions only grew as other examples of "aggressive accounting procedures" became public, bringing down such companies as Arthur Andersen and WorldCom. These fears were based on my intimate knowledge of the plot points of Brooks's 1967 film *The Producers,* for which he won the Academy Award for best original screenplay. More recently, Brooks turned *The Producers* into a Broadway smash. The show swept the 2001 Tony Awards, while tickets for the show vacuumed out my wallet.

For anyone living in a cave (more on that later), *The Producers* is the story of down-and-out ex–big shot Broadway impresario Max Bialystock and his protégé, the former accountant Leo Bloom. When we first meet Bloom, he is a current accountant, sent to examine Bialystock's ledgers. The notion hits Bloom that under the right circumstances a producer could make more money with a flop than with a hit. Just get investors to put up way more money than the play actually costs to stage. The key is that the play must close early: if it ran long enough, it might appear to make money, but there could never be enough profits to pay off the backers. Bloom explains his "academic accounting theory" to Bialystock, who realizes Bloom's inherent genius, and they go on to sell 25,000 percent of a new show, the musical romp *Springtime for Hitler,* to investors.

Enron CEO Kenneth Lay apparently stopped watching *The Producers* just before Bloom clarifies that, in reality, "you can only sell 100 percent of anything." Enron, after all, kept selling parts of itself to other parts of itself and claimed the acquisitions as assets—voilà, the old 110-percent (at least) effort. Lay therefore missed the part where *Springtime for Hitler,* which Bialystock predicted would close on page four of the script, becomes an unfortunate hit, and he and Bloom go to jail (for blowing up the theater to try to get out of their accounting mess).

Despite Brooks's obvious tutelage of today's accounting criminals, I dismissed my initial suspicions of him as mere paranoia. But in July I happened on an episode of *Get Smart,* the late-1960s sitcom about bumbling secret agent Maxwell Smart. The plotline concerned a plague sweeping the nation's potato crop—potatoes look fine on the outside but, like Lay and his cronies, are empty inside their jackets. Smart discovers that Siegfried, head of the evil enemy agency Kaos, is using a bioengineered bacterium to attack our potatoes [see "Biological Warfare against Crops," by Paul Rogers, Simon Whitby and Malcolm Dando; SCIENTIFIC AMERICAN, June 1999]. Siegfried explains that the bacterium enters the potato, eats the insides, burps and dies, leaving no trace. Now for the really frightening part—Siegfried is spreading the potato-destroying bacteria using crop dusters. And who co-created *Get Smart*? That's right: Mel Brooks, who is clearly an inspiration to corporate schemers and cave-dwelling evildoers (see, I told you we'd get to people living in caves).

By the way, "Brooks" is not even his real name. He was born Melvin Kaminsky, yet he lives freely among us using a false identity. I find this whole situation highly suspicious. And I know for a fact that guys whose names end in "-sky" are not to be trusted. SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

How is caffeine removed to produce decaffeinated coffee?

—Rick Woolley, Everett, Wash.

Fergus M. Clydesdale, head of the food science department at the University of Massachusetts at Amherst, provides this answer:

There are currently three main processes, all of which begin with moistening the green or roasted beans to make the caffeine soluble. Decaffeination is typically carried out at 70 to 100 degrees Celsius.

In the first method, called water processing, the moistened coffee beans are soaked in a mixture of water and green-coffee extract that has previously been caffeine-reduced. Osmosis draws the caffeine from the highly caffeine-concentrated beans into the less caffeine-concentrated solution. Afterward, the decaffeinated beans are rinsed and dried. The extracted caffeine-rich solution is passed through a bed of charcoal that has been pretreated with a carbohydrate. The carbohydrate blocks sites in the charcoal that would otherwise absorb sugars and additional compounds that contribute to the coffee's flavor but permits the absorption of caffeine. The caffeine-reduced solution, which still contains compounds that augment the taste and aroma, can then be infused into the beans. The water process is natural, in that it does not employ any harmful chemicals, but it is not very specific for caffeine, extracting some noncaffeine solids and reducing flavor. It eliminates 94 to 96 percent of the caffeine.

An alternative method extracts caffeine with a chemical solvent. The liquid solvent is circulated through a bed of moist, green coffee beans, removing the caffeine. The solvent is recaptured in an evaporator, and the beans are washed with water. Finally, the beans are steamed to remove chemical residues. Solvents, such as methylene chloride, are more specific for caffeine than charcoal is, extracting 96 to 97 percent of the caffeine and leaving behind nearly all the noncaffeine solids.

In the third approach, carbon dioxide is circulated through the beans in drums operating at roughly 250 to 300 times atmospheric pressure. At these pressures, carbon dioxide takes on unique supercritical properties, having a density similar to that of a liquid but with the diffusivity of a gas, allowing it to penetrate the beans and dissolve the caffeine. These attributes also significantly lower the pumping costs for carbon dioxide. The caffeine-rich carbon dioxide exiting the extraction vessel is channeled through charcoal or water to absorb the caffeine and is then returned to the extraction vessel. Carbon dioxide is popular because it has a relatively low pressure critical point, it is nontoxic, and it is naturally abundant. Supercritical carbon dioxide decaffeination is more expensive, but it extracts 96 to 98 percent of the caffeine.

SARA CHEN

Why is spider silk so strong?

