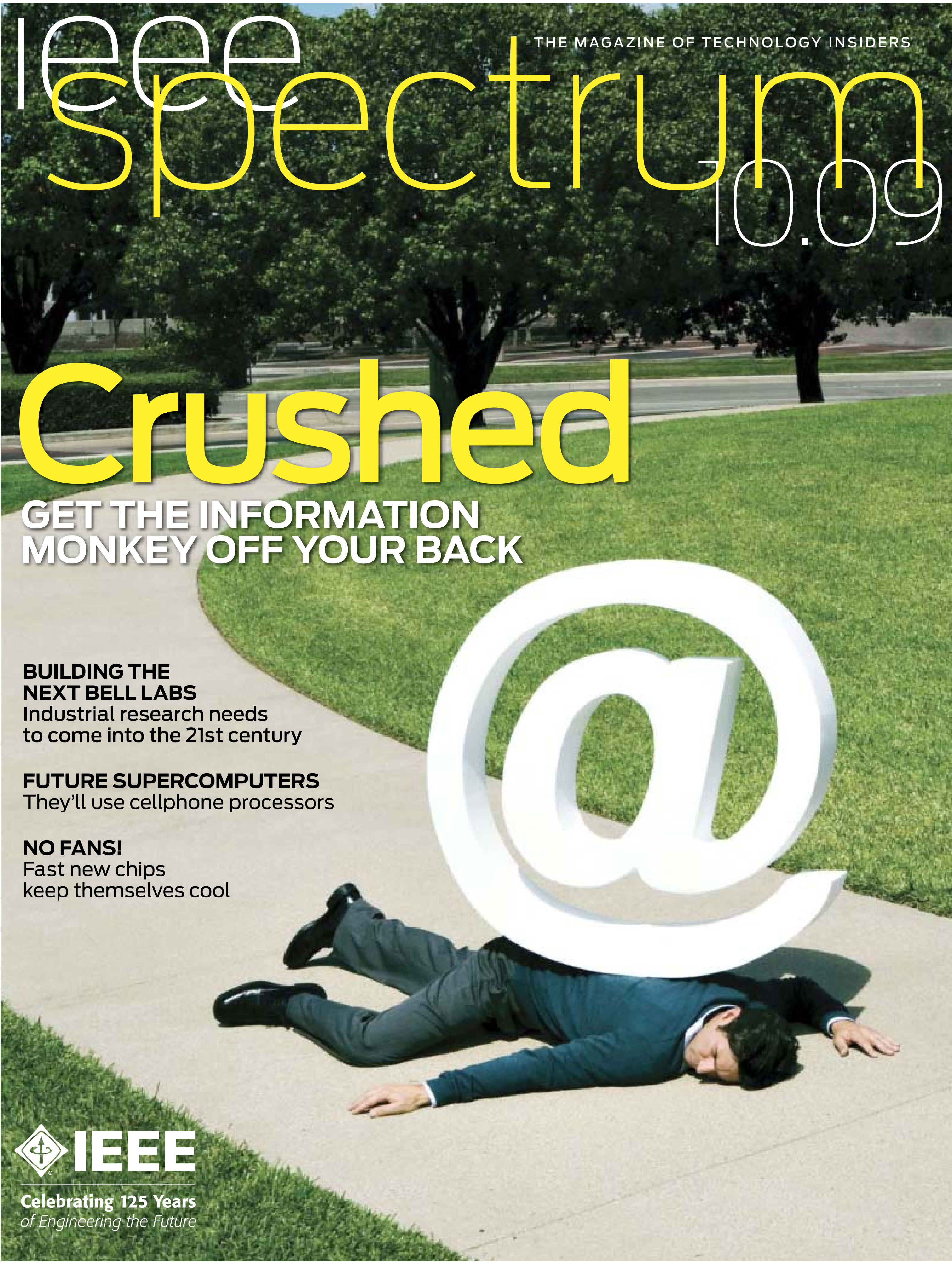
ieee spectrum
THE MAGAZINE OF TECHNOLOGY INSIDERS
10.09



Crushed

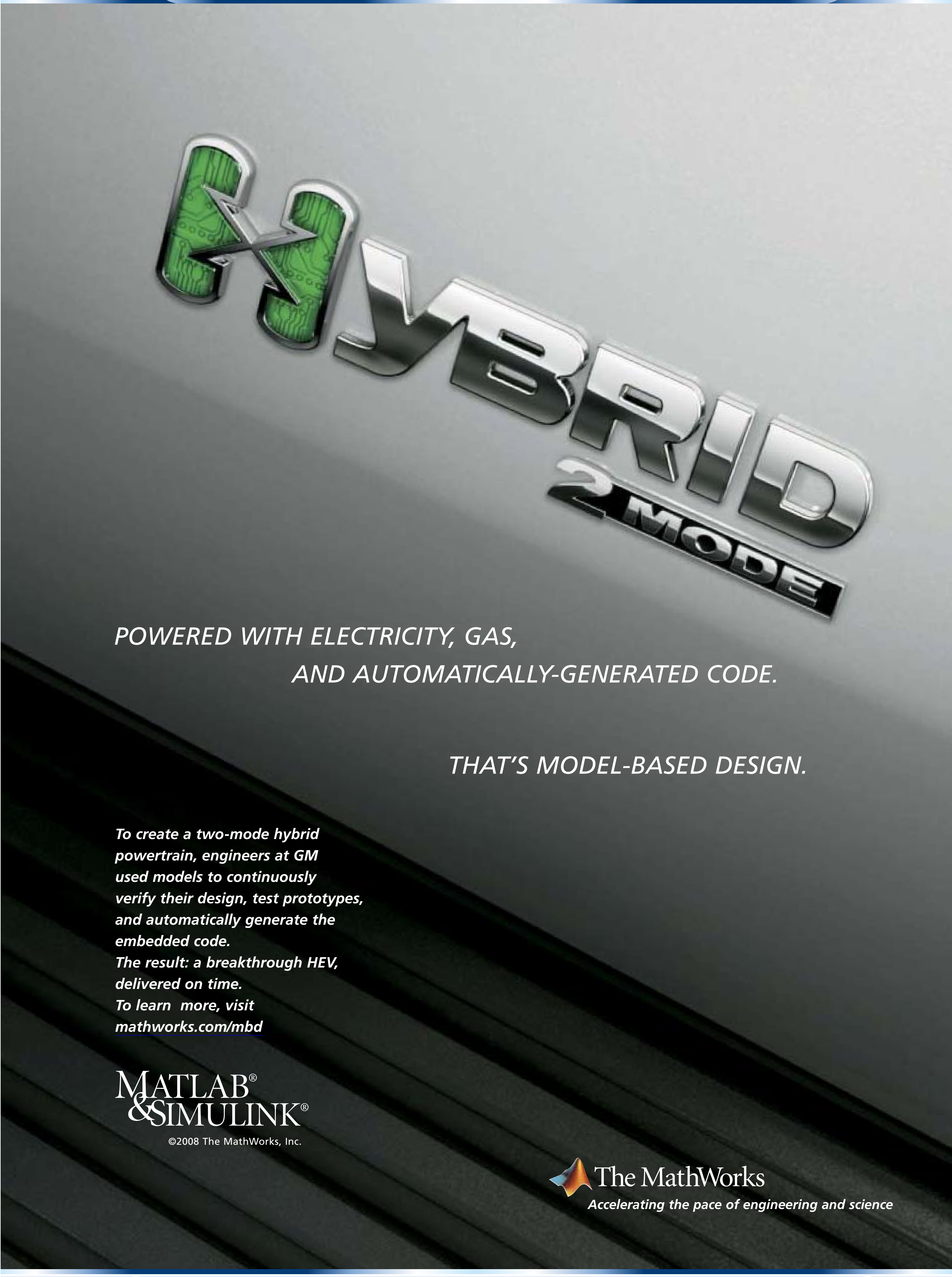
GET THE INFORMATION MONKEY OFF YOUR BACK

BUILDING THE NEXT BELL LABS
Industrial research needs to come into the 21st century

FUTURE SUPERCOMPUTERS
They'll use cellphone processors

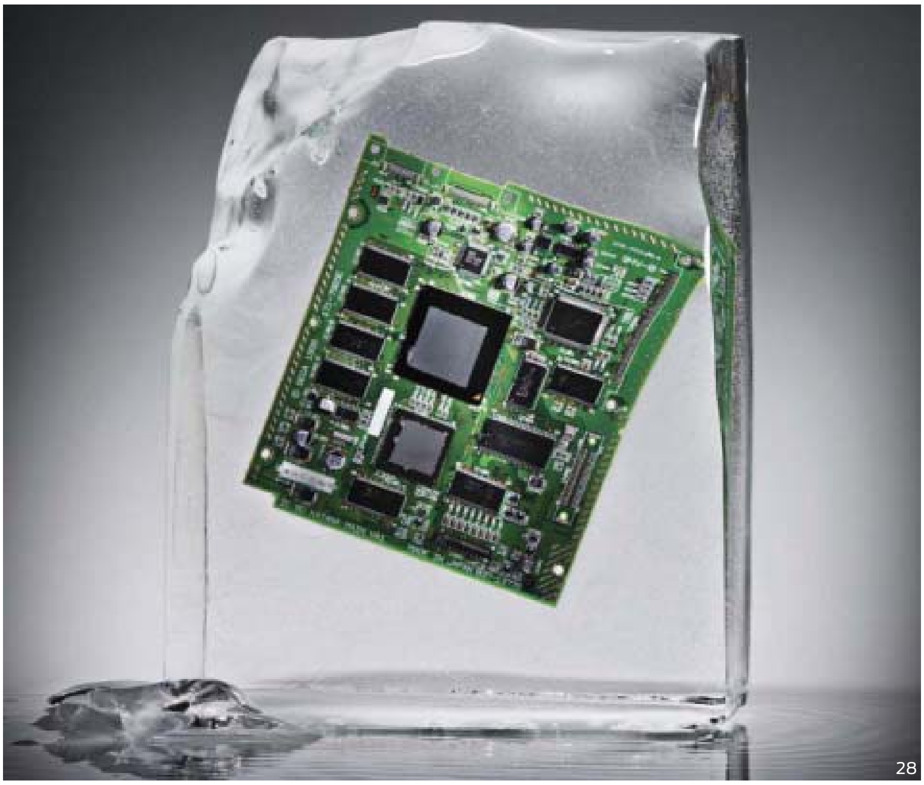
NO FANS!
Fast new chips keep themselves cool

IEEE
Celebrating 125 Years
of Engineering the Future


HYBRID
2 MODE
POWERED WITH ELECTRICITY, GAS,
AND AUTOMATICALLY-GENERATED CODE.
THAT'S MODEL-BASED DESIGN.
To create a two-mode hybrid powertrain, engineers at GM used models to continuously verify their design, test prototypes, and automatically generate the embedded code.
The result: a breakthrough HEV, delivered on time.
To learn more, visit mathworks.com/mbd
MATLAB® & SIMULINK®
©2008 The MathWorks, Inc.
The MathWorks
Accelerating the pace of engineering and science

volume 46 number 10 international

ieee spectrum 10.09


28

LET'S CHILL: Overheated chips can degrade the performance and shorten the life span of state-of-the-art electronics. New research in thermal management looks for ways to cool them down.

COVER STORY

24 INFOGLUT
E-mail, tweets, and Facebook updates are destroying our productivity and our leisure. *By Nathan Zeldes*

18 A REAL CLOUD COMPUTER
Climate scientists dearly want to track cloud systems with their models. But first, a very different sort of supercomputer must be built. *By Michael Wehner, Lenny Oliker & John Shalf*

28 CHILL OUT
Heat can be a CPU's worst enemy, but researchers are devising new materials and chip designs to keep the microprocessors cool. *By Alexander A. Balandin*

34 THE RECESSION'S SILVER LINING
The recession has not been kind to the semiconductor industry. But there may be an opportunity in this crisis. *By Pushkar Apte & George Scalise*

40 MAY THE FORCE OF LIGHT BE WITH YOU
Light exerts enough force to throw nanoscale switches, making superfast photonic signal processing possible. *By Hong X. Tang*

COVER: FREDRIK BRODÉN

THIS PAGE: PHOTO: DAN SAELINGER; STYLING: WENDY SCHELAH

UPDATE

9 THE BRAIN, UNPLUGGED
Researchers report prototype wireless neural interfaces. *By Prachi Patel*

10 THE SPASER NANOLASER

12 A REIMAGINED DATABASE

12 EXOSKELETONS ON THE MARCH

OPINION

5 SPECTRAL LINES
IE6 is the 8-track tape of Generation Web. It's time to move on—or get left behind. *By Harry Goldstein*

8 FORUM
Readers comment on the energy efficiencies of island empires and incandescent bulbs.

17 TECHNICALLY SPEAKING
Bioengineering technology is maturing, and so is its vocabulary. *By Paul McFedries*

DEPARTMENTS

3 BACK STORY
How fighting information overload turned into a full-time job.

4 CONTRIBUTORS

14 HANDS ON
Google's Street View vans don't go everywhere. Now you can make your own panoramas with a do-it-yourself camera array. *By Roy D. Ragsdale*

16 GEEK LIFE
Airlines are finally making it easy to plug in at the airport and go online in the air. *By Prachi Patel*

16 BOOKS
There are plenty of patent handbooks but none written by engineers, for engineers—until now. *Reviewed by Kirk Teska*

52 THE DATA
Mobile devices are producing exabytes of data, and more and more of it resides in the Internet cloud. *By Steven Cherry*

spectrum Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

volume 46 number 10 international

ieee spectrum 10.09

PHOTO: BRE PETTIS

View exclusive online-only content now by clicking on the headlines

WWW.SPECTRUM.IEEE.ORG

AVAILABLE 1 OCTOBER ON IEEE SPECTRUM ONLINE

WATCH THE AUTOMATED BARTENDER

The folks at NYC Resistor, a hacker collective in Brooklyn, N.Y., have come up with an automated bartender based on Arduino, a collection of open-source hardware designs and software-development tools. As DIY guru Bre Pettis shows us in a new video, the barbot takes drink orders via an LED control panel, moving a glass from one filling station to the next and adding alcohol, mixers, and bitters in succession. When finished, the bot signals the customer to imbibe with a wink of its LEDs.

ONLINE FEATURES:

EIGHT-BIT ENCORE: A number of DIYers and hardware hackers are transforming 1980s video-game systems into custom music-making machines. Chiptunes, or songs in which all sounds are synthesized in real time, were a central element of early video games. Today an entire chiptunes community has blossomed, complete with live concerts, festivals, and films. *IEEE Spectrum*'s audio slide show will guide you through the chiptune scene.

ALSO ONLINE

- Webinars
- Podcasts
- News
- Videos
- Blogs
- Jobs
- Career Accelerator
- IEEE Xplore® digital library
- Interviews
- Opinions
- More!

TECH INSIDER WEBINARS

Check out all webinars, including these below, at http://www.spectrum.ieee.org/webinar.

Episode III: Robotics and Academic Research

Getting the Most From MATLAB and Simulink With Symbolic Computation

Reducing Physical Verification Cycle Times With Debug Innovation

Windows Embedded: The Platform for Connected, Efficient Industrial Devices

WWW.IEEE.ORG/THEINSTITUTE

AVAILABLE 6 OCTOBER ON THE INSTITUTE ONLINE

JOIN A SOCIETY AND BOOST YOUR CAREER

Four members discuss how belonging to an IEEE society can be invaluable to your career.

IEEE XPLORE GETS EVEN BETTER

A mobile version of the IEEE Xplore digital library has been rolled out, and there is more content to search now that more than 250 000 applied physics articles have been added.

NEW IEEE MILESTONES

Read about the tech histories behind the Milestones that will honor the IBM Thomas J. Watson Research Center, ferrite materials, the integrated circuit, and the Speak & Spell toy, one of the earliest handheld electronic devices.

SEMICONDUCTOR SMARTS

Learn the latest in semiconductor/insulator interfaces and the physics of insulating thin films at the IEEE Semiconductor Interface Specialists Conference, being held from 3 to 5 December in Arlington, Va.

IEEE SPECTRUM (ISSN 0018-9235) is published monthly by The Institute of Electrical and Electronics Engineers, Inc. All rights reserved. © 2009 by The Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, New York, NY 10016-5997, U.S.A. The editorial content of IEEE Spectrum magazine does not represent official positions of the IEEE or its organizational units. Canadian Post International Publications Mail (Canadian Distribution) Sales Agreement No. 40013087. Return undeliverable Canadian addresses to: Circulation Department, IEEE Spectrum, Box 1051, Fort Erie, ON L2A 6C7. Cable address: ITRIPLEE. Fax: +1 212 419 7570. INTERNET: spectrum@ieee.org. ANNUAL SUBSCRIPTIONS: IEEE Members: $21.40 included in dues. Libraries/institutions: $399. POSTMASTER: Please send address changes to IEEE Spectrum, c/o Coding Department, IEEE Service Center, 445 Hoes Lane, Box 1331, Piscataway, NJ 08855. Periodicals postage paid at New York, NY, and additional mailing offices. Canadian GST #125634188. Printed at W224-N3322 Duplainville Rd., Pewaukee, WI 53072-4195, U.S.A. IEEE Spectrum circulation is audited by BPA Worldwide. IEEE Spectrum is a member of American Business Media, the Magazine Publishers of America, and the Society of National Association Publications.

Spectrum | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

back story

NATHAN ZELDES

OFF-LINE: When Nathan Zeldes isn't fighting information overload, he's collecting slide rules and other historical computing artifacts, like this Fuller Calculator.

Overlord of Overload

IN THE mid-1990s Nathan Zeldes was manager of computing productivity at Intel Corp.'s Israel site. His job, basically, was to figure out how employees could interact more effectively by using computers.

He understood how complicated it was going to be when he realized that e-mail was part of the problem, rather than the solution. For example, there was a senior manager who refused to open any messages that came with attachments. He interviewed the manager and found that hundreds of messages swamped the man every day and he simply couldn't handle the load. Zeldes interviewed other Intel managers, heard similar stories, and set about trying to identify the root causes.

He set up a pilot program within Intel's Israel operations to make e-mail less burdensome. It worked—so well, in fact, that Intel's European offices invited him to come and implement it there. And then it was rolled out to Intel offices worldwide.

The press got wind of this effort in 2001 and revealed Zeldes as the man behind the movement; within weeks, pleas for help began coming in from companies all over the world. An informal network of people started taking the idea of addressing information overload seriously. The group formalized its existence last year as the nonprofit Information Overload Research Group, with Zeldes as president. Zeldes continued to work within Intel on new types of solutions until last year, when he left the company. Today, as a consultant, he helps a variety of organizations deal with information overload.

As Senior Editor Tekla S. Perry discovered, Zeldes practices what he preaches. Thanks to good e-mail habits, Zeldes says, his e-mail traffic is down to a couple of dozen messages a day; he clears his in-box in a preset time slot every afternoon. By moving messages into "to do" folders, he makes sure his in-box never gets bigger than one screenful.

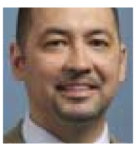
"Working with Zeldes in editing his article for this issue ("Infoglut"), I quickly learned not to expect an instantaneous response to my many e-mails," Perry says. "I simply had to be patient, and I'd inevitably hear back from him within 24 to 36 hours. Which, as it turned out, was just fine." ❑

CITING ARTICLES IN IEEE SPECTRUM

IEEE Spectrum publishes two editions. In the international edition, the abbreviation INT appears at the foot of each page. The North American edition is identified with the letters NA. Both have the same editorial content, but because of differences in advertising, page numbers may differ. In citations, you should include the issue designation. For example, the first Update page is in *IEEE Spectrum,* Vol. 46, no. 10 (INT), October 2009, p. 9, or in *IEEE Spectrum*, Vol. 46, no. 10 (NA), October 2009, p. 13.

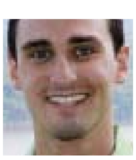
contributors

PUSHKAR APTE is vice president of technology programs and **GEORGE SCALISE** is president of the Semiconductor Industry Association. They describe the effort to push semiconductor R&D past the end of Moore's Law in "The Recession's Silver Lining" [p. 34].



ALEXANDER A. BALANDIN is chairman of materials science and engineering at the University of California, Riverside. In "Chill Out" [p. 28], he describes new materials and designs to keep chips cool. He notes that people often confuse "photons," which are quanta of light, with "phonons," which measure crystal lattice vibrations and are the focus of Balandin's research. Once, an overeager assistant painstakingly changed *phonon* to *photon* on each flyer for a lecture Balandin was giving, assuming it was a typo.

FREDRIK BRODÉN shot the cover photo and the opening image for "Infoglut" [p. 24]. On a hot summer's day in Dallas—29 °C—the model on the cover had to lie on the pavement wearing a wool sweater. "He could only lie down for about 30 seconds at a time," Brodén says. Fortunately, the @ symbol on his back wasn't much of a burden. It was carved out of Styrofoam and held in place with fishing line.



ROY D. RAGSDALE studied computer science at the United States Military Academy, in West Point, N.Y. In "DIY Street-View Camera" [p. 14], he describes how to produce panoramas like those of Google's Street View. Ragsdale will report to his first unit, in Germany, next year.

DAN SAELINGER created the icy photos for "Chill Out" [p. 28]. The ice-encased circuit board in our table of contents was one of three prepared for the shoot, because "you never know with ice," says Saelinger. "It changes quickly under the lights." It took him several days to build up layers of ice for the electronics in the freezer. He enhanced the effect with dry ice and spray-on faux frost.

HONG X. TANG won a National Science Foundation Career award this year for using the pressure of light to operate nano-mechanical devices, work he describes in "May the Force of Light Be With You" [p. 40]. The award, worth US $400 000 over five years, is meant to encourage young scientists. Tang came to Yale in 2006 as an assistant professor of electrical and mechanical engineering.

MICHAEL WEHNER, LENNY OLIKER, and **JOHN SHALF** work at Lawrence Berkeley National Laboratory. Oliker and Shalf have backgrounds in computer science and electrical engineering, as you might expect for members of a team proposing a supercomputer for modeling climate change—"A Real Cloud Computer" [p. 18]. Wehner, however, was trained in nuclear physics and designed bombs before switching to climate research. Their supercomputer idea is now being taken seriously, but when Wehner first proposed it at a 2008 conference, even he felt it was fanciful. "I was prepared to be laughed off the stage," he recalls.

CLOCKWISE FROM TOP RIGHT: ADAM HARVEY; LIN FENG; REED RAGSDALE; STEVEN CHILDERS; UNIVERSITY OF CALIFORNIA, RIVERSIDE spectral lines

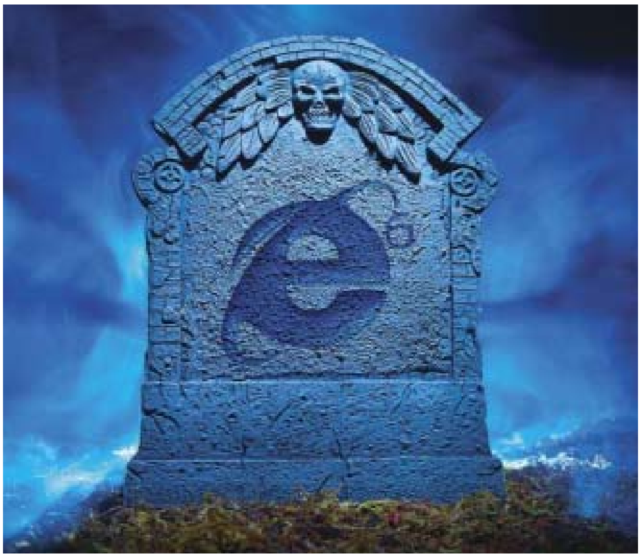


PHOTO-ILLUSTRATION: RANDI SILBERMAN KLETT; TOMBSTONE: DON FARRALL/GETTY IMAGES

We Come to Bury IE6

SCIENTISTS, programmers, academics, brokers, journalists, and countless other knowledge workers spend a good portion of their time on the Web, gathering and analyzing information, tracking down sources, and sifting through data. Engineers are no exception. According to a recent study by market research firm Outsell, in Burlingame, Calif., engineers spend almost a quarter of their time working with external information, much of it found using vendor sites, search engines like Google, or info troves like the IEEE Xplore digital library.

No tech company would expect its engineers to hunt down that information using an 8-year-old computer. Yet some let their workers limp along with an 8-year-old browser that has inspired a movement dedicated to taking it behind the proverbial shed and putting it down. I refer to Microsoft's Internet Explorer 6, which will be virtually unusable within a couple of years, possibly much sooner.

