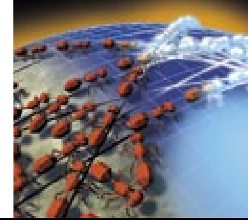
**Mind-Machine Meld**
Memory implants by 2030 p. 20

**Will They Think Like Us?**
Aim: 100 trillion instructions a second p. 12

**Swarm Smarts**
Lessons from the hive mind p. 40



# SCIENTIFIC AMERICAN REPORTS

SPECIAL EDITION ON ROBOTICS

www.SciAm.com

# YOUR FUTURE WITH ROBOTS

## How Smart Machines Will Change Everything

**Robots at Home**
Bill Gates's Future Vision

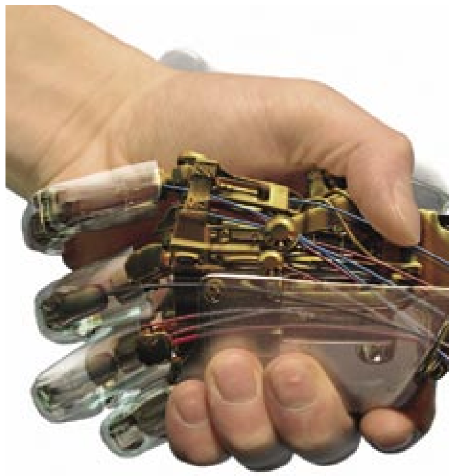
**Mind Control**
Command Robots by Thought

**Artificial Muscles**
Flexible-Polymer Power

Page Intentionally Blank

## letter from the editor

# Aye, Robot



KENN BROWN AND CHRIS WREN *Mondolithic Studios*

THEY HAVE BEEN part of our collective imagination almost since we began to set down words. Mechanical beings sparked to life in the myths of ancient Greece, the Middle East, China and the Nordic countries. Today we call them robots—from *robota,* meaning "drudgery" or "hard work" in Czech and related languages. As that name implies, so far these useful machines have been limited in their applications to the sorts of repetitive tasks best suited to automatons—tirelessly turning screw after screw in a factory assembly line, for instance.

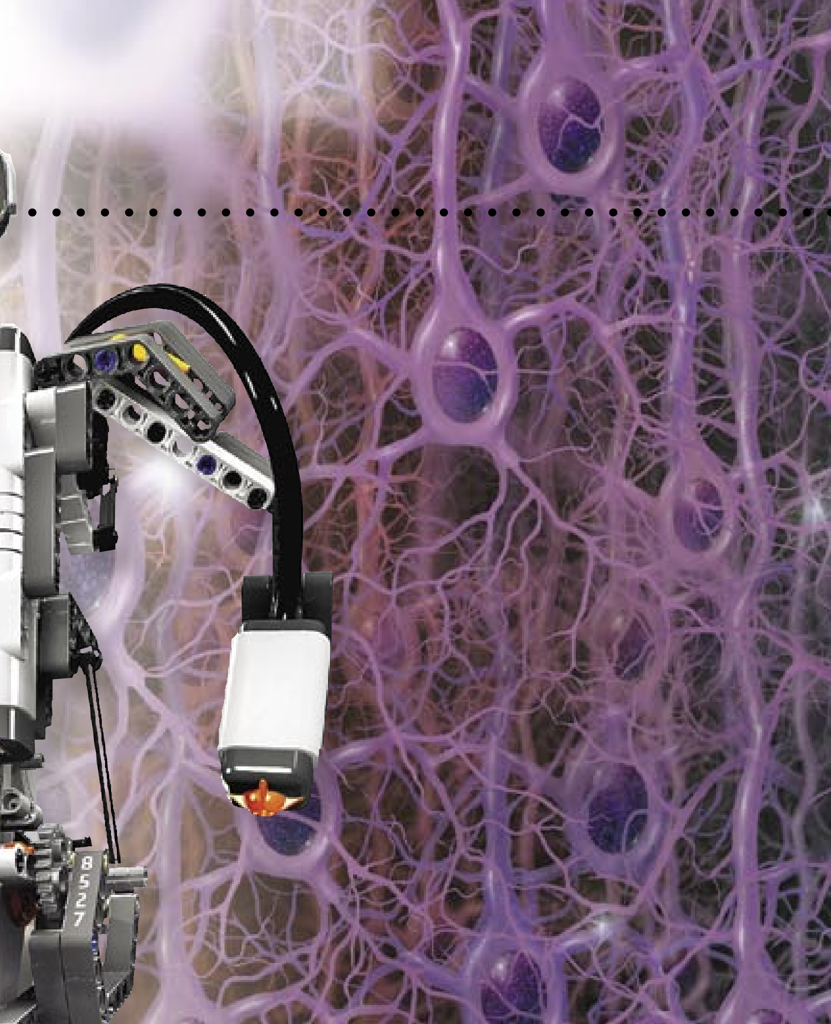
Now robots are beginning to enter our lives in much more personal ways. Already robo-vacuums such as the Roomba are easing housework, and digital pets such as Tamagotchis and the e-dog Aibo are serving as electronic companions. Experts envision far more in the short years ahead. Bill Gates writes in his feature article "A Robot in Every Home," starting on page 4, of nothing less than a transformation of domestic life. It is only a matter of time. After all, he adds: "Some of the world's best minds are trying to solve the toughest problems of robotics, such as visual recognition, navigation and machine learning." Two million personal robots were in use worldwide in 2004, and seven million more will be installed by this year, according to one estimate.

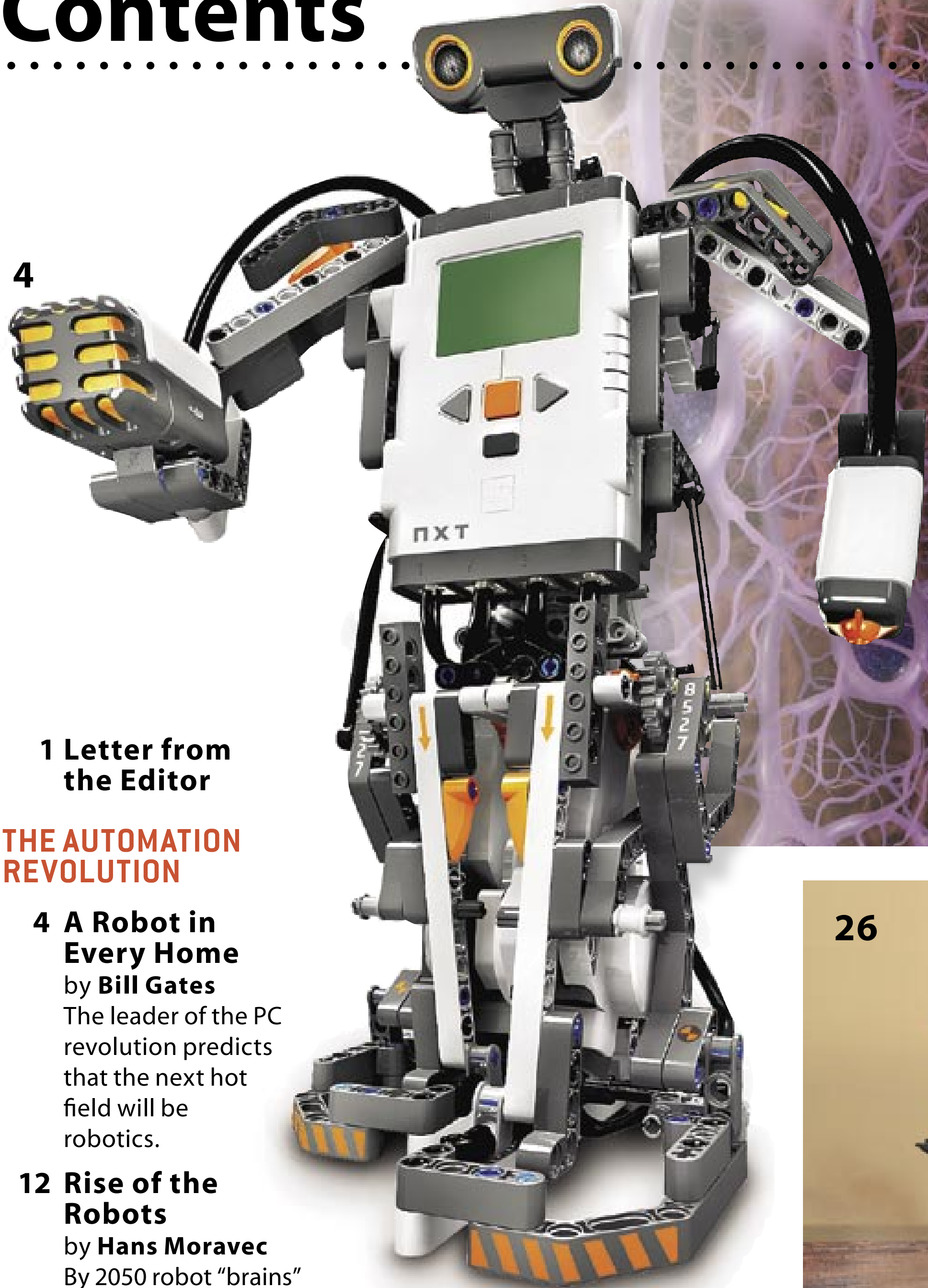
To expand further, they will require adaptive, complex processors, as Hans Moravec describes in "Rise of the Robots," beginning on page 12. By 2050 robot "brains" that execute 100 trillion instructions per second will start to rival human intelligence. Robots will also need to become more physically flexible and adaptable. In "Artificial Muscles," Steven Ashley describes springy polymers that could even produce power with movement. Turn to page 64.

At the same time that robots will be acquiring more human attributes, people will be adopting electronic implants to improve skills such as memory, according to "The Coming Merging of Mind and Machine," by Ray Kurzweil, starting on page 20. We will also be using the power of thought to direct machines, say Miguel A. L. Nicolelis and John K. Chapin in "Controlling Robots with the Mind"; turn to page 72. Indeed, the differences between maker and creation grow less distinct all the time.

Mariette DiChristina
Executive Editor
*Scientific American*
editors@SciAm.com

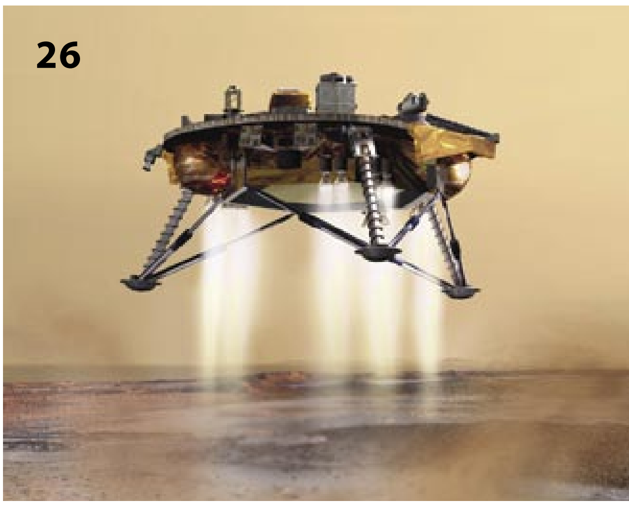
# Contents





4



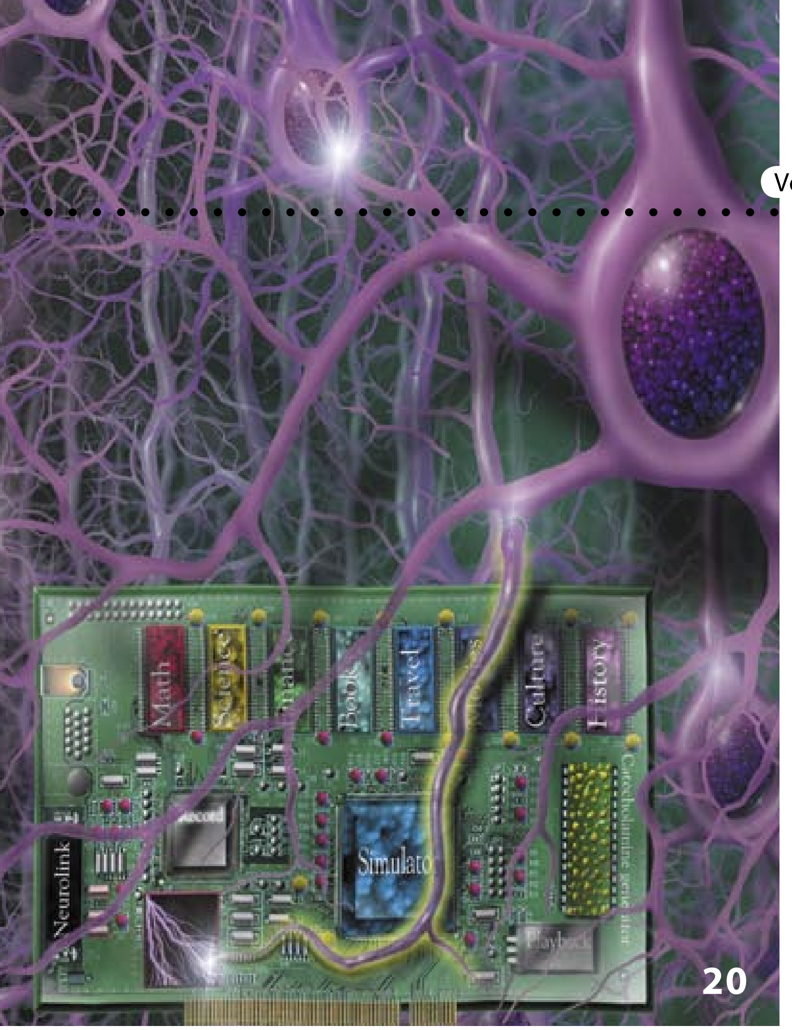


26



Cover image by Kenn Brown and Chris Wren, Mondolithic Studios
The articles in this special edition are updated from previous issues of *Scientific American*.

SCIENTIFIC AMERICAN
REPORTS
Volume 18, Number 1, 2008




20

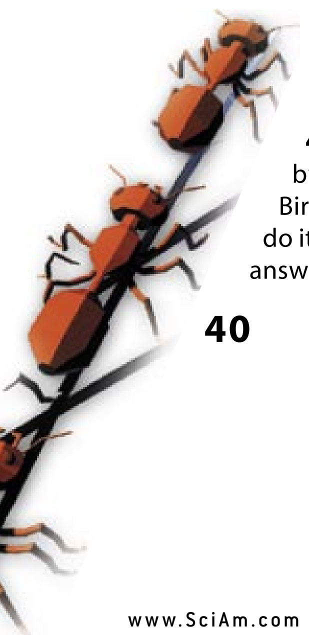
## MODERN MOBILITY



58

## STRENGTH IN NUMBERS





40

Scientific American Special (ISSN 1936-1513), Volume 18, Number 1, 2008, published by Scientific American, Inc., 415 Madison Avenue, New York, NY 10017-1111.  Canadian BN No. 127387652RT; QST No. Q1015332537. To purchase additional quantities: U.S., $10.95 each; elsewhere, $13.95 each. Send payment to Scientific American, Dept. RBT2008, 415 Madison Avenue, New York, NY 10017-1111. **Inquiries: fax 212-355-0408 or telephone 212-451-8442.** Printed in U.S.A.

# A ROBOT IN EVERY HOME

## The leader of the PC revolution predicts that the next hot field will be robotics

**By Bill Gates**

Imagine being present at the birth of a new industry. It is an industry based on groundbreaking new technologies, wherein a handful of well-established corporations sell highly specialized devices for business use and a fast-growing number of start-up companies produce innovative toys, gadgets for hobbyists and other interesting niche products. But it is also a highly fragmented industry with few common standards or platforms. Projects are complex, progress is slow, and practical applications are relatively rare. In fact, for all the excitement and promise, no one can say with any certainty when—or even if—this industry will achieve critical mass. If it does, though, it may well change the world.

Of course, the paragraph above could be a description of the computer industry during the mid-1970s, around the time that Paul Allen and I launched Microsoft. Back then, big, expensive mainframe computers ran the back-office operations for major companies, governmental departments and other institutions. Researchers at leading universities and industrial laboratories were creating the basic building blocks that would make the information age possible. Intel had just introduced the 8080 microprocessor, and Atari was selling the popular electronic game Pong. At homegrown computer clubs, enthusiasts struggled to figure out exactly what this new technology was good for. But what I really have in mind is something much more contemporary: the emergence of the robotics industry, which is developing in much the same way that the computer business did 30 years ago. Think of the manufacturing robots currently used on automobile assembly lines as the equivalent of yesterday's mainframes. The industry's niche products include robotic arms that perform surgery, surveillance robots deployed in Iraq and Afghanistan that dispose of roadside bombs, and domestic robots that vacuum the floor. Electronics companies have made robotic toys that imitate people or dogs or dinosaurs, and hobbyists are anxious to get their hands on the latest version of the Lego robotics system.

AMERICAN ROBOTIC: Although a few of the domestic robots of tomorrow may resemble the anthropomorphic machines of science fiction, a greater number are likely to be mobile peripheral devices that perform specific household tasks.


GRANT
WOOD

Meanwhile some of the world's best minds are trying to solve the toughest problems of robotics, such as visual recognition, navigation and machine learning. And they are succeeding. At the 2004 Defense Advanced Research Projects Agency (DARPA) Grand Challenge, a competition to produce a robotic vehicle capable of navigating autonomously a rugged 142-mile course through the Mojave Desert, the top competitor traveled just 7.4 miles before breaking down. In 2005 five vehicles covered the complete distance. And in November 2007 six vehicles completed a 60-mile course through a simulated urban environment in which they were required to merge with moving traffic, traverse busy intersections, avoid obstacles and find parking. (In another intriguing parallel between the robotics and computer industries, DARPA also funded the work that led to the creation of Arpanet, the precursor to the Internet.)

What is more, the challenges facing the robotics industry are similar to those we tackled in computing three decades ago. Robotics companies have no standard operating software that could allow popular application programs to run in a variety of devices. The standardization of robotic processors and other hardware is limited. Whenever somebody wants to build a new robot, they usually have to start from square one.

Despite these difficulties, when I talk to people involved in robotics—from university researchers to entrepreneurs, hobbyists and high school students—the level of excitement and expectation reminds me so much of that time when Paul Allen and I looked at the convergence of new technologies and dreamed of the day when a computer would be on every desk and in every home. And as I look at the trends that are now starting to converge, I can envision a future in which robotic devices will become a nearly ubiquitous part of our day-to-day lives. I believe that technologies such as distributed computing, voice and visual recognition, and wireless broadband connectivity will open the door to a new generation of autonomous devices that enable computers to perform tasks in

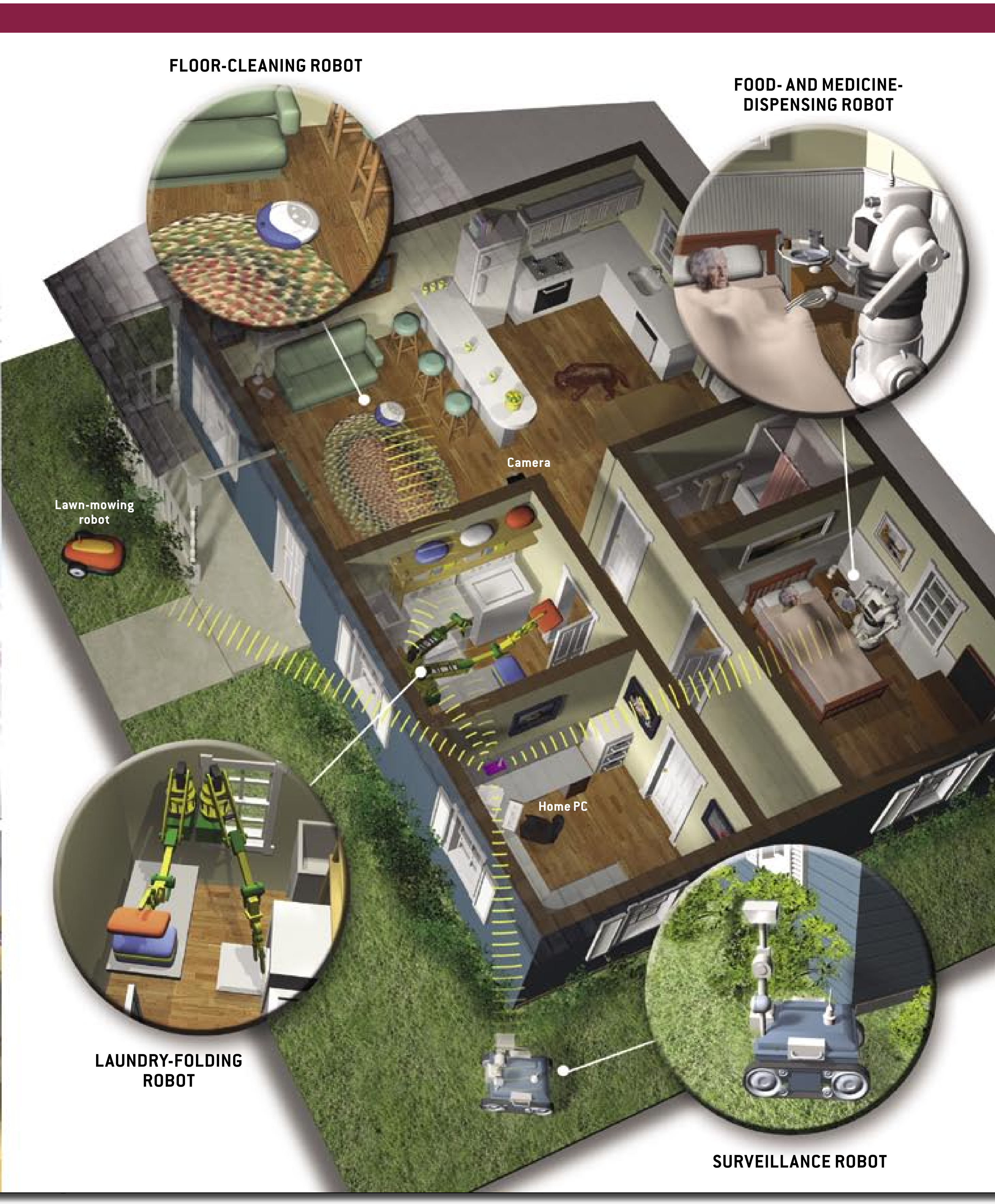
## Overview/*The Robotic Future*

- The robotics industry faces many of the same challenges that the personal computer business faced 30 years ago. Because of a lack of common standards and platforms, designers usually have to start from scratch when building their machines.
- Another challenge is enabling robots to quickly sense and react to their environments. Recent decreases in the cost of processing power and sensors are allowing researchers to tackle these problems.
- Robot builders can also take advantage of new software tools that make it easier to write programs that work with different kinds of hardware. Networks of wireless robots can tap into the power of desktop PCs to handle tasks such as visual recognition and navigation.

