THE BRAINY SECRETS OF MUSIC'S POWER • THE PHOTONIC CONNECTION

SCIENTIFIC AMERICAN

NOVEMBER 2004
WWW.SCIAM.COM

The Major Flaws of the New Missile Shield

BLACK HOLE COMPUTERS

Stephen Hawking Was Wrong. Matter Goes In. Answers Come Out

Growing New Parts for Damaged Hearts

The Truth about Sudden Climate Change

november 2004

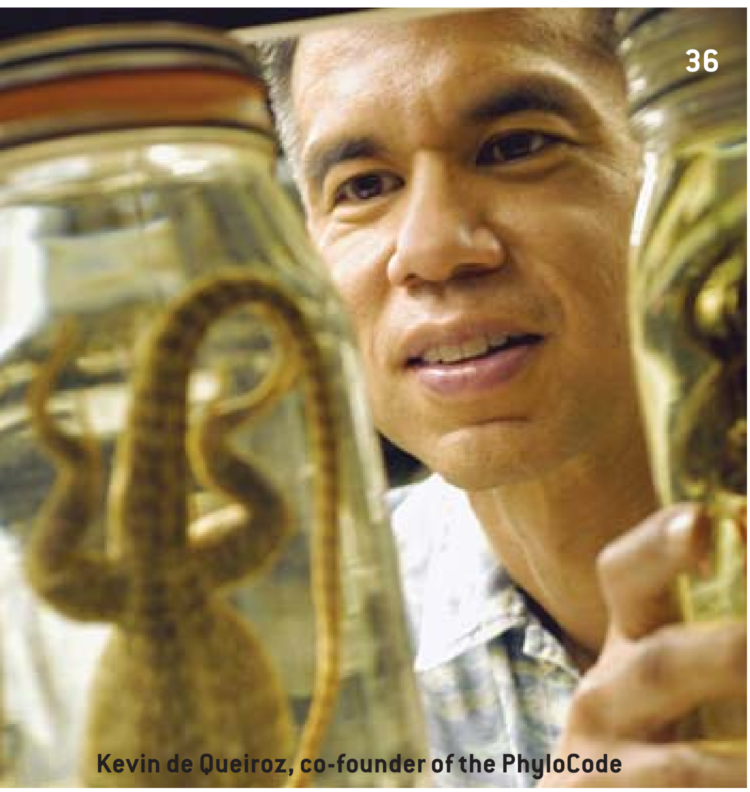
contents

SCIENTIFIC AMERICAN Volume 291 Number 5

features

62 Climate can get unstable

departments



Kevin de Queiroz, co-founder of the PhyloCode

columns

Cover image by Kenn Brown; photograph by Kay Chernush (*at left*)

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. 40012504. Canadian BN No. 127387652RT; QST No. Q1015332537. Publication Mail Agreement #40012504. Return undeliverable mail to Scientific American, P.O. Box 819, Stn Main, Markham, ON L3P 8A2. Subscription rates: one year $34.97, Canada $49 USD, International $55 USD. 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.

Political Science

Since this column began three years ago, *Scientific American* has commented on President George W. Bush's science-related policies—often, though not always, critically. Many readers have complained that we shouldn't mix politics and science. If only it were within our power to keep them separate. Science bears on some of the most important issues of the day, and increasingly, politics bears on how science is done. Although science is not, and should not be, the only factor guiding our choice on November 2, it is up to scientifically literate citizens to take it into account.

Embryonic stem cells. This issue brings out the sharpest disagreements between the candidates. Bush all but banned federal funding for studying these cells and opposes both therapeutic and reproductive cloning. Senator John Kerry of Massachusetts vows to lift the restrictions and to allow therapeutic cloning with ethical oversight.

Climate change. Bush now accepts, as Kerry long has, that the planet is warming and that humans are at least partly responsible. Both men disavow the Kyoto Protocol. What distinguishes them is where to go from here. Bush has set modest goals for industry to meet (or not) voluntarily. Kerry calls for mandatory emissions restrictions, higher car fuel-economy standards and renewed negotiations on international environmental treaties.

National missile defense. Both candidates support it. Bush's advocacy has been more forceful and consistent; a prototype system is now being deployed. Although physicists and engineers fault the design [see "Holes in the Missile Shield," by Richard L. Garwin, on page 70], Bush argues that an imperfect system is better than none and will improve with time. Kerry asserts that deployment is premature and supports more rigorous testing.

AIDS. Both candidates deserve credit. Bush committed the country to the first substantial global initiative: $15 billion over five years. Kerry has a distinguished record on the subject, often working with Senator Bill Frist of Tennessee, now the Senate majority leader. The differences lie in how to spend the money. Bush's initiative requires organizations to buy brand-name drugs, rather than cheaper generics, and to emphasize sexual abstinence even if other approaches would save more lives. Kerry promises more money with fewer strings.

Science policy. Bush speaks highly of science and has increased research funding. The fields of defense, medicine, space and nanotechnology have done well, basic science less so. But his five-year budget plan imposes cuts. And many scientists say his administration has suppressed scientific reports and stuffed technical committees with ideologues or lobbyists. Kerry has echoed these critiques. He calls for increases in funding, although his overall budget numbers don't add up any better than Bush's do.

We hope that whoever wins will reestablish a tradition of nonpartisan thinking about science. Both the political left and right have been guilty of ignoring scientific findings they don't like and elevating their gut feelings above the judgment of people who spend their lives studying a subject. It's one thing to oppose embryonic cell research on moral grounds, and quite another to deny, as have some opponents, its great potential for curing disease. It's one thing to consider global warming a low priority compared with other problems, and quite another to deny that it is a problem at all. It's one thing to want a missile shield, and quite another to deny scientific critiques that could make it better. Science may not be the only factor in decision making, but the dilemmas and tradeoffs it identifies should be faced squarely.

THE EDITORS editors@sciam.com

PHOTOGRAPH BY RON CHAPPEL *Thinkstock/Getty Images*; PHOTOILLUSTRATION BY EMILY HARRISON

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*:
www.sciam.com/permissions
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

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

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

Southeast
Publicitas North America, Inc.
404-262-9218
fax: 404-262-3746

Southwest
Publicitas North America, Inc.
972-386-6186
fax: 972-233-9819

Direct Response
Special Aditions
914-461-3269
fax: 914-461-3433

Belgium
Publicitas Media S.A.
+32-(0)2-639-8420
fax: +32-(0)2-639-8430

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

France and Switzerland
PEM-PEMA
+33-1-46-37-2117
fax: +33-1-47-38-6329

Germany
Publicitas Germany GmbH
+49-211-862-092-0
fax: +49-211-862-092-21

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

India
Convergence Media
+91-22-2414-4808
fax: +91-22-2414-5594

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

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

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

Sweden
Publicitas Nordic AB
+46-8-442-7050
fax: +46-8-442-7059

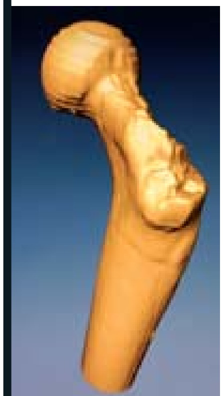
U.K.
The Powers Turner Group
+44-207-592-8331
fax: +44-207-630-9922

On the Web

WWW.SCIAM.COM

UPDATED EVERY WEEKDAY

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



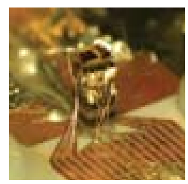
Case for Ancient Upright-Walking Ancestor Gets Legs

Researchers say that a new fossil analysis bolsters the theory that a chimp-size primate that lived in Kenya's Tugen Hills some six million years ago walked on two legs. As such, the creature, known as *Orrorin tugenensis*, may be one of the earliest human forebears on record.

Long Thought Immune, Felines Contract Bird Flu

The threat from avian influenza may be broadening. Surprising evidence shows that one subtype of the virus can infect domestic cats, which health officials have long considered immune to bird flus. The fact that they are not could spell trouble for humans, too.

Researchers Unveil Smallest Atomic Clock Yet



Scientists have manufactured the world's tiniest atomic clock, with inner machinery about the size of a grain of rice. Requiring very little power to run, the device loses only one second every 300 years and could one day provide precise timekeeping for portable applications such as wireless communications devices and Global Positioning System (GPS) receivers.

Ask the Experts

Does damp or wet weather really make arthritis pain worse? If so, how?

Donald A. Redelmeier, a professor of medicine at the University of Toronto, explains.

Exclusive Online Issue: Amazing Animals

A menagerie of marvels—warring ants, cultured chimps, brainy birds, deadly dragons

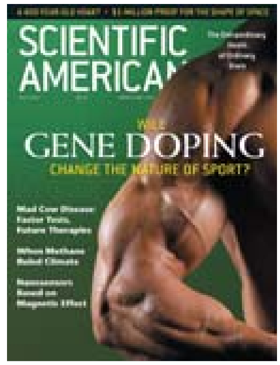
Purchase today for $5 at Scientific American Digital

www.sciamdigital.com

GREG GRIECO *Pennsylvania State University (femur)*; NIST *(atomic clock)*

Letters

EDITORS@SCIAM.COM



IF GENE DOPING and a 500-year-old mystery manuscript weren't intriguing enough, July's issue included nothing less than God and taxes.

In "God's Number Is Up" [Skeptic], Michael Shermer used Bayesian probabilities to peg the likelihood of a supreme being at 2 percent, which led William Orem of Waltham, Mass., to write, "Until a theist presents some credible evidence for the existence of his or her particular god, atheism is the rational position of default. As long as gods stay in the magical realm, the question of their existence is not scientifically insoluble; it is a nonquestion." John Walsh of Arlington, Tex., registered his amazement that "Shermer finds the probability of the existence of God a mere mathematical puzzle, while the dogged search for dark matter and dark energy based on inference gains popularity." And although no one denies the existence of taxes, it is just as certain that everyone will have an opinion about them, as with genetic modification.

TAX ATTACK

Over the past couple of years, I've noticed that some of your articles have been crossing the line between objective science and subjective politics.

This month's By the Numbers column ("Undercutting Fairness") by Rodger Doyle is an example. The unabashed bent that the author displays toward "progressive" taxation seems out of place in *Scientific American*.

Whereas the appropriateness, the practicality and the politics of progressive versus flat versus regressive taxation might promote a lively debate, using loaded words like "fairness" and starting the article from the foregone conclusion that a progressive tax is best leaves a bad taste in my mouth. Please keep politics out of my favorite magazine.

Kevin Kretz
Bernardsville, N.J.

Doyle describes the federal income tax as "efficient and mostly fair, thanks to its progressive rates." Is he kidding? First of all, what definition of "efficient" could one possibly use to describe a complex system of tax codes that costs taxpayers $8 billion annually just to operate the Internal Revenue Service and pay its approximately 116,000 employees, along with another estimated $225 billion for compliance?

Second, whether or not progressive rates are "fair" is clearly a matter of opinion. Considering an individual's tax contribution only as a percentage of income distorts the simple reality that high earners contribute a lot in taxes and buy a lot of goods and services (which also, by the way, provides work for other people).

Becky Brown
Altamonte Springs, Fla.

You assumed progressive tax rates are "fair." Can you provide a basis for that assumption? We generally don't give discounts, based on income, for other services or admission fees. Isn't that fair?

Also, can you provide (by the numbers) some measure of "progressivity" versus other measures of "good outcomes," such as rates of college graduates, income, reading levels, SAT scores, water quality, divorce rates and so on by comparing states or other localities?

Bob Smith
Oak Ridge, Tenn.

DESIGNER GENES

Regarding "Gene Doping," by H. Lee Sweeney: it was disappointing to see this article emphasize the possible abuse by athletes of genetically enhanced muscles. Wouldn't it have been better to speak in terms of the benefits to all those who suffer from muscular dystrophy and aging muscles? You're seeing the glass half-empty instead of half-full.

Susan Godzac
Erie, Pa.

Sweeney's article on gene doping raised several questions for me. The first was entirely selfish: Can he develop a gene therapy to repair my torn knee cartilage and ligaments?

Other questions of more relevance to society followed. What effect will gene doping have on warfare and on law enforcement? It seems entirely predictable that nations will attempt to produce super soldiers, and those intending to break the law will try to use gene doping to gain an advantage over citizens and law enforcers. These activities will inevitably lead to "arms races" among the medical–scientific establishments of nations to produce ever more invincible soldiers. There is also reason to expect that a different sort of arms race will develop between criminal organizations (and the medical personnel they can hire or blackmail) and physicians working with the justice system.

What changes and adjustments will these developments require of society, law and individuals? It seems prudent to begin thinking about these matters now.

Robert Henley
Diamond Springs, Calif.

When I read Sweeney's article, I was enthused about the possibilities of the muscle-growth gene therapy. If athletes want to abuse the potential, so what? The possible benefits far outweigh the fact that sports will be changed forever. I have to believe that even the most ardent sports fan would prefer that his grandparents enjoy a physically active life for their entire span instead of spending years as a prisoner in their weakened bodies.

Nathan Bell
Leetonia, Ohio

SPACE REFLECTIONS

Regarding "The Shapes of Space," by Graham P. Collins: If the universe is a 3-sphere, shouldn't an astronomer looking into very deep space see the same distant galaxies as the ones that would be seen by looking in the exact opposite direction, that is, the deep sky looking exactly celestial north should be the same (mirror reversed) as that seen looking exactly celestial south? I have often wondered if this is so, and perhaps modern instruments are powerful enough that this can be checked. Has anyone done this? Maybe time has to be included in the real universe (making a 4-sphere), which makes it more complex—does anyone know?

Robin A. Cox
Scarborough, Ontario

COLLINS REPLIES: Astronomers have indeed looked for such patterns, but none have been detected. Last May a group reported finding no evidence of such repetitions in the microwave background, implying that the universe extends for at least about 75 billion light-years before it can curve back on itself. In October 2003, however, another group reported evidence, based on a different analysis of the microwave background, that the universe could be like a dodecahedron of a diameter of about 90 billion light-years, with opposite faces identified (a space known as the Poincaré dodecahedral space).

These issues are also discussed in "Is Space Finite?" by Jean-Pierre Luminet, Glenn D. Starkman and Jeffrey R. Weeks; SCIENTIFIC AMERICAN, April 1999, updated in "The Once and Future Cosmos," SCIENTIFIC AMERICAN Special Issue, Vol. 12, No. 2; 2002.

GENE THERAPY could duplicate the effects of a natural mutation that blocks the antigrowth factor myostatin, resulting here in "double-muscled," exceptionally lean cattle.

YANN ARTHUS-BERTRAND *Corbis*

CEREBRATING VOYNICH

I was delighted by Gordon Rugg's approach in "The Mystery of the Voynich Manuscript." Rugg's attention to the history of the acceptance of certain views underlines the social and personal forces that have shaped the sciences.

We must pay attention to the politics and provenance of "fact," the process of "expert reasoning," as Rugg puts it. Increasingly, we must join the perspectives of available disciplines: biology, psychology, history, linguistics and many other fields. Only such cooperation can pry truth from process and allow us in some small way to escape our wiring.

Certainly "forms of sensibility," as Kant put it, affect our knowledge. Because we are in a better position than any other age to examine such forms and categories, we should do so in a more systematic way. I hope a reliance on such investigations will become a larger element in our scientific literature.

John J. Ronan
Magnolia, Mass.

ERRATA In "Big Air," by Mark Fischetti [Working Knowledge], the location of the John Wanamaker Grand Court Organ in Philadelphia is Lord & Taylor, not Hecht's.

In "When Methane Made Climate," by James F. Kasting, the photograph of Methanosarcinales should be credited to Everly Conway de Macario and Alberto J. L. Macario.

50, 100 & 150 Years Ago
FROM *SCIENTIFIC AMERICAN*

Puzzling Courtship ▪ Mysterious Aether ▪ Mythic Life-Forms

NOVEMBER 1954

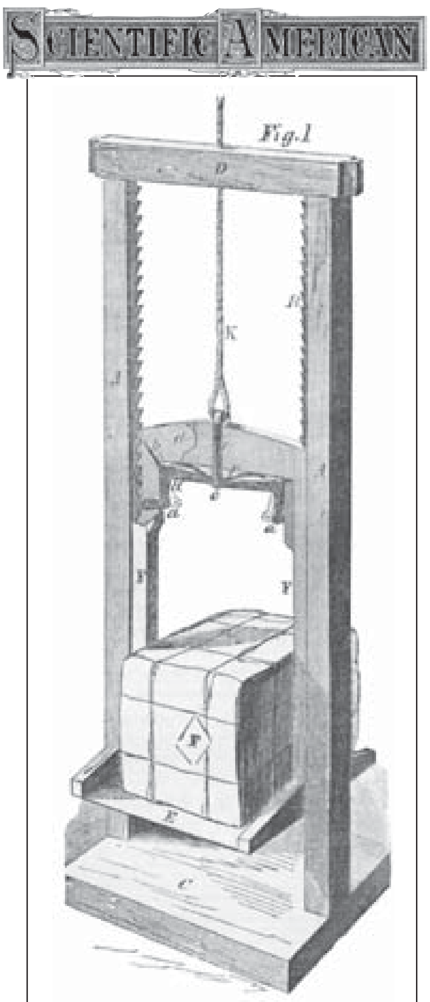
THE COURTSHIP OF ANIMALS—"When a golden pheasant cock displays his brilliant plumage before the hen, we are accustomed to say he is courting her. Just what this expression means when applied to a nonhuman animal is far from clear; the idea is so obviously anthropomorphic that zoologists have been reluctant to pursue it seriously by taking up the study of animals' so-called 'courtship' activities. Yet these strange, often grotesque activities are there, like Mount Everest, and they have challenged some of us to explore them. In contrast to such clearly motivated behavior as feeding or flight from predators, the courtship postures of animals are altogether puzzling, because it is difficult to see not only what circumstances cause them but what functions they serve. —N. Tinbergen" [*Editors' note: Nikolaas Tinbergen won the 1973 Nobel Prize in Physiology or Medicine.*]

ANTIBODIES—"My fellow Australian microbiologist Frank Fenner and I suggested in 1949 a new theory of antibody production which has been called 'the self-marker hypothesis.' It is an untidy theory, because it postulates happenings for which there is only the slightest evidence. In this theory there are in the scavenger cells of the body (technically the cells of the reticulo-endothelial system) certain structures which we call 'recognition units.' These units can recognize a certain fairly limited range of 'self-markers.' Recognition comes simply by virtue of a complementary structural pattern—like the lock-and-key relationship between an enzyme and the molecule with which it interlocks. —Sir Macfarlane Burnet" [*Editors' note: Burnet won the 1960 Nobel Prize in Physiology or Medicine.*]

NOVEMBER 1904

THE COMPLEAT HI-TECH ANGLER—"Steel fishing rods have been brought to such a state of perfection that they are now being sold extensively in place of those of bamboo. It is said that they are handier to carry and are better balanced, and much more sensitive, and can be weighted to suit the most fastidious taste. These rods are made of the finest tempered steel tubing, japanned."

COSMIC ETHER—"Miss Agnes Clerke writes in Knowledge, 'The glory of the heavens is transitory, but the impalpable, invisible ether inconceivably remains. Such as it is to-day, it already was when the *Fiat Lux* was spoken; its beginning must have been coeval with that of time. It is evasive of common notice, while obtrusive to delicate scrutiny. It does not perceptibly arrest, absorb, or scatter light. Looking, however, below the surface of things, we find the semi-fabulous quintessence to be unobtrusively doing all the world's work. The potencies of matter are rooted in it; the substance of matter is latent in it.'"



THE OTIS ELEVATOR, for safety, 1854

NOVEMBER 1854

ARE WE ALONE?—"Sir David Brewster, who supposes the stars to be inhabited, as being 'the hope of the Christian,' asks, 'is it necessary that an immortal soul be hung upon a skeleton of bone; must it see with two eyes, and rest on a duality of limbs? May it not rest in a Polyphemus with one eye ball, or an Argus with a hundred? May it not reign in the giant forms of the Titans, and direct the hundred hands of Briareus?' Supposing it were true, what has that to do with the hope of the Christian? Nothing at all. This speculating in the physical sciences, independent of any solid proofs one way or the other, and dragging in religion into such controversies, neither honors the Author of religion, nor adds a single laurel to the chaplet of the sciences; nor will we ever be able to tell whether Mars or Jupiter contain a single living object."

THE SAFETY ELEVATOR—"The annexed figure is a perspective view of the improved elevator of Elisha G. Otis, of Yonkers, N.Y., who has taken measures to secure a patent. A platform is attached to a frame which works between two vertical racks, the upper part of the frame having pawls, which catch into the racks when the lifting power is taken off. By this improvement, if the rope should break, the platform will be sustained, and no injury or accident can possibly occur, as the weight is prevented from falling. This excellent platform elevator was on exhibition in the Crystal Palace during the past season, and was much admired."

news SCAN

POLICY

Damage Control

A CRACKDOWN TO PREVENT CONFLICTS OF INTEREST AT THE NIH **BY DAVID LABRADOR**

"Drastic changes" is how Elias A. Zerhouni, director of the National Institutes of Health, described new rules to address allegations of conflicts of interest at the renowned U.S. agency. Scientists, watchdog groups and the 18,000 employees of the NIH are hoping that the limits, announced in June but still being refined, will quell the scandal, foster impartial research and reclaim the public's trust—without driving away top-notch talent.

UNDER SIEGE: National Institutes of Health director Elias A. Zerhouni testified before Congress several times (shown here in January) to address charges against the agency.

DENNIS COOK *AP Photo*

The roots of the current trouble reach back to 1995, when then director Harold Varmus lifted restrictions on outside collaboration in a bid to attract top researchers. The freedom allowed staff to work for, own stock in and even serve on the boards of outside companies that they didn't deal with directly as officers of the government, making the agency a much more desirable workplace. Even so, says Paul W. Kincade, president of the Federation of American Societies for Experimental Biology, "it's important to realize that top scientists like Dr. Zerhouni forgo many thousands of dollars in salary to hold a government job at the NIH." Permitting outside income provided an incentive, and disclosure and approval rules seemed adequate to prevent abuse.

Then, in December 2003, the *Los Angeles Times* reported that many NIH scientists, including institute directors, had been receiving improper consulting payments from drug companies. The fees amounted to millions of dollars over 10 years.

Congressional investigations soon alleged more serious abuses. A clinical trial of a company's drug was apparently derailed shortly after the researcher in charge of the trial received consulting fees from a competing firm. The NIH's National Cancer Institute paid a large legal settlement for a university shortly before the director who approved the payment won a substantial cash award from the very same school. More broadly, a survey of a few corporations revealed that more than one third of the consulting deals between drug companies and agency researchers were never reported to the NIH.

"The conflict-of-interest rules were greatly diminished or weakened," argues Judith

news SCAN

RESTORING INTEGRITY

The National Institutes of Health announced new standards to prevent conflicts of interest, which will probably be strengthened at the behest of the Office of Government Ethics. Among the restrictions:

- **Outside consulting is now forbidden for senior NIH staff, including directors and deputy directors. All other researchers, including the heads of laboratories, are exempt.**
- **No employee can serve on a board or receive stock in a drug or biotechnology company as compensation. About 5,000 individuals (out of 18,000) cannot own any stock in relevant industries.**
- **Employees who are permitted outside income must limit their earnings to 25 percent of their government salary and cannot spend more than 400 hours a year moonlighting.**
- **No NIH employee may accept an award or begin a collaboration without approval of a new, central NIH ethics committee.**

S. Bond, president of the American Society for Biochemistry and Molecular Biology, "and there was little effort to enforce those rules." Vera Hassner Sharav, president of the Alliance for Human Research Protection, a group advancing medical ethics and responsibility, puts it more bluntly: "The culture was that they had carte blanche to do as they do and the more research the better—it didn't matter how, what or why. They were simply turning a blind eye and a deaf ear."

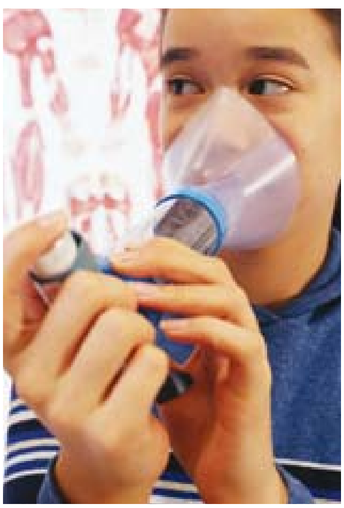
Zerhouni, who has been director of the NIH since May 2002, proposed his drastic changes at the end of June. They severely limit who at the NIH can collaborate with private industry and how much time and money can accrue to those who do. Although most researchers would still be allowed to work with drug companies, a single committee must now approve all outside collaboration before it can begin. "Dr. Zerhouni took prompt action, and I think that action is going to go a long way to demonstrating that when there are charges of wrongdoing that are merited, they are even more serious to the scientists than they are to the public," Kincade says.

Others do not think Zerhouni's proposals went far enough. "If they're not mandated and overseen, then they won't change anything," Sharav says. Merrill Goozner, project director at the Center for Science in the Public Interest, a consumer advocacy group based in Washington, D.C., thinks that "there is a good scientific justification—given the NIH's mission—to actually ban all outside consulting and contracting on the part of NIH scientists."

Such concerns have registered at the Office of Government Ethics, which must approve the new regulations. The office's acting director, Marilyn L. Glynn, wrote in a 20-page letter that she contemplated the possibility of an "absolute prohibition on consulting with drug companies." She wrote that the proposals that have been announced to date "could give the appearance that some level of misuse of office is tolerable." Zerhouni's office is working with Glynn's office to develop the "supplemental standards of conduct" that Glynn has called for. Still, Kincade worries that even the restrictions proposed last June "could compromise the ability of the NIH to attract and retain its most valuable assets"—namely, the scientists.

The NIH, which in September imposed a one-year moratorium on consulting, should clarify the rules in 2005. But ultimately these issues will have to be decided by the scientists themselves. Government posts will never pay as well as drug company jobs. The most the NIH can do is to set clear standards and to enforce them—and hope its employees continue to choose public service over private gain.

David Labrador is a freelance writer and researcher based in New York City.



ASTHMA and other allergies may stem from overly hygienic conditions.

LWA-STEPHEN WELSTEAD *Corbis*

ALLERGY

Breathing with Hepatitis

DOES EXPOSURE TO A LIVER-INFLAMING VIRUS PREVENT ASTHMA? BY LISA MELTON

According to the hygiene hypothesis, the soaring rates of asthma, hay fever, eczema and other allergies in the past two decades have resulted from the overly sanitized conditions of industrial countries. Because children are exposed to fewer bacteria and viruses, the theory goes, their immune systems tend to overreact to otherwise harmless substances such as pollen and dander.

The hypothesis, however, fails to explain why some people are more susceptible than others or why those in dirty environments still develop asthma. But now a genetic study has pointed out a plausible mechanism for allergy development: it suggests that the hepatitis A virus, which thrives in polluted environments, may protect people from asthma.

Recognizing that allergies tend to run in families, a team led by Dale T. Umetsu and Rosemarie DeKruyff at the Stanford University School of Medicine searched for a genetic component. "We knew that finding a susceptibility gene for asthma in humans would be a formidable task, so we decided to simplify the problem and use a mouse model," Umetsu says. Their studies cast up one gene, *TIM-1,* that predisposed the mice to asthma.

But the researchers got more than they bargained for. "*TIM-1* is also the receptor used by the hepatitis A virus to infect human cells," Umetsu explained at a Novartis Foundation symposium in London this past June. This was a crucial discovery, because Paolo Matricardi, now at Bambino Gesú Children's Hospital in Rome, had found that allergies occur much less often in people exposed to the hepatitis A virus. Hepatitis A spreads by exposure to the stool of an infected person and causes jaundice and flulike symptoms; the illness usually clears up on its own.

The Stanford team found that humans carry either a long or a short variant of *TIM-1*. In a study of 375 people, the researchers saw that those who carried the long version of *TIM-1* and who had been infected with the hepatitis A virus were one fourth as likely to suffer from asthma as those with the shorter version. (Overall about 7 to 12 percent of U.S. children develop asthma.) Protection seems to hinge on two crucial factors: inheriting the right version of the *TIM-1* gene and succumbing to hepatitis A.

These tantalizing results highlight the importance of genetic interactions with the environment in the development of asthma. In the U.S., because nearly two thirds of whites and blacks and almost half of Asians carry the protective version of *TIM-1,* a spell of hepatitis A infection might keep allergies at bay. But whereas before the 1970s virtually everyone had been infected with the virus, today's healthier and more sanitized existence means that only 25 to 30 percent of people in developed countries have been exposed.

If hygiene is key, why do children in inner cities suffer from severe asthma? One of the leading theories is that exposure to cockroach feces triggers the problem, which seems to contradict the hepatitis A theory and the hygiene hypothesis. But Matricardi points out that in the 20th century fecal contamination has also declined in depressed areas, just as it has everywhere else in the U.S. "These children are unlucky: they now have the susceptibility to develop allergies and are still exposed to cockroach and mouse allergens, dust mites and cigarette smoke," he remarks.

"Exposure to hepatitis A is one possible

ASTHMA ON THE RISE

- **Number of asthmatics worldwide:** 100 million to 150 million
- **Number of asthma sufferers in the U.S.:** 17 million
- **Number in the U.S. in 1980:** 7 million
- **Percent of people affected by asthma in industrial countries:** 10 to 20
- **Average increase per decade, worldwide:** 50 percent

SOURCES: World Health Organization; National Institutes of Health; Centers for Disease Control and Prevention. Data are from 2000 and 2001.

mechanism for how the hygiene hypothesis works," says geneticist William Cookson of the Wellcome Trust Center for Human Genetics in Oxford, England. Umetsu's group has produced "interesting work, but the study needs follow-up," Cookson adds. "It would be nice to see it replicated and generalized." Of course, nobody would choose to go back to the bad old days of dubious drainage and rampant infections to fend off asthma. The Stanford team is currently testing whether vaccination will do the trick instead.

Lisa Melton is based in London.

SENSORS

Volcanic Sniffing

QUANTUM-CASCADE LASER MAY DETECT IMPENDING ERUPTIONS **BY CHARLES CHOI**

In A.D. 79 Mount Vesuvius erupted, annihilating the cities of Pompeii and Herculaneum and killing thousands who did not evacuate in time. To avert a similar fate for present-day Naples, which lies six miles west of the still active Vesuvius, as well as for the cities near volatile Mount Etna in Sicily, a novel laser system could soon forecast volcanic eruptions up to months in advance.

Current methods to predict eruptions have downsides. Seismometers can monitor tremors and other ground activity that signal a volcano's awakening, but their readings can prove imprecise or complicated to interpret. Scanning for escaping gases can reveal whether magma is moving inside, but the instruments used to analyze such emissions are often too delicate and bulky for life outside a laboratory. "You have to collect samples from the volcano, bring them to a lab, and often wait through backlogs of weeks to months before analysis," explains Frank Tittel, an applied physicist at Rice University.

A more promising technique for early detection focuses on changes in carbon isotopes in carbon dioxide. The ratio between carbon 12 and carbon 13 is roughly 90 to one in the atmosphere, but it can differ appreciably in volcanic gases. A ratio change by as little as 0.1 part per million could signal an influx of carbon dioxide from magma either building under or rising up through the volcano.

Lasers can help detect this change: carbon 12 and 13 absorb light at slightly different mid-infrared wavelengths. The lasers must continuously tune across these wavelengths. Previously investigators used lead-salt lasers, which require liquid-nitrogen cooling and thus are impractical in the field. Furthermore, they are low-power devices, generating less than millionths of a watt, and can emit frequencies in an unstable manner. Other isotope scanning techniques are similarly lab-bound.

Tittel and other scientists in the U.S. and Britain, in partnership with the Italian government, have devised a volcano-monitoring system around a quantum-cascade laser. Such a semiconductor laser can produce high power across a wide frequency. Moreover, they are rugged and do not require liquid-nitrogen

MOUNT VESUVIUS, shown here in 1944, still threatens Naples.

BETTMANN/CORBIS

A WATERFALL OF ELECTRONS

Quantum-cascade lasers consist of thin, nanoscale layers of semiconducting materials that provide several energy levels to an electron. An excited electron cascades down the levels to lose energy, emitting a laser photon at each step. In this way, a single electron can emit dozens of photons, making quantum-cascade lasers more powerful than standard semiconductor lasers, in which only one photon per excited electron is emitted. Moreover, by adjusting the size of the layers during fabrication, researchers can make an electron emit photons of different frequencies. The combination of high power, wide-frequency bandwidth and compact size makes the quantum-cascade laser ideal for volcano monitoring.

cooling, making them compact enough to fit inside a shoe box.

The researchers first tried out their device on gas emissions from Nicaraguan craters in 2000. The new field tests will check its performance and accuracy in harsh volcanic locales. Dirk Richter, a research engineer at the National Center for Atmospheric Research in Boulder, Colo., says it would prove difficult to design a system "to work in one of the worst and most challenging environments possible on earth," but "if there's one group in the world that dares to do this, that's Frank Tittel's group."

If the instrument works, the plan is to deploy early-warning systems of lasers around volcanoes, with each device transmitting data in real time. False alarms should not occur, because carbon isotope ratios in magma differ significantly from those in the crust. The changes that the laser helps to detect also take place over weeks to months, providing time to compare data from other instruments, as well as ample evacuation notice. "Our system aims at avoiding a catastrophe like the Vesuvius eruption," says team member Damien Weidmann, a physicist at the Rutherford Appleton Laboratory in Oxfordshire, England. Field tests for the prototype are planned for the spring of 2005 in the volcanic Alban Hills region southeast of Rome, near the summer home of Pope John Paul II, as well as for volcanic areas near Los Alamos, N.M.

Charles Choi is a freelance writer based in New York City.

SPACE

Flawed Revelations?

CONTAMINATION MAY UNDERMINE GENESIS DATA **BY BARRY E. DIGREGORIO**

After traveling 1.5 million kilometers beyond Earth to obtain bits of the solar wind, NASA's first automated sample-return mission, Genesis, ended in a crash in the Utah desert on September 8. Researchers do not know just why the parafoil failed to deploy, but they say they feel confident that they can still accomplish the major goals of the mission despite the damaged capsule. Any conclusion stemming from the mission, however, may remain dubious because of the mere possibility of contamination.

Genesis had onboard an estimated 20 micrograms of solar-wind particles

LOOK OUT BELOW: Genesis capsule lay in a shallow crater shortly after impact.

NASA/JPL; USAF 388 RANGE SQUADRON

collected over three years in space. These particles came from the sun's visible surface, called the photosphere. Ninety-nine percent of it consists of the original material from the primitive nebula that coalesced into the sun and planets; an analysis of the samples would therefore provide a baseline of information about the building blocks of our solar system.

The original retrieval plan had called for a helicopter to snag Genesis's parachute in midair. Once secured, the capsule would have been placed in a specialized shipping container and transported to a clean room at the NASA Johnson Space Center in Houston.

But because the parafoil failed to open, Genesis slammed into the desert at 311 kilometers an hour, half-burying itself. On impact, sand, dust and other contaminants entered the capsule. Engineers carefully excavated it and brought it to the nearby U.S. Army Dugway Proving Grounds, where an initial inspection with mirrors revealed a 7.6-centimeter-wide hole in the sample container. Peering inside revealed more bad news: many of the hexagonal wafers, which collected the solar-wind samples, were broken. (Genesis carried a total of 275 sample wafers, each 10 centimeters on a side, placed on five collector arrays.) Broken up, they were even more directly exposed to contamination.

Roger Wiens, project flight payload leader for Genesis, notes that "this kind of contamination can overwhelm the amount of signal we have in the solar-wind sample. We will certainly be doing some cleaning of the samples, probably with ultrapure water." Wiens says that the particles get embedded in the wafers to a depth of about 50 nanometers and that the penetration might provide some protective distance from small amounts of surface contamination. Bruce Barraclough, a principal investigator on the Genesis mission, remains optimistic: "I am sure that some, if not most, of the science will be able to be recovered from the samples."

Gilbert V. Levin, a former astrobiologist on two NASA Mars missions,

thinks the Genesis team is mistaken. "It is unlikely that any such sample recovery will be without controversy," remarks Levin, now chairman of Spherix, a Beltsville, Md., biotechnology company. He points to the Martian meteorite ALH 84001 as an example. In 1996 some scientists concluded that fossilized forms of unusual bacteria were inside; however, other researchers attributed the forms to earthly contamination. The Martian debate "is a strong harbinger of the same issue arising with the Genesis material," Levin says.

Meanwhile determining the cause of the crash will take several months, and how the incident might affect future sample-return missions is unclear. An investigation might prove useful for NASA's Stardust craft, which is carrying precious samples of comet Wild 2. It will land on Earth on January 15, 2006. Although no midair retrieval is planned—Stardust is less sensitive to shock than Genesis—the capsule will still rely on a parachute. According to Stardust principal investigator Donald Brownlee, the sample-return capsule has parachute-releasing pyrotechnic charges similar to those that failed on Genesis. "One thing the Genesis team is looking at is the temperature readings from the battery area on the spacecraft while it was gathering samples," Brownlee states. The sun may have cooked the batteries, causing them to malfunction. So far the Stardust team has not observed any temperature anomalies with the craft's batteries. Concerns will remain high, however, until the Genesis investigation board finds out exactly what went wrong.

Barry E. DiGregorio is an astrobiologist for the Cardiff Center for Astrobiology in Wales.

COMPUTING

Connection Blues

A HOLE FOR EXTERNAL CONTROL OF BLUETOOTH DEVICES BY WENDY M. GROSSMAN

My mobile phone, lying on the table in front of me, flashes "Connecting" a couple of times and then falls back to blank normality. Adam Laurie looks up from his laptop and says, "Do you have a phone book entry 'marca03?' "

Yes, I do.

Laurie, a security expert, co-organizer of the annual hacker conference Defcon and head of the London-based data security company AL Digital, has just Bluesnarfed my phone. That is, he's hacked my phone's Bluetooth connection to demonstrate that he can access my information without my knowledge or consent. This flaw exists in many manufacturers' Bluetooth devices, and it represents an increasing danger as mobile phones become all-purpose communicators that can handle payments and banking transactions.

As a personal networking standard, Bluetooth allows devices to connect to one another over short distances. Bluetooth replaces cables and infrared connections, enabling computers, cell phones, PDAs, keyboards, printers and other devices to communicate with one another. (It is not to be confused with 802.11, otherwise known as Wi-Fi, which permits wireless Internet and local-area networking.)

BLUETOOTH computer can swipe cell phone data.

THE IMAGE BANK/GETTY IMAGES

The creators of Bluetooth were conscientious about security. Data in transit are encrypted. Depending on the built-in features, a Bluetooth connection can often be configured so that the device talks only to specified other devices and is not discoverable except by them. The problem is, this setting is not always available or easy to use. Just like Wi-Fi networks

OPENING UP WITH A HANDSHAKE

Keeping connections open is the main reason why Bluesnarfing—the hijacking of a Bluetooth device—is possible. (Activating a unit's security setting, such as its "nondiscoverable" mode, may deter casual attacks but not determined assaults.) Most likely Bluesnarfing will become more common thanks to the current fad of Bluejacking, an exchange of messages between anonymous strangers that depends on Bluetooth remaining on. Before pairing up, Bluetooth devices authenticate themselves to each other via a "handshake"—presenting a "name" of up to 248 characters. Bluejacking substitutes messages for these names. But passing the name opens up the device—and therein lies the security hole. People often do not realize that if their device pairs with another, then that remote device has access to all the data stored on the phone as long as they keep Bluetooth on.

in residential neighborhoods, many Bluetooth connections are left open and vulnerable.

In his attack, Laurie convinced my phone that it was paired with his laptop, even though his laptop does not appear on my list of authenticated devices. He has made use of the fact that Bluetooth devices have a common standard. Bluetooth serves as the conduit for familiar services—such as voice, file transfer, printing and faxing—and relies on customized sets of protocols referred to as profiles. Laurie will not say exactly how he exploits the profiles, but he does explain that he is using Bluetooth to access flaws in the manufacturers' implementation of those services. He adds that most of the necessary software for his eavesdropping is readily available on the Internet and otherwise has legitimate purposes, such as utilities for data backup and short message service (SMS) text.

To most people, the data at risk don't sound like much at first. "People think it doesn't matter," Laurie says, "but usually they find a few entries in their phones they don't want the world to see." This will be even truer as functions and storage space continue to grow to include e-mail, recordings, photographs and other forms of data.

While attempting to duplicate Laurie's work, Martin Herfurt, a researcher at Salzburg Research in Austria, stumbled onto something even worse: Bluebugging. It relies on the same pairing double cross as Bluesnarfing, but it then connects to that device's Bluetooth profile for a serial port—the traditional spot for modem connections. You can then send the "AT" commands familiar from the old dial-up days to take control of the device. Standard utilities enable you to use the phone to call premium rate numbers, send SMS text (which also may be charged at premium rates) and connect to the Internet. You can even get the hijacked phone to call you without the owner's knowledge and thereby listen in on nearby conversations.