While sites like YouTube, Facebook, and Digg inch closer to dropping support for IE6 and Microsoft itself urges customers to upgrade, a coalition of Web sites organized under the moniker IE6 No More (http://ie6nomore.com) has already decided not to waste any more effort on the ancient browser: "As any web developer will tell you, working with IE6 is one of the most difficult and frustrating things they have to deal with on a daily basis, taking up a disproportionate amount of their time. Beyond that, IE6's support for modern web standards is very lacking...." Or as *IEEE Spectrum* Twitter follower The GT put it: "Loathe IE6 with a passion. As a web developer, too much time & $ wasted on IE6 hacks & security precautions. Bad for Web 2.0."

Included as part of Microsoft Windows XP in 2001, IE6 instantly attained dominance—as much as 80 percent of the browser market. That share has steadily diminished over the years, but IE6 still has anywhere from a 10 to 25 percent market share, depending on whose statistics you believe.

Many Web users who have not installed newer browsers want to upgrade but can't. They are employees of large corporations whose IT departments support only IE6. The social bookmarking site Digg asked its users how many of them using IE6 are doing so because they prefer it. The answer: 7 percent. A whopping 70 percent said they're saddled with IE6 because their companies won't let them upgrade.

Lots of IE6 visitors to *Spectrum*'s Web site are probably in the same boat. Recently, an IEEE member called me to ask why articles he'd printed out at work weren't formatted correctly. He was using IE6, for which we have done some, but not all, of the coding on our site necessary to accommodate it. When I asked him if he could upgrade to any other browser, he told me that his company, a major U.S. utility, doesn't allow employees to download software of any sort on their PCs because of security concerns.

The irony is that IE6, because of its ubiquity and longevity, is a favorite target for malicious hackers. The Copenhagen-based security firm Secunia has reported 158 security vulnerabilities for IE6 since its release. Of the 13 IE6 advisories Secunia issued in 2008, one-fourth involve the exposure of sensitive information or ways to bypass security.

Companies whose employees rely on an insecure, outmoded browser need to rethink their reluctance to upgrade, and fast.

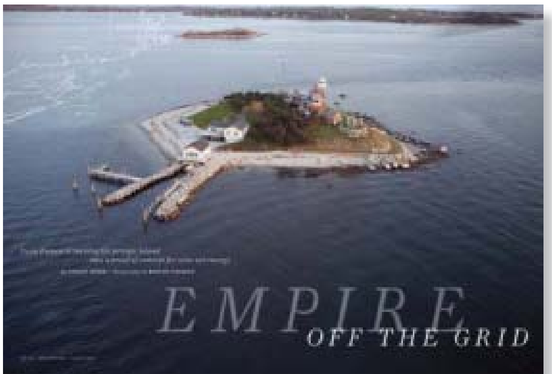
Even as people increasingly rely on the Web to do their jobs, big changes to the Web itself are in the works. As *Spectrum* has recently reported, Microsoft and Google are both radically revamping their browsers to protect both browsers and operating systems from attack, facilitating a more secure and fluid exchange of applications and data between the Web and the desktop. And as Ben Parr points out on the social-networking blog Mashable, every browser except IE6 is now—or soon will be—benefiting from the new HTML 5 Web standard.

HTML 5 will make it easy to run Web applications on your desktop even when you're off-line. It will make it a simple matter to move files from page to page, including to and from Web apps. Embedding video directly into Web pages will be as simple as embedding images is today. Android, BlackBerry, Safari, and other mobile browsers will particularly benefit from HTML 5 and its geolocation capabilities. Chrome, Firefox, and Opera already support some elements of the HTML 5 standard.

But IE6 won't. Ever. Add that fact to the growing number of Web sites that aren't going to optimize for IE6, and that puts workers whose companies don't upgrade their browsers now at an ever-growing disadvantage.

—HARRY GOLDSTEIN

forum



LETTERS do not represent opinions of IEEE. Short, concise letters are preferred. They may be edited for space and clarity. To post your comments online, go to http://www.spectrum.ieee.org. Or write to Forum, *IEEE Spectrum*, 3 Park Ave., 17th Floor, New York, NY 10016-5997, U.S.A.; fax, +1 212 419 7570; e-mail, n.hantman@ieee.org.

ONE NATION, UNDER KAMEN

I ENJOYED THE article describing Dean Kamen's electrical system for his private island ["Empire Off the Grid," August], but I think it's a bit misleading to call the island "energy independent." While the technology is impressive, it does not provide enough power for winter heating, and it does not address the energy needs for transportation to the island or for manufacturing and transporting everything required to live there. The location also offers some significant advantages not available to most, such as unobstructed wind and a sizable area of land that can be devoted to energy production. That's not to belittle what Kamen has achieved, but these issues do suggest a worthy objective for an alternative-energy prize: building a truly energy-independent community.

Such a community would have the density and location of a typical urban or suburban community and produce enough power to meet all its internal needs. It would also need to produce enough power or energy-intensive goods for "export" in sufficient quantity to balance over time the energy required for making and transporting anything the community had to "import." These are the challenges we must meet if we are to truly gain independence from fossil fuels and other unsustainable sources.

WARREN MONTGOMERY
IEEE Member
DeKalb, Ill.

SHEDDING LIGHT

I'VE JUST noted what appears to be a discrepancy in the August 2009 issue. Referring to incandescent bulbs in "The LED's Dark Secret," author Richard Stevenson says, "They still waste 90 percent of their power, delivering roughly 16 lumens per watt." This statement implies that incandescents are around 10 percent efficient. In "Empire Off the Grid," Sally Adee states that "a traditional 60-watt incandescent bulb lasts about 1500 hours and has only a 2.5 percent efficiency because it wastes most of its energy as heat."

So which is it? Is a typical incandescent bulb 2.5 or 10 percent efficient in converting electricity into light? Perhaps you can shed some light on this discrepancy.

ERIC SLATER
IEEE Life Member
Long Beach, Calif.

Senior Editor Philip E. Ross and Associate Editor Sally Adee respond: By the broadest possible measure, an incandescent bulb can be said to waste 90 percent of its energy as heat. However, because much of the remainder radiates from the filament in wavelengths that either do not reach the human eye or cannot be fully detected by it, the effective efficiency is much lower than Stevenson's statement implies—around 2.5 percent, or 3.5 percent for halogen lamps.

PAPER PEOPLE

WITH REGARD to "The Kindlers Are Coming to Get Me" [Spectral Lines, August], I agree completely with the author's preference for paper books over e-books. I, too, spend all day in front of a screen, so reading from a screen is not my idea of relaxation. There is something about hard copy, whether a newspaper or a book, that allows me to unwind. Electronic copies have their place, but they cannot replace hard copy when it comes to reading for pleasure.

DAVID LEBLANC
IEEE Member
Fredericton, N.B., Canada

ENERGY ON RESERVE

WHAT HAPPENS when the wind doesn't blow?" asks Peter Fairley in "Germany's Green-Energy Gap" [July]. Do the German Alps have regions where water could be stored at high elevations? In 1971, I worked on the construction of a project to draw water from the Thompson River, in British Columbia, Canada, and elevate it more than a kilometer for use by concentrator mills at mines in Highland Valley.

Pumping water uphill is an established technology. So is hydroelectric generation. Wind, tidal, and solar power generated at peak output periods could be used to pump water into high-elevation reservoirs, storing energy for hydroelectric generation at night or when the wind isn't blowing.

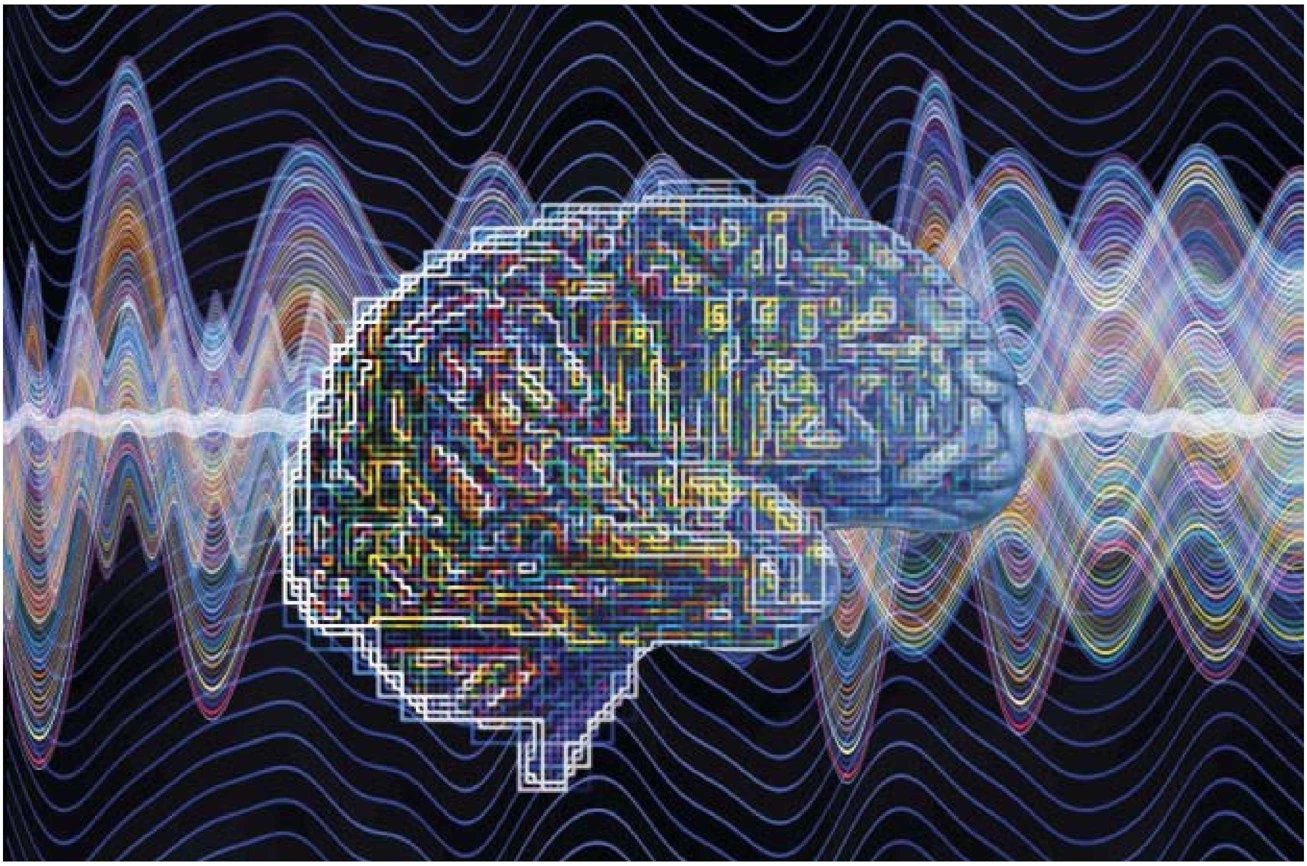
KELLY MANNING
IEEE Member
Victoria, B.C., Canada

ARCHIVES AVAILABLE

OVER 40 years' worth of copies of *IEEE Spectrum* are available, for the cost of shipping, from Mrs. John M. Blivin, 840 Denecourt Ave., Jackson, MI 49202-2916. The issues were collected by her late husband, John M. Blivin, an IEEE Life Member.

update

more online at www.spectrum.ieee.org



WIRELESS FOR WETWARE: Wireless neural interfaces would be safer and allow more of the brain to be monitored.
ILLUSTRATION: MEHAU KULYK/GETTY IMAGES

The Brain-Machine Interface, Unplugged

Researchers report prototype wireless neural interfaces

IN EXPERIMENTS and even limited human clinical trials, electrode arrays implanted on the brain's surface have given monkeys and humans the ability to move objects with their thoughts. The experiments are proof that brain-computer interfaces could improve the lives of severely paralyzed people. But these systems rely on wires snaking out from the skull, which would affect a person's mobility and leave an opening in the scalp prone to infection.

Wireless brain-machine interfaces would be much more practical and could be implanted in several different areas of the brain to tap into more neurons. A typical scheme would have electrodes penetrating brain tissue, picking up neuronal electrical impulses, called spikes. A chip would amplify and process the signals and transmit them over a broadband RF connection through the skull to a receiver. Then, just as in wired systems, algorithms would decode these signals into commands for operating a computer or a robot.

The key requirement for such a system is that it consume very little power to keep the heat down. "Most of the guidelines for implantable devices say that you should not raise the surrounding tissue temperature by more than 1 °C; otherwise, you'll kill the cells you're trying to record from," says Reid Harrison, an electrical and computer engineering professor at the University of Utah, in Salt Lake City.

Sending the complex analog impulses as they are would take up too much bandwidth. So it will

update

be necessary to convert them into a simpler, robust form as close as possible to that of the neurons, says Brown University neuroengineer Arto Nurmikko. He and some of his colleagues were associated with the now-defunct Foxborough, Mass., start-up Cyberkinetics Neurotechnology Systems, which did the first human clinical trials of an implanted brain-computer interface. Now his team has a promising wireless interface scheme, which they presented last month at the IEEE Engineering in Medicine and Biology Conference (EMBC).

The researchers have implanted a 16-electrode microchip into the arm-control region of a monkey's brain. The chip amplifies neural signals and sends them via a flexible wire ribbon through the skull to another chip beneath the scalp. Here, the signals are digitized, and a diode laser transmits the data as infrared light pulses through the skin to a receiver. The system uses 12 milliwatts of power in its current configuration; expanding it to a 100-electrode scheme, which would allow for controlling a prosthetic, would require 30 mW, the safety limit for cranial use.

Working along similar lines, Utah's Harrison, together with engineers and neuroscientists at Stanford, has designed an 8-mW radio chip with 100 amplifiers that could be connected to a 100-electrode array. A radio transmitter sends data at 350 kilobits per second but only over a distance of 5 centimeters. Increasing the range to 4 meters shoots the power needs past 63 mW. To stay within the transmitter's low RF bandwidth, Harrison designed the amplifiers to send signals only when a neuronal spike exceeds a certain voltage threshold. He is now working on extracting other neuronal spike features, such as

BRAIN PORTS: These are various versions of a wireless neural interface by University of Utah researchers.
PHOTOS: REID HARRISON/UNIVERSITY OF UTAH

width or height, to be able to distinguish between different neurons.

The implant in each of these schemes is powered by an electromagnetic induction coil outside the skull. A cap or headband containing a coil no more than a few centimeters wide could send power to the device. An external power source for neural chips makes more sense than batteries, Harrison says, because the chips consume a hundred times as much current as pacemakers do. This means their batteries would need to be replaced much more often than a pacemaker battery, whose typical life span is seven years.

Wireless neural implants open up the possibility of embedding multiple chips in the brain, enabling them to read more and different types of neurons and allowing more complicated thoughts to be converted into action. Thus, for example, a person with a paralyzed arm might be able to play sports. "When we hit a tennis ball or kick a soccer ball, we plan things first...and then execute an action based on input," Nurmikko says. "The brain is furiously calculating what it's going to do with this thing that's coming at you."

Eventually, you would want to listen in on hundreds or even thousands of neurons. But then infrared or RF transmission bandwidth would be a constraint, observes Babak Ziaie, an electrical and computer engineering professor at Purdue University, in West Lafayette, Indiana. At the EMBC meeting, Ziaie presented an optical approach: An LED array, possibly attached to the skull, could convert the electrodes' signals into light pulses that are captured by a high-speed camera chip and reconverted into electrical signals. He plans to test the scheme on animals this fall.

—Prachi Patel

The Spaser Nanolaser

Surface plasmon resonance lasers are the smallest nanolasers yet

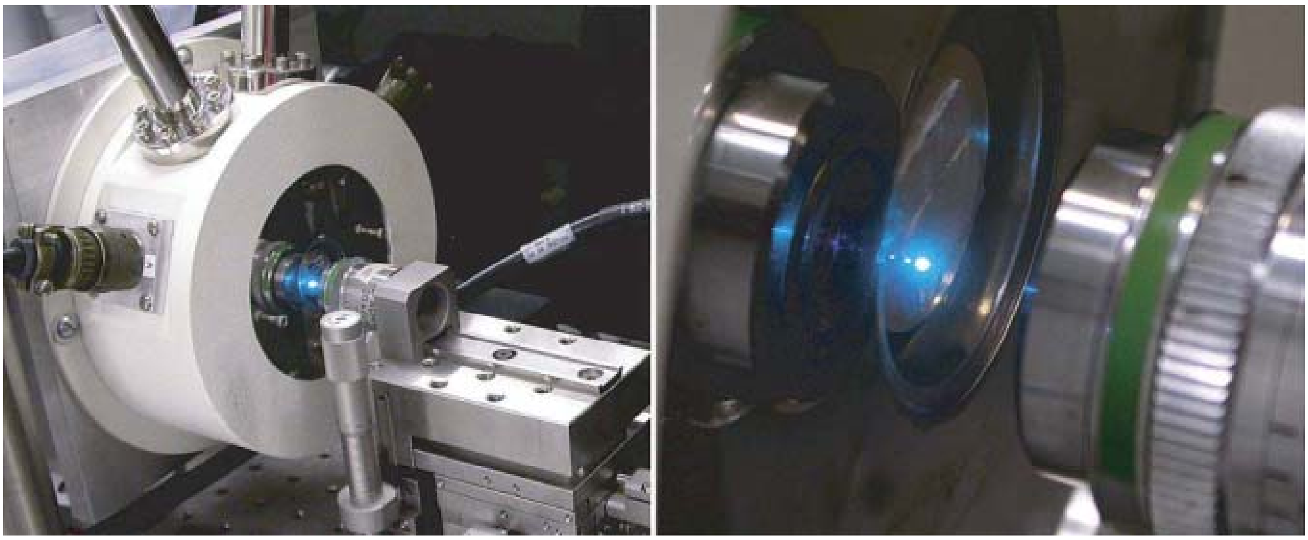
In the quest to make computer processors smaller and faster, computing with light instead of relatively slow electrons has long been a tantalizing goal. One roadblock has been the inability to make lasers tiny enough so that several thousand of them could fit easily on a chip. In late August, two groups of researchers reported the construction of a new kind of nanometer-scale laser. Surface plasmon resonance nanolasers, or spasers, are the smallest lasers yet made, and their creators say the devices could pave the way toward ultrafast optical computing.

Spasers can "bridge the worlds of electronics and optics at truly molecular-length scales," says mechanical engineering professor Xiang Zhang, of the University of California, Berkeley, who led one of the groups with research associate Rupert Oulton.

Surface plasmons are oscillations of electrons that form at the junction between an insulator and a metal. Metals have lots of free electrons, which oscillate when light shines on them. But when the metal borders a dielectric, the movement

"We're sorry, you deserved so much better."

—Gordon Brown, prime minister of the UK, apologizing for the persecution of famed computer scientist Alan Turing



of the electrons is curtailed because they cannot enter the dielectric. That forces the electrons to move in waves of density along the junction.

Significantly, light will produce plasmons even in structures much smaller than the light's wavelength. It was this phenomenon that the two research groups exploited in creating their nanoscale lasers.

Mark Stockman of Georgia State University, in Atlanta, and David Bergman of Tel Aviv University first proposed the mechanism of a spaser in 2003. The idea was to build a device that operates similarly to a laser to generate and amplify surface plasmons. In a laser, light reflects back and forth through a special material called a gain medium, stimulating the emission of more light of the same phase with each pass. In a spaser, it is the waves of electrons that are amplified, which are then converted to light.

A team led by Mikhail Noginov of Norfolk State University, in Virginia, including researchers at Purdue University, in Indiana, and Cornell University, in New York, has demonstrated the spaser in gold nanoparticles surrounded by a silica shell that has been impregnated with green dye. When light hits the nanoparticles, plasmons form at the gold-silica interface. The plasmons are amplified by their interaction with the dye-filled shell, which acts as the gain medium, and laser light emerges. The nanoparticle system is just 44 nanometers in diameter, less than one-tenth the size of the 530-nm wavelength light emitted by it.

Noginov's new laser relies on organic dye molecules, which would make it difficult to incorporate into existing technology. Future work may involve creating a spaser-based nanolaser driven by an electrical source instead of a light source, which would make the lasers more practical for computer and electronics applications, says Noginov.

The UC Berkeley researchers led by Zhang and Oulton have created a spaser using a different approach, which they call hybrid plasmonics. "We have found a way to tightly confine light and at the same time sustain it for a long time in a very small space," says Oulton.

Their team's spaser consists of a cadmium sulfide semiconductor nanowire on top of a silver substrate. The nanowire and silver are separated by a tiny magnesium fluoride insulating layer that's 5 nm thick. A laser feeds light into the semiconductor nanowire, which acts like a waveguide and transmits the light's electromagnetic modes—harmonic frequencies of laser light that are not limited by diffraction. The plasmons that form on the silver substrate couple with the modes in the waveguide and generate laser light within the magnesium fluoride gap. Despite the gap's puny 5-nm dimensions—about the size of a single protein molecule—the wavelength of laser light emitted is 489 nm.

LIGHT FANTASTIC: A spaser, one of the smallest lasers ever made, shines brightly.
PHOTOS: XIANG ZHANG LAB/UC BERKELEY

The spaser approach "looks wonderful," says Dick Slusher, a prominent laser physicist not involved with either team, who is currently the director of the Georgia Tech Quantum Institute.

Though it will be a long time before spasers are powering optical computer chips, there are many intermediate steps that could be useful. "We would like to use plasmon lasers to do something you cannot do with conventional lasers," such as analyzing individual molecules, says Oulton.

—Saswato R. Das

75 °C The temperature engineers incorrectly figured the Chandrayaan-1 orbiter would experience 100 kilometers from the moon, according to *The Times of India*. Because of the miscalculation, India's first lunar mission overheated and was shut down in August.

update

Ingres and VectorWise Claim Database Speedup

Firms rebuild database functions from scratch

To hear those in the know tell it, database management systems leave a lot to be desired: They require too much hardware, and they make poor use of it. Some advocate a shift in the type of hardware used—away from the CPU toward graphics-processing units [see "Data Monster," September 2009]. But open-source database firm Ingres, of Redwood City, Calif., and start-up VectorWise, of Amsterdam, see the answer in a better use of the CPU.

Their prototype software has shown more than a 10-fold improvement in performance and an 80-fold improvement in some tasks.

To get such an improvement, database luminary Peter Boncz and others at the Dutch national math and computer science research institute, Centrum Wiskunde und Informatica (CWI), took a close look at how modern CPUs work and used what they found to make a database system from scratch. They formed VectorWise in 2008 and joined forces with Ingres and Intel this year.

Database systems today are written "for the machine of 20 years ago," says Bill Maimone, Ingres's chief technology officer. They can't easily take advantage of a modern processor's ability to perform a single instruction on a large set of data, and they're at the mercy of the relatively slow movement of data on and off the CPU.

To solve these problems, CWI computer scientists came up with versions of database operations that work on sets of 100 to 1000 values, or vectors, instead of on one database value at a time. As a result, some operations that take tens or hundreds of CPU clock cycles in other databases take just a handful in the VectorWise system.

The scientists also constructed the system so that all the work is done on data in the CPU's cache, where the processor cores can quickly get at it, instead of in main memory, which can take hundreds of clock cycles to fetch. This required them to compress the data in some parts of the cache and come up with fast decompression algorithms so that the process of fetching data didn't bog down.

Both tasks were helped by VectorWise's use of a database storage scheme called column-store. Data is sent from storage to the CPU as strings of values from the same attribute domain—for instance, a list consisting only of salaries rather than a record containing employee names, salaries, and other data, explains Daniel Abadi, an assistant professor of computer science at Yale University. Column-store makes it easier to perform vector calculations because all the needed values are stored contiguously. Column-store data is also easier to compress, he notes, because it has more inherent order to it.

Abadi calls VectorWise "a company to watch," but its software won't be for everyone. One limitation, he points out, is that the new system is designed to run on a single machine with a database of less than 10 terabytes. That would rule out most databases used by big retail firms.