## THE ROBOT AND THE PC CAN BE FRIENDS

Linking domestic robots to PCs could provide many benefits. An office worker, for example, could keep tabs on the security of his home, the cleaning of his floors, the folding of his laundry, and the care of his bedridden mother by monitoring a network of household robots on his desktop PC. The machines could communicate wirelessly with one another and with a home PC.

DON FOLEY


FLOOR-CLEANING ROBOT
FOOD- AND MEDICINE-DISPENSING ROBOT
Camera
Lawn-mowing robot
Home PC
LAUNDRY-FOLDING ROBOT
SURVEILLANCE ROBOT

the physical world on our behalf. We may be on the verge of a new era, when the PC will get up off the desktop and allow us to see, hear, touch and manipulate objects in places where we are not physically present.

## From Science Fiction to Reality

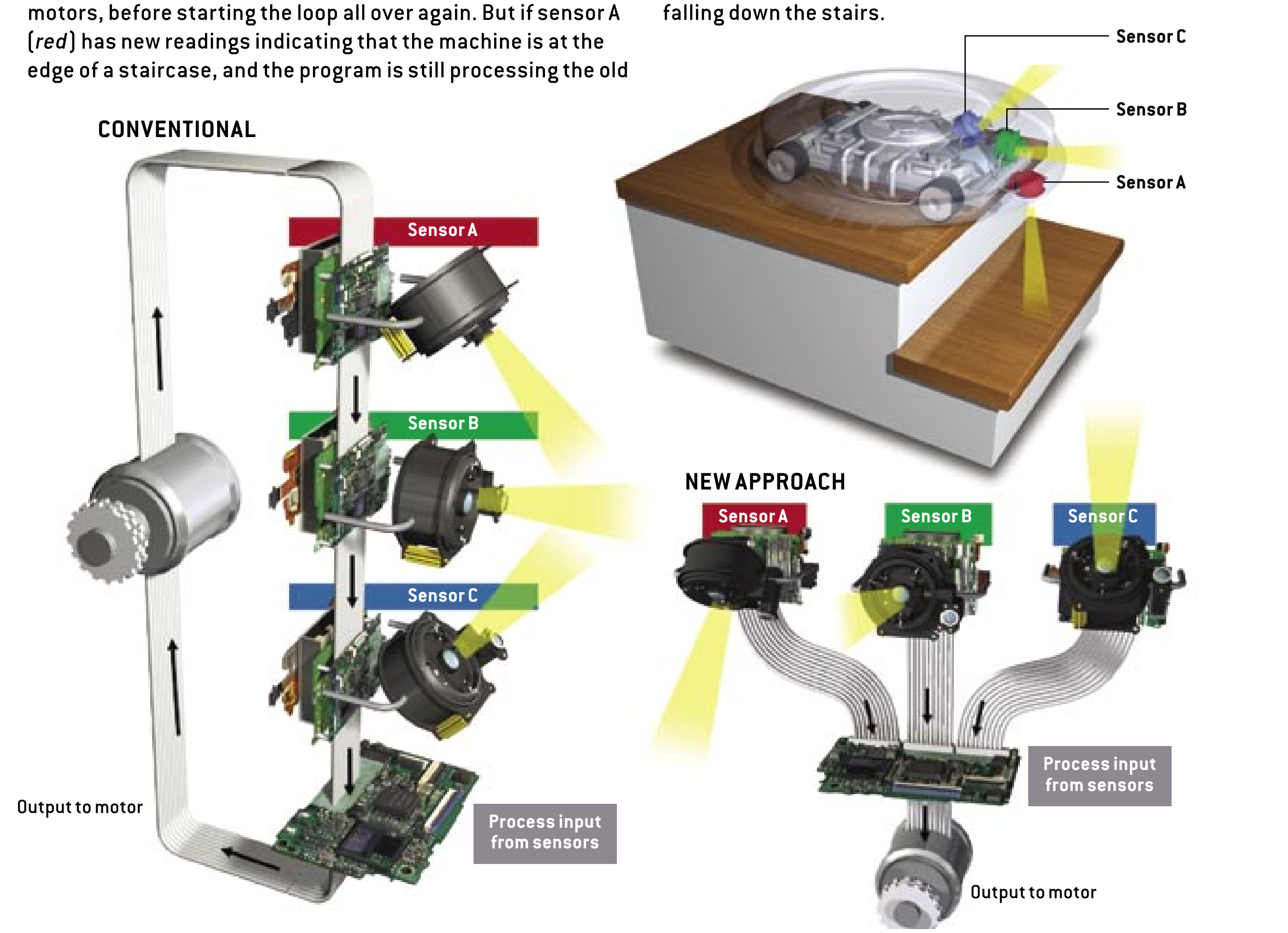
THE WORD "ROBOT" was popularized in 1921 by Czech playwright Karel Čapek, but people have envisioned creating robotlike devices for thousands of years. In Greek and Roman mythology, the gods of metalwork built mechanical servants made from gold. In the first century A.D., Heron of Alexandria—the great engineer credited with inventing the first steam engine—designed intriguing automatons, including one said to have the ability to talk. Leonardo da Vinci's 1495 sketch of a mechanical knight, which could sit up and move its arms and legs, is considered to be the first plan for a humanoid robot.

Over the past century, anthropomorphic machines have become familiar figures in popular culture through books such as Isaac Asimov's *I, Robot,* movies such as *Star Wars* and television shows such as *Star Trek*. The popularity of robots in fiction indicates that people are receptive to the idea that these machines will one day walk among us as helpers and even as companions. Nevertheless, although robots play a vital role in industries such as automobile manufacturing—where there is about one robot for every 10 workers—we have a long way to go before real robots catch up with their science-fiction counterparts.

One reason for this gap is that it has been much harder than expected to give robots the capabilities that humans take for granted—for example, the abilities to orient themselves with respect to the objects in a room, to respond to sounds and interpret speech, and to grasp objects of varying sizes, textures and fragility. Even something as simple as tell-

### BETTER PROGRAMMING MEANS FEWER TUMBLES

Handling data from multiple sensors—for example, the three infrared sensors pictured on the robot at the right—can pose a dilemma. Under the conventional approach (*below*), the program first reads the data from all the sensors, then processes the input and delivers commands to the robot's motors, before starting the loop all over again. But if sensor A (*red*) has new readings indicating that the machine is at the edge of a staircase, and the program is still processing the old sensor data, the robot may take a nasty fall. A better approach to dealing with this problem of concurrency is to write a program with separate data paths for each sensor (*bottom right*). In this design, new readings are processed immediately, enabling the robot to hit the brakes before falling down the stairs.




DON FOLEY

ing the difference between an open door and a window can be devilishly tricky for a robot.

But researchers are starting to find the answers. One trend that has helped them is the increasing availability of tremendous amounts of computer power. One megahertz of processing power, which cost more than $7,000 in 1970, can now be purchased for just pennies. The price of a megabit of storage has seen a similar decline. The access to cheap computing power has permitted scientists to work on many of the hard problems that are fundamental to making robots practical. Today, for example, voice-recognition programs can identify words quite well, but a far greater challenge will be building machines that can understand what those words mean in context. As computing capacity continues to expand, robot designers will have the processing power they need to tackle issues of ever greater complexity.

Another barrier to the development of robots has been the high cost of hardware, such as sensors that enable a robot to determine the distance to an object as well as motors and servos that allow the robot to manipulate objects with strength and delicacy. But prices are dropping fast. Laser range finders used in robotics to measure distance with precision that cost about $10,000 a few years ago can be purchased today for about $2,000. And new, more accurate sensors based on ultrawideband radar are available for even less.

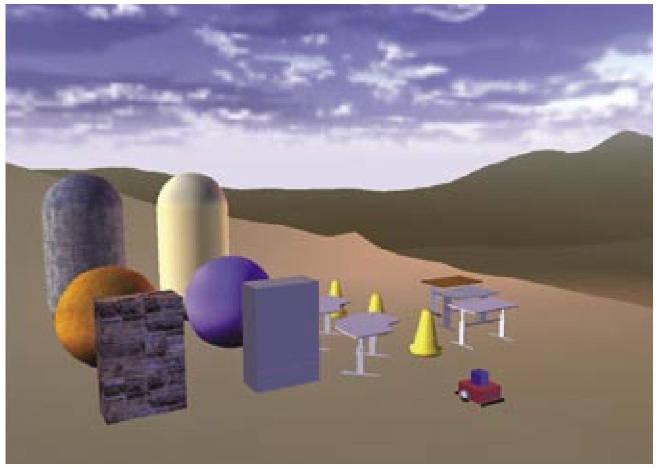
Now robot builders can also add Global Positioning System chips, video cameras, array microphones (which are better than conventional microphones at distinguishing a voice from background noise), and a host of additional sensors for a reasonable expense. The resulting enhancement of capabilities, combined with expanded processing power and storage, allows today's robots to do things such as vacuum a room or help to defuse a roadside bomb—tasks that would have been impossible for commercially produced machines just a few years ago.

## A BASIC Approach

IN FEBRUARY 2004 I visited a number of leading universities, including Carnegie Mellon University, Cornell University and the University of Illinois, to talk about the powerful role that computers can play in solving some of society's most pressing problems. My goal was to help students understand how exciting and important computer science can be, and I hoped to encourage a few of them to think about careers in technology. At each university, after delivering my speech, I had the opportunity to see some of the most interesting research projects in the school's computer science department. Almost without exception, I was shown at least one project that involved robotics.

At that time, my colleagues at Microsoft were also hearing from people in academia and at commercial robotics firms who wondered if our company was doing any work in robotics that might help them with their own development efforts. We were not, so we decided to take a closer look. I asked Tandy Trower, a member of my strategic staff and a 26-year Microsoft veteran, to speak with people across the robotics community. What he found was universal enthusiasm for the potential of robotics and an industry-wide desire for tools that would make development easier. "Many see the robotics industry at a technological turning point where a move to PC architecture makes more and more sense," Tandy wrote in his report to me after his fact-finding mission. "As Red Whittaker, leader of [Carnegie Mellon's] entry in the DARPA Grand Challenge, recently indicated, the hardware capability is mostly there; now the issue is getting the software right."

Back in the early days of the personal computer, we realized that we needed an ingredient that would allow all of the pioneering work to achieve critical mass, to coalesce into a real industry capable of producing truly useful products on a commercial scale. What was needed, it turned out, was Microsoft BASIC. When we created this programming language in the 1970s, we provided the common foundation that enabled programs developed for one set of hardware to run on another. BASIC also made computer programming much easier, which brought more and more people into the industry. Although a great many individuals made essential contributions to the development of the personal computer, Microsoft BASIC was one of the key catalysts that made the PC revolution possible.



COMPUTER TEST-DRIVE of a mobile device in a three-dimensional virtual environment helps robot builders analyze and adjust the capabilities of their designs before trying them out in the real world. Part of the Microsoft Robotics Studio software development kit, this tool simulates the effects of forces such as gravity and friction.

COURTESY OF MICROSOFT

THE AUTHOR

*BILL GATES* is co-founder and chairman of Microsoft, the world's largest software company. While attending Harvard University in the 1970s, Gates developed a version of the programming language BASIC for the first microcomputer, the MITS Altair. In his junior year, Gates left Harvard to devote his energies to Microsoft, the company he had begun in 1975 with his childhood friend Paul Allen. In 2000 Gates and his wife, Melinda, established the Bill & Melinda Gates Foundation, which focuses on improving health, reducing poverty and increasing access to technology around the world.



BIRTH OF AN INDUSTRY: Robot makers have so far introduced a variety of useful machines, but the designs are wildly different. Boss (*above*), an autonomous vehicle built by Carnegie Mellon's Tartan Racing team, won the 2007 DARPA Urban Challenge, traveling 60 miles in a simulated city environment without the aid of a human driver. iRobot, a company based in Burlington, Mass., manufactures the Packbot EOD (*opposite page*), which assists with bomb disposal in Iraq, as well as the Roomba (*right*), which vacuums hardwood floors and carpets. And Lego Mindstorms (*this page, far right*), a tool set for building and programming robots, has become the best-selling product in the history of the Lego Group, the Danish toy maker.

After reading Tandy's report, it seemed clear to me that before the robotics industry could make the same kind of quantum leap that the PC industry made 30 years ago, it, too, needed to find that missing ingredient. So I asked him to assemble a small team that would work with people in the robotics field to create a set of programming tools that would provide the essential plumbing so that anybody interested in robots could easily write robotic applications that would work with different kinds of hardware. The goal was to see if it was possible to provide the same kind of foundation for integrating hardware and software into robot designs that Microsoft BASIC provided for computer programmers.

Tandy's robotics group has drawn on a number of advanced technologies developed by a team working under the direction of Craig Mundie, Microsoft's chief research and strategy officer. One such technology will help solve one of the most difficult problems facing robot designers: how to simultaneously handle all the data coming in from multiple sensors and send the appropriate commands to the robot's motors, a challenge known as concurrency. A conventional approach is to write a traditional, single-threaded program—a long loop that first reads all the data from the sensors, then processes this input and finally delivers output that determines the robot's behavior, before starting the loop all over again. The shortcomings are obvious: if your robot has fresh sensor data indicating that the machine is at the edge of a precipice, but the program is still at the bottom of the loop calculating trajectory and telling the wheels to turn faster based on previous sensor input, there is a good chance the robot will fall down the stairs before it can process the new information.

Concurrency is a challenge that extends beyond robotics. Today as more and more applications are written for distributed networks of computers, programmers continue to struggle to figure out how to efficiently orchestrate code running on many different servers at the same time. And as computers with a single processor are replaced by machines with multiple processors and "multicore" processors—integrated circuits with two or more processors joined together for enhanced performance—software designers will need a new way to program desktop applications and operating systems that solves the problem of concurrency.

One approach to concurrency is multithreaded programming that allows data to travel along many paths. But as any developer who has written multithreaded code can tell you, this is one of the hardest tasks in programming. The answer that Craig's team has devised is something called the concurrency and coordination runtime (CCR). The CCR is a library of functions—sequences of software code that perform specific tasks—that makes it easy to write multithreaded applications that coordinate a number of simultaneous activities. Designed to help programmers take advantage of the power of multicore and multiprocessor systems, it is now being used to program scientific modeling applications, to construct sensor networks and to develop software for financial transaction companies. The CCR also turns out to be ideal for robotics. By drawing on this library to write their programs, robot designers can dramatically reduce the chances that one of their creations will run into a wall because its software is too busy sending output to its wheels to read input from its sensors.

In addition to tackling the problem of concurrency, the work that Craig's team has done will also simplify the writ-


ing of distributed robotic applications through a technology called decentralized software services (DSS). DSS enables developers to create applications in which the services—the parts of the program that read a sensor, say, or control a motor—operate as separate processes that can be orchestrated in much the same way that text, images and information from several servers are aggregated on a Web page. Because DSS allows software components to run in isolation from one another, if an individual component of a robot fails, it can be shut down and restarted—or even replaced—without having to reboot the machine. Combined with broadband wireless technology, this architecture makes it easy to monitor and control a robot from a remote location using a Web browser.

What is more, a DSS application controlling a robotic device does not have to reside entirely on the robot itself but can be distributed across more than one computer. As a result, the robot can be a relatively inexpensive device that delegates complex processing tasks to the high-performance hardware found on today's home PCs. I believe this advance will pave the way for an entirely new class of robots that are essentially mobile, wireless peripheral devices that tap into the power of desktop PCs to handle processing-intensive tasks such as visual recognition and navigation. And because these devices can be networked together, we can expect to see the emergence of groups of robots that can work in concert to achieve goals such as mapping the seafloor or planting crops.

These technologies are a key part of a software development kit built by Tandy's team for the robotics industry. The kit also includes tools that make it easier to create robotic applications using a wide range of programming languages. One example is a simulation tool that lets robot builders test their applications in a three-dimensional virtual environment before trying them out in the real world. The software development kit was created to provide an affordable, open platform that allows robot developers to readily integrate hardware and software into their designs, and it has been downloaded more than 150,000 times since it was released in 2006. We are also working with a number of universities to support robotic research programs. One example is the Institute for Personal Robots in Education at the Georgia Institute of Technology and Bryn Mawr College, which was created to explore the use of robots as a way to engage students in the study of engineering, math and science.

## Should We Call Them Robots?

HOW SOON WILL ROBOTS become part of our day-to-day lives? According to the International Federation of Robotics, about two million personal robots were in use around the world in 2004, and another seven million will be installed by the end of this year. In South Korea the Ministry of Information and Communication hopes to put a robot in every home there by 2013. The Japanese Robot Association predicts that by 2025, the personal robot industry will be worth more than $50 billion a year worldwide, compared with about $5 billion today.

As with the PC industry in the 1970s, it is impossible to predict exactly what applications will drive this new industry. It seems quite likely, however, that robots will play an important role in providing physical assistance and even companionship for the elderly. Robotic devices will probably help people with disabilities get around and extend the strength and endurance of soldiers, construction workers and medical professionals. Robots will maintain dangerous industrial machines and handle hazardous materials. They will enable health care workers to diagnose and treat patients who may be thousands of miles away, and they will be a central feature of security systems and search-and-rescue operations.

Although a few of the robots of tomorrow may resemble the anthropomorphic devices seen in *Star Wars,* most will look nothing like the humanoid C-3PO. In fact, as mobile peripheral devices become more and more common, it may be increasingly difficult to say exactly what a robot is. Because the new machines will be so specialized and ubiquitous—and look so little like the two-legged automatons of science fiction—we probably will not even call them robots. But as these devices become affordable to consumers, they could have just as profound an impact on the way we work, communicate, learn and entertain ourselves as the PC has had over the past 30 years. SA

### MORE TO EXPLORE

More information about robotics in general is available at:

Center for Innovative Robotics: **www.cir.ri.cmu.edu**

DARPA Grand Challenge: **www.darpa.mil/grandchallenge**

International Federation of Robotics: **www.ifr.org**

The Robotics Alliance Project: **www.robotics.nasa.gov**

Robotics Industries Association: **www.roboticsonline.com**

The Robotics Institute: **www.ri.cmu.edu**

The Tech Museum of Innovation: **www.thetech.org/robotics**

Technical details and other information about Microsoft Robotics Studio can be found at **http://msdn.microsoft.com/robotics**

# RISE OF THE ROBOTS

## By 2050 robot "brains" based on computers that execute 100 trillion instructions per second will start rivaling human intelligence

By Hans Moravec

In recent years the mushrooming power, functionality and ubiquity of computers and the Internet have outstripped early forecasts about technology's rate of advancement and usefulness in everyday life. Alert pundits now foresee a world saturated with powerful computer chips, which will increasingly insinuate themselves into our gadgets, dwellings, apparel and even our bodies.