Some of the affected manufacturers have fixed their phone software. Meanwhile Laurie is working with the Bluetooth creators to help improve security on the next generation of standards. But the incident is a good reminder of a basic problem: going from cable to wireless adds a whole new layer of invisible risk.

Wendy M. Grossman is based in London.

DATA POINTS: ALL WET

The largest raindrops on record fell in Brazil in 1995 and on the Marshall Islands in 1999. University of Washington researchers now conclude that the Brazil drops, measured after a fire, formed by condensing onto giant smoke particles and that the Marshall Islands drops resulted from collisions with other drops inside a water-heavy cloud. A device that records a raindrop's shadow as it passes through a laser beam indicated the sizes. Despite the typical tear shape attributed to raindrops, the form of the drop actually depends on its size: a spherical drop tends to flatten and distort as it gets bigger.

Average size in millimeters (mm) of a raindrop: 1 to 2

Approximate shape of drop when its diameter is about:
1 mm: sphere
2 mm: hamburger bun
5 mm: parachute

Diameter of the largest recorded:
Raindrop: 8.8 mm
Hailstone: 178 mm

Crop and property damage in 2003 caused by:
Rain: $78.2 million*
Hail: $632.5 million

*Damage from rain excludes flooding.

SOURCES: National Oceanic and Atmospheric Administration; Geophysical Research Letters, July 13, 2004; Wikipedia.org

ENTOMOLOGY

Her Majesty's Secret Service

FOR THE COLONY'S SAKE, workers will crack down on cheaters.

Workers in some species of social insects police one another to prevent individuals from laying their own eggs, as opposed to helping rear the queen's. Researchers assumed that policing is selected for in colonies that contain multiple queens or mates, which means that workers share more genes on average with a queen's offspring than with one another's and so have an incentive to force others to invest in the queen's young. But a survey of research on 50 species of ants, bees and wasps finds that once-mated single queen colonies are just as likely to crack down on cheaters as those with multiple queens or mates. The surveyors, Robert Hammond and Laurent Keller of the University of Lausanne in Switzerland, point out that policing should also arise if unchecked cheating imposes significant costs on the hive, such as workers loafing. The drive for efficiency thus seems to outweigh relatedness in leading to actual policing, Hammond says. The work appears online in the September *Public Library of Science Biology*. —*JR Minkel*

SETI

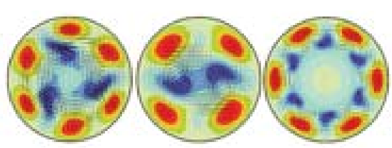
Check the "In" Box

Instead of phoning home, E.T. might be better off writing. Searching for extraterrestrial intelligence typically means exploring the sky for radio messages, but such waves grow weaker as they cross space. Beaming more than a basic note across the stars requires an antenna the size of Earth, calculate electrical engineer Christopher Rose of Rutgers University and physicist Gregory Wright of Antiope Associates in Fair Haven, N.J. On the other hand, they calculate roughly 10 billion billion nanosize bits of data—all Earth's current written and electronic information—could be inscribed via scanning tunneling microscopy within a gram of matter. Interstellar mail could therefore prove far more efficient over long distances, though much slower. In the September 2 *Nature*, Rose and Wright suggest that resting points for parcels could lie in orbits near the Earth, moon, asteroid belt and sun or on surfaces in the inner solar system. —*Charles Choi*

PHYSICS

Piped-in Turbulence

Turn a faucet from low to high, and the initially smooth, clear flow of water breaks and goes cloudy without warning. More than a century after the field's pioneering experiments, the mechanism that creates and feeds such turbulence in pipes is still poorly understood. Simulations have suggested that periodically changing forward-moving eddies might constitute the building blocks of turbulent flow, but no one had observed such fleeting waves. Now a group led by researchers from Delft University of Technology in the Netherlands has done just that, by shining a laser across a turbulent flow and using cameras to track microscopic tracer beads in the water. The beads outlined vortices that pushed sluggish water to the center of the pipe, creating a slow-moving streak down the center and fast-moving streaks around it—the predicted eddies. The September 10 *Science* has more. —*JR Minkel*



WATER through pipes, shown in cross section in these simulations, reveals characteristic eddy patterns of fast (*red*) and slow (*blue*) flow that produce turbulence.

CHRISTIAN KÖNIG (*top*); CASIMIR VAN DOORNE AND BJÖRN HOF, REPRINTED WITH PERMISSION FROM B. HOF ET AL. IN *SCIENCE*, VOL. 305; © 2004, AAAS (*bottom*); ILLUSTRATION BY MATT COLLINS

NEUROBIOLOGY

Ear-y Coincidence

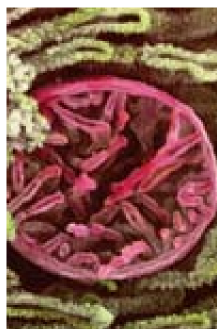
In one example of brain specialization, the left hemisphere usually processes speech sounds, letting the right side handle changes in tone. New work on babies indicates that the ears may conribute to auditory brain specialization. Specifically, the right ear—corresponding to the left hemisphere—has a stronger built-in response to rapid sound changes than the right-brained left ear, which prefers tones. Audiologists Yvonne Sininger of the University of California at Los Angeles and Barbara Cone-Wesson of the University of Arizona exposed infants to clicking noises and tone pairs in both ears and listened for the echolike amplification of acoustic energy produced by hair cells in the cochlea. The right ear tended to amplify the speechlike clicks, and the left ear preferred to amplify tones, even though the brain's auditory regions are undeveloped in infants. Sininger plans to test adults and those with deafness for the effect, described in the September 10 *Science*. —*JR Minkel*

THIS EAR is for tone.

GENETICS

Invasion of the Sequences

Mitochondria, the powerhouses of the cell, have their own DNA, evidence that their progenitors once lived free and colonized cells. They may still be acting as invaders. Investigators at the Pasteur Institute in Paris scanned the human genome in the nucleus and discovered 211 DNA fragments that matched sequences in mitochondria. The researchers propose that these fragments get integrated mistakenly during routine DNA repair work. This genetic invasion appears ongoing—27 fragments are unique to *Homo sapiens* and must have colonized the chromosomes since humans diverged from the rest of the primates in the past four to six million years. The penchant for mitochondrial DNA to lodge in nuclear genes could trigger disease, the researchers say in the September *Public Library of Science Biology*—some fragments mutated tumor-suppressing genes. The invading mitochondrial DNA could also help track human migration and evolution, because ethnic groups possess a unique mix of fragments. —*Charles Choi*



MITOCHONDRIA can act as genetic invaders.

HEALTH

Beating a Deadly Gas

Carbon monoxide kills or injures more people than any other poison. Half of those who survive serious encounters suffer brain damage, but mysteriously, the damage happens days to weeks after the gas clears a victim's blood. Investigators at the University of Pennsylvania Medical Center deduce that the reason for the delay lies with a by-product of a chain reaction triggered by carbon monoxide. This by-product changes the shape and ionic charge of myelin basic protein (MBP), a major ingredient in the protective sheath around nerves. White blood cells attack this altered protein, but they also go after normal MBP molecules, too. Feeding rats MBP before carbon monoxide poisoning averted brain injury by rendering their immune systems tolerant to the protein. The findings could open up novel therapies for carbon monoxide poisoning, as stated in the September 1 online report from the *Proceedings of the National Academy of Sciences USA*. —*Charles Choi*

NICK KELSH *Corbis* (*top*); P. MOTTA AND T. NAGURO/SCIENCE PHOTO LIBRARY (*bottom*)

BRIEF POINTS

- **Hydrocarbon fuel may be more abundant than previously thought and may have nonbiological origins. Iron oxide, calcium carbonate and water turned into methane after being squeezed and heated as they would be in the upper mantle.**

  Proceedings of the National Academy of Sciences USA online, September 13, 2004

- **The impending loss of species may extend well beyond those on the endangered list. Some 6,300 species may disappear because the organisms they depend on die off.**

  Science, September 10, 2004

- **Modifying the backbone structure of an aminocoumarin, a type of antibiotic, with chemical components (amino and methyl groups, sugars and others) can lead to hundreds of new antibiotic varieties.**

  American Chemical Society meeting, August 25, 2004

- **Mice that had their visual signals rerouted to their auditory brain region reacted to fear-inducing light as if they had heard it rather than seen it. Such plasticity shows that emotional responses can be quickly learned by alternative regions.**

  Nature Neuroscience, September 2004

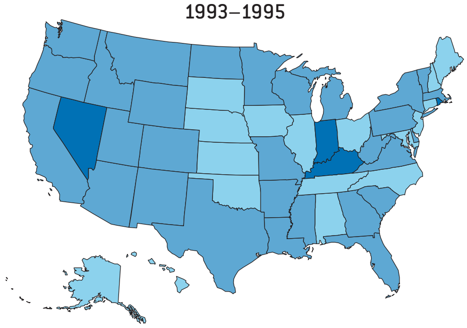
BY THE NUMBERS

Getting Sicker

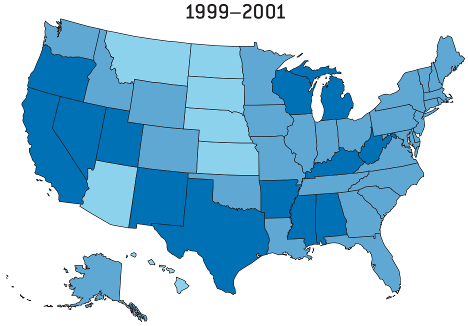
STATE BUDGET CONSTRAINTS THREATEN PUBLIC HEALTH **BY RODGER DOYLE**

American health care is getting worse. That conclusion comes from a comprehensive state-by-state analysis conducted by *Governing* magazine and the University of Richmond and funded by the Pew Charitable Trusts in Philadelphia. The study found dramatic improvements in the efficacy of treatment but declining access to treatment. The information is consistent with the trends in several key indicators, including growing prevalence of diabetes and obesity. Matthew M. Zack and his colleagues at the Centers for Disease Control and Prevention found that those most affected are the middle-aged and those with only a high school education. The Zack group used the CDC's Behavioral Risk Factor Surveillance surveys to develop trend data by state for the period 1993 to 2001 on the number of unhealthy days, both mental and physical. These data, shown on the maps, record a significant decline in health between 1993–1995 and 1999–2001.

1993–1995



1999–2001



Number of Unhealthy Days in an Average Month

Fewer than 5.0 | 5.0 to 6.0 | More than 6.0

SOURCE: Matthew M. Zack et al. Data were gathered in response to such questions as "For how many days during the past 30 days was your physical health not good?"

The states, which bear the major responsibility for administering public health programs, have had to divert some of their resources to fight bioterrorism. But the heart of the problem is Medicaid, which targets the disadvantaged. The federal government customarily pays 59 percent of Medicaid costs, with the states paying the balance, which in fiscal year 2004 amounted to an estimated $120 billion. Twenty years ago Medicaid accounted for 8 percent of state expenditures, but it is likely to go over 25 percent in the next year or so.

As a result, states have cut back on many essential programs. Efforts to combat diseases such as tuberculosis and asthma are grossly underfunded. Lack of community-based services for the mentally ill has left tens of thousands on the streets or in jail. Some states, such as Kentucky, have slashed long-term health care for the elderly. Others, such as Texas, are cutting back on preventive health care for children, which will probably add to the health care burden in the long run.

Positive efforts do stand out to be sure: Nevada has strengthened its mental health program, and Arkansas has improved programs for the elderly. Constructive developments such as these, however, have not been enough to stop the general downward trend.

The rising costs of medical care, including prescriptions, have contributed to the states' fiscal woes. Perhaps more important, however, is the declining number of individuals covered by health insurance, which has forced the states to insure more people. The drop, driven mostly by the erosion of employer-based coverage, is not easily remedied, because, according to a 1974 federal law, states cannot mandate that employers cover all workers. Hawaii, the only state exempted from the law because it mandated employee coverage before the act was passed, had the lowest number of unhealthy days reported for both the 1993–1995 and 1999–2001 periods.

Rodger Doyle can be reached at rdoyle2@adelphia.net

RODGER DOYLE

IN SICKNESS AND HEALTH

Selected health indicators, best/worst states

Percent not insured:
Minnesota 8.7/Texas 24.6

Infant death rate per 1,000:
New Hampshire 3.8/Delaware 10.7

Percent infants not vaccinated:
Massachusetts 9.3/Texas 25.2

Percent adults smoking cigarettes daily:
Utah 8.6/Kentucky 26.8

Percent who are heavy drinkers:
Tennessee 2.2/Wisconsin 8.6

Percent with diabetes:
Colorado 4.7/Mississippi 11.0

Percent with asthma:
Hawaii 5.6/Maine 9.9

SOURCES: U.S. Census Bureau, National Center for Health Statistics, and the Centers for Disease Control and Prevention. All data are for 2003, except infant deaths, which are for 2001.

FURTHER READING

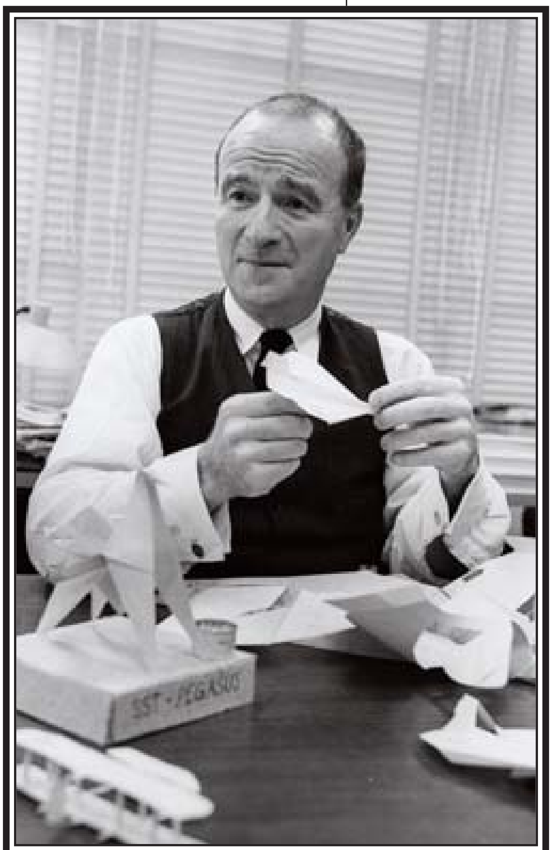
A Case of Neglect: Why Health Care Is Getting Worse, Even Though Medicine Is Getting Better. Katherine Barrett, Richard Greene and Michele Mariani in *Governing*; February 2004. http://governing.com/gpp/2004/intro.htm

Worsening Trends in Adult Health-Related Quality of Life and Self-Rated Health—United States, 1993–2001. Matthew M. Zack, David G. Moriarty, Donna F. Stroup, Earl S. Ford and Ali H. Mokdad in *Public Health Reports*, Vol. 119, No. 5, pages 493–505; September 2004.

IN MEMORIAM

Gerard Piel (1915-2004)

Gerard Piel, the former publisher of *Scientific American* who helped to redefine the modern era of science journalism, died on September 5 at the age of 89. His death resulted from the lingering complications of a stroke suffered this past February.



GERARD PIEL during this magazine's paper airplane contest in the 1960s.

WERNER WOLFF *Black Star*

Although Piel never formally studied science—he graduated from Harvard magna cum laude as a history major—his influence over science literacy worldwide cannot be overstated. For six years in the 1940s, he worked as the science editor at *Life* magazine, reaching an audience of millions. Then, in 1947 he, fellow journalist Dennis Flanagan and a group of investors purchased *Scientific American*. The already venerable magazine had published accounts of inventions and discoveries since 1845, but its new owners had still higher ambitions.

Piel as publisher and Flanagan as editor remade the magazine into a more sophisticated, authoritative monthly in which the feature articles were routinely written by experts in those fields and yet were intended for an educated general public instead of a purely professional one. This new formula for *Scientific American* proved hugely successful, and Piel saw the magazine's readership flourish into the 1980s.

Gerard Piel believed strongly that the entire world should participate and share in the benefits of scientific enterprise. For that reason, he aggressively sought out international partners who could bring *Scientific American* to other countries and languages. Thanks in part to his initiative, editions of *Scientific American* are today published in 18 nations. Piel became a roving ambassador of goodwill for science, and he occasionally tapped other staffers of the magazine to join him: longtime employees fondly remember when, in the late 1970s, Piel brought the entire staff with him on trips through the then Soviet Union and China.

For Piel, scientific reason also deserved to be a beacon for political reform and good governance. Throughout the 1970s the magazine published a string of articles arguing for international nuclear disarmament and a reassessment of the military strategies of mutually assured destruction.

In 1984 Piel stepped away from his day-to-day duties as publisher and assumed the role of chairman of *Scientific American*, which he held for another decade. Nevertheless, he continued to write and speak about science. During this time he authored *The Age of Science: What Scientists Learned in the 20th Century* (2001) and *Only One World* (1992) and co-edited *The World of René Dubos: A Collection of His Writings* (1990). These books joined such earlier works as *Science in the Cause of Man* (1961) and *The Acceleration of History* (1972).

In addition, Piel was named to many other positions of note throughout his career, including presidency of the American Association for the Advancement of Science and membership on the boards of the American Museum of Natural History, the Henry J. Kaiser Foundation and the Mayo Clinic.

Gerard Piel's panache showed in the character of his hats and the bouquet of his cigars. A wrestler in college, he carried that fitness and determination with him throughout life. He became a landmark figure in journalistic letters, and he directly promoted the growth of science as much as any one person could.

As a small tribute, www.sciam.com/ontheweb will host some recollections of the man from his former co-workers and others who knew him. *Scientific American* itself mourns his passing but is privileged to stand as his memorial. —*The Editors*

Flying Carpets and Scientific Prayers

Scientific experiments claiming that distant intercessory prayer produces salubrious effects are deeply flawed By MICHAEL SHERMER

In late 1944, as he cajoled his flagging troops to defeat the Germans in the Battle of the Bulge, General George S. Patton turned to his chief chaplain for help.

Patton: Chaplain, I want you to publish a prayer for good weather. I'm tired of these soldiers having to fight mood and floods as well as Germans. See if we can't get God to work on our side.

Chaplain: Sir, it's going to take a pretty thick rug for that kind of praying.

Patton: I don't care if it takes the flying carpet. I want the praying done.

Although few attribute Patton's subsequent success to a divine miracle, a number of papers have been published in peer-reviewed scientific journals in recent years claiming that distant intercessory prayer leads to health and healing. These studies are fraught with methodological problems.

Suspicions of fraud. In 2001 the *Journal of Reproductive Medicine* published a study by three Columbia University researchers claiming that prayer for women undergoing in vitro fertilization resulted in a pregnancy rate of 50 percent, double that of women who did not receive prayer. ABC News medical correspondent Timothy Johnson cautiously enthused, "A new study on the power of prayer over pregnancy reports surprising results, but many physicians remain skeptical." One of those skeptics was from the University of California at Irvine, a clinical professor of gynecology and obstetrics named Bruce Flamm, who not only found numerous methodological errors in the experiment but also discovered that one of the study's authors, Daniel Wirth, a.k.a. John Wayne Truelove, is not an M.D. but an M.S. in parapsychology who has since been indicted on felony charges for mail fraud and theft, to which he has pled guilty. The other two authors have refused to comment, and after three years of inquiries from Flamm, the journal removed the study from its Web site, and Columbia University launched an investigation.

Lack of controls. Many of these studies failed to control for such intervening variables as age, sex, education, ethnicity, socioeconomic status, marital standing, degree of religiosity and ignored the fact that most religions have sanctions against such insalubrious behaviors as sexual promiscuity, alcohol and drug abuse, and smoking. When such variables are controlled for, the formerly significant results disappear. One study on recovery from hip surgery in elderly women did not control for age; another study on church attendance and recovery from illness did not consider that people in poor health are less likely to attend church.

Outcome differences. In a highly publicized study of cardiac patients prayed for by born-again Christians, of 29 outcome variables measured only six showed a significant difference between the prayed-for and nonprayed-for groups. In related studies, different outcome measures were significant. To be meaningful, the same measures need to be significant across studies because if enough outcomes are measured, some will show significant correlations by chance.

Scientific prayer makes God a celestial lab rat.

Operational definitions. When experiments are carried out to determine the effects of prayer, what precisely is being studied? For example, what type of prayer is being employed? (Are Christian, Jewish, Muslim, Buddhist, Wiccan and shaman prayers equal?) Who or what is being prayed to? (Are God, Jesus and a universal life force equivalent?) What is the length and frequency of the prayer? (Are two 10-minute prayers equal to one 20-minute prayer?) How many people are praying, and does their status in the religion matter? (Is one priestly prayer identical to 10 parishioner prayers?) Most prayer studies either lack such operational definitions or lack consistency across studies in such definitions.

The ultimate fallacy is theological: if God is omniscient and omnipotent, he should not need to be reminded or inveigled into healing someone. Scientific prayer makes God a celestial lab rat, leading to bad science and worse religion. SA

Michael Shermer is publisher of Skeptic *(www.skeptic.com) and author of* The Science of Good and Evil.

BRAD HINES

What's in a Name?

Not much at the moment, thinks biologist Kevin de Queiroz, but names could be made to reflect our modern understanding of life's origins and complexity   By CHRISTINE SOARES

It is drizzling and unusually cold for July in Paris. Still, Kevin de Queiroz, his wife, Molly Morris, and half a dozen old friends are huddled at sidewalk café tables on the rue Linné, flush with enthusiasm. In honor of the locale, de Queiroz raises his glass to Carl von Linné, the 18th-century Swedish botanist who also went by the name Linnaeus. "He's not the problem," sighs de Queiroz, careful, even among allies, to distinguish between the historical figure and today's rules for naming organisms, which are still largely based on Linnaeus's method. For 20 years, de Queiroz and a core group of like-minded colleagues have argued that Linnean-style classification was fine for the static, divinely created universe of the 18th century but that it just doesn't work in an evolving world. They have spent the past week in Paris, along with some 60 fellow systematic biologists from 11 countries, polishing an alternative.

KAY CHERNUSH

KEVIN DeQUEIROZ: ORGANIZING ORGANISMS

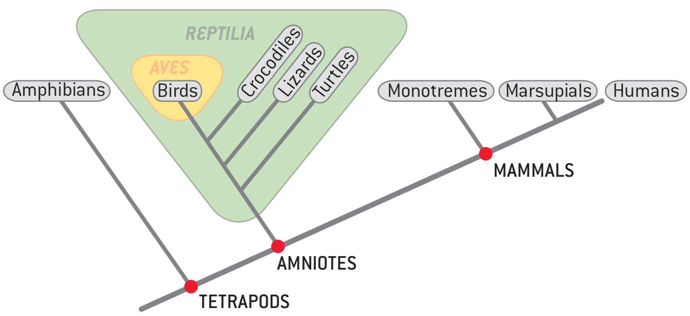
- **Growing up in Los Angeles, he became fascinated by animals during hikes with his two brothers and engineer father.**
- **On the PhyloCode's future: "If an idea is good, it will take hold. Even if it doesn't because it isn't the right time, it will resurface later."**

Their proposed "PhyloCode" is a system for naming organisms based on evolutionary relationships (phylogeny), rather than grouping them by shared characteristics. Instead of arranging such groups in a descending hierarchy of kingdoms, phyla, classes, orders, families and so forth, PhyloCode proponents want to denote groupings descended from a common ancestor. These nested "clades" more accurately depict the reality of the branching tree of life, PhyloCoders contend. The new nomenclature would also provide a language "for scientists to talk about what we really want to talk about," says Jacques Gauthier, who originated the idea with de Queiroz.

The pair's frustration with Linnean classification first emerged in 1983, when they were graduate students at the University of California at Berkeley. Asked to prepare a taxonomy of lizards and snakes, de Queiroz and Gauthier ran up against the system's ranking requirements. Current rules for naming and classifying animals or plants are governed, respectively, by the International Codes of Zoological or of Botanical Nomenclature. The ICZN and ICBN mandate that an organism fit into a series of categorical ranks. A group of species constitutes a genus, with one of the species serving as the defining "type" of that genus. Similarly, a group of genera make up a family defined by a single type genus and so on up the ranks memorized by generations of high school students.

Members of a given rank are "diagnosed" by the rank's defining characteristics. Hence, a member of the Chordate phylum must have a backbone. The system becomes awkward when it encounters an intermediate organism, one that has features of more than one group. And nothing about the ranks indicates where and when characteristics originated. To de Queiroz and Gauthier, forcing interrelated creatures and their assorted common ancestors into separate ranks seemed arbitrary and artificial.

"That's when we first came up with ideas about how to define names phylogenetically," de Queiroz says. Philosophically and scientifically, it just made more sense, he explains. Linnaeus lived a century before Charles Darwin published *On the Origin of Species*. Taxonomists have not ignored the implications of evolution. The nomenclature based on Linnaeus's system has expanded since the 19th century to include supertribes, subcohorts and myriad other gradations to accommodate newly discovered organisms. But for de Queiroz and Gauthier, jury-rigging the old system was insufficient. In a series of papers published between 1987 and 1994, they declared the time had come "to complete the Darwinian revolution" with a nomenclature based in evolutionary theory.



PHYLOCODE might put birds in a "clade" named *Aves*, within a larger *Reptilia* clade that encompasses a reptile ancestor and all its descendants. Current classification makes birds and reptiles separate, equal "classes."

In the meantime, both were becoming esteemed reptile experts. De Queiroz, now 48, is curator of reptiles and amphibians at the National Museum of Natural History, and Gauthier is professor of vertebrate paleontology at Yale University. During those years, their idea also started to garner adherents.

By 1998 the concept had enough momentum for 30 converts to convene a workshop at Harvard University to start shaping a code for phylogenetic nomenclature. De Queiroz teamed with Ohio University botanist Philip Cantino to draft the actual PhyloCode, and they posted the document online for comments. Another workshop in 2002 at Yale further refined the idea. Critics began to take the notion seriously as well. Heckled at meetings and assailed in scientific journals, de Queiroz is a poised and eloquent defender of the faith.

"For people like me, the whole appeal of going into science is to think about new things," de Queiroz says. "If we're forced to keep thinking in accepted forms, then what's the point of doing science?" Opponents insist that there is nothing wrong with the current form of nomenclature. Organisms can be arranged into phylogenetic clades without changing their present names and thereby losing useful information about their larger group memberships. "We don't want to change their names; we want to change what they mean," de Queiroz counters. He does not see the point in preserving misleading information.

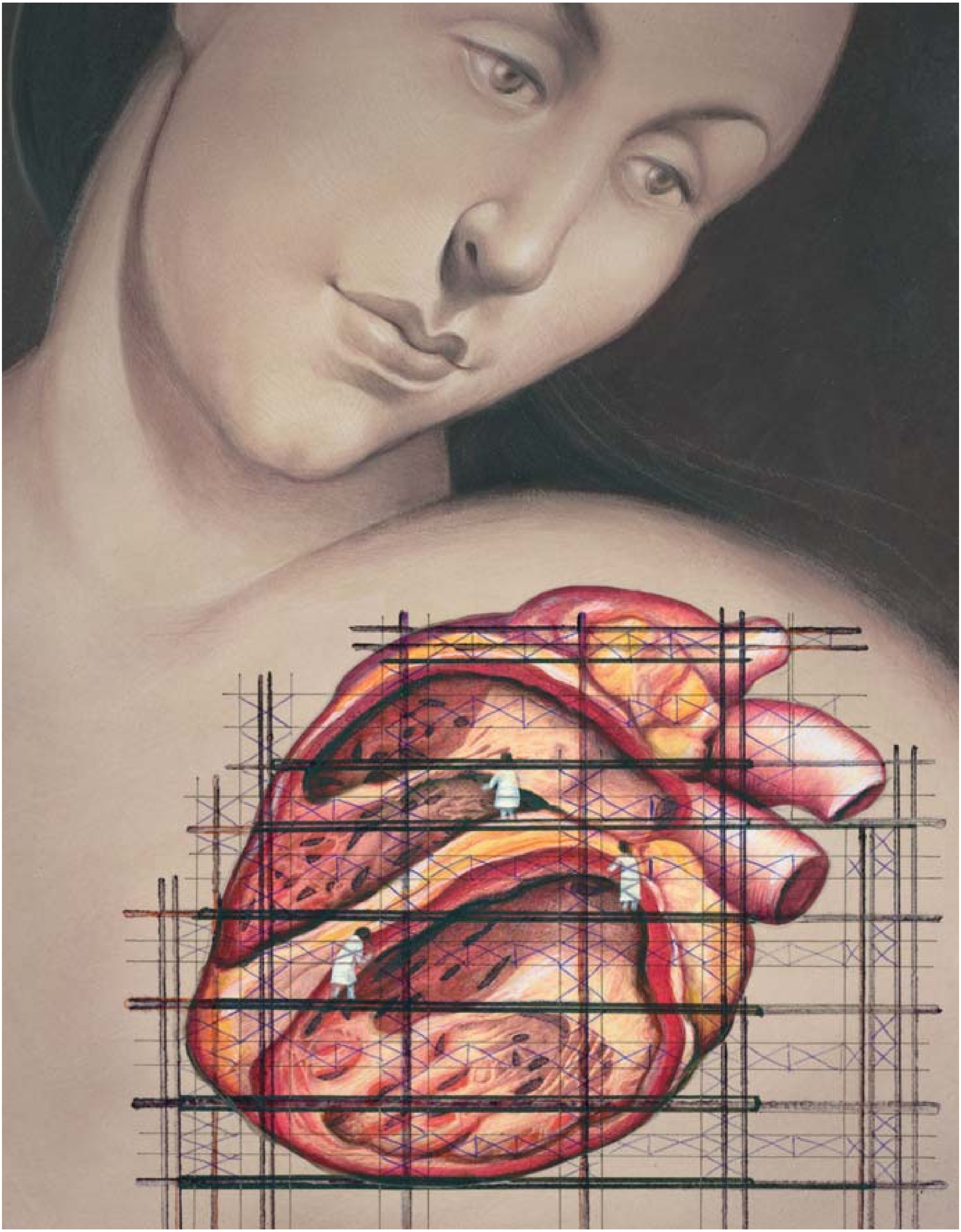
For example, the constituents of "class Mammalia" have shifted fairly dramatically at least 10 times, depending on which characteristics—such as mammary glands or certain bone structures—taxonomists used to identify its members. Under the PhyloCode, the rank-free name Mammalia would be "fixed to a point in the tree," de Queiroz explains. It would thus be defined by its members rather than by a checklist of attributes. One proposed PhyloCode definition for Mammalia reads: "the least inclusive clade containing the common ancestor of *Homo sapiens* + *Didelphis marsupialis* [opossum] + *Ornithorhynchus anatinus* [platypus] + *Tachyglossus aculeatus* [echidna]." Proponents believe that what the PhyloCode definitions may lack in physical description of the included organisms, they more than make up for in precision.

De Queiroz realized that to prove its merits, the PhyloCode had to become fully operational, so last summer supporters convened the first ever meeting of the International Society for Phylogenetic Nomenclature. Gathering in a small auditorium off the main gallery of the French National Museum of Natural History, 70 scientists inaugurated the society and elected de Queiroz its first president. With troops of children filing past to see a blockbuster mammoth exhibit, the PhyloCoders spent four full days debating conventions of the code and proposed names and definitions under it.

The symposium's published proceedings, expected in 2006, will serve as the official starting point for PhyloCode nomenclature, just as Linnaeus's 1758 *Systema Naturae* and 1753 *Species Plantarum* mark the beginnings of the ICZN and ICBN. De Queiroz and the other PhyloCoders are content for now to peacefully coexist with the current codes. That is just fine with the International Society for Zoological Nomenclature, according to its executive secretary, Andrew Polaszek. He finds the PhyloCode "impossible to apply—with so many unwieldy aspects it's unusable." But Polaszek admits that enthusiasm for the PhyloCode, particularly among younger scientists, made him realize how user-unfriendly the 250-year-old ICZN had become. PhyloCode's appeal gave Polaszek "a profound sense of urgency" for the need to streamline the ICZN's arcane language and to make the code accessible online. But he foresees no changes to the essential Linnean system.

As de Queiroz hurries to the PhyloCode meeting's final session along streets named for historic systematists—Linné, Jussieu, Buffon—he observes that future generations of scientists will be the ones to decide which code is most sensible. "If the principles were incorporated into the current nomenclature systems and PhyloCode died out, I wouldn't mind," de Queiroz declares. "The PhyloCode doesn't matter—it's the principle that matters." SA

TOMMY MOORMAN



REBUILDING BROKEN HEARTS

Biologists and engineers working together in the fledgling field of tissue engineering are within reach of one of their greatest goals: constructing a living human heart patch

By Smadar Cohen and Jonathan Leor

A heart broken by love usually heals with time, but damage to cardiac muscle caused by a heart attack gets progressively worse. Unlike liver or skin, heart tissue cannot regenerate, so the scar left after a heart attack remains a noncontractile dead zone.

By hobbling the heart muscle's normal synchronous contractions, the scar, known as an infarct, also increases strain on the healthy parts of the muscle, leading to further cell death and deformation of the cardiac wall. This cycle of deterioration can cause an infarct to double in size within just months.

Medical interventions are allowing more people to survive the crisis of a heart attack. But at least a third of these will experience the subsequent steady weakening of their injured hearts, termed heart failure, for which there is only one cure at present: transplantation—a complicated, expensive alternative limited by a severe shortage of donors. Last year in the U.S., for example, more than 550,000 new cases of heart failure were diagnosed, but only about 2,000 transplants were performed. For the remainder of patients, quality of life steadily erodes, and less than 40 percent will survive five years after the initial attack.

If doctors could repair an infarct in the human heart, or even just halt its expansion, they would transform millions of lives. Thus, building a patch of living human heart tissue has become one of the most urgent goals in tissue engineering. It is also one of the most ambitious. Cardiac muscle fibers must organize themselves in parallel, then form physical and neural connections in order to conduct the electrical signals that allow the fibers to synchronize contractions. Skin and cartilage are far less complex, and growing them in the lab is also simpler because those tissues do not require internal vasculature. For thicker structures such as heart muscle, finding a way to integrate the requisite blood supply into a three-dimensional piece of tissue remains a major obstacle.

Still, the prospect of "building" any kind of living tissue outside the body was widely considered outlandish just 15 years ago. Since that time, cell biologists and materials engineers have brought novel insights and techniques from their respective disciplines to the challenge and made substantial progress. In our own collaboration, for example, engineering principles played a crucial role in enabling us to develop a

ANITA KUNZ

scaffold that encourages heart cells and blood vessels to grow, even in the dead zone of an infarct.

Laying the Groundwork

A MYOCARDIAL INFARCTION, popularly known as a heart attack, usually happens because a major blood vessel supplying the heart's left ventricle is suddenly blocked by an obstruction, such as a clot. Part of the cardiac muscle, or myocardium, is deprived of blood and therefore oxygen, which kills the heart's contractile muscle cells (called cardiomyocytes) and leaves a swath of dead tissue. The size of this infarct will depend on the size of the area fed by the blood vessel that was blocked.

Because myocytes rarely divide, surviving cells cannot repopulate the area by replicating themselves. Local stem cells, which act as progenitors of new cells in some other tissues, are proving elusive in the heart and seem unable to heal the wound on their own. Instead, noncontractile fibrous cells gradually replace an infarct's dead myocytes. Healthy myocytes adjacent to the infarct may also die, causing the infarct to expand further. In this process, known as remodeling, the ventricle wall in the area of the infarct becomes thinner and eventually distends [*see illustration on opposite page*] or even ruptures.

In the past few years, researchers have attempted to regrow heart tissue in an infarct zone by transplanting stem cells from other tissues, such as bone marrow or skeletal muscle. The hope was that these cells would either adapt to their surroundings and begin producing new cardiomyocytes or at least help to spur any natural regenerative capacity the heart itself might possess. Unfortunately, trials of this approach have had limited success. Most of the stem cells do not survive the transplant. Those that do tend to congregate at the edges of the infarct but fail to make physical contact with adjacent healthy tissue or to conduct the electrical signals that allow heart cells to synchronize their contractions.

These implanted cells cannot thrive in the infarct primarily because the damaged area lacks the vital natural infrastructure that normally supports living cells. In healthy tissue, this extracellular matrix is composed of structural proteins, such as collagen, and complex sugar molecules known as polysaccharides, such as heparan sulfate. The extracellular matrix both generates growth-signaling chemicals and provides physical support for cells.

Aware of the importance of extracellular matrix, tissue engineers have long sought an ideal substitute to serve as a platform for growing living tissues. Such a material could form a scaffold to support cells, allowing them to thrive, divide and organize themselves into a three-dimensional tissue as they do in nature. The structure would solve the problem of transplanted cells migrating away from a scarred area. But after the cells have established themselves and begun secreting their own extracellular matrix, the scaffold should dissolve, leaving behind only healthy tissue. Perhaps most important, the scaffold should allow—better still, promote—rapid vascularization within the new tissue. Blood vessels delivering oxygen to every cell and carrying away their wastes are essential to the cells' survival once they are transplanted into the living host.

During the late 1980s, one of us (Cohen) had the pleasure of working with Robert Langer, a pioneer in the field of tissue engineering [see "Tissue Engineering: The Challenges Ahead," by Vacanti and Langer; SCIENTIFIC AMERICAN, April 1999], in his lab at the Massachusetts Institute of Technology. At the time, the very idea of building living tissue was dismissed by many as a dream. Moreover, cells had always been the domain of biologists, and we were chemical engineers. But it was a time of breakthroughs in both disciplines: biologists were gaining new insights into the way cells interact with materials, whereas engineers were achieving the ability to synthesize new types of polymers. In the nearly 20 years since, tissue engineers have been experimenting with a wide variety of materials, both synthetic and natural, to create the optimal platform for living cells to grow into a whole, functioning tissue.

Among the most popular synthetics have been degradable polyesters composed of lactide or glycolide, or a combination of the two. Although these have proved generally safe to use within the human body, they have several drawbacks. Because most of these materials repel water, living cells do not adhere to them well, and scaffolds made of these polymers tend to crumble rather than degrade at a steady rate. Acidic by-products of their degradation can cause local tissue in-

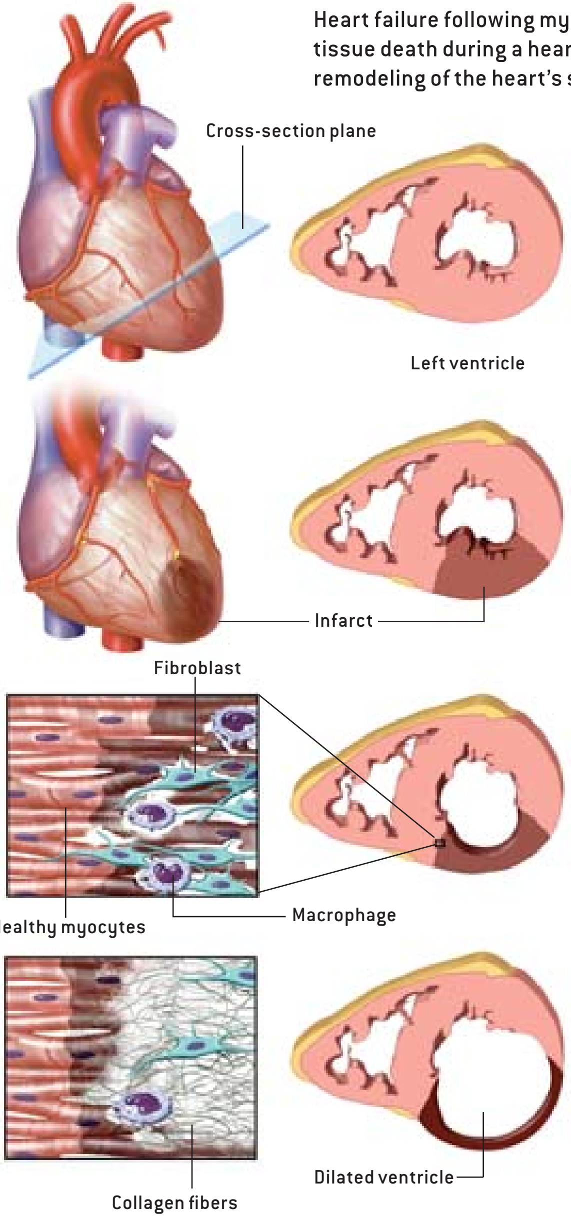
Implanted cells cannot thrive in the INFARCT because the area lacks vital natural INFRASTRUCTURE.

Overview/*Mending Hearts*

- Scarred cardiac muscle will lead to heart failure in millions of heart attack survivors unless the damaged area can be restored or replaced with new tissue.
- Constructing living tissue brings together the biologist's understanding of cell behavior and the material chemist's mastery of engineering.
- Tissue engineers, already able to coax heart muscle regeneration in vivo, are building on what they have learned to generate working heart muscle in the lab.