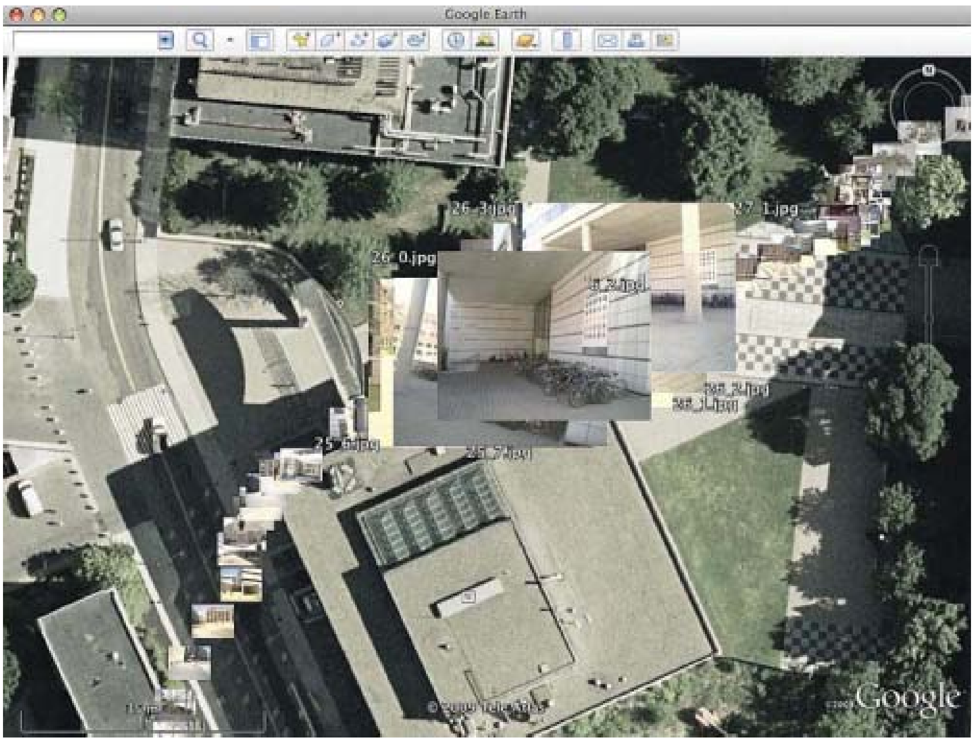
—Samuel K. Moore

YOSHIKAZU TSUNO/AFP/GETTY IMAGES

On the March

An army of exoskeletons is coming. Cyberdyne, the Tsukuba, Japan–based maker of robotic exoskeletons, is making a big push to increase the supply. Yoshiyuki Sankai, Cyberdyne's CEO, says this month the company expects to ship 80 to 90 exoskeletons in Japan, where there are now just 20 in use. The robotic suits are used mostly to help rehabilitate patients by enhancing the strength of limbs that have been weakened by injury or disease. Sankai says the company will be introducing a new version of its exoskeleton this fall at an event in Kyoto. See http://www.spectrum.ieee.org/robotics/medical-robots/exoskeletons-are-on-the-march.

hands on



OFF THE GRID: A do-it-yourself camera array lets you create your own street views of places where Google's cameras don't go.

ALL IMAGES: ROY D. RAGSDALE

DIY STREET-VIEW CAMERA

Create Google Street View–like panoramas with cheap webcams and open-source software

IF YOU use Google Maps, you're probably familiar with its Street View feature, which shows actual ground-level photos of many cities around the world. Google creates the images by mounting special cameras on vehicles and driving them around.

Now wouldn't it be great if you could have your own Street View–like camera? You could hike a trail and later share the photos with friends. The photos would carry GPS tags, so you could display them on Google Earth and include annotations—good water here, poison ivy there. Realtors could display whole neighborhoods to potential clients. A country club could offer a virtual tour of its golf course. Architects could monitor progress at a construction site.

Last year, as part of a "disruptive technologies" course at the United States Military Academy, in West Point, N.Y., I set out to develop a prototype. I thought such a system would have many applications in the battlefield, for example, helping soldiers patrol dangerous routes. My system—I call it PhotoTrail—uses off-the-shelf components and open-source software. It consists of webcams, a GPS receiver, a notebook computer, and imaging software.

For the camera system, I chose the Microsoft LifeCam NX-6000, which is small and has UVC (USB video class) compatibility. It was also cheap (although it lists for US $79.95, I got it for $25 new). It has a megapixel video resolution and shoots 8-megapixel still images.

The NX-6000 has a lens with a 71-degree field of view. In order to stitch images together for 360-degree panoramas, I bought eight units, for a total of 568 degrees of coverage, allowing a healthy image overlap. To connect all the cameras to the notebook, I used two D-Link USB hubs ($25 each), which ran unpowered.

For the GPS receiver, I chose the GlobalSat BU-353, a self-contained waterproof device with good signal reception and accuracy, which costs a mere $37. If you attach it to a USB port, the GPS coordinates will appear in a log file, using a standard GPS encoding scheme.

Construction was straightforward. On a flat octagonal heavy-cardboard base, I glued small posts for the cameras' clips to latch onto. I aligned each unit and then placed the USB hubs and the GPS receiver in the middle. I secured the cables with Velcro and sandwiched everything with another piece of cardboard. The whole thing's the size of a small pizza box, weighing less than 1 kilogram. Excluding the notebook (a 2-gigahertz machine with 512 megabytes of RAM running Ubuntu Linux), the hardware cost about $300.

To start capturing images, I installed a UVC driver and a device driver compatible with the camera array. For the capture itself, I used luvcview, a small open-source webcam program by Logitech. (Uvccapture, also by Logitech, lets you take still shots, but it was incompatible with this camera.)

I had set the camera array on video capture, so I needed to tweak luvcview's source code to get still images from the video feed. The tweaks call for the array to capture a few frames and then stabilize itself so that the images are in focus and have good light contrast. I wrote a Python script to capture the eight 1280-by-1024 JPEG files. That capture takes about 8 seconds. Images captured within that time frame can be considered a single cluster to be stitched together.

Digital cameras normally add data about the photograph, but because luvcview operates at the file level, these images have no such metadata. So I wrote a Python script to read the date and time the file was created. I then used Exiftool, a command-line image metadata editor, to put the date and time into the file.

The images also need to be GPS-tagged. Gpicsync, an open-source tool, can automatically get the latitude and longitude data from a GPS receiver's log and add the coordinates to the image's metadata field. Gpicsync also lets you transform this image set into a single file that you can view using Google Earth.

I used two tools to generate panoramas. The first, autopano-sift, identifies common features in different images and aligns them along a horizon line. Another tool, hugin, uses those common elements to effectively stitch the images into a single panorama. I again used gpicsync to GPS-tag the panorama and generate a Google Earth file. To see the panorama as a 360-degree image and zoom in and pan about the scene, you can use PTViewer.

JEEP CAM: Mounted on a moving vehicle, the camera array [detail] can capture images of streets and their surroundings—just as Google does to produce its Street View panoramas.

On my underpowered computer, it took 15 minutes to stitch each panorama. It's a long time. But you can do the capture first and the stitching later, or transmit the images to a more powerful server for remote processing.

With all this development work done, it was time to test the prototype. During a trip to the Boston area, I walked around the MIT campus holding the system above my head. Passersby didn't seem bothered. I guess students attached to weird contraptions are a common sight there. On Google Earth, I can retrace my route and see the surroundings with great detail [see photo, opposite page].

I also mounted the array on a Jeep [see photos above] and drove around West Point, capturing images while driving up to 100 kilometers per hour. I programmed it to take one set of images every 20 seconds. In an hour I had 300 MB of data from 180 sets of images. Unless the Jeep is stationary, the images can't be clustered into panoramas. (Recall that it takes 8 seconds to grab a single set.) Still, the individual images are perfectly clear and on a par with those available on Google Street View.

I'm now working on some improvements. One idea is to replace the notebook with a smaller computer, such as one based on the Pico-ITX board, and shrink the camera system (the actual CCD, or charge-coupled device, and lens elements are no bigger than a fingernail). Eventually, you could build a camera system small enough to be integrated into a headband or hat.

The software could use some tweaks as well. I'm planning to write an Adobe Flash application to allow the user to see the panoramas as 360-degree images and be able to navigate from one panorama to another, just as in Google Street View.

The U.S. Army is currently evaluating my prototype. Eventually, a contractor could produce a field version for tests. Meanwhile, as this article goes to print, I'm preparing to travel far and wide. If I have space in my backpack, I'll have the camera capturing my journey, step by step. —Roy D. Ragsdale

TOOLBOX FOR BUILDING PHOTOTRAIL

HARDWARE

Microsoft LifeCam NX-6000 camera (8)
D-Link USB hub (2)
GlobalSat BU-353 GPS receiver
Laptop running Ubuntu Linux

SOFTWARE

UVC driver: *http://linux-uvc.berlios.de*
Webcam device driver: *http://linuxtv.org*
luvcview: *http://www.quickcamteam.net*
Exiftool: *http://www.sno.phy.queensu.ca/~phil/exiftool*
gpicsync: *http://code.google.com/p/gpicsync*
autopano-sift: *http://user.cs.tu-berlin.de/~nowozin/autopano-sift*
hugin: *http://hugin.sourceforge.net*
PTViewer: *http://www.fsoft.it/panorama/ptviewer.htm*

geek life

THE WELL-CONNECTED TRAVELER

More and more airports and airplanes offer Wi-Fi and power

CALL IT the airport electric-outlet slide: You stoop, crane your neck, do a quarter turn, and repeat until you find an open power socket. Then you make a mad dash to get to it before someone else.

Broadband Internet and plug-in power aren't luxuries anymore. In fact, road warriors apparently care more about providing sustenance for their laptops than for themselves. Recently, when American Airlines queried frequent fliers on the airport amenities that were most important to them, Wi-Fi came at the top of the list, and the availability of power came second.

Ironically, sometimes power is the harder one to provide. Airports have been quick to add Wi-Fi, and more than 200 of them offer it for free, according to Wififreespot.com. But new electrical outlets can cost thousands apiece, so airports usually add them slowly, during renovations, if at all. Even when there are outlets, it can be hard to find them. Jeff Sandquist, a Microsoft employee, started a wiki called AirPower to list them. The page took off within days, and its readers regularly supplement it.

Airlines are starting to provide power themselves—at least on board, if not in the airport. Tiny San Francisco–based Virgin America Airlines offers power at every seat; two of the largest U.S. airlines, Delta and American, have outlets at each first- and business-class seat and in a few rows in coach.

Wireless access is coming as well. British Airways, Portugal's TAP, and Ireland's Ryanair already offer mobile service from provider OnAir, based in Geneva. OnAir uses GSM and Wi-Fi to let travelers make phone calls, send text messages, and connect to the Internet. Other airlines planning to offer OnAir include Hong Kong Airlines, Egyptair, and Qatar Airways.

In the United States, where it's against regulations to use cellphones in flight, Virgin America already offers Wi-Fi on every plane, and Delta hopes to have Internet access on more than 300 airplanes by the end of this year. American plans to have 318 planes equipped with Wi-Fi, a little over half its fleet, by 2010.

In today's ultraconnected world, you're expected to read e-mail and answer business calls even on a tropical island vacation, but at least everyone understands you'll be off-line while in transit. Soon, even that little respite will be gone. —PRACHI PATEL

books

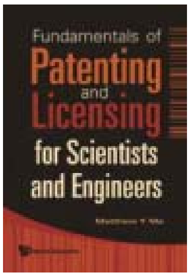
A New Patent Perspective

A book about patents by an engineer, for engineers

No doubt about it, engineers and laypersons view the world differently. And patent attorneys view it differently from both those groups. Matthew Y. Ma, author of *Fundamentals of Patenting and Licensing for Scientists and Engineers*, is a senior member of IEEE with years of experience as an engineer and scientist. The result is a book about patents for scientists and engineers by one of their peers.

I am in favor of anything that educates engineers about the world of patents. Engineers who are knowledgeable about patenting make my job as a patent attorney easier, and bringing them up to speed was the goal of my own book (*Patent Savvy for Managers,* Nolo Press, 2007). Ma even uses several of the same patent examples in his book that I used in mine, including the famous—or infamous, depending on your point of view—Amazon "one-click" patent.

As its name suggests, *Fundamentals* covers nearly everything under the patent sun, including a chapter on innovation harvesting, which discusses various scenarios in which you should and should not undertake patent protection.



Fundamentals of Patenting and Licensing for Scientists and Engineers
By Matthew Y. Ma; World Scientific, 2009; 292 pp.; US $68; ISBN 978-981-283-420-1

The book could stand more rigorous copyediting, and readers may occasionally find themselves lost in a thicket of quoted statutes and government regulations. But such minuses are far outweighed by the clear, concise fashion in which Ma explains concepts that engineers—and even my law school students—may find difficult, such as dominant patents.

Add this book to your library. In today's complex and commercial landscape, engineers and scientists who refuse to become patent savvy will likely be eclipsed by those who are. —KIRK TESKA

VIRGIN AMERICA

technically speaking BY PAUL McFEDRIES

Hacking DNA

But the real challenge will start when we enter the synthetic biology phase of research.... This would be a field with...hardly any limitations to building..."synthetic" organisms, like a "new better mouse."

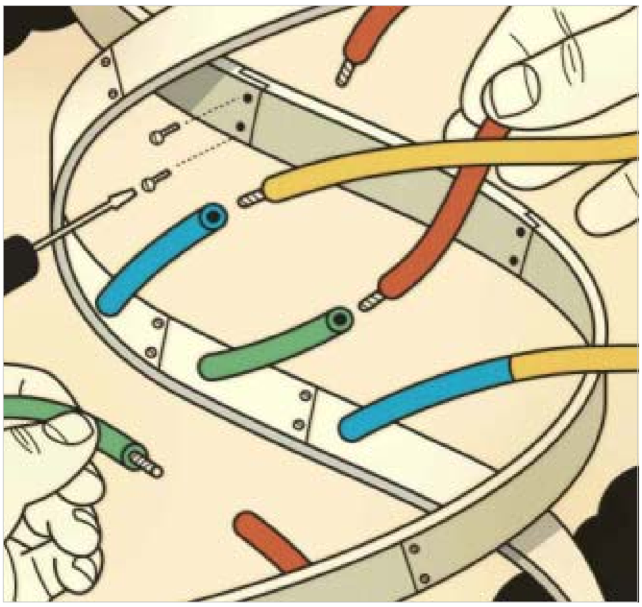
—Waclaw Szybalski, "In Vivo and In Vitro Initiation of Transcription," 1974

AN OXYMORON is a phrase that combines two contradictory words, such as "jumbo shrimp" and "deafening silence." Appropriately, the word *oxymoron* is itself an oxymoron, from the Greek *oxymōros,* which means "pointedly foolish," but the roots of the word are *oxys,* "sharp," and *mōros,* "dull."

A relatively new oxymoron is *synthetic biology,* coined by the geneticist Waclaw Szybalski in 1974. Synthetic biology (also called **synbio**) uses engineering methods to produce something new by treating a living system not so much as a biological entity but as a kind of technology. Hence synthetic biology is also called *biological engineering* or just *bioengineering.*

If a synthetic biologist focuses on genetic modifications using custom-built genes instead of naturally occurring ones, then he or she is practicing **synthetic genomics**. A biological system augmented with nonnatural components is a **semiotic system**.

Engineers require standardized parts, and **bioneers** are no different. BioBrick biological parts were designed at MIT by Tom Knight and his colleagues to be incorporated into living cells for the construction of new biological systems. The term *BioBricks* was trademarked in 2006 by the BioBricks Foundation, and now **biobricks** is commonly used as a synonym for **bioparts**, which consist of DNA sequences that have a standard structure and perform a well-defined function. The most basic is the **biobrick part**, which encodes a standard biological function and acts as a kind of biological building block; next in the hierarchy is the **biobrick device**, which consists of several biobrick parts that together perform some function (defined not by nature, but by the engineer);



HARRY CAMPBELL

finally, there's the **biobrick system**, a collection of biobrick parts and devices that perform some high-level function. My favorite example is the wonderfully named **repressilator**, created by the physicist Michael B. Elowitz and the geneticist Stanislas Leibler. This biobrick system contains three genes connected in a feedback loop in which each gene represses the next at regular intervals. The result is the biological equivalent of an electrical oscillator.

Biobricks require host cells to ride in, often called the **chassis** or the **hardware**. The host role is often filled by the famous *E. coli* bacterium. But *E. coli*, yeast, and other natural cells are complex and can interfere with whatever function the inserted **genetic device** is trying to perform. So scientists sometimes use as a chassis a **minimal cell**, which has the minimum number of components to allow the cell to support the synthetic system. Other bioneers prefer the **cell-free** approach, where the synthetic biological system rests in a bath of biochemical components in which the device can do its thing.

Today synthetic biology is the domain of big-time scientists, but biobrick parts will usher in a new era of **DIY biology**, where **biohackers** and other members of the **open-source biology** movement mess with DNA and build new biological systems biobrick by biobrick.

What new **genetic machines** will we see with all this molecular tinkering? Some examples include advanced **biosensors** engineered to detect diseases; **biofuels** derived from *biomass* (like sugarcane); the **biofabrication** of scaffolds used in tissue engineering; and **biodrugs** that are cheaper and more powerful than existing medications.

Synthetic biology will surely also generate a new category of evildoer—the *bioterrorist* who will try to **biohack** nasty pathogens that will wreak havoc on society. Actually, that era may already be upon us with the recent announcement from the firm Ginkgo BioWorks of the BioBrick Assembly Kit (see http://ginkgobioworks.com/biobrickassemblykit.html), which for a mere US $235 gives you a collection of biobrick parts and all the equipment you need to break into **biohackerdom**.

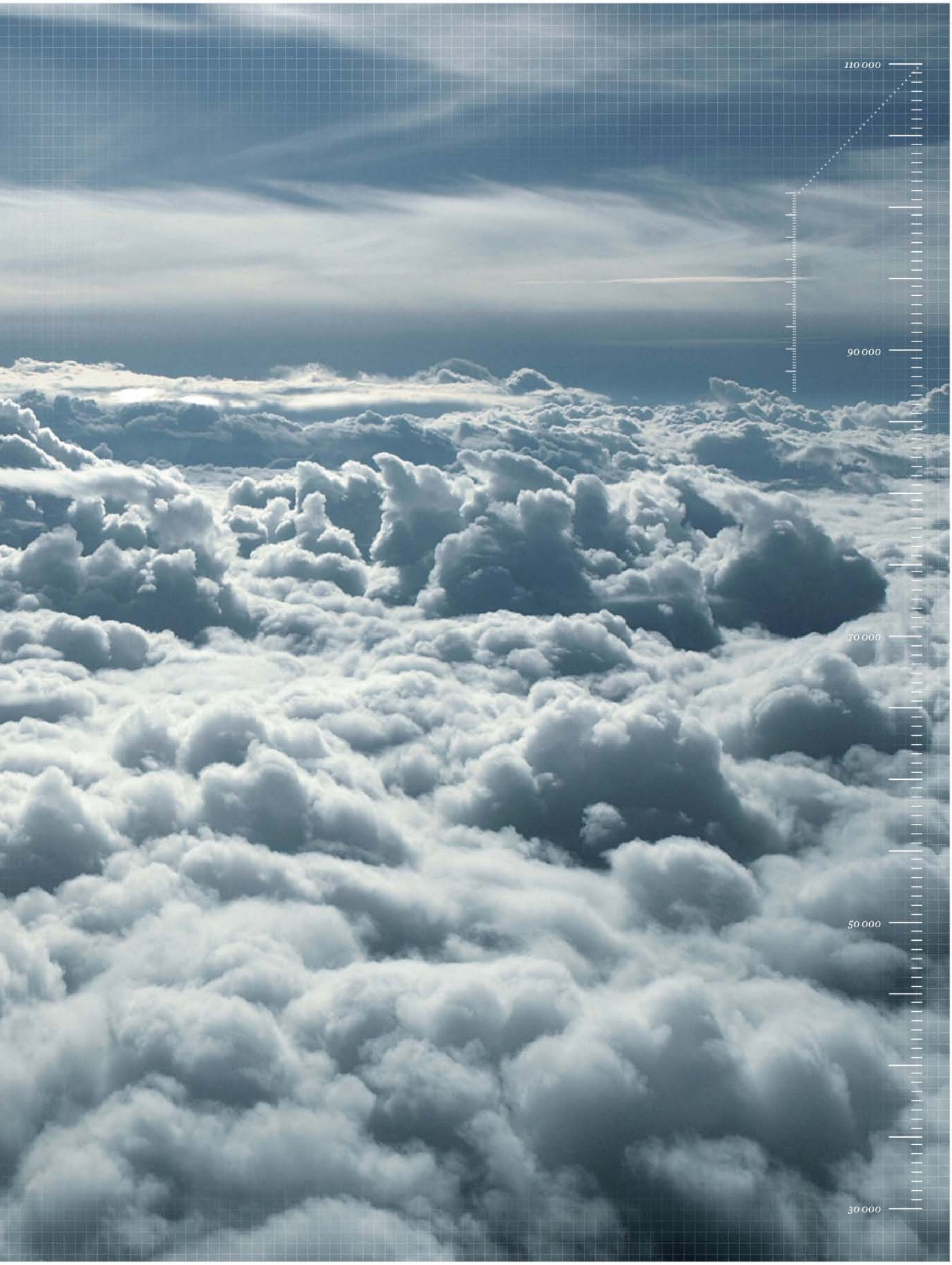
While the fearsome possibilities of **dark biology** aren't to be dismissed—and indeed synthetic biology as a whole must leap a number of ethics and security hurdles before it becomes the megabuck industry that everyone predicts—I can't wait for the day we finally see that "new better mouse." ❏

A Real Cloud Computer

A NEW CLIMATE-MODELING SUPERCOMPUTER WILL USE THE PROCESSORS NOW FOUND IN CELLPHONES AND OTHER PORTABLE DEVICES

BY MICHAEL WEHNER, LENNY OLIKER & JOHN SHALF

ieee spectrum Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags



ieee spectrum Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

ATTEMPTS TO CALCULATE THE WEATHER numerically have a long history. The first effort along these lines took place not in some cutting-edge university or government lab but on what the lone man doing it described as "a heap of hay in a cold rest billet." Lewis Fry Richardson, serving as an ambulance driver during World War I and working with little more than a table of logarithms, made a heroic effort to calculate weather changes across central Europe from first principles way back in 1917. The day he chose to simulate had no particular significance—other than that a crude set of weather-balloon measurements was available to use as a starting point for his many hand calculations. It's no surprise that the results didn't at all match reality.

Three decades (and one world war) later, mathematician John von Neumann, a computer pioneer, returned to the problem of calculating the weather, this time with electronic assistance, although the limitations of the late-1940s computer he was using very much restricted his attempt to simulate nature. The phenomenal advances in computing power since von Neumann's time have, however, improved the accuracy of numerical weather forecasting and allowed it to become a routine part of daily life. Will it rain this afternoon? Ask the weatherman, who in turn will consult a computer calculation.

Like weekly weather forecasting, climate simulation has benefited greatly from the steady advance of computational power. Nonetheless, there's still a long way to go. In particular, predicting the influence of clouds remains a weak link in the chain of reasoning used to make projections about changes in Earth's climate. Part of the reason is that the resolution of the global climate models in use today is too coarse to simulate individual cloud systems. To gauge their effect, today's models must rely on statistical approximations; some climatologists would be much happier if they could model cloud systems directly. The problem is that the computing oomph for that isn't available today. And it probably won't be anytime soon.