Yet a closely related goal has remained stubbornly elusive. In stark contrast to the largely unanticipated explosion of computers into the mainstream, the entire endeavor of robotics has failed rather completely to live up to the predictions of the 1950s. In those days experts who were dazzled by the seemingly miraculous calculational ability of computers thought that if only the right software were written, computers could become the artificial brains of sophisticated autonomous robots. Within a decade or two, they believed, such robots would be cleaning our floors, mowing our lawns and, in general, eliminating drudgery from our lives.

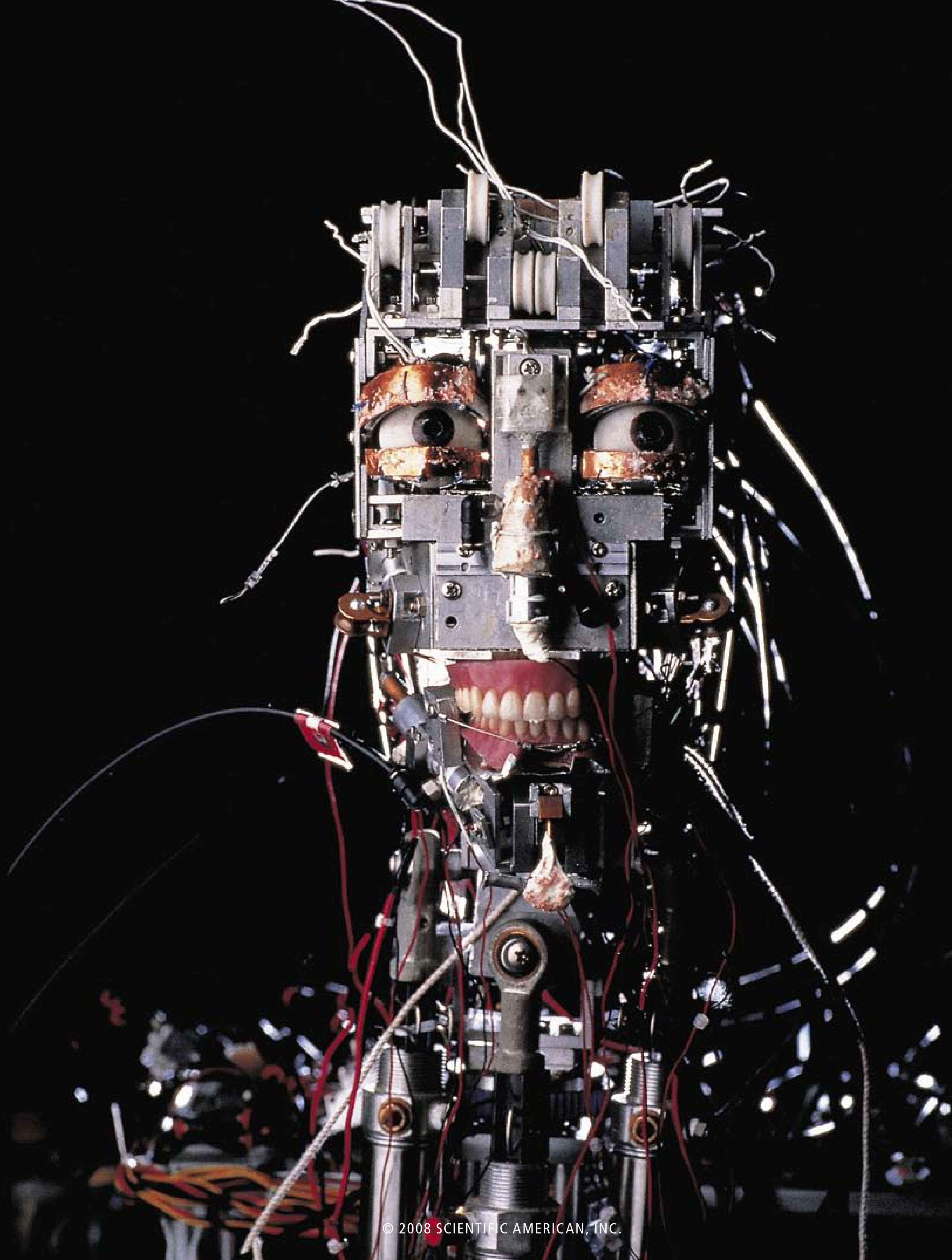
Obviously, it hasn't turned out that way. It is true that industrial robots have transformed the manufacture of automobiles, among other products. But that kind of automation is a far cry from the versatile, mobile, autonomous creations that so many scientists and engineers have hoped for. In pursuit of such robots, waves of researchers have grown disheartened and scores of start-up companies have gone out of business.

PETER MENZEL

It is not the mechanical "body" that is unattainable; articulated arms and other moving mechanisms adequate for manual work already exist, as the industrial robots attest. Rather it is the computer-based artificial brain that is still well below the level of sophistication needed to build a humanlike robot.

Nevertheless, I am convinced that the decades-old dream of a useful, general-purpose autonomous robot will be realized in the not too distant future. By 2010 we will see mobile robots as big as people

FACE ROBOT at the Science University of Tokyo is used in research on how machines can show and respond to emotional expressions. Nonverbal communication will be important in later generations of robots because it will enable them to interact more smoothly with humans.



but with cognitive abilities similar in many respects to those of a lizard. The machines will be capable of carrying out simple chores, such as vacuuming, dusting, delivering packages and taking out the garbage. By 2040, I believe, we will finally achieve the original goal of robotics and a thematic mainstay of science fiction: a freely moving machine with the intellectual capabilities of a human being.

## Reasons for Optimism

IN LIGHT OF what I have just described as a history of largely unfulfilled goals in robotics, why do I believe that rapid progress and stunning accomplishments are in the offing? My confidence is based on recent developments in electronics and software, as well as on my own observations of robots, computers and even insects, reptiles and other living things over the past 30 years.

The single best reason for optimism is the soaring performance in recent years of mass-produced computers. Through the 1970s and 1980s, the computers readily available to robotics researchers were capable of executing about one million instructions per second (MIPS). Each of these instructions represented a very basic task, like adding two 10-digit numbers or storing the result in a specified location in memory.

In the 1990s computer power suitable for controlling a research robot shot through 10 MIPS, 100 MIPS and has lately reached 50,000 MIPS in a few high-end desktop computers with multiple processors. Apple's MacBook laptop computer, with a retail price at the time of this writing of $1,099, achieves about 10,000 MIPS. Thus, functions far beyond the capabilities of robots in the 1970s and 1980s are now coming close to commercial viability.

For example, in October 1995 an experimental vehicle called Navlab V crossed the U.S. from Washington, D.C., to San Diego, driving itself more than 95 percent of the time. The vehicle's self-driving and navigational system was built around a 25-MIPS laptop based on a microprocessor by Sun Microsystems. The Navlab V was built by the Robotics Institute at Carnegie Mellon University, of which I am a member. Similar robotic vehicles, built by researchers elsewhere in the U.S. and in Germany, have logged thousands of highway kilometers under all kinds of weather and driving conditions. Dramatic progress in this field became evident in the DARPA Grand Challenge contests held in California. In October 2005 several fully autonomous cars successfully traversed a hazard-studded 132-mile desert course, and in 2007 several successfully drove for half a day in urban traffic conditions.

In other experiments within the past few years, mobile robots mapped and navigated unfamiliar office suites, and computer vision systems located textured objects and tracked and analyzed faces in real time. Meanwhile personal computers became much more adept at recognizing text and speech.

Still, computers are no match today for humans in such functions as recognition and navigation. This puzzled experts for many years, because computers are far superior to us in calculation. The explanation of this apparent paradox follows from the fact that the human brain, in its entirety, is not a true programmable, general-purpose computer (what computer scientists refer to as a universal machine; almost all computers nowadays are examples of such machines).

**Within a decade or two, experts in the 1950s believed, robots would be cleaning our floors, mowing our lawns and, in general, eliminating drudgery from our lives.**

To understand why this is requires an evolutionary perspective. To survive, our early ancestors had to do several things repeatedly and very well: locate food, escape predators, mate and protect offspring. Those tasks depended strongly on the brain's ability to recognize and navigate. Honed by hundreds of millions of years of evolution, the brain became a kind of ultrasophisticated—but special-purpose—computer.

The ability to do mathematical calculations, of course, was irrelevant for survival. Nevertheless, as language transformed human culture, at least a small part of our brains evolved into a universal machine of sorts. One of the hallmarks of such a machine is its ability to follow an arbitrary set of instructions, and with language, such instructions could be transmitted and carried out. But because we visualize numbers as complex shapes, write them down and perform other such functions, we process digits in a monumentally awkward and inefficient way. We use hundreds of billions of neurons to do in minutes what hundreds of them, specially "rewired" and arranged for calculation, could do in milliseconds.

A tiny minority of people are born with the ability to do seemingly amazing mental calculations. In absolute terms, it's not so amazing: they calculate at a rate perhaps 100 times that of the average person. Computers, by comparison, are millions or billions of times faster.

THE AUTHOR

*HANS MORAVEC* is an adjunct professor at Carnegie Mellon University. He constructed his first mobile robot—an assemblage of tin cans, batteries, lights and a motor—at age 10. His current work focuses on enabling robots to determine their position and to navigate by a three-dimensional awareness of their surroundings. Since 2004 Moravec has been chief scientist of Seegrid Corporation, founded to commercialize "tuggers" and other robots for warehouses and factories.

## Can Hardware Simulate Wetware?

THE CHALLENGE facing roboticists is to take general-purpose computers and program them to match the largely special-purpose human brain, with its ultraoptimized perceptual inheritance and other peculiar evolutionary traits. Today's robot-controlling computers are much too feeble to be applied successfully in that role, but it is only a matter of time before they are up to the task.

Implicit in my assertion that computers will eventually be capable of the same kind of perception, cognition and thought as humans is the idea that a sufficiently advanced and sophisticated artificial system—for example, an electronic one—can be made and programmed to do the same thing as the human nervous system, including the brain. This issue is controversial in some circles right now, and there is room for brilliant people to disagree.

At the crux of the matter is the question of whether biological structure and behavior arise entirely from physical law and whether, moreover, physical law is computable—that is to say, amenable to computer simulation. My view is that there is no good scientific evidence to negate either of these propositions. On the contrary, there are compelling indications that both are true.

Molecular biology and neuroscience are steadily uncovering the physical mechanisms underlying life and mind but so far have addressed mainly the simpler mechanisms. Evidence that simple functions can be composed to produce the higher capabilities of nervous systems comes from programs that read, recognize speech, guide robot arms to assemble tight components by feel, classify chemicals by artificial smell and taste, reason about abstract matters, and so on. Of course, computers and robots today fall far short of broad human or even animal competence. But that situation is understandable in light of an analysis, summarized in the next section, that concludes that today's computers are only powerful enough to function like insect nervous systems. And, in my experience, robots do indeed perform like insects on simple tasks.

Ants, for instance, can follow scent trails but become disoriented when the trail is interrupted. Moths follow pheromone trails and also use the moon for guidance. Similarly,


THIRD-GENERATION robots will have computer "brains" that process around five million MIPS, or millions of instructions per second, giving them intelligence similar to that of monkeys. They will perform a variety of routine domestic and manual chores.

SPACE CHANNEL/PHILIP SAUNDERS/GORDON CHAPMAN

SHUTTLE ROBOT moves people over a predefined area, locating itself with respect to magnets in a grid pattern on the ground. Based in Utrecht, the Netherlands, the company that built the machine, Frog, took its name from the acronym for "free ranging on grid."

FROG NAVIGATION SYSTEMS

many commercial robots can follow guide wires installed below the surface they move over, and some orient themselves using lasers that read bar codes on walls.

If my assumption that greater computer power will eventually lead to human-level mental capabilities is true, we can expect robots to match and surpass the capacity of various animals and then finally humans as computer-processing rates rise sufficiently high. If on the other hand the assumption is wrong, we will someday find specific animal or human skills that elude implementation in robots even after they have enough computer power to match the whole brain. That would set the stage for a fascinating scientific challenge—to somehow isolate and identify the fundamental ability that brains have and that computers lack. But there is no evidence yet for such a missing principle.

The second proposition, that physical law is amenable to computer simulation, is increasingly beyond dispute. Scientists and engineers have already produced countless useful simulations, at various levels of abstraction and approximation, of everything from automobile crashes to the "color" forces that hold quarks and gluons together to make up protons and neutrons.

## Nervous Tissue and Computation

IF WE ACCEPT THAT computers will eventually become powerful enough to simulate the mind, the question that naturally arises is: What processing rate will be necessary to yield performance on a par with the human brain? To explore this issue, I have considered the capabilities of the vertebrate retina, which is understood well enough to serve as a Rosetta stone roughly relating nervous tissue to computation. By comparing how fast the neural circuits in the retina perform image-processing operations with how many instructions per second it takes a computer to accomplish similar work, I believe it is possible to at least coarsely estimate the information-processing power of nervous tissue—and by extrapolation, that of the entire human nervous system.

The human retina is a patch of nervous tissue in the back of the eyeball half a millimeter thick and approximately two centimeters across. It consists mostly of light-sensing cells, but one tenth of a millimeter of its thickness is populated by image-processing circuitry that is capable of detecting edges (boundaries between light and dark) and motion for about a million tiny image regions. Each of these regions is associated with its own fiber in the optic nerve, and each performs about 10 detections of an edge or a motion each second. The results flow deeper into the brain along the associated fiber.

From long experience working on robot vision systems, I know that similar edge or motion detection, if performed by efficient software, requires the execution of at least 100 computer instructions. Therefore, to accomplish the retina's 10 million detections per second would necessitate at least 1,000 MIPS.

The entire human brain is about 75,000 times heavier than the 0.02 gram of processing circuitry in the retina, which implies that it would take, in round numbers, 100 million MIPS (100 trillion instructions per second) to emulate the 1,500-gram human brain. Personal computers in 2008 are just about a match for the 0.1-gram brain of a guppy, but a typical PC would have to be at least 10,000 times more powerful to perform like a human brain.

## Brainpower and Utility

THOUGH DISPIRITING to artificial-intelligence experts, the huge deficit does not mean that the goal of a humanlike artificial brain is unreachable. Computer power for a given price doubled each year in the 1990s, after doubling every 18 months in the 1980s and every two years before that. Prior to 1990 this progress made possible a great decrease in the cost and size of robot-controlling computers. Cost went from many millions of dollars to a few thousand, and size went from room-filling to handheld. Power, meanwhile, held steady at about 1 MIPS. Since 1990 cost and size reductions have abated, but power has risen to about 10,000 MIPS for a home computer. At the present pace, only about 20 or 30 years will be needed to close the gap. Better yet, useful robots don't need full human-scale brainpower.

Commercial and research experiences convince me that the mental power of a guppy—about 10,000 MIPS—will suffice to guide mobile utility robots reliably through unfamiliar surroundings, suiting them for jobs in hundreds of thousands of industrial locations and eventually hundreds of millions of homes. A few machines with 10,000 MIPS are here already, but most industrial robots still use processors with less than 1,000 MIPS.

## Robot vacuum cleaners of the future would run unattended on a schedule that minimized disturbances, and they would remember recharging locations and frequently empty their dust loads.

Commercial mobile robots have found few jobs. A paltry 10,000 work worldwide, and the companies that made them are struggling or defunct. (Makers of robot manipulators are not doing much better.) The largest class of commercial mobile robots, known as automatic guided vehicles (AGVs), transport materials in factories and warehouses. Most follow buried signal-emitting wires and detect end points and collisions with switches, a technique developed in the 1960s.

It costs hundreds of thousands of dollars to install guide wires under concrete floors, and the routes are then fixed, making the robots economical only for large, exceptionally stable factories. Some robots made possible by the advent of microprocessors in the 1980s track softer cues, like magnets or optical patterns in tiled floors, and use ultrasonics and infrared proximity sensors to detect and negotiate their way around obstacles.

The most advanced industrial mobile robots, developed since the late 1980s, are guided by occasional navigational markers—for instance, laser-sensed bar codes—and by preexisting features such as walls, corners and doorways. The costly labor of laying guide wires is replaced by custom software that is carefully tuned for each route segment. The small companies that developed the robots discovered many industrial customers eager to automate transport, floor cleaning, security patrol and other routine jobs. Alas, most buyers lost interest as they realized that installation and route changing required time-consuming and expensive work by experienced route programmers of inconsistent availability. Technically successful, the robots fizzled commercially.

In failure, however, they revealed the essentials for success. First, the physical vehicles for various jobs must be reasonably priced. Fortunately, existing AGVs, forklift trucks, floor scrubbers and other industrial machines designed for accommodating human riders or for following guide wires can be adapted for autonomy. Second, the customer should not have to call in specialists to put a robot to work or to change its routine; floor cleaning and other mundane tasks cannot bear the cost, time and uncertainty of expert installation. Third, the robots must work reliably for at least six months before encountering a problem or a situation requiring downtime for reprogramming or other alterations. Customers routinely rejected robots that after a month of flawless operation wedged themselves in corners, wandered away lost, rolled over employees' feet or fell down stairs. Six months, though, earned the machines a sick day.

Robots exist that have worked faultlessly for years, perfected by an iterative process that fixes the most frequent failures, revealing successively rarer problems that are corrected in turn. Unfortunately, that kind of reliability has been achieved only for prearranged routes. An insectlike 10 MIPS is just enough to track a few handpicked landmarks on each segment of a robot's path. Such robots are easily confused by minor surprises such as shifted bar codes or blocked corridors (not unlike ants thrown off a scent trail or a moth that has mistaken a streetlight for the moon).

### A Sense of Space

ROBOTS THAT chart their own routes emerged from laboratories worldwide in the mid-1990s, as microprocessors reached 100 MIPS. Most build two-dimensional maps from sonar or laser rangefinder scans to locate and route themselves, and the best seem able to navigate office hall-

SEEGRID "TUGGER" autonomously pulls a loaded cart along a memorized route. It sees the world through four stereoscopic cameras on its navigation "head," which glimpses a few thousand visual features per second. (The two diagonally positioned lenses of one camera are visible.) A human "trains" the tugger by leading it through a new route. Thereafter the tugger can automatically retrace the route, matching what it sees with a 3-D grid map built from glimpses during training.


MARC URBAND *Seegrid*

ROBOT VISION conveys the key elements of a scene in a map useful for navigation. These interior and overhead views are from a 3-D map made in 2001 by a robot with stereoscopic computer vision as it drove down the center of an L-shaped hallway.

ways for days before becoming disoriented. Of course, they still fall far short of the six-month commercial criterion. Too often different locations in the coarse maps resemble one another. Conversely, the same location, scanned at different heights, looks different, or small obstacles or awkward protrusions are overlooked. But sensors, computers and techniques are improving, and success is in sight.

My efforts are in the race. In the 1980s at Carnegie Mellon we devised a way to distill large amounts of noisy sensor data into reliable maps by accumulating statistical evidence of emptiness or occupancy in each cell of a grid representing the surroundings. The approach worked well in two dimensions and still guides many of the robots described above.

Three-dimensional maps, 1,000 times richer, promised to be much better but for years seemed computationally out of reach. In 1992 we used economies of scale and other tricks to reduce the computational costs of three-dimensional maps 100-fold. Continued research led us to found a company, Seegrid, that sold its first dozen robots by late 2007. These are load-pulling warehouse and factory "tugger" robots that, on command, autonomously follow routes learned in a single human-guided walk-through. They navigate by three-dimensionally grid-mapping their route, as seen through four wide-angle stereoscopic cameras mounted on a "head," and require no guide wires or other navigational markers.

## Robot, Version 1.0

IN 2008 desktop PCs offer more than 10,000 MIPS. Seegrid tuggers, using slightly older processors doing about 5,000 MIPS, distill about one visual "glimpse" per second. A few thousand visually distinctive patches in the surroundings are selected in each glimpse, and their 3-D positions are statistically estimated. When the machine is learning a new route, these 3-D patches are merged into a chain of 3-D grid maps describing a 30-meter "tunnel" around the route. When the tugger is automatically retracing a taught path, the patches are compared with the stored grid maps. With many thousands of 3-D fuzzy patches weighed statistically by a so-called sensor model, which is trained offline using calibrated example routes, the system is remarkably tolerant of poor sight, changes in lighting, movement of objects, mechanical inaccuracies and other perturbations.

Seegrid's computers, perception programs and end products are being rapidly improved and will gain new functionalities such as the ability to find, pick up and drop loads. The potential market for materials-handling automation is large, but most of it has been inaccessible to older approaches involving buried guide wires or other path markers, which require extensive planning and installation costs and create inflexible routes. Vision-guided robots, on the other hand, can be easily installed and rerouted.

## Fast Replay

PLANS ARE AFOOT to improve, extend and miniaturize our techniques so that they can be used in other applications. On the short list are consumer robot vacuum cleaners. Externally these may resemble the widely available Roomba machines from iRobot. The Roomba, however, is a simple beast that moves randomly, senses only its immediate obstacles and can get trapped in clutter. A Seegrid robot would see, explore and map its premises and would run unattended, with a cleaning schedule minimizing owner disturbances. It would remember its recharging locations, allowing for frequent recharges to run a powerful vacuum motor, and also would be able to frequently empty its dust load into a larger container.