HEART FAILURE: FROM CRISIS TO CHRONIC ILLNESS

Heart failure following myocardial infarction can result from massive tissue death during a heart attack, but more often it is caused by a gradual remodeling of the heart's shape.



HEALTHY HEART

The heart's left ventricle pumps newly oxygenated blood to the rest of the body, and its walls are normally thick with muscle fibers called myocytes.

ACUTE INFARCTION

When a blood vessel feeding the heart muscle is blocked, myocytes die from oxygen deprivation. The resulting swath of dead muscle tissue is called an infarct.

SCAR FORMATION

Within hours to days, enzymes in the infarct zone begin degrading the extracellular matrix. Meanwhile macrophages move in to consume dead myocytes, and collagen-producing fibroblasts take their place. The formerly thick, muscular ventricle wall becomes thin and rigid. As healthy myocytes die at the border of the scarred area, the infarct can keep growing, doubling in size in just a few months.

VENTRICULAR REMODELING

The scarred heart's contractions become stilted, like the gait of a person with one leg in a cast. To compensate for the added strain, remaining healthy muscle may thicken at first. Ultimately, though, overload causes additional cells to die and the entire wall of the left ventricle to dilate, thinning further as it distends. The failing heart becomes progressively less able to pump adequate amounts of blood to the body.

TERESE WINSLOW

flammation as well as affect the viability of transplanted cells. Newer synthetic water-based gels do not have most of these problems, and they do resemble natural extracellular matrix in texture. Still, these hydrogels lack chemical features found in natural extracellular matrix proteins, such as collagen, which provide cells with important functional cues.

Along with collagen itself, other extracellular matrix proteins, such as fibronectin, have also been tested as possible scaffold materials. Whereas these proteins do contain amino acids to which living cells readily adhere, they lack sufficient mechanical strength to support a large number of cells, and collagen in particular is quickly consumed by enzymes in the body. In addition, depending on their source, proteins can provoke immune rejection, which would only add further dangers and hardships to the lives of patients already suffering heart failure.

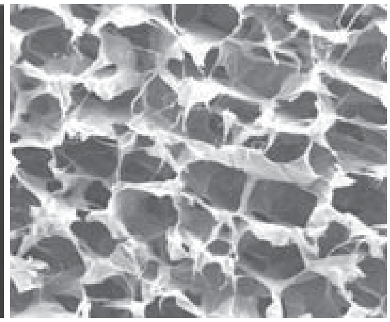
Therefore, we decided to try building a scaffold from a different kind of natural polymer: a polysaccharide called alginate that is derived from algae. It is biocompatible, meaning that it does not provoke the body's immune system. And when a particular type of alginate is dissolved in water, then exposed to positively charged calcium ions, its molecules crosslink to form a hydrogel that is 98 percent water, with a gelatinous consistency and elastic properties similar to those of natural extracellular matrix.

But to use this alginate hydrogel as a scaffold, we needed to give it shape and internal structure, while enhancing its

ENGINEERING A TISSUE SCAFFOLD

Alginate scaffolds

Spongelike structure

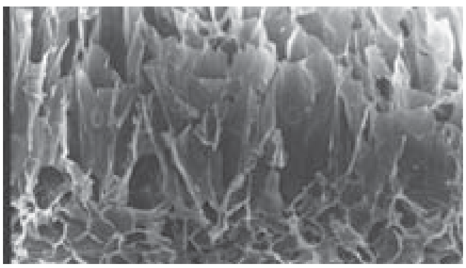


Scaffolds provide physical support and guidance for living cells to organize themselves into a tissue. Ideally, the structure consists mostly of pores that are highly interconnected, with a diameter of at least 200 microns (the average size of a capillary), to permit blood vessel penetration and cell interactions.

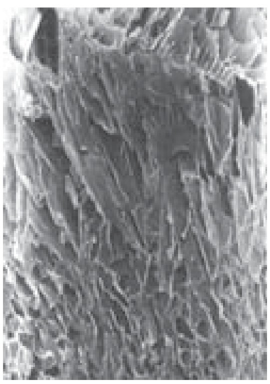
We chose alginate, an algae derivative, as our scaffold material for its chemical resemblance to natural extracellular matrix. But we had to devise a way to turn a viscous alginate-water solution into a solid scaffold, whose shape (*near right*) and internal architecture (*far right*) could be precisely controlled.

Knowing that the water in our alginate hydrogel would form ice crystals if frozen and that different cooling methods might dramatically influence the crystals' shapes, we experimented with freeze-drying techniques to create our scaffolds. As expected, freezing the hydrogel produced a spongelike architecture of ice crystals separated by thin solid walls of alginate. Subliming away the crystals left pores in a variety of shapes, sizes and orientations, reflecting the speed and direction of the crystals' growth as heat was transferred from the alginate solution to its cooling medium (*below*).

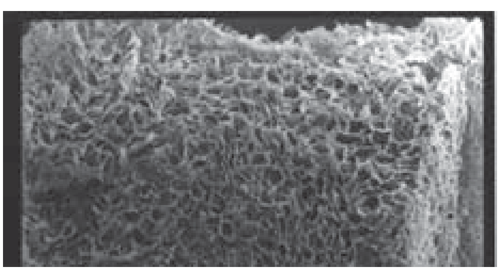
FREEZING REGIMES



In an oil bath at –35 degrees Celsius, ice formed fastest at the bottom of the sample, producing tiny, densely packed, interconnected pores there. Larger, elongated pores above follow the cooling front's direction.



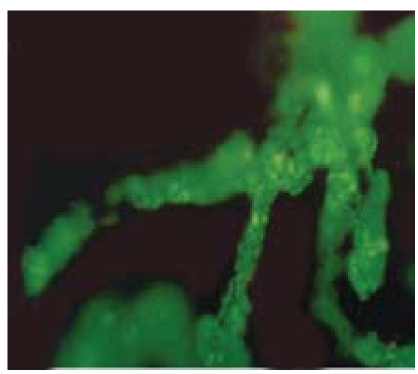
In liquid nitrogen at –196 degrees C, a similar gradient appears from bottom to top. The complex pore shapes and directions near the top of the sample may result from nitrogen's high volatility, producing multidirectional cooling fronts where the cold vapor meets the alginate solution.



In a freezer at –20 degrees C, the alginate solution first cooled to –10 degrees C, warmed suddenly to –2 degrees C, then slowly cooled to –20 degrees C. The sharp temperature spike suggests that the water released its heat and began crystallizing simultaneously throughout the sample, as reflected in its uniform, interconnected pores.

PORE ARCHITECTURE

Our ability to plan and control scaffold architecture with these freezing techniques is so important because pore structure has a fundamental influence on the forming tissue's function. Elongated pores may promote blood vessel formation, for example. When we used liquid nitrogen to create scaffolds containing long channels and then seeded these with fluorescently marked endothelial cells (*green, below*), the cells arranged themselves into capillarylike structures within two weeks.



LILIA SHAPIRO AND SMADAR COHEN *Ben-Gurion University* (*top left*); MICHAL SHACHAR, RONIT BASHER AND SMADAR COHEN *Ben-Gurion University* (*top right*); FROM S. ZMORA ET AL. IN *BIOMATERIALS*, VOL. 23; 2002, USED WITH PERMISSION FROM ELSEVIER AND SHARON ZMORA AND SMADAR COHEN *Ben-Gurion University* (*middle, top to bottom*); SIGALIT AMITAY-SHAPRUT AND SMADAR COHEN *Ben-Gurion University* (*bottom right*) mechanical strength so that it would keep its shape under the weight of seeded cells. To achieve this, we devised a new technique for solidifying alginate that was inspired by engineering principles.

We started by pouring the alginate solution into a variety of templates, then froze these using three different cooling methods, each of which produces distinctive temperature gradients within the solution during freezing. In all the frozen samples, the resulting structure comprised ice crystals separated by thin alginate walls. When we sublimed away the ice crystals, we were left with a spongelike scaffold whose tiny pores mirrored the shape of the crystals. As suspected, we found that by varying the freezing methods, we can control the density of the pores, their size and direction, and their degree of interconnection [*see box on opposite page*].

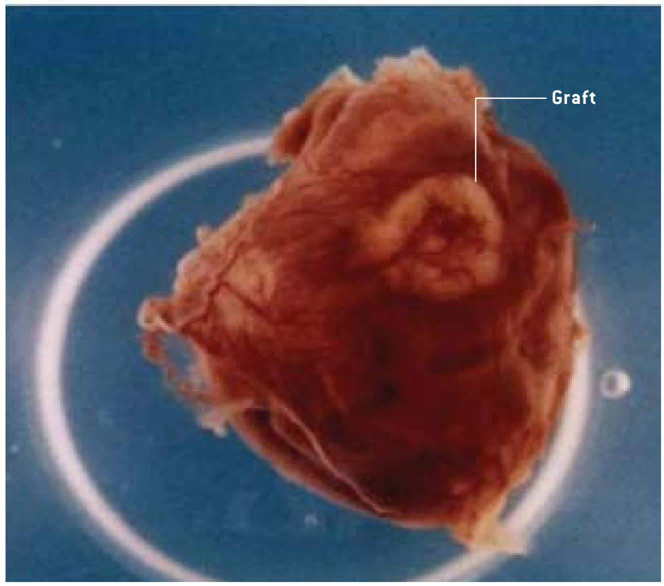
Interconnected pores are particularly important because they allow living cells, when first "seeded" in the scaffold, to pass easily throughout the structure. Free, continuous passage of nutrients and wastes to and from the cells when they are incubating is also vital. And we have learned that the degree of pore interconnection critically influences the ability of new blood vessels to penetrate the forming tissue once it is transplanted into the host. Finally, the unique internal architecture of these scaffolds, resembling foam or a beehive, contributes to their mechanical strength. Even when pores constitute more than 95 percent of their volume, the scaffolds can withstand significant external pressure.

So we now had the ability to create a scaffold with the exact shape and structure we desired, one that did not activate the immune system, was made from a natural material using nontoxic chemistry and had good mechanical durability, yet disintegrated within the body in a reasonable time. Still, it remained to be seen whether living cells would find our scaffold to be an adequate substitute for the missing extracellular matrix in an actual infarction.

Building a Tissue

BEFORE IMPLANTING our scaffolds in lab animals, we wanted to see how heart cells would take to the alginate material in vitro, that is, outside the body. We took cells from the hearts of rat embryos—which, unlike mature cardiomyocytes, are still capable of dividing—and suspended them in a liquid medium containing nutrients. Next we infused this suspension into round scaffolds six millimeters in diameter and one millimeter tall. With the help of some mild centrifugal force, the cells rapidly penetrated the scaffold's pores, distributing themselves evenly in less than 30 minutes.

We had achieved an initial goal—PROTECTING a heart that had SUFFERED infarction and preventing further deterioration.



CELL-SEEDED SCAFFOLD, two months after implantation in a rat's heart, has integrated into an infarcted area. Local blood vessels penetrated the graft extensively, sustaining mature heart cells within the scaffold and preventing the infarct from expanding.

FROM J. LEOR ET AL. IN *CIRCULATION*, VOL. 102, NO. 19; 2000, © AHA/LWW

Speed is an important advantage in preserving the viability of these cells that are so sensitive to a lack of oxygen, and the cells' homogeneous distribution enables us to load a large number of them onto the scaffold. As a result, cell density in our scaffolds was 10^8 cells per cubic centimeter—similar to the density of mature native heart muscle tissue.

We transferred our seeded scaffolds into a special incubator called a bioreactor, which maintains ideal humidity and atmospheric conditions while continuously circulating nutrient-containing medium in and around the scaffolds. We monitored the cells' metabolism closely and after just 48 hours detected beating myocytes. After seven days, it was time to take the next step: transplanting the scaffolds into living hearts.

THE AUTHORS

SMADAR COHEN and *JONATHAN LEOR* have been collaborating for six years to create a cardiac muscle patch. Cohen, professor of biotechnology engineering at Ben-Gurion University of the Negev in Israel, studies how cells are affected by external cues. She also designs and synthesizes polymer biomaterials for tissue engineering and controlled-release drug delivery. Leor is a cardiologist at Sheba Medical Center and director of the Tel Aviv University Neufield Cardiac Research Institute. His interest in the complications of acute myocardial infarction drew him to investigating possible heart muscle regeneration through cell transplants, tissue engineering and gene therapy.

MICROSPHERES can be incorporated throughout a scaffold by mixing them into the alginate solution before freeze-drying. Only three microns in diameter, the microspheres accelerate blood vessel formation—without getting in the way—by steadily releasing growth factors.

We anesthetized and operated on adult rats that had experienced myocardial infarction in the left ventricle seven days earlier. In each animal, it was easy to detect the infarct: the pale scar was clearly visible and not contracting. We placed the cell-seeded scaffolds directly onto the infarcts, closed the surgical incision and waited.

After two months, we exposed the rats' hearts again and were stunned to see massive growth of new blood vessels from the healthy heart tissue into the implanted biografts [*see illustration on preceding page*]. The engineered heart transplants had integrated well with the scar tissue, and the alginate scaffolds had begun to dissolve, with natural extracellular matrix appearing in their place. The embryonic cardiac cells had developed into mature muscle fibers, some of which were organized in a parallel structure similar to that of native heart tissue. Mechanical connections and electrical synapses necessary for heart cells to contract and conduct nerve signals were also present between the fibers.

Before the transplants, we had measured the rats' cardiac function using echocardiography and did the same for a control group of rats with infarcts who would receive sham surgery but no transplant. Two months later we examined all the rats by echocardiography again. In the control group, we saw the typical scenario of heart deterioration: considerable dilation of the left ventricle and significant loss of heart function. In contrast, the transplant group all had more or less the same results as they had had immediately after infarction: the size of the left ventricle and thickness of its wall, as well as heart function, were unchanged.

We had achieved an initial goal of this research—protecting a heart that had suffered infarction and preventing further deterioration that would lead to heart failure. Still, many questions remain unanswered. The mechanism by which this treatment protected the heart muscle is still not clear, because the grafted tissue was not yet contributing to the heart's contractions. It seems that just by preventing the infarction from growing and by artificially thickening the heart wall in the infracted area, the graft could have helped prevent the usual remodeling of the ventricle.

We believe that growth of new blood vessels in the area of the infarct also contributed greatly to slowing down the tissue deterioration. New blood vessels were greatest in both number and size when we implanted scaffolds populated by cells, but one of the surprises we found in these experiments was that unseeded scaffolds also encouraged new blood vessel growth into an infarct.

The alginate scaffold might encourage growing blood vessels simply by providing support as they penetrate the damaged area. We also suspect that the material itself may help recruit stem cells to aid in regeneration because alginate's chemical structure is similar to heparan sulfate's, an important polysaccharide in natural extracellular matrix. To test this idea, we recently tried injecting alginate hydrogel directly into rats' infarcts. Even in hydrogel form, alginate preserved ventricle structure and function, apparently by acting as a substitute for the extracellular matrix and thereby promoting angiogenesis.

Of course, along with many other researchers in this field, we are also working to identify potential sources of heart cells for use in human transplantation. Because the patient's own mature heart cells do not replicate, they are not an option. Possible donor cells that might be coaxed to become cardiomyocytes include embryonic stem cells and "adult" stem cells from bone marrow or umbilical cord blood. Still, all donor cells would be recognized as foreign by the patient's immune system, necessitating the use of immunosuppressive drugs. Autologous—that is, the patient's own—cells would be preferable to avoid the problem of immune rejection. These might include stem cells and precursor cells derived from bone marrow, muscle or fat or embryonic stem cells created from the patient's cells through so-called therapeutic cloning. Or the local cardiac stem cell may yet be isolated.

Roads to Rebuilding Hearts

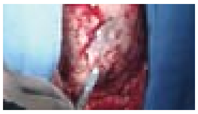
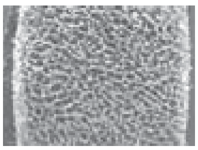
OUR PROGRESS has been encouraging and has suggested several possible approaches to using our alginate scaffolds to protect and regenerate human hearts damaged by myocardial infarction. Within three years, for example, we believe that we could certainly be ready to test unseeded alginate scaffolds in human patients who have suffered myocardial infarction. Our recent experiments in pigs have confirmed what we saw in rats: even without cells, the alginate scaffold alone seems to prevent a new infarct from expanding and the ventricle wall from remodeling. As a result, unseeded scaffolds could be especially effective in preventing cardiac failure from ever beginning in patients whose hearts have yet to undergo significant remodeling.

The apparent ability of alginate to help foster angiogenesis

ANAT PERETS AND SMADAR COHEN *Ben-Gurion University*

APPROACHES TO PATCHING HEART MUSCLE

Tissue engineers are exploring several interrelated methods for patching human heart muscle. Each technique has certain advantages, but insights gained from every experimental approach help to advance the entire field.

	TECHNIQUE	ADVANTAGES	DISADVANTAGES
	**Cell injection** Stem or precursor cells are delivered to infarct via catheter or direct injection	■ Easy delivery ■ Injected cells may induce formation of extracellular matrix and blood vessels	■ Low cell survival ■ Cells do not produce new, functioning myocytes
	**Cultured tissue** Cardiomyocytes are grown in thin sheets, layered to form a patch and implanted surgically	■ Relatively easy to grow in lab ■ More stable than injection of dissociated cells	■ Sheets lack vasculature so only small, thin constructs are possible ■ Extremely fragile
	**Porous scaffolds** Cells seeded onto 3-D scaffold made of natural or synthetic polymers are cultured in a bioreactor and then implanted surgically	■ Structure supports cell organization and promotes vascularization ■ Certain materials may promote vascularization	■ Lag time between implantation and vascularization of the tissue causes cell death
	**3-D cell printing** Ink-jet-like device dispenses layers of cells suspended in hydrogel in desired patterns; constructs are cultured and then implanted surgically	■ Multiple cell types can be precisely positioned ■ Cells are free to move and organize	■ Early-stage research that has yet to prove functional in vivo
	**Injectable scaffolds** Polymer hydrogels, alone or containing cell suspension, are delivered directly into infarct by catheter or injection	■ Easy delivery ■ May foster regeneration by providing temporary substitute for extracellular matrix	■ Limited control over tissue formation

NABIL DIB *Arizona Heart Institute* AND JONATHAN DINSMORE *GenVec, Inc.* (*heart injection*); FROM T. SHIMIZU ET AL. IN *CIRCULATION RESEARCH*, VOL. 90, NO. 3, PAGE e40; 2002, © AHA/LWW (*tissue squares*); MICHAL SHACHAR, RONIT BASHER AND SMADAR COHEN *Ben-Gurion University* (*porous scaffolds*); B. TEFFT, COURTESY OF T. BOLAND *Clemson University* (*ear*); K. STAUB, COURTESY OF K.J.L. BURG *Clemson University* (*injectable scaffolds*)

also suggests that we might enhance the survival of transplanted cells by implanting a scaffold first and waiting for vascularization, *then* seeding the scaffold with cells. We have tried this in vivo tissue formation in rats with promising results. Vascularization was also significantly enhanced when we incorporated controlled-release microspheres containing growth factors into the scaffold [*see illustration on opposite page*]. Unfortunately, we noticed that prevascularization of the scaffold reduced the space available for transplanted cells, so we are now working to improve our ability to tailor the angiogenesis using different types of growth factors.

At present, the in vitro approach to tissue engineering still allows the greatest control over the tissue's shape, composition and function. In addition, for patients whose infarct has ruptured, an entire piece of the heart may have to be replaced. We would need to fill that hole with a real piece of tissue; implantation of an empty porous scaffold will not work. Hence, we still face the problem of keeping a transplanted tissue alive until local vascularization is adequate. With the experience we have gained so far, we are now exploring the possibility of creating a prevascularized graft.

We have created a capillary bed in vitro by seeding an alginate scaffold with endothelial cells, which normally line blood vessel walls, then culturing the construct in a bioreactor. Next, we will culture endothelial cells and cardiomyocytes together on a scaffold, to attempt to form capillaries within a piece of myocardial tissue. If we succeed, we would still have to see whether this capillary bed becomes functional after transplantation and, if so, how quickly. If it connects with local vasculature rapidly, then the transplanted tissue's chances of survival should be excellent.

Many other investigators are working to overcome this hurdle of creating a prevascularized tissue, employing a variety of different strategies [see "Body Building," by Christine Soares; News Scan, SCIENTIFIC AMERICAN, May]. We are counting on the fact that we are not the only ones attempting cardiac tissue engineering. If every possible approach gets a chance to prove its merits, the entire field will be able to learn and progress. It may take another 15 years to achieve, but the dream of building a living piece of human heart is certainly no longer outlandish. SA

MORE TO EXPLORE

Tailoring the Pore Architecture in 3-D Alginate Scaffolds by Controlling the Freezing Regime during Fabrication. Sharon Zmora, Rachel Glickis and Smadar Cohen in *Biomaterials*, Vol. 23, pages 4087–4094; October 2002.

Tissue Engineering: Current State and Perspectives. Erin Lavik and Robert Langer in *Applied Microbiology and Biotechnology*, Vol. 65, No. 1, pages 1–8; July 2004.

Myocardial Tissue Engineering: Creating a Muscle Patch for a Wounded Heart. Jonathan Leor and Smadar Cohen in *Annals of the New York Academy of Sciences*, Vol. 1015, pages 312–319; May 2004.

BLACK HOLE COMPUTER may sound absurd but is proving to be a useful conceptual tool for researchers studying cosmology and fundamental physics. And if physicists are able to create black holes in particle accelerators—as some predict will be possible within a decade—they may actually observe them perform computation.

In keeping with the spirit of the age, researchers can think of the laws of physics as computer programs and the universe as a computer

BY SETH LLOYD AND Y. JACK NG

BLACK HOLE COMPUTERS

What is the difference between a computer and a black hole? This question sounds like the start of a Microsoft joke, but it is one of the most profound problems in physics today. Most people think of computers as specialized gizmos: streamlined boxes sitting on a desk or fingernail-size chips embedded in high-tech coffeepots. But to a physicist, all physical systems are computers. Rocks, atom bombs and galaxies may not run Linux, but they, too, register and process information. Every electron, photon and other elementary particle stores bits of data, and every time two such particles interact, those bits are transformed. Physical existence and information content are inextricably linked. As physicist John Wheeler of Princeton University says, "It from bit."

JEAN-FRANCOIS PODEVIN

Black holes might seem like the exception to the rule that everything computes. Inputting information into them presents no difficulty, but according to Einstein's general theory of relativity, getting information out is impossible. Matter that enters a hole is assimilated, the details of its composition lost irretrievably. In the 1970s Stephen Hawking of the University of Cambridge showed that when quantum mechanics is taken into account, black holes do have an output: they glow like a hot coal. In Hawking's analysis, this radiation is random, however. It carries no information about what went in. If an elephant fell in, an elephant's worth of energy would come out—but the energy would be a hodgepodge that could not be used, even in principle, to re-create the animal.

That apparent loss of information poses a serious conundrum, because the laws of quantum mechanics preserve information. So other scientists, including Leonard Susskind of Stanford University, John Preskill of the California Institute of Technology and Gerard 't Hooft of the University of Utrecht in the Netherlands, have argued that the outgoing radiation is not, in fact, random—that it is a processed form of the matter that falls in [see "Black Holes and the Information Paradox," by Leonard Susskind; Scientific American, April 1997]. This past summer Hawking came around to their point of view. Black holes, too, compute.

Black holes are merely the most exotic example of the general principle that the universe registers and processes information. The principle itself is not new. In the 19th century the founders of statistical mechanics developed what would later be called information theory to explain the laws of thermodynamics. At first glance, thermodynamics and information theory are worlds apart: one was developed to describe steam engines, the other to optimize communications. Yet the thermodynamic quantity called entropy, which limits the ability of an engine to do useful work, turns out to be proportional to the number of bits registered by the positions and velocities of the molecules in a substance. The invention of quantum mechanics in the 20th century put this discovery on a firm quantitative foundation and introduced scientists to the remarkable concept of quantum information. The bits that make up the universe are quantum bits, or "qubits," with far richer properties than ordinary bits.

Analyzing the universe in terms of bits and bytes does not replace analyzing it in conventional terms such as force and energy, but it does uncover new and surprising facts. In the field of statistical mechanics, for example, it unknotted the paradox of Maxwell's demon, a contraption that seemed to allow for perpetual motion. In recent years, we and other physicists have been applying the same insights to cosmology and fundamental physics: the nature of black holes, the fine-scale structure of spacetime, the behavior of cosmic dark energy, the ultimate laws of nature. The universe is not just a giant computer; it is a giant quantum computer. As physicist Paola Zizzi of the University of Padova says, "It from qubit."

Overview/*Cosmic Computers*

- Merely by existing, all physical systems store information. By evolving dynamically in time, they process that information. The universe computes.
- If information can escape from black holes, as most physicists now suspect, a black hole, too, computes. The size of its memory space is proportional to the square of its computation rate. The quantum-mechanical nature of information is responsible for this computational ability; without quantum effects, a black hole would destroy, rather than process, information.
- The laws of physics that limit the power of computers also determine the precision with which the geometry of spacetime can be measured. The precision is lower than physicists once thought, indicating that discrete "atoms" of space and time may be larger than expected.

When Gigahertz Is Too Slow

The confluence of physics and information theory flows from the central maxim of quantum mechanics: at bottom, nature is discrete. A physical system can be described using a finite number of bits. Each particle in the system acts like the logic gate of a computer. Its spin "axis" can point in one of two directions, thereby encoding a bit, and can flip over, thereby performing a simple computational operation.

The system is also discrete in time. It takes a minimum amount of time to flip a bit. The exact amount is given by a theorem named after two pioneers of the physics of information processing, Norman Margolus of the Massachusetts Institute of Technology and Lev Levitin of Boston University. This theorem is related to the Heisenberg uncertainty principle, which describes the inherent trade-offs in measuring physical quantities, such as position and momentum or time and energy. The theorem says that the time it takes to flip a bit, t, depends on the amount of energy you apply, E. The more energy you apply, the shorter the time can be. Mathematically, the rule is $t \geq h/4E$, where h is Planck's constant, the main parameter of quantum theory. For example, one type of experimental quantum computer stores bits on protons and uses magnetic fields to flip them. The operations take place in the minimum time allowed by the Margolus-Levitin theorem.

From this theorem, a huge variety of conclusions can be drawn, from limits on the geometry of spacetime to the computational capacity of the universe as a whole. As a warm-up, consider the limits to the computational power of ordinary matter—in this case, one kilogram occupying the volume of one liter. We call this device the ultimate laptop.

Its battery is simply the matter itself, converted directly to energy per Einstein's famous formula $E = mc^2$. Putting all this energy into flipping bits, the computer can do 10^{51} operations per second, slowing down gradually as the

EXTREME COMPUTING

What is a computer? That is a surprisingly complex question, but whatever precise definition one adopts, it is satisfied not just by the objects people commonly call "computers" but also by everything else in the world. Physical objects can solve a broad class of logic and mathematics problems, although they may not accept input or give output in a form that is meaningful to humans. Natural computers are inherently digital: they store data in discrete quantum states, such as the spin of elementary particles. Their instruction set is quantum physics.

ORDINARY LAPTOP

SPEED: 10^9 hertz ■ MEMORY: 10^{12} bits

INPUT

A keyboard and associated circuitry encode information as voltage pulses in a wire.

COMPUTATION

The pulses interact, guided by devices such as transistors, which perform logical operations such as NOT.

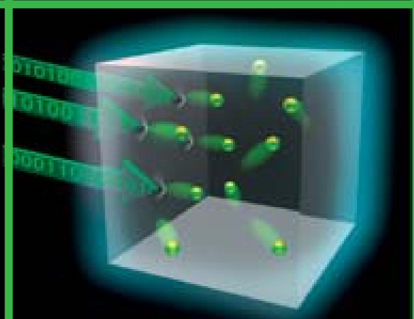
OUTPUT

The pulses, having been processed, are translated into meaningful patterns of light.

ULTIMATE LAPTOP

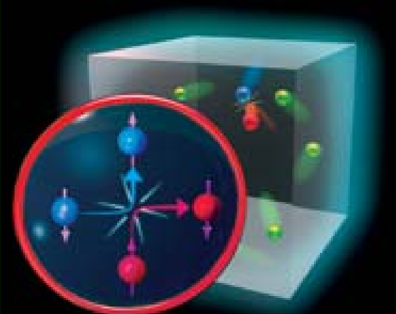
SPEED: 10^{20} hertz ■ MEMORY: 10^{31} bits

INPUT



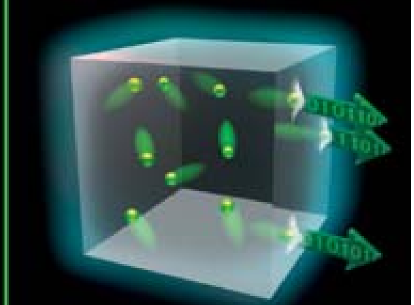
Consisting of one kilogram of hot plasma in a one-liter box, this device accepts data encoded as particle positions, velocities and spins.

COMPUTATION



The particles interact. Collisions can be arranged to perform operations such as NOT: a collision can cause particles to flip.

OUTPUT

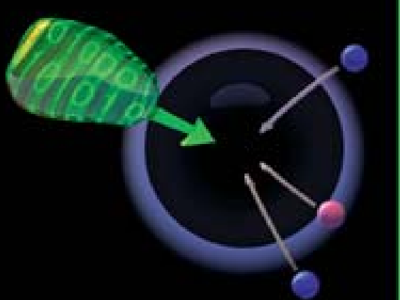


As particles leave the volume, their properties can be measured and translated. The system slowly winds down as its energy degrades.

BLACK HOLE

SPEED: 10^{35} hertz ■ MEMORY: 10^{16} bits

INPUT

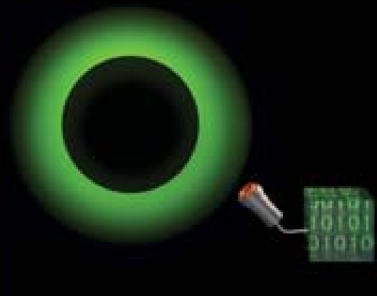


This black hole consists of one kilogram in a volume 10^{-27} meter in radius. Data and instructions are encoded in matter and dropped in.

COMPUTATION

On their descent, particles interact much as in the ultimate laptop, except that gravity also plays a role. The governing laws are not yet understood.

OUTPUT



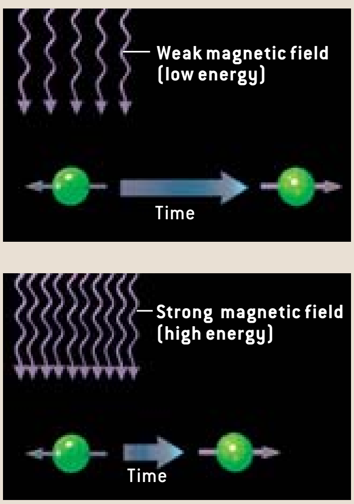
The hole emits radiation, named after physicist Stephen Hawking. New theories suggest that the radiation carries the computational output.

ALFRED T. KAMAJIAN

By preparing the material that falls into a black hole, A HACKER COULD PROGRAM IT to perform any desired computation.

energy degrades. The memory capacity of the machine can be calculated using thermodynamics. When one kilogram of matter is converted to energy in a liter volume, its temperature is one billion kelvins. Its entropy, which is proportional to the energy divided by the temperature, corresponds to 10^{31} bits of information. The ultimate laptop stores information in the microscopic motions and positions of the elementary particles zipping around inside it. Every single bit allowed by the laws of thermodynamics is put to use.

Whenever particles interact, they can cause one another to flip. This process can be thought of in terms of a programming language such as C or Java: the particles are the variables, and their interactions are operations such as addition. Each bit can flip 10^{20} times per second, equivalent to a clock speed of 100 giga-gigahertz. In fact, the system is too fast to be controlled by a central clock. The time it takes a bit to flip is approximately equal to the time it takes a signal to travel from one bit to its neighbor. Thus, the ultimate laptop is highly parallel: it acts not as a single processor but as a vast array of processors, each working almost independently and communicating its results to the others comparatively slowly.



FIRST LAW of quantum computation is that computation takes energy. The spin of a proton encodes a single bit, which can be inverted by applying a magnetic field. The stronger the field is—the more energy it applies—the faster the proton will flip.

By comparison, a conventional computer flips bits at about 10^9 times per second, stores about 10^{12} bits and contains a single processor. If Moore's law could be sustained, your descendants would be able to buy an ultimate laptop midway through the 23rd century. Engineers would have to find a way to exert precise control on the interactions of particles in a plasma hotter than the sun's core, and much of the communications bandwidth would be taken up in controlling the computer and dealing with errors. Engineers would also have to solve some knotty packaging problems.

In a sense, however, you can already purchase such a device, if you know the right people. A one-kilogram chunk of matter converted completely to energy—this is a working definition of a 20-megaton hydrogen bomb. An exploding nuclear weapon is processing a huge amount of information, its input given by its initial configuration and its output given by the radiation its emits.

From Nanotech to Xennotech

IF ANY CHUNK of matter is a computer, a black hole is nothing more or less than a computer compressed to its smallest possible size. As a computer shrinks, the gravitational force that its components exert on one another becomes stronger and eventually grows so intense that no material object can escape. The size of a black hole, called the Schwarzschild radius, is directly proportional to its mass.

A one-kilogram hole has a radius of about 10^{-27} meter. (For comparison, a proton has a radius of 10^{-15} meter.) Shrinking the computer does not change its energy content, so it can perform 10^{51} operations per second, just as before. What does change is the memory capacity. When gravity is insignificant, the total storage capacity is proportional to the number of particles and thus to the volume. But when gravity dominates, it interconnects the particles, so collectively they are capable of storing less information. The total storage capacity of a black hole is proportional to its surface area. In the 1970s Hawking and Jacob Bekenstein of the Hebrew University of Jerusalem calculated that a one-kilogram black hole can register about 10^{16} bits—much less than the same computer before it was compressed.

In compensation, the black hole is a much faster processor. In fact, the amount of time it takes to flip a bit, 10^{-35} second, is equal to the amount of time it takes light to move from one side of the computer to the other. Thus, in contrast to the ultimate laptop, which is highly parallel, the black hole is a serial computer. It acts as a single unit.

How would a black hole computer work in practice? Input is not problematic: just encode the data in the form of matter or energy and throw it down the hole. By properly preparing the material that falls in, a hacker should be able to program the hole to perform any desired computation. Once the material enters a hole, it is gone for good; the so-called event horizon demarcates the point of no return. The plummeting particles interact with one another, performing computation for a finite time before reaching the center of the hole—the singularity—and ceasing to exist. What happens to matter

ALFRED T. KAMAJIAN

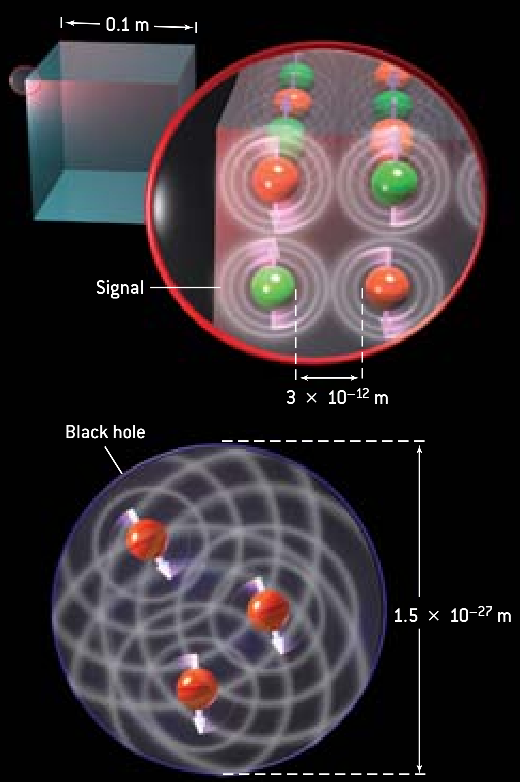
as it gets squished together at the singularity depends on the details of quantum gravity, which are as yet unknown.

The output takes the form of Hawking radiation. A one-kilogram hole gives off Hawking radiation and, to conserve energy, decreases in mass, disappearing altogether in a mere 10^{-21} second. The peak wavelength of the radiation equals the radius of the hole; for a one-kilogram hole, it corresponds to extremely intense gamma rays. A particle detector can capture this radiation and decode it for human consumption.

Hawking's study of the radiation that bears his name is what overturned the conventional wisdom that black holes are objects from which nothing whatsoever can escape [see "The Quantum Mechanics of Black Holes," by Stephen W. Hawking; SCIENTIFIC AMERICAN, January 1977]. The rate at which black holes radiate is inversely related to their size, so big black holes, such as those at the center of galaxies, lose energy much more slowly than they gobble up matter. In the future, however, experimenters may be able to create tiny holes in particle accelerators, and these holes should explode almost immediately in a burst of radiation. A black hole can be thought of not as a fixed object but as a transient congregation of matter that performs computation at the maximum rate possible.

CLASSIFYING COMPUTERS

The ultimate laptop and black hole computer embody two different approaches to increasing computing power. The ultimate laptop is the supreme parallel computer: an array of processors working simultaneously. The black hole is the supreme serial computer: a single processor executing instructions one at a time.



Ultimate laptop consists of a collection of particles that encode and process bits. Each can execute an instruction in 10^{-20} second. In that time, signals can move a distance of only 3×10^{-12} meter, which is roughly the spacing between particles. Therefore, communication is much slower than computation. Subregions of the computer work almost independently.

Black hole computer also consists of a collection of particles. Because of gravity, they encode fewer bits, giving more energy per bit. Each can execute an instruction in 10^{-35} second, which is the time it takes for a signal to cross the hole. Therefore, communication is as fast as computation. The computer operates as a single unit.

ALFRED T. KAMAJIAN

Escape Plan

THE REAL QUESTION is whether Hawking radiation returns the answer of the computation or merely gibberish. The issue remains contentious, but most physicists, including Hawking, now think that the radiation is a highly processed version of the information that went into the hole during its formation. Although matter cannot leave the hole, its information content can. Understanding precisely how is one of the liveliest questions in physics right now.

Last year Gary Horowitz of the University of California at Santa Barbara and Juan Maldacena of the Institute for Advanced Study in Princeton, N.J., outlined one possible mechanism. The escape hatch is entanglement, a quantum phenomenon in which the properties of two or more systems remain correlated across the reaches of space and time. Entanglement enables teleportation, in which information is transferred from one particle to another with such fidelity that the particle has effectively been beamed from one location to another at up to the speed of light.

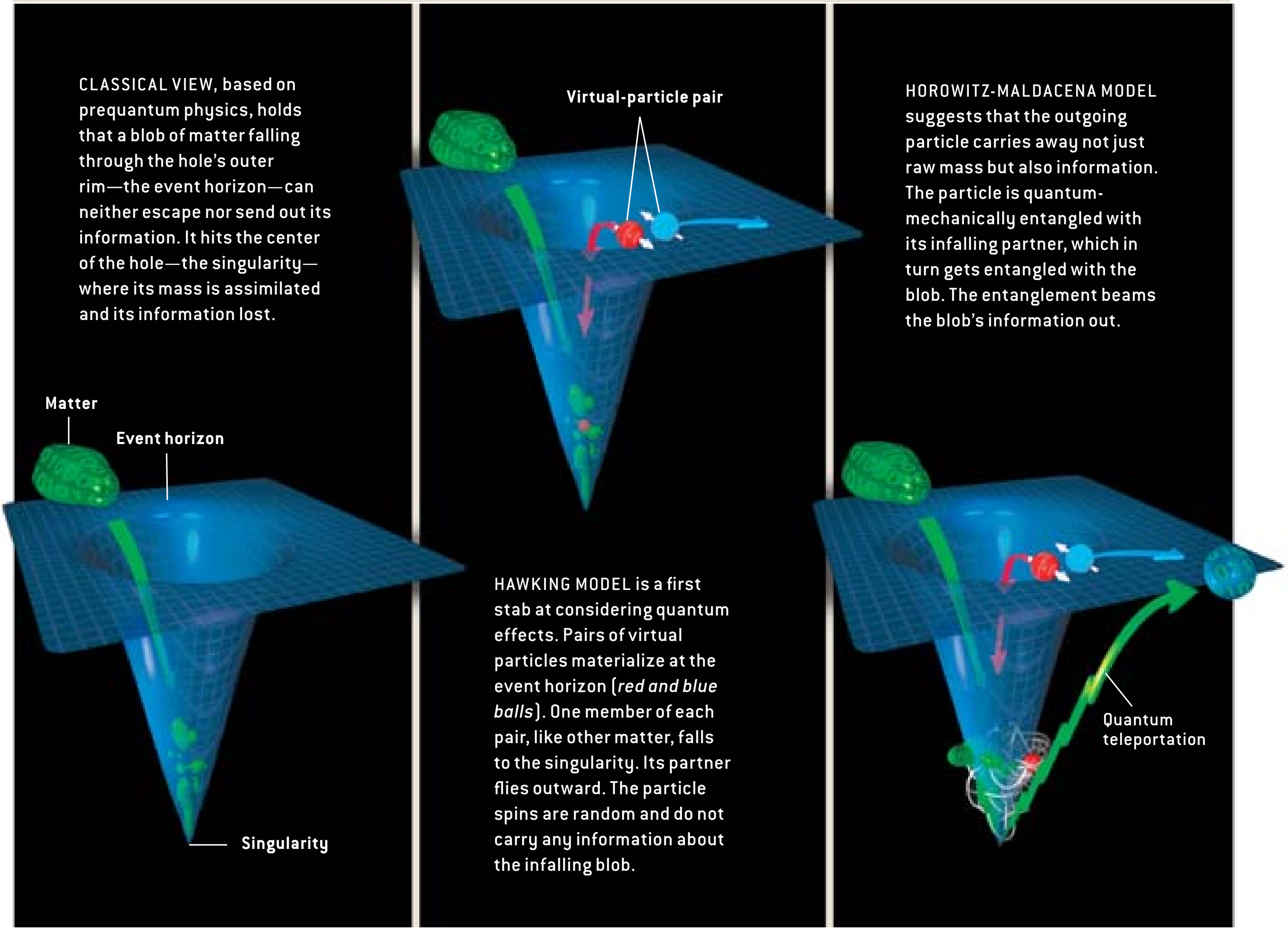
The teleportation procedure, which has been demonstrated in the laboratory, first requires that two particles be entangled. Then a measurement is performed on one of the particles jointly with some matter that contains information to be teleported. The measurement erases the information from its original location, but because of entanglement, that information resides in an encoded form on the second particle, no matter how distant it may be. The information can be decoded using the results of the measurement as the key [see "Quantum Teleportation," by Anton Zeilinger; SCIENTIFIC AMERICAN, April 2000].