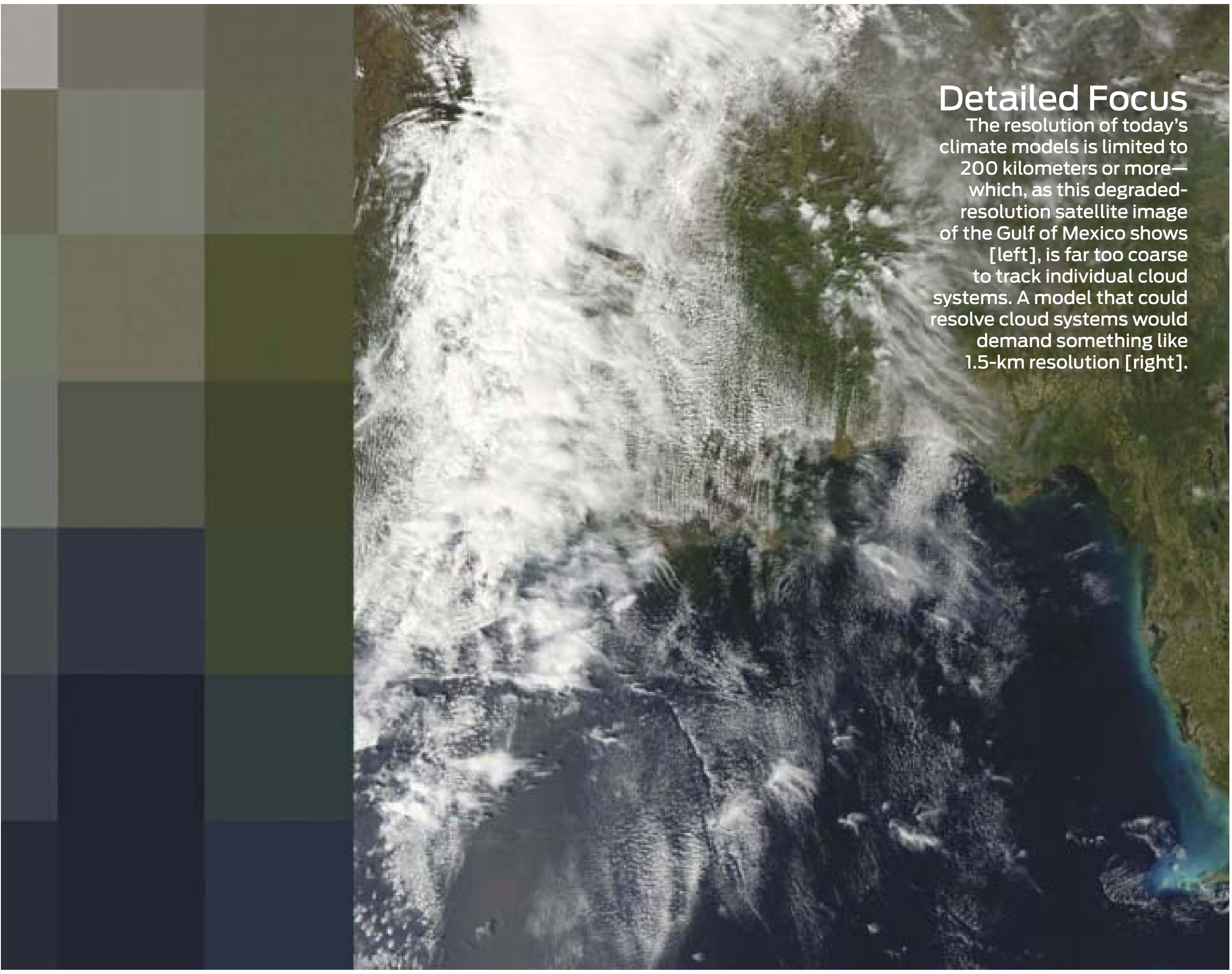
Microprocessor clock speeds are no longer increasing with each new generation of chip fabrication. So to obtain more computational horsepower, the usual strategy is to gang together many processors, each working on a piece of the problem at hand. But that solution has drawbacks, not the least of which is that it multiplies electrical demands. Indeed, the cost of the power required to run such computer systems can exceed their capital costs. This is an industry-wide problem. Companies with large computing needs, such as Google, will build facilities near hydroelectric dams to get inexpensive electricity for their data centers, which can consume 40 megawatts or more.

This power crisis also means that high-performance computing for such things as climate modeling is not going to advance at anything like the pace it has during the last two decades unless fundamentally new ideas are applied. Here we describe one possible approach. Rather than constructing supercomputers as most of them are made now, using as building blocks the kinds of microprocessors found in fast desktop computers or servers, we propose adopting designs and design principles drawn, oddly enough, from the portable-electronics marketplace. Only then will it be possible to reduce the power consumption and cost of a next-generation supercomputer to a manageable level.

BACK IN THE 1970S AND 1980S, the high-performance computing industry was focused on building the equivalent of Ferraris—high-end machines designed to drive circles around the kinds of computing hardware a normal person could buy. But by the late 1980s and early 1990s, research and development in the rapidly growing personal-computer industry dramatically improved the performance of standard microprocessors. The ensuing pace of advance was so quick that clusters of ordinary processors, the Fords and Volkswagens of the industry, all driving in parallel, soon proved as powerful as specially designed supercomputers—and at a fraction of the cost.

That development benefited the many scientific groups that required high-performance computers to get them around the next bend on the research frontier. The equipment may not

PREVIOUS PAGES, CLOUDS: IRINA EFREMOVA/ISTOCKPHOTO



Detailed Focus

The resolution of today's climate models is limited to 200 kilometers or more—which, as this degraded-resolution satellite image of the Gulf of Mexico shows [left], is far too coarse to track individual cloud systems. A model that could resolve cloud systems would demand something like 1.5-km resolution [right].

NASA

have been quite as efficient as custom-built supercomputers—either computationally or as measured by the power they consumed—but there was no sticker shock, and they were relatively easy to trade in for newer models when the time came.

Although it now seems obvious that using mass-produced commercial microprocessors was the best way to build supercomputers, that lesson took a while to sink in. Over and over since the mid-1990s, some research group would attempt to build special-purpose computing hardware to solve a particular problem with great computational efficiency and elegance. But such efforts would regularly be eclipsed by brute-force improvements in the performance of mass-produced microprocessors, for which transistor density and clock speeds were doubling every 18 months or so. That's why for more than a decade only the boldest, most ambitious research programs have veered from using commercial off-the-shelf technology for their high-performance computing needs.

But the improvements in clock speeds that have for so long come from shrinking the size of transistors have now come to an end. The reason, broadly speaking, is that to maintain the historical connection between reduced transistor size and increased speed would require too much power—the chip would cook. So to boost computing performance, manufacturers are increasingly putting more than one microprocessor core on each chip.

What's more, research and development investments today are rapidly shifting from desktop PCs toward handheld devices, which are built around simple low-power cores. Because they are designed to satisfy just the requirements of the application at hand, they draw substantially less power than do desktop microprocessors, which are intended for general use. The performance of such embedded processors is more limited, of course, but they can offer greater computational and energy efficiency. And these low-power processors are now starting to offer the capability for double-precision floating-point operations, which makes them suitable for scientific computing.

Just as the architects of today's supercomputers first started advocating the use of building blocks based on desktop-computing technology in the early 1990s, we are now urging our colleagues to use commonplace designs—processors intended for portable, embedded-computing applications.

At first blush it might seem strange to want to build a supercomputer using the kind of pared-down processors found in, say, cellphones. But when you consider that both applications are constrained by the amount of power used, this notion begins to makes sense. Indeed, in both cases efficiency is everything. The approach we're suggesting follows logically from supercomputer pioneer Seymour Cray's 1995 statement "Don't put anything in that isn't necessary."

IEEE Spectrum | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

Consider also that the embedded-processor market has recently taken off with the emergence of high-performance portable electronics like smartphones, which can perform very sophisticated tasks. Now, for example, you can buy something that handles speech recognition, gaming, Web browsing, even today's cutting-edge augmented-reality apps—and fits in the palm of your hand.

The availability of such wonders helps to explain why portable devices now account for a larger share of the microprocessor market than personal computers do. And because embedded microprocessors are comparatively simple to design, they can be quickly adapted. Whereas you would expect the vendor of a desktop or server processor to develop just one new CPU core every two years or so, the maker of a typical embedded processor may generate as many as 200 variants every year.

The companies serving the embedded-chip market can keep up that breakneck pace because they have the tools to offer rapid turnaround when a client requests a power-efficient, semicustom chip for its latest mobile gizmo. And those companies can also provide software—compilers, debuggers, profiling tools, even complete Linux operating systems—tailored to each specific chip they sell.

Supercomputer makers would do well to leverage these capabilities to build power-efficient number-crunching machines. This approach differs from past attempts to customize chips for supercomputers, because we envision using designs and design tools that have already been tested in industry. Doing so would allow the construction of unique supercomputers, each one tailored to meet the particular requirements of the task at hand. "What interesting problems can I solve with the best machine available?" will no longer be the question. Instead, researchers will ask, "What kind of hardware do I need to address the scientific issue that's important to me?" And getting the money to build and run it will finally be within the realm of possibility.

Take the problem of predicting climate change, which requires knowledge of the future evolution of land surfaces, the ocean, and the atmosphere, the last of these being the toughest to model. Most of today's atmospheric models slice and dice Earth along lines of latitude and longitude. They do so at a surprisingly coarse resolution—often with computational grid cells that measure two or more degrees across in latitude or longitude. That's to say, one cell—one pixel in the resulting global maps—might measure a couple of hundred kilometers on a side. A lot of interesting weather phenomena take place within areas of such size—the formation of clouds, thunderstorms, perhaps even entire weather fronts—things that now have to be accounted for in some way other than modeling their dynamics using the equations of motion.

The strategy employed to keep track of the weather within each cell involves various statistical averages and parameterizations, which introduce a great deal of uncertainty. For this reason, climatologists would dearly like to improve the resolution of their global atmospheric models, ideally reducing the cell size to just a kilometer or two. But up until now that's been almost inconceivable because the computing demands would become so great. Indeed, trying to satisfy those demands with the kinds of supercomputers available now would be astronomically expensive; the electricity used alone would probably blow the budget.

Five years ago, a group of us—computer researchers from the University of California, Berkeley, and the Lawrence Berkeley National Laboratory—began informal discussions on the future of computing. We arrived at the idea that the building blocks for tomorrow's computing systems of all scales were likely to be simple microprocessor cores, like those found now in battery-powered consumer electronics. Even the powerful supercomputers used for such things as climate modeling could be built this way. To investigate that possibility, the three of us looked carefully at a well-regarded climate model (a version of something called the Community Atmospheric Model) to estimate what it would take to reduce its spatial resolution to the point where individual cloud systems could be simulated.

Our target was to have the horizontal resolution be no more than 1.5 kilometers, with the atmosphere separated into 100 layers, rather than just the 25 or so that are typically used. That would make for about 20 billion computational cells. We have a name for the hypothetical supercomputer capable of running such a climate model while using a reasonable amount of electricity: the Green Flash. It sounds like a superhero, and in some sense it would be one.

How fast would the Green Flash have to run? To keep with the superhero theme, we could say that it would have to run faster than a speeding bullet, which travels about 1000 times as fast as a person can walk. That is, the Green Flash would need to simulate atmospheric motions at a rate about 1000 times as fast as the weather plays out in reality. To model Earth's climate over the course of a century, for example, should take no longer than a month or so. Otherwise the computer would prove too slow to be useful to climate scientists, who must run their models many times to adjust various internal parameters, and to compare different climate scenarios against one another and against one or more millennium-long baseline simulations. To achieve the desired 1000-fold speed advantage over nature would require, at the very least, 10^{16} floating-point operations per second (flops) using double-precision arithmetic. That's 10 petaflops.

Thankfully, the atmosphere lends itself to modeling on a machine using many different processors working in unison. That's because you can calculate changes in the weather at one spot without having to know what's happening at distant locales. All you really need is information about what's going on directly to the north, south, east, and west.

You might, for example, divide the globe into 2 million horizontal patches, each one further subdivided into 10 vertical slices.

Processor Pick

To run a cloud-resolving climate model would require a gargantuan supercomputer. If it were built using general-purpose processors (like AMD's Opteron), the power consumption would be enormous. As the numbers below demonstrate, picking a low-power chip, say Tensilica's XTensa LX2, which is typically used for embedded computing, makes much more sense.

PROCESSOR	CLOCK SPEED	GFLOPS/CORE	CORES NEEDED	POWER
AMD Opteron	2.8 GHz	5.6	1 700 000	179 MW
Tensilica XTensa LX2	500 MHz	1	10 000 000	3 MW

ieee spectrum Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

Each of these 20 million domains would span about 1000 cloud-resolving grid cells, demanding a sustained computational rate of 500 megaflops—well within the capabilities of one of today's high-end embedded processors. Modeling climate in this way would also require about 5 megabytes of random-access memory for each processor, and you would need to be able to read and write to that memory at a rate of about 5 gigabytes per second.

Because the computations for one region within an atmospheric model hinge only on what's happening in the adjacent regions, the communication requirements between processors are modest. Each second, about 1 GB of data would have to flow back and forth between one processor and a few others, a rate that should be possible with modern embedded hardware.

Having not yet constructed a Green Flash supercomputer, we can provide only a sketch of how one could be put together. The basic building block might be something like the CPUs available from Tensilica, a company based in Santa Clara, Calif., that tailors its relatively simple embedded-processor designs to the needs of each customer. Tensilica's XTensa LX2 processor cores, for example, can easily be run at 500 megahertz and can carry out 1 gigaflops without breaking a sweat.

It wouldn't be too difficult to integrate, say, 32, 64, or even 128 of these processor cores, along with enough local memory and data-transfer capacity for the climate-modeling problem, within a single package. But you'd need 10 million or so cores to achieve the required computational might.

Using such low-cost chips would slash the bill for processors and memory, even though you'd still have to pay the usual amounts for hard disks, software, a building to house this gargantuan computer, staff to tend it, coffee to keep them going, and so forth. In all, a supercomputer built along these lines might set you back a couple hundred million dollars. But that's no more than the cost of some of the existing supercomputing facilities at U.S. national laboratories—or of many big-budget motion pictures. Compare that with what the required supercomputer would cost if it were to be built following the conventional approach, using high-end desktop or server microprocessors.

American Micro Devices' Opteron processor core, for example, can easily outperform one of Tensilica's embedded processors. "Only" 1.7 million AMD Opterons would be needed to carry out the 10 petaflops required to simulate Earth's atmosphere at cloud-resolving resolution. So you might then be tempted to assemble a climate-modeling supercomputer out of standard dual-core Opteron chips. But that would be foolish. Even if you could foot the initial bill, which we estimate would be more than a billion dollars, you probably couldn't afford to run it, because it would consume almost 200 MW, costing perhaps $200 million a year for electricity. In contrast, a Tensilica-equipped Green Flash would sip power, relatively speaking, requiring only 3 MW, a level many large computer centers are already equipped to handle.

These numbers assume that the millions of microprocessors involved are working at high computational efficiency. That is, each of the needed calculations shouldn't take too many floating-point operations to accomplish. That might be difficult to achieve, particularly with a generic processor like the Opteron. Cores intended for embedded applications are much more promising in this regard because their circuitry can be tailored to the scientific task at hand. Need to include another hardware-multiply unit? No sweat. Want your microprocessor to carry out the most common instructions in climate modeling especially fast? That's possible too. But there's a danger if you go overboard: The development of a better algorithm to solve the problem can make your costly supercomputer obsolete in one stroke.

Our concept for the Green Flash is to build in the processing power, memory, and data-transfer pipes needed for each computational domain. Designers would further tune the circuitry to the climate-modeling problem, most likely using prototype hardware built out of field-programmable gate arrays (FPGAs). But they would take care not to make the circuitry too specific to one particular climate model. That way the resulting supercomputer could be reprogrammed easily to do whatever calculations are desired as the numerical recipes used to simulate the oceans and atmosphere improve with time. And the same supercomputer could be used to run many different climate models efficiently, allowing scientists to compare one against another—which helps enormously in evaluating the credibility of the results.

Building a Green Flash supercomputer would obviously be a huge undertaking, so we're approaching it with baby steps. One of the most significant took place late last year, when we and our coworkers simulated a design for a possible Green Flash processor using something called the Research Accelerator for Multiple Processors, or RAMP, an FPGA-based hardware emulator that our colleagues in the department of electrical engineering and computer science at UC Berkeley have assembled for research in computer architecture.

We came up with the design in collaboration with Chris Rowan and others at Tensilica. The design uses that company's XTensa LX2 processor core as its basic building block. The RAMP simulation of our science-optimized processor core ran a coarsened version of a global climate model that David Randall and his colleagues at Colorado State University, in Fort Collins, devised expressly to resolve clouds. The point was not to find out about the climate; it was just to test our RAMP-emulated designs.

Randall and his colleagues are excited about the prospect of a Green Flash to run their model. Using the fastest supercomputers available right now and operating at a somewhat marginal resolution for tracking individual clouds (4 km), their model outpaces nature by only a factor of four—0.4 percent of what's required. So only with something like the Green Flash will such models ever prove truly useful for long-term climate prediction.

The results of our RAMP experiments should allow us to run a climate model on an emulated Green Flash chip by the end of this year, which in turn will show just how well the real chip would perform. And if all goes well, we hope to have a small prototype system that includes something like 64 or 128 processors and their interconnections built about a year later.

Building a complete supercomputer and optimizing the code to run on it will keep us and many other members of the computer-science community busy well beyond that—perhaps for a decade or more. That's why it's so important to begin the process now. Otherwise, changes in Earth's climate could well get ahead of us. ❑

TO PROBE FURTHER The authors describe their work establishing the requirements for the Green Flash in more detail in "Towards Ultra-High Resolution Models of Climate and Weather," which appeared in the May 2008 issue of the International Journal of High-Performance Computing Applications. *The material posted at http://www.lbl.gov/cs/html/greenflash.html provides recent updates on their research.*

Spectrum Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

INFOGLUT

It's the disease of the new millennium. How do we treat it?

BY NATHAN ZELDES

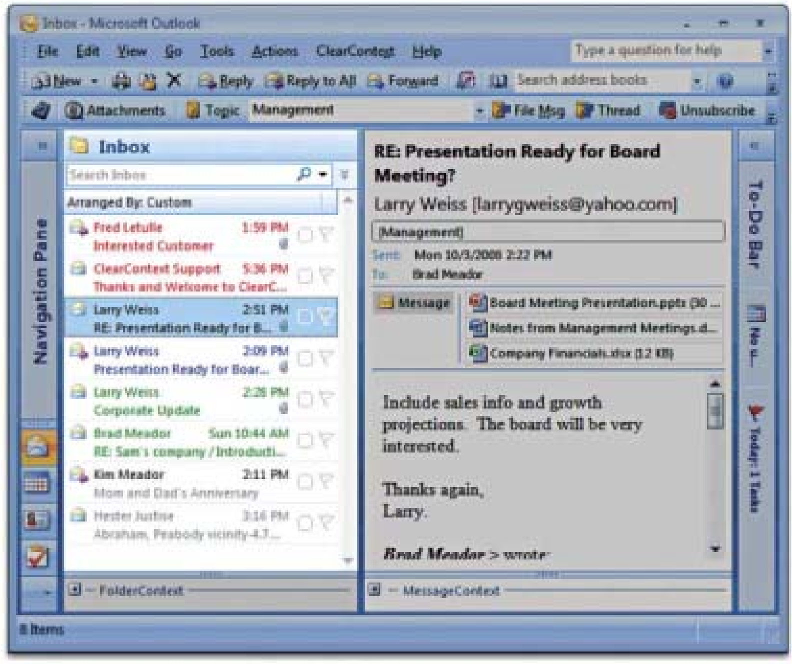
INFORMATION, THE VERY THING THAT MAKES IT POSSIBLE TO BE an engineer, a doctor, a lawyer, or any other kind of modern information worker, is threatening our ability to do our work. How's that for irony?

The global economy may run on countless streams, waves, and pools of information, but unrestrained, that tidal wave of data is drowning us. It washes away our productivity and creativity, swamps our social lives, and can even shipwreck our relationships.

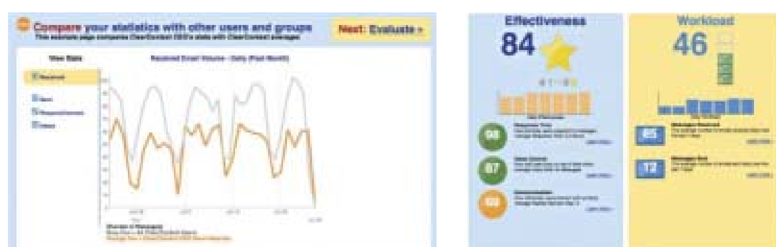
Some of us actually call it "quality time" when we sit on the sofa with our kids scrolling through e-mail on our BlackBerries. Take a few days of well-earned vacation and you spend them dreading the thousand e-mails that await your return—e-mails you'll spend a day clearing while putting off real work. On your next trip, you try to head off that problem by taking your computer along so you can chip away at the inflow late at night in your hotel room. The result? Thoughts of work cloud your enjoyment of what should be a respite from office life.

But information overload isn't just about having too much e-mail, voice mail, and text messages. It's a much more complex problem, and its effects take a toll on companies' bottom lines and on their employees' well-being. The time that information workers invest in coping with this overload is significant; at Intel, where I was until recently a principal engineer, we assessed the loss due to unnecessary e-mails and unproductive interruptions at 8 hours a week. In a 1996 Reuters survey of 1300 managers worldwide, two out of three respondents associated information overload with loss of job satisfaction and tension with colleagues, and 42 percent attributed ill health to this stress. Today, more than 10 years later, the numbers would likely be even higher.

FREDRIK BRODÉN



ORGANIZE Tools to help individual users better manage their e-mail are entering the market. One such tool is ClearContext Professional [above] from ClearContext Corp., which automatically sorts messages by importance and includes tools to let you easily consolidate e-mail by project.



ANALYZE ClearContext Professional also includes e-mail analysis tools [above] and, just for fun, lets you compare your e-mail behavior with that of other users. It doesn't, however, answer the question of whether getting a speedy average response time is good or bad.

Academic researchers have been studying the problem for years, and now at last organizations are beginning to wake up and take action to mitigate it. Some are deploying training and behavioral change programs, trying their hands at setting up quotas and encouraging alternatives to e-mail, and experimenting with interruption-free "quiet time" blocks. Change is in the air.

WHATEVER YOU WANT TO CALL IT—infomania? infoglut?—it's a combination of two elements: queued messaging overload and interruptions or distractions. Queued messaging overload can happen anywhere you have a queue of incoming messages, most notably your e-mail in-box. Some of the messages are critical, most are not, but they all accumulate until you deal with them. Information workers typically receive 50 to 200 work-related e-mails daily. Surveys at Intel showed that people spend some 20 hours a week processing work-related e-mail messages, of which about a third are unnecessary. Processing this third took workers about 2 hours a week.

Interruptions and distractions take many forms. They include ringing phones, text messages, instant messages, the chime that alerts you to incoming e-mail, and, of course, a colleague dropping by your office to chat. Any of these will break your chain of thought and may make you drop your current task to start another. The myth that this is okay because people can multitask is just that; ample research proves that the brain simply doesn't work that way.

Even when the interrupting task is related to work, you still waste time as your brain switches from one task to another and back again. Field research by Gloria Mark and her colleagues at the University of California, Irvine, shows that information workers are interrupted on average every 3 minutes. Even if it takes the brain only a minute to get back in gear, that's a lot of wasted time.

Constant distractions also make us stupid. Research clearly demonstrates that interruptions degrade accuracy, judgment, creativity, and effective management. The psychiatrist Edward Hallowell coined the term *attention deficit trait* to describe this phenomenon and found that it makes people perform far below their full potential.

Creative thinking, essential to many engineering jobs, requires long stretches of uninterrupted time. Programmers are known for working odd hours, when they can have the quiet they need to concentrate. Other professionals find that their best thinking takes place on airplanes and in hotels during business trips, when they're somewhat disconnected. But even this time is shrinking fast as remote access becomes ubiquitous. At one company, researchers found that recipients read 70 percent of e-mails within 6 seconds of arrival. In the battle between creative thought and distractions, creativity is losing.

WILLIAM SHOCKLEY KNEW THE VALUE of isolation. In 1948, shortly after his colleagues John Bardeen and Walter Brattain invented the point-contact transistor in Shockley's absence, he became so upset that he holed up in a hotel room. He knew he needed a quiet place to think. Some days later he emerged, having worked out the basic design for the far superior junction transistor that became the key to modern electronics.

Today few can afford the luxury of such isolation. While just about everybody agrees that electronic messaging is critical to modern business and that some interruptions are vital to workplace interactions, clearly they've become too much of a good thing. This glut affects Fortune 500 corporations, tiny companies, schools, government agencies, churches, and nonprofits. Just about everyone, in other words.

The irony of all this constant communication is that we're not communicating well at all. Consider the meeting where everyone's eyes are glued to their BlackBerries or laptops. They're sifting through e-mail or scanning reports or updating spreadsheets; nobody is paying attention to the business at hand. Long ago, e-mail used to guarantee a next-day response; today employees respond to many of their messages slowly or not at all. In the process, they may delay progress on key projects. Catherine Durnell Cramton, an associate professor in the School of Management at George Mason University, in Fairfax, Va., identified e-mail silence as one of the biggest challenges facing geographically dispersed teams. Let's say Jack fails to answer Jill's e-mail asking him to weigh in on an important question. She may misinterpret his silence as indifference, when in fact he may be just too swamped or distracted to fashion a coherent response. Misunderstandings like that can hamper a team's performance.