Commercial success will provoke competition and accelerate investment in manufacturing, engineering and re-

HANS MORAVEC

**Asked why there are candles on the table, a third-generation robot might reply that it put them there because its owner likes candlelit dinners and it likes to please its owner.**

search. Vacuuming robots ought to beget smarter cleaning robots with dusting, scrubbing and picking-up arms, followed by larger multifunction utility robots with stronger, more dexterous arms and better sensors. Programs will be written to make such machines pick up clutter, store, retrieve and deliver things, take inventory, guard homes, open doors, mow lawns, play games, and so on. New applications will expand the market and spur further advances when robots fall short in acuity, precision, strength, reach, dexterity, skill or processing power. Capability, numbers sold, engineering and manufacturing quality, and cost-effectiveness will increase in a mutually reinforcing spiral. Perhaps by 2010 the process will have produced the first broadly competent "universal robots," as big as people but with lizardlike 20,000-MIPS minds that can be programmed for almost any simple chore.

Like competent but instinct-ruled reptiles, first-generation universal robots will handle only contingencies explicitly covered in their application programs. Unable to adapt to changing circumstances, they will often perform inefficiently or not at all. Still, so much physical work awaits them in businesses, streets, fields and homes that robotics could begin to overtake pure information technology commercially.

A second generation of universal robot with a mouselike 100,000 MIPS will adapt as the first generation does not and will even be trainable. Besides application programs, such robots would host a suite of software "conditioning modules" that would generate positive and negative reinforcement signals in predefined circumstances. For example, doing jobs fast and keeping its batteries charged will be positive; hitting or breaking something will be negative. There will be other ways to accomplish each stage of an application program, from the minutely specific (grasp the handle underhand or overhand) to the broadly general (work indoors or outdoors). As jobs are repeated, alternatives that result in positive reinforcement will be favored, those with negative outcomes shunned. Slowly but surely, second-generation robots will work increasingly well.

A monkeylike five million MIPS will permit a third generation of robots to learn very quickly from mental rehearsals in simulations that model physical, cultural and psychological factors. Physical properties include shape, weight, strength, texture and appearance of things, and ways to handle them. Cultural aspects include a thing's name, value, proper location and purpose. Psychological factors, applied to humans and robots alike, include goals, beliefs, feelings and preferences. Developing the simulators will be a huge undertaking involving thousands of programmers and experience-gathering robots. The simulation would track external events and tune its models to keep them faithful to reality. It would let a robot learn a skill by imitation and afford a kind of consciousness. Asked why there are candles on the table, a third-generation robot might consult its simulation of house, owner and self to reply that it put them there because its owner likes candlelit dinners and it likes to please its owner. Further queries would elicit more details about a simple inner mental life concerned only with concrete situations and people in its work area.

Fourth-generation universal robots with a humanlike 100 million MIPS will be able to abstract and generalize. They will result from melding powerful reasoning programs to third-generation machines. These reasoning programs will be the far more sophisticated descendants of today's theorem provers and expert systems, which mimic human reasoning to make medical diagnoses, schedule routes, make financial decisions, configure computer systems, analyze seismic data to locate oil deposits, and so on.

Properly educated, the resulting robots will become quite formidable. In fact, I am sure they will outperform us in any conceivable area of endeavor, intellectual or physical. Inevitably, such a development will lead to a fundamental restructuring of our society. Entire corporations will exist without any human employees or investors at all. Humans will play a pivotal role in formulating the intricate complex of laws that will govern corporate behavior. Ultimately, though, it is likely that our descendants will cease to work in the sense that we do now. They will probably occupy their days with a variety of social, recreational and artistic pursuits, not unlike today's comfortable retirees or the wealthy leisure classes.

The path I've outlined roughly recapitulates the evolution of human intelligence—but 10 million times more rapidly. It suggests that robot intelligence will surpass our own well before 2050. In that case, mass-produced, fully educated robot scientists working diligently, cheaply, rapidly and increasingly effectively will ensure that most of what science knows in 2050 will have been discovered by our artificial progeny! SA

## MORE TO EXPLORE


**AI: The Tumultuous History of the Search for Artificial Intelligence.** Daniel Crevier. Basic Books, 1993.

**Robot: Mere Machine to Transcendent Mind.** Hans Moravec. Oxford University Press, 1999.

**Probabilistic Robotics (Intelligent Robotics and Autonomous Agents).** Sebastian Thrun, Wolfram Burgard and Dieter Fox. MIT Press, 2005.

# THE COMING MERGING OF MIND AND MACHINE

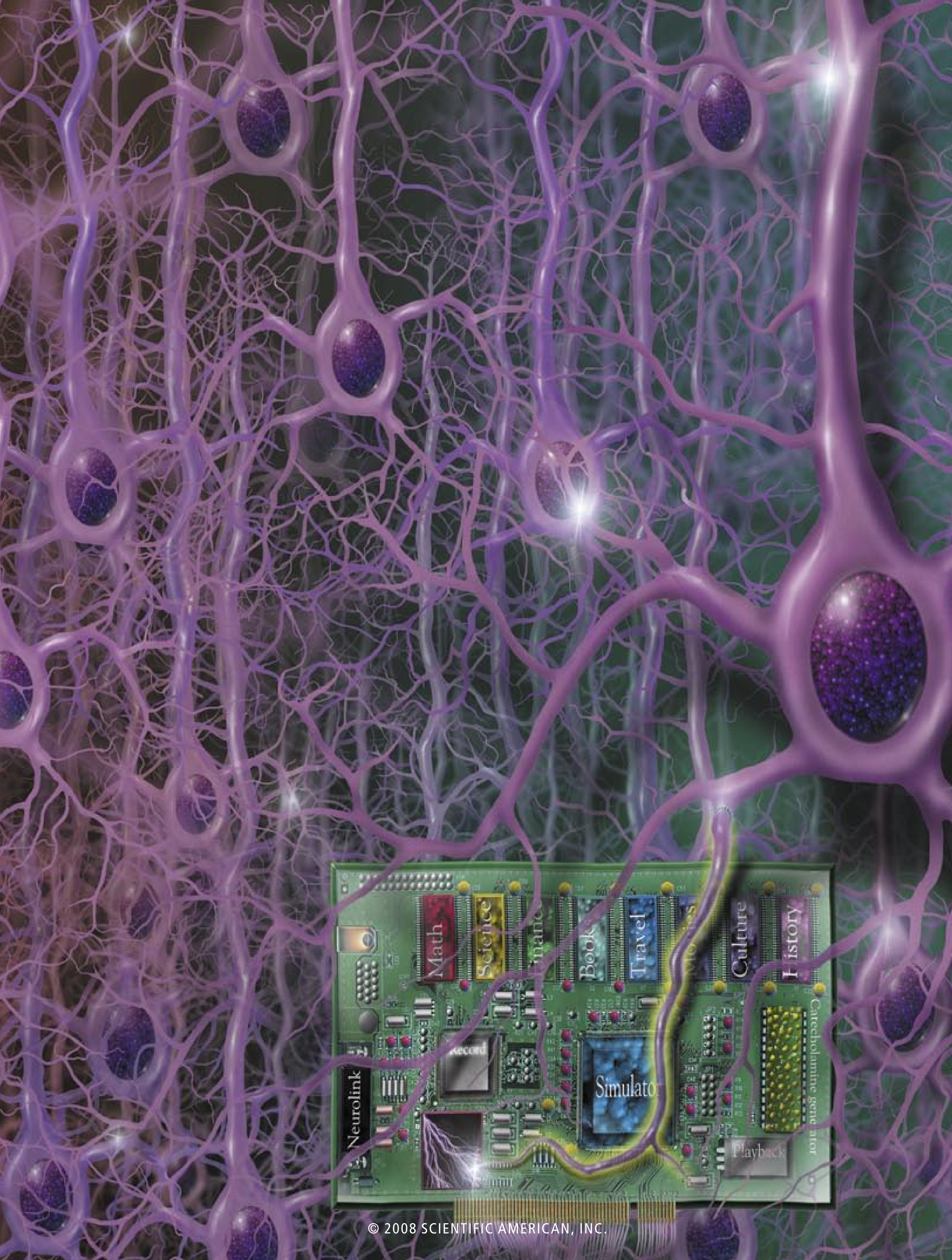
**The accelerating pace of technological progress means that our intelligent creations will soon eclipse us—and that their creations will eventually eclipse them** **By Ray Kurzweil**

Sometime early in this century the intelligence of machines will exceed that of humans. Within a quarter of a century, machines will exhibit the full range of human intellect, emotions and skills, ranging from musical and other creative aptitudes to physical movement. They will claim to have feelings and, unlike today's virtual personalities, will be very convincing when they tell us so. By around 2020 a $1,000 computer will at least match the processing power of the human brain. By 2029 the software for intelligence will have been largely mastered, and the average personal computer will be equivalent to 1,000 brains.

Once computers achieve a level of intelligence comparable to that of humans, they will necessarily soar past it. For example, if I learn French, I can't readily download that learning to you. The reason is that for us, learning involves successions of stunningly complex patterns of interconnections among brain cells (neurons) and among the concentrations of biochemicals known as neurotransmitters that enable impulses to travel from neuron to neuron. We have no way of quickly downloading these patterns. But quick downloading will allow our nonbiological creations to share immediately what they learn with billions of other machines. Ultimately, nonbiological entities will master not

Within a quarter of a century, the author maintains, neural implants will be available that interface directly with our brain cells. The implants would enhance sensory experiences and improve our memory and thinking.

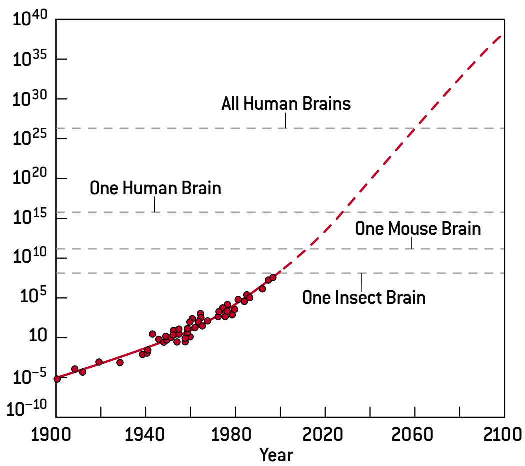
EDMOND ALEXANDER


Math
Science
Book
Travel
Culture
History
Neurolink
Playback

only the sum total of their own knowledge but all of ours as well.

As this happens, there will no longer be a clear distinction between human and machine. We are already putting computers—neural implants—directly into people's brains to counteract Parkinson's disease and tremors from multiple sclerosis. We have cochlear implants that restore hearing. A retinal implant is being developed in the U.S. that is intended to provide at least some visual perception for some blind individuals, basically by replacing certain visual-processing circuits of the brain. A team of scientists at Emory University implanted a chip in the brain of a paralyzed stroke victim that allowed him to use his brainpower to move a cursor across a computer screen.

In the 2020s neural implants will improve our sensory experiences, memory and thinking. By 2030, instead of just phoning a friend, you will be able to meet in, say, a virtual Mozambican game preserve that will seem compellingly




The accelerating rate of progress in computing is demonstrated by this graph, which shows the amount of computing speed that $1,000 (in constant dollars) would buy, plotted as a function of time. Computer power per unit cost is now doubling every year.

real. You will be able to have any type of experience—business, social, sexual—with anyone, real or simulated, regardless of physical proximity.

## How Life and Technology Evolve

TO GAIN INSIGHT into the kinds of forecasts I have just made, it is important to recognize that information technology is advancing exponentially. An exponential process starts slowly, but eventually its pace increases extremely rapidly. (A fuller documentation of my argument is contained in my recent book *The Singularity Is Near.*)

The evolution of biological life and the evolution of technology have both followed the same pattern: they take a long time to get going, but advances build on one another, and progress erupts at an increasingly furious pace. We are entering that explosive part of the technological evolution curve right now.

Consider: It took billions of years for Earth to form. It took two billion more for life to begin and almost as long for molecules to organize into the first multicellular plants and animals about 700 million years ago. The pace of evolution quickened as mammals inherited Earth some 65 million years ago. With the emergence of primates, evolutionary progress was measured in mere millions of years, leading to *Homo sapiens* perhaps 500,000 years ago.

The evolution of technology has been a continuation of the evolutionary process that gave rise to us—the technology-creating species—in the first place. It took tens of thousands of years for our ancestors to figure out that sharpening both sides of a stone created useful tools. Then, earlier in this past millennium, the time required for a major paradigm shift in technology had shrunk to hundreds of years.

The pace continued to accelerate during the 19th century, during which technological progress was equal to that of the 10 centuries that came before it. Advancement in the first two decades of the 20th century matched that of the entire 19th century. Today significant technological transformations take just a few years; for example, the World Wide Web, already a ubiquitous form of communication and commerce, did not exist just 20 years ago. One decade ago almost no one used search engines.

Computing technology is experiencing the same exponential growth. Over the past several decades a key factor in this expansion has been described by Moore's Law. Gordon Moore, a cofounder of Intel, noted in the mid-1960s that technologists had been doubling the density of transistors on integrated circuits every 12 months. This meant computers were periodically doubling both in capacity and in speed per unit cost. In the mid-1970s Moore revised his observation of the doubling time to a more accurate estimate of about 24 months, and that trend has persisted through the years.

After decades of devoted service, Moore's Law will have run its course around 2019. By that time, transistor features will be just a few atoms in width. But new computer architectures will continue the exponential growth of computing. For example, computing cubes are already being designed that will provide thousands of layers of circuits, not just one as in today's com-

LAURIE GRACE

puter chips. Other technologies that promise orders-of-magnitude increases in computing density include nanotube circuits built from carbon atoms, optical computing, crystalline computing and molecular computing.

We can readily see the march of computing by plotting the speed (in instructions per second) per $1,000 (in constant dollars) of 49 famous calculating machines spanning the 20th century [*see illustration on opposite page*]. The graph is a study in exponential growth: computer speed per unit cost doubled every three years between 1910 and 1950 and every two years between 1950 and 1966 and is now doubling every year. It took 90 years to achieve the first $1,000 computer capable of executing one million instructions per second (MIPS). Now we add an additional MIPS to a $1,000 computer every day.

## Why Returns Accelerate

WHY DO WE see exponential progress occurring in biological life, technology and computing? It is the result of a fundamental attribute of any evolutionary process, a phenomenon I call the Law of Accelerating Returns. As order exponentially increases (which reflects the essence of evolution), the time between salient events grows shorter. Advancement speeds up. The returns—the valuable products of the process—accelerate at a nonlinear rate. The escalating growth in the price performance of computing is one important example of such accelerating returns.

A frequent criticism of predictions is that they rely on an unjustified extrapolation of current trends, without considering the forces that may alter those trends. But an evolutionary process accelerates because it builds on past achievements, including improvements in its own means for further evolution. The resources it needs to continue exponential growth are its own increasing order and the chaos in the environment in which the evolutionary process takes place, which provides the options for further diversity. These two resources are essentially without limit.

The Law of Accelerating Returns shows that by around 2020 a $1,000 personal computer will have the processing power of the human brain—20 million billion calculations per second. The estimates are based on regions of the brain that have already been successfully simulated. By 2055, $1,000 worth of computing will equal the processing power of all human brains on Earth (of course, I may be off by a year or two).

## Programming Intelligence

THAT'S THE PREDICTION for processing power, which is a necessary but not sufficient condition for achieving human-level intelligence in machines. Of greater importance is the software of intelligence.

One approach to creating this software is to painstakingly program the rules of complex processes. Another approach is "complexity theory" (also known as chaos theory) computing, in which self-organizing algorithms gradually learn patterns of information in a manner analogous to human learning. One such method, neural nets, is based on simplified mathematical models of mammalian neurons. Another method, called genetic (or evolutionary) algorithms, is based on allowing intelligent solutions to develop gradually in a simulated process of evolution.

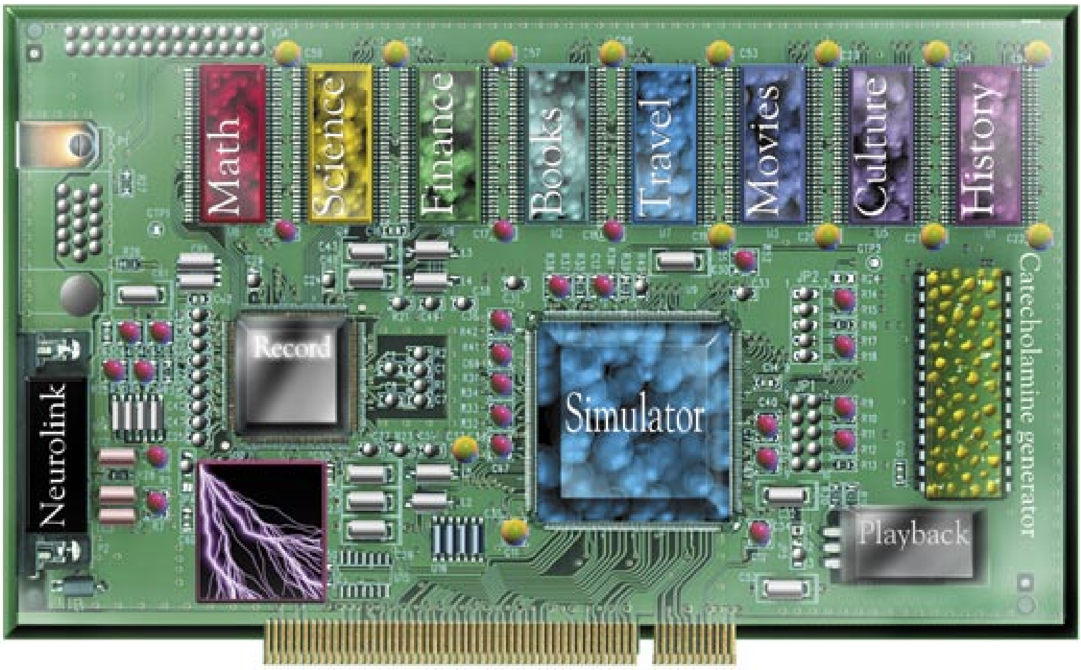
Ultimately, however, we will learn to program intelligence by copying the best intelligent entity we can get our hands on: the human brain itself. We will reverse-engineer the human brain, and fortunately for us it's not even copyrighted!

The most immediate way to reach this goal is by destructive scanning: take a brain frozen just before it was about to expire and examine one very thin slice at a time to reveal every neuron, interneuronal connection and concentration of neurotransmitters across each gap between neurons (these gaps are called synapses). One condemned killer has already allowed his brain and body to be scanned, and all 15 billion bytes of him can be accessed on the National Library of Medicine's Web site (www.nlm.nih.gov/research/visible/visible_gallery.html). The resolution of these scans is not nearly high enough for our purposes, but the data at least enable us to start thinking about these issues.

By 2055 a $1,000 personal computer will have as much processing power as **all human brains combined**.

THE AUTHOR

*RAY KURZWEIL* is CEO of Kurzweil Technologies, Inc. He led teams that built the first print-to-speech reading machine, the first omni-font ("any" font) optical-character-recognition system, the first text-to-speech synthesizer, the first music synthesizer capable of re-creating the grand piano and the first commercially marketed large-vocabulary speech-recognition system.




The author argues that neural implants will confer on humans an important advantage that only machines now possess: instant downloading of knowledge. Memories of events could be played back exactly as they occurred, rather than being colored by emotions. Simulations could make fantasies indistinguishable from reality.

EDMOND ALEXANDER

We also have noninvasive scanning techniques, including high-resolution magnetic resonance imaging (MRI) and others. Recent scanning methods can image individual interneuronal connections in a living brain and show them firing in real time. The increasing resolution and speed of these techniques will eventually enable us to resolve the connections among neurons. The rapid improvement is again a result of the Law of Accelerating Returns, because massive computation is the main element in higher-resolution imaging.

Another approach would be to send microscopic robots (or "nanobots") into the bloodstream and program them to explore every capillary, monitoring the brain's connections and neurotransmitter concentrations.

## Fantastic Voyage

ALTHOUGH SOPHISTICATED robots that small are still a couple of decades away at least, their utility for probing the innermost recesses of our bodies would be far-reaching. They would communicate wirelessly with one another and report their findings to other computers. The result would be a noninvasive scan of the brain taken from within.

Most of the technologies required for this scenario already exist, though not in the microscopic size required. Miniaturizing them to the tiny sizes needed, however, would reflect the essence of the Law of Accelerating Returns. For example, the transistors on an integrated circuit have been shrinking by a factor of approximately five in each linear dimension every 10 years.