A similar procedure might work for black holes. Pairs of entangled photons materialize at the event horizon. One of the photons flies outward to become the Hawking radiation that an observer sees. The other falls in and hits the singularity together with the matter that formed the hole in the first place. The annihilation of the infalling photon acts as a measurement, transferring the information contained in the matter to the outgoing Hawking radiation.

The difference from laboratory teleportation is that the results of this "mea-

Evolution of Black Hole Theory

"Objects so dense that nothing, not even light, can escape"—this definition of black holes has become a cliché of newspaper articles and freshman astronomy lectures. But it is probably wrong. Physicists have argued since the mid-1970s that energy can leak out of a black hole, and most now think that information (which describes the form that the energy takes) can, too. These diagrams show a black hole from a hypothetical viewpoint outside spacetime.



surement" are not needed to decode the information that was teleported. Horowitz and Maldacena argued that the annihilation does not have a variety of possible outcomes—only one. An observer on the outside can calculate this unique outcome using basic physics and thereby unlock the information. It is this conjecture that falls outside the usual formulation of quantum mechanics. Though controversial, it is plausible. Just as the initial singularity at the start of the universe may have had only one possible state, so it is possible that the final singularities inside black holes have a unique state. This past June one of us (Lloyd) showed that the Horowitz-Maldacena mechanism is robust; it does not depend on what exactly the final state is, as long as there is one. It still seems to lead to a small loss of information, however.

Other researchers have proposed escape mechanisms that also rely on weird quantum phenomena. In 1996 Andrew Strominger and Cumrun Vafa of Harvard University suggested that black holes are composite bodies made up of multidimensional structures called branes, which arise in string theory. Information falling into the black hole is stored in waves in the branes and can eventually leak out. Earlier this year Samir Mathur of Ohio State University and his collaborators modeled a black hole as a giant tangle of strings. This "fuzzyball" acts as a repository of the information carried by things that fall into the black hole. It emits radiation that reflects this information. Hawking, in his recent approach, has argued that quantum fluctuations prevent a well-defined event horizon from ever forming [see "Hawking a Theory," by Graham P. Collins; News Scan, October]. The jury is still out on all these ideas.

ALFRED T. KAMAJIAN

Understanding how information could LEAVE A BLACK HOLE is one of the liveliest questions in physics right now.

Cyberspacetime

THE PROPERTIES of black holes are inextricably intertwined with those of spacetime. Thus, if holes can be thought of as computers, so can spacetime itself. Quantum mechanics predicts that spacetime, like other physical systems, is discrete. Distances and time intervals cannot be measured to infinite precision; on small scales, spacetime is bubbly and foamy. The maximum amount of information that can be put into a region of space depends on how small the bits are, and they cannot be smaller than the foamy cells.

Physicists have long assumed that the size of these cells is the Planck length (l_P) of 10^{-35} meter, which is the distance at which both quantum fluctuations and gravitational effects are important. If so, the foamy nature of spacetime will always be too minuscule to observe. But as one of us (Ng) and Hendrik van Dam of the University of North Carolina at Chapel Hill and Frigyes Károlyházy of Eötvös Loránd University in Hungary have shown, the cells are actually much larger and, indeed, have no fixed size: the larger a region of spacetime, the larger its constituent cells. At first, this assertion may seem paradoxical—as though the atoms in an elephant were bigger than those in a mouse. In fact, Lloyd has derived it from the same laws that limit the power of computers.

The process of mapping the geometry of spacetime is a kind of computation, in which distances are gauged by transmitting and processing information. One way to do this is to fill a region of space with a swarm of Global Positioning System satellites, each containing a clock and a radio transmitter [*see illustration on next page*]. To measure a distance, a satellite sends a signal and times how long it takes to arrive. The precision of the measurement depends on how fast the clocks tick. Ticking is a computational operation, so its maximum rate is given by the Margolus-Levitin theorem: the time between ticks is inversely proportional to the energy.

The energy, in turn, is also limited. If you give the satellites too much energy or pack them too closely together, they will form a black hole and will no longer be able to participate in mapping. (The hole will still emit Hawking radiation, but that radiation has a wavelength the size of the hole itself and so is not useful for mapping features on a finer scale.) The maximum total energy of the constellation of satellites is proportional to the radius of the region being mapped.

Thus, the energy increases more slowly than the volume of the region does. As the region gets bigger, the cartographer faces an unavoidable tradeoff: reduce the density of satellites (so they are spaced farther apart) or reduce the energy available to each satellite (so that their clocks tick more slowly). Either way, the measurement becomes less precise. Mathematically, in the time it takes to map a region of radius R, the total number of ticks by all the satellites is R^2/l_P^2. If each satellite ticks precisely once during the mapping process, the satellites are spaced out by an average distance of $R^{1/3}l_P^{2/3}$. Shorter distances can be measured in one subregion but only at the expense of reduced precision in some other subregion. The argument applies even if space is expanding.

This formula gives the precision to which distances can be determined; it is applicable when the measurement apparatus is just on the verge of becoming a black hole. Below the minimum scale, spacetime geometry ceases to exist. That level of precision is much, much bigger than the Planck length. To be sure, it is still very small. The average imprecision in measuring the size of the observable universe is about 10^{-15} meter. Nevertheless, such an imprecision might be detectable by precise distance-measuring equipment, such as future gravitational-wave observatories.

From a theorist's point of view, the broader significance of this result is that it provides a new way to look at black holes. Ng has shown that the strange scaling of spacetime fluctuations with the cube root of distances provides a back-door way to derive the Bekenstein-Hawking formula for black hole memory. It also implies a universal bound for all black hole computers: the number of bits in the memory is proportional to the square of the computation rate. The proportionality constant is Gh/c^5—mathematically demonstrating the linkage between information and the theories of special relativity (whose defining parameter is the speed of light, c), general relativity (the gravitational con-

THE AUTHORS

SETH LLOYD and *Y. JACK NG* bridge the two most exciting fields of theoretical physics: quantum information theory and the quantum theory of gravity. Lloyd, professor of quantum-mechanical engineering at the Massachusetts Institute of Technology, designed the first feasible quantum computer. He works with various teams to construct and operate quantum computers and communications systems. Ng, professor of physics at the University of North Carolina at Chapel Hill, studies the fundamental nature of spacetime. He has proposed various ways to look for the quantum structure of spacetime experimentally. Both researchers say their most skeptical audience is their family. When Lloyd told his daughters that everything is made of bits, one responded bluntly: "You're wrong, Daddy. Everything is made of atoms, except light." Ng has lost credibility on the subject because he is always having to turn to his sons for help with his computer.

Computing Spacetime

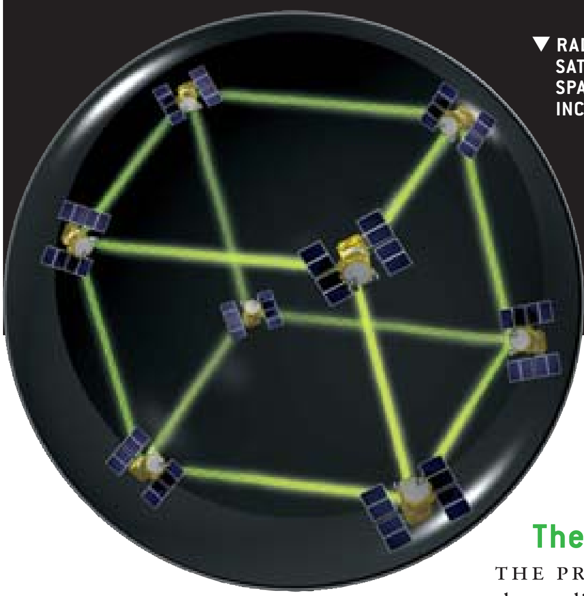
Measuring distances and time intervals is a type of computation and falls under the same constraints that computers do. It turns out that measurement is a much more slippery process than physicists had thought.

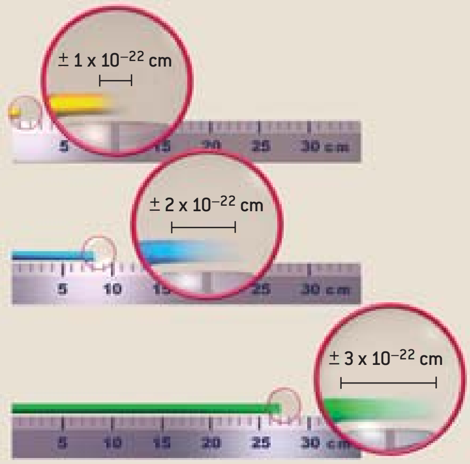
TO MAP A VOLUME of space, you might use a constellation of Global Positioning System satellites. They make measurements by sending signals and timing their arrival. For maximum precision, you need lots of satellites. But the number of satellites is limited: too many, and the entire system will collapse to a black hole.

To measure a region twice the size, you can use twice as many satellites. Because the volume is eight times as great, the satellites must be spaced farther apart. Each covers a larger subregion and can devote less attention to individual measurements, reducing their precision.

▲ RADIUS: 100 km
SATELLITES: 4
SPACING: 90 km

▼ RADIUS: 200 km
SATELLITES: 8
SPACING: 150 km
INCREASE IN ERROR: 26%





MEASUREMENT UNCERTAINTY is thus not fixed but can vary with the size of the object being measured. The larger the object is, the fuzzier its detailed structure. That differs from everyday life, in which the measurement imprecision is independent of the object and depends only on how finely subdivided your ruler is. It is as though your choice of what to measure affects the fine-scale structure of spacetime.

ALFRED T. KAMAJIAN

stant, G) and quantum mechanics (h).

Perhaps most significantly, the result leads directly to the holographic principle, which suggests that our three-dimensional universe is, in some deep but unfathomable way, two-dimensional. The maximum amount of information that any region of space can store seems to be proportional not to its volume but to its surface area [see "Information in the Holographic Universe," by Jacob D. Bekenstein; SCIENTIFIC AMERICAN, August 2003]. The holographic principle is normally thought to arise from the unknown details of quantum gravity, yet it also follows directly from the fundamental quantum limits to the precision of measurement.

The Answer Is ... 42

THE PRINCIPLES of computation can be applied not just to the most compact computers (black holes) and tiniest possible computers (spacetime foam) but also to the largest: the universe. The universe may well be infinite in extent, but it has existed a finite length of time, at least in its present form. The observable part is currently some tens of billions of light-years across. For us to know the results of a computation, it must have taken place within this expanse.

The above analysis of clock ticks also gives the number of operations that can have occurred in the universe since it began: 10^{123}. Compare this limit with the behavior of the matter around us—the visible matter, the dark matter and the so-called dark energy that is causing the universe to expand at an accelerated rate. The observed cosmic energy density is about 10^{-9} joule per cubic meter, so the universe contains 10^{72} joules of energy. According to the Margolus-Levitin theorem, it can perform up to 10^{106} operations per second, for a total of 10^{123} operations during its lifetime so far. In other words, the universe has performed the maximum possible number of operations allowed by the laws of physics.

To calculate the total memory capacity of conventional matter, such as atoms, one can apply the standard methods of statistical mechanics and cosmology. Matter can embody the most information when it is converted to energetic, massless particles, such as neutrinos or photons, whose entropy density is proportional to the cube of their temperature. The energy density of the par-

The universe has performed the **MAXIMUM POSSIBLE NUMBER OF OPERATIONS** allowed by the laws of physics.

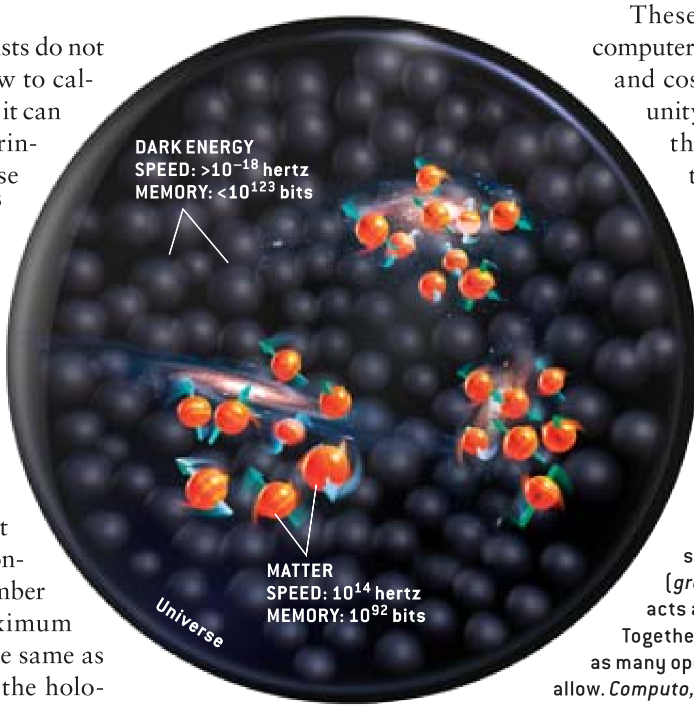
ticles (which determines the number of operations they can perform) goes as the fourth power of their temperature. Therefore, the total number of bits is just the number of operations raised to the three-fourths power. For the whole universe, that amounts to 10^{92} bits. If the particles contain some internal structure, the number of bits might be somewhat higher. These bits flip faster than they intercommunicate, so the conventional matter is a highly parallel computer, like the ultimate laptop and unlike the black hole.

As for dark energy, physicists do not know what it is, let alone how to calculate how much information it can store. But the holographic principle implies that the universe can store a maximum of 10^{123} bits—nearly the same as the total number of operations. This approximate equality is not a coincidence. Our universe is close to its critical density. If it had been slightly more dense, it might have undergone gravitational collapse, just like the matter falling into a black hole. So it meets (or nearly meets) the conditions for maxing out the number of computations. That maximum number is R^2/l_P^2, which is the same as the number of bits given by the holographic principle. At each epoch in its history, the maximum number of bits that the universe can contain is approximately equal to the number of operations it could have performed up to that moment.

Whereas ordinary matter undergoes a huge number of operations, dark energy behaves quite differently. If it encodes the maximum number of bits allowed by the holographic principle, then the overwhelming majority of those bits have had time to flip no more than once over the course of cosmic history. So these unconventional bits are mere spectators to the computations performed at much higher speeds by the smaller number of conventional bits. Whatever the dark energy is, it is not doing very much computation. It does not have to. Supplying the missing mass of the universe and accelerating its expansion are simple tasks, computationally speaking.

What is the universe computing? As far as we can tell, it is not producing a single answer to a single question, like the giant Deep Thought computer in the science-fiction classic *The Hitchhiker's Guide to the Galaxy.* Instead the universe is computing itself. Powered by Standard Model software, the universe computes quantum fields, chemicals, bacteria, human beings, stars and galaxies. As it computes, it maps out its own spacetime geometry to the ultimate precision allowed by the laws of physics. Computation is existence.

These results spanning ordinary computers, black holes, spacetime foam and cosmology are testimony to the unity of nature. They demonstrate the conceptual interconnections of fundamental physics. Although physicists do not yet possess a full theory of quantum gravity, whatever that theory is, they know it is intimately connected with quantum information. It from qubit. SA



UNIVERSE IS A COMPUTER consisting of two types of components. Matter (*red*) is highly dynamic; it acts as a high-speed parallel computer. Dark energy (*gray*) appears to be nearly static; it acts as a lower-speed serial computer. Together the components have performed as many operations as the laws of physics allow. *Computo, ergo sum.*

ALFRED T. KAMAJIAN

MORE TO EXPLORE

Ultimate Physical Limits to Computation. Seth Lloyd in *Nature,* Vol. 406, pages 1047–1054; August 31, 2000. Preprint available at **arxiv.org/abs/quant-ph/9908043**

From Computation to Black Holes and Space-Time Foam. Y. Jack Ng in *Physical Review Letters,* Vol. 86, No. 14, pages 2946–2949; April 2, 2001. Erratum, Vol. 88, No. 13, article 139902(E); March 14, 2002. **gr-qc/0006105**

Computational Capacity of the Universe. Seth Lloyd in *Physical Review Letters,* Vol. 88, No. 23, article 237901Z; June 10, 2002. **quant-ph/0110141**

The Black Hole Final State. Gary T. Horowitz and Juan Maldacena in *Journal of High Energy Physics,* JHEP02(2004)008, 2004. **hep-th/0310281**

Information: The New Language of Science. Hans Christian von Baeyer. Harvard University Press, 2004.

Abrupt Climate Change

By Richard B. Alley

In the Hollywood disaster thriller *The Day after Tomorrow,* a climate catastrophe of ice age proportions catches the world unprepared. Millions of North Americans flee to sunny Mexico as wolves stalk the last few people huddled in freeze-dried New York City. Tornadoes ravage California. Giant hailstones pound Tokyo.

Are overwhelmingly abrupt climate changes likely to happen anytime soon, or did Fox Studios exaggerate wildly? The answer to both questions appears to be yes. Most climate experts agree that we need not fear a full-fledged ice age in the coming decades. But sudden, dramatic climate changes have struck many times in the past, and they could happen again. In fact, they are probably inevitable.

Inevitable, too, are the potential challenges to humanity. Unexpected warm spells may make certain regions more hospitable, but they could magnify sweltering conditions elsewhere. Cold snaps could make winters numbingly harsh and clog key navigation routes with ice. Severe droughts could render once fertile land agriculturally barren. These consequences would be particularly tough to bear because climate changes that occur suddenly often persist for centuries or even thousands of years. Indeed, the collapses of some ancient societies—once attributed to social, economic and political forces—are now being blamed largely on rapid shifts in climate.

The specter of abrupt climate change has attracted serious scientific investigation for more than a decade, but it has only recently captured the interest of

Winter temperatures plummeting six degrees Celsius and sudden droughts scorching farmland around the globe are not just the stuff of scary movies. Such striking climate jumps have happened before—sometimes within a matter of years

SLIM FILMS

SUDDEN FLIP-FLOPS are an unavoidable element of earthly climate.

filmmakers, economists and policymakers. Along with more attention comes increasing confusion about what triggers such change and what the outcomes will be. Casual observers might suppose that quick switches would dwarf any effects of human-induced global warming, which has been occurring gradually. But new evidence indicates that global warming should be more of a worry than ever: it could actually be pushing the earth's climate faster toward sudden shifts.

Jumping Back and Forth

SCIENTISTS MIGHT NEVER have fully appreciated the climate's ability to lurch into a radically different state if not for ice cores extracted from Greenland's massive ice sheet in the early 1990s. These colossal rods of ice—some three kilometers long—entomb a remarkably clear set of climate records spanning the past 110,000 years. Investigators can distinguish annual layers in the ice cores and date them using a variety of methods; the composition of the ice itself reveals the temperature at which it formed.

Such work has revealed a long history of wild fluctuations in climate—long deep freezes alternating with brief warm spells. Central Greenland experienced cold snaps as extreme as six degrees Celsius in just a few years. On the other hand it achieved roughly half of the heating sustained since the peak of the last ice age—more than 10 degrees C—in a mere decade. That jump, which occurred about 11,500 years ago, is the equivalent of Minneapolis or Moscow acquiring the relatively sultry conditions of Atlanta or Madrid.

Not only did the ice cores reveal what happened in Greenland, but they also hinted at the situation in the rest of the world. Researchers had hypothesized that the 10-degree warming in the north was part of a warming episode across a broad swath of the Northern Hemisphere and that this episode enhanced precipitation in that region and far beyond. In Greenland itself, the thickness of the annual ice layers showed that, indeed, snowfall had doubled in a single year. Analyses of old air bubbles caught in the ice corroborated the prediction of increased wetness in other areas. In particular, measurements of methane in the bubbles indicated that this swamp gas was entering the atmosphere 50 percent faster during the intense warming than it had previously. The methane most likely entered the atmosphere as wetlands flooded in the tropics and thawed up north.

The cores also contained evidence that helped scientists fill in other details about the environment at various times. Ice layers that trapped dust from Asia indicated the source of prevailing winds, for instance. Investigators concluded that the winds must have been calmer during warm times because less wind-blown sea salt and ash from faraway volcanoes accumulated in the ice. And the list of clues goes on [see "Greenland Ice Cores: Frozen in Time," by Richard B. Alley and Michael L. Bender; SCIENTIFIC AMERICAN, February 1998].

Intense, abrupt warming episodes appeared more than 20 times in the Greenland ice records. Within several hundreds or thousands of years after the start of a typical warm period, the climate reverted to slow cooling followed by quick cooling over as short a time as a century. Then the pattern began again with another warming that might take only a few years. During the most extreme cold conditions, icebergs strayed as far south as the coast of Portugal. Lesser challenges probably drove the Vikings out of Greenland during the most recent cold snap, called the Little Ice Age, which started around A.D. 1400 and lasted 500 years.

Sharp warm-ups and cool-downs in the north unfolded differently elsewhere in the world, even though they may have shared a common cause. Cold, wet times in Greenland correlate with particularly cold, dry, windy conditions in Europe and North America; they also coincide with anomalously warm weather in the South Atlantic and Antarctica. Investigators pieced together these regional histories from additional clues they found in the ice of high mountain glaciers, the thickness of tree rings, and the types of pollen and shells preserved in ancient mud at the bottoms of lakes and oceans, among other sources.

The evidence also revealed that abrupt shifts in rainfall have offered up challenges rivaling those produced by tem-

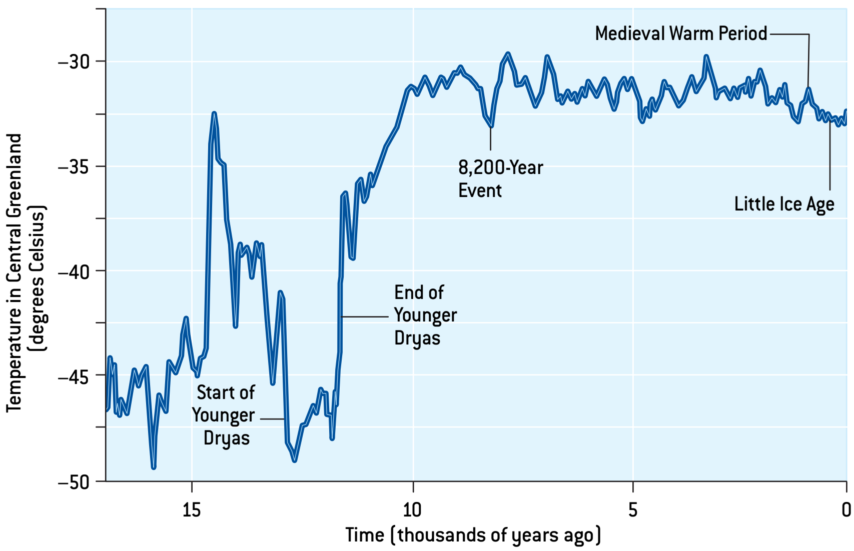
Global warming should be more of a worry than ever: it could be pushing the earth's climate faster toward sudden shifts.

Overview/*Inevitable Surprises*

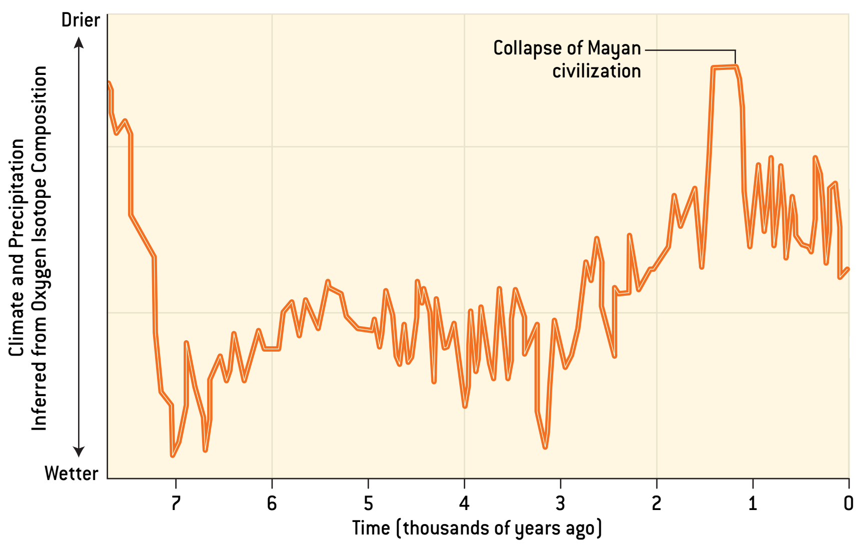
- Most climate change policy discussions and research efforts center on global warming. But another problem looms as well: climate has suddenly flip-flopped in the past and will surely do so again.
- A regional drought could, for instance, arrive one summer and stay for decades, wiping out rich agricultural lands across Asia and North America; weather patterns in Europe could shift in a matter of years, making that area's climate more like Siberia's.
- Scientists cannot yet predict when such abrupt changes will occur, but most climate experts warn that global warming and human activities may be propelling the world faster toward sudden, long-lasting climate changes.

PAST AS PROLOGUE?

Abrupt climate change has marked the earth's history for eons. Ice cores from Greenland, for instance, reveal that wild temperature swings (*top left*) punctuated the gradual warming that brought the planet out of the last ice age starting about 18,000 years ago. Fossil shells in lake sediments from Mexico's Yucatán Peninsula record sudden and severe droughts (*bottom left*) because a diagnostic ratio of oxygen isotopes in the shells shoots up when more water evaporates from the lake than falls as rain. Societies have often suffered as a result of these rapid shifts (*photographs*).



Viking settlement, now in ruins, was among those in Greenland abandoned during an abrupt cold spell called the Little Ice Age.



Mayan rain god (*statue in foreground*) was apparently no match for the drought now widely blamed for the collapse of Mayan civilization about 1,100 years ago.

ELIZA JEWETT; SOURCES: M. STUIVER ET AL. AND K. CUFFEY ET AL. (*top graph*) AND D. HODELL ET AL. (*bottom graph*) *National Geophysical Data Center*; RUDY BRUEGGEMANN (*top photo*); FREELANCE CONSULTING SERVICES PTY LTD/CORBIS (*bottom photo*)

perature swings. Cold times in the north typically brought drought to Saharan Africa and India. About 5,000 years ago a sudden drying converted the Sahara from a green landscape dotted with lakes to the scorching, sandy desert it is today. Two centuries of dryness about 1,100 years ago apparently contributed to the end of classic Mayan civilization in Mexico and elsewhere in Central America. In modern times, the El Niño phenomenon and other anomalies in the North Pacific occasionally have steered weather patterns far enough to trigger surprise droughts, such as the one responsible for the U.S. dust bowl of the 1930s.

Point of No Return

BE THEY WARM SPELLS, cold snaps or prolonged droughts, the precipitous climate changes of the past all happened for essentially the same reason. In each case, a gradual change in temperature or other physical condition pushed a key driver of climate toward an invisible threshold. At the point that threshold was crossed, the climate driver—and thus the climate as well—flipped to a new and different state and usually stayed there for a long time [*see box on next page*].

Crossing a climate threshold is similar to flipping a canoe. If you are sitting in a canoe on a lake and you lean gradually to one side, the canoe tips, too. You are pushing the canoe toward a threshold—the position after which the boat

CROSSING THE THRESHOLD

Global warming alters ambient conditions little by little. But even this kind of slow, steady change can push climate drivers, such as well-established ocean currents or patterns of rainfall, to a critical point at which they lurch abruptly into a new and different state. That switch brings with it an associated shift in climate—with potentially challenging consequences to people and societies. Once a climate driver crosses its so-called threshold, the changes that ensue could persist for millennia. Many thresholds may still await discovery; here are three that scientists have identified:

CLIMATE DRIVER	THRESHOLD CROSSING	RESULTING CLIMATE SHIFT	SOCIAL CONSEQUENCES
Ocean currents in the North Atlantic carry warmth northward from tropics, keeping western Europe's winters mild (*see box on opposite page*).	Freshening of surface waters in the far north slows down these currents, possibly stopping them altogether.	Temperatures plummet in the region, and climate in Europe and the eastern U.S. becomes more like Alaska's.	Agriculture suffers in regions around the world, and key navigation routes become clogged with ice.
Rainwater that is recycled through plants (absorbed by their roots and returned to the air through evaporation from their leaves) provides much of the precipitation in the world's grain belts.	A minor dry spell wilts or kills too many plants, and recycled rainfall disappears, reinforcing the drying in a vicious cycle.	A potentially mild dry spell is enhanced and prolonged into a severe drought.	Parched land can no longer support crops; famine strikes those who cannot trade for the remaining grain in the world market.
Currents in the Pacific Ocean determine major patterns of sea-surface temperature, which in turn control regional weather patterns.	Natural phenomena, such as El Niño, cause subtle changes in sea-surface temperatures, although scientists are still not sure why.	Weather patterns on adjacent continents shift, triggering severe storms or droughts where they typically do not occur.	Some croplands dry up while other places incur damage from intense storms.

can no longer stay upright. Lean a bit too far, and the canoe overturns.

Threshold crossings caused history's most extreme climate flips—and point to areas of particular concern for the future. To explain the icy spells recorded in Greenland's ice cores, for example, most scientists implicate altered behavior of currents in the North Atlantic, which are a dominant factor in that region's long-term weather patterns.

Eastern North America and Europe enjoy temperate conditions (like today's) when salty Atlantic waters warmed by southern sunshine flow northward across the equator. During far northern winters, the salty water arriving from the south becomes cold and dense enough to sink east and west of Greenland, after which it migrates southward along the seafloor. Meanwhile, as the cooled water sinks, warm currents from the south flow northward to take its place. The sinking water thereby drives what is called a conveyor belt circulation that warms the north and cools the south.

Ice cores contain evidence that sudden cold periods occurred after the North Atlantic became less salty, perhaps because freshwater lakes burst through the walls of glaciers and found their way to the sea. Researchers identify this rush of freshwater as the first phase of a critical threshold crossing because they know freshening the North Atlantic can slow or shut off the conveyor, shifting climate as a result.

Diluted by water from the land, seawater flowing in from the south would become less salty and thus less dense, possibly to the point that it could freeze into sea ice before it had a chance to sink. With sinking stopped and the conveyor halted, rain and snow that fell in the north could not be swept into the deep ocean and carried away. Instead they would accumulate on the sea surface and freshen the North Atlantic even more. The conveyor then would stay quiet, leaving nearby continents with climates more like Siberia's [*see box on opposite page*].

Chilling Warmth

EIGHT THOUSAND YEARS have passed since the last of the biggest North Atlantic cold snaps. Could it be that humans are actually "leaning" in the right direction to avoid flipping the climate's canoe? Perhaps, but most climate experts suspect instead that we are rocking the boat—by changing so many aspects of our world so rapidly. Particularly worrisome are human-induced increases in atmospheric concentrations of greenhouse gases, which are promoting global warming [see "Defusing the Global Warming Time Bomb," by James Hansen; SCIENTIFIC AMERICAN, March; www.sciam.com/ontheweb].

The United Nations–sanctioned Intergovernmental Panel on Climate Change has predicted that average global temperatures will rise 1.5 to 4.5 degrees C in the next 100 years. Many computer models that agree with this assessment also predict a slowdown of the North Atlantic conveyor. (As ironic as it may sound, gradual warming could lead to a sudden cooling of many degrees.) Uncertainties abound, and although a new ice age is not thought credible, the resulting changes could be notably larger than they were during the Little Ice Age, when the Thames in London froze and glaciers rumbled down the Alps.

THE AUTHOR

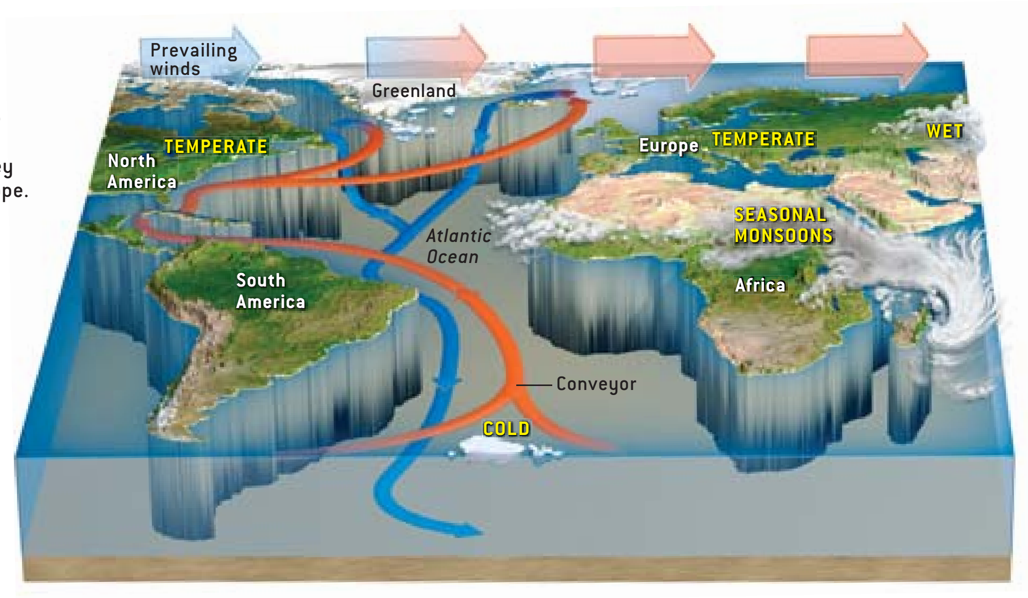
RICHARD B. ALLEY is professor of geosciences at Pennsylvania State University and an associate at the Earth System Science Center there. Since earning his Ph.D. in geology from the University of Wisconsin–Madison in 1987, Alley has focused his work on understanding how glaciers and ice sheets record climate change, how their flow affects sea level, and how they erode and deposit sediment. Alley has spent three field seasons in Antarctica and five in Greenland. During those visits, he helped to harvest many of the key ice cores that are revealing the history of the earth's climate over the past 110,000 years. In 2001 Alley won the national Phi Beta Kappa Science Award for his nontechnical book *The Two-Mile Time Machine,* which details the results from a two-mile-long ice core from Greenland.

MELTING TOWARD A COLD SNAP?

As global warming continues to heat up the planet, many scientists fear that large pulses of freshwater melting off the Greenland ice sheet and other frozen northern landmasses could obstruct the so-called North Atlantic conveyor, the system of ocean currents that brings warmth to Europe and strongly influences climate elsewhere in the world. A conveyor shutdown—or even a significant slowdown—could cool the North Atlantic region even as global temperatures continue to rise. Other challenging and abrupt climate changes would almost certainly result.

CONVEYOR ON

Salty ocean currents (*red*) flowing northward from the tropics warm prevailing winds (*large arrows*) as they blow eastward toward Europe. The heat-bearing currents, which are dense, become even denser as they lose heat to the atmosphere. Eventually the cold, salty water becomes dense enough to sink near Greenland. It then migrates southward along the seafloor (*blue*), leaving a void that draws more warm water from the south to take its place.

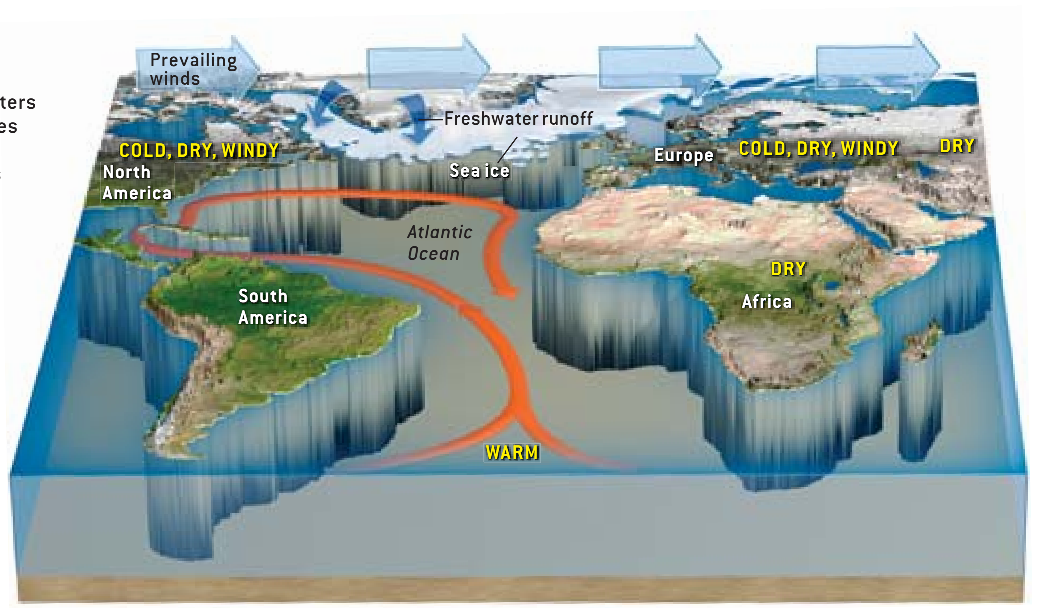


RESULTING CLIMATE

When the North Atlantic conveyor is active, temperate conditions with relatively warm winters enable rich agricultural production in much of Europe and North America. Seasonal monsoons fuel growing seasons in broad swaths of Africa and the Far East. Central Asia is wet, and Antarctica and the South Atlantic are typically cold.

CONVEYOR OFF

If too much freshwater enters the North Atlantic, it dilutes the salty currents from the south. Surface waters no longer become dense enough to sink, no matter how cold the water gets, and the conveyor shuts down or slows. Prevailing winds now carry frigid air eastward (*large arrows*). This cold trend could endure for decades or more—until southern waters become salty enough to overwhelm the fresher water up north, restarting the conveyor in an enormous rush.



RESULTING CLIMATE

As the conveyor grows quiet, winters become harsher in much of Europe and North America, and agriculture suffers. These regions, along with those that usually rely on seasonal monsoons, suffer from droughts sometimes enhanced by stronger winds. Central Asia gets drier, and many regions in the Southern Hemisphere become warmer than usual.

DAVID FIERSTEIN

Tricky Predictions

No credible predictions of abrupt climate changes have ever been issued—nor should you expect one anytime soon. That is because rapid changes are inherently more difficult to forecast than global warming or other gradual changes.

One major stumbling block has to do with the very nature of abrupt change. A rapid shift occurs when a slow but steady force, such as global warming, moves a crucial component of the climate system past a point of no easy return. Crossing such a threshold triggers a sudden switch to a new state—much the way leaning over too far in a canoe suddenly dumps you in the lake. Knowing exactly how far you can tip the canoe without overturning is almost impossible, however, especially as wind and waves rock the boat. Similarly, it is exceedingly tough to recognize when an aspect of climate is approaching a critical threshold.

Researchers have attempted to gain insight into the factors that operate near a tipping point through computer modeling. These efforts have revealed much about what rocks the climate's canoe, but uncertainties still abound. To determine how accurately their computer models will forecast climate change, scientists check to see how well the programs simulate real-world changes from the past. Many models match the basic types of previous climate anomalies reasonably well—in other words, they reliably reproduce cold spells or droughts or flooding at the locations and times that have been recorded in annual layers of ice and sediment. Some of them even do a decent job of simulating shifts in storm tracks, wind patterns, seasonal precipitation and other finer details.

BALANCING ACT: The earth typically experiences a given climate for millennia or more. Then, at a moment that is nearly impossible to predict, some aspect of the climate system teeters too far to one side and global conditions tumble into a starkly different state.