The very paradigm of work planning has changed: Where we used to be plan driven—we had a plan and spent our time executing it—we are now "interrupt driven." We respond, sometimes on the spot, to any request for action. This unplanned shift of priorities can derail progress on the primary job.

YOUR BLACKBERRY GOES OFF DURING your anniversary dinner—for the third time. Sound familiar? These days many of us have to make ourselves available to our jobs literally 24/7. And with hundreds of queued, unread messages weighing on our minds, we spend an increasing fraction of our evenings, weekends, and vacations processing mail, to the detriment of our well-being and that of our families.

Judiciously applied work-from-home options can significantly enhance both productivity and work-life balance—if handled correctly. Working for Intel in Israel, I had many late-evening meetings with colleagues up to 10 time zones away. It helped enormously that I could sit in on those teleconferences from home after dinner rather than staying in the office. Intel also allowed me one telecommuting day each week, which I put to good use. However, bringing work home like that made it a challenge to keep my weekends and evenings work-free, what with the stream of e-mail continuing to flow in.

So what's the answer? At this point, every organization recognizes that information overload is a problem, and a small but growing fraction of organizations are actually doing something about it. Intel, for instance, has been trying different approaches to the problem since 1995. That's long enough to have figured out a few remedies that work—and understand why others don't do as well. At first glance the causes of and fixes for information overload would seem obvious: People send too many messages—if only they'd send less! And to be sure, part of the issue is the thoughtless use of communication channels. People write long messages where shorter ones would do, or hit "Reply All" where one recipient would suffice. Such bad habits lead many organizations to think they can solve the problem simply by issuing memos of "Top 10 E-mail Tips" advising people not to do all that. If only it were that easy.

Senders of superfluous e-mail know full well that it will be deleted; after all, they do the same thing when they're the recipients. Why, then, do they send it? The reasons run deep in the murky undercurrents of organizational culture. People may hit Reply All because they think sending a message at midnight will impress the boss, or they may be trying to cover themselves and create a paper trail in an organization where mistrust is a factor. The situation calls to mind the "tragedy of the commons" scenario: Everyone would prefer that there be fewer messages, but nobody can afford to be the first to cut back on sending them.

IT'S SURPRISING HOW LITTLE has been done on an organization-wide level to fight a problem as big as information overload, considering that the cost of fixing it is trivial

Unfortunately, organizational culture evolves much more slowly than technology does. New information channels appear and are adopted with little attention to the behavioral outcomes. When a new device makes it possible, for instance, to communicate with workers who are on vacation, nobody stops to question whether applying this capability might contribute to employee burnout. The time has come to change this shortsighted approach. Before adopting any new technology, we should figure out how best to use it in the cultural context it will inhabit.

GIVEN THAT INFORMATION OVERLOAD arises from a variety of sources, it's not too surprising that the solutions also run the gamut, from the simple to the complex. At the simple end are guidelines on e-mail management that employees are encouraged, but not required, to adopt. Although these guidelines don't remove the problem's root causes, they are easy to implement and often do have a positive impact. In 2007, Intel's worldwide IT group circulated a carefully chosen e-mail guideline to its employees every few weeks (for example, "Make a long story short—add a management summary to lengthy messages"). Meanwhile, the company offered prizes for employees' own improvement ideas. The program indeed increased awareness, improved behavior, and reduced the reported time for e-mail processing, and it has been adopted by additional groups in the company.

At the other end of the complexity scale was a program called YourTime, deployed across most of Intel in 2000. It was based on a waterfall model, in that it started at the top management level, which was exposed to the required training and thinking, and then moved down level by level, with each manager at each level training his or her staff, who then trained their own staffs, and so on, all the way to the bottom of the hierarchy. At each level people received awareness training, held a team discussion to identify changes in the context of their own work, and took skill coaching that made them more proficient in the effective use of e-mail. The program sought to teach individuals the skills required for faster in-box processing while helping teams to define "group contracts"—mutually agreed-upon behaviors and expectations that would reduce the misuse and abuse of e-mail. This program was also quite successful—for a while.

The problem with such training drives is that they are inevitably forgotten in a year or two. To maintain the impact, you must reinforce the training periodically. That's not hard to do, as the trainings are relatively inexpensive to implement, but it does take some effort and time.

In recent years companies have been experimenting with more radical solutions. One approach is to apply quotas to the e-mail messages a worker can send. The manager of one British company went so far as to ban e-mail altogether for internal messaging. *Continued on page 46*

spectrum Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

ieee spectrum Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

NEW MATERIALS AND DESIGNS CAN KEEP CHIPS COOL

Chill Out

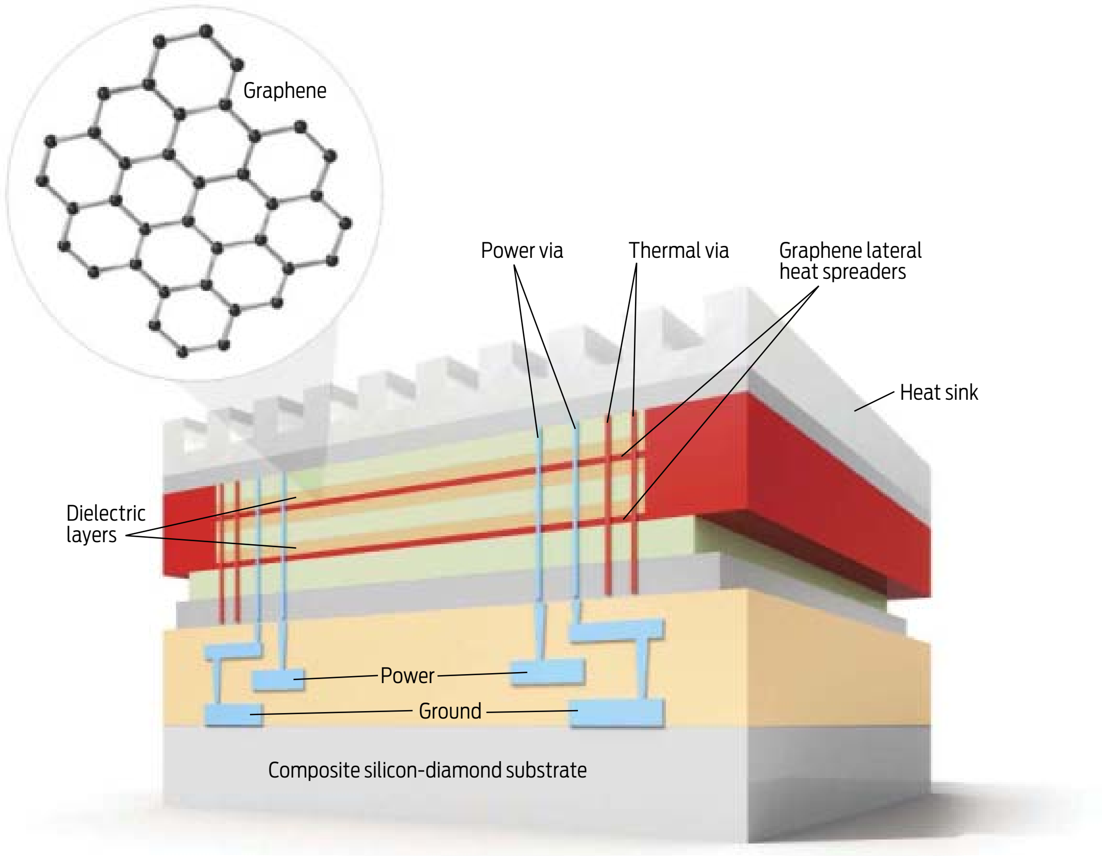
My wife recently gained a new appreciation for my work. She was trying to transfer family videos from old videocassettes to DVDs using various gadgets and software running on her laptop. Inevitably, though, the process always yielded a blank screen somewhere in the recording. These interruptions occurred because the laptop kept overheating. Only after she placed it on a stack of books with a fan blowing directly on it could the computer handle the job.

Heat is one of the worst enemies of electronics. Sit on the sofa with your laptop and you quickly feel the heat on your lap. Often though, overheating can be hard to diagnose. You may notice random errors occurring no matter what program you're running. This is especially true if you use your computer to play advanced video games, which can really tax the microprocessor and the graphics card. If your machine frequently experiences fatal errors or "the blue screen of death" on such occasions, chances are it has thermal management problems. And overheating doesn't just degrade a computer's performance; it can also shorten its useful life.

By Alexander A. Balandin

PHOTO: DAN SAELINGER; STYLING: WENDY SCHELAH; ELECTRONICS: COURTESY OF TEKSERVE

spectrum Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags



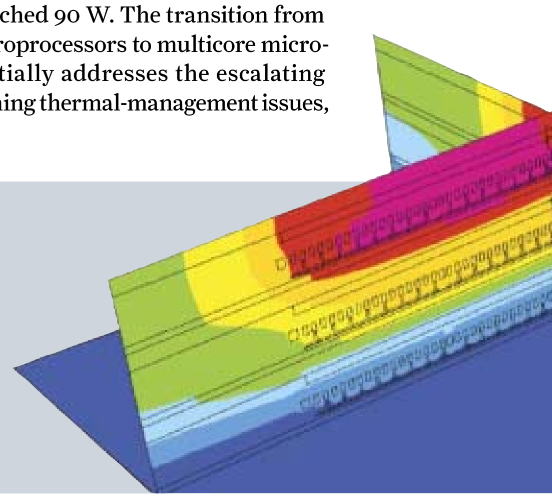
MIGHTY MATERIAL: Graphene is a one-atom-thick layer of carbon arranged in a honeycomb lattice. The material's high electron mobility and high thermal conductivity could lead to chips that are not only faster but also better at dissipating heat. This schematic shows a three-dimensional stacked chip with layers of graphene acting as heat spreaders. At present, though, graphene is extremely costly to manufacture.

The electronics industry has tended to deal with overheating more at the system level—using cooling fans, for instance, to regulate rising temperatures inside computers. Until recently, heat was not considered a major problem to be addressed in the design of the ICs themselves. But higher circuit densities and faster clock speeds are making chips run so hot that manufacturers can no longer ignore the problem. According to the International Technology Roadmap for Semiconductors (ITRS), which reflects a consensus of chip manufacturers worldwide, managing heat generation within ICs will be a crucial issue in developing the next generation of electronics. Growing concern has in turn sparked new research into chip designs and novel materials that would allow electronics to run much cooler, thereby improving their performance and extending their life span.

Heat is an unavoidable by-product when operating any electronic device. Electronic circuits contain many sources of heat, including the millions and even billions of transistors that are routinely packed into modern ICs as well as the interconnects—the labyrinthine connections linking these components together. In the past, packaging engineers, not chip designers, were the ones who dealt with overheating. They would position components so that any excess heat would move first from the die to a heat sink, and then they'd use a flow of air—from a cooling fan, say—to dissipate heat into the surroundings. Still running too hot? Just use a bigger fan. For years, such coping strategies were sufficient. But now, with the electronics industry aggressively shrinking chip features below 50 nanometers and moving toward three-dimensional integrated circuits, the era of the big-fan solution has passed.

One good way to assess a microprocessor's propensity to heat up is by looking at its thermal design power (TDP), which represents the maximum sustained power that must be dissipated when the chip is performing a typical task. With each new generation of microprocessor, the TDP has grown. In the first Pentiums, for example, the TDP was below 20 watts; in the Pentium 4 it reached 90 W. The transition from single-core microprocessors to multicore microprocessors partially addresses the escalating TDP and worsening thermal-management issues,

TOP: BRYAN CHRISTIE DESIGN; BOTTOM: CFD RESEARCH CORP.



HEAT WAVE: These simulations show temperature distributions in a conventional PC [far right] and in a proposed three-tier, three-dimensional chip [near right]. Areas of red and pink indicate substantial temperature rises. In a traditional two-dimensional IC, the entire die would rest on a heat sink, which dissipates excess heat. But in 3-D chips, some tiers are separated from the heat sink by many thermally resistive layers of material, so a large hot spot can form.

as the new multicore chips gain performance not by increasing clock frequencies but by adding processors. But other big heat issues remain, including the appearance of hot spots within the chip, where heat fluxes can climb as high as 1000 W per square centimeter. What's worse, according to ITRS projections, within the next five years, up to 80 percent of microprocessor power will be consumed by interconnect wiring in regions of the chips that are particularly difficult to cool.

Designs intended to improve performance in state-of-the-art ICs can also unintentionally aggravate overheating. For example, the silicon-on-insulator technology used in some advanced chips incorporates buried layers of silicon dioxide to boost transistor speeds and reduce power consumption. But the thermal conductivity of SiO_2 is only one-tenth to one-hundredth that of regular silicon substrates. And that diminished ability to conduct heat translates into faster temperature rise and earlier device failure.

Three-dimensional electronics, which are currently under development, likewise pose a real nightmare for heat removal. Compared with conventional, 2-D designs, ICs containing stacked dies promise higher integration density, lower interconnection complexity, and a smaller chip area. Researchers are considering various 3-D architectures, but they all share one thing: They are hard to keep cool. Any theoretical performance gain from 3-D electronics will be impossible to realize if thermal issues aren't addressed. Simulations show that a rise of 10 °C in the operating temperature of such an IC increases circuit delay by about 5 percent; doubling the heat density degrades performance by more than 30 percent.

Some of the thermal phenomena taking place in newer chips have never been observed before. For example, a material's thermal conductivity can change when it is formed into a nanostructure. To understand why, you need to know a little bit about what's going on at the atomic scale.

Most of the heat conduction in semiconductors (such as silicon) and electrical insulators (such as silicon dioxide) takes place by virtue of the vibrational motion of atoms in the crystal lattices of these materials. These complicated vibrations can be reduced to quantized modes, known as phonons, which in many ways resemble elemental particles. A given material can be characterized by the average distance that the phonons will travel in it at a given temperature. In silicon at room temperature, phonons travel around 50 to 100 nanometers. But if the device's features are on the order of 50 nm—as they already are in state-of-the-art microprocessors—the phonons will be subject to additional scattering at the boundaries and interfaces inside the chip, which in turn decreases the chip's ability to shed heat.

As device features approach just a few nanometers, the behavior of phonons gets even more interesting. In such cases, the phonons become spatially confined between layers of different materials and exhibit properties different from those in bulk crystals. While this confinement may prove problematic in many instances, it may also lead to new opportunities to improve device performance, by allowing the transport of the phonons to be precisely controlled and engineered.

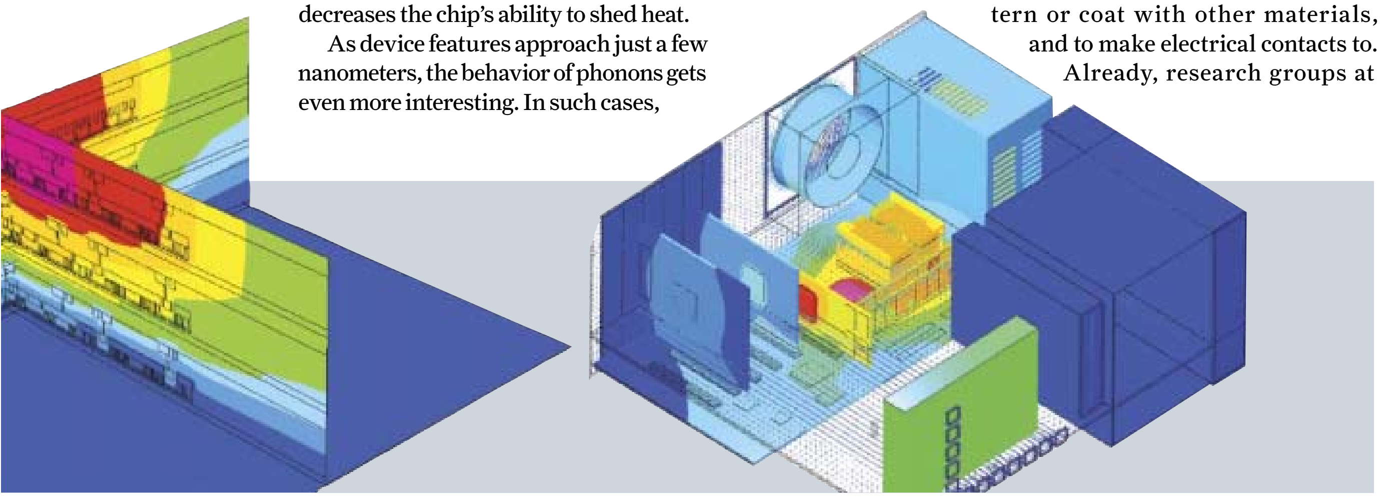
These and other factors have led engineers to search for innovative ways to remove heat from electronic devices. One promising approach is to incorporate materials that have a high thermal conductivity into the chip. These can be used as fundamental building blocks for the electronics, or they can serve as specialized heat spreaders, which as their name implies help to move heat so that it doesn't concentrate and form a hot spot.

Two candidates for such applications are diamond and carbon nanotubes. Diamond, with a room-temperature thermal conductivity in the range of 1000 to 2200 W per meter per kelvin, is the best thermal conductor among bulk crystals and has been considered for use in heat spreaders and in composite silicon-diamond substrates. The room-temperature thermal conductivity of carbon nanotubes is even higher than that of diamonds: around 3000 to 3500 W/m·K. Because carbon nanotubes are also excellent electrical conductors, researchers are exploring using them as transistor channels and interconnects and as thermal vias, which are channels used to move heat vertically through the layers of a chip or a 3-D stack to the heat sinks above or below.

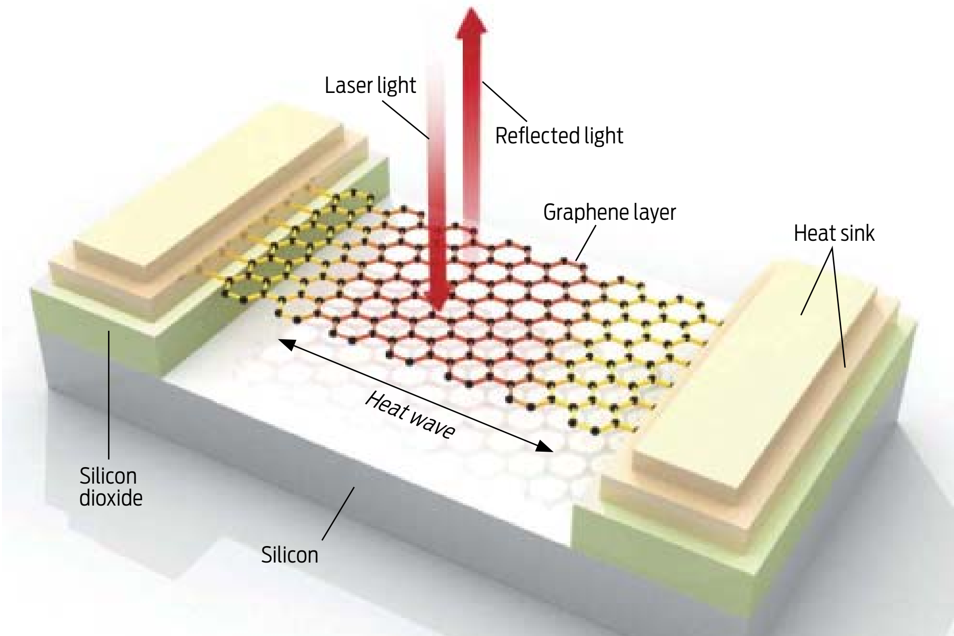
Potentially better than either diamond or carbon nanotubes is graphene, which is a one-atom-thick sheet of carbon arranged in a honeycomb lattice. The material exhibits unique electrical properties. In particular, it has extremely high electron mobility at room temperature. Electron mobility defines how easily an applied electric field moves the electrons in a given material. Higher mobility translates into faster devices.

The first freestanding graphene was produced in 2004, when a research team from the University of Manchester, in England, and the Institute for Microelectronics Technology, in Chernogolovka, Russia, used Scotch tape to peel off a single layer of carbon atoms from a sample of bulk graphite. Since then, the investigation of this material's unique properties has grown at an astonishing pace worldwide.

Researchers have proposed using graphene in carbon or hybrid silicon-carbon electronics and in ultrasensitive detectors. The new material could avoid many of the challenges that carbon nanotubes face. For one, its planar geometry means that it's much easier to grow or deposit on a substrate, to pattern or coat with other materials, and to make electrical contacts to. Already, research groups at



HEATING UP GRAPHENE



Measuring the thermal conductivity of something with a thickness of just one atomic layer is tricky. At the University of California, Riverside, we approached the problem this way: First, we prepared samples to measure, each consisting of a long graphene flake suspended across a trench in a silicon wafer and attached to heat sinks. We then heated the graphene flake with a laser. The heat wave propagated from the middle of the graphene to the heat sinks.

To measure the temperature at the center of the hot spot, we came up with an unconventional use for a micro Raman spectrometer. Ordinarily, Raman spectroscopy is an optical technique used to identify materials. The Raman spectrum of graphene has a clear peak, referred to as a G peak; the spectral position of the peak depends on the temperature of the sample. So by measuring the exact position of the G peak, we were able to use our spectrometer as a thermometer.

Only a small portion of the laser light ended up dissipating in the graphene. Most of the light passed through it and was reflected back. We determined the fraction of the light dissipated in the sample by comparing the Raman intensity of the graphene with that of bulk graphite. Knowing the temperature rise, the dissipated light power, and the geometry of the graphene flake, we then determined the graphene's thermal conductivity. The measured values exceeded 3000 watts per meter per kelvin near room temperature and depended on the size of the graphene flake. We learned later that our results agreed with physicist Paul G. Klemens's predictions years earlier. *—A.A.B.*

the University of Manchester, Columbia University, IBM, and elsewhere have succeeded in fabricating field-effect transistors (FETs) from single and multiple layers of graphene. My group at the University of California, Riverside, in collaboration with researchers at Rensselaer Polytechnic Institute, as well as an independent group of researchers at IBM, have experimentally demonstrated that graphene FETs are less noisy—that is, they produce clearer signals—than their carbon-nanotube counterparts and can operate within the strict noise limits for data processing and communication applications.

Although graphene still has significant drawbacks, including the lack of a commercial manufacturing method, its potential as an electronic material looks promising. Recent developments suggest that it may also aid with thermal management. Last year, my group in the electrical engineering department at UC Riverside teamed up with researchers from the physics department to carry out the first measurements of the material's thermal conductivity and found it to be above 3000 W/m·K near room temperature—higher than that of diamond and on a par with that of carbon nanotubes [see sidebar, "Heating Up Graphene"].

Impressed by such high values, we then set out to find a theoretical explanation. It turns out that physicist Paul G. Klemens, an emeritus professor at the University of Connecticut, had already suggested a theory. Back in 2000, well before all the recent excitement about graphene had begun, Klemens had predicted that graphene would have a higher thermal conductivity than bulk graphite, provided the graphene sheet is large enough. In simple language, this happens because the low-energy phonons, which transfer the bulk of the heat, scatter less in a 2-D system like graphene than in a 3-D stack of graphite. In an ideal, pure sample of graphene, the phonons' travel distance would be limited only by the size of the graphene flake. Building on this and other work, we then developed our own detailed theory that takes into account the new data on graphene and explains why its thermal conductivity is so sensitive to the width of the graphene flake.

Graphene could be used in a number of ways to dissipate heat. For instance, it would work better than carbon nanotubes for making heat-dissipating, electrically conducting interconnects between transistors. Used as a heat spreader in a 3-D chip, it could be paired with thermal vias made of carbon fiber or carbon nanotubes. In situations where graphene's high electrical conductivity would prove a problem, it could be combined with an unconventional electrical insulator like synthetic diamond. If necessary, graphene can be converted into an electrical insulator by irradiating it with low-energy electrons.

In short, incorporating graphene into chip designs could yield devices that are faster, less noisy, and run cooler. Of course, the usefulness of graphene—or, for that matter, of diamond or carbon nanotubes—will depend not just on their physical characteristics but also on their cost and their compatibility with existing chip-manufacturing technology. It's also possible that the high thermal conductivity of graphene may decrease when it is embedded between layers of other materials or when the number of atomic layers gets too high. The exact extent of

BRYAN CHRISTIE DESIGN

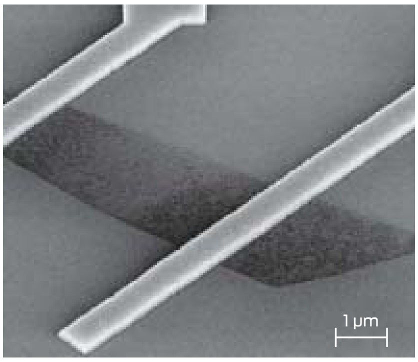
such changes is a matter of ongoing research. Still, the results so far provide reasons to be optimistic about graphene.