The capabilities of these embedded nanobots would not be limited to passive roles such as monitoring. Eventually they could be built to communicate directly with the neuronal circuits in our brains, enhancing or extending our mental capabilities. We already have electronic devices that can communicate with neurons by detecting their activity and either triggering nearby neurons to fire or suppressing them from firing. The embedded nanobots will be capable of reprogramming neural connections to provide virtual-reality experiences and to enhance our pattern recognition and other cognitive faculties.

To decode and understand the brain's information-processing methods (which, incidentally, combine both digital and analog methods), it is not necessary to see every connection, because there is a great deal of redundancy within each region. We are already applying insights from early stages of this reverse-engineering process. For example, in speech recognition, we have decoded and copied the brain's early stages of sound processing.

Perhaps more interesting than this scanning-the-brain-to-understand-it approach would be scanning the brain for the purpose of downloading it. We would map the locations, interconnections and contents of all the neurons, synapses and neurotransmitter concentrations. The entire organization, including the brain's memory, would then be re-created on a digital-analog computer.

To do this, we would need to understand local brain processes, and progress is already under way. Theodore W. Berger and his co-workers at the University of Southern California have built integrated circuits that precisely match the processing characteristics of substantial clusters of neurons. Carver A. Mead and his colleagues at the California Institute of Technology have built a variety of integrated circuits that emulate the digital-analog characteristics of mammalian neural circuits. There are simulations of the visual-processing regions of the brain, as well as the cerebellum, the region responsible for skill formation.

Developing complete maps of the human brain is not as daunting as it may sound. The Human Genome Project seemed impractical when it was first proposed. At the rate at which it was possible to scan genetic codes 20 years ago, it would have taken thousands of years to complete the genome. But in accordance with the Law of Accelerating Returns, the ability to sequence DNA has doubled every year, and the project was completed on time in 2003.

By the third decade of this century, we will be in a position to create complete, detailed maps of the computationally relevant features of the human brain and to re-create these designs in advanced neural computers. We will provide a variety of bodies for our machines, too, from virtual bodies in virtual reality to bodies comprising swarms of nanobots, as well as humanoid robots.

If we **download someone's "mind file"** into a suitable medium, will the entity that emerges be conscious?

## Will It Be Conscious?

SUCH POSSIBILITIES prompt a host of intriguing issues and questions. Suppose we scan someone's brain and reinstate the resulting "mind file" into a suitable computing medium. Will the entity that emerges from such an operation be conscious? This being would appear to others to have very much the same personality, history and memory. For some, that is enough to define consciousness. For others, such as physicist and author James Trefil, no logical reconstruction can attain human consciousness, although Trefil concedes that computers may become conscious in some new way.

At what point do we consider an entity to be conscious, to be self-aware, to have free will? How do we distinguish a process that is conscious from one that just acts *as if* it is conscious? If the entity is very convincing when it says, "I'm lonely, please keep me company," does that settle the issue?

If you ask the "person" in the machine, it will strenuously claim to be the original person. If we scan, let's say, me and reinstate that information into a neural computer, the person who emerges will think he is (and has been) me (or at least he will act that way). He will say, "I grew up in Queens, New York, went to college at M.I.T., stayed in the Boston area, walked into a scanner there and woke up in the machine here. Hey, this technology really works."

But wait, is this really me? For one thing, old Ray (that's me) still exists in my carbon-cell-based brain.

Will the new entity be capable of spiritual experiences? Because its brain processes are effectively identical, its behavior will be comparable to that of the person it is based on. So it will certainly claim to have the full range of emotional and spiritual experiences that a person claims to have.

No objective test can absolutely determine consciousness. We cannot objectively measure subjective experience (this has to do with the very nature of the concepts "objective" and "subjective"). We can measure only correlates of it, such as behavior. The new entities will appear to be conscious, and whether or not they actually are will not affect their behavior. Just as we debate today the consciousness of nonhuman entities such as animals, we will surely debate the potential consciousness of nonbiological intelligent entities. From a practical perspective, we will accept their claims. They'll get mad if we don't.

Before this century is over, the Law of Accelerating Returns tells us, Earth's technology-creating species—us—will merge with our own technology. And when that happens, we might ask: What is the difference between a human brain enhanced a millionfold by neural implants and a nonbiological intelligence based on the reverse-engineering of the human brain that is subsequently enhanced and expanded?

The engine of evolution used its innovation from one period (humans) to create the next (intelligent machines). The subsequent milestone will be for the machines to create their own next generation without human intervention.

An evolutionary process accelerates because it builds on its own means for further evolution. Humans have beaten evolution. We are creating intelligent entities in considerably less time than it took the evolutionary process that created us. Human intelligence—a product of evolution—has transcended it. So, too, the intelligence that we are now creating in computers will soon exceed the intelligence of its creators. SA

# ROBOTS V

# WHO SHOULD

Unmanned spacecraft are exploring the solar system more cheaply and effectively than astronauts are

**By Francis Slakey**

The National Aeronautics and Space Administration has a difficult task. It must convince U.S. taxpayers that space science is worth $16.25 billion a year. To achieve this goal, the agency conducts an extensive public-relations effort that is similar to the marketing campaigns of America's biggest corporations. NASA has learned a valuable lesson about marketing in the 21st century: to promote its programs, it must provide entertaining visuals and stories with compelling human characters. For this reason, NASA issues a steady stream of press releases and images from its human spaceflight program.

Every launch of the space shuttle is a media event. NASA presents its astronauts as ready-made heroes, even when their accomplishments in space are no longer groundbreaking. Perhaps the best example of NASA's public-relations prowess was the participation of John Glenn, the first American to orbit Earth, in the 1998 shuttle mission STS-95. Glenn's return to space at the age of 77 made STS-95 the most avidly followed mission since the Apollo moon landings. NASA claimed that Glenn went up for science—he served as a

*Continued on page 28*

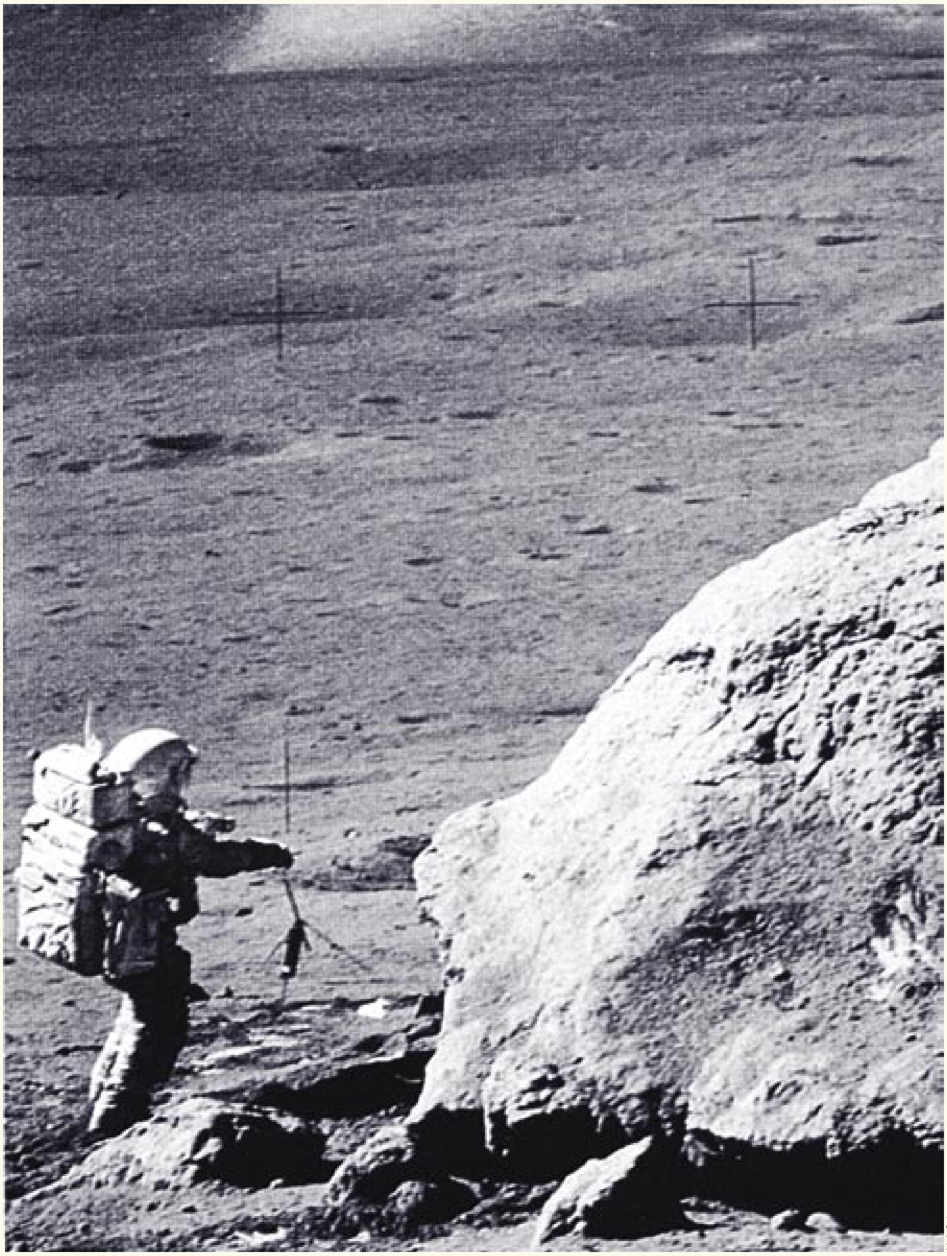
NASA/JPL/CALTECH

NASA'S SPIRIT ROVER uses a spectrometer mounted on its robotic arm to examine Martian rocks. Spirit and Opportunity, twin rovers exploring opposite sides of Mars, landed in January 2004. Their original 90-day mission has been extended five times, and the robots may remain active through 2009. Together they have sent back more than 200,000 images of Mars.

# s. HUMANS EXPLORE SPACE?

Astronaut explorers can perform science in space that robots cannot

By Paul D. Spudis



NASA

*APOLLO 17* ASTRONAUT Harrison Schmitt investigates a huge boulder at the Taurus-Littrow landing site on the moon in 1972. Schmitt, a geologist, made important discoveries about the moon's composition and history, thus demonstrating the value of astronauts as space explorers.

Criticism of human spaceflight comes from many quarters. Some people point to the high cost of manned missions. They contend that the National Aeronautics and Space Administration has a full slate of tasks to accomplish and that human spaceflight is draining funds from more important missions. Other critics question the scientific value of sending people into space. Their argument is that human spaceflight is an expensive "stunt" and that scientific goals can be more easily and satisfactorily accomplished by robotic spacecraft.

But the actual experience of astronauts and cosmonauts over the past 47 years has decisively shown the merits of people as explorers of space. Human capability is required in space to install and maintain complex scientific instruments and to conduct field exploration. These tasks take advantage of human flexibility, experience and judgment. They demand skills that are unlikely to be automated within the foreseeable future. A program of purely robotic exploration is inadequate in addressing the important scientific issues that make the planets worthy of detailed study.

Many of the scientific instruments sent into space require careful emplace-

*Continued on page 29*

*Slakey, continued from page 26*
guinea pig in various medical experiments—but it was clear that the main benefit of Glenn's space shuttle ride was publicity, not scientific discovery.

NASA is still conducting grade-A science in space, but it is being done by unmanned probes rather than astronauts. In recent years the Pathfinder rover has scoured the surface of Mars, and the Galileo spacecraft has surveyed Jupiter and its moons. The Hubble Space Telescope and other orbital observatories are bringing back pictures of the early moments of creation. But robots aren't heroes. No one throws a ticker-tape parade for a telescope. Human spaceflight provides the stories that NASA uses to sell its programs to the public. And that's the main reason NASA spends nearly a quarter of its budget to launch the space shuttle about half a dozen times every year.

The space agency is now saddled with the International Space Station, the budget-hemorrhaging "laboratory" orbiting Earth. NASA says the station provides a platform for space research and helps to determine how people can live and work safely in space. This knowledge could be used to plan a manned mission to Mars or the construction of a base on the moon. But these justifications for the station are largely myths. Here are the facts, plain as potatoes: The International Space Station is not a platform for cutting-edge science. Unmanned probes can explore Mars and other planets more cheaply and effectively than manned missions can. And a moon colony would be a silly destiny.

NASA'S UNMANNED PHOENIX LANDER, launched in August 2007, is scheduled to touch down near the Martian north pole in late May. This artist's conception shows the spacecraft in its final descent, using pulsed rocket engines to control its speed. After landing, Phoenix will use a robotic arm to dig for signs of past microbial life in the icy soil.

NASA/JPL/CALTECH/UNIVERSITY OF ARIZONA

Scientific organizations around the world have determined that space station experiments are a waste of time and money.

## The Myth of Science

IN 1990 THE American Physical Society, an organization of 41,000 physicists, reviewed the experiments then planned for the International Space Station. Many of the studies involved examining materials and fluid mechanics in the station's microgravity environment. Other proposed experiments focused on growing protein crystals and cell cultures on the station. The physical society concluded, however, that these experiments would not provide enough useful scientific knowledge to justify building the station. Thirteen other scientific organizations, including the American Chemical Society and the American Crystallographic Association, drew the same conclusion.

Since then, the station has been redesigned and the list of planned experiments has changed, but the research community remains overwhelmingly opposed. To date, at least 20 scientific organizations from around the world have determined that the space station experiments in their respective fields are a waste of time and money. All these groups have recommended that space science should instead be done through robotic and telescopic missions.

These scientists have various reasons for their disapproval. For researchers in materials science, the station is simply

*Continued on page 30*

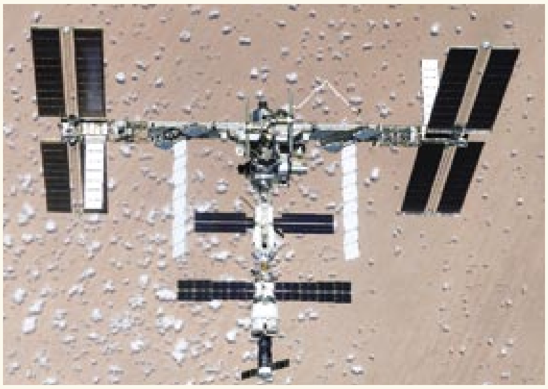
*Spudis, continued from page 27*
ment and alignment to work properly. Astronauts have successfully deployed instruments in Earth orbit—for example, the Hubble Space Telescope—and on the surface of Earth's moon. In the case of the space telescope, the repair of the originally flawed instrument and its continued maintenance have been ably accomplished by space shuttle crews on servicing missions. From 1969 to 1972 the Apollo astronauts carefully set up and aligned a variety of experiments on the lunar surface, which provided scientists with a detailed picture of the moon's interior by measuring seismic activity and heat flow. These experiments operated flawlessly for eight years until shut down in 1977 for fiscal rather than technical reasons.

Elaborate robotic techniques have been envisioned to allow the remote emplacement of instruments on planets or moons. For example, surface rovers could conceivably install a network of seismic monitors. But these techniques have yet to be demonstrated in actual space operations. Very sensitive instruments cannot tolerate the rough handling of robotic deployment. Thus, the auto-deployed versions of such networks would very likely have lower sensitivity and capability than their human-deployed counterparts do.

The value of humans in space becomes even more apparent when complex equipment breaks down. On several occasions astronauts have been able to repair hardware in space, saving missions and the precious scientific data that they produce. When Skylab was launched in 1973, the lab's thermal heat shield was torn off and one of its solar panels was lost. The other solar panel, bound to the lab by restraining ties, would not release. But the first Skylab crew—astronauts Pete Conrad, Joe Kerwin and Paul Weitz—installed a new thermal shield and deployed the pinned solar panel. Their heroic efforts saved not only the mission but also the entire Skylab program.

Of course, some failures are too severe to be repaired in space, such as the damage caused by the explosion of an oxygen tank on the *Apollo 13* spacecraft in 1970. But in most cases when spacecraft equipment malfunctions, astronauts are able to analyze the problem, make on-the-spot judgments and come up with innovative solutions. Machines are capable of limited self-repair, usually by switching to redundant systems that can perform the same tasks as the damaged equipment, but they do not possess as much flexibility as people. Machines can be designed to fix expected problems, but so far only people have shown the ability to handle unforeseen difficulties.

Astronauts have been able to repair hardware in space, saving missions and the precious scientific data that they produce.



ASTRONAUTS ONBOARD the space shuttle *Endeavour* shot this photograph of the International Space Station, with Earth below, as they departed after nine days of work in August 2007. Construction of the station began in 1998 and is scheduled for completion in 2010.

NASA

## Astronauts as Field Scientists

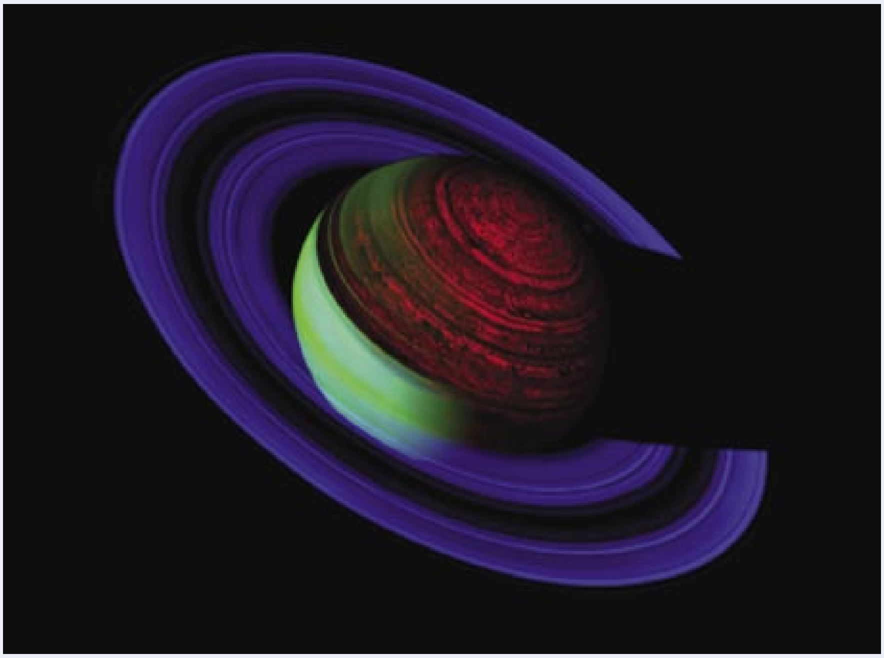
EXPLORATION has two stages: reconnaissance and field study. The goal of reconnaissance is to acquire a broad overview of the compositions, processes and history of a given region or planet. Questions asked during the reconnaissance phase tend to be general—for instance, What's there? Examples of geologic reconnaissance are an orbiting spacecraft mapping the surface of a planet and an automated lander measuring the chemical composition of the planet's soil.

The goals of field study are more ambitious. The object is to understand planetary processes and histories in detail. This requires observation in the field, the creation of a conceptual model, and the formulation and testing of hypotheses. Repeated visits must be made to the same geographic location. Field study is an open-ended, ongoing activity; some field sites on Earth have been studied continuously for more than 100 years and still provide scientists with important new insights. Field study is not a simple matter of collecting data: it requires the guiding presence of human intelligence. People are needed in the field to analyze the overabundant data and determine what should be collected and what should be ignored.

The transition from reconnaissance to field study is fuzzy. In any exploration, reconnaissance dominates the earliest phases. Because it is based on broad questions and simple, focused tasks, reconnaissance is the type of exploration best suited to robots. Unmanned orbiters can provide general information about the atmosphere, surface features and magnetic fields of a planet. Rovers can traverse the planet's surface, testing the physical and chemical properties of the soil and collecting samples for return to Earth.

But field study is complicated, interpretive and protracted. The method of solving the scientific puzzle is often not apparent immediately but must be formulated, applied and modified during

*Continued on page 31*



ABLE TO SEE wavelengths of light that are invisible to the human eye, the unmanned Cassini spacecraft captured the images used to create this false-color mosaic of Saturn, which shows heat rising through bands of clouds (*red*) on the windy planet's night side. Robotic probes can observe planets from many different angles over lengthy periods.

NASA/JPL/UNIVERSITY OF ARIZONA

*Slakey, continued from page 28*

too unstable a platform. Vibrations caused by the movements of astronauts and machinery jar sensitive experiments. The same vibrations make it difficult for astronomers to observe the heavens and for geologists and climatologists to study Earth's surface as well as they could with unmanned satellites. The cloud of gases vented from the station interferes with experiments in space nearby that require near-vacuum conditions. And last, the station orbits only 400 kilometers (250 miles) overhead, traveling through a region of space that has already been studied extensively.