But even if the models get the general nature of the climate change right, they represent important parameters imperfectly. In particular, the abrupt changes of the past were usually larger and more widespread than the models indicate. Most of them underestimate the amount of moisture lost in the Sahara over the past few thousand years, for example. The models also seem to have trouble simulating both the great warmth of the polar regions during the time of the dinosaurs and the extreme cold at the peak of the most recent ice age.

The simplest reason for these mismatches may be that the models are typically less sensitive than the climate is, perhaps because they omit key feedbacks and responses. Climate thresholds that no one has yet considered might also explain the discrepancies. Locating these thresholds would undoubtedly prove helpful, even though more potential climate flips could be unveiled in the process. On the bright side, new discoveries might instead reveal that the likelihood for abrupt change is actually lower than scientists suspect or that one change might offset another. *—R.B.A.*

Perhaps of greater concern than cold spells up north are the adverse effects that would probably strike other parts of the world concurrently. Records of climate across broad areas of Africa and Asia that typically benefit from a season of heavy monsoons indicate that these areas were particularly dry whenever the North Atlantic region was colder than the lands around it. Even the cooling from a conveyor slowdown might be enough to produce the drying. With billions of people relying on monsoons to water crops, even a minor drought could lead to widespread famine.

The consequences of future North Atlantic freshening and cooling may make life more difficult even for people living in regions outside the extreme cold or drought. Unease over such broad impacts spurred the U.S. Department of Defense to request that a think tank called the Global Business Network assess the possible national security implications of a total shutdown of the North Atlantic conveyor. Many scientists, including me, think that a moderate slowdown is much more likely than a total shutdown; either way, the seriousness of the potential outcome makes considering the worst-case implications worthwhile. As the Global Business Network report states, "Tensions could mount around the world.... Nations with the resources to do so may build virtual fortresses around their countries, preserving resources for themselves. Less fortunate nations ... may initiate in struggles for access to food, clean water, or energy."

Floods and Droughts

EVEN IF A SLOWDOWN of the North Atlantic conveyor never happens, global warming could bring about troubling threshold crossings elsewhere. The grain belts that stretch across the interiors of many midlatitude continents face a regional risk of prolonged drought. Most climate models produce greater summertime drying over these areas as average global temperatures rise, regardless of what happens in the North Atlantic. The same forecasts suggest that greenhouse-induced warming will increase rainfall overall, possibly in the form of more severe storms and flooding; however, those events—significant problems on their own—are not expected to offset the droughts.

Summer drying could cause a relatively mild drought to worsen and persist for decades or more. This transition would occur because of a vulnerability of the grain belts: for precipitation, they rely heavily on rainfall that local plants recycle rather than on new moisture

MATT COLLINS

blown in from elsewhere. The plants' roots normally absorb water that would otherwise soak through the ground to streams and flow to the sea. Some of that water then returns to the air by evaporating through their leaves. As the region begins to suffer drier summers, however, the plants wilt and possibly die, thereby putting less water back into the air. The vital threshold is crossed when the plant population shrinks to the point that the recycled rainfall becomes too meager to sustain the population. At that point more plants die, and the rainfall diminishes further—in a vicious cycle like the one that turned the Sahara into a desert 5,000 years ago. The region has shown no signs of greening ever since.

Scientists fear they have not yet identified many of the thresholds that, when crossed, would lead to changes in regional climates. That knowledge gap is worrisome, because humans could well be doing many things to tip the climate balance in ways we will regret. Dancing in a canoe is not usually recommended, yet dance we do: We are replacing forests with croplands, which increases how much sunlight the land reflects; we are pumping water out of the ground, which changes how much water rivers carry to the oceans; and we are altering the quantities of trace gases and particulates in the atmosphere, which modifies the characteristics of clouds, rainfall and more.

Facing the Future

NEGATIVE CONSEQUENCES of a major climate shift can be mitigated if the change occurs gradually or is expected. Farmers anticipating a drought can drill wells, or learn to grow crops less dependent on water, or simply cut their losses and move elsewhere. But unexpected change can be devastating. A single, surprise drought year may at first bankrupt or starve only the most marginal farmers, but damage worsens as the drought lengthens—especially if no one had time to prepare. Unfortunately, scientists have little ability to predict when abrupt climate change will occur and what form it will take [*see box on opposite page*].

Despite the potentially enormous consequences of a sudden climate transformation, the vast majority of climate research and policymaking has addressed gradual shifts—most notably by calling for global reductions of carbon emissions as a way to slow the gradual rise in global temperatures. Although such reductions would probably help limit climate instability, thought should also be given specifically to avoiding abrupt changes. At one extreme, we might decide to ignore the prospect altogether and hope that nothing happens or that we are able to deal with whatever does happen; business-as-usual did sink the *Titanic,* but many other unprepared ships have crossed the North Atlantic unscathed. On the other hand, we might seriously alter our behavior to keep the human effects on climate small enough to make a catastrophic shift less likely. Curbing global warming would be a step in the right direction. Further investigation of climate thresholds and their vulnerabilities to human activities should illuminate other useful actions.

A third strategy would be for societies to shore up their abilities to cope with abrupt climate change before the next surprise is upon us, as suggested by the U.S. National Research Council. The authors of the council's report pointed out that some former societies have bent in response to climate change when others have broken. The Viking settlers in Greenland abandoned their weakening settlement as the onset of the Little Ice Age made their way of life marginal or unsustainable, while their neighbors, the Thule Inuit, survived. Understanding what separates bending from breaking could prove constructive. Plans designed to help ease difficulties if a crisis develops could be made at little or no cost. Communities could plant trees now to help hold soil during the next windy dry spell, for example, and they could agree now on who will have access to which water supplies when that resource becomes less abundant.

For now, it appears likely that humans are rocking the boat, pushing certain aspects of climate closer to the thresholds that could unleash sudden changes. Such events would not trigger a new ice age or otherwise rival the fertile imaginations of the writers of the silver screen, but they could pose daunting challenges for humans and other living things on earth. It is well worth considering how societies might increase their resiliency to the potential consequences of an abrupt shift—or even how to avoid flipping the climate canoe in the first place. SA

It appears likely that humans are pushing certain aspects of climate closer to the thresholds that could unleash sudden changes.

MORE TO EXPLORE

Climate Change 2001: The Scientific Basis. Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, 2001. Available at **www.ipcc.ch**

Abrupt Climate Change: Inevitable Surprises. National Research Council, Committee on Abrupt Climate Change. Richard B. Alley, chair. National Academy Press, 2002.

Global Climatic Impacts of a Collapse of the Atlantic Thermohaline Circulation. Michael Vellinga and Richard A. Wood in *Climatic Change,* Vol. 54, No. 3, pages 251–267; August 2002.

Rapid Climate Change. Spencer Weart in *The Discovery of Global Warming.* Harvard University Press, 2003. Essay available at **www.aip.org/history/climate/pdf/rapid.pdf**

An Abrupt Climate Change Scenario and Its Implications for United States National Security. Peter Schwartz and Doug Randall. October 2003. Available at **www.gbn.org/ArticleDisplayServlet.srv?aid=26231**

MODIFIED MINUTEMAN MISSILE was the interceptor rocket in a series of controversial tests of the national missile defense system from 1997 to 2002. Fired from Kwajalein Atoll in the central Pacific Ocean, the interceptors destroyed mock warheads in five of eight attempts, but the tests did not realistically simulate a ballistic-missile attack.

Holes in the Missile Shield

The national missile defense now being deployed by the U.S. should be replaced with a more effective system

By Richard L. Garwin

Photography by Paul Shambroom

This fall, perhaps by the time you read this, President George W. Bush is expected to declare that the first phase of the long-awaited national missile defense is operational. The Pentagon's Missile Defense Agency (MDA) plans to install six interceptor rockets—designed to strike a ballistic missile in midcourse—in silos at Fort Greely in Alaska by mid-October. Ten more will be deployed at Fort Greely and four more at Vandenberg Air Force Base in California by the end of 2005. Over the following years the MDA intends to bolster this rudimentary midcourse defense with more interceptors, advanced radars and surveillance satellites. The reason for the deployment is to counter the threat that a rogue state—namely, North Korea or Iran—will attempt to hit the U.S. with nuclear or biological weapons delivered on intercontinental ballistic missiles (ICBMs).

But despite the more than $80 billion spent by the U.S. on missile defense since 1985, this system will not provide significant protection for many years, if ever. The political pressure to claim that the U.S. is secure against a rogue nation's attack has led to a defense that will not counter even the earliest threats from the emerging missile powers. The MDA's midcourse system is built to intercept long-range missiles fired thousands of kilometers from the U.S.; it can do nothing to stop a short- or medium-range missile launched from a ship off America's coasts. What is more, the interceptor rockets would most likely prove inadequate against long-range missiles as well, because an enemy could easily equip its ICBMs with fundamentally simple and highly effective countermeasures.

A strong defense against ballistic missiles is a worthy goal. The destructive capacity of nuclear warheads is so enormous that it would be unconscionable not to explore methods for preventing them from hitting the U.S. But instead of rushing to construct a flawed system, military contractors, technologists and politicians should pay more attention to evaluating the relative magnitudes of the threats and assessing the capabilities

of the proposed defenses. The Pentagon must focus on the more immediate danger of short- and medium-range ballistic and cruise missiles, and the funds now being devoted to the MDA's midcourse defense should be shifted to the development of alternative systems that would have a real chance of stopping ICBMs.

Missile Defense Basics

MISSILE INTERCEPTION schemes can be divided into three broad categories: terminal, boost phase and midcourse. Terminal defense involves stopping the missile's warhead in the final phase of its trajectory, typically when it is less than a minute away from its target. An important consideration for terminal defense is making sure that the intercept occurs before the warhead gets close enough to damage the target. Thus, the task of protecting a city's buildings and people is far more difficult than that of preserving enough hardened concrete missile silos to retaliate against a first strike (thereby deterring such an attack). When defending a city, the interceptors must preclude a nuclear blast from a larger area and destroy the warheads at a higher altitude. And because a city is much more valuable than a missile silo, the reliability of the intercept must be correspondingly higher.

For example, a one-megaton nuclear warhead would need to be intercepted at an altitude of at least 10 kilometers to prevent the city from being incinerated by the heat of the hydrogen bomb. Further, the interceptor rocket could not be launched until the warhead enters the atmosphere, allowing the defense to distinguish between the heavy weapon and any light decoys accompanying it. These constraints mean that the interceptors cannot be based more than 50 kilometers from the city. So unless the Pentagon is prepared to carpet the nation with interceptors, it is clear that terminal defense is not an appropriate response to the threat of a few nuclear-armed ICBMs. Even a perfect defense of many cities would simply lead to the targeting of an undefended city.

Boost-phase intercept requires disabling the ballistic missile in the first few minutes of its flight, while it is still ascending. This strategy puts high demands on the interceptor. After liftoff, a typical ICBM arcs upward at an average acceleration of about three *g*'s (three times the acceleration of gravity at the earth's surface), reaching a speed of seven kilometers per second in 250 seconds. Suppose the interceptor has 200 seconds to get to the ICBM (that is, it is launched within a minute of the ICBM's launch) and must travel 500 kilometers from its base to hit the enemy missile. The interceptor would be able to cover the distance if it accelerates at a constant rate of about three and a half *g*'s for the first 100 seconds, then moves at a burnout speed of 3.33 kilometers per second for the next 100 seconds. If the interceptor has to fly 1,000 kilometers, the acceleration and burnout speed would need to be twice as great.

To down ICBMs launched anywhere in North Korea, boost-phase interceptors based on ships off the country's coast or in a neighboring nation might have to travel as far as 1,000 kilometers, so they would need a burnout speed of six to eight kilometers per second. Iran, though, is a much larger country, so interceptors would need burnout speeds of about 10 kilometers per second. And because the interceptors must reach this speed in as little as 50 seconds, they would require an average acceleration of 20 *g*'s. Building such interceptors would not be an insuperable task; in the 1960s the U.S. tested a small missile that briefly accelerated at an average rate of 260 *g*'s. Boost-phase interceptors show more promise than an alternative called the airborne laser, which would attempt to disable ICBMs by focusing laser beams on them as they ascend. Extremely expensive to build and operate, the aircraft-mounted laser would have no effect on missiles more than 300 or so kilometers away.

Boost-phase intercept becomes still more difficult when defending against ICBMs launched from Russia or China. These countries are so vast that offshore interceptors could not reach the missiles while they are ascending. The interceptors would have to be placed in orbit, which greatly increases the expense of the system. Although space-based boost-phase interceptors were part of the original Strategic Defense Initiative proposed by President Ronald Reagan two decades ago, by the late 1990s the Pentagon had focused instead on midcourse systems, which attempt to destroy a missile's warhead while it is hurtling above the atmosphere at the top of its arc. For this reason, the missile defense now being deployed by the MDA is the most mature technology—but it is definitely not the most effective.

The U.S. system will not counter even the earliest threats from the emerging missile powers.

Overview/*National Missile Defense*

- The missile defense system now being deployed relies on intercepting an intercontinental ballistic missile in midcourse, while the warhead is hurtling hundreds of kilometers above the earth's surface.
- The system's main weakness is that an attacker could easily load a ballistic missile with dozens of decoys. Unable to distinguish between the decoys and the warhead, the interceptors would soon be overwhelmed.
- A more effective strategy is to try to destroy ballistic missiles while they are still ascending. But the U.S. must also deploy a defense against short- and medium-range missiles fired from ships off the nation's coasts.

PAUL SHAMBROOM *nukephoto.com*

PROTOTYPE X-BAND RADAR located at Kwajalein Atoll was used in the tests of the missile defense system to track the mock warheads. Because the prototype's tracking abilities were limited, the interceptors also received location data from radar beacons placed on the warheads.

Hit to Kill

HOW WOULD the MDA's system work? The first indication of the launch of an ICBM against the U.S. would come from military satellites designed to detect the hot flame of a large rocket motor. In operation since the 1970s, these Defense Support Program (DSP) satellites are located 36,000 kilometers above the earth in geosynchronous orbits—because each has an orbital period of 24 hours, it remains fixed over a single point on the equator. Together the satellites view almost the entire planet in the infrared part of the spectrum and are thus able to see the launch of every ballistic missile of significant size and range. Each DSP satellite scans the earth's surface every 10 seconds, providing a rough position for the ICBM—to an accuracy of about one kilometer—from the time the missile breaks cloud cover at an altitude of about 10 kilometers until the burnout of its rocket motor some 200 to 300 seconds later. Over parts of the earth there is stereo DSP coverage.

The MDA's system would fire several interceptors against each missile to destroy it in midcourse, long after its engine had stopped firing and its nuclear warhead had separated from the rocket. The intercept would take place in the vacuum of space, hundreds of kilometers above the earth's surface, and the target would be the warhead, which is encased within a reentry vehicle to protect it against the fiery heat of reentering the atmosphere. Because the interceptors must be given a precise aim point in space and time to allow them to home in on the warhead, and because the DSP satellites cannot detect a missile after its rocket engine shuts off, midcourse intercept requires accurate radar tracking of the missile's trajectory.

To cover the North Pacific trajectories that would most likely be used by North Korean ICBMs, the Pentagon originally planned to build an advanced radar station on Shemya Island at the tip of Alaska's Aleutian chain. The strong winds and violent seas of that area made on-site construction difficult, though, so the MDA is now building the $900-million radar on a floating platform off the coast of Texas. On its expected completion by the end of 2005, it will be towed to its operational site near the Aleutian chain. (Until then, the MDA's system must rely on the U.S. Air Force's Cobra Dane early-warning radar on Shemya Island.) The new radar will send out microwaves in the X-band part of the spectrum. With a wavelength of about three centimeters, these waves are shorter than those used by conventional radars. Such short waves are necessary to narrow the radar's beam so that it can pinpoint a missile's warhead amid the "threat cloud," which also includes the ICBM's last rocket stage and, most likely, dozens of inflatable balloon decoys designed to mimic the warhead to radar and even visible or infrared sensors.

Of course, the new X-band radar near the Aleutians would be of no use in protecting the U.S. against ICBMs launched from Iran, which would fly over Europe and the North Atlantic. The MDA plans to remedy this deficiency in the coming years by deploying improved sensors and interceptors at a wider array of sites. Ultimately, the DSP system will be supplanted by a new space-based infrared system consisting of satellites in geosynchronous orbit that will provide better tracking of ICBMs during their ascent. In addition, the Pentagon is developing a constellation of satellites in low earth orbit that would be able to track missiles during the midcourse phase using infrared and visible-light sensors. The communications network that links the satellites, radar and interceptors with the command and control center at Cheyenne Mountain in Colorado will also be upgraded in stages.

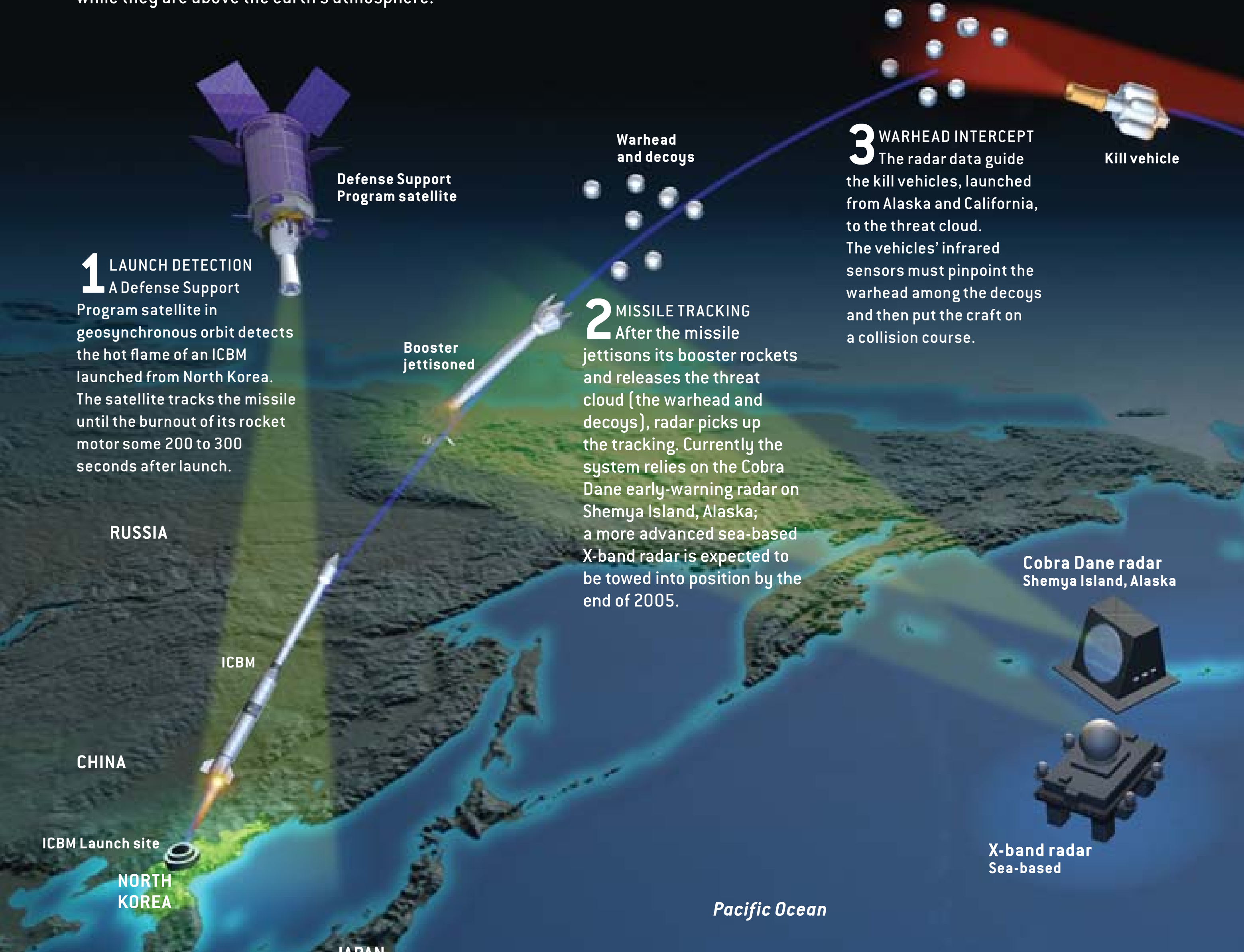
Earlier missile defense approaches—the Safeguard system operated by the U.S. in the mid-1970s to protect 150 ICBM silos in North Dakota and the still operating Russian system that defends Moscow—relied on nuclear-armed interceptors that were designed to detonate once they were close enough to the enemy missile to obliterate it. But guidance systems have improved so much in the past few decades that it has become possible for interceptors to destroy ballistic missiles simply by colliding with them. This technique avoids the need for nuclear

THE AUTHOR

RICHARD L. GARWIN has worked with the U.S. government since 1950 on the technology of nuclear weapons, missiles, air defense and missile defense. He is an experimental physicist with a research background in nuclear and particle physics, condensed-matter physics and the detection of gravity waves. A recipient of the National Medal of Science and the Enrico Fermi Award, he chaired the U.S. State Department's Arms Control and Nonproliferation Advisory Board from 1994 to 2001. In 1998 he was one of the nine members of the Rumsfeld Commission to Assess the Ballistic Missile Threat to the United States.

Stopping a Missile with a Missile

The Pentagon's midcourse intercept system is designed to destroy incoming warheads while they are above the earth's atmosphere.



1 LAUNCH DETECTION A Defense Support Program satellite in geosynchronous orbit detects the hot flame of an ICBM launched from North Korea. The satellite tracks the missile until the burnout of its rocket motor some 200 to 300 seconds after launch.

2 MISSILE TRACKING After the missile jettisons its booster rockets and releases the threat cloud (the warhead and decoys), radar picks up the tracking. Currently the system relies on the Cobra Dane early-warning radar on Shemya Island, Alaska; a more advanced sea-based X-band radar is expected to be towed into position by the end of 2005.

3 WARHEAD INTERCEPT The radar data guide the kill vehicles, launched from Alaska and California, to the threat cloud. The vehicles' infrared sensors must pinpoint the warhead among the decoys and then put the craft on a collision course.

X-BAND RADAR

Because the intercept system must precisely track the warheads and decoys, the X-band radar steers a narrow beam by shifting the phases of the signals transmitted from thousands of antenna elements. The radar can distinguish objects as close as 15 centimeters from one another, but if the warheads and decoys are enclosed in radar-reflecting aluminized balloons, the system may not be able to determine which is which. The floating platform for the radar will be roughly the size of two football fields; the radar's face (*at right*) will be about 12 to 15 meters wide.

KILL VEHICLE

Built by Raytheon, the 64-kilogram kill vehicle is about 140 centimeters long and 60 centimeters in diameter. Its infrared seeker, designed to guide the vehicle toward its target in the final seconds before the intercept, is attached to a telescope (*top of photograph at right*). The vehicle has four side thrusters for adjusting its course.

COMMAND CENTER

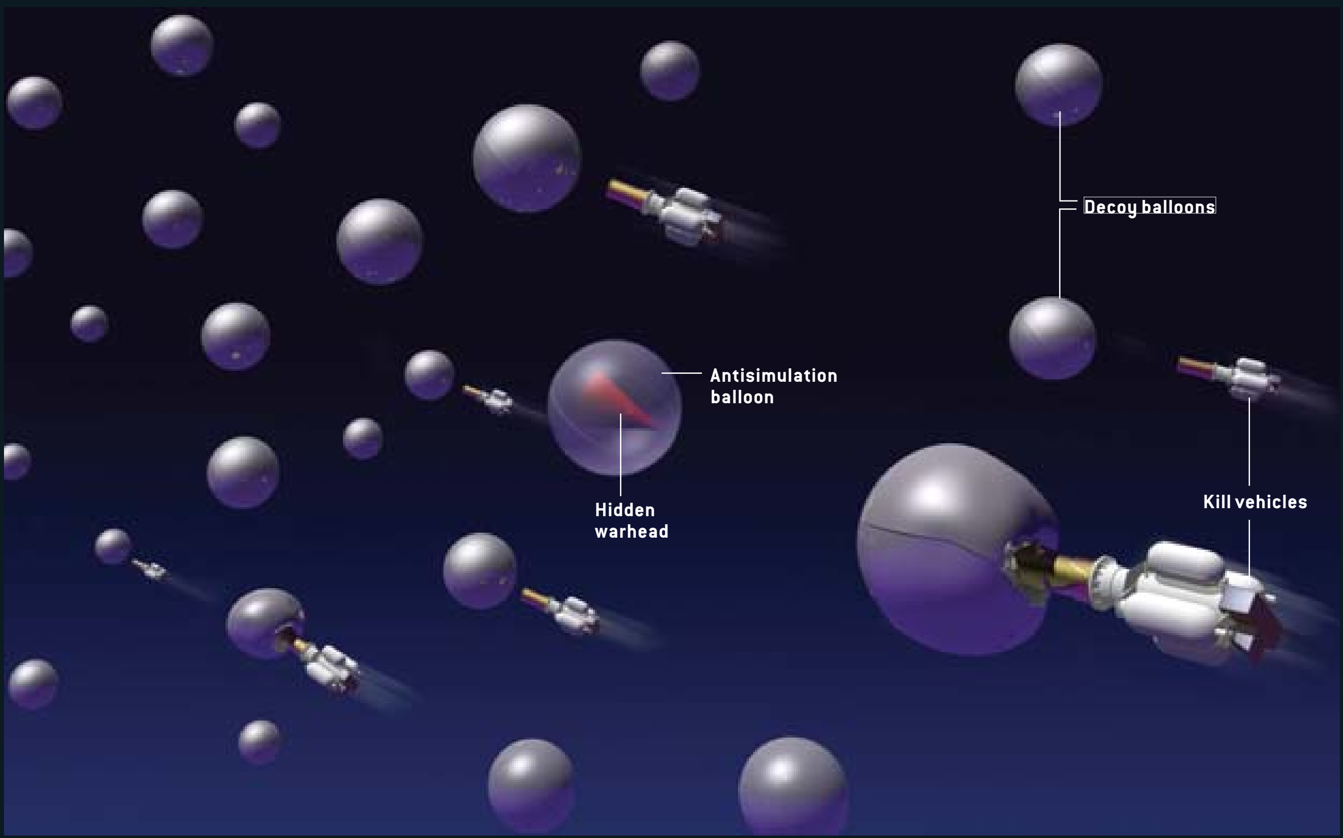
Command and control for the intercept system will be based at the U.S. Air Force's Cheyenne Mountain Operations Center near Colorado Springs, Colo. During tests of the missile defense system, officials monitored the operations from a control room at Kwajalein Atoll (*below*).

INTERCEPTOR ROCKET

The first of the interceptor boosters, a 17-meter-long three-stage rocket, was installed in July in its silo at Fort Greely, Alaska. A total of six interceptors will be deployed at Fort Greely by this fall and 10 more by the end of next year. Four interceptors will be based at Vandenberg Air Force Base in California.

The Problem of Countermeasures

The Achilles' heel of the national missile defense now being deployed is the availability of simple countermeasures. Even a relatively unsophisticated missile power such as North Korea or Iran could equip its ICBMs with balloon decoys designed to mimic the missile's warhead, which itself could be enclosed in an antisimulation balloon to make the decoys more effective. Each decoy balloon would contain a small heater to give it the same infrared signature as the balloon containing the warhead. The defense system would be forced to target every decoy and exhaust its supply of interceptors.



detonations in space or the atmosphere that would disrupt communications and pose environmental hazards. In the MDA's system, each interceptor carries a payload called a kill vehicle, which uses infrared sensors to home in on the ICBM warhead. Once the kill vehicle comes close enough that the warhead is in the sensor's field of view, the vehicle can use small side thrusters to put it on a collision course. The Pentagon has demonstrated this so-called hit-to-kill capability in several tests since 1999, although the MDA and its critics agree that the Defense Department did not realistically simulate a ballistic-missile attack.

In this approach, the closing velocity at intercept is so high that the entire kill vehicle becomes an effective projectile. Even a stationary kill vehicle colliding with an ICBM warhead moving at seven kilometers per second delivers a tremendous wallop of kinetic energy—nearly 25 million joules for every kilogram of the vehicle's mass. In contrast, the energy density of high explosive is only about four million joules per kilogram. Adding explosive to the kill vehicle is obviously not necessary; it would be better to add mass in the form of improved guidance systems that would increase the probability of a hit on the target.

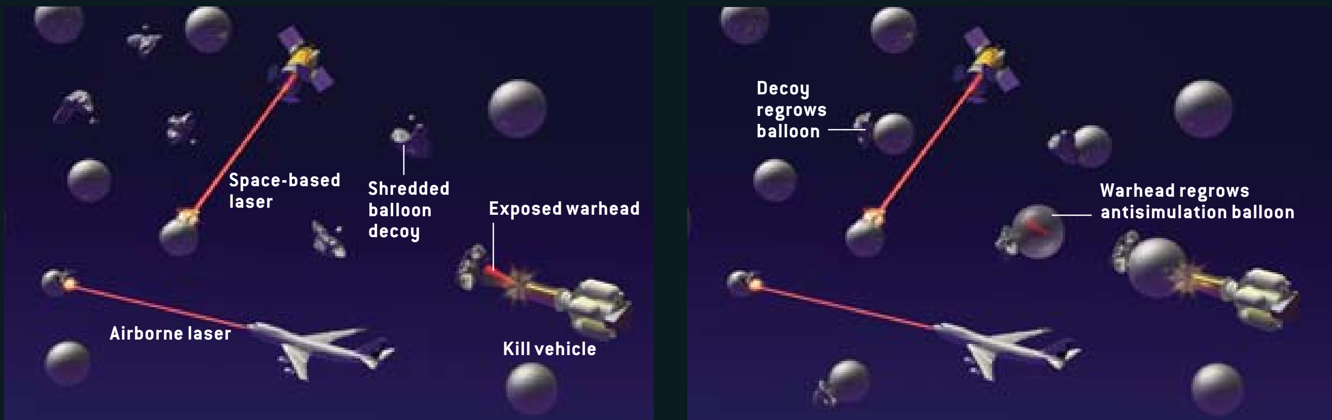
Countermeasures

NATIONAL MISSILE DEFENSE must contend with an adversary who has high stakes at risk and does not want our activity to succeed. Thus, it would be false security to field a system that would work only against an opponent who did not make use of readily available countermeasures. One obvious countermeasure would be to reduce the radar and infrared signatures of the ballistic missile and its warhead to make it harder for the interceptors to home in. For example, putting the warhead in a reentry vehicle shaped like a sharply pointed cone and coated with radar-absorbing material could significantly shrink the object's appearance on X-band radar. Also, an attacker could cool the black shroud of the warhead using liquid nitrogen, making it invisible to the kill vehicle's infrared sensor.

Another countermeasure would be to load each ICBM with dozens of decoys designed to look just like the warhead. If the ICBM releases the decoys and warhead at the end of powered flight, the paths of the lightweight decoys would be indistinguishable from that of the heavier warhead when they are traveling through

ALFRED T. KAMAJIAN; MISSILE DEFENSE AGENCY (*interceptor rocket*) (*preceding pages*); ALFRED T. KAMAJIAN (*above*)

Could the defense be modified to distinguish between the warhead and the decoys? Ultimately, the U.S. could deploy an early-arriving interceptor that would detonate, creating a wave of gas that would move the light decoys but not the balloon containing the heavy warhead; the X-band radar would detect the motion and direct another interceptor to the warhead (*left*). But the balloons could be equipped with small explosives that would fire in the opposite direction, countering the thrust before any motion could be detected (*right*).



Powerful lasers based on aircraft or in space could also pop the individual balloons by heating them (*left*). But the attacker could frustrate this defense by designing the decoys and warhead to inflate new balloons if the original ones collapse (*right*).

the vacuum of space. The attacker could also put heaters in the decoys to give them the same infrared signature as the warhead. To ease the task of decoy building, the attacker could create an antisimulation warhead—a weapon dressed to look like a decoy. For instance, the warhead could be enclosed in a radar-reflecting aluminized balloon that appears identical to dozens of empty decoy balloons. If the kill vehicles cannot distinguish between the warheads and the decoys, hundreds of interceptors would have to be launched and the missile defense would quickly be overwhelmed [*see box on these two pages*].

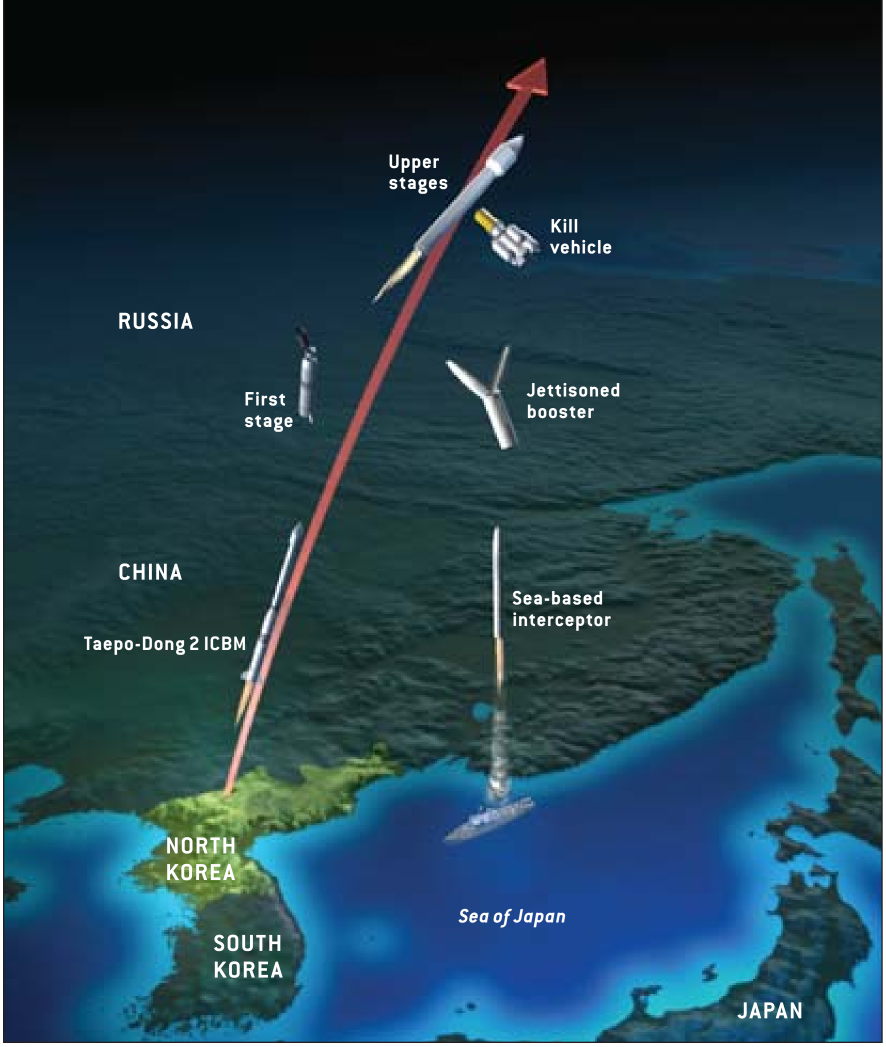
The fundamental weakness of midcourse intercept is that the countermeasures are all too simple. The money and skill needed to implement them are trivial compared with the effort required to design, build and care for the ICBMs themselves. Unfortunately, the MDA makes the artful assumption that North Korea (which has not yet tested an ICBM capable of carrying a nuclear warhead, although the CIA has expected such a test since before 1998) will not field any countermeasures that could defeat the U.S. interceptors. I am so persuaded of the effectiveness of these countermeasures—specifically, decoys and antisimulation balloons—that beginning in 1999 I strongly urged the Ballistic Missile Defense Organization (the predecessor of the MDA) to abandon the midcourse defense and assign higher priority to boost-phase intercept instead.

The only sure way of defeating these countermeasures is to intercept the missile earlier in its flight, because credible decoys cannot be released from the ICBM while the rocket is still firing—they would soon be left behind. An attacker could still attempt to fool the interceptors by launching dummy missiles designed to look like a warhead-carrying ICBM, but because each dummy rocket would have to include at least two stages to be a credible replica, this effort would be considerably more expensive than releasing a few dozen balloons. Another advantage of boost-phase intercept is its ability to stop the delivery of biological weapons, which would most likely be packaged in hundreds of small bomblets that would be released from the ICBM just after its ascent. Because the bomblets would streak separately toward the U.S., no midcourse or terminal defense system could halt such an attack.

Proponents of the MDA's system have argued that they intend to eventually in-

Boost-Phase Intercept

To destroy an ICBM before the warhead and decoys can be released—that is, while the missile is ascending—the kill vehicle must intercept the rocket within about four minutes of launch. The interceptor must be based within 1,000 kilometers of the initial trajectory of the ICBM. In the case of North Korea, the interceptors could be based on ships in the Sea of Japan. To shoot down Iranian ICBMs, the interceptors could be launched from the Caspian Sea or the Persian Gulf.



ALFRED T. KAMAJIAN

corporate boost-phase intercept into the national missile defense. But the creation of a layered defense, which attempts to intercept the missile at multiple stages of its trajectory, is not necessarily a cost-effective strategy. Each of the defensive layers has a price, and investing in boost-phase intercept will make the U.S. much safer than using the same funds to build or expand the flawed midcourse system. Unfortunately, the technology development for boost-phase intercept is still in the preliminary stages. My 1999 discussions with missile defense officials were not continued, and the MDA delayed several years before initiating a formal boost-phase program.

In 2003 the American Physical Society (APS) released an analysis of the potential of boost-phase intercept. The study was written by a highly qualified panel of physicists and engineers, many with years of experience in missile defense. Although news accounts claimed the study was negative on boost-phase defense, a careful reading shows that the analysis is in reasonable agreement with my qualitative estimates of 1999. I judged that 14-ton interceptors would be required, with burnout speeds ranging from eight to 11 kilometers per second. The guidance system would home in on the ICBM's flame and then the missile's body, and the interceptor would smash into the missile hard enough to disable its rocket engines. The APS study group analyzed in considerable detail the evasive maneuvers available to the ICBM and the capabilities required of the interceptor to cope with them.

The APS group showed how difficult it would be to plan the intercept of an ICBM from North Korea in such a way as to guarantee that the warhead—which might still be functional after the impact—does not strike somewhere in the U.S. or in some other country. (If the ICBM is hit near the end of the boost phase, the warhead might still have enough momentum to reach North America.) In my own writings, I have emphasized that the intercept should still be considered successful if the warhead were to reenter the atmosphere over some random point in the U.S. instead of its target city. Because the average population density of the U.S. is only about 1 percent of the peak density in cities, the intercept would effectively devalue the enemy's missile force by a factor of 100, which should be enough to deter the launch.

Space Wars

AS THE MDA now decides which boost-phase options to pursue, some officials are dusting off the old plans for space-based interceptors that were part of the original Strategic Defense Initiative. The Brilliant Pebbles concept, for example, envisioned a constellation of interceptors in low earth orbit, each equipped with enough fuel to propel itself toward an ascending ICBM and to counter any evasive maneuvers that the missile might attempt. Representative Curt Weldon of Pennsylvania, a strong advocate of national missile defense, recently cautioned supporters of space-based interceptors not to oppose the deployment of land- and sea-based interceptors, because this internecine struggle is likely to delay both programs. But many missile defense supporters make no pretense about their utmost goal—to deploy a system that could counter Chinese and ultimately Russian

ICBMs—and only space-based interceptors could achieve that aim.

A space-based system, however, would be extremely costly and vulnerable. If the interceptors were placed in low earth orbit, they would circle the globe every 90 minutes or so; the U.S. would need to deploy more than 1,000 of them to ensure that a sufficient number would be near North Korea when even a single ICBM is launched. Although the space-based interceptors would not have to be as large as those launched from land or sea, they would be useless in orbit without powerful rocket engines to reach the ascending ICBMs in time. The APS group estimated that the interceptors would have to weigh between 600 and 1,000 kilograms. And because it costs $20,000 to send just one kilogram into orbit, the price tag of the space-based intercept system could easily run into the tens of billions of dollars.

Furthermore, the fact that the space-based system could also threaten Chinese and Russian ICBMs may compel those governments to take preemptive steps. China may appear particularly vulnerable because it has only about two dozen nuclear-armed ICBMs capable of reaching North America. If the U.S. puts thousands of boost-phase interceptors into orbit, China would no doubt build more long-range missiles, because the space-based system can be defeated by launching many ICBMs at once from a small region. China would also have every incentive to destroy the orbiting interceptors. Unlike a preemptive strike on land- or sea-based systems, an attack against a space weapon would not result in human casualties and might not be considered an act of war by the international community.

One way to destroy a satellite in low earth orbit is to launch a pellet cloud to orbital altitude to shred the interceptor as it passes through. The Chinese could also use ground-based rockets to shoot down the interceptors one by one. Or they could put microsatellite space mines into orbit, each within a few tens of meters of an interceptor and ready to detonate at a moment's notice. The same countermeasures would be even more cost-effective against another proposal for boost-phase defense, the space-based laser, which would be larger and more vulnerable than the interceptors.