Managing heat in a nanoscale device becomes even more complex as chip designers continue to boost IC speeds while shrinking transistor channels. The relentless push of Moore's Law, which holds that the number of transistors per chip doubles roughly every 18 months, has led to ever higher speeds and drive currents and ever smaller chip features.

The speed and current limits of a chip are proportional to the electron mobility of the semiconductor materials used in its construction. In silicon at room temperature, electron mobility is mainly limited by phonons, which cause the electrons to scatter. A higher electron mobility usually corresponds to weaker electron-phonon interactions and thus less heat generation.

But efforts to increase electron mobility—and thus the speed and current a device can handle—can end up degrading the device's ability to dissipate heat. One approach, already used in conventional chips, is to create strain in the atomic lattice of the semiconductor substrate by forming it out of two materials with slightly mismatched lattice spacings. Often a combination of silicon and silicon germanium is used, the latter having slightly larger spacing between atoms.

A device that includes silicon germanium can have serious problems dissipating heat. That's because the thermal conductivity of SiGe, and of other alloys, is an order of magnitude smaller than that of its constituent semiconductors. So having a layer of SiGe in a device makes it more difficult for heat to escape from the transistor channel to the heat sink. In nanoscale chip architectures, problems caused by overheating in the transistor channels can cancel out any gains in mobility achieved by straining the lattice.



TOUGH STUFF: Graphene can sustain high electric currents and conduct heat extremely efficiently. This scanning electron microscopy image shows a single-atomic-layer graphene flake connected to metal bars, a basic building block for many electronic devices.

There may be ways to both enhance electron mobility *and* improve heat removal by devising ways to control the flow of phonons in the device. Such methods are termed phonon engineering. Our theoretical work suggests that sandwiching silicon nanowires or ultrathin silicon films between layers of diamond can increase electron mobility if the interface is good. Diamond not only has a high thermal conductivity, it is also acoustically harder than silicon (acoustical hardness is a product of a material's sound velocity and its mass density) and can be exploited to modify the phonon dispersion within the material. The acoustical mismatch between the silicon and diamond acts to partially suppress the phonons and electron-phonon scattering that would otherwise slow electrons down. Initial experimental results are encouraging, but the quality of the interface continues to be an issue. The rapid progress in depositing synthetic diamond films on silicon achieved in industry and research labs gives us hope that composite wafers with optimized phonon properties may soon become a reality. In other studies, we are investigating the usefulness of graphene for redirecting and controlling the phonons.

What other ways are there to deal with excessive heat? One promising new technology is thermoelectric cooling, either of the entire chip or of local hot spots. A thermoelectric cooler is essentially a heat pump that transfers heat from one side to the other when current is run through it. By design, it has no moving parts or liquid components and can be very robust and compact. Such devices are already being used to cool solid-state lasers and LEDs—and beer: The same basic technology can be found in those portable coolers that plug into the cigarette lighter of your car.

The main problem with thermoelectric cooling of ICs is finding a good material to use. One of the best bulk thermoelectric compounds is bismuth telluride, but it can't be integrated with silicon using conventional chip-fabrication technology. So investigators are seeking thermoelectric materials that are more compatible with manufacturing techniques but still retain their ability to wick away heat. Recent reports suggest that silicon nanowires or tiny blobs of silicon germanium and silicon (called quantum dots) may work.

A variation on this theme is to attach thermoelectric coolers to the chip's heat sinks instead of relying on air to carry away the heat. In January, a team led by researchers at Intel demonstrated the first chip-scale thermoelectric refrigerator.

Another approach is to introduce a liquid cooling system comparable to that in your car's engine or what was used in the early Cray supercomputers, which were cooled by Freon circulating through pipes. For state-of-the-art ICs, it may be possible to cool individual heat-dissipating elements on the chip using liquids. Liquid cooling avoids the thermal resistance that occurs at the interface of two solids. Exactly how to manage the microscale (if not nanoscale) plumbing remains an area of active research.

Efforts also continue to improve the thermal interface materials between the chip and the heat sink. The right kind of interface material can greatly improve heat removal and reduce the thermal resistance between the chip and its packaging. Common interface materials consist of an oil or grease embedded with ceramic or metal particles. But here again, fillers made from graphene look attractive. Even a tiny amount of graphene can lead to a substantial increase in thermal conductivity.

Although researchers are optimistic about these recent developments in thermal management, it will take years until any of the new materials and designs find their way into commercial chips. So you may be wondering what you can do in the meantime to keep your computer from overheating. Here's one easy fix: Open up your machine and remove any dust, which can accumulate and prevent heat from dissipating properly. Or follow my wife's example and aim a good strong fan at your computer. With any luck, that will tide you over until the more elegant solutions I've described here finally arrive. ❑

TO PROBE FURTHER For more about Alexander A. Balandin's work on thermal management in electronic devices, see the Web site of the Nano-Device Laboratory at the University of California, Riverside: http://ndl.ee.ucr.edu. His work on phonon engineering is supported by the U.S. Air Force's Office of Scientific Research. Balandin is also associated with two multi-institutional research consortia that are investigating materials and designs for future electronics: the Center on Functional Engineered Nano Architectonics (http://www.fena.org) and the Interconnect Focus Center (http://www.ifc.gatech.edu).

GUANXIONG LIU, NANO-DEVICE LABORATORY, UNIVERSITY OF CALIFORNIA, RIVERSIDE

spectrum Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

The RECESSION'S SILVER LINING

How the semiconductor industry can use the recession to create the next technology renaissance

By **Pushkar Apte & George Scalise** *Illustrations by* **Keith Negley**

COUNTLESS RESEARCH INSTITUTIONS contributed to the digital, wireless, and mobile technologies that underpin our modern world. But none contributed more than Bell Telephone Laboratories, which logged an astonishing share of the key advances of the 20th century, including the transistor, the cellphone, the digital signal processor, the laser, the Unix operating system, and motion-picture sound. We no longer have Bell Labs to fund research with long-term payback. That has prompted many to wonder: Who will pay for such research now, and where will it be done?

We say: Governments and corporations must share the burden, and they must do it in structured collaborations among universities, companies, and government agencies in which intellectual property is freely available to all participants.

We also say, the sooner we can get started, the better. The recession has left R&D spending in free fall. This year, the global semiconductor industry is expected to spend just US $200 billion on research—$50 billion less than in 2008. And times are really tough in the semiconductor equipment industry, whose R&D operations will shrivel like a salted leech from $34 billion in 2007 down to a pitiful $10 billion in 2009.

In the United States, a few basic sciences are getting a reprieve, thanks to the federal stimulus package. Of the $787 billion designated, $10 billion went to the National Institutes of Health for life-sciences research; on the other hand, there has been a steady federal funding decline for physical sciences in the United States.

In the long run, however, even the life sciences are unlikely to benefit in any meaningful way from that load of cash. A one-time infusion helps, but it also creates a classic feast-or-famine problem: The money needs to be spent by September of next year. And because there's no follow-up money to keep these programs going beyond that time, officials can't start major long-term initiatives.

But the recession isn't what's causing this problem; it's only revealing an intensifying trend in the semiconductor industry. Revolutionary innovation has been missing in action for about 40 years as the industry instead focused on incremental advances. The industry could get away with short-term research because those incremental advances got the companies where they needed to be, financially speaking.

Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

Limiting that funding to incremental research is why there hasn't been a "transistor moment" in 50 years. So, painful as it is, this economic gloom might actually turn out to be a good thing. It offers the industry, for the first time in decades, an opportunity to rethink its most basic strategies, down to the engine that keeps it all going—innovation.

Innovation has often been a catalyst for economic recovery. It happened in the 1930s, when DuPont invented one of the major materials of the 20th century: neoprene. Two years after its introduction, neoprene, a synthetic rubber, was in every car and plane built in the United States, and 50 years after that it was in knee braces and wet suits. And again in the 1980s, small steel mills like Geneva and Nucor rose from the ashes of Big Steel. Today many developed countries have stupendous R&D resources and infrastructures and are eager to use them to pursue very high potential payoffs, especially in semiconductors. So the basic factors are in place to use this recession to establish a new model of semiconductor R&D that could usher in the next generation of innovation.

But there's a problem. The innovation strategies that semiconductor companies large and small have developed over the past half century are grounded in the business practices and research conventions of a bygone era. Unless their strategies evolve to meet these changes, many of those companies will die a slow and avoidable death.

There is a way out. It just doesn't look anything like the old way out, and it will make some of those companies uncomfortable, at least initially. At the Semiconductor Industry Association (SIA), in San Jose, Calif., we have developed a new model for innovation. Our model is counterintuitive—it asks companies to share intellectual property and invest in research that also benefits competitors, something that's anathema in today's standard industry practices. But our approach has been successfully tested by corporations like IBM, Intel, Micron, and Xilinx, among many others. When companies have embraced it, they've seen encouraging results. For example, there has already been a significant breakthrough that can be largely attributed to this model: the graphene-based BiSFET logic device, which operates at a fraction of the power of today's typical devices. The concept is being developed by researchers at the University of Texas at Austin, and if it works as well as the simulations imply, it could change the world.

Revolutionary INNOVATION has been *missing in action* for almost 40 years

The BiSFET, we hope, is only the beginning. But it could be the end, if we can't convince semiconductor executives of the value of our model. (And not everyone is on board with the idea of a "shared innovation environment.") To better understand the barriers, we interviewed top management at some leading semiconductor companies and universities. Our subjects represented a broad cross-section of the industry, some of which use our model, and some of which do not. These institutions represent a perfect microcosm of the stumbling blocks—and the rewards for letting go of the stifling old ways and making the leap.

BIG CHANGES ARE AFOOT. The semiconductor industry has been the greatest single source of industrial innovation in recent history. But many of the advances have been incremental, such as the shift to high-*k* dielectric materials and the move from aluminum to copper for on-chip interconnects. But as the old saw has it, after you've gone from the buggy to the car, building a better buggy whip won't do you any good. In electronics, building a better triode won't help. What the industry needs now is more like the shift from vacuum tubes to semiconductors. That's because two trends are driving the semiconductor industry to a momentous inflection point.

First, the customer is changing. Several hundred million individual consumers, many of them in the developing world, have joined the global economy in just the past few years. Individuals have replaced companies and governments as the domi-

ieee spectrum Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

nant buyers of cellphones, laptops, digital cameras, and other high-tech goodies. In fact, large corporate IT departments are no longer the world's primary technology consumers [see time line, "Vectors of Change"]. These hundreds of millions of customers are atomized into many fragments, they don't have monolithic tastes, and most important, they're much more cost-conscious than big companies are.

Second, the metal-oxide-semiconductor transistor—the basic building block of the entire edifice of modern semiconductors—is approaching fundamental physical limits. The "next big thing" won't be a linear progression of faster and faster computing and communication, laid out in road maps of the sort we've issued over the past 20 years. In fact, we don't know what this next big thing is going to look like, because it could come from anywhere.

Engineering will never return to its isolation in a bubble of mechanics and computers. And therein lies the rub: Today the field has become fantastically inter- and multidisciplinary. So engineers and their companies need to be fluent in a growing panoply of languages: neuroscience, biology, geophysics, and more. The reason is that the next great innovation might come from a neuroscientist whose circuits can mimic the functions of a synapse or from a geologist whose algorithms model the flow of magma inside volcanoes.

However, no one company can be expected to keep tabs on every significant development in academic science and technology. Indeed, conventional semiconductor R&D and strategic marketing departments are often mired in short-term firefights, product deadlines, and meeting the next quarter's financial goals.

WE THINK the SIA model can do for research in the 21st century what the Bell Labs model did in the 20th. With our model, the industry can draw on a kind of nationwide "neural network" of academic research. To understand what we're proposing, you need a quick lesson in university-industry interaction as it has existed for the past five decades.

In the existing system, a company consults individual professors with specific research questions, or it invests in local colleges mainly to burnish its image. In the first model, a company hires a star professor or researcher as a consultant and might also fund one or two graduate students for a small, proprietary project. The typical scale of this engagement is \$50 000 to \$100 000. In the second model, a company invests similar or potentially larger amounts to build goodwill in the community and to supplement its local talent base for recruitment.

These partnerships have yielded incremental advances. But to get to the next big paradigm, we need to innovate the way we innovate. To do that, we have developed our research model, exemplified here by the Nanoelectronics Research Initiative (NRI), one of 11 national centers we have set up to solve the technology showstoppers waiting to meet us in the future.

Here's how they work. The research takes place not inside one particular company but across multiple universities and various disciplines, all tied together with a common goal. Each center is funded at several million dollars a year, with about 50 universities, 250 faculty, and 450 graduate and postdoctoral students. Companies "buy in" to the research conducted there and then share early results. All the interdisciplinary research centers operate with a nonexclusive intellectual property (IP) model. What that means is that all sponsoring companies have the right to use the IP without paying any royalties, but the university owns it. More on that later.

For NRI specifically, the funding comes from five companies, two U.S. federal agencies, and four state governments. Together, these organizations have invested a total of \$20 million per year for the past four years. The NRI focuses on radically new semiconductor logic devices, ones not based on metal-oxide-semiconductor field-effect transistors, or MOSFETs, as virtually all modern chips are. In particular, NRI-hosted research has already produced the new device mentioned earlier: the BiSFET, or bilayer pseudospin field-effect transistor (not to be confused with the *bistable* field-effect transistor, or BISFET).

Some background: One of the most urgent needs in technology today is for ultralow-power devices. Vacuum tubes could have never been used to build a personal computer. A cellphone or MP3 player created with the bipolar junction or *n*-type MOS semiconductor technology that was common 30 years ago would suck up so much power that it couldn't be powered with batteries. All digital information processing is based on variations in electronic charge (for instance, in the capacitor of a dynamic RAM cell), which correspond to a 1 or 0 state. Manipulating charge requires power, which generates heat. Just as previous technology transitions from vacuum tubes to solid-state devices to integrated circuit chips were all driven by power consumption, so will the next transition.

The BiSFET, described by Sanjay Banerjee and Leonard Franklin Register and their colleagues at UT Austin, is in the earliest research phase but offers tremendous potential. The BiSFET could substitute for a MOSFET transistor in logic and memory applications. Like a MOSFET transistor, it can switch and it can amplify. Where the BiSFET stands alone, however, is in its phenomenal power parsimony: It needs only one-hundredth to one-thousandth the power of a standard MOSFET, mainly because it would switch at much lower voltages than a MOSFET.

BiSFETs will not be drop-in replacements for MOSFETs, but in principle, BiSFET-based circuits could replace CMOS circuits in any application. Behind the BiSFET is a theoretical concept that's not new in physics, but it had been completely beyond the ken of the semiconductor industry. Unlike the silicon channel in a MOSFET, the BiSFET channel is based on graphene, an exotic material consisting of single atomic sheets of the element carbon. Think of these layers as unrolled carbon nanotubes. Also, unlike a CMOS field-effect transistor, which has three terminals—source, drain, and gate—the BiSFET has four terminals: source, drain, and a top and bottom gate, which sandwich two electrically coupled layers of graphene between them. Though the two gates function as one, they must be biased differently to create electrons in one graphene layer and positively charged holes in the other. Interactions between these electrons and holes leads to what's known as an electron-hole condensate, an esoteric state of matter in quantum physics, in which the particles tend to lose their individuality and display collective behavior. The basic idea has been around for decades, but according to the rules of their strange physical makeup, these condensates could be realized only with exotic materials and at cryogenic temperatures.

The proposed graphene devices require just 25 millivolts, a scant one-fortieth of the operating voltage of today's "low-power" devices. This device could operate at room temperature and require a thousandth of the power of current devices. The BiSFET is as yet only a concept based on novel predicted physics in a novel material system. We still need experimental verification of the underlying phenomena on which the device is based.

VECTORS OF CHANGE

GOVERNMENT AND CORPORATE APPLICATIONS once dominated technology products but now bow to the demands of the individual consumer. The convergence of what were once standalone devices reflects the growing demands of an increasingly fragmented, cost-conscious consumer population.

CATEGORIES: ● Military ● Research ● Corporate ● Consumer

● ● ● **IBM 701 DEFENSE CALCULATOR** (1952) The company's first production computer designed primarily for scientific calculation is built at IBM's Poughkeepsie laboratory. Only 19 of the room-size computers are manufactured for use by U.S. national laboratories, aircraft manufacturers, and the U.S. Weather Bureau.

Mobile Telephony A (MTA)

● **MOBILE TELEPHONY A (MTA)** (1956) Ericsson develops the first fully automatic mobile phone system. The 160-megahertz system, commercially released in Sweden and with a total of 125 subscribers, requires no manual control but weighs 40 kilograms.

● **IBM 608 CALCULATOR** (1957) The 608, with its 3000 transistors and punch-card input/output mechanism, begins the transition of IBM's line of small and intermediate electronic calculators from vacuum-tube to transistor operation. The calculator is available for purchase for US $83 210.

● ● **TX-2** (1958) Built at MIT's Lincoln Laboratory, the TX-2 computer's then-huge 64 kilobits of memory is credited with advancing both artificial intelligence and human-computer interaction.

● ***TENNIS FOR TWO*** (1958) Before arcades or video

games, visitors waited in line at Brookhaven National Laboratory to play *Tennis for Two*, a forerunner of the modern video game.

● **TELSTAR LINK** (1962) The first transatlantic satellite link via the AT&T Telstar satellite relays the first telephone calls and television images through space.

● ● **AUDIOCASSETTE** (1963) Philips introduces the first audiocassette in Europe. In 1964, the cassette is introduced in the United States. The system is initially designed for dictation and

portable use, not music, so its audio quality needs improvement.

● **SONY CV-2000** (1965) Sony introduces the first videotape recorder for home use.

● **MICRAL N, THE FIRST PC** (1972) The Micral N computer, developed by French engineers, is the earliest commercial personal computer based on a microprocessor, the Intel 8008.

● ***PONG*** (1972) Atari kick-starts the coin-operated video-game industry with *Pong*, the earliest electronic ping-pong video game.

● **ETHERNET** (1973) Xerox PARC begins work on Ethernet technology, forming the computer networking technology framework for future corporate local area networks.

Color TV becomes available in England in 1967 with BBC2.

Four computers are connected to form the ARPANET. The Internet is born on October 29, 1969.

1950 1960 1970

The bottom line is that behind this breakthrough were NRI-assembled teams that included physicists, materials scientists, and electrical engineers specializing in device design. The successful application of graphene's alchemical properties to semiconductor physics could have happened only within the interdisciplinary research architecture we have created.

IN RESEARCH, AS IN LIFE, there's no such thing as one size fits all. When we queried tech companies about our model, we found that few of them would be willing to adopt it. Impediments are often relics of the mind-set created in the last century.

Two main arguments came up again and again: Technology managers said they did not want to share intellectual property or research with competitors, and they did not want to spend money on what they could learn by attending conferences. A more fundamental issue was that many companies, particularly ones forced into short-term strategies, do not consider university research an important part of their business strategy.

By definition, the research performed in a collaborative university environment is shared by many players, including competitors—and potential future competitors. One perceived nightmare scenario, for a corporation, is that of a university professor or student forming her own company to exploit a tech breakthrough. Why should a company invest in research that also benefits its rivals? In terms of time and money, IP is the proverbial "giant sucking sound." Of course IP is critical, but what's often misunderstood is that its value depends entirely on the maturity of the technology. Guarding product IP like Cerberus at the gates of hell is not necessarily a wise strategy, especially for early-stage research, which occurs years before an innovation can be brought to market.

The problem here is that semiconductor companies are behaving as if they were pharmaceutical companies. With pharmaceutical discoveries, the early-stage IP is the most important; it would be unthinkable to share the development costs of a Prozac or a Celexa with a competitor. But in the semiconductor industry, no early-stage IP is ever "ready to wear." There's lots of cutting, fitting, altering, refitting, and realtering before it's ready for the runway. Xilinx chief technology officer Ivo Bolsens put it very well when he told us, "There are a hundred decisions and innovations that I will need to add before I can take an excellent academic idea and make it into a product."

Consider carbon nanotubes. These basic building blocks can be used in many different ways to develop countless different technologies and products. Patenting something so basic would be akin to patenting a brick. Builders can use the same brick to make castles and cottages. The outcomes are vastly different and do not depend in any way on whether that builder has the patent on the brick. And in that sense, the BiSFET device is a stellar example of the kind of early IP that companies are so unwilling to share. No one has even created the device yet—it's certainly not ready

CLOCKWISE FROM LEFT: ERICSSON; BROOKHAVEN NATIONAL LABORATORY; GEOFF BLACK/ISTOCKPHOTO; ALCATEL-LUCENT

BETAMAX (1975) Sony introduces the SL-6300 video system.

VHS (1976) JVC introduces the Video Home System, the Betamax competitor that goes on to win the format wars.

APPLE I (1976) Apple demonstrates its first computer at the Homebrew Computer Club in Palo Alto, Calif. Unlike the kit-based hobbyist computers at the time, the Apple I is a fully assembled circuit board containing about 30 chips.

COMMODORE PET (1977) Commodore's Personal Electronic Transactor is the first complete, full-featured home computer.

APPLE II (1977) The first personal computer able to generate color graphics includes a keyboard, power supply, 4K of memory, two game paddles, and a demo cassette. It sells for US $1 298.

MOTOROLA DYNATAC 8000X (1983) Motorola is the first to receive Federal Communications Commission approval for handheld mobile phones in the U.S. market.

NEC ULTRALITE (1988) Japan-based NEC Corp. launches the first MS-DOS–based portable computer in a notebook size. The product's major problems lead to its ultimate failure, but it opens the door for the ensuing laptop revolution.

POWERBOOK (1991) Apple Computer replaces its Macintosh Portable—the company's first battery-powered portable personal computer—with the PowerBook, and strikes gold.

SMARTPHONE (1996) Nokia introduces the first Web-enabled "smartphone," the 9000 Communicator mobile phone with full PDA functionality. It becomes the world's best-selling "office computer in your pocket."

BLACKBERRY (2002) Research in Motion releases the BlackBerry smartphone, which supports mobile calling, e-mail, text messaging, and Web browsing. The device soon earns the nickname "CrackBerry," because users depend heavily on the device.

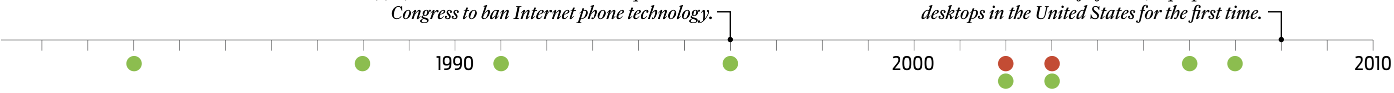
SKYPE (2003) Estonian-written software that allows users to make voice calls over the Internet for free, Skype is sold to eBay two years after the first public beta version is released. The software still allows voice over Internet telephony at no charge.

Wii (2006) Nintendo launches Wii, the first game console with a remote that detects movement in three dimensions.

iPHONE (2007) Apple introduces the iPhone, the first Internet-connected multimedia smartphone with a multitouch screen instead of keys. The phone's initial selling price of $599 is quickly lowered to $399, creating a consumer outcry.

In 1996, U.S. telecommunications companies ask Congress to ban Internet phone technology.

In the second half of 2008, laptops outsell desktops in the United States for the first time.



for commercialization. Like the brick, it could lead to a hundred different architectures. And we hope it will.

The other belief, that companies can gain access to early-stage R&D results at conferences, is even easier to dispel. What companies don't understand is that by the time their researchers read it at a conference, it's already too little, too late, and too limited. Too little because you see only the tip of the iceberg in the final results; too late because by the time it's in a paper the research has already been picked over for two years; and too limited—this is the most important point—*because you see only the path that resulted in the positive outcome*. You want to be engaged with the full research, not just the condensed summary, boiled down to 20 PowerPoint slides and 20 minutes. You miss all the paths that were taken that were not successful—and that alone is worth the price of admission, because knowing all the dead ends to avoid could save a company millions. These kinds of negative results never get published at conferences.