Despite the scientific community's disapproval, NASA went ahead with experiments on the space station. The agency has been particularly enthusiastic about studying the growth of protein crystals in microgravity; NASA claims the studies may spur the development of better medicines. But the American Society for Cell Biology has bluntly called for the cancellation of the crystallography program. The society's review panel concluded that the proposed experiments were not likely to make any serious contributions to the knowledge of protein structure.

## The Myth of Economic Benefit

HUMAN SPACEFLIGHT is extremely expensive. A single flight of the space shuttle costs about $450 million. The shuttle's cargo bay can carry up to 23,000 kilograms (51,000 pounds) of payload into orbit and can return 14,500 kilograms back to Earth. Suppose that NASA loaded up the shuttle's cargo bay with confetti before launching it into space. Even if every kilogram of confetti miraculously turned into a kilogram of gold during the trip, the mission would still lose $80 million.

The same miserable economics hold for the International Space Station. Over its history the station underwent five major redesigns and fell 11 years behind schedule. NASA has spent over three times the $8 billion that the original project was supposed to cost in its entirety.

NASA had hoped that space-based manufacturing on the station would offset some of this expense. In theory, the microgravity environment could allow the production of certain pharmaceuticals and semiconductors that would have advantages over similar products made on Earth. But the high price of sending anything to the station has dissuaded most companies from even exploring the idea.

So far the station's only economic beneficiary has been Russia, one of America's partners in the project. NASA paid $660 million over four years to the Russian Space Agency so it could finish construction of key modules of the station. The money was needed to make up for funds the Russians could not provide because of their country's economic collapse. U.S. Congressman James Sensenbrenner of Wisconsin, who sits on the House Science Committee, bitterly referred to the cash infusion as "bailout money" for Russia.

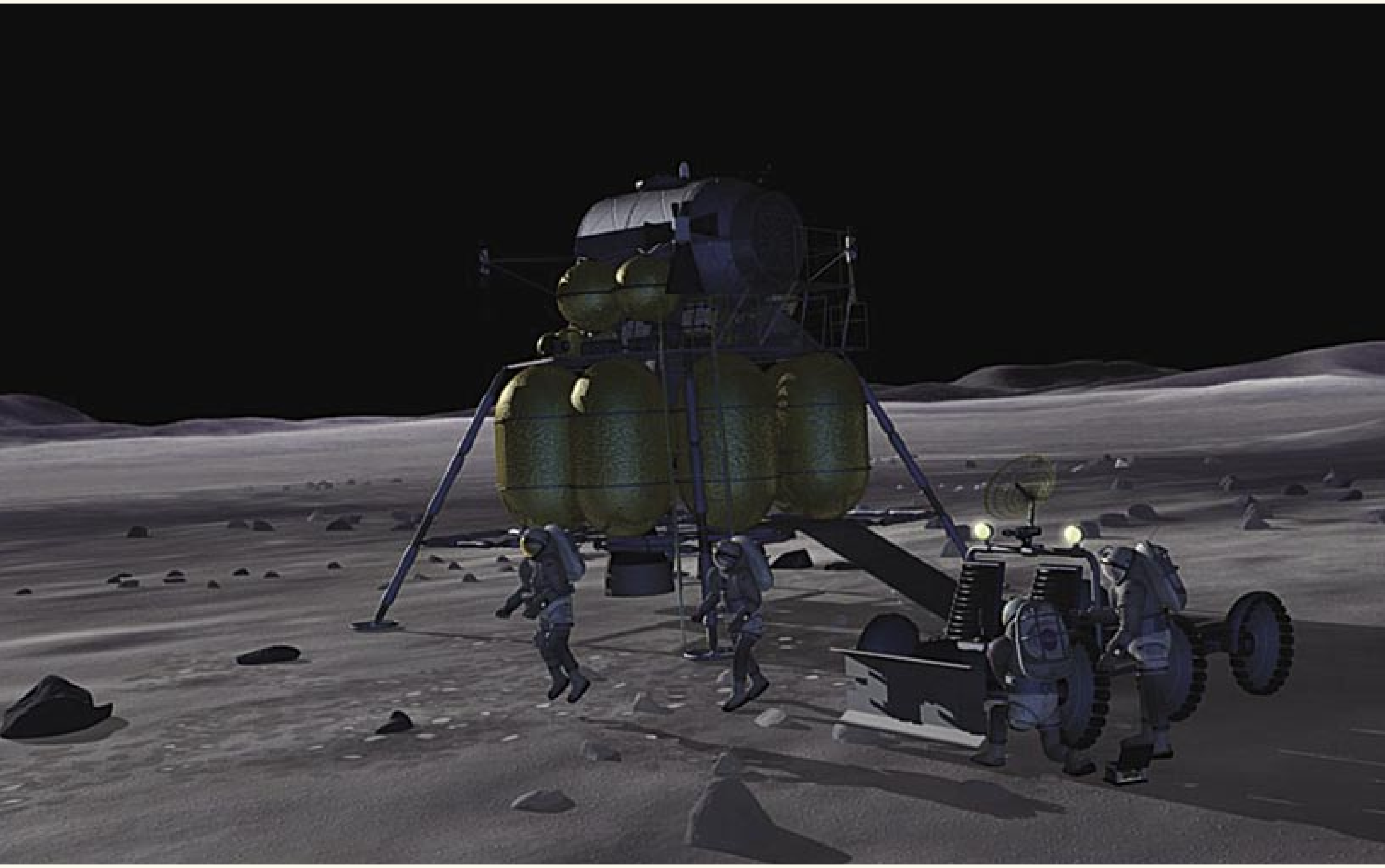
But what about long-term economic benefits? NASA has maintained that the ultimate goal of the space station is to serve as a springboard for a manned mission to Mars. Such a mission would probably cost at least as much as the station; even the most optimistic experts estimate that sending astronauts to the Red Planet would cost tens of billions of dollars. Other estimates run as high as $1 trillion. The only plausible economic benefits of a Mars mission would be in the form of technology spin-offs, and history has shown that such spin-offs are a poor justification for big-money space projects.

In January 1993 NASA released an internal study that examined technology spin-offs from previous missions. According to the study, "NASA's technolo-

*Continued on page 32*

THE AUTHOR

*FRANCIS SLAKEY* is Upjohn Professor of Physics and Biology at Georgetown University and associate director of public affairs for the American Physical Society. He received his Ph.D. in physics in 1992 from the University of Illinois, where his research focused on the optical properties of high-temperature superconductors. He writes and lectures on the subject of science policy; his commentaries have appeared in the *New York Times* and the *Washington Post*.



*Spudis, continued from page 29*
the course of the study. Most important, fieldwork nearly always involves uncovering the unexpected. A surprising discovery may lead scientists to adopt new exploration methods or to make different observations. But an unmanned probe on a distant planet cannot be redesigned to observe unexpected phenomena. Although robots can gather significant amounts of data, conducting science in space requires *scientists*.

It is true that robotic missions are much less costly than human missions; I contend that they are also much less capable. The unmanned Luna 16, 20 and 24 spacecraft launched by the Soviet Union in the 1970s are often praised for returning soil samples from the moon at little cost. But the results from those missions are virtually incomprehensible without the paradigm provided by the results from the manned Apollo program. During the Apollo missions, the geologically trained astronauts were able to select the most representative samples of a given locality and to recognize interesting or exotic rocks and act on such discoveries. In contrast, the Luna samples were scooped up indiscriminately by the robotic probes. We understand the geologic makeup and structure of each Apollo site in much greater detail than those of the Luna sites.

For a more recent example, consider the Mars Pathfinder mission, which was widely touted as a major success. Although Pathfinder discovered an unusual, silica-rich type of rock, because of the probe's limitations we do not know whether this composition represents an igneous rock, an impact breccia or a sedimentary rock. Each mode of origin would have a widely different implication about the history of Mars. Because the geologic context of the sample is unknown, the discovery has negligible scientific value. A trained geologist

*Continued on page 33*

BY 2020 NASA plans to send humans to the moon's surface in a lander that can carry four astronauts. Humans are more adept than robots at deploying instruments and repairing equipment, such as this lunar rover.

NASA/JOHN FRASSANITO AND ASSOCIATES

THE AUTHOR

*PAUL D. SPUDIS* is a staff scientist at the Lunar and Planetary Institute in Houston. He earned his Ph.D. in geology from Arizona State University in 1982 and worked for the U.S. Geological Survey's astrogeology branch until 1990. His research has focused on the moon's geologic history and on volcanism and impact cratering on the planets. He has served on numerous committees advising NASA on exploration strategies and is the author of *The Once and Future Moon* (Smithsonian Institution Press, 1996).

*Slakey, continued from page 30*
gy-transfer reputation is based on some famous examples, including Velcro, Tang and Teflon. Contrary to popular opinion, NASA created none of these." The report concluded that there had been very few technology-transfer successes at NASA over the previous three decades.

## The Myth of Destiny

NOW IT'S TIME to get personal. When I was seven years old, I had a poster of the Apollo astronauts on my bedroom wall. My heroes had fearlessly walked on the moon and returned home in winged glory. They made the universe seem a bit smaller; they made my eyes open a bit wider. I was convinced that one day I would follow in their footsteps and travel to Mars.

So, what happened? I went to Mars three times—twice with the Viking landers in the late 1970s and then with the Mars Pathfinder mission in July 1997. I wasn't alone: millions of people joined me in front-row seats to watch Pathfinder's rugged Sojourner rover scramble over the Martian landscape. I've also traveled to Jupiter's moons with the Galileo spacecraft and seen hints of a liquid ocean on Europa. In 2004 I went to Saturn with the Cassini probe and got a close-up view of the planet's rings.

In recent years there have been tremendous strides in the capabilities of unmanned spacecraft. NASA's Discovery program has encouraged the design of compact, cost-effective probes that can make precise measurements and transmit high-quality images. Mars Pathfinder, for example, returned a treasure trove of data and pictures for only \$265 million. And NASA's New Millennium program is testing advanced technologies such as microsatellites and inertial compasses.

Robotic spacecraft still need human direction, of course, from scientists and engineers in control rooms on Earth. Unlike astronauts, mission controllers are usually not celebrated in the press. But if explorers Lewis and Clark were alive today, that's where they would be sitting. They would not be interested in spending their days tightening bolts on a space station.

Building a manned base on the moon makes even less sense. Unmanned spacecraft can study the moon quite efficiently, as the Lunar Prospector probe has shown. It is not our destiny to build a moon colony any more than it is to walk on our hands.

## What's Next?

FOR THE PRESENT, NASA appears committed to maintaining its human spaceflight program, whatever the cost. But in the next decade the space agency may discover that it does not need human characters to tell compelling stories. Mars Pathfinder proved that an unmanned mission can thrill the public just as much as a shuttle flight. The Pathfinder Web site had 720 million hits in one year. Maybe robots can be heroes after all.

Instead of gazing at posters of astronauts, children are now playing with toy models of Mars rovers. The next generation of space adventurers is growing up with the knowledge that one can visit another planet without boarding a spacecraft. Decades from now, when those children are grown, some of them will lead the next great explorations of the solar system. Sitting in hushed control rooms, they will send instructions to far-flung probes and make the final adjustments that point us toward the stars.

Instead of gazing at posters of astronauts, children are now playing with toy models of Mars rovers.

TO LEARN MORE about how planets in our solar system formed, NASA's unmanned Dawn spacecraft will visit two of the largest protoplanets: Ceres and Vesta, asteroids orbiting in the main belt between Mars and Jupiter. Launched in September 2007, Dawn will arrive at Vesta in 2011—a trip too long for humans to make.

NASA/JPL/WILLIAM K. HARTMANN/COURTESY OF U.C.L.A.

*Spudis, continued from page 31*
could have made a field identification of the rock in a few minutes, giving context to the subsequent chemical analyses and making the scientific return substantially greater.

## Exploring Space by Remote Control

HUMAN DEXTERITY and intelligence are the prime requirements of field study. But is the physical presence of people really required? Telepresence—the remote projection of human abilities into a machine—may permit field study on other planets without the danger and logistical problems associated with human spaceflight. In telepresence the movements of a human operator on Earth are electronically transmitted to a robot that can reproduce the movements on another planet's surface. Visual and tactile information from the robot's sensors give the human operator the sensation of being present on the planet's surface, "inside" the robot. As a bonus, the robot surrogate can be given enhanced strength, endurance and sensory capabilities.

IN NASA'S VISION of the future, astronauts and robots will work side by side in space. This artist's conception shows an astronaut and his robotic assistant collecting rock samples on the surface of Mars. An astronaut trained as a geologist can make quick field identifications.

NASA

**The desire to probe new territory, to see what's over the hill, is a natural human impulse—with a rational basis.**

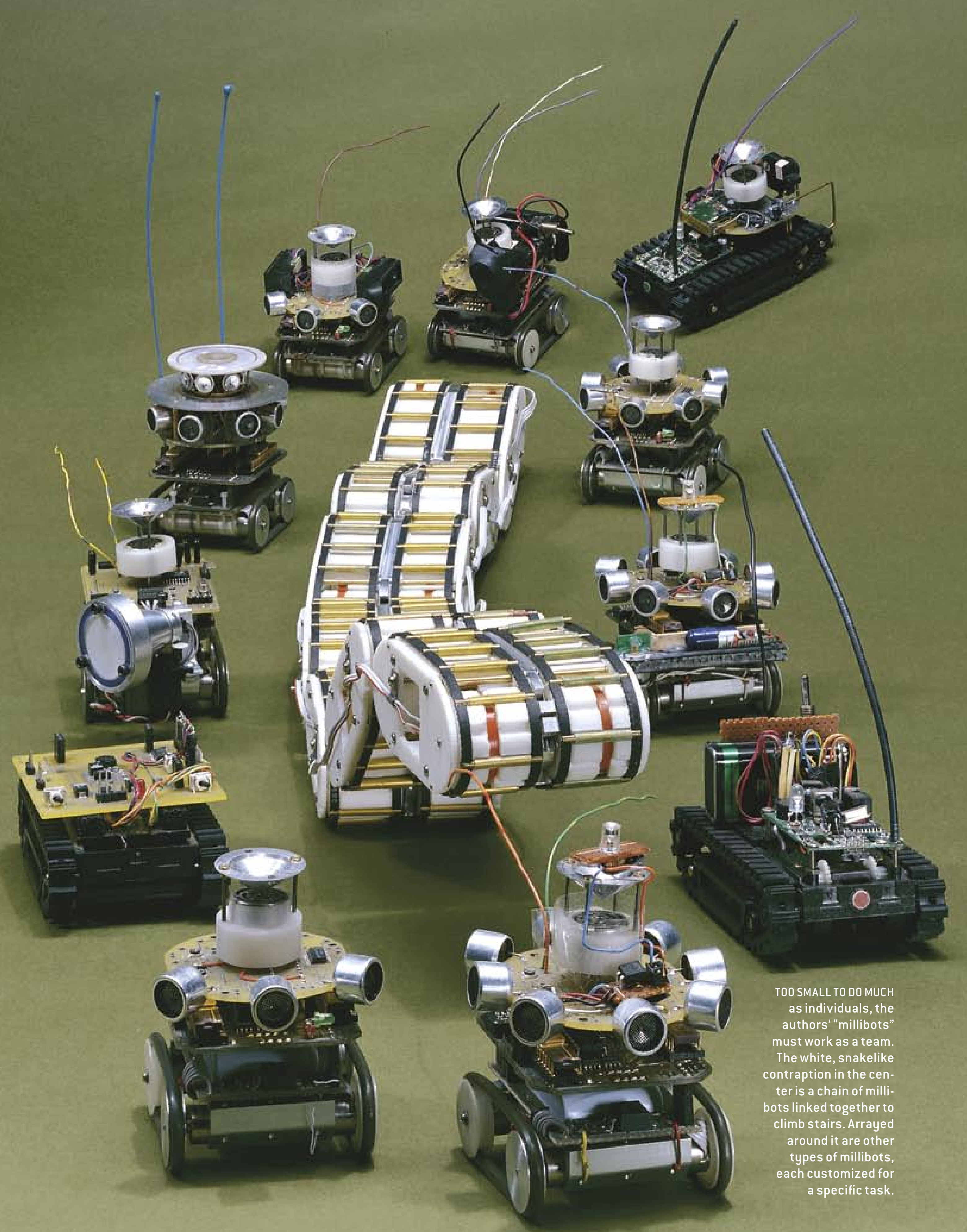
If telepresence is such a great idea, why do we need humans in space? For one, the technology is not yet available. Vision is the most important sense used in field study, and no real-time imaging system developed to date can match human vision, which provides 20 times more resolution than a video screen. But the most serious obstacle for telepresent systems is not technological but psychological. The process that scientists use to conduct exploration in the field is poorly understood, and one cannot simulate what is not understood.

Finally, there is the critical problem of time delay. Ideally, telepresence requires minimal delays between the operator's command to the robot, the execution of the command and the observation of the effect. The distances in space are so vast that instantaneous response is impossible. A signal would take 2.6 seconds to make a round-trip between Earth and its moon. The round-trip delay between Earth and Mars can be as long as 40 minutes, making true telepresence impossible. Robotic Mars probes must rely on a cumbersome interface, which forces the operator to be more preoccupied with physical manipulation than with exploration.

## Robots and Humans as Partners

CURRENTLY NASA is focusing on the construction of the International Space Station. The station is not a destination, however; it is a place to learn how to roam farther afield. Although some scientific research will be done there, the station's real value will be to teach astronauts how to live and work in space. Astronauts must master the process of in-orbit assembly so they can build the complex vehicles needed for interplanetary missions. In the coming decades, the moon will also prove useful as a laboratory and test bed. Astronauts at a lunar base could operate observatories and study the local geology for clues to the history of the solar system. They could also use telepresence to explore the moon's inhospitable environment and learn how to mix human and robotic activities to meet their scientific goals.

The motives for exploration are both emotional and logical. The desire to probe new territory, to see what's over the hill, is a natural human impulse. This impulse also has a rational basis: by broadening the imagination and skills of the human species, exploration improves the chances of our long-term survival. Judicious use of robots and unmanned spacecraft can reduce the risk and increase the effectiveness of planetary exploration. But robots will never be replacements for people. Some scientists believe that artificial-intelligence software may enhance the capabilities of unmanned probes, but so far those capabilities fall far short of what is required for even the most rudimentary forms of field study. SA



TOO SMALL TO DO MUCH as individuals, the authors' "millibots" must work as a team. The white, snakelike contraption in the center is a chain of millibots linked together to climb stairs. Arrayed around it are other types of millibots, each customized for a specific task.

FOR ROBOT DESIGNERS THESE DAYS, SMALL IS BEAUTIFUL

By Robert Grabowski, Luis E. Navarro-Serment and Pradeep K. Khosla

# AN ARMY OF SMALL ROBOTS

A group of terrorists has stormed into an office building and taken an unknown number of people hostage. They have blocked the entrances and covered the windows. No one outside can see how many they are, what weapons they carry or where they are holding their hostages. But suddenly a SWAT team bursts into the room and captures the assailants before they can even grab their weapons. How did the commandos get the information they needed to move so confidently and decisively?

The answer is a team of small, coordinated robots. They infiltrated the building through the ventilation system and methodically moved throughout the ducts. Some were equipped with microphones to monitor conversations, others with small video cameras, still others with sensors that sniffed the air for chemical or biological agents. Working together, they radioed this real-time information back to the authorities.

This is roughly the scenario that the Defense Advanced Research Projects Agency (DARPA) presented to robotics researchers in 1998. Their challenge was to develop tiny reconnaissance robots that soldiers could carry on their backs and scatter on the floor like popcorn. On the home front, firefighters and search-and-rescue workers could toss these robots through windows and let them scoot around to look for trapped victims or sniff out toxic materials. For now, these scenarios remain well beyond the state of the art. Yet the vision of minirobots has captured the attention of leading robot designers. Rather than concentrating on a few large platforms bristling with sensors (like Swiss Army knives on wheels), the focus these days is shifting toward building fleets of small, light and simple robots.

In principle, Lilliputian robots have numerous advantages over their bulkier cousins. They can crawl through pipes, inspect collapsed buildings and hide in inconspicuous niches. A well-organized group of them can exchange sensor information to map objects that cannot be easily comprehended from a single vantage point. They can come to the aid of one another to scale obstacles or recover from a fall. Depending on the situation, the team leader can send in a bigger or smaller number of robots. If one robot fails, the entire mission is not lost; the rest can carry on.

But diminutive robots require a new design philosophy. They do not have the luxury of abundant power and space, as do their larger cousins, and they cannot house all the components necessary to execute a given mission. Even carrying something as compact as a video camera can nearly overwhelm a little robot. Consequently their sensors, processing power and physical strength must be distributed among several robots, which must then work in unison. Such robots are like ants in a colony: weak and vulnerable on their own but highly effective when they join forces.