The Weakest Link

EVEN THE LEADERS of the MDA claim very little for the national missile defense that is now operational. Testifying before the House Armed Services Committee last March, Lieutenant General Ronald Kadish, longtime head of the MDA and its predecessor agency, noted that "what we do in 2004 and 2005 is only the starting point—the beginning—and it involves very basic capability." My assessment, however, is that the present missile defense approach is utterly useless against ICBMs of new or existing nuclear powers because midcourse countermeasures are so effective.

What is more, the primary missile threat to the U.S. is not ICBMs. If a nation such as North Korea or Iran is intent on attacking an American city, it is far more likely to do so using short-range missiles launched from ships near the U.S. coasts. In a press briefing in 2002 Secretary of Defense Donald H. Rumsfeld noted: "Countries have placed ballistic missiles in ships—dime a dozen—all over the world. At any given time, there's any number off our coasts, coming, going. On transporter-erector-launchers, they simply erect it, fire off a ballistic missile, put it down, cover it up. Their radar signature's not any different than 50 others in close proximity." Despite this acknowledgment, however, the Defense Department has no system planned for deployment that could defend against these missiles.

The irrelevance of the operational missile defense has become apparent even to longtime supporters of the goal. Conservative columnist George Will recently wrote that "a nuclear weapon is much less likely to come to America on a rogue nation's ICBM—which would have a return address—than in a shipping container, truck, suitcase, backpack or other ubiquitous thing." But even in the unlikely case of an undeterrable ICBM launch from an irresponsible power, the midcourse system is not the best defense. The MDA's efforts should be shifted to boost-phase intercept, and if the goal is to stop ICBMs from North Korea and Iran, then land- or sea-based interceptors show the most promise. In all these cases, the vulnerability of the defending system must be taken into account, which effectively rules out the use of space-based weapons. In missile defense, as in so many other fields, the system is only as strong as its weakest link. SA

The primary missile threat is not ICBMs, but short-range missiles launched from ships.

MORE TO EXPLORE

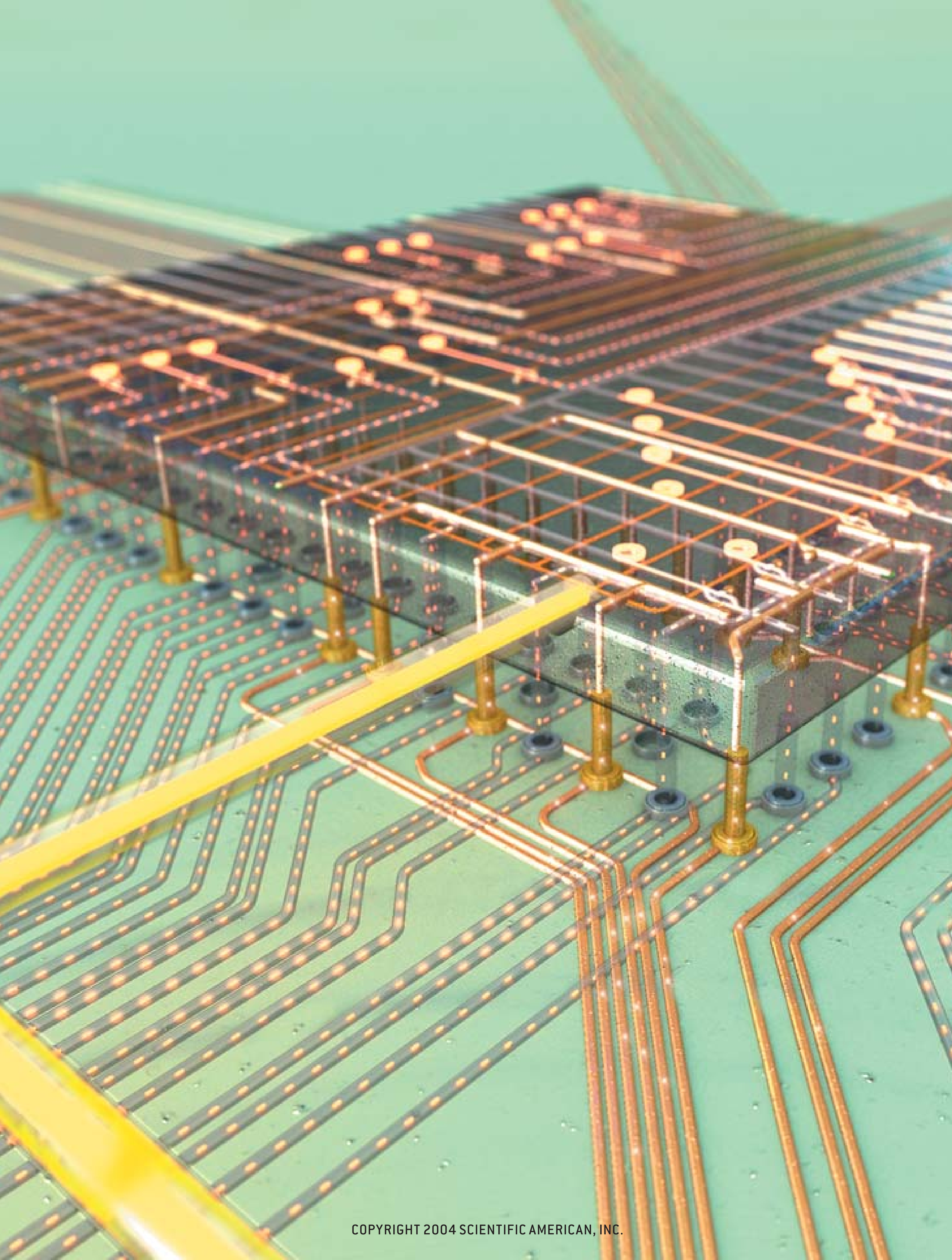
Anti-Ballistic-Missile Systems. Richard L. Garwin and Hans A. Bethe in *Scientific American,* Vol. 218, No. 3, pages 21–31; March 1968.

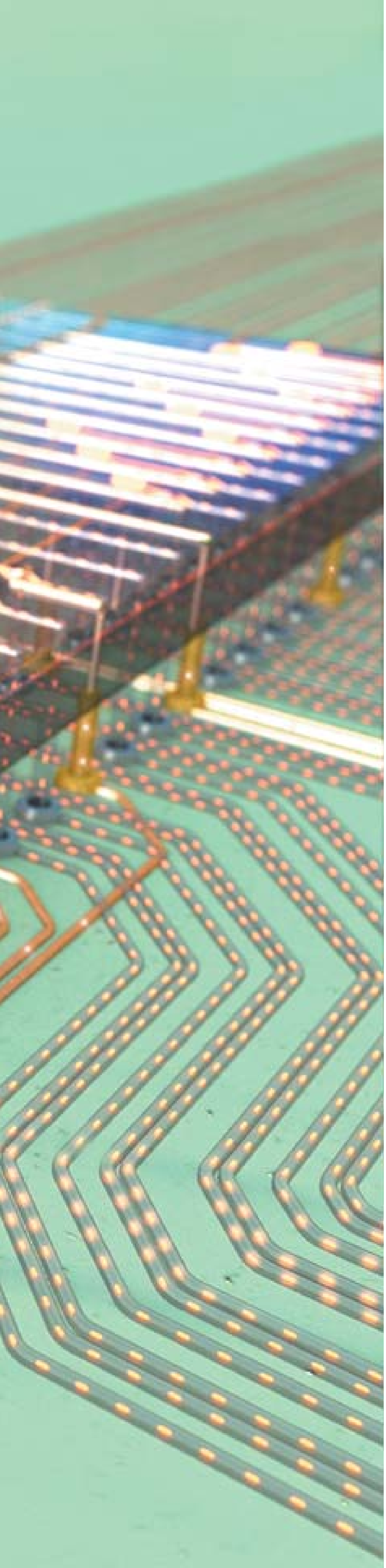
Cooperative Ballistic Missile Defense. Richard L. Garwin. Presented at the Secretary of State's Open Forum, November 18, 1999. Available at **www.fas.org/rlg/991117.htm**

Countermeasures: A Technical Evaluation of the Operational Effectiveness of the Planned U.S. National Missile Defense System. A. M. Sessler et al. Union of Concerned Scientists and MIT, April 2000. Available at **www.ucsusa.org/publications/report.cfm?publicationID=132**

Report of the APS Study Group on Boost-Phase Intercept Systems for National Missile Defense. D. Kleppner et al. in *Reviews of Modern Physics,* Vol. 76, No. 3, pages S1–S424; July 2004. Available at **www.aps.org/public_affairs/popa/reports/nmd03.cfm**

Technical Realities: An Analysis of the 2004 Deployment of a U.S. National Missile Defense System. L. Gronlund, D. C. Wright, G. N. Lewis and P. E. Coyle III. Union of Concerned Scientists, May 2004. Available at **www.ucsusa.org/global_security/missile_defense/page.cfm?pageID=1403**





Computing at the Speed of Light

Emerging ways to make photonic connections to electronic microchips may dramatically change the shape of computers in the decade ahead

By W. Wayt Gibbs

FUTURISTIC MICROPROCESSOR might communicate with the rest of the computer via light as well as electricity. Recently invented devices such as microcavity lasers, silicon optical modulators and translucent polymer pillars could be combined to move bits seamlessly from the electronic realm to the photonic, and back again.

BRYAN CHRISTIE DESIGN

Since about 1995, microprocessors have been outrunning the other parts of computer systems by ever increasing margins. The latest processors churn through instructions at up to 3.6 gigahertz (GHz); some operations, such as arithmetic, run at double that rate. But the wiring on the motherboard that connects the processor to its memory chips and other pieces of the system plods along at 1 GHz or less. So the brain of the machine spends as much as 75 percent of its time idle, waiting for instructions and data that are stuck in traffic.

"In the coming years, the imbalance between microprocessor performance and memory access will be driven to a crisis point," physicist Anthony F. J. Levi of the University of Southern California argued in a detailed analysis three years ago. He noted that the plastic material in printed circuit boards squelches high frequencies: for every 2-GHz increase in electrical signal bandwidth, signal strength falls 10-fold. As clock rates rise, so do power consumption, heat production and electromagnetic interference. Those are already three of the biggest headaches for system designers. And International Sematech, an industry consortium, forecasts that processor-to-peripheral links must accelerate by roughly 2 GHz every two years just to keep the bottleneck from tightening further.

"Our engineers think they will eventually be able to squeeze 20 GHz out of wires as long as 20 inches," says Michael Morse, a photonics researcher at Intel. According to the Sematech roadmap, 20 GHz would be just sufficient for the 32-nanometer generation of microchips, three steps down the road from the 90-nanometer chips that arrived earlier this year. Mark T. Bohr, director of process architecture at Intel, reports that his company is on track to bring that generation to market by 2010.

The stage thus appears set for photonic connections, which exchange data via laser light, to take over for copper wiring in the next decade. "I'm a big fan of optical connections at the system level," says Patrick P. Gelsinger, Intel's chief technical officer, although he remains unconvinced that they will take over the very high speed but short-distance hop from the processor to the memory bank. Exactly when the transition will occur, in which connections, and at what price, depends in large measure on how the photonic devices are made.

Data already often move between electronic and photonic forms at the periphery of a computer system, on their way to or from a CD or DVD, display monitor, mouse, camera, stereo amplifier or fiber-optic network. But the core of most computers—the processor, the main memory, and the motherboard that connects those to the various peripheral devices—remains an all-electron show.

The reason for this is simple: optical interconnections, though often many times faster than copper wires and traces, tend to be 10 to 100 times more expensive. For some applications, such as switching thousands of telephone calls or shuttling billions of Internet packets, capacity trumps cost. That is why long-distance communications in rich countries now travel primarily over optical fiber. And it is why Cisco spent half a billion dollars over the past four years to create an optical router, debuted this past May, whose 30 fiber-optic lines run at 40 gigabits per second (Gbps)—in principle, enough aggregate bandwidth to handle the Internet traffic of 1.6 million DSL-equipped households. For distances greater than 100 meters, nothing beats the switching speed of light. But over short links, such as those in office networks and inside computer cases, copper still reigns king.

A change of regime now seems more likely, however, because scientists have at last succeeded in making a wide range of photonic devices that could be manufactured by existing microchip factories and thus be cost-competitive. "We want to drive optics all the way down to chip-to-chip communications," says Mario Paniccia, head of Intel's silicon photonics research group.

If that happens, computers may look and operate very differently a decade from now. Some changes will be of the "faster, smaller" variety. Video cameras and portable video players might plug their fiber-optic cables into the photonic successors to USB ports. Some machines may have holographic disk drives that can archive hundreds of gigabytes on one removable CD-size platter. For those people fortunate enough to have a direct connection to the international fiber-optic telecommunications grid, an optical network card may provide Internet access at more than a gigabit per second (Gbps)—about 1,000 times the speed of today's DSL and cable modems.

Other changes could be more dramatic. The maximum practical speed of electronic connections falls off quickly as cable lengths increase. So memory chips and graphics cards have to be close to the processor that shovels information to them. "But once you have data in the optical realm, distance doesn't matter," Paniccia observes. "A low-cost photonics technology is low cost both at one foot and at 1,000 miles." Many of the components of a computer that are now crammed into a two-foot-high rectangular box could in principle be spread across a car, throughout a building or all over a city, with data flowing among them on pulses of light.

Overview/*Optical Computing*

- Computer engineers expect that within the next decade, the copper wiring that now connects components inside machines will reach the practical limits of its bandwidth.
- Until recently, connecting microchips with light meant using lasers and detectors made from exotic semiconductors. Such devices are affordable only for niche applications, such as high-speed telecommunications hubs. But this year engineers unveiled new classes of photonic devices that could be made in the same factories used to make low-cost microchips.
- Researchers have also begun demonstrating novel schemes for guiding laser pulses to and from microprocessors and circuit boards.
- Because optical connections can run at very high bandwidth over both long and short distances, the addition of photonics could fundamentally change the shape of computers over the long term.

Opening the Bottleneck

CURRENT OPTICAL CHIPS, of the kind used as lasers in CD players and as photodetectors in telecommunications switches, are manufactured from III-V semiconductors. These compounds pair one or more elements from the third column of the periodic table (such as aluminum, gallium or indium) with an element from the fifth column (typically phosphate, arsenic or antimony).

At first glance, III-V chips might seem ideal for photonics. Electrons move faster in them than in silicon, so III-V processors can operate at much higher frequencies. And they not only emit laser light from cavities in their surface but also con-

THE PHOTONIC PC: WHAT'S HERE AND WHAT'S AHEAD

Computers today already use optical devices at a few spots around the periphery of the machine. But photonic components now in labs or near release will probably make their way to the core of machines within the next decade.

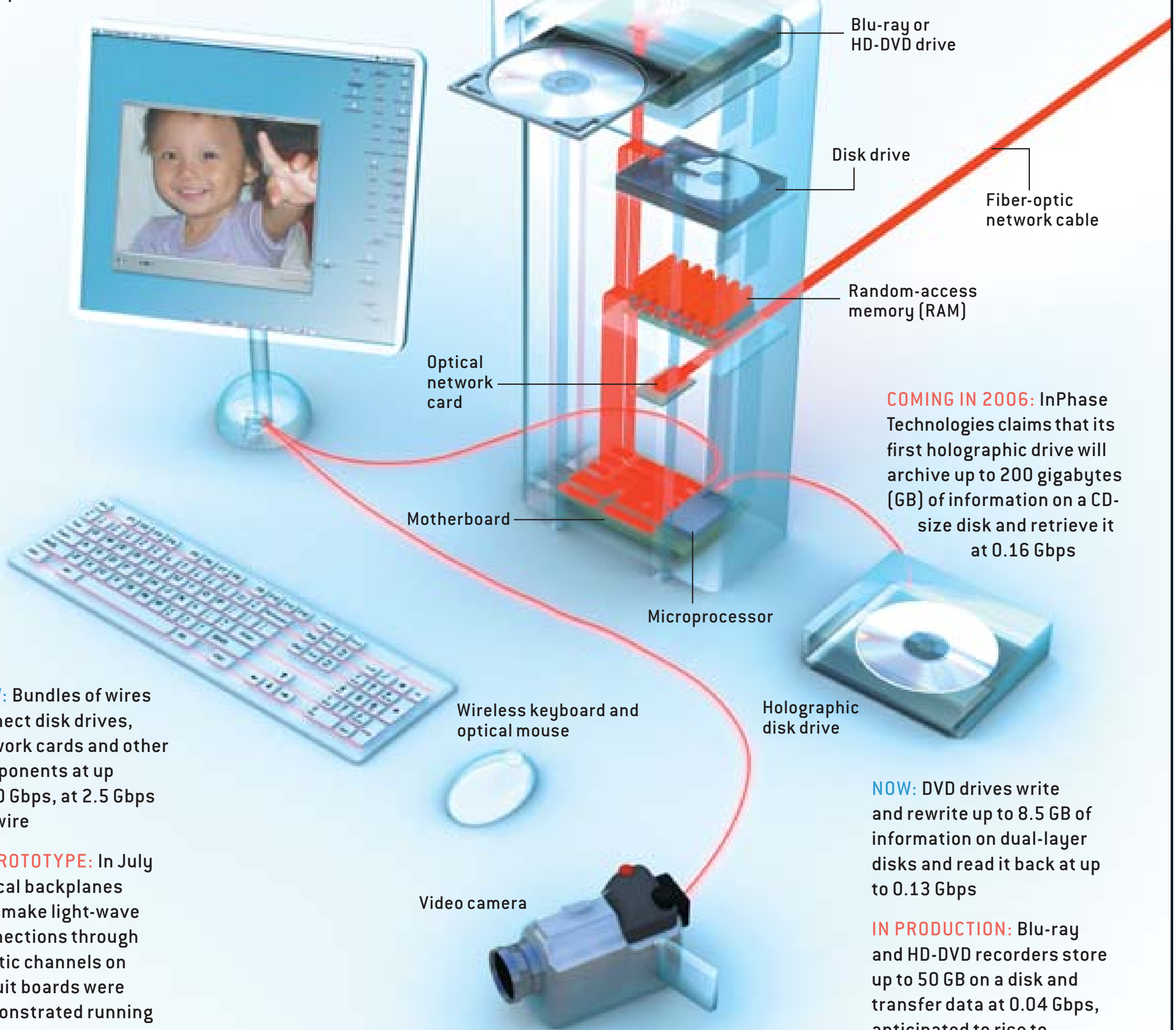
NOW: Microprocessors operate at three gigahertz (GHz) or more but often idle while waiting for data to arrive from RAM running at just 0.4 GHz. Newer processors can retrieve data from memory over multiple wires, at aggregate rates up to 51 gigabits per second (Gbps)

IN PROTOTYPE: Photonic connections between memory and processor operating at 1.25 GHz per waveguide

IN LABS: New methods for fabricating thousands of polymer pillars onto processors, allowing both high-speed electrical and optical connections to a motherboard

NOW: Fiber-optic networks run at speeds up to 10 Gbps, but most machines use lower-cost Ethernet connections of just 0.1 Gbps

IN PROTOTYPE: Silicon-optical modulators built at Intel could lead to affordable optical networks that run at 2.5 to 10 Gbps



COMING IN 2006: InPhase Technologies claims that its first holographic drive will archive up to 200 gigabytes (GB) of information on a CD-size disk and retrieve it at 0.16 Gbps

NOW: Bundles of wires connect disk drives, network cards and other components at up to 40 Gbps, at 2.5 Gbps per wire

IN PROTOTYPE: In July optical backplanes that make light-wave connections through plastic channels on circuit boards were demonstrated running at 8 Gbps per waveguide

NOW: DVD drives write and rewrite up to 8.5 GB of information on dual-layer disks and read it back at up to 0.13 Gbps

IN PRODUCTION: Blu-ray and HD-DVD recorders store up to 50 GB on a disk and transfer data at 0.04 Gbps, anticipated to rise to 0.32 Gbps

NOW: USB 2.0 connects a mouse, video camera and other gadgets to the computer at about 0.48 Gbps, over distances as long as five meters

IN PRODUCTION: High-priced fiber-optic connectors manufactured by Xanoptix transfer data at up to 245 Gbps, over distances up to 500 meters

BRYAN CHRISTIE DESIGN

vert incoming flashes to electrical signals at blistering speeds. For that reason, photonics researchers have turned first to III-Vs to build optical integrated circuits.

Using indium phosphide, for example, a group led by Daniel Blumenthal and Larry Coldren at the University of California at Santa Barbara last year constructed a "photon copier." The device accepts photonic bits at one wavelength, regenerates them if they have faded, and uses a tunable laser to translate them to a different wavelength without ever converting the information into electronic form. Such a device would be very handy in a future photonic computer.

But in comparison with silicon, III-V semiconductors are finicky and recalcitrant materials to manufacture, and that makes them expensive. A microchip that costs $5 to make from silicon, using the standard complementary metal oxide semiconductor (CMOS) process, would cost about $500 to fabricate from indium phosphide. And with the performance of silicon continually improving, "competing against mainstream CMOS is like lying on a railroad track," bemoans Ravindra A. Athale, who manages photonics programs at the Defense Advanced Research Projects Agency. "Sooner or later the train runs you over."

If photonics is ever to find its way onto $100 motherboards, it must hop aboard that train. So in recent years, much of the research in optical computing has focused on finding CMOS-compatible ways to integrate electronic and photonic devices. This strategy has begun to pay off.

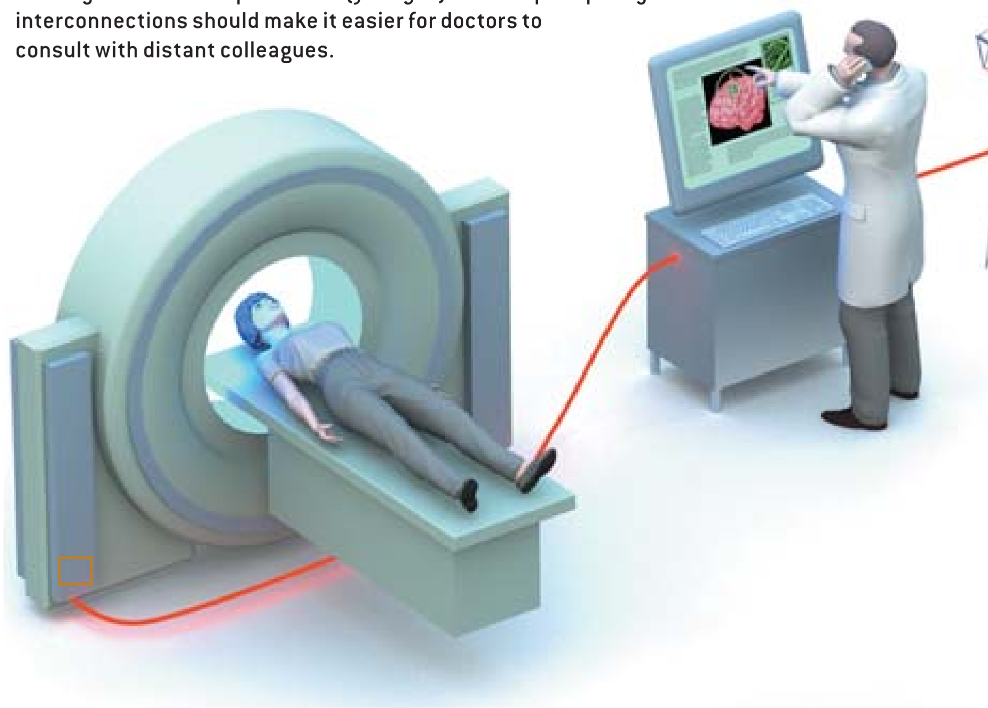
"We are at a stage in the field now that was unthinkable two years ago," says Salvatore Coffa, who directs the silicon photonics laboratory at STMicroelectronics in Catania, Sicily. "We are talking about going to market soon with the first silicon-based device with optical functions."

All Aboard the CMOS Express

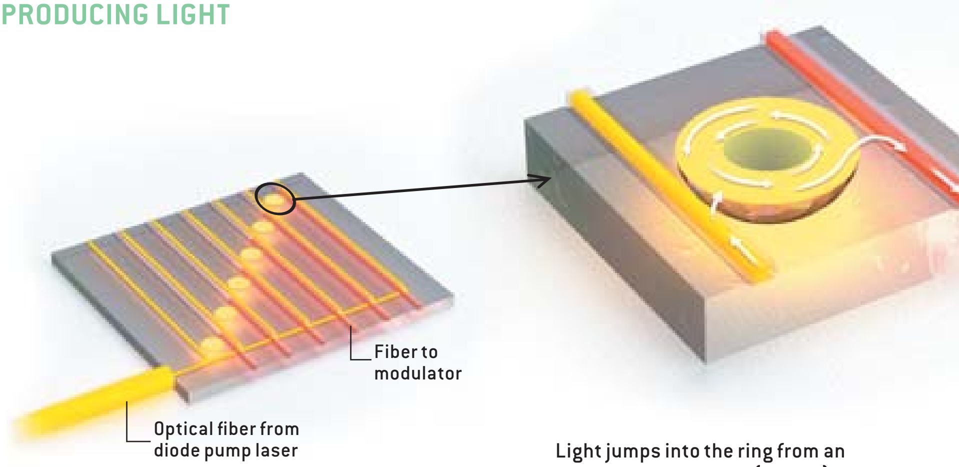
AT LEAST THREE WAYS exist to make photonic components passengers on the CMOS train, and each has been making impressive progress. The most conservative approach, called hybrid integration,

FROM WIRES TO WAVEGUIDES

Photonic microchips will likely find their first uses in special-purpose computers that must quickly process huge amounts of data, such as those used in medical imaging. An MRI scanner, for example, might one day use new kinds of microscopic lasers (*below left*) and silicon modulators (*below center*) to send its images over an optical fiber to a computer. New kinds of connectors are being developed to bring such vast amounts of data directly to the central processor (*far right*). Such super-speedy interconnections should make it easier for doctors to consult with distant colleagues.

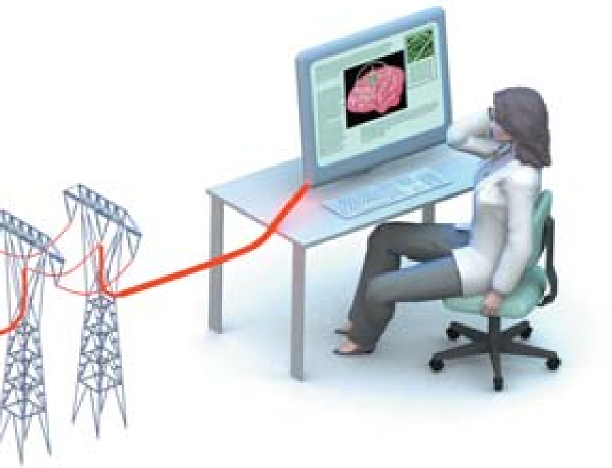


PRODUCING LIGHT

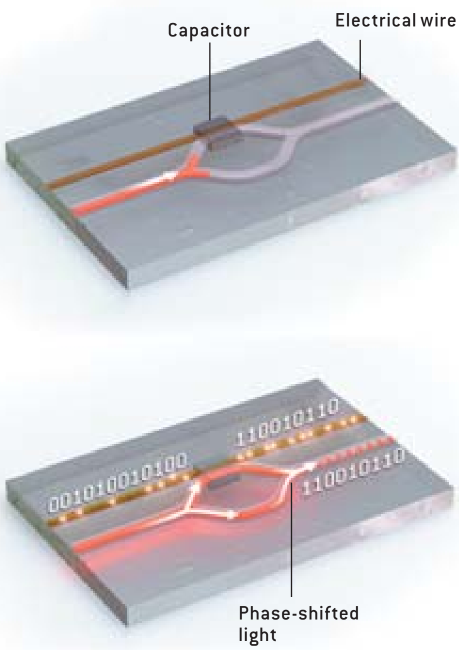


Microlasers built by Kerry J. Vahala of the California Institute of Technology can be constructed by the thousands on standard silicon microchips. The tiny rings can purify light pumped in from an inexpensive diode laser and change its color to match the standard wavelength used by other photonic components.

Light jumps into the ring from an ultrathin optical fiber (*yellow*) that passes nearby. Light of a particular frequency resonates within the ring (*arrows*) and stimulates it to emit a laser beam on a different fiber (*red*). In a working microchip, the ring lasers would probably be made on the rims of holes, with connecting fibers embedded in the chip surface.

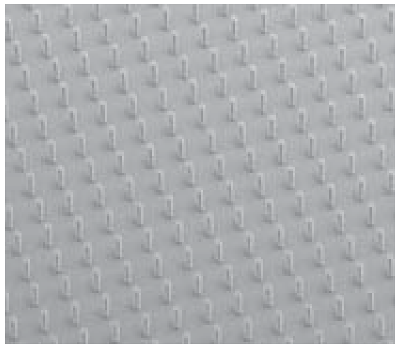


MODULATING AN OPTICAL SIGNAL

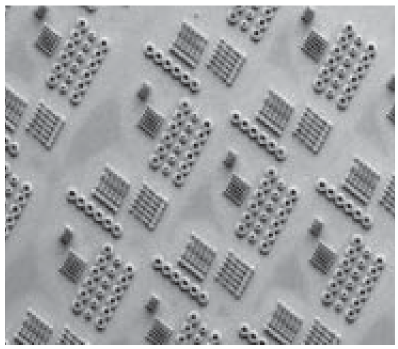


An optical modulator, such as one made last year by Mario Paniccia and his co-workers at Intel, takes digital bits in electrical form and encodes them onto a light beam. First the beam is split into two arms (*top*). Digital signals arrive at a capacitor on one or both of the arms. The signal alters the phase of light passing near each capacitor. When the arms recombine, the phase-shifted light beams interfere, creating pulses in the outgoing beam (*bottom*).

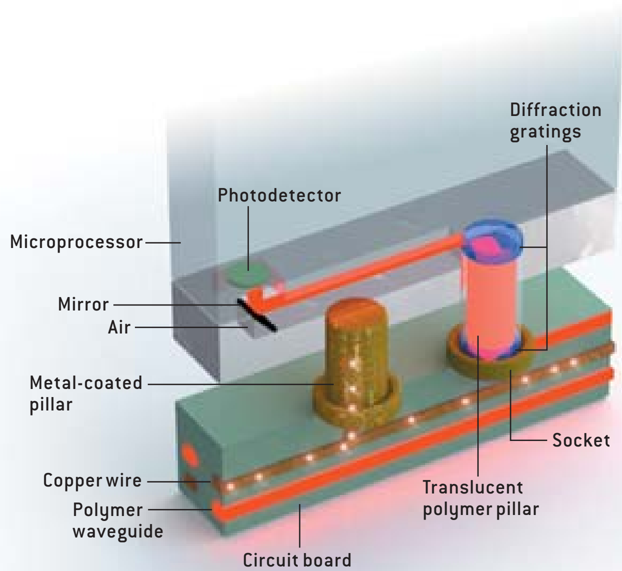
PHOTONIC MICROPROCESSING



Minuscule pillars made of translucent plastic could connect microprocessors to the circuit boards on which they rest. Thousands of such pillars, developed by James D. Meindl's group at the Georgia Institute of Technology, would be attached to the bottom of the processor.



The pillars would fit into corresponding sockets on the surface of the circuit board.



Both electronic and photonic signals could pass through pillars, carrying data to and from the microprocessor. Standard electrical connections would be made from wires to metal-coated pillars. Light pulses would flow through polymer waveguides, be turned by plastic diffraction gratings or metallic mirrors, and be detected by silicon photodiodes.

BRYAN CHRISTIE DESIGN; MUHANNAD S. BAKIR AND JAMES D. MEINDL (*polymer pillars*) is closest to commercial success, as it has already produced chips for the telecommunications industry.

Hybrid devices cram logic-bearing silicon microchips into a small package that also houses III-V chips, which perform all of the optical duties. A CMOS plant would have to be significantly modified before managers could let gallium arsenide or indium phosphide anywhere near their multibillion-dollar equipment, because those compounds can contaminate a silicon production line. But the two halves of a hybrid device can be manufactured in separate factories and then assembled later.

Xanoptix, a small start-up in Merrimack, N.H., has used this technique to mate gallium arsenide lasers to silicon control chips. The result is a thumb-size optical connector that looks similar to a USB plug. But whereas USB cables top out at less than 0.5 Gbps, Xanoptix says its photonic jack can pump up to 245 Gbps through a pencil-width bundle of 72 optical fibers.

The hybrid approach faces a long-term problem, however—the faster microprocessors become, the hotter they run. The hottest spots on some chips already rise above 80 Celsius, the temperature at which III-V lasers start to burn out. So hybrid optoelectronic chips may find a niche in slower external connectors and peripheral devices, rather than at the center of the computer.

Intel has ordered its scientists to stick to CMOS, in the hope that it might one day build entire photonic systems right into microprocessors or motherboard chips using its existing factories. To make this so-called monolithic integration work, engineers have been tricking silicon and the few other elements that are CMOS-friendly into emitting, manipulating and detecting light.

That first step is a doozy: "We think we can do everything in silicon—except the laser," Paniccia says. Silicon by itself lacks the quantum-mechanical wherewithal to make light. Coffa's group at STMicroelectronics has discovered a way around part of that problem, however. By infusing small amounts of cerium or erbium into a layer of silicon dioxide laced with silicon nanocrystals, the researchers constructed silicon chips that glow green or blue in response to a small voltage.

Because the luminescence is incoherent, these are light-emitting diodes (LEDs), not lasers. "But they are as efficient as gallium arsenide LEDs," Coffa reports. And because they are CMOS-compatible, he adds, "we can incorporate them directly into our existing electronic parts." By next year, STMicroelectronics plans to introduce silicon-based optocouplers that allow computers to control high-voltage machinery.

The silicon LEDs might also serve as a light source for a CMOS-compatible laser demonstrated earlier this year by Kerry J. Vahala and his co-workers at the California Institute of Technology. Vahala and others have been experimenting with microscopic disks of silicon dioxide perched on silicon pillars. By smoothing the edges of each disk and carefully controlling its diameter, Vahala turned it into the optical equivalent of a whispering gallery. Light passing through a nearby optical fiber leaks into the disk and circles its edge over and over, building in intensity by a factor of a million or more [*see illustration on page 84*] before it is emitted as laser light.

Incoherent sources such as LEDs could feed light to the disks, or they could be used to purify laser light coming in from outside the chip, regenerate it as it fades and tune it to a new wavelength. "Instead of etching disks, we could create holes. The laser would then form on the interior rim of the void," Vahala says. That would make it easier to connect the devices to waveguides and other photonic components on the surface. "These 'microcavity' lasers could be sources of carrier signals for information launched from the chip," he suggests.

For that to work, engineers will need a way to transfer information from electronic to optical form. Until this year, that was hard to do in silicon, which offered only a slow and unsteady lever for manipulating light. But in February, Paniccia's team unveiled a way to use silicon to modulate a laser beam—to flicker it in step with a digital signal—50 times faster.

CHANGING THE SHAPE

Over the long term, high-speed photonic connections may obviate the need to pack all the components of a computing system into a single box. "With an electrical signal, it is a totally different physical phenomenon to go four meters than it is to go four inches," notes Ravindra A. Athale of the Defense Advanced Research Projects Agency. "But once you launch into the optical domain, there is no significant difference between four inches and four meters"—or even 400 meters. So optically connected pieces could be widely scattered and still function as a seamless machine.

In a car, for example, multiple processors, memory banks and disk drives could be embedded within the body and linked by fiber optics. "I understand that Daimler Chrysler is already planning to have laser-based, gigabit optical links inside their 2005-model S-class Mercedes cars," Athale reports.

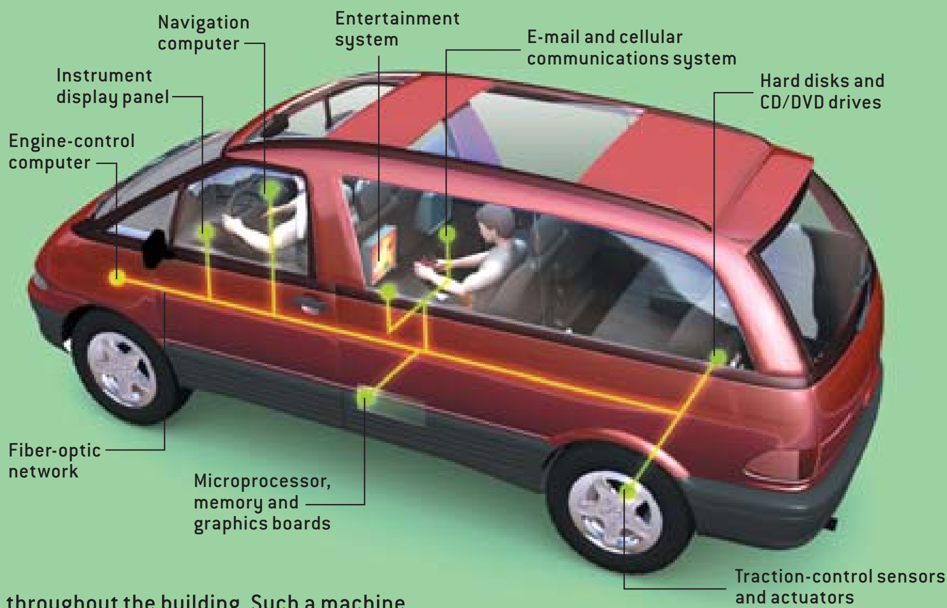
In principle, each worker in a future office could use a computer assembled by making photonic connections among parts spread

"There are about 100 modulators on this," Paniccia says, as he picks up a silicon chip the size of a postage stamp. We are in Optics Lab 1A at Intel's research center in Santa Clara. "And here's one in action," he continues, motioning to a workstation at the far end of the room. The computer is playing a high-definition DVD of *Terminator 3*. As it processes the video stream, the machine sends a copy of each bit down an Ethernet cable and into a tiny circuit board containing a single modulator.

Although it is powered by a laser, the device works on the same principle as AM radio [*see illustration on preceding page*]. It splits the microscopic laser beam into two arms. A CMOS capacitor under each arm—electrically connected to the Ethernet cable—stores and releases static electricity. "When those regions are highly charged, electrons interact with the light," Paniccia explains, shifting the relative position of the light waves. As the two halves rejoin, the peaks and troughs in their waves interfere, causing the out-

BRYAN CHRISTIE DESIGN

OF COMPUTERS



throughout the building. Such a machine could be temporarily "upgraded" at a keystroke to a faster processor or a larger memory bank, if a particular task demanded extra resources.

The ramifications of photonic links for computer architecture are still largely speculative. "System architects tend not to think along these lines because they know that such a technology does not yet exist," Athale says. But it is on its way.

put beam to pulse in the same pattern as the bits in the video stream.

Those pulses carry the data over a single fiber, thin as spider silk, that leaves the chip and connects to a photodetector attached to a second computer a few feet away. The two computers show Arnold Schwarzenegger leaping from his car in perfect synchrony.

To date, the modulator has run at rates up to 2.5 GHz. "But we can make it much smaller, and we are confident that we can scale it up to 10 GHz," Paniccia asserts. "By combining all these elements in a single chip, we'll be able to make this," he says, raising an Ethernet plug up to his eye: "little silicon-optical devices you can plug in anywhere and a $250 network interface card that replaces a $25,000 router."

"Of course, if we're going to transmit at 10 GHz, we need to be able to receive at that speed, too," Morse points out. Silicon is as transparent as glass at the infrared wavelengths typically used in photonic devices. But with the addition of germanium to the mix—which chipmakers have begun doing anyway to help speed up their processors—CMOS-compatible photodetectors have been built to convert the light pulses back into electronic bits.

The Best of Both Worlds

AS A PHOTONIC MATERIAL, silicon has come a long way in two years. But it has much farther yet to go if it is to handle optical data at more than 20 GHz. So it may be that a relatively new method of pulling photonics into the electronics, known as polylithic integration, will prove the most economical.

The general idea is to attach a CMOS processor to the motherboard with a dense array of both optical and electronic connections. Light could then be pumped into the processor from small (and thus relatively affordable) III-V chips, mounted a safe distance away so they do not overheat.

James D. Meindl and Muhannad S. Bakir of the Georgia Institute of Technology, working with Anthony V. Mulé of Intel, have demonstrated several polylithic schemes. One is called a sea of leads: thousands of microscopic S-shaped springs of metal are etched onto the processor as a final step in its manufacture. Electrical signals pass through the metal springs; light signals shoot through the holes in their centers and hit diffracting gratings that deflect the pulses into waveguides buried within the chip or motherboard.

In a second scheme, the processor rests on thousands of transparent plastic pillars, which fit into circular plastic sockets on the circuit board [*see illustration on page 85*]. Meindl's group has fabricated regular arrays of pillars that are just five microns wide and 12 microns apart. The researchers have also demonstrated how some of the cylinders and sockets could be coated with metal to make electrical, rather than optical, connections.