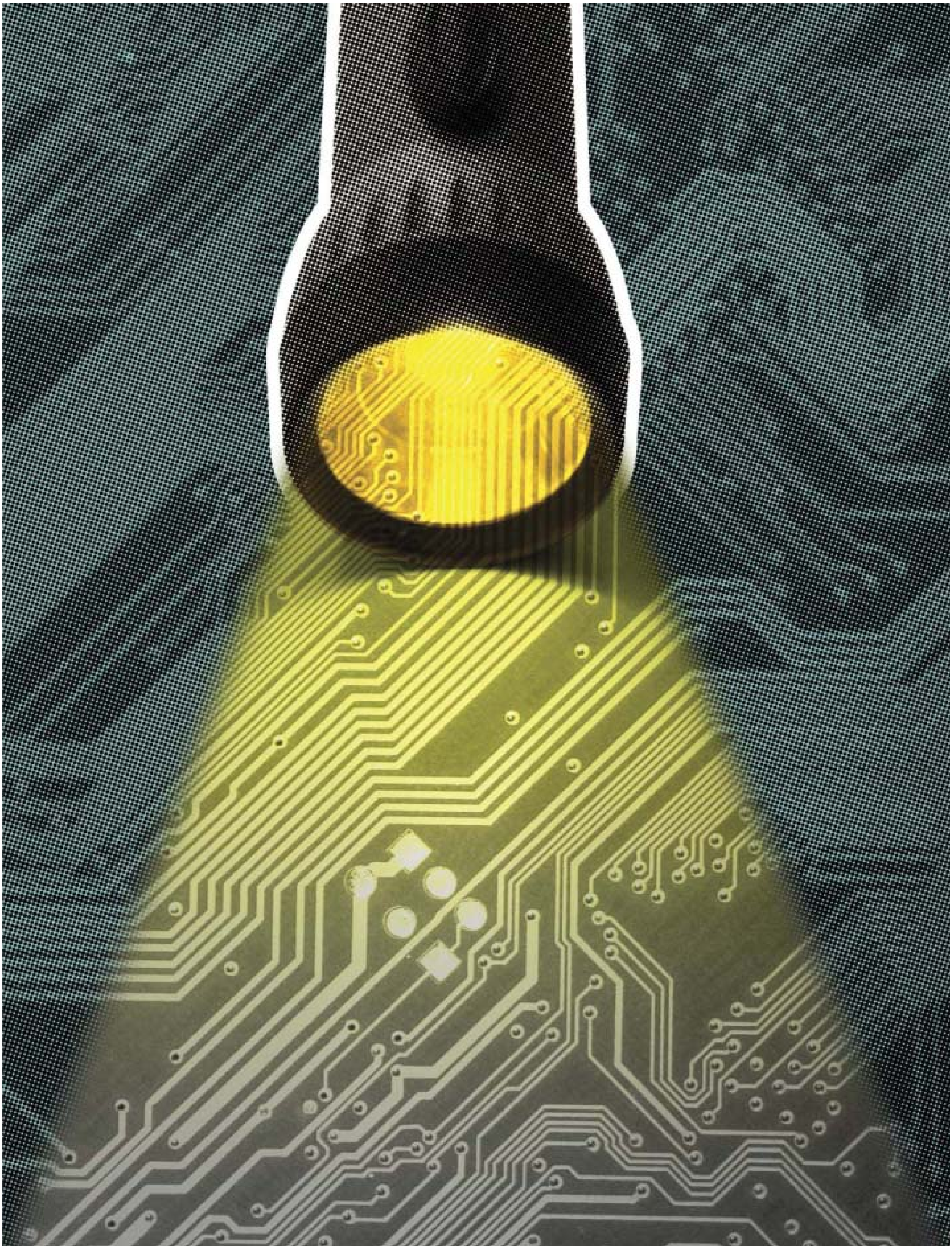
ANY COMPANY would be thrilled to achieve a 10 percent reduction in power between product generations. That number is typical of what evolutionary advances can accomplish at their best. Our national centers, by contrast, have enabled revolutionary and discontinuous advances in the last four to five years that haven't been seen for the last four or five decades.

With devices that perform far better than today's devices and yet consume a thousandth of the power, we could drastically reduce the consumption of power-hungry server farms that run today's critical Internet applications but consume enough power for a small city. We could realize "green" residential and transportation systems, a huge opportunity—or perhaps even a necessity, given that the world in 2050 may need 28 terawatts of power, compared with the 15 TW we use today. We might enable a new generation of personal electronics that turn our beloved iPhones into dinosaurs. We might build implantable medical devices that never need external charging, which means they wouldn't require invasive surgery just to change the battery. The breakthrough research in the centers may even enable radical concepts like "energy scavenging," where the chip survives entirely on power it draws from its surroundings—that is, from the movements of the person wearing the device.

But none of this will be possible until companies let go of their outdated notions and downright misconceptions.

The challenge today is in finding sources of disruptive scientific innovation. At Bell Labs and the Xerox Palo Alto Research Center, the seeds were planted for today's technology revolution. No one has the resources to replicate these today, but we believe we can make an alternative model of innovation, updated for the 21st century. It may very well be the key to an epochal change. ❑

CLOCKWISE FROM TOP LEFT: LENSCAP/ALAMY; COMPUTER HISTORY MUSEUM; NOKIA; ED UTHMAN



Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

MAY THE FORCE OF LIGHT BE WITH YOU

Light can exert enough force to flip switches on a silicon chip

BY HONG X. TANG

IN 1867, JULES VERNE IMAGINED spaceships propelled by the pressure of light. In 1871 James Clerk Maxwell predicted that such pressure actually existed, and in 1900 Pyotr Lebedev confirmed that prediction experimentally. But 109 years later, engineers have yet to find a practical use for this odd force.

Maybe it's just too weak. Light pressure equals beam power divided by *c*, the speed of light. A one-milliwatt laser pointer, therefore, presses its object with the force of 3.3 piconewtons. You could lift a penny with laser pointers—you'd just need 30 *billion* of them.

Yet such weakness becomes a strength when you're trying to nudge something of nanometer size and picogram mass. And the optics for directing light on such a scale already exist, in the form of miniature waveguides, couplers, and beam splitters, all of which are now routinely laid down on silicon-on-insulator substrates. We use these nanophotonic devices for optical processing. Why not also use them to exploit the ultra-light touch that light itself can provide? Why not use light as an actuator, reaching right into the guts of an integrated circuit to throw tiny switches, either to control electronic circuits or, better yet, to reroute light itself, and the data that it carries?

PHOTO ILLUSTRATION: BRANDON PALACIO; FLASHLIGHT: TOM GRILL/GETTY IMAGES; CIRCUIT BOARD: GETTY IMAGES

Such a marriage of nanomechanics and nanophotonics would bring us a giant step closer to optical chips. That's important, because light has a vastly wider bandwidth than electricity, which would enable it to get around the critical bottleneck in computing: the connections between processors. If light could act directly on circuit elements without first being converted to electricity, entire systems would run faster. You could imagine yoking together the multicore processors in a chip, making them run much faster and more efficiently than they can now. And if we really master the technology, optically controlled switches might ultimately supplant transistors, ushering in an era of all-optical computers. The resulting speedup would be stupendous, even by the standards of modern computing.

AT THE BEGINNING OF 2007, shortly after I joined the engineering faculty at Yale, I began to assemble a team to find ways of using light to drive silicon devices on a nanometer scale. We would be building on nearly two decades of research.

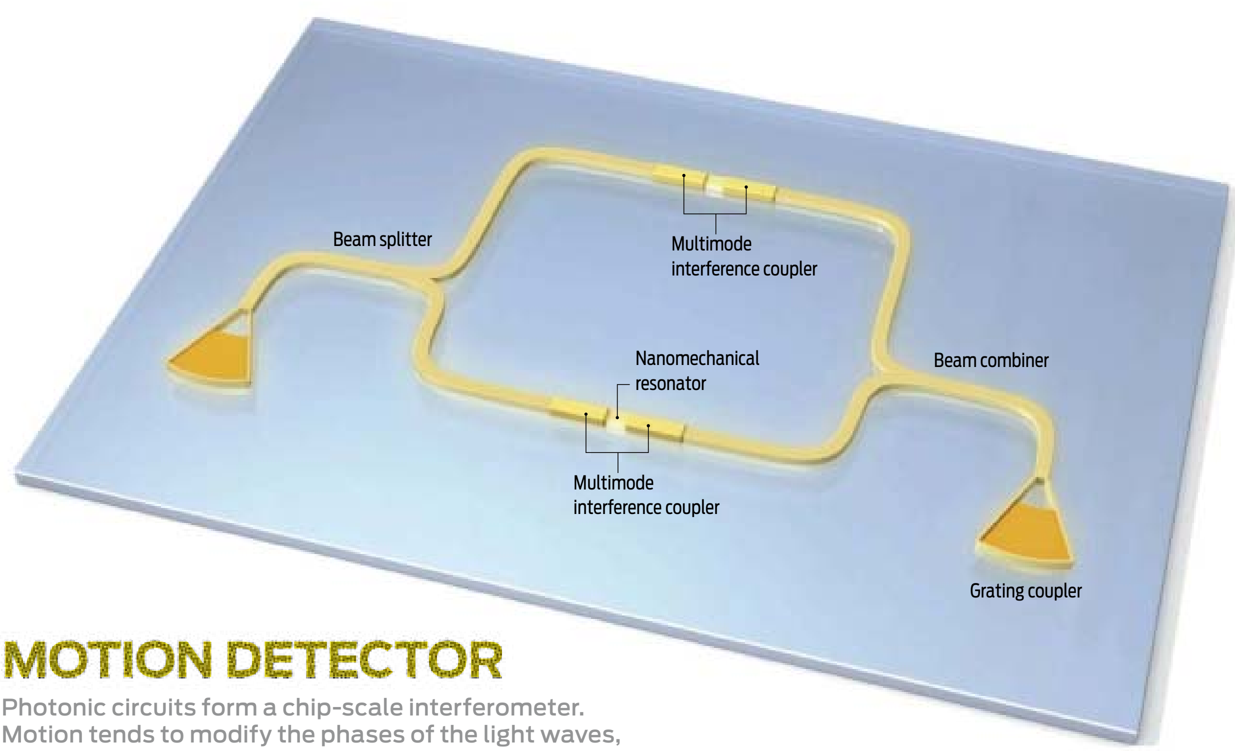
The quest had begun with microelectromechanical systems, or MEMS, which as the name implies are built in dimensions measured in micrometers. Today's engineers find it easy to build resonators—tiny tuning forks, basically—at that scale by simply embedding them within electronic circuits. The circuits drive the resonators through electromechanical coupling, typically by pairing them with electrical plates—one fixed, the other on the movable MEMS. In such a scheme, a current applied between the plates alters the gap between the plates, which changes the capacitance and further induces a current that oscillates in response to the motion of the plates. Basically, you feed in a steady current and get an oscillating one in response. In a cellphone, for example, such oscillators are used in filters, picking out the desired signal from the swath of frequencies your antenna pulls in.

The smaller the size of the plates and the gap between them, the faster the oscillation can be and thus the higher the frequency that can be isolated. At the nanoscale, such oscillators attain the frequencies needed for microwave communications. However, the high frequencies give rise to complex impedances both mechanical and electrical. Because the structure is so small, the impedances are often badly mismatched.

How badly are they mismatched? In high-frequency circuitry you normally want every component to have an impedance of around 50 ohms. At these tiny dimensions, though, you're going to end up with an impedance that's millions of times as great, which means that essentially none of your signal will get through.

Scaling down from MEMS to nanoelectromechanical systems, or NEMS, also brings on other, more fundamental problems. First, in the nano realm, the oscillators are so fast that the conventional electronic circuitry they work with can't keep up. Second, the oscillators' signals are so faint that they can get drowned out by the random noise that's endemic in any electronic circuit. A NEMS device can barely make itself heard over noise that's just one-thousandth as strong as what you'd find in a typical IC.

HERE IS WHERE PHOTONICS saves the day: Photons, unlike electrons, don't interact with each other and so are immune to cross talk. Moreover, because light has a much greater bandwidth, or carrying capacity, photonic signals can carry far more bits per second than electronic signals, while dissipating much less power. Also, it's easy to route



MOTION DETECTOR

Photonic circuits form a chip-scale interferometer. Motion tends to modify the phases of the light waves, allowing an inference of the resonators' oscillation.

ILLUSTRATION: EMILY COOPER

IEEE spectrum Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

photons on a chip; you just build a photonic wire—an optical nanofiber, essentially—on the top layer of the silicon in your chip. Finally, because we are in a purely optical domain, impedance is no problem.

The one big problem with those wonderfully noninteracting photons is that because they *are* noninteracting, there's no obvious way to use one of them to control another. So that's where the mechanical force of light comes in.

One such force, of course, is straightforward radiation pressure, like that of sunlight pressing on the sail of a future interstellar spacecraft. However, you can't harvest much momentum that way, and what you do collect will press in only one direction, which means you can't use it to both push and pull things, an important consideration (as we'll explain later). We therefore chose to exploit a different kind of optical force, one that often gets short shrift in university courses in optics. It's called the gradient optical force, and it arises from an incredibly evanescent bonding between light waves [see illustration, "A Light Touch"].

The gradient optical force was first used in the 1970s, in "optical tweezers," which were designed to manipulate molecules in a kind of optical microscope. That such a force might also be used on a chip was suggested in 2005 by John Joannopolous's group at MIT, together with Federico Capasso's group at Harvard. Using theories ultimately derived from Maxwell's equations, they concluded that it would be possible to generate a gradient force in the piconewtons, more than enough to get a nanometer-scale oscillator thrumming. The researchers based their calculations on a device involving two parallel waveguides, which are light-conducting channels engineered to confine waves of a given frequency in a beam so that it can travel through the guide with very little loss. Even though the two waveguides kept their beams separate, the bonding of the optical fields between the beams was surprisingly strong.

We found in 2007 that we could exploit the same principle to get oscillation using one waveguide rather than two. In the single-waveguide case, the optical field around the waveguide must be asymmetrical, in order to create the imbalance that's needed to exert a net force.

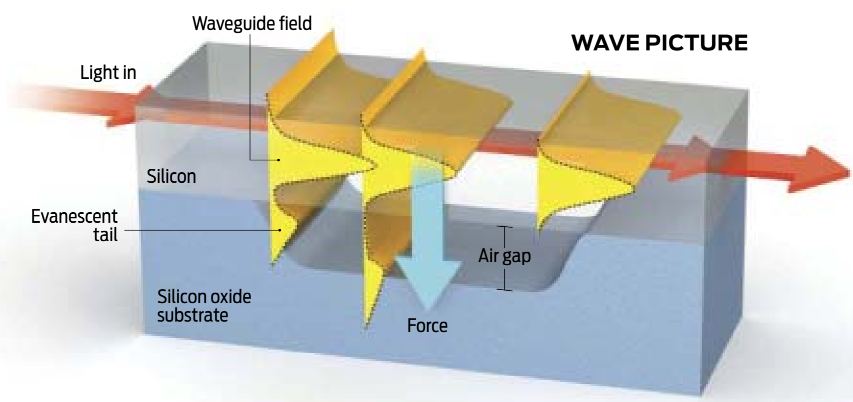
Happily enough, you can create such a waveguide on a chip by etching away the oxide under a slab of silicon to form a slab, which looks rather like a tiny bridge. This slab can swing back and forth—that is, it can oscillate. Now when light propagates along the waveguide, it encounters asymmetry: The bottom boundary of the waveguide is separated from the silicon oxide substrate by a thin air gap—a dielectric substrate—whereas the top boundary borders on air alone. So, as Maxwell's equations predict, a net optical force arises in the direction perpendicular to the waveguide. The force is substantial—well, for our purposes at least.

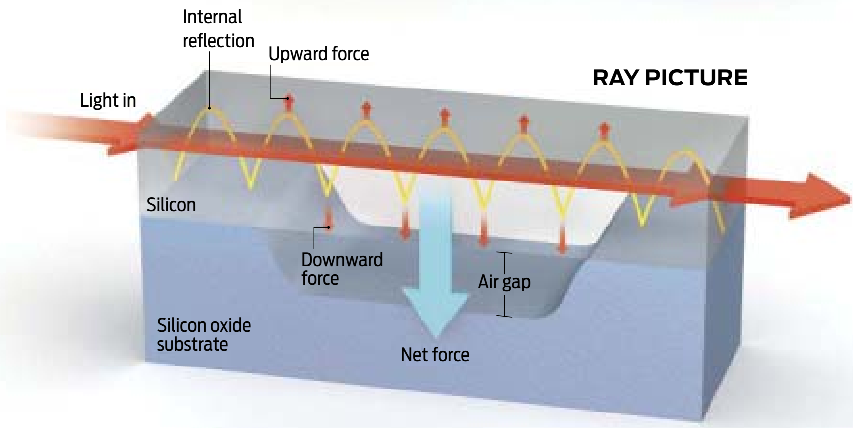
It may be easier to visualize the physics by means of ray optics. In a symmetric waveguide, the high refractive index in the middle guarantees that the light rays will bounce back and forth equally on the top and bottom surfaces of the guide. This perfect symmetry yields no net optical force. However, when a silicon substrate is present along the bottom of the waveguide, the rays are more strongly attracted toward the bottom surface, because the refractive index of the substrate—1.45—is higher than air's 1.00. So think of it like this: The rays drag more forcibly along the bottom surface, applying a net force to the substrate.

A LIGHT TOUCH

There are two ways to picture the gradient optical force exerted in a waveguide. According to Maxwell's equations [top illustration] the asymmetry between the open air on top of the waveguide and the thin air gap underneath will distort the optical field [yellow], creating a downward force.

Or you can resort to ray optics [bottom]. While in a conventional waveguide the light reflects equally strongly off the top and bottom edges, here the proximity of the silicon oxide substrate, with its relatively high refractive index, causes the rays to drag. They therefore push down harder than they push up. *ILLUSTRATION: EMILY COOPER*





ieee spectrum Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

So we've got ourselves a force and a mobile waveguide with which to harvest it. The next question is how to get light into the chip. It isn't easy to generate photons in silicon, so they have to be fed to the silicon waveguide from an external laser, through standard optical fibers. That scheme turns out to be harder to carry out than it might seem.

Silicon waveguides are normally made just several hundred nanometers wide, so that at a wavelength of 1550 nanometers, a telecommunications standard, they'll support only a single optical mode—a stable pattern that's guided within the waveguide's structure so that the signal doesn't disperse. A standard optical fiber, however, is dozens of times as wide. When you hook such a waveguide to such a fiber, the modes won't match up properly, and only about 0.1 percent of the light's power transfers. It isn't enough.

There are various tricks you can use to match the waveguide to the fiber, but because these tricks involve micromachining to an extremely high tolerance and at a very small scale, it is quite difficult. And the operations must be performed on each device, one at a time.

We found a better solution. In his lab at Caltech, Axel Scherer had developed some couplers for precisely this purpose: matching the light wave in an optical fiber to one in a silicon waveguide. Such couplers consist of optical gratings that are directly fabricated on silicon. Michael Hochberg, then a graduate student in Scherer's lab, began collaborating with us in 2007 and walked us through the design of these couplers.

This grating coupler has a horn structure with openings at either end: a large one that connects to the optical fiber and another with a cross-sectional area just one-thousandth as big that connects to the waveguide. In this scheme, the fibers are aligned from the top of the wafer, and you can test hundreds of devices repeatedly. You can provide light on demand, piping it throughout the chip, bending, splitting, and recombining it at will. And because the coupler ensures that the light goes from the waveguide to whatever device it is driving on the chip, very little of the light leaks out because of diffraction.

Mechanical devices require proper anchoring so that they can oscillate for a long time without losing too much energy. However, you'd rather not have such an anchor in a photonic design because it would tend to disturb the guided light wave, causing photons to scatter and thus be lost. We had to come up with several approaches to overcome this problem.

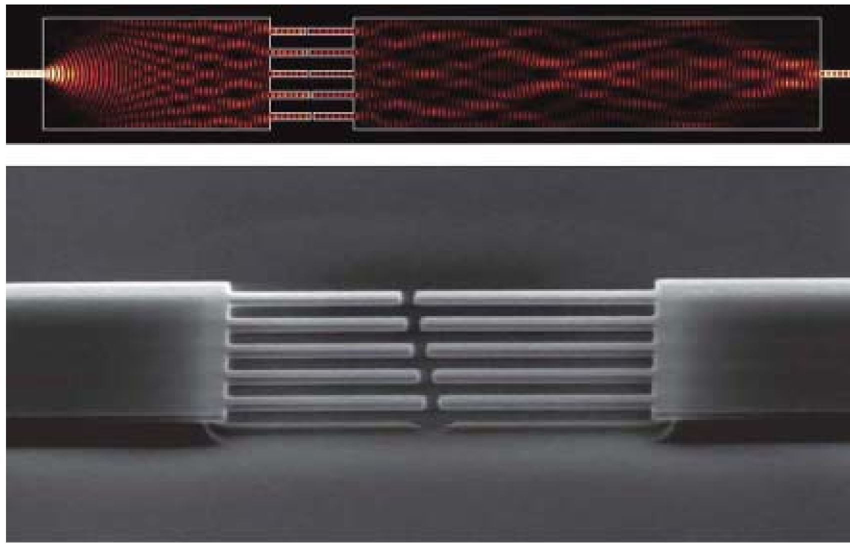
First, we built an interface between the stationary waveguide and the mobile waveguide, providing rigid mechanical support and thus focusing the light waves. This structure is called a multimode interference coupler, and by working effectively as an in-plane lens, it keeps the loss of light below 1 decibel.

Next, we built a nanomechanical interferometer on a silicon chip. We wanted this device not only because its ability to detect motion very sensitively is inherently useful, but also because it would help us keep track of what's going on in the various nanomechanical systems we devise.

We built a Mach-Zehnder interferometer, which splits one light beam evenly into two [see illustration, "Motion Detector"]. One beam goes through a moving part of the device we're building—say, the mobile, bridgelike part of the oscillator we described above. The other beam serves as the reference. A movement of the chip will cause the mobile part to vibrate, changing the effective refractive index through which the light is traveling and thus shifting its phase with respect to that of the reference beam. When the two beams recombine, their waves will interfere, forming a pattern from which we infer the degree of movement.

Our on-chip interferometer can measure movement to a sensitivity of 2×10^{14} meter, in a frequency range around 1 hertz. The spring constant of our devices—the measure of stiffness—is typically in the range of 1 to 10 newtons per meter. Therefore, the force sensitivity is on the order of 0.02 piconewtons, enough to resolve the gradient optical force—if it exists.

By the end of 2007 we had designed and redesigned the device—complete with interferometer, grating, and support coupler—many times. We were getting a reasonable amount of light through the measurement system, and almost all the components were in place. Yet even so, we still did not see the effect of the optical force. At this moment, our group suddenly received an infusion of new talent: Mo Li, a longtime friend and collaborator from my days as a graduate student at Caltech, where we'd both worked under Michael Roukes, a pioneer of nanoelectromechanical systems. We also welcomed



HARP STRINGS

A single photonic bus feeds 10 nanoscale cantilever waveguides, each of a different length and thus vibrating at its own rate, like a harp string. Each cantilever can register a deflection of just a fraction of the diameter of an atom. Together they can measure complex patterns of motion. *ILLUSTRATION: COURTESY OF MO LI, WOLFRAM PERNICE, TANG GROUP*

Wolfram Pernice, a gifted photonic device designer who, after obtaining his Ph.D. from Oxford, joined us as a postdoctoral fellow.

By January 2008, Li had put together many complete chips with movable beams, but he could observe no evidence of the optical force in the vibration of those beams. It became a blind game. We couldn't know whether the force was absent or simply undetected.

So we decided to separate the two problems of generating and detecting the optical force. To confirm that the interferometer was actually working, we actuated the device through the brute-force method of putting a sample chip on a piezoelectric disk, energizing the disk, and mechanically shaking the silicon photonic interferometer. The interferometer clearly indicated mechanical resonance.

Pernice worked around the clock and came up with new designs for the grating couplers and multimode interference couplers that allowed for much lower transmission losses, and therefore much greater sensitivity. Finally we were able to detect nanomechanical resonance at 10 megahertz without resorting to the piezodisk shaker. We therefore determined that the resonance had been induced purely by optical force.

The signal strength was enormous. With a slight increase of laser intensity, the beam resonance rose markedly, in nonlinear fashion. This result suggested that we could produce a force that was actually greater than we needed to move the picogram parts that make up a nanomachine. The actuation turned out to be just as efficient as that commonly used in MEMS devices, such as those that include inertial sensors of the kind used as gyroscopes. Of course, our NEMS system worked at a far higher frequency.