## Whegs, Golf Balls and Tin Cans

RESEARCHERS HAVE TAKEN various approaches to the problems of building robots at this scale. Some have adopted a biological approach to mimic the attributes of insects and animals. For example, robot designers at Case Western Reserve University have developed a highly mobile platform modeled after a cockroach. It uses a hybrid of wheels and legs ("whegs") to scoot across uneven terrain. A team from the University of Michigan at Ann Arbor has come up with a two-legged robot with suction cups at the ends of its articulated limbs that allow it to climb walls, much like a caterpillar.

Biology has inspired not only the physical shape of the robots but also their control systems. Roboticists at the Massachusetts Institute of Technology have invented robots the size of golf balls that forage for food in the same fashion as ants. They use simple light sensors to express "emotions" to one another and to make decisions collectively. This type of research takes its cue from the work of famous robot scientist Rodney A. Brooks. In the behavior-based control algorithms that he pioneered, each robot reacts to local stimuli. There is no central plan, no colonel commanding the troops. Instead the team's action emerges as a consequence of the combination of individuals interacting with one another. As innovative as this approach is, many problems remain before it can

MARK PORTLAND

bear fruit. Deliberate missions require deliberate actions and deliberate plans—something that emergent behavior cannot reliably provide, at least not yet.

On the more deliberate side, researchers at the University of Minnesota have developed scouts, robots that can be launched like grenades through windows. Shaped like tin cans, these two-wheeled devices are equipped with video cameras that allow them to be teleoperated by a controlling user. Similarly, PARC (formerly known as Xerox PARC) in Palo Alto, Calif., has created a highly articulated snake robot that can be guided via remote video by a user. It literally crawls over obstacles and through pipes. Like the scouts, though, these robots currently lack sufficient local sensing and must rely on a human operator for decision making. This handicap makes them unwieldy for deployment in large numbers.

A few small robot platforms have become commercially available. Khepera, a hockey-puck-size robot developed in Switzerland, has become popular among researchers interested in behavior-based control. Hobbyists, too, are experimenting with the technology. Living Machines in Lompoc, Calif., puts out a tiny programmable robot known as PocketBot. Along the same lines, Lego Mindstorms, an extension to the popular Lego toy bricks, allows the general public to build and operate simple robots. They are being used in science projects and college contests. But the sensing and control for these commercial designs remain extremely rudimentary, and they lack the competence for complex missions.

## Power Shortage

AT CARNEGIE MELLON UNIVERSITY, the emphasis is on flexibility. We built a team of about a dozen "millibots," each about five centimeters on a side. This is the scale at which we could still use off-the-shelf components for sensing and processing, although we had to custom-design the circuit boards and controllers. Each robot consists of three main modules: one for mobility, one for control and one for sensing. The mobility module sits on the bottom. Its two motors drive treads made from small O-rings. The present version can move across office floors and rugs at a maximum speed of about 20 centimeters a second, or about a sixth of normal human walking speed. As

HE'S GOT THE WHOLE ROBOT in his hands: One of the authors (Grabowski) holds a millibot. This particular design is about as small as designers could make it from off-the-shelf components.

MARK PORTLAND

we develop new mobility platforms, we can snap them into place without having to redesign the rest of the robot.

The middle module provides processing and control. The current design contains an eight-bit microcontroller akin to the ones used in personal computers of the early 1980s. Though no match for modern desktop computers, these processors can still perform real-time control for the robot. The sensing module, which sits on top, includes sonar and near-infrared sensors for measuring the distance to nearby obstacles; a mid-infrared sensor (like those used in motion detectors) for detecting warm bodies; a video camera for surveillance; and a radio modem for communicating with other robots or the home base.

Perhaps the most severe limitation on these and other small robots is power. Batteries are bulky and heavy. They do not scale well: as its size is reduced, a battery reaches a threshold at which it cannot supply the power needed to move its own weight. The two rechargeable NiMH cellular-phone batteries on our millibots take up about a third of the available space. They provide enough power for only a limited array of sensors and a run time of between 30 and 90 minutes, depending on the complexity of the mission. Larger batteries would increase the run time but crowd out necessary components. Small-robot design is all about compromise. Speed, duration and functionality compete with weight, size and component availability.

To deal with these constraints, we have adopted two design methodologies for the millibots: specialization and collaboration. The former means that a robot is equipped with only enough sensing and processing for a specific task, allowing it to make optimal use of the available room and power. In a typical mission, some millibots are charged with making maps of the surroundings. Others provide live feedback for the human operator or carry sensors specific to that mission. To get the job done, the robots must collaborate.

## Where Are We?

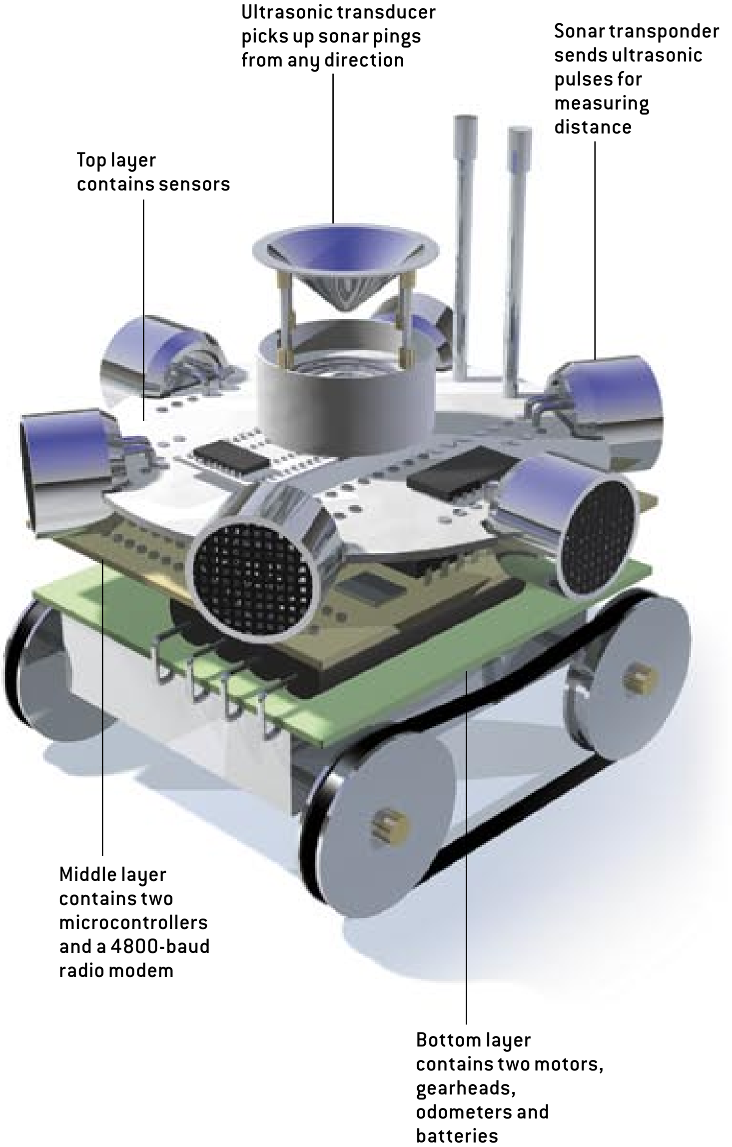
ONE VITAL TASK that requires collaboration is localization: figuring out the team's position. Larger robots have the luxury of several techniques to ascertain their position, such as Global Positioning System (GPS) receivers, fixed beacons and visual landmark recognition. Moreover, they have the processing

## Overview/*Millibots*

- Small robots will one day complement their larger, pricier cousins. Bots the size of Matchbox cars could scurry down pipes and crawl through the debris of collapsed buildings—very useful skills in espionage, surveillance, and search and rescue.
- Limited by size and battery power, small robots do not have the capabilities of a single larger robot. They must divvy up tasks and work together as a team, which is not as easy as it might sound. Engineers have had to develop new techniques for tasks such as ascertaining position and mapping territory.

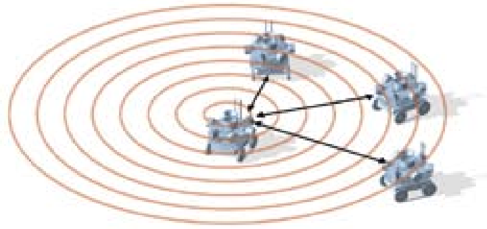
# FINDING THEIR WAY IN THE WORLD

## ANATOMY OF A MILLIBOT

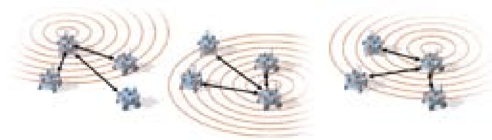



## LOCALIZATION

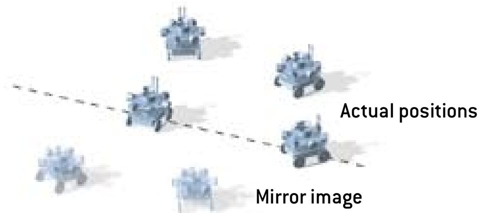
One robot simultaneously sends out an ultrasonic and radio pulse. The others receive the radio pulse instantaneously and the sound pulse shortly after. The time difference is a measure of the distance.



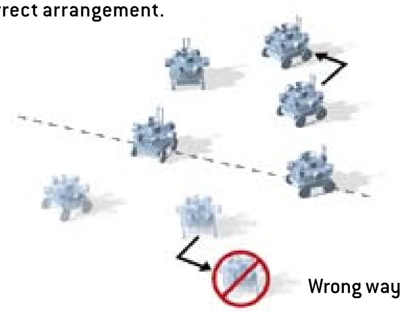
The robots take turns sending and receiving pulses.



A computer uses the distance measurements to deduce the position of each robot. One caveat is that mirror-image arrangements give the same set of measurements.



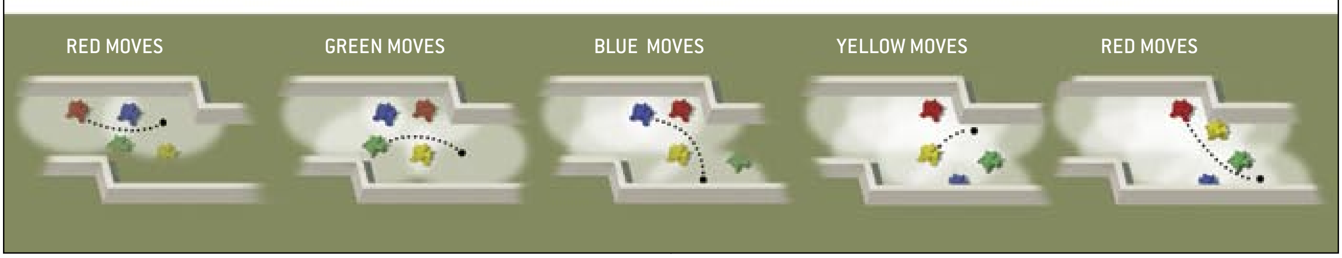

This ambiguity is resolved by having one of the robots take a left turn and measuring its new position, which will differ depending on which mirror image is the correct arrangement.




## MAPPING STRATEGY

By using one another as reference points, millibots can find their way through an unknown space. In this example, three robots fix themselves in place and act as beacons. The fourth robot surveys the area using its sonar. When it is done, the robots switch roles.

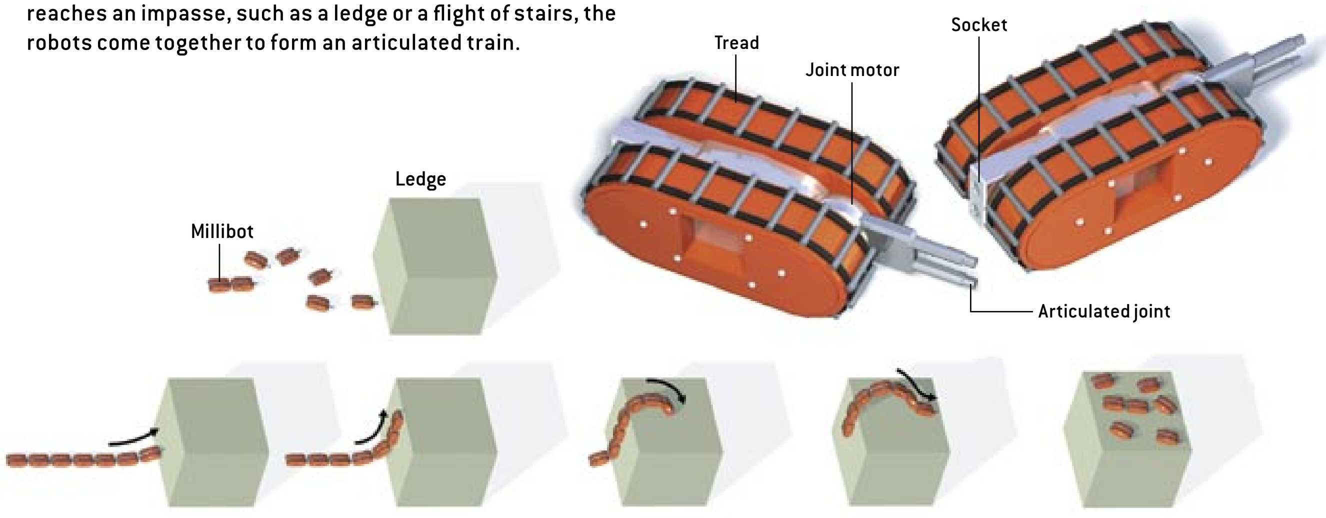
The lead robots become the new beacons, and the rearmost millibot begins moving around and taking data. The maps thus collected can be stitched together to generate a larger composite map of the entire area.




BRYAN CHRISTIE DESIGN

## MILLIBOT CHAIN

During normal operation, individual millibots explore their space and share information to build maps. When the team reaches an impasse, such as a ledge or a flight of stairs, the robots come together to form an articulated train.




An assembled train of robots pushes up against the edge of the obstacle.

Motorized joints lift the forwardmost robots against the side of the obstacle.

As the topmost robots reach over the lip of the object, they bend forward to gain traction.

Once the lead robots have gained enough traction, they pull up the bottom ones.

Having scaled the obstacle, the robots then go their separate ways.

power to match current sensor information to existing maps.

None of these techniques works reliably for midget robots. They have a limited sensor range; the millibot sonar can measure distances out to about two meters. They are too small to carry GPS units. Dead reckoning—the technique of tracking position by measuring the wheel speed—is frustrated by their low weight. Something as seemingly inconsequential as the direction of the weave of a rug can dramatically influence their motion, making odometry readings inaccurate, just as a car's odometer would fail to give accurate distances if driven on an ice-covered lake.

So we have had to come up with a new technique: a miniaturized version of GPS. Rather than using satellites, this technique utilizes sound waves to measure the distances between robots in the group. Each millibot is equipped with an ultrasonic transducer in addition to its radio modem. To determine distance, a millibot simultaneously emits a radio pulse and an ultrasonic signal, which radiate in all directions. The other robots listen for the two signals. The radio wave, traveling at the speed of light, arrives essentially instantaneously. The sound, moving at roughly 340 meters a second, arrives a few milliseconds later, depending on the distance between the robot sending the signal and the robot receiving it. A cone-shaped piece of metal on the sensing module reflects ultrasound down onto a transducer, allowing the robots to detect sound from any direction. The process is analogous to measuring the distance to an approaching storm by timing the interval between lightning and thunder.

By alternating their transmitting and listening roles, the robots figure out the distances between them. Each measurement takes about 30 milliseconds to complete. The team leader—either the home base or a larger robot, perhaps the mother bot that deployed the millibots—collects all the information and calculates robot positions using trilateration. Trilateration resembles the better-known technique of triangulation, except that it relies on distances rather than compass headings to get a fix on position. In two dimensions, each range estimate indi-


**THE AUTHORS**

*ROBERT GRABOWSKI, LUIS E. NAVARRO-SERMENT* and *PRADEEP K. KHOSLA* began working together on the millibot project in the summer of 1999. Grabowski is a principal investigator at the MITRE Corporation. He served eight years in the U.S. Navy working with nuclear reactors and received a Ph.D. from Carnegie Mellon University. He has tinkered with electronics all his life and still enjoys playing with Legos and taking apart old VCRs. Navarro-Serment is a project scientist with the Robotics Institute at Carnegie Mellon, where he received his Ph.D. His background is in industrial automation and control systems; he used to head the electrical engineering department of the Guadalajara campus of the Monterrey Institute of Technology and Higher Education in Mexico. He is an avid amateur astronomer. Khosla is dean of the College of Engineering at Carnegie Mellon and is a member of the National Academy of Engineering. He made his name in robotics by developing the first direct-drive manipulator arms, which are now used in most automated factories.


The authors thank the rest of the millibot team—Chris Paredis, Ben Brown, Curt Bererton and Mike Vande Weghe—for their invaluable contributions.


BRYAN CHRISTIE DESIGN

cates that another robot lies somewhere on a circle around the transmitting robot. The intersection of two or more circles marks the potential location of other robots [*see box on page 37*]. The algorithm finds the arrangement of robots that best satisfies all the circle intersections and range measurements.

One thing that complicates the procedure is that more than one arrangement of robots may match the data. Another is that range measurements are prone to error and uncertainty. Ultrasonic signals echo off floors and walls, creating ambiguity in the distance readings. In fact, depending on the geometry, wave interference can cause the signal to vanish altogether. For this reason, we developed an algorithm that combines the ultrasonic ranging with dead reckoning, which, despite its problems, provides enough additional information to resolve the ambiguities. The algorithm estimates the measurement error and computes the set of robot positions that minimizes the overall error.

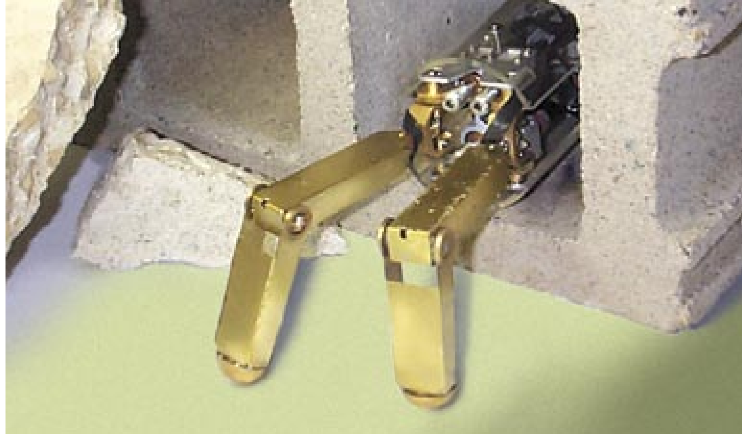
The advantage of this localization method is that the millibots do not need fixed reference points to navigate. They can enter an unfamiliar space and survey it on their own. During mapping, a few selected millibots serve as beacons. These robots remain stationary while the others move around, mapping and avoiding objects while measuring their position relative to the beacons. When the team has fully explored the area around the beacons, the robots switch roles. The exploring robots position themselves as beacons, and the previous set begins to explore. This technique is similar to the children's game of leapfrog, and it can be executed without human intervention.

## Chain of Command

OBSTACLES PRESENT small robots with another reason to collaborate. By virtue of its size, a little robot is susceptible to the random clutter that pervades our lives. It must deal with rocks, dirt and loose paper. The standard millibot has a clearance of about 15 millimeters, so a pencil or twig can stop it in its tracks. To get around these limitations, we have come up with a newer version of the millibots that can couple together like train cars. Each of these new millibots, about 11 centimeters long and six centimeters wide, looks like a miniature World War I–style tank. Typically they roam around independently and are versatile enough to get over small obstacles. But when they need to cross a ditch or scale a flight of stairs, they can link up to form a chain.

What gives the chain its versatility is the coupling joint between millibots. Unlike a train couple or a trailer hitch on a car, the millibot coupling joint contains a powerful motor that can rotate the joint up or down with enough torque to lift several millibots. To climb a stair, the chain first pushes up against the base of the stair. One of the millibots near the center of the chain then cantilevers up the front part of the chain. Those millibots that reach the top can then pull up the lower ones [*see box on opposite page*]. Right now this process has to be remotely controlled by humans, but eventually the chain should be able to scale stairs automatically.

Already researchers' attention has begun to turn from



YOU COULD THROW THIS ROBOT through a window, and after it landed, it would start zipping around. Designed by a University of Minnesota team, TerminatorBot (a variant of the "scout" robots) is a bit smaller than a beer can. It has two arms that can pull it along, climb stairs and manipulate objects.

RICHARD VOYLES *University of Minnesota*

hardware development toward the design of better control systems. The emphasis will shift from the control of a few individuals to the management of hundreds or thousands—a fundamentally different challenge that will require expertise from related fields such as economics, military logistics and even political science.