Bristling with tens of thousands of such minuscule pillars, a microprocessor 10 or 15 years from now may throb with infrared flashes even as it hums with high-frequency electrons. Microchip factories may etch transistors and wires in the spaces between lasers and waveguides. The long separation between laboratory photonics and consumer electronics seems to be closing at last, and our machines will be better for it. SA

W. Wayt Gibbs is senior writer.

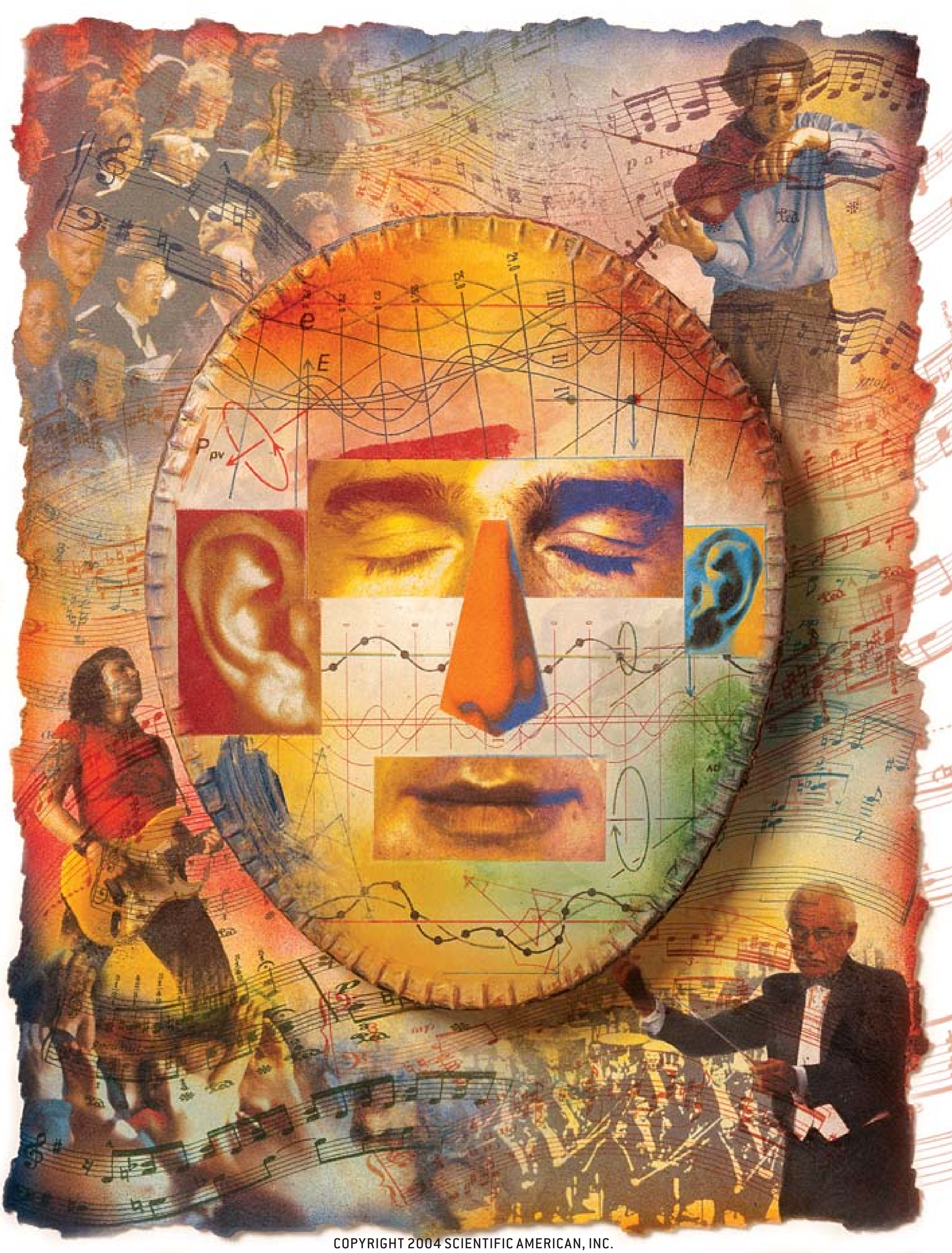
MORE TO EXPLORE

A High-Speed Silicon Optical Modulator Based on a Metal-Oxide-Semiconductor Capacitor. Mario Paniccia et al. in *Nature*, Vol. 427, pages 615–618; February 12, 2004.

Ultralow-Threshold Microcavity Raman Laser on a Microelectronic Chip. Kerry J. Vahala et al. in *Optics Letters*, Vol. 29, No. 11, pages 1224–1226; June 1, 2004.

Sea of Polymer Pillars Electrical and Optical Chip I/O Interconnections for Gigascale Integration. M. S. Bakir and J. D. Meindl in *IEEE Transactions On Electron Devices*, Vol. 51, No. 7, pages 1069–1077; July 2004.

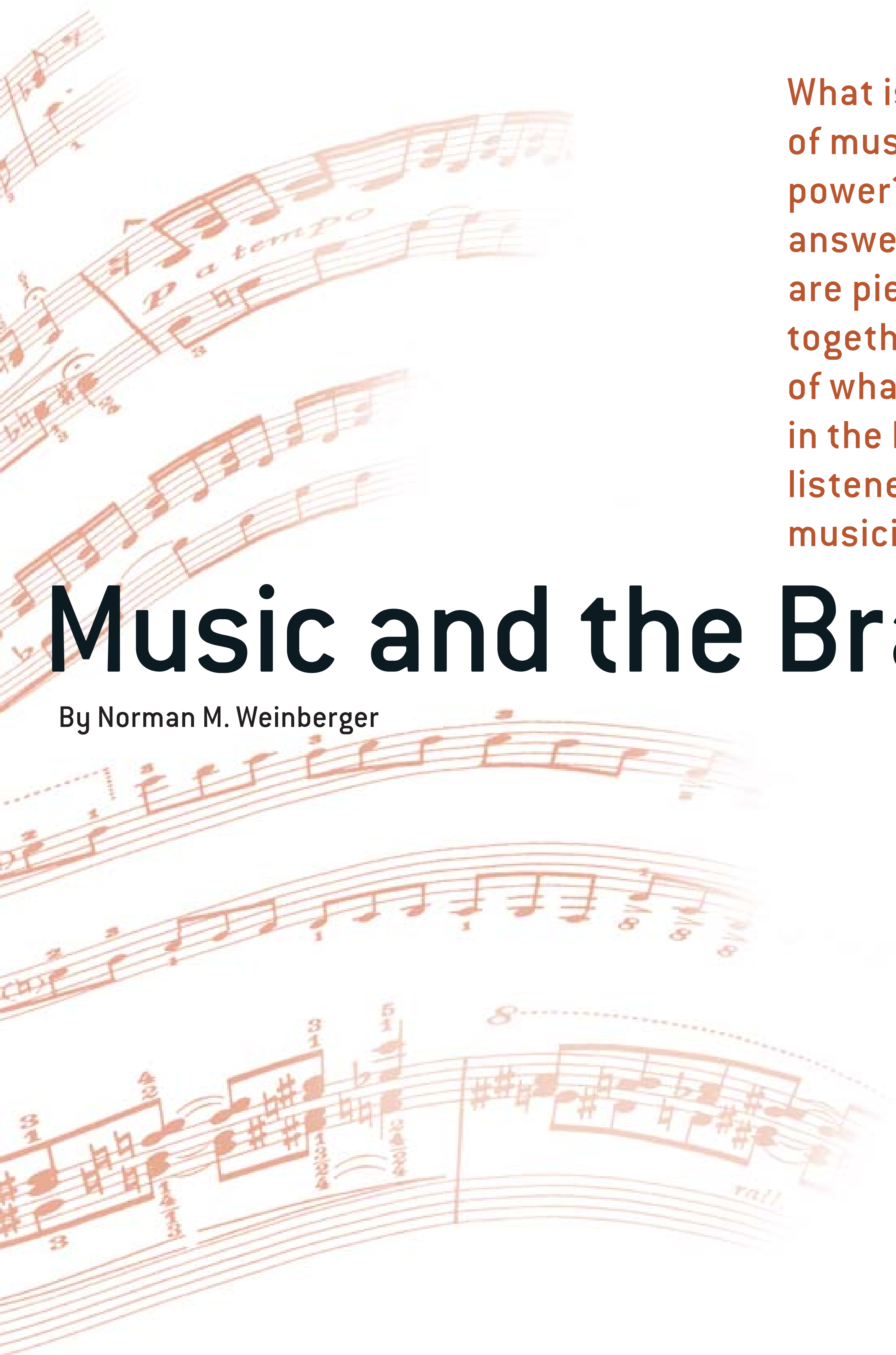
Silicon Photonics. Edited by Lorenzo Pavesi and David J. Lockwood. Springer, 2004.



What is the secret of music's strange power? Seeking an answer, scientists are piecing together a picture of what happens in the brains of listeners and musicians

Music and the Brain

By Norman M. Weinberger



JOHN STEWART

MUSIC SURROUNDS US—AND WE WOULDN'T HAVE IT ANY OTHER WAY.

An exhilarating orchestral crescendo can bring tears to our eyes and send shivers down our spines. Background swells add emotive punch to movies and TV shows. Organists at ballgames bring us together, cheering, to our feet. Parents croon soothingly to infants.

And our fondness has deep roots: we have been making music since the dawn of culture. More than 30,000 years ago early humans were already playing bone flutes, percussive instruments and jaw harps—and all known societies throughout the world have had music. Indeed, our appreciation appears to be innate. Infants as young as two months will turn toward consonant, or pleasant, sounds and away from dissonant ones [*see box on page 94*]. And when a symphony's denouement gives delicious chills, the same kinds of pleasure centers of the brain light up as they do when eating chocolate, having sex or taking cocaine.

Therein lies an intriguing biological mystery: Why is music—universally beloved and uniquely powerful in its ability to wring emotions—so pervasive and important to us? Could its emergence have enhanced human survival somehow, such as by aiding courtship, as Geoffrey F. Miller of the University of New Mexico has proposed? Or did it originally help us by promoting social cohesion in groups that had grown too large for grooming, as suggested by Robin M. Dunbar of the University of Liverpool? On the other hand, to use the words of Harvard University's Steven Pinker, is music just "auditory cheesecake"—a happy accident of evolution that happens to tickle the brain's fancy?

Neuroscientists don't yet have the ultimate answers. But in recent years we have begun to gain a firmer understanding of where and how music is processed in the brain, which should lay a foundation for answering evolutionary questions. Collectively, studies of patients with brain injuries and imaging of healthy individuals have unexpectedly uncovered no specialized brain "center" for music. Rather music engages many areas distributed throughout the brain, including those that are normally involved in other kinds of cognition. The active areas vary with the person's individual experiences and musical training. The ear has the fewest sensory cells of any sensory organ—3,500 inner hair cells occupy the ear versus 100 million photoreceptors in the eye. Yet our mental response to music is remarkably adaptable; even a little study can "retune" the way the brain handles musical inputs.

Why is music—universally beloved and uniquely powerful in its ability to wring emotions—so pervasive and important to us?

Overview/*The Musical Brain*

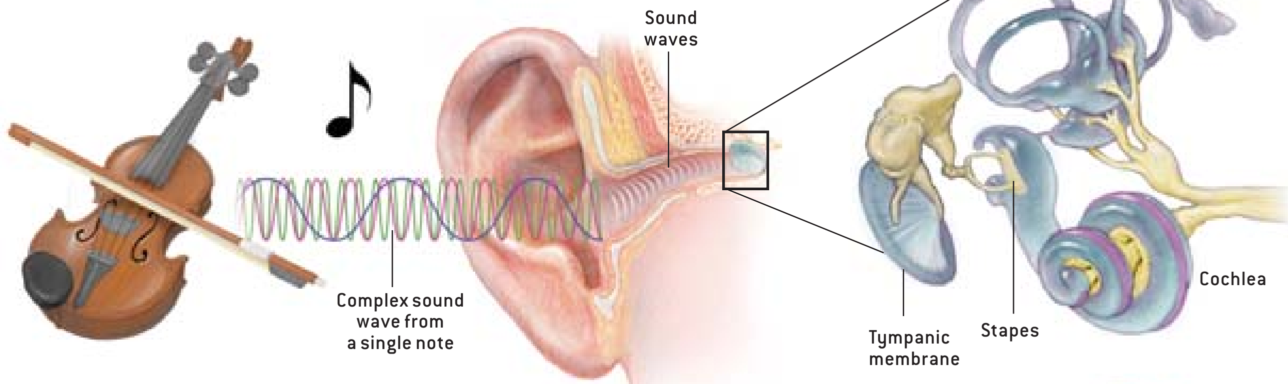
- Music has been ubiquitous in human societies throughout the world since the dawn of culture. Appreciation for music appears to be innate; infants as young as two months will turn toward pleasant sounds.
- Many different regions of the brain respond to the perceptual and emotional aspects of music, and the brain alters itself to react more strongly to musical sounds that become important to an individual.
- Scientists who study how music is processed in the brain are laying the groundwork to understand the underlying reasons for music's power and importance to humans.

Inner Songs

UNTIL THE ADVENT of modern imaging techniques, scientists gleaned insights about the brain's inner musical workings mainly by studying patients—including famous composers—who had experienced brain deficits as a result of injury, stroke or other ailments. For example, in 1933 French composer Maurice Ravel began to exhibit symptoms of what might have been focal cerebral degeneration, a disorder in which discrete areas of brain tissue atrophy. His conceptual abilities remained intact—he could still hear and remember his old compositions and play scales. But he could not write music. Speaking of his proposed opera *Jeanne d'Arc,* Ravel confided to a friend, "…this opera is here, in my head. I hear it, but I will never write it. It's over. I can no longer write my music." Ravel died four years later, following an unsuccessful neurosurgical procedure. The case lent credence to the idea that the brain might not have a specific center for music.

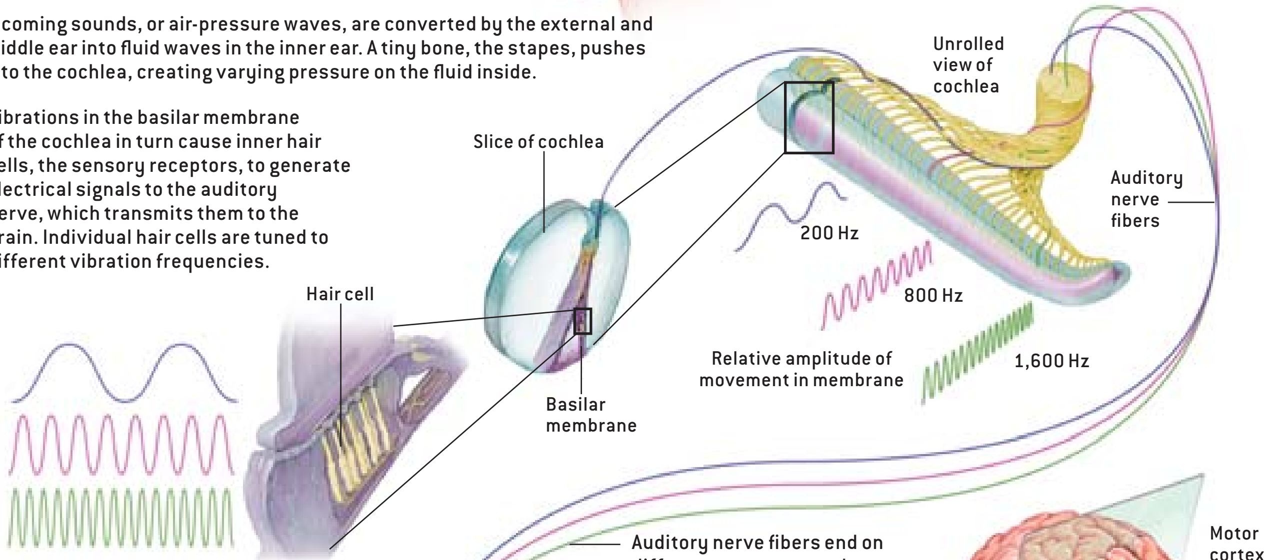
SINGING IN THE BRAIN

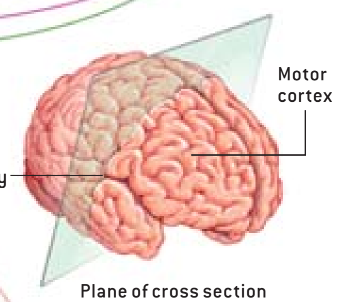
When a person listens to music, the brain's response involves a number of regions outside the auditory cortex, including areas normally involved in other kinds of thinking. A person's visual, tactile and emotional experiences all affect where the brain processes music.



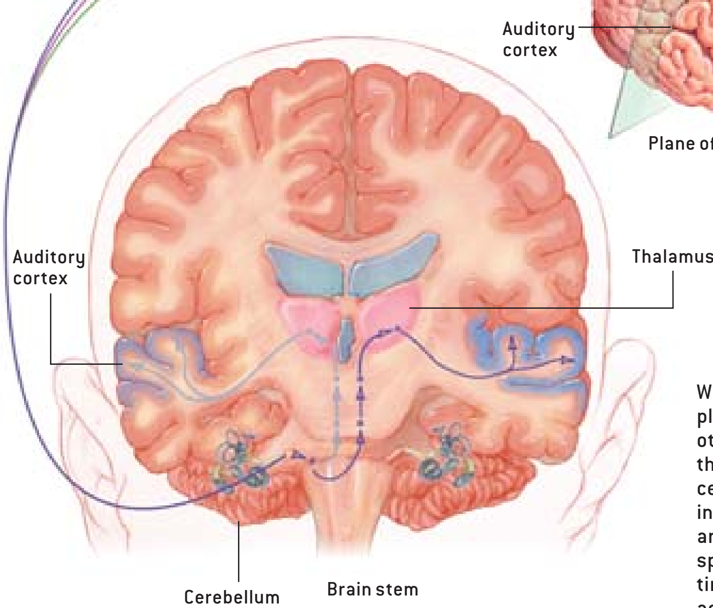
Incoming sounds, or air-pressure waves, are converted by the external and middle ear into fluid waves in the inner ear. A tiny bone, the stapes, pushes into the cochlea, creating varying pressure on the fluid inside.

Vibrations in the basilar membrane of the cochlea in turn cause inner hair cells, the sensory receptors, to generate electrical signals to the auditory nerve, which transmits them to the brain. Individual hair cells are tuned to different vibration frequencies.





The brain processes music both hierarchically and in a distributed manner. Within the overall auditory cortex, the primary auditory cortex, which receives inputs from the ear and lower auditory system via the thalamus, is involved in early stages of music perception, such as pitch (a tone's frequency) and contour (the pattern of changes in pitch), which is the basis for melody. The primary auditory cortex is "retuned" by experience so that more cells become maximally responsive to important sounds and musical tones. This learning-induced retuning affects further cortical processing in areas such as secondary auditory cortical fields and related so-called auditory association regions, which are thought to process more complex music patterns of harmony, melody and rhythm.

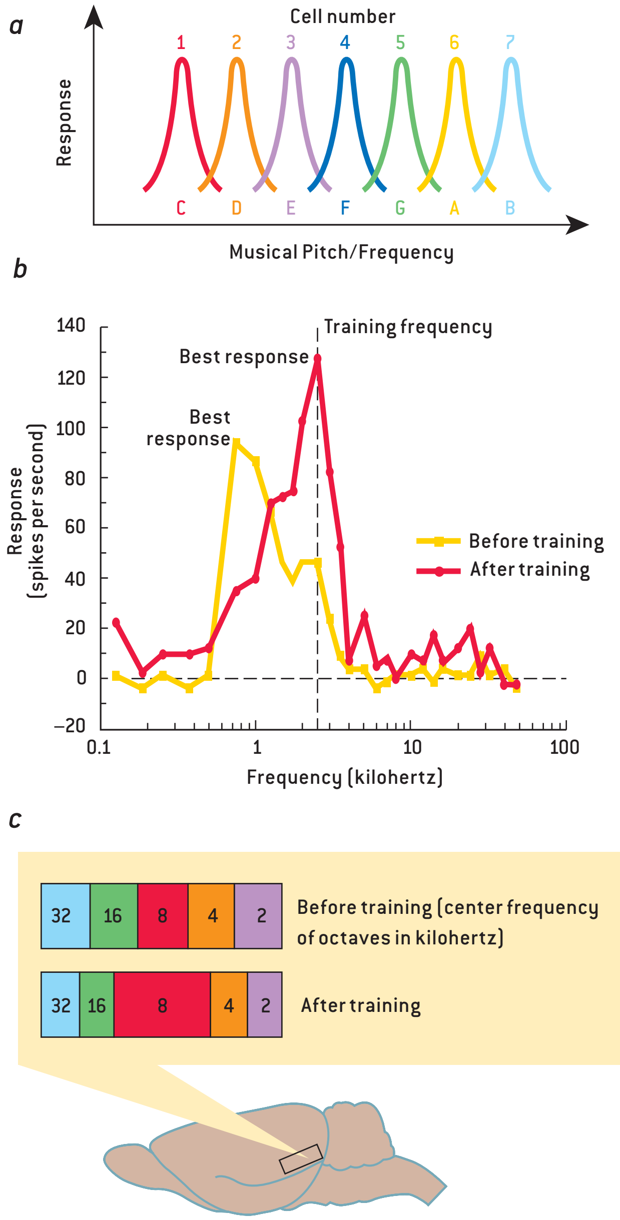


While a musician is playing an instrument, other areas, such as the motor cortex and cerebellum, which are involved in the planning and performance of specific, precisely timed movements, are active as well.

ANDREW SWIFT

RETUNING THE BRAIN

Individual brain cells each respond optimally to a particular pitch or frequency (*a*). Cells shift their original tuning when an animal learns that a specific tone is important (*b*). This cellular adjustment "edits" the "frequency map" of a rat's brain so that a greater area of the cortex processes an important tone—for instance, it expands the map for eight kilohertz when that is the important frequency (*c*).



The experience of another composer additionally suggested that music and speech were processed independently. After suffering a stroke in 1953, Vissarion Shebalin, a Russian composer, could no longer talk or understand speech, yet he retained the ability to write music until his death 10 years later. Thus, the supposition of independent processing appears to be true, although more recent work has yielded a more nuanced understanding, relating to two of the features that music and language share: both are a means of communication, and each has a syntax, a set of rules that govern the proper combination of elements (notes and words, respectively). According to Aniruddh D. Patel of the Neurosciences Institute in San Diego, imaging findings suggest that a region in the frontal lobe enables proper construction of the syntax of both music and language, whereas other parts of the brain handle related aspects of language and music processing.

Imaging studies have also given us a fairly fine-grained picture of the brain's responses to music. These results make the most sense when placed in the context of how the ear conveys sounds in general to the brain [*see box on preceding page*]. Like other sensory systems, the one for hearing is arranged hierarchically, consisting of a string of neural processing stations from the ear to the highest level, the auditory cortex. The processing of sounds, such as musical tones, begins with the inner ear (cochlea), which sorts complex sounds produced by, say, a violin, into their constituent elementary frequencies. The cochlea then transmits this information along separately tuned fibers of the auditory nerve as trains of neural discharges. Eventually these trains reach the auditory cortex in the temporal lobe. Different cells in the auditory system of the brain respond best to certain frequencies; neighboring cells have overlapping tuning curves so that there are no gaps. Indeed, because neighboring cells are tuned to similar frequencies, the auditory cortex forms a "frequency map" across its surface [*see box at left*].

The response to music per se, though, is more complicated. Music consists of a sequence of tones, and perception of it depends on grasping the relationships between sounds. Many areas of the brain are involved in processing the various components of music. Consider tone, which encompasses both the frequencies and loudness of a sound. At one time, investigators suspected that cells tuned to a specific frequency always responded the same way when that frequency was detected.

But in the late 1980s Thomas M. McKenna and I, working in my laboratory at the University of California at Irvine, raised doubts about that notion when we studied contour, which is the pattern of rising and falling pitches that is the basis for all melodies. We constructed melodies consisting of different contours using the same five tones and then recorded the responses of single neurons in the auditory cortices of cats. We found that cell responses (the number of discharges) varied with the contour. Responses depended on the location of a given tone within a melody; cells may fire more vigorously when that tone is preceded by other tones rather than when it is the first. Moreover, cells react differently to the same tone when it is part of an ascending contour (low to high tones) than when it is part of a descending or more complex one. These findings show that the pattern of a melody matters: processing in the auditory system is not like the simple relaying of sound in a telephone or stereo system.

Although most research has focused on melody, rhythm (the relative lengths and spacing of notes), harmony (the relation of two or more simultaneous tones) and timbre (the char-

LAURIE GRACE

acteristic difference in sound between two instruments playing the same tone) are also of interest. Studies of rhythm have concluded that one hemisphere is more involved, although they disagree on which hemisphere. The problem is that different tasks and even different rhythmic stimuli can demand different processing capacities. For example, the left temporal lobe seems to process briefer stimuli than the right temporal lobe and so would be more involved when the listener is trying to discern rhythm while hearing briefer musical sounds.

The situation is clearer for harmony. Imaging studies of the cerebral cortex find greater activation in the auditory regions of the right temporal lobe when subjects are focusing on aspects of harmony. Timbre also has been "assigned" a right temporal lobe preference. Patients whose temporal lobe has been removed (such as to eliminate seizures) show deficits in discriminating timbre if tissue from the right, but not the left, hemisphere is excised. In addition, the right temporal lobe becomes active in normal subjects when they discriminate between different timbres.

Brain responses also depend on the experiences and training of the listener. Even a little training can quickly alter the brain's reactions. For instance, until about 10 years ago, scientists believed that tuning was "fixed" for each cell in the auditory cortex. Our studies on contour, however, made us suspect that cell tuning might be altered during learning so that certain cells become extra sensitive to sounds that attract attention and are stored in memory.

To find out, Jon S. Bakin, Jean-Marc Edeline and I conducted a series of experiments during the 1990s in which we asked whether the basic organization of the auditory cortex changes when a subject learns that a certain tone is somehow important. Our group first presented guinea pigs with many different tones and recorded the responses of various cells in the auditory cortex to determine which tones produced the greatest responses. Next, we taught the subjects that a specific, nonpreferred tone was important by making it a signal for a mild foot shock. The guinea pigs learned this association within a few minutes. We then determined the cells' responses again, immediately after the training and at various times up to two months later. The neurons' tuning preferences had shifted from their original frequencies to that of the signal tone. Thus, learning retunes the brain so that more cells respond best to behaviorally important sounds. This cellular adjustment process extends across the cortex, "editing" the frequency map so that a greater area of the cortex processes important tones. One can tell which frequencies are important to an animal simply by determining the frequency organization of its auditory cortex [*see box on opposite page*].

The retuning was remarkably durable: it became stronger over time without additional training and lasted for months. These findings initiated a growing body of research indicating that one way the brain stores the learned importance of a stimulus is by devoting more brain cells to the processing of that stimulus. Although it is not possible to record from single neurons in humans during learning, brain-imaging studies can detect changes in the average magnitude of responses of thousands of cells in various parts of the cortex. In 1998 Ray Dolan and his colleagues at University College London trained human subjects in a similar type of task by teaching them that a particular tone was significant. The group found that learning produces the same type of tuning shifts seen in animals. The long-term effects of learning by retuning may help explain why we can quickly recognize a familiar melody in a noisy room and also why people suffering memory loss from neurodegenerative diseases such as Alzheimer's can still recall music that they learned in the past.

Learning retunes the brain, so that more cells respond best to behaviorally important sounds.

Even when incoming sound is absent, we all can "listen" by recalling a piece of music. Think of any piece you know and "play" it in your head. Where in the brain is this music playing? In 1999 Andrea R. Halpern of Bucknell University and Robert J. Zatorre of the Montreal Neurological Institute at McGill University conducted a study in which they scanned the brains of nonmusicians who either listened to music or imagined hearing the same piece of music. Many of the same areas in the temporal lobes that were involved in listening to the melodies were also activated when those melodies were merely imagined.

Well-Developed Brains

STUDIES OF MUSICIANS have extended many of the findings noted above, dramatically confirming the brain's ability to revise its wiring in support of musical activities. Just as some training increases the number of cells that respond to a sound when it becomes important, prolonged learning produces more marked responses and physical changes in the brain. Musicians, who usually practice many hours a day for years, show such effects—their responses to music differ from those of nonmusicians; they also exhibit hyperdevelopment of certain areas in their brains.

THE AUTHOR

NORMAN M. WEINBERGER, who received his Ph.D. in experimental psychology from Case Western Reserve University, works in the department of neurobiology and behavior at the University of California, Irvine. He is a founder of U.C.I.'s Center for the Neurobiology of Learning and Memory and of MuSICA (Music and Science Information Computer Archive). A pioneer in the field of learning and memory in the auditory system, Weinberger is on the editorial board of *Neurobiology of Learning and Memory* and *Music Perception*.

Born to Rock?

MUSICAL COMMUNICATION, or "motherese," is common in all cultures.

Although many people think they are musically impaired, we are all musical to some degree. In fact, to find someone with a "musical brain," we need only look at any infant. Even before babies have acquired language, they exhibit a marked capacity for reacting to music. Perhaps that is why parents and others instinctively communicate with infants in a musical manner, using wide ranges of pitch and melodiclike phrases, often called "motherese." All cultures use motherese.

Beyond reacting positively to such communication, infants appear to encourage the performance of their mothers. In a 1999 study by Laura-Lee Balkwill and William F. Thompson, then both at York University in Toronto, North American and East Indian mothers sang the same song both with their infant present and absent. Others later were able to judge accurately in which of the two recordings the infant was present. The study showed as well that at least some musical cues appear to play across cultures. Listeners to the recordings could tell if the infant had been present or not regardless of whether they heard the song in their own language or in another.

How do we know infants are aware of music when they can't yet talk? We use objective measures of their behavior. For example, an infant sits on his mother's lap. To the left and right are two loudspeakers and adjacent transparent plastic boxes. Each box is ordinarily dark, but when the tot turns his head toward one it rewards him by lighting up and activating an animated toy, such as a dog or monkey. During testing, a researcher manipulates puppets or other objects directly in front of the baby to attract attention. A musical stimulus (which can be a single tone or a melody) plays repeatedly from one loudspeaker. At random times, the experimenter pushes a hidden button that changes the stimulus. If the infant notices the difference and turns toward the speaker, he is rewarded with the sight of the toy.

Such tests have revealed that infants differentiate between two adjacent musical tones as well as adults. Babies also notice changes in both tempo, the speed at which music is played, and rhythm. And they recognize a melody when it is played in a different key. Underscoring such studies, Sandra Trehub of the University of Toronto recently found that babies as young as two to six months prefer consonant sounds to dissonant ones. Music learning appears to begin even earlier, however—in utero. Peter Hepper of Queen's University in Belfast found that about two weeks before birth, fetuses recognized the difference between the theme music of the "Neighbors" TV show, heard daily by their mothers for weeks, and a novel song. *—N.M.W.*

Christo Pantev, then at the University of Münster in Germany, led one such study in 1998. He found that when musicians listen to a piano playing, about 25 percent more of their left-hemisphere auditory regions respond than do so in nonmusicians. This effect is specific to musical tones and does not occur with similar but nonmusical sounds. Moreover, the authors found that this expansion of response area is greater the younger the age at which lessons began. Studies of children suggest that early musical experience may facilitate development. In 2004 Antoine Shahin, Larry E. Roberts and Laurel J. Trainor of McMaster University in Ontario recorded brain responses to piano, violin and pure tones in four- and five-year-old children. Youngsters who had received greater exposure to music in their homes showed enhanced brain auditory activity, comparable to that of unexposed kids about three years older.

Musicians may display greater responses to sounds, in part because their auditory cortex is more extensive. Peter Schneider and his co-workers at the University of Heidelberg in Germany reported in 2002 that the volume of this cortex in musicians was 130 percent larger. The percentages of volume increase were linked to levels of musical training, suggesting that learning music proportionally increases the number of neurons that process it.

In addition, musicians' brains devote more area toward motor control of the fingers used to play an instrument. In 1995 Thomas Elbert of the University of Konstanz in Germany and his colleagues reported that the brain regions that receive sensory inputs from the second to fifth (index to pinkie) fingers of the left hand were significantly larger in violinists; these are precisely the fingers used to make rapid and complex movements in violin playing. In contrast, they observed no enlargement of the areas of the cortex that handle inputs from the right hand, which controls the bow and requires no special finger movements. Nonmusicians do not exhibit these differences. Further, Pantev, now at the Rotman Research Institute at the University of Toronto, reported in 2001 that the brains of professional trumpet players react in such an intensified manner only to the sound of a trumpet—not, for example, to that of a violin.

Musicians also must develop greater ability to use both hands, particularly for keyboard playing. Thus, one might expect that this increased coordination between the motor regions of the two hemispheres has an anatomical substrate. That seems to be the case. The anterior corpus callosum, which contains the band of fibers that interconnects the two motor areas, is larger in musicians than in nonmusicians. Again, the extent of increase is greater the earlier the music lessons began. Other studies suggest that the actual size of the motor cortex, as well as that of the cerebellum—a region at the back of the brain involved in motor coordination—is greater in musicians.

IT INTERNATIONAL/eSTOCK PHOTO

Ode to Joy—or Sorrow

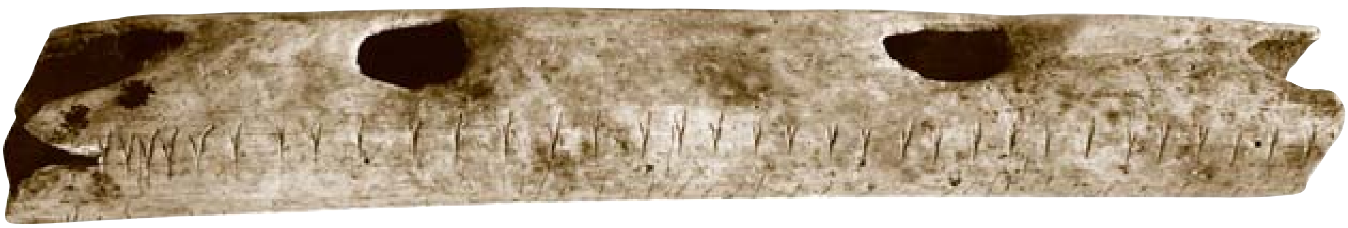
BEYOND EXAMINING HOW the brain processes the auditory aspects of music, investigators are exploring how it evokes strong emotional reactions. Pioneering work in 1991 by John A. Sloboda of Keele University in England revealed that more than 80 percent of sampled adults reported physical responses to music, including thrills, laughter or tears. In a 1995 study by Jaak Panksepp of Bowling Green State University, 70 percent of several hundred young men and woman polled said that they enjoyed music "because it elicits emotions and feelings." Underscoring those surveys was the result of a 1997 study by Carol L. Krumhansl of Cornell University. She and her coworkers recorded heart rate, blood pressure, respiration and other physiological measures during the presentation of various pieces that were considered to express happiness, sadness, fear or tension. Each type of music elicited a different but consistent pattern of physiological change across subjects.

Until recently, scientists knew little about the brain mechanisms involved. One clue, though, comes from a woman known as I. R. (initials are used to maintain privacy), who suffered bilateral damage to her temporal lobes, including auditory cortical regions. Her intelligence and general memory are normal, and she has no language difficulties. Yet she can make no sense of nor recognize any music, whether it is a previously known piece or a new piece that she has heard repeatedly. She cannot distinguish between two melodies no matter how different they are. Nevertheless, she has normal emotional reactions to different types of music; her ability to identify an emotion with a particular musical selection is completely normal! From this case we learn that the temporal lobe is needed to comprehend melody but not to produce an emotional reaction, which is both subcortical and involves aspects of the frontal lobes.

An imaging experiment in 2001 by Anne Blood and Zatorre of McGill sought to better specify the brain regions involved in emotional reactions to music. This study used mild emotional stimuli, those associated with people's reactions to musical consonance versus dissonance. Consonant musical intervals are generally those for which a simple ratio of frequencies exists between two tones. An example is middle C (about 260 hertz, or Hz) and middle G (about 390 Hz). Their ratio is 2:3, forming a pleasant-sounding "perfect fifth" interval when they are played simultaneously. In contrast, middle C and C sharp (about 277 Hz) have a "complex" ratio of about 8:9 and are considered unpleasant, having a "rough" sound.

What are the underlying brain mechanisms of that experience? PET (positron emission tomography) imaging conducted while subjects listened to consonant or dissonant chords showed that different localized brain regions were involved in the emotional reactions. Consonant chords activated the orbitofrontal area (part of the reward system) of the right hemisphere and also part of an area below the corpus callosum. In contrast, dissonant chords activated the right parahippocampal gyrus. Thus, at least two systems, each dealing with a different type of emotion, are at work when the brain processes emotions related to music. How the different patterns of activity in the auditory system might be specifically linked to these differentially reactive regions of the hemispheres remains to be discovered.

In the same year, Blood and Zatorre added a further clue to how music evokes pleasure. When they scanned the brains of musicians who had chills of euphoria when listening to music, they found that music activated some of the same reward systems that are stimulated by food, sex and addictive drugs.



BONE FLUTE from a site in France dates back at least 32,000 years—evidence that humans have been playing music since the dawn of culture.

ALEXANDER MARSHACK

Overall, findings to date indicate that music has a biological basis and that the brain has a functional organization for music. It seems fairly clear, even at this early stage of inquiry, that many brain regions participate in specific aspects of music processing, whether supporting perception (such as apprehending a melody) or evoking emotional reactions. Musicians appear to have additional specializations, particularly hyperdevelopment of some brain structures. These effects demonstrate that learning retunes the brain, increasing both the responses of individual cells and the number of cells that react strongly to sounds that become important to an individual. As research on music and the brain continues, we can anticipate a greater understanding not only about music and its reasons for existence but also about how multifaceted it really is. SA

MORE TO EXPLORE

The Origins of Music. Edited by Nils L. Wallin, Björn Merker and Steven Brown. MIT Press, 1999.

The Psychology of Music. Second edition. Edited by Diana Deutsch. Academic Press, 1999.

Music and Emotion: Theory and Research. Edited by Patrik N. Juslin and John A. Sloboda. Oxford University Press, 2001.

The Cognitive Neuroscience of Music. Edited by Isabelle Peretz and Robert J. Zatorre. Oxford University Press, 2003.

A Split at the Core

Physics is forcing the microchip industry to redesign its most lucrative products. That is bad news for software companies

By W. Wayt Gibbs

It was never a question of whether, but only of when and why. When would microprocessor manufacturers be forced to slow the primary propulsive force for their industry—namely, the biennial release of new chips with much smaller transistors and much higher clock speeds that has made it so attractive for computer users to periodically trade up to a faster machine? And would it be fundamental physics or simple economics that raised the barrier to further scaling? The answers are: in 2004, and for both reasons.

Production difficulties bedeviled almost every major semiconductor firm this year, but none were more apparent than the travails of Intel, the flagship of the microchip business. The company delayed the release of "Prescott," a faster version of its Pentium 4 processor, by more than six months as it worked out glitches in the fabrication of the 125-million-transistor chip. When Prescott did finally arrive, analysts were generally unimpressed by its performance, which was only marginally superior to the previous, 55-million-transistor Pentium 4. The company recalled defective batches of another microchip, postponed the introduction of new notebook processors, and pushed to next year a four-gigahertz Pentium model that it had promised to deliver this autumn.

The decision of greatest portent, however, was the one that Intel took in May to halt work on its next-generation Pentium 4 and Xeon processors. "They were probably a couple of years in design," estimates William M. Siu, manager of the company's desktop platforms group and the executive who proposed the cancellation. "It was obviously a significant decision," he says—not just because of the lost investment but because it means that the Pentium microarchitecture, the central engine both of Intel's business and of about three quarters of the world's computers, has reached the end of its life earlier than planned.

Beginning next year, all new Intel microprocessor designs for desktop and server computers will have not one but two "cores," or computational engines, on the same chip. Some high-end machines already have two or more microprocessors working side by side, as separate chips on a circuit board. But integrating multiple processors into one "multicore" chip involves a much more dramatic design change.

"When you bring those processors onto a single chip and reduce their interaction time to fractions of a nanosecond, that changes the whole equation," observes Justin R. Rattner, who joined Intel in 1973 and now directs its systems technology lab. "This is a major inflec-

ELIZA JEWETT

tion point" in computer architecture, he emphasizes. The shift to multicore processing has considerable ramifications for how computers are sold, how they are upgraded and—most significantly—how they are programmed.

Dodging the Danger Zone

"WE ARE NOT THE FIRST to do multicore," acknowledges Bob Liang, head of Intel's architecture research lab. In 2001 IBM introduced a dual-core processor, the Power4. "But we will be the first to bring it to a mass market," Liang claims. To make good on that promise, Intel will have to beat AMD, which in August demonstrated a dual-core version of its fast-selling Opteron processor and promised to have the chips in volume production by mid-2005. Meanwhile Sun Microsystems is rushing to develop "Niagara," a new microprocessor for network servers that boasts eight identical cores.