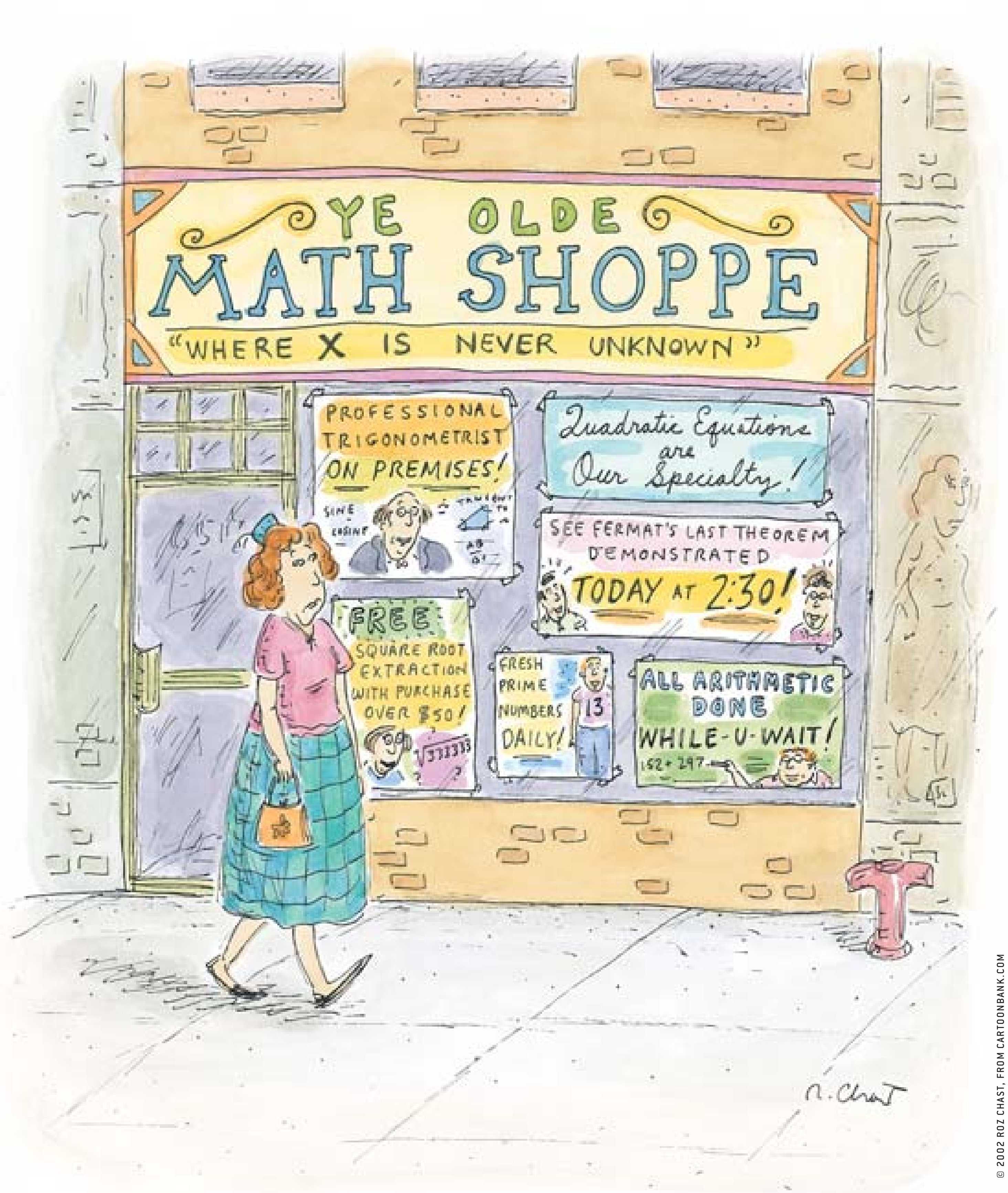
—D. Gray, Corinna, Maine

Biologist William K. Purves of Harvey Mudd College offers an explanation:

Dragline silk, the silk that forms the radial spokes of a spider's web, is composed of two proteins, making it strong and tough—yet elastic. Each protein contains three regions with distinct properties. The first forms an amorphous (noncrystalline) matrix that is stretchable, giving the silk elasticity. When an insect strikes the web, the matrix stretches, absorbing the kinetic energy of the insect's impact. Embedded in the amorphous parts of both proteins are two kinds of crystalline regions that toughen the silk. Both regions are tightly pleated and resist stretching, and one of them is rigid. It is thought that the pleats of the less rigid crystalline regions anchor the rigid crystals to the matrix.

A spider's dragline is only about one tenth the diameter of a human hair, but it is several times stronger than steel, on a weight-for-weight basis. The recent movie *Spider-Man* drastically underestimates the strength of silk—real dragline silk would not need to be nearly as thick as the strands deployed by our web-swinging hero. SA

For a complete text of these and other answers from scientists in diverse fields, visit www.sciam.com/askexpert.


YE OLDE
MATH SHOPPE
"WHERE X IS NEVER UNKNOWN"
PROFESSIONAL TRIGONOMETRIST ON PREMISES!
SINE
COSINE
Quadratic Equations are Our Specialty!
SEE FERMAT'S LAST THEOREM DEMONSTRATED
TODAY AT 2:30!
FREE
SQUARE ROOT EXTRACTION WITH PURCHASE OVER $50!
FRESH PRIME NUMBERS DAILY!
13
ALL ARITHMETIC DONE
WHILE-U-WAIT!
R. Chast