The next question involved the *sign* of the force. Just as electrically charged objects can either attract or repel each other, depending upon the sign of the charges, theory predicts that the gradient optical force should also be either attractive or repulsive, depending on the relative phase of the interacting light waves. Yet our devices exhibited only the attractive force. In other words, they were "monopolar."

To create a bipolar device, we went back to the original MIT/Harvard theory, which considers the optical forces that come into play between two opposing waveguides. As two waveguides approach each other in parallel, that is, side to side, the theory predicts that the waves from each waveguide will overlap to form a bonding (symmetric) mode or an antibonding (asymmetric) mode. When the waves are in phase, the optical force is attractive; when they are out of phase, the force is repulsive.

To demonstrate these effects, we developed a butterfly-shaped circuit having two "wings." In the left wing, the light is split equally into a top and a bottom arm, with the one on top having a longer waveguide to delay the light. This delay shifts the phase of the light waves. We then directed waves from each arm to recombine at the center of the butterfly, where the two suspended waveguides formed nanomechanical structures. If the light waves are out of phase at this point, the recombination should produce a repulsive effect. As for the identical right wing of the butterfly, its purpose is to make the whole structure symmetrical. That way, when the light proceeds into the right wing, the phase can be reversed, and the relative phase difference between both wings is maintained.

If the optical path difference between the two arms is an even multiple of the half wavelength, the waves will arrive at the nanomechanical devices in phase. The result should be an attractive force. If the delay is an odd multiple, you should get a repulsive force.

Everything worked according to theory. Better still, we found that by adjusting the relative phase of the interacting waveguide, we could tune the force from repulsive to attractive, or vice versa. (In practical experiments, we adjust the wavelength of the light source. This is equivalent to adjusting the delay phase in the wings.) In this fashion we have demonstrated all the predicted properties of the gradient optical force. Even more important, because this bipolar force can either push or pull, it allows us to manipulate components in both directions. We thus have complete control.

We achieved this push-me-pull-you control in late 2008. Now, instead of being limited to setting a vibrating beam in motion, we can push a nanomechanical lever in one direction—for instance, to open an optical switch and then pull it back again, closing the switch.

THE WAVEGUIDE BEAMS PROVE that optical force can be used to throw switches inside silicon optical circuits. Now we are extending that principle to other NEMS devices. Recently we embedded a series of cantilevers in photonic circuits, got all the cantilevers to resonate, and detected all their resonances at once [see illustration, "Harp Strings"]. We did all this by sending a light beam via a waveguide, which branched off into subsidiary channels.

With this setup we can integrate many sensors on a single photonic bus. Such a design is fully compatible with the standard wafer-scale processes used to fabricate chips, so large arrays could be mass-produced in a straightforward and low-cost way.

The light-force interaction brings NEMS device development to a true circuit level, making possible all sorts of applications. Such a device would detect signals so faint that it could measure the weight of just a sprinkling of molecules on top of it. It truly would weigh the molecules, not just detect them, because it would measure the change in the frequency of the resonator.

Further out, we expect to use the nanomechanics and nanophotonics on a given chip to achieve dual-mode sensing—that is, an optical spectrometer and a resonant mass sensor. If you have a fluorescent molecule, you can use the spectrometer to determine what the molecule is by its color, and you can use the resonant mass sensor to tell how many molecules there are.

Of course, you could also use light pressure to process RF analog signals better than is possible today, by combining optical and mechanical filtering. Optical filtering can, for instance, let only microwaves through, whereas mechanical filtering—using the oscillating beams we discussed earlier—can filter at a lower frequency range.

We can imagine using optical force to reroute light on the fly, allowing a photonic circuit to perform at a blindingly fast speed, far beyond anything that electronic controls can manage. This capability would go a long way toward realizing the dream of an all-optical computer, able to exploit the immense bandwidth of light to its fullest. Maybe if you left your hard drive at home, you could read it at a comfortable rate over the Internet—if we're still using hard drives by that time!

The most intractable bottleneck in today's high-end computers comes from having to use electronic signals to control photons. The sky will be the limit when we can at last use light to steer light. ❑

TO PROBE FURTHER To learn more, go to http://www.eng.yale.edu/tanglab.

Spectrum | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

Infoglut

Continued from page 27

Others have merely limited the number of messages a person may send in a day. One of the most sophisticated systems for influencing e-mail usage is Attent, from Seriosity, based in Palo Alto, Calif. It works by charging "postage" to send an e-mail, paid in a virtual currency denominated in "serios." The more urgent your e-mail, the more serios you attach to it; the recipients can then reuse the serios to send their own messages. Research into such systems is ongoing, and the opportunities for refinement include varying the postage according to parameters like the number of recipients, the recipient's organizational role (a senior manager might charge more postage for his or her attention), and the length of the message.

What about shielding employees from interruption? Many engineers secure thinking time on their own by working odd hours (say, coming in at 6 a.m.). A more structured approach is scheduling quiet time, an experiment described in detail by Leslie Perlow, a professor at Harvard Business School, in her book *Finding Time* (Cornell University Press, 1997). At a Fortune 500 company that manufactures computing hardware, she blocked out three mornings a week for the engineers in a design team to work without interruption, posting signs during the quiet periods to remind them of this commitment. She reported that the policy led to faster completion of the design project as well as a less harassing work environment.

In 2007 and 2008 Intel conducted a pilot of this methodology, albeit for only one morning per week, with a team of 300 design engineers and their managers. Results were encouraging: In surveys, 45 percent of respondents said they found the methodology effective as it was, and 71 percent recommended that Intel extend it to other groups, possibly with some modifications. People applied the quiet hours in different ways. We had expected that the quiet hours would be most useful for the designers, but even people in support roles benefited from having one morning a week when they could catch their breath, plan, and deal with the accumulation of tasks that were not related to their primary roles. Following the pilot, the company has gone on to try the approach with other groups.

Companies can also institute what has come to be called a Zero E-mail Day. The catchy name is in fact a misnomer: The idea isn't to ban e-mail on a given day. Rather, it's an attempt to break the e-mail addiction by getting everyone in a work group to agree to collaborate on the chosen day, by walking across the aisle, talking to coworkers, and solving problems in real time, rather than shooting an e-mail to someone just two cubicles away.

The most successful Zero E-mail Day program I know of was undertaken at PBD Worldwide Fulfillment Services, a company in Alpharetta, Ga., whose business involves warehousing products and filling orders. In this case, the CEO made it clear that he was passionate about the project. The results included enthusiastic employees, delighted customers, and a significant reduction in e-mail volume during the rest of the week.

An obvious next step would be to enlist technology to help prevent interrupting people at the wrong time. Development along these lines is happening in various quarters. Microsoft Research has prototyped a tool called Priorities, which analyzes incoming messages to predict their criticality, examines the recipient's current activity, and takes action accordingly. A message deemed to be urgent from a sender known to be *Continued on page 48*

ieee spectrum Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

Infoglut

Continued from page 46

important to the recipient may trigger an immediate alert or be forwarded to the recipient's mobile device, while delivery of a less urgent message may be deferred.

Another simple, automated approach is just-in-time coaching. Clearly, many breaches of e-mail etiquette are the result of simple oversight: hitting Reply All instead of Reply, forgetting to attach a file, or leaving the subject line blank. These situations are easily detectable by software, such as the E-mail Effectiveness Coach, a homegrown tool Intel had used in the early years of this decade. This ran in the background, and whenever the user clicked Send, it checked the message for etiquette problems; if one was found, a friendly alert popped up to give the user the opportunity to correct it. For example, when a user composed a message referring to attachments and the tool noticed that no files were attached, the alert would say, "Did you notice your message contains a reference to an attachment, but there are no attached files?" and then provide the option to abort the Send operation and go back to fix the problem.

There is also a growing body of products that automate the classification and handling of messages in the in-box of the individual user. A good example is ClearContext Professional, from San Francisco–based ClearContext Corp. It analyzes a user's e-mail history to identify the important messages and correspondents; provides in-box views that sort and color-code messages by importance, topic, and so forth; and places messages, contacts, meetings, and tasks into one contextual framework where all things related to a given project are presented and managed together. No one tool is best for everyone, but there is enough choice that anyone can find a tool that matches his or her work style.

WHILE MANY PERSONAL TOOLS EXIST, IT'S surprising how little has been done at an organization-wide level to fight a problem as big as information overload, considering that the cost of fixing it is trivial compared to the potential benefit. This failure may in part be due to the critical role of electronic communications in today's workplace and beyond; many people feel horrified by the thought of any interference with the free flow of information.

This thinking, however, is wrong. In reality, there is a continuum between doing nothing and preventing all communication. We need to discover the optimal balance of communication and thinking time, human interaction and concentration, useful messages and junk. Convincing individuals and organizations to actually do something is not easy, but a slowly growing number of cases show that people can manage information with good results. What is most needed are managers with the vision and leadership to move their organizations to make the changes.

So, whatever organization you're in, try to identify ways to mitigate information overload—not just for yourself, but for your entire organization. Try to convince your coworkers and your managers to create a serious program, either using the tools and approaches that are already out there or inventing new ones. The main thing is to renounce the attitude that this is how things are and nothing can be done.

A lot can be done; let's do it. ❑

TO PROBE FURTHER Visit the Information Overload Research Group's site at http://www.iorgforum.org. Nathan Zeldes's personal Web site is http://www.nzeldes.com.

THE UNIVERSITY OF UTAH

The Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, seeks applications to fill at least two tenure-track positions at the assistant, associate or full professor level for an interdisciplinary research cluster in ***Micro and Nanosystem Integration and Packaging***.

We are particularly interested in candidates with background in electronic micro/nano-system integration and packaging, biocompatible materials and packaging, solid state devices, reliability, testing, and micro/nano system modeling and simulation. Information on department research activities and curricula may be found on the web at **www.ece.utah.edu**. The web site also has information on two more positions available in the department. Information on the College of Engineering can be found at **www.coe.utah.edu**. Successful candidates will conduct research with tenure track appointments in the Department of Electrical and Computer Engineering, but may also be appointed in other departments such as Materials Science, Bioengineering or Mechanical Engineering. Suitable candidates may be considered for joint appointments with the College of Science or the Medical School at the University of Utah.

These positions are part of the Utah Science, Technology and Research Initiative (USTAR), which was funded by the Utah State Legislature to attract focused teams of outstanding researchers who have the potential to help build major research programs and create new technology that can ultimately lead to commercial products and/or new industries for Utah. The USTAR initiative is also supporting a new interdisciplinary building which will house a new nanofabrication laboratory and characterization facilities that will cater to solid state devices, MEMS, sensor and packaging research and development, as well as the handling of biomedical samples. The building will facilitate communication for researchers such as the ones hired under this solicitation, from engineering, sciences and the medical school, as well as offering lab access for selected industrial stake holders. Information on the USTAR initiative can be found under **www.ustar.utah.gov**. Candidates for this initiative should have a demonstrated track record of successful, funded projects and an interest or track record in technology commercialization, entrepreneurial or industrial experience.

The positions are also associated with and partially supported by the ***Fraunhofer Institute for Reliability and Microintegration IZM***, and leverage a strong collaborative and international research program with a Fraunhofer IZM branch laboratory in Utah. Fraunhofer support includes in-house access to Fraunhofer infrastructure, know-how, and resources. Selected positions may be associated with joint Fraunhofer appointments, possibly at a center director's or co-director's level.

Résumés with names, contact information for at least three references, and statements for research and teaching goals should be sent to Ms. Debbie Sparks, USTAR Faculty Search Committee, University of Utah, Electrical and Computer Engineering Department, 50 South Central Campus Drive, Room 3280, Salt Lake City, UT 84112-9206. Email applications are accepted at **dsparks@ece.utah.edu**. Applications will be reviewed starting September 1, 2009, and will be accepted until the positions are filled.

Faculty responsibilities include developing and maintaining an internationally recognized research program, effective classroom teaching at the undergraduate and graduate levels, and professional service. Applicants must hold a Ph.D. by the time of appointment. The University of Utah values candidates who have experience working in settings with students from diverse backgrounds and possess a strong commitment to improving access to higher education for historically underrepresented students. The University is an AA/EO employer, encourages applications from women and minorities, and provides reasonable accommodations for known disabilities of applicants and employees.

microfab University of Utah · inip · Fraunhofer Institut Zuverlässigkeit und Mikrointegration · USTAR

THE HONG KONG POLYTECHNIC UNIVERSITY

香港理工大學

The Hong Kong Polytechnic University is the largest government-funded tertiary institution in Hong Kong, with a total student headcount of about 28,090, of which 14,260 are full-time students, 10,050 are part-time students, and 3,780 are mixed-mode students. It offers programmes at Doctorate, Master's, Bachelor's degrees and Higher Diploma levels. The University has 27 academic departments and units grouped under six faculties, as well as 2 independent schools and 2 independent research institutes. It has a full-time academic staff strength of around 1,300. The total consolidated expenditure budget of the University is in excess of HK$4 billion per year.

DEPARTMENT OF BUILDING SERVICES ENGINEERING

The Department of Building Services Engineering (BSE) is looking for an enthusiastic, established and suitably qualified academic in electric energy systems and technologies or lighting engineering to further advance our excellent reputation through research, teaching and publications.

The Department is one of the constituent departments of the Faculty of Construction and Land Use. It offers a full range of programmes leading to the awards of Doctor of Philosophy, Master of Philosophy, Master of Science and Bachelor of Engineering. It has an undergraduate student body of over 400 full-time students, over 200 part-time students and a postgraduate student body of over 350. BSE currently has a total full-time academic staff establishment of 32 engaging in a wide range of active research areas including but not limited to HVAC, Building Energy Studies, Building Environment, Fire and Safety Engineering, Electrical and Lighting Systems and Facility Management. The University and the Department highly encourage a multi-cultural environment concerning both staff and student mix. For more information regarding the Department, please visit the departmental homepage at http://www.bse.polyu.edu.hk.

Assistant Professor in Electrical Engineering

The appointee will be required to (a) undertake teaching up to master's degree level in Electrical Services and related subjects; (b) be an active researcher in terms of quality publications, external research grants and applied research; (c) supervise undergraduate and postgraduate research/design projects; and (d) provide leadership in developing a programme, a subject discipline, or a research programme, etc.

Applicants should (a) have a PhD degree in an appropriate discipline related to Electrical Engineering; (b) have a good publication record in refereed journals and good potential in bidding for research grants; (c) preferably have professional qualification in a recognised and relevant professional body; and (d) have several years of research experience in an electrical discipline, preferably related to electric power or lighting.

A start-up package will be provided for initiation of a new research programme or research area.

Remuneration and Conditions of Service

Salary offered will be commensurate with qualifications and experience. Initial appointment will be made on a fixed-term gratuity-bearing contract. Re-engagement thereafter is subject to mutual agreement. Remuneration package will be highly competitive. Applicants should state their current and expected salary in the application.

Application

Please submit application form via email to hrstaff@polyu.edu.hk; by fax at (852) 2764 3374; or by mail to **Human Resources Office, 13/F, Li Ka Shing Tower, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong**. If you would like to provide a separate curriculum vitae, please still complete the application form which will help speed up the recruitment process. Application forms can be obtained via the above channels or downloaded from http://www.polyu.edu.hk/hro/job.htm. **Recruitment will continue until the position is filled**. Details of the University's Personal Information Collection Statement for recruitment can be found at http://www.polyu.edu.hk/hro/jobpics.htm.

Your Career Starts With Us!

RWTH AACHEN UNIVERSITY

Full Professor (W3) Theory of Electrodynamics

Faculty of Electrical Engineering and InformationTechnology

We are seeking qualified applicants for teaching and research in the area of Theory of Electricity with special emphasis in one or more of the following areas:

- Maxwell's Theory with all extensions as well as classical and quantum dynamics,
- Theory, characterization, models and basic circuits of ultra-small and ultra-fast electronic, microwave and optoelectronic devices,
- Multi-physics and multi-scale simulation of nano- and heterogeneous -systems including the coupling between electrical, thermal, mechanical and optical sub-systems.

The starting date of the professorship is Juli 1, 2011. A Ph.D. degree is required; additionally, Habilitation (post-doctoral lecturing qualification), an exemplary record of research achievement as an assistant / an associate / a junior professor or university researcher and/or an outstanding career outside academia are highly desirable. Ability in and commitment to teaching are essential and it is expected that applicants will be able to teach in German after two to three years. In addition, significant experience in leading positions in industry or in renowned research organizations with activities that are relevant to industry are highly appreciated.

The application should include supporting documents to demonstrate success in teaching.
Please send a cover letter stating research aims and a CV to: Dekan der Fakultät für Elektrotechnik und Informationstechnik der RWTH Aachen, Prof. Dr. Michael Vorländer, Templergraben 55, 52062 Aachen, Germany. The deadline for applications is November 30, 2009.

The RWTH Aachen aims to increase the number of women in areas in which they are under-represented, thus women are strongly encouraged to apply. For further information please see: http://www.rwth-aachen.de/equality
The RWTH Aachen aims to integrate persons with disabilities, thus persons with disabilities are strongly encouraged to apply. For further information please see: http://www.rwth-aachen.de/disabilities

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY POSITIONS IN **Electronic and Computer Engineering**

The Department of Electronic and Computer Engineering invites applications for several faculty positions for the Assistant Professor rank. Applicants should have a PhD with demonstrated strength in research and commitment to teaching. We are particularly interested in qualified applicants with relevant research experience in areas related to solar cells and solid state lighting, bioengineering and RFIC design. However applications are also encouraged from candidates whose research programs are nontraditional and interdisciplinary and whose instructional programs will bring innovation to the curriculum. Outstanding candidates at the Professor and Associate Professor ranks may also be considered for these positions on an exceptional basis.

The Hong Kong University of Science and Technology is a truly international university in Asia's world city, Hong Kong, and its Engineering School has been consistently ranked among the world's top 25 since 2004. The high quality of our faculty, students and facilities provides outstanding opportunities for faculty to develop highly visible research programs. All formal instruction is given in English and all faculty members are expected to conduct research and teach both undergraduate and graduate courses. The Department has excellent computing resources, state-of-the-art teaching and research laboratories and currently has about 40 faculty members, 813 undergraduate students and 388 postgraduate students.

Starting rank and salary will depend on qualifications and experience. Fringe benefits including medical and dental benefits, annual leave and housing will be provided where applicable. Initial appointment at Associate Professor and Assistant Professor ranks will normally be on a three-year contract. A gratuity will be payable upon successful completion of contract. Reappointment will be subject to mutual agreement.

Applications including full curriculum vitae, list of publications, names of five referees addressed to: Professor Hoi Sing Kwok, Chairman of Search Committee should be sent by email to **eesearch@ust.hk**. Applications will be considered until all the positions are filled.

More information about the Department is available on the website **http://www.ece.ust.hk/**.

(Information provided by applicants will be used for recruitment and other employment-related purposes.)

Imperial College London

Institute of Biomedical Engineering
Centre for Medical Engineering Solutions in the Management of Osteoarthritis

Chair in Medical Device Engineering

Competitive Salary; the minimum professorial salary is £66,150 p.a.

Applications are invited for the Chair in Medical Device Engineering. The post is within the Institute of Biomedical Engineering, at Imperial College London, based in South Kensington.

The successful applicant will be expected to provide expertise in the development of devices for the early detection and monitoring of osteoarthritis, in particular in developing sensors, both for joint implants and "smart clothing" in order to assist treatment and rehabilitation. The post will require experience of sensor development and the deployment into medical devices.

Details of this appointment can be obtained at http://www3.imperial.ac.uk/employment and online applications at http://www3.imperial.ac.uk/employment (select "Job Search"). Please quote reference number **EN20090717.**

Alternatively, you may email your application and CV to: **Maria Monteiro**, at m.monteiro@imperial.ac.uk, Telephone: +44 (0)207 594 5498.

Please contact Dr Rob Fenton, Research & Development Director, to discuss this post informally at: r.fenton@imperial.ac.uk

Closing date: 31 October 2009

Committed to equality and valuing diversity. We are also an Athena Silver SWAN Award winner and a Stonewall Diversity Champion.

جامعة كارنيجي ميلون في قطر
Carnegie Mellon Qatar
School of Computer Science
Qatar Foundation

Computer Science Faculty Positions

Carnegie Mellon University in Qatar invites applications for teaching-track positions at all levels in the field of Computer Science. These career-oriented renewable appointments involve teaching international undergraduate students, and maintaining a significant research program. Candidates must have a Ph.D. in Computer Science or related field, substantial exposure to Western-style education, outstanding teaching record and excellent research accomplishments or potential.

Specifically, we are seeking candidates with expertise in databases, data mining, web technology and human-computer interaction. Truly exceptional candidates in other areas also will be considered.

The position offers a competitive salary, foreign service premium, research seed grant, excellent international health coverage and allowances for housing, transportation, dependent schooling and travel.

Carnegie Mellon is internationally recognized as a leader in research and higher education. In 2004, the university established itself in Education City, a state-of-the-art campus that is home to six top universities. Collaboration opportunities with internationally-known researchers and world-class businesses are abundant.

For further information or to apply, visit
http://www.qatar.cmu.edu/cs/positions/

THE UNIVERSITY OF UTAH

The Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, seeks applications to fill two tenure-track positions at the assistant professor level. Outstanding applicants with significant experience may also be considered at the associate or full professor level.

We are particularly interested in candidates with expertise in electromagnetics and communications. Information on department research activities and curricula may be found on the web at **www.ece.utah.edu**. The web site also has information on additional positions available for the Utah Science, Technology and Research Initiative. Faculty responsibilities include developing and maintaining an internationally recognized research program, effective classroom teaching at the undergraduate and graduate levels, and professional service.

Résumés with names and contact information for at least three references should be sent to Ms. Debbie Sparks, Faculty Search Committee, University of Utah, Electrical and Computer Engineering Department, 50 South Central Campus Drive, Room 3280, Salt Lake City, UT 84112-9206. Email applications are accepted at **dsparks@ece.utah.edu**. Applications will be reviewed starting September 1, 2009, and will be accepted until the positions are filled. Applicants must hold a Ph.D. by the time of appointment. The University of Utah values candidates who have experience working in settings with students from diverse backgrounds and possess a strong commitment to improving access to higher education for historically underrepresented students. The University is an AA/EO employer, encourages applications from women and minorities, and provides reasonable accommodations for known disabilities of applicants and employees.

the data

Forecast for Cloud Computing: Up, Up, and Away

Mobile data to increase 14-fold by 2014, much of it in the cloud

BY 2014, cellphones and other mobile devices will send and receive more data each month than they did in all of 2008. Three-fourths of the total will come from Internet access and nearly all the rest from audio and video streaming.

A big part of the increase in mobile data will come from cloud computing applications. Utility software (such as maps), will lead the way, followed closely by productivity tools (especially for sales, data sharing, and collaboration), then social networking and search. So predicts ABI Research, a telecom analysis firm in Oyster Bay, N.Y.

According to senior analyst Mark Beccue, the number of people subscribing to mobile cloud apps will rise from 71 million to nearly a billion by 2014. The firm defines an app as pertaining to the cloud when it "is no longer dependent on the device for data storage or processing power." Understood this way, much online-game playing doesn't take place in the cloud. Still, games will be the fifth most popular application area.

Asia will have by far the largest number of mobile cloud subscribers, but North America will pull in nearly as much in revenue, because high-paying enterprises will have a larger slice of the pie there.

But what will it be like for consumers? Beccue expects mobile banking to be big, and he says the cloud will also let us remotely "change thermostat settings, turn lights on or off, and record TV shows. [Security products maker] Schlage even has a keyless entry system that now lets you buzz someone through your front door via your phone." —*Steven Cherry*

IPHONE: APPLE


HFSS
TWELVE
Experience the high-performance computing (HPC) benefits of HFSS™v12, the industry-standard simulation tool for 3D full-wave electromagnetics. HFSS v12 is enhanced with breakthrough technology including a new parallel computing option that makes it possible to solve enormous electromagnetic field problems faster than you ever thought possible. HFSS v12 is further enhanced with new features such as volumetric meshing, curvilinear elements and mixed element orders. With HFSS you can calculate S-parameter derivatives with respect to variations in geometry, materials, and boundary conditions to evaluate design sensitivity and to speed optimization. Solve localized E and H-fields, currents, s-parameters and near and far radiated fields in your antennas, RF/microwave component designs, on-chip embedded passives, IC packages, and high-speed interconnects with even greater performance!
12
HFSS.COM
ANSOFT
A Subsidiary of ANSYS, Inc.