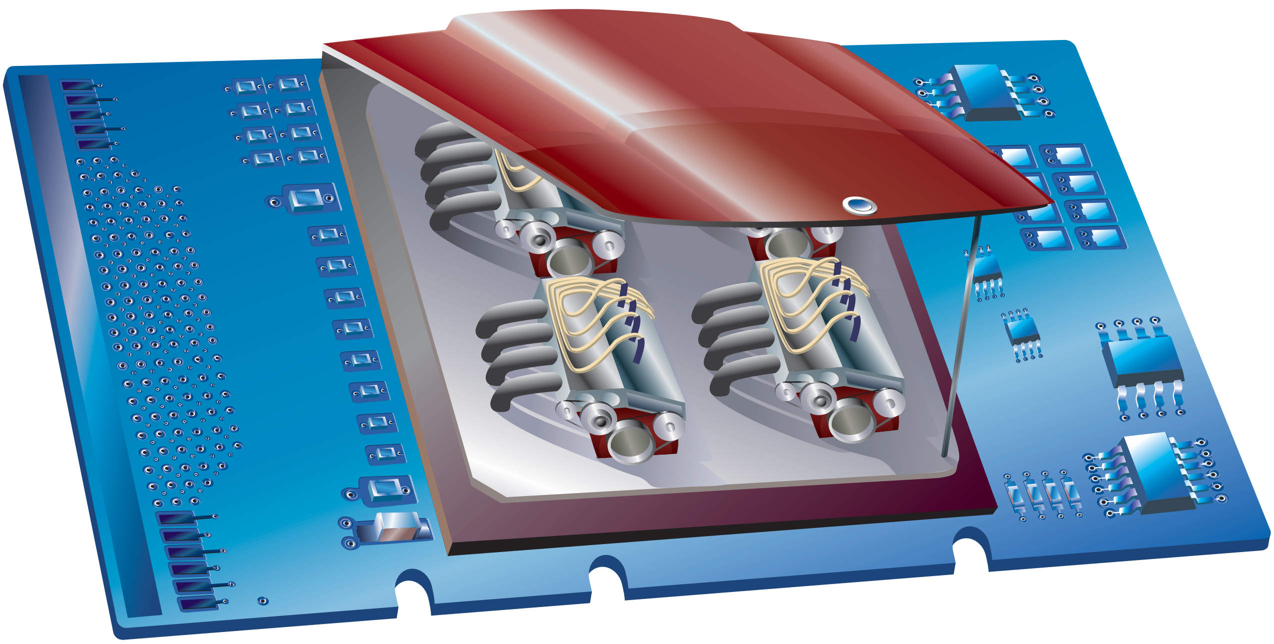
"The basic idea is to run them slower [than the single cores in today's processors] and make them simpler but to use more of them," says Stephen S. Pawlowski, who runs Intel's microprocessor technology lab. "Slower" and "simpler" are words rarely heard in the microchip industry—they give marketers migraines—yet that strategy may offer the only practical course around serious technical and economic obstacles.

Engineers have been able to continue shrinking the smallest logic gates, albeit with difficulty and some delays. The next generation of processors, expected in 2006, will knock the length of logic gates down from 50 nanometers (billionths of a meter) to 35. "We're now making test chips with half a billion transistors on that process," reports Mark T. Bohr, Intel's director of process technologies. Bohr says the industry is on track to produce chips with 18-nanometer-long gates by the end of the decade. So the number of switches that fit on a chip—the so-called transistor budget—is rising as quickly as ever [see "The First Nanochips," by G. Dan Hutcheson; SCIENTIFIC AMERICAN, April].

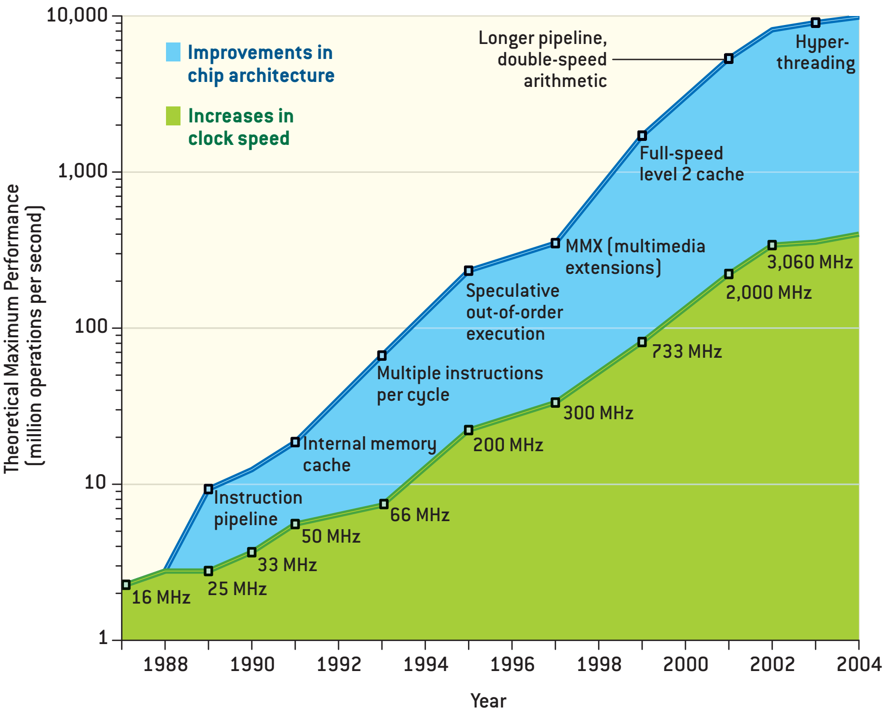
Heat and power budgets are tightening rapidly, however. The peak energy consumption of a microprocessor has soared to well over 110 watts in recent years as chipmakers have cranked up the clock frequencies at which processors run [*see lower chart on next page*]. Most of that energy ends up as heat; a new Pentium 4 can generate more heat, per square centimeter, than an iron set on "cotton." As engineers scale down transistor sizes and pack them more densely onto the thumbnail-size processor die, operating temperatures will rise further unless clock speeds stabilize.

"One limit we face is the threshold voltage of transistors, which is determined by their ability to shut off current," Bohr explains. He compares a transistor to a valve. "We used to have to turn the wheel three or four times to get it fully open or fully closed. Now we're dealing with valves that turn off if you move them just a few degrees to the left or right." The sensitivity of the transistors makes them leaky. Even when turned off, each typically draws 100 nanoamperes of current, Bohr says. That may not sound like much. "But multiply that times 100 million transistors, and it adds up."

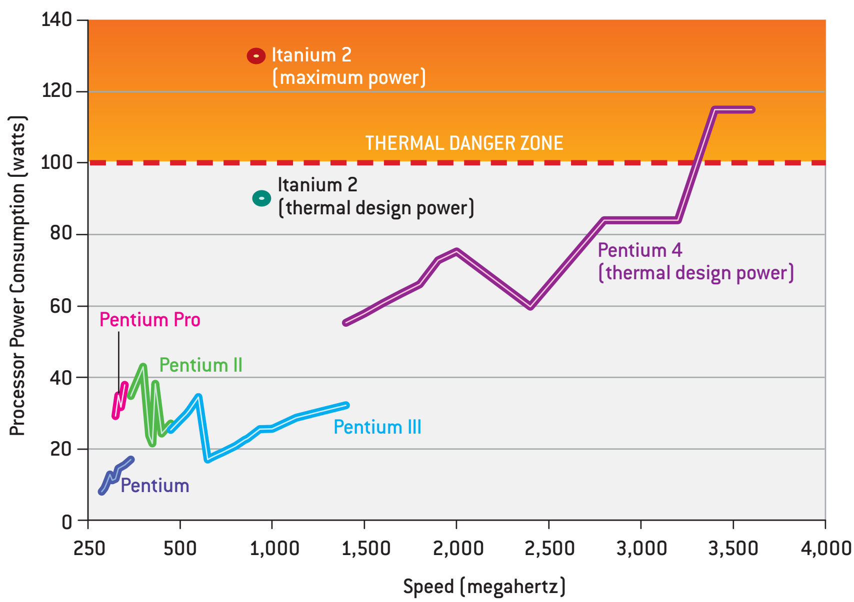
"When you push a certain thermal density, probably 100 watts per square centimeter, then you are really stressing the silicon," Pawlowski notes. Thanks to elaborate metal heat sinks and multiple fans, the hottest microprocessors still operate safely below that physical limit. "But the real issue is the cost of getting the heat out of the box, which is becom-



PEEK UNDER THE HOOD of a state-of-the-art microprocessor late next year, and you are likely to find two or more separate computing engines, running in parallel on a single slab of silicon. Such "multicore" designs alleviate old headaches for microchip engineers—but create new ones for software developers.



STEADILY INCREASING PERFORMANCE of Intel microprocessors has come mainly from quickening the clock pulse that sets the pace for the devices' transistors. But the company's engineers say that they will not be able to raise chip frequencies as quickly as in the past; further performance gains will have to come from innovations in the "architecture," or internal layout, of future processors. In fact, the next major generation of Intel processors will most likely use a radically different multicore architecture that may initially run at clock speeds below the current maximum frequency of 3,600 megahertz (MHz).



LATEST GENERATIONS OF INTEL PROCESSORS, the Pentium 4 and Itanium 2, are capable of burning more power than current heat sinks are designed to dissipate. To keep the chips from damaging themselves, Intel designers added circuits that monitor the processor temperature and throttle back the clock speed if a device runs too hot. Future generations of Intel processors will have multiple "cores" that run at lower speeds, generating less heat and spreading it more broadly across the surface of the chip.

ing prohibitive," he continues. "That is what creates this 'power wall.'"

Intel could require computer makers to switch from air-cooled to liquid-cooled machines. Apple Computer took that approach this year with its Power Mac G5 systems. But it adds to the cost. Intel sells roughly 50 times as many Pentium 4 systems as Apple sells G5s, in large part because Apple demands a premium price.

In any case, heat and power are not the only concerns. "When you have transistors on opposite corners of a chip, and you need to send a signal from one to the other, then those electrons have to flow through a copper wire," Bohr says. "The speed at which the electrons can flow is limited by the resistance and the capacitance of that wire. And while most wires in a chip are getting shorter, which is helpful, the wires are also getting thinner, which increases the delays caused by resistance and capacitance. So interconnects are inherently becoming more and more of a bottleneck."

That goes double for the relatively slow connection between the processor and the main memory bank. A microprocessor running at 3.6 gigahertz can execute several instructions each time its clock ticks, once every 277 trillionths of a second. But the system typically takes about 400 times that long to fetch information from the main memory. "The processor is just sitting there, waiting an eternity for each piece of data to come back from memory," Rattner observes.

Microprocessor architects have used on-chip memory caches and a technique called instruction-level parallelism to keep the processor busy working on instructions B and C while instruction A is waiting for its data to arrive. But that technique is nearly exhausted. "We're on the wrong side of a square law," Rattner says. "It is taking an exponential increase in transistors—and dramatic increases in the amount of power and chip area—to get even a modest increase in instruction-level parallelism."

Hence the dramatic change of strategy. Because the transistor budget is still rising by tens of millions of switches with each generation, engineers can ex-

ELIZA JEWETT; SOURCE: INTEL CORPORATION

A NEW STRATEGY: E PLURIBUS UNUM

COMPANY	DEVELOPMENT OF MULTICORE PROCESSORS
AMD	Demonstrated its first working dual-core processor in August; the chip is scheduled to reach market next summer
Cisco Systems	Released a new network router in May that uses 192-core processors to handle 1.2 trillion bits per second of Internet traffic
IBM	Was first to sell a dual-core processor, the Power4, in 2001; Power5 processor introduced in May also sports two cores
Intel	Has prototype dual-core Itanium 2 chips; announced in May that new desktop and server microprocessor designs will have multiple cores
Sun Microsystems	UltraSparc-IV processor unveiled in February is dual-core; "Niagara" chip scheduled to appear by early 2006 will have eight cores

ploit higher levels of parallelism by divvying the chip into multiple cores.

Intel's first dual-core chips will probably run at lower frequencies than the fastest Pentiums. But clock speeds will still continue to rise, asserts Patrick P. Gelsinger, Intel's chief technical officer, "just much more gradually than in the past." Intel recently relabeled its chips with abstract model numbers instead of the megahertz ratings it has used for 15 years. Gelsinger predicts that from now on, 70 percent of performance gains will come from architectural improvements—mainly parallelism—rather than from additional megahertz.

Life in a Parallel Universe

IN PRINCIPLE, multicore processors could work more efficiently and more flexibly than today's single-core chips do. A notebook processor might have eight cores; a program customized for such a chip could divide itself into many "threads," each running simultaneously on a different core. Alternatively, the operating system might turn off some of the cores to extend battery life.

"The cores don't have to be identical," Siu points out. Building a variety of different cores could help deal with the fact that most existing software has no idea how to exploit a multicore processor. "You could have a big single-threaded core that can run legacy applications and also a bunch of small cores sitting on the side to run new [multicore-savvy] applications," Pawlowski elaborates.

But then he pauses to think about that prospect: "Quite frankly, it is going to take the software community a long time to start working on that. Unfortunately, very few people have played in this space."

"One of the big problems with parallelism for 40 years has been that it is hard to think about it and hard to do," says David J. Kuck, director of the KAI Software Lab, a company that Intel bought to help it make this transition. "When these threads are handing things back and forth, everyone gets lost."

A parallel processor deprives the programmer of one of the most valuable tools for debugging: repeatability. "A threaded [parallel] program is not a deterministic thing," Kuck explains. "It may execute one way one time and a different way the next time, just because of subtle timing differences in the machine's state. So most [software executives] think: 'Oh, my God, I just don't want to face this.' That holds all the way up and down the line from Oracle

MORE TO EXPLORE

International Technology Roadmap for Semiconductors. International Sematech, 2003. Available at **http://public.itrs.net/Files/2003ITRS/Home2003.htm**

Architecting the Era of Tera. Intel Research and Development, 2004. Available at **ftp://download.intel.com/labs/nextnet/download/Tera_Era.pdf**

and Microsoft to the little guys," he says.

Even setting aside the difficulty of rewriting software in parallel form, "there are some applications where you won't get any boost from multicore. So it's just lost," acknowledges Glenn J. Hinton, director of microarchitecture development at Intel.

But certainly many kinds of tasks could run dramatically faster when redesigned for multicore chips. When converting a home movie to DVD format, for example, several frames can be processed in parallel. Rendering 3-D scenes, manipulating photographs, running scientific models, searching through databases and similar tasks can all be more quickly conquered once divided. A few specialized tasks could exploit as many cores as chipmakers can throw at them.

For general-purpose computing, however, "there is a point of diminishing returns," Pawlowski avers. "Sixteen cores is not clearly better than eight."

The most worrisome long-term question for the microprocessor industry may be whether the shift to multicore processors will discourage its customers from upgrading to newer computers. Today's computers are more than fast enough to handle most popular software. Demand for speedier machines has already begun to flag. In July, Intel reported that its inventory of unsold products had risen by 15 percent; the bad news knocked 11 percent off its share price.

A major design change adds uncertainty to apathy as a reason that computer owners might postpone an upgrade. It is not yet clear whether customers who buy the first dual-core machines and replace much of their software to suit the new architecture will have to repeat the process three years later to take advantage of "quad-core" machines. Faced with that prospect, many users—and for that matter, many software companies—could decide that the new architecture is simply not worth the hassle. On the other hand, the most obvious lesson from the history of computing is that every leap in performance is never enough for long. SA

W. Wayt Gibbs is senior writer.

WORKINGKNOWLEDGE

PACEMAKERS

Keep the Beat

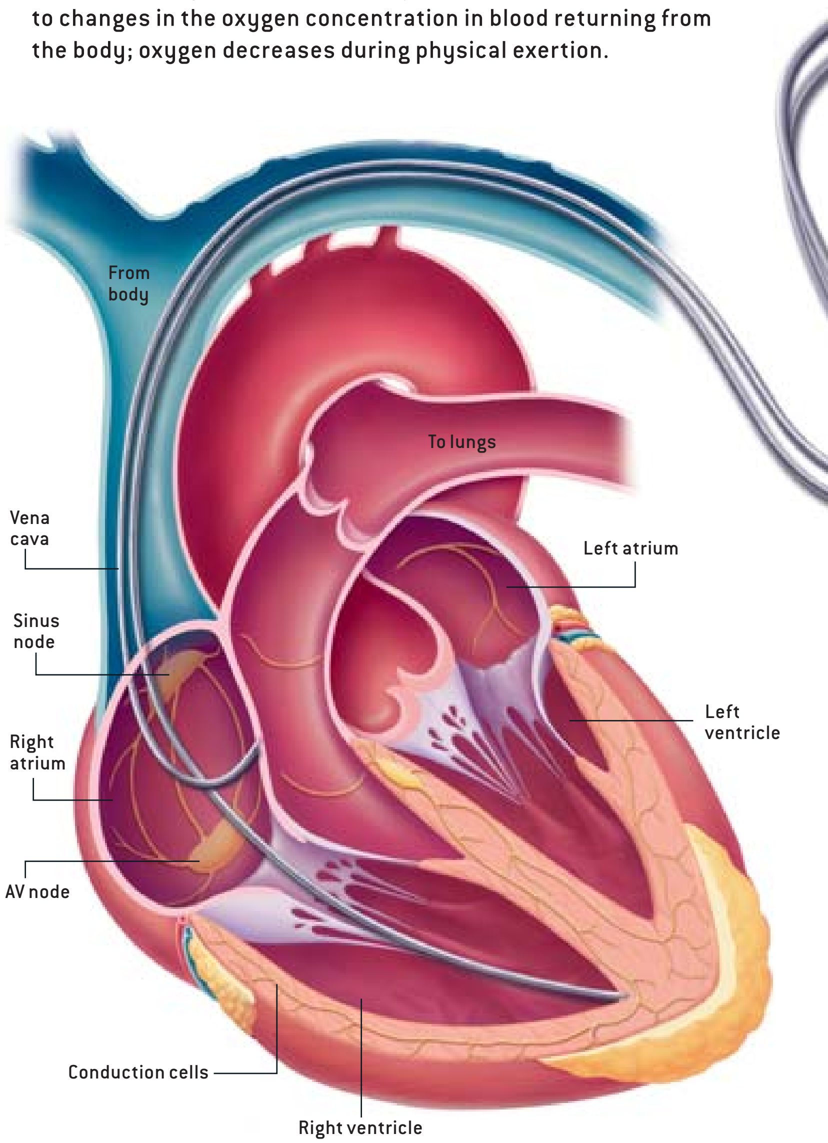
Almost four million people worldwide have pacemakers, which were first deployed experimentally 50 years ago to speed hearts that beat too slowly.

In a healthy heart, electrically conductive muscle cells fire for each contraction of the atria and ventricles [*see illustration*]. In cases of distress, a pacemaker sends an electrical impulse to select points in the heart to speed up or coordinate the firing. Historically pacemakers could deliver one pulse rate, but most modern instruments are rate-responsive—they track conduction activity as well as body motion and adjust heart rate accordingly. Doctors can now also implant special pacemakers that stimulate each ventricle directly, resynchronizing ventricles that contract out of phase. This condition often occurs in people whose cardiac muscle has been weakened by a heart attack.

Pacemakers have shrunk to the size of a matchbox, thanks to better electronics and software but also to a reservoir of steroid at the tip of the wire that sparks the heart. By reducing inflammation after implantation, the drug keeps the electrode close to viable heart muscle, thus lowering the voltage the pacemaker must generate and allowing for smaller components. "The steroid has been key," says Toby Markowitz, a Medtronic distinguished scientist who oversees pacemaker research at the Minneapolis company. When batteries fade, usually after six to 10 years, the entire unit (but not the leads) is replaced, because the battery is fused with the case and because pacemaker technology generally improves markedly over that interval.

Since the 1980s a technical spin-off has also gained prominence: the implantable cardioverter defibrillator (ICD). The size of a small pager, it can shock atria or ventricles that are quivering (fibrillating) or racing out of control so the patient's heart can reset itself to normal, preventing the sudden cardiac arrest that kills 450,000 Americans every year. About 80,000 ICDs were implanted in 2003. ICDs can provide pacing, too, and Markowitz says they could one day replace pacemakers if they come down in size and cost—and if insurers, particularly Medicare, approve the switch. —*Mark Fischetti*

ELECTRICAL CONDUCTION triggers each heartbeat. Muscle cells at the sinus node depolarize as the atria fill with blood, setting off a chain of depolarization—an electrical impulse wave—toward the atrioventricular (AV) node. The AV node slows the impulse until the atria fill the ventricles. It then forwards the impulse around the ventricles, prompting them to contract simultaneously. The sinus node adjusts heart rate in response to changes in the oxygen concentration in blood returning from the body; oxygen decreases during physical exertion.



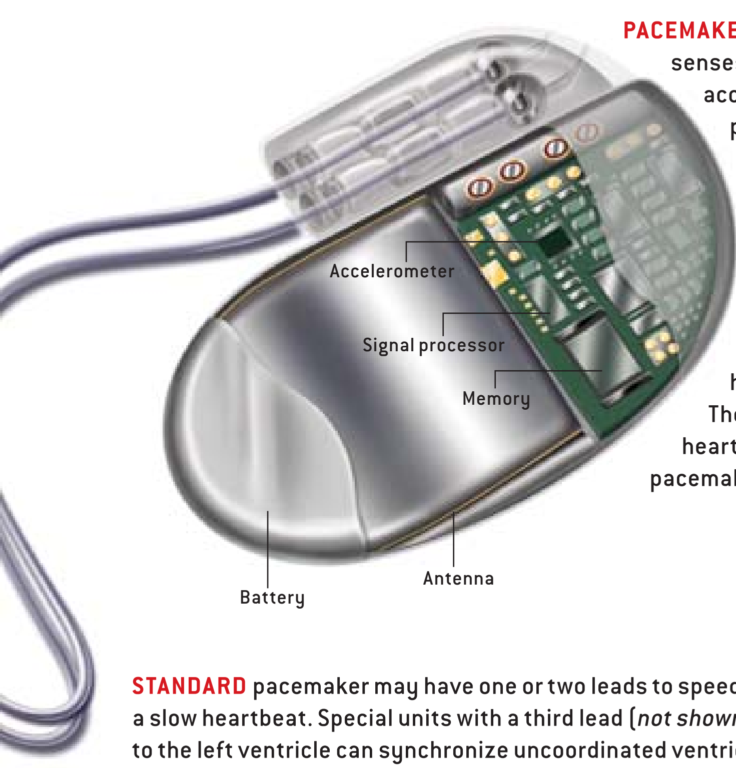
ALICE CHEN

DID YOU KNOW. . .

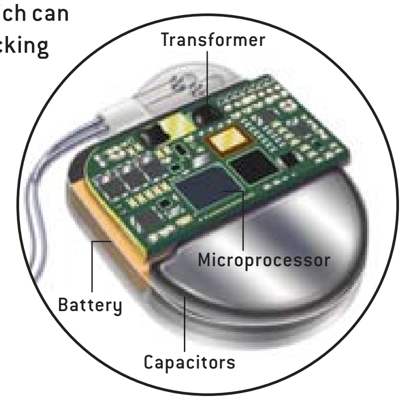
➤ BACKUP: All cardiac cells can beat on their own. Cells in the sinus node beat most quickly, though, and therefore start the contraction impulses that sweep across the heart, establishing a resting pulse of 60 to 100 beats per minute (bpm). If the sinus is damaged, atrium cells will take over the conductor duties but can only set a pace of 40 to 80 bpm. If they fail, AV node cells can muster 30 to 40 bpm—just enough to keep a human alive.

➤ YOU MAKE THE CALL: Patients can check their pacemakers remotely. The patient places wristbands on each arm and holds a magnetic wand against his chest, over the pacemaker. He then places his telephone handset on a transmitter pad connected to the wristbands and wand and calls the doctor, who tests the pacemaker's functions and battery life. New units can send data to a doctor's computer.

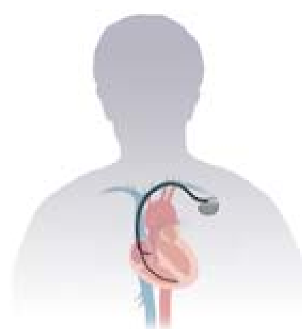
➤ DON'T GO THERE: According to Medtronic, pacemaker wearers need not worry about electromagnetic interference from microwave ovens, electric blankets or airport metal detectors. Antitheft arches by store doorways are fine if patients don't linger next to them. Trouble can arise by getting too close to items that produce strong magnetic fields, such as large stereo speakers and gasoline engines. Cell phones should not be kept in a breast pocket. Magnetic resonance imagers are off limits. Moving away from a source ends the complication; pacemaker damage rarely occurs.



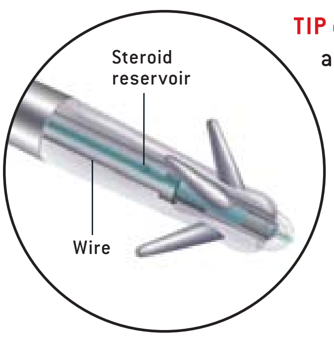
PACEMAKER circuitry emits a 1- to 5-volt impulse if the signal processor senses that the heart's pulse is too slow, stopped or uncoordinated. An accelerometer tells the unit to quicken the heartbeat if it senses physical activity. A technician can retrieve information from the memory using a magnetic wand that communicates through the skin via radio-frequency signals to an antenna. Implantable cardioverter defibrillators (*inset*) correct quivering (fibrillation) by the atria or ventricles, which can cause cardiac arrest, by shocking the heart with 30 joules or more of energy, resetting the heart's conduction system. The unit can also stop a racing heartbeat (tachycardia) and provide pacemaker functions.



STANDARD pacemaker may have one or two leads to speed up a slow heartbeat. Special units with a third lead (*not shown*) to the left ventricle can synchronize uncoordinated ventricle contractions, which plague some people who have weak or damaged hearts.



MAIN CASE of a pacemaker or ICD is implanted in the chest, and leads are threaded through veins to the heart.

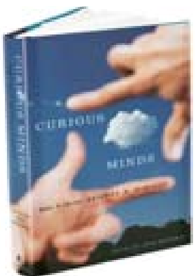


TIP of a wire lead is fixed into heart tissue and delivers a small electrical impulse that causes conduction cells to fire. The wire also senses heartbeat, tracked by the pacemaker. A reservoir of steroid is slowly released for several months after implantation, minimizing inflammation and the chance of rejection.

Have an idea for a topic? Send it to workingknowledge@sciam.com

Terrified by a Tyrannosaur

AND OTHER INSIGHTS INTO WHY PEOPLE CHOOSE SCIENCE **BY RICHARD MILNER**



CURIOUS MINDS: HOW A CHILD BECOMES A SCIENTIST
Edited by John Brockman
Pantheon Books, New York, 2004 ($23.95)

When the late evolutionist and polymath Stephen Jay Gould was a toddler, he became fascinated and terrified by the towering *Tyrannosaurus rex* skeleton at the American Museum of Natural History. Gould later claimed to have been instantly "imprinted" on the monstrous saurian, like a duckling on its mama. The little boy decided on the spot to become a paleontologist—years before he even learned the word.

In John Brockman's *Curious Minds: How a Child Becomes a Scientist,* a collection of 27 autobiographical essays by leading savants, Harvard psychologist Steven Pinker scoffs at this oft-told story. Pinker relates that Gould dedicated his first book: "For my father, who took me to see the *Tyrannosaurus* when I was five," and admires Gould's "genius ... for coming up with that charming line." But he doesn't buy it.

Pinker goes on to tell his own childhood story, with the caveat that long-term memory is notoriously malleable and that we often concoct retrospective scenarios to fit satisfying scripts of our lives. So don't believe anything in this book, he warns, including his own self-constructed mythology; many children are exposed to books and museums, but few become scientists. Pinker concludes that perhaps the essence of who we are from birth shapes our childhood experiences rather than the other way around.

Nevertheless, when Brockman asked Pinker and others to trace the roots of their adult obsessions for this book, he received some unexpected and entertaining responses. Primatologist Robert Sapolsky, for example, haunted the Bronx Zoo and the natural history museum, as Gould did, but fell in love with living primates rather than fossil bones. He didn't want to just study mountain gorillas, he recalls of his childhood crush on monkeys and apes, "I wanted to *be* one." For the past few decades, Sapolsky has spent half of each year in his physiology lab and the other half among wild baboon troops in East Africa.

Some people, such as theoretical psychologist Nicholas Humphrey, are simply born into science. His grandfather, Nobel laureate A. V. Hill, often took him along to the physiology lab. Grandfather Hill—quoting his friend Ivan Pavlov—taught young Nicholas that "facts are the air of a scientist. Without them you can never fly." Among frequent visitors to the family home were his great-uncles Maynard and Geoffrey Keynes, members of British science's aristocracy, as well as his great-aunt Margaret, a granddaughter of Charles Darwin. He recalls how their long-term houseguest, an adolescent, "bossy" Stephen Hawking, once marched up and down the hallways clutching a military swagger stick, barking at a "platoon of hapless classmates." Science was Humphrey's birthright.

Richard (*The Selfish Gene*) Dawkins, one of England's preeminent Darwinians, admits that he never cared for science or the natural world during his early years. He was inspired, however, by the fanciful children's books about Dr. Dolittle by Hugh Lofting. The good doctor was a Victorian gentleman who held intelligent conversations with mice and parrots and whales. An adventurous sort, he traveled the world to learn the secrets of faraway places. When the adult Dawkins encountered the life and works of Charles Darwin, he welcomed him as an old friend and hero of his youth. Dolittle and Darwin, he opines, "would have been soul brothers."

Lynn Margulis's early interest in the wonders of the microscopic world began when she was a "boy crazy" adolescent, who was amazed to learn that some minuscule creatures never need sex in order to reproduce. Enter a teenage heartthrob:

IRVING MILNER

AT AGE 12, Stephen Jay ("Fossilface") Gould and pal Richard ("Dino") Milner were already fascinated by evolutionary biology. Queens, New York, school yard, 1953.

the budding astrophysicist Carl Sagan. ("Tall, handsome in a sort of galooty way, with a shock of brown-black hair, he captivated me.") She was 16 when they met; eventually they married.

Sagan's fascination with "billions and billions" of cosmic bodies resonated with her own fixation on the billions of microcosms to be observed through the microscope. Margulis's study subjects have included a tiny animal in a termite's gut that is made up of five distinct genomes cobbled together. She has argued that we and other animals are composite critters, whose every cell harbors long-ago invaders—minute symbiotic organisms that became part of our makeup. Her innovative approach to evolution has profoundly influenced biology.

Harvard psychologist and neurologist Howard Gardner says his youth was notable for its *lack* of any clues indicating a future in science: "I did not go around gathering flowers, studying bugs, or dissecting mice … I neither assembled radios nor tore apart cars." Yet, for others, there was a decisive turning point. And some could clearly remember it.

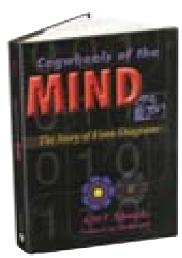
I was fortunate in having been a childhood friend of Steve Gould's and can vouch for the sincerity of his conviction that his extraordinary career as a paleontologist, historian of science and evolutionary theorist began when that *T. rex* followed him into his nightmares. Once, during our junior high school days, I stood with him beneath that iconic carnosaur in the museum, observing his reverence and awe on revisiting the shrine of his inspiration. Professor Pinker, of course, is free to believe that I'm making this up for my own psychological reasons. SA

Richard Milner is an associate in anthropology at the American Museum of Natural History. His new book, Darwin's Universe, *will be published by the University of California Press in 2005.*

THE EDITORS **RECOMMEND**

COGWHEELS OF THE MIND: THE STORY OF VENN DIAGRAMS

by A.W.F. Edwards. Foreword by Ian Stewart. Johns Hopkins University Press, Baltimore, Md., 2004 ($25)



Three overlapping circles intersect to create eight distinct areas: for well over a century, the classic Venn diagram has helped delineate relationships in fields as diverse as theoretical physics and business strategy. Edwards, a statistician at the University of Cambridge, provides an insightful history of the diagrams, which were developed by English logician John Venn (1834–1923), discussing their presence in everything from Christian iconography to tennis balls and flags.

J. ROBERT OPPENHEIMER AND THE AMERICAN CENTURY

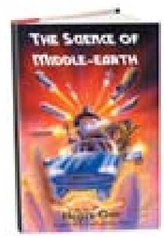
by David C. Cassidy. Pi Press, New York, 2004 ($27.95)

The "American Century" was a concept put forward in 1941 by publisher Henry Luce, who declared that in the years ahead the U.S. must "exert upon the world the full impact of our influence, for such purpose as we see fit and by such means as we see fit." The eminent theoretical physicist J. Robert Oppenheimer, as director of the Manhattan Project to build the atomic bomb, had much to do with forging the alliance between science and government that contributed greatly to the realization of Luce's concept. Cassidy, professor in the natural sciences program at Hofstra University, covers this ground admirably in his thoughtful biography of Oppenheimer. Telling the story against a background of the events of the time, he takes Oppenheimer from a cosseted childhood through his distinguished career as a scientist and science administrator to disgrace in 1954 when the Atomic Energy Commission withdrew his security clearance after politicized hearings on his loyalty and leftist affiliations. "The Oppenheimer case," Cassidy writes, "cast in stark relief the subservient position imposed on civilian research, especially physics, during the darkest days of the Cold War. It was not primarily about Oppenheimer as an individual but about the existence of free inquiry in a 'garrison state,' and what role a scientific adviser might have within a system of militarized science beyond providing weapons of ever greater destructive power."

THE SCIENCE OF MIDDLE-EARTH: EXPLAINING THE SCIENCE BEHIND THE GREATEST FANTASY EPIC EVER TOLD!

by Henry Gee. Cold Spring Press, Cold Spring Harbor, N.Y., 2004 ($14)

How did Frodo's mithril coat ward off the fatal blow of an orc? How was Legolas able to count the number of riders crossing the plains of Rohan from five leagues away? Could Balrogs fly? Gee, a senior editor at *Nature* (who says he read *The Lord of the Rings* about once a year between the ages of 10 and 25), elucidates and expands on the scientific aspects of J.R.R. Tolkien's world in this fascinating book. Many commentators have noted Tolkien's use of philology and cultural history to create believable languages for his elves and orcs. Now Gee shows how scientific precepts can make the wonders of Middle-earth even richer. In a closing essay, he argues that "Tolkien's own worldview was closer to the true spirit of science than that held by many who propose to promote the public understanding of science."



The books reviewed are available through www.sciam.com

Wild Life

OF BEERS, BEARS, FISH AND ASSAULT WITH A DEADLY REPTILE **BY STEVE MIRSKY**

Human beings have close encounters with other species on a daily basis. For example, in the past 24 hours I have shaken a large spider from a bath towel; taken care not to hit a frog with my car; carried a hitchhiking grasshopper on my bike; been leapt on by a big dog; been rubbed against by a small cat; fed upon a chicken; been fed upon by a mosquito. Such a day is fairly tame. But this past summer you couldn't swing a dead alligator without hitting a news story about wild interactions between *Homo sapiens* and other organisms.

Take the guy who swung the *live* alligator. Here's the beginning of the Associated Press account, out of Port Orange, Fla., in July: "A man hit his girlfriend with a three-foot alligator and threw beer bottles at her during an argument in the couple's mobile home, authorities said." This relatively short sentence (which is also what a judge will probably give the gator swinger) is so evocative that further comment would only dull its luster.

A few weeks later a better use of beer was discovered by an innocent ursine in the Pacific Northwest. "A black bear," the Reuters coverage began, "was found passed out at a campground in Washington State recently after guzzling down three dozen cans of a local beer." (Animal alcohol abuse is not unknown. In 2002 elephants in India found some rice beer and went on a drunken rampage. Many birds have found that flying and fermented berries don't mix.) Our buzzed bear, who busted into coolers at Baker Lake Resort, northeast of Seattle, drank his dinner discriminatingly: after trying another brand, the bear limited his libations to Rainier Beer. And that brewer appreciated what the market would, well, you know. Rainier named the bear its unofficial spokesbeast and then literally named the bear, through a contest. The winning entry was Brewtus, although I was hoping for either Logy Bear or, of course, Booze Booze.

And it's no surprise that bears have taken to drink, for it seems that more people have taken to bears—to eat. But the bears have a secret weapon. The Centers for Disease Control and Prevention's *Morbidity and Mortality Weekly Report* noted in July that "during 1997–2001, a total of 72 cases of trichinellosis … were reported to CDC; the majority of these infections were associated with eating wild game, predominantly bear." The bears harbor the trichinella roundworm, which happily infects people who keep too low a fire on Smokey. It's not exactly a rampaging epidemic, with an average of about 12 cases a year. But even that number puts the lie to the the old adage "Sometimes you get the bear [pronounced "bar" in old adages], sometimes the bear gets you." More accurately, "Sometimes you get the bear, sometimes the bear gets you, and every once in a while you get the bear, a high fever, chills, aches, diarrhea and a visit to a parasitologist."

Meanwhile bears and people share a taste for salmon. So the first sentence of an August press release from the American Chemical Society (ACS) was of interest both to us and to any literate bears: "Farm-raised salmon contain much higher levels of flame retardants than most wild salmon, and some wild Chinook have the highest levels of all, according to new research."

The release, discussing a report in the ACS journal *Environmental Science & Technology,* revealed that many salmon now come with the secret ingredient polybrominated diphenyl ethers (PBDE), "used widely as flame-retardant additives in electronics and furniture." A salmon industry press release countered with the headline "Salmon Flame Retardant Study Shows No New Data; Consumers Should Not Be Alarmed." Or five-alarmed. Anyway, is it still okay to regularly consume salmon with FDNY, I mean, PBDE? The two sides disagree, no surprise, so I offer this helpful salmon rule of thumb: if you can't cook it, don't eat it. Fin. SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

ASK THE EXPERTS

How do scientists know the composition of Earth's interior? —J. Gerber

Arthur Lerner-Lam, associate director for seismology, geology and tectonophysics at the Lamont-Doherty Earth Observatory at Columbia University, explains:

Because we cannot extract samples directly from the deep Earth, we must deduce its composition. Scientists look at the clues hidden in rocks that are igneous (volcanic) or metamorphic (those that have changed after exposure to high pressures and temperatures underground). They also examine proxies for composition and structure, such as the three-dimensional variation of the velocity of seismic waves produced by earthquakes, which are sampled by networks of seismometers on the surface of the planet.

The late Francis Birch, the eminent Harvard geophysicist, and his colleagues and students worked out the basic methodology that brings these distinct observations together. Birch showed how the stiffness of rocks changes under the extreme conditions of pressure and temperature deep within planets. Because the speed of seismic waves depends on the stiffness of the medium through which they propagate, it is possible to calculate temperature and composition from maps of seismic velocity. Most current research is based on Birch's work, and it has even been extended to the most extreme temperature and pressure conditions of Earth's core. For example, much of our understanding of the large- and small-scale convection patterns driving plate tectonics has come about by using Birch-type proxies for temperature and composition.

How can we improve our knowledge of the other planets? Manned and unmanned missions to the moon and Mars deployed seismometers, which provided tantalizing, but ultimately limited, information before they stopped operating (although the Spirit and Opportunity rovers continue to transmit chemical analyses and pictures of the Red Planet back to Earth). Almost all planetary landing missions now in the design stage include seismological instrumentation, and some even plan to return rock samples to Earth. We hope the best geoscience is yet to come.

How does decanting red wine affect its taste? And why not decant white? —J. Eastley, Pittsburgh, Pa.

Andrew L. Waterhouse, a professor in the department of viticulture and enology at the University of California at Davis, provides an answer:

Decanting—simply pouring wine into another container—can mellow the flavor of harsh younger red wines by exposing them to oxygen; it also serves to remove sediments in older vintages. White wines, which are aged for shorter periods, do not benefit from decanting.

Some young red wines—between three and 10 years old—can be astringent if consumed directly after opening the bottle because they are maintained in an environment relatively free of oxygen during aging. Over time, these conditions result in a "closed" character for the wine that comes from the accumulation of particular compounds. A wine's aroma changes during the first 10 to 30 minutes after opening. Decanting accelerates this so-called breathing process, which increases the wine's aroma from natural fruit and oak, by allowing a few volatile substances to evaporate. Decanting also appears to "soften" the taste of the tannins that cause harshness and astringency, although chemists have not detected noticeable changes to the tannins. Less dramatic changes result from just uncorking the bottle 15 to 60 minutes before pouring. In older wines, decanting also serves to remove sediments, which are harmless but make wine look cloudy and taste gritty. Unlike younger wines, older wines should be served immediately after decanting—the "bottle bouquet" can be fleeting.

The author would like to thank Kay Bogart for her help in preparing this answer. SA

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

MATT COLLINS