One of the ways we envision large-scale control is through hierarchy. Much like the military, robots will be divided into smaller teams controlled by a local leader. This leader will be responsible to a higher authority. Already millibots are being directed by larger, tanklike robots whose Pentium processors can handle the complex calculations of mapping and localization. These larger robots can tow a string of millibots behind them like ducklings and, when necessary, deploy them in an area of interest. They themselves report to larger all-terrain-vehicle robots in our group, which have multiple computers, video cameras, GPS units and a range of a few hundred kilometers. The idea is that the larger robots will deploy the smaller ones in areas that they cannot access themselves and then remain nearby to provide support and direction.

To be sure, small robots have a long way to go. Outside of a few laboratories, no small-robot teams are roaming the halls of buildings searching for danger. Although the potential of these robots remains vast, their current capabilities place them just above novelty—which is about where mobile phones and handheld computers were a decade or two ago. As the technology filters down from the military applications and others, we expect the competence of the small robot to improve significantly. Working as teams, they have a full repertoire of skills; their modular design allows them to be customized to particular missions; and, not least, they are fun to work with. SA

### *MORE TO EXPLORE*


**Behavior-Based Robotics (Intelligent Robotics and Autonomous Agents).** Ronald C. Arkin. MIT Press, 1998.

**Heterogeneous Teams of Modular Robots for Mapping and Exploration.** Robert Grabowski, Luis Navarro-Serment, Christopher J. J. Paredis and Pradeep K. Khosla in special issue on heterogeneous multirobot systems, *Autonomous Robots*, Vol. 8, No. 3, pages 293–308; June 2000.

**Millibot Trains for Enhanced Mobility.** H. Benjamin Brown, J. Michael Vande Weghe, Curt A. Bererton and Pradeep K. Khosla in *IEEE/ASME Transactions on Mechatronics*, Vol. 7, No. 4, pages 452–461; December 2002.

# SWARM SMARTS

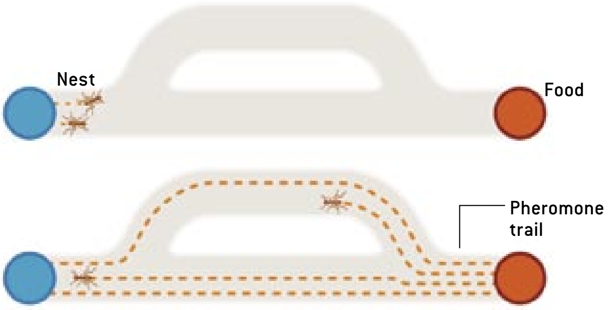
By Eric Bonabeau and Guy Théraulaz

**Using ants and other social insects as models, computer scientists have created software agents that cooperate to solve complex problems, such as the rerouting of traffic in a busy telecom network**

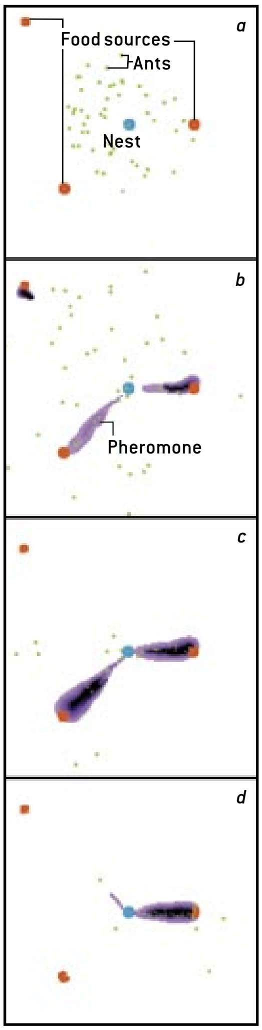
Insects that live in colonies—ants, bees, wasps, termites—have long fascinated everyone from naturalists to artists. Maurice Maeterlinck, the Belgian poet, once wrote, "What is it that governs here? What is it that issues orders, foresees the future, elaborates plans and preserves equilibrium?" These, indeed, are puzzling questions.

Each insect in a colony seems to have its own agenda, and yet the group as a whole appears to be highly organized. Apparently the seamless integration of all individual activities does not require any supervision. In fact, scientists who study the behavior of social insects have found that

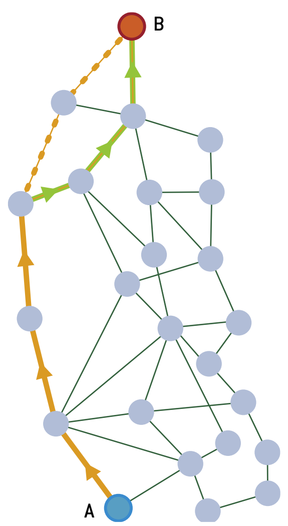
BRYAN CHRISTIE DESIGN




PHEROMONE TRAILS enable ants to forage efficiently. Two ants leave the nest at the same time (*top*), each taking a different path and marking it with pheromone. The ant that took the shorter path returns first (*bottom*). Because this trail is now marked with twice as much pheromone, it will attract other ants more than the longer route will.




DIFFERENT FOOD SOURCES are raided sequentially because of pheromone evaporation. In this computer simulation, three identical sources of food are located at unequal distances from a nest. After foraging randomly (*a*), the ants begin to raid the food sources that are closest (*b*, *c*). As those supplies dwindle, the concentration of pheromone along their trails decreases through evaporation (*d*). The ants will then exploit the farther source.




NETWORK TRAFFIC can be rerouted on the fly with software agents that mimic ants. A transmission that needs to travel from A to B must go through a number of intermediate nodes. If a portion of the shortest path (*orange*) between the two locations is congested, the system must redirect the transmission through an alternative (*green*). Software agents can perform this rerouting automatically in a manner that is similar to how ants raid different food sources (*a–d, above left*). In the analogy, a congested path is like a depleted food source.

cooperation at the colony level is largely self-organized: in numerous situations the coordination arises from interactions among individuals. Although these interactions might be simple (one ant merely following the trail left by another), together they can solve difficult problems (finding the shortest route among countless possible paths to a food source). This collective behavior that emerges from a group of social insects has been dubbed "swarm intelligence."

A growing community of researchers has been devising new ways of applying swarm intelligence to diverse tasks. The foraging of ants has led to a novel method for rerouting network traffic in busy telecommunications systems. The cooperative interaction of ants working to build their nests leads to more effective control algorithms for groups of robots. The way in which insects cluster their colony's dead and sort their larvae can aid in analyzing banking data. And the division of labor among honeybees could help streamline assembly lines in factories.

## Virtual Foraging

ONE OF THE EARLY STUDIES of swarm intelligence investigated the foraging behavior of ants. It had long been known that the ant "highways" often seen in nature (and in people's kitchens) are laid down by individual ants depositing pheromone, a chemical attractant, which increases the probability that other ants will follow the same path to the food source. In the 1990s Jean-Louis Deneubourg of the Free University of Brussels in Belgium, a pioneer in the field, showed that the trail-laying and trail-following behavior of ants was also a good strategy for finding the shortest path between a nest and a food source.

In experiments with the Argentine ant *Linepithema humile*, Deneubourg and his colleagues constructed a bridge with two branches, one twice as long as the other, that separated a nest from a food source. Within just a few minutes the colony usually selected the shorter branch. Deneubourg found that the ants lay and follow trails of pheromone as they forage. The first ants returning to the nest from the food source are those that have taken the shorter path in both directions, from the nest to the food and back. Because this route is the first to be doubly marked with pheromone, nestmates are attracted to it.

If, however, the shorter branch is presented to the colony after the longer branch, the ants will not take it because the longer branch has already been marked with pheromone. But computer scientists can overcome this problem in an artificial system by introducing pheromone decay: when the chemical

**THE AUTHORS**

*ERIC BONABEAU* and *GUY THÉRAULAZ* study the behaviors of social insects and the application of those behaviors in the design of artificial "swarm intelligence" systems. Bonabeau is chief scientific officer at Icosystem Corporation in Cambridge, Mass. He received a Ph.D. in theoretical physics and advanced degrees in computer science and applied mathematics from the University of Paris XI (Paris-Sud). Théraulaz is a research director at the Research Center on Animal Cognition of CNRS at Paul Sabatier University in Toulouse, France. He received a Ph.D. in neuroscience and animal behavior from the University of Provence.

JAMES GARY (*preceding pages*); SIMULATION DONE IN STARLOGO, COURTESY OF MITCH RESNICK *M.I.T. Media Lab* (*this page*)

# Traveling Sales Ants

In the traveling salesman problem, a person must find the shortest route by which to visit a given number of cities, each exactly once. The classic problem is devilishly difficult: for just 15 cities [*see top illustration below*] there are billions of route possibilities.

Researchers have utilized experiments with antlike agents to derive a solution. The approach relies on the artificial ants laying and following the equivalent of pheromone trails [*see illustrations on opposite page*].

Envision a colony of such ants, each independently hopping from city to city, favoring nearby locations but otherwise traveling randomly. After completing a tour of all the cities, an ant goes back to the links it used and deposits pheromone. The amount of the chemical is inversely proportional to the overall length of the tour: the shorter the distance, the more pheromone each of the links receives. Thus, after all the ants have completed their tours and spread their pheromone, the links that belonged to the highest number of short tours will be richest with the chemical. Because the pheromone evaporates, links in long routes will eventually contain significantly less of the substance than those in short tours will.

The colony of artificial ants is then released to travel over the cities again, but this time the insects are guided by the earlier pheromone trails (high-concentration links are favored) as well as by the intercity distances (nearby locations have priority), which the ants can obtain by consulting a table storing those numbers. In general, the two criteria—pheromone strength and intercity distance—are weighted roughly equally.

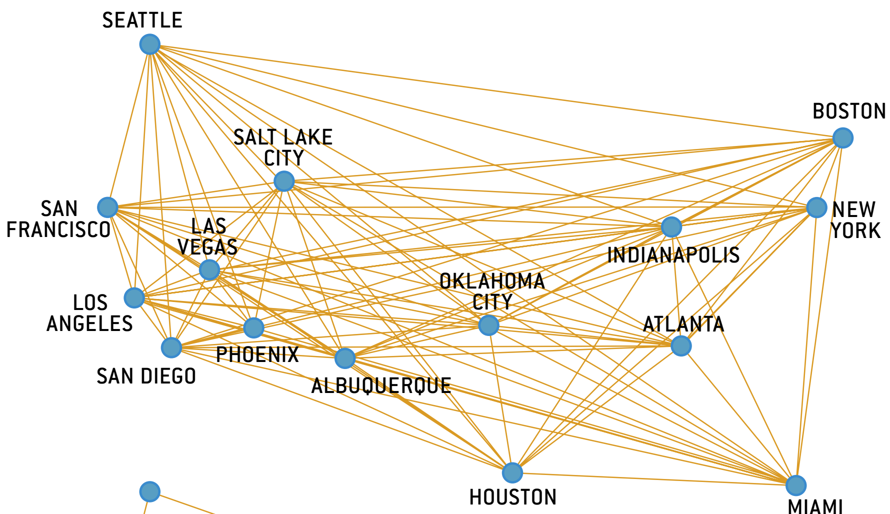
Marco Dorigo of the Free University of Brussels and his colleagues implemented this ant-based system in software. Of course, the methodology assumes that the favored links, when taken together, will lead to an overall short route. Dorigo found that after repeating the process (tour completion followed by pheromone reinforcement and evaporation) numerous times, the artificial ants are indeed able to obtain progressively shorter tours, such as that shown in the bottom illustration below.

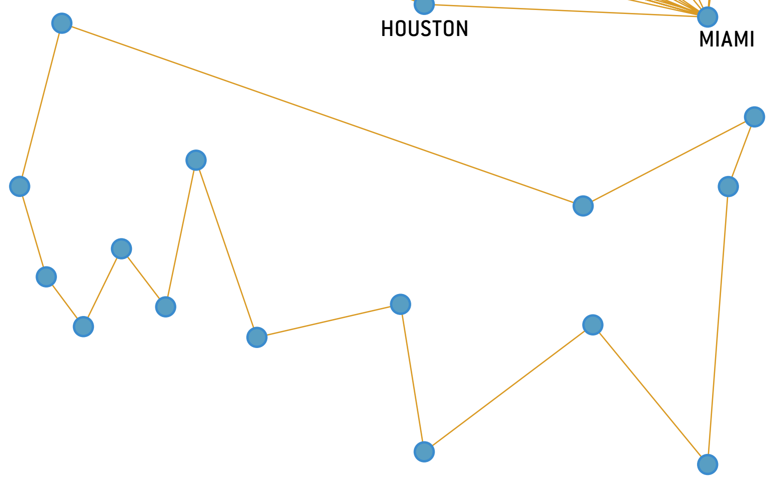
Nevertheless, a difficulty arises when many routes happen to use a link that, as it turns out, is not part of a short tour. (In fact, such a link might belong to many, many long routes.) Dorigo discovered that although this popular link might bias the search for several iterations, a better connection will eventually replace it. This optimization is a consequence of the subtle interplay between reinforcement and evaporation, which ensures that only the better links survive. Specifically, at some point an alternative connection that is part of a short route would be selected by chance and would become reinforced more than the popular link, which would then lose its attractiveness to the artificial ants as its pheromone evaporated.

Another problem occurs when a short route contains a very long link that initially is less likely to be used. But Dorigo showed that even though the connection might be a slow starter, once it has been selected it will quickly become reinforced more than other, competing links.

It is important to note that this ant-based method is effective for finding short routes but not necessarily the shortest one. Nevertheless, such near-optimal solutions are often more than adequate, particularly because obtaining the best route can require an unwieldy amount of computation. In fact, determining the exact solution quickly becomes intractable as the number of cities increases.

In addition, Dorigo's system has one advantage: its inherent flexibility. Because the artificial ants are continuously exploring different paths, the pheromone trails provide backup plans. So whenever one of the links breaks down (bad weather between Houston and Atlanta, for instance), a pool of alternatives already exists. *—E.B. and G.T.*



BRYAN CHRISTIE DESIGN

evaporates quickly, longer paths will have trouble maintaining stable pheromone trails. The software ants can then select a shorter branch even if it is discovered belatedly. This property is highly desirable in that it prevents the system from converging on mediocre solutions. (In *L. humile,* the pheromone concentrations do decay but at a very slow rate.)

In a computer simulation of pheromone evaporation [*see middle illustration on page 42*], researchers presented identical food sources to an artificial colony at different distances from the nest. At first the virtual ants explored their environment randomly. Then they established trails that connected all of the food sources to the nest. Next they maintained only the trails of the sources closest to the nest, leading to the exploitation of those supplies. With the depletion of that food, the software ants began to raid the farther sources.

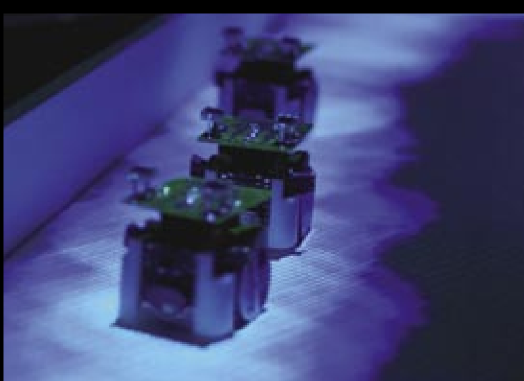
Extending this ant model, Marco Dorigo, a computer scientist at the Free University of Brussels, and his colleagues devised a way to solve the famous "traveling salesman problem" [*see box on preceding page*]. The problem calls for finding the shortest route that goes through a given number of cities exactly once. This test is appealing because it is easy to formulate and yet extremely difficult to solve. It is "NP-complete": the solution requires a number of computational steps that grows faster than the number of cities raised to any finite power (NP stands for *n*ondeterministic *p*olynomial). For such problems, people usually try to find an answer that is good enough but not necessarily the best (that is, a route that is sufficiently short but perhaps not the shortest). Dorigo showed that he could obtain near-optimal routes by using artificial ants that are tweaked so that the concentration of pheromone they deposit varies with the overall distances they have traveled.

Similar approaches have been successful in a number of other optimization tasks. For instance, artificial ants provide the best solution to the classic quadratic assignment problem, in which the manufacture of a number of goods must be assigned to different factories so as to minimize the total distance over which the items need to be transported between facilities. In a related application, David Gregg of Unilever in the U.K. and Vincent Darley of BiosGroup in Santa Fe, N.M., reported that they developed an ant-based method for decreasing the time it takes to perform a given amount of work in a large Unilever plant. The system must efficiently schedule various storage tanks, chemical mixers, packing lines and other equipment.

In addition to solving optimization problems that are basically static, or nonvarying, antlike agents can also cope with glitches and dynamic environments—for example, a factory where a machine breaks down. By maintaining pheromone trails and continuously exploring new paths, the ants serendipitously set up a backup plan and thus are prepared to respond to changes in their environment. This property, which may explain the ecological success of real ants, is crucial for many applications.

Consider the dynamic unpredictability of a telephone network. A phone call from A to B generally has to go through a number of intermediate nodes, or switching stations, requiring a mechanism to tell the call where it should hop next to establish the A-to-B connection. Obviously the algorithm for this process should avoid congested areas to minimize delays, and backup routes become especially valuable when conditions change dramatically. Bad weather at an airport or

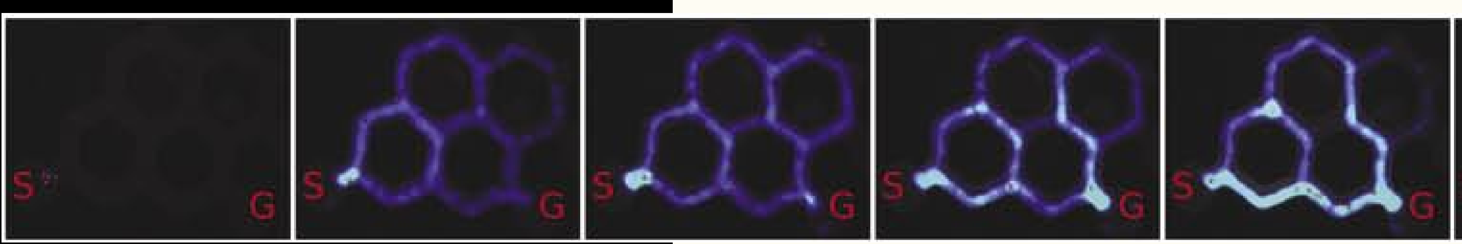
## Swarm Robots



SIMON GARNIER

Hardware miniaturization and costs are strong constraints to the development of complex autonomous robots. An alternative strategy is to design simple robots that cooperate to accomplish tasks. Coordinating the activities of multiple robots is not easy, however, especially when large numbers of them are involved. In such cases, the "swarm

MIMICKING ANTS, tiny robots leave traces of "pheromone" light behind as they explore a network (*left*). The robots can detect and are attracted to the light traces left by other robots. Over time, the most popular path between the starting area (*S*) and the goal area (*G*) becomes the most brightly lit, and the other routes fade away (*below, left to right*).



a phone-in competition on TV will lead to transient local surges of network traffic, requiring on-the-fly rerouting of calls through less busy parts of the system.

To handle such conditions, Ruud Schoonderwoerd and Janet Bruten, both then at Hewlett-Packard's research laboratories in Bristol, England, and Owen Holland, then at the University of the West of England, invented a routing technique in which antlike agents deposit bits of information, or "virtual pheromone," at the network nodes to reinforce paths through uncongested areas. Meanwhile an evaporation mechanism adjusts the node information to disfavor paths that go through busy areas.

Specifically, each node keeps a routing table that tells phone calls where to go next depending on their destinations. Antlike agents continually adjust the table entries, or scores, to reflect the current network conditions. If an agent experiences a long delay because it went through a highly congested portion of the network, it will add just a tiny amount of "pheromone" to the table entries that would send calls to that overloaded area. In mathematical terms, the scores for the corresponding nodes would be increased just slightly. On the other hand, if the agent went quickly from one node to another, it would reinforce the use of that path by leaving a lot of "pheromone"—that is, by increasing the appropriate scores substantially. The calculations are such that even though a busy path may by definition have many agents traveling on it, their cumulative "pheromone" will be less than that of an uncongested path with fewer agents.

The system removes obsolete solutions by applying a mathematical form of evaporation: all of the table entries are decreased regularly by a small amount. This process and the way in which the antlike agents increase the scores are designed to work in tandem so that busy routes experience more evaporation than reinforcement, whereas uncongested routes undergo just the opposite.

Any balance between evaporation and reinforcement can be disrupted easily. When a previously good route becomes congested, agents that follow it are delayed, and evaporation overcomes reinforcement. Soon the route is abandoned, and the agents discover (or rediscover) alternatives and exploit them. The benefits are twofold: when phone calls are rerouted through the better parts of a network, the process not only allows the calls to get through expeditiously but also enables the congested areas to recover from the overload.

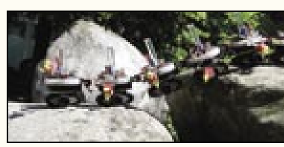
Several companies have explored this approach for handling the traffic on their networks. France Télécom and British Telecommunications took an early lead in applying ant-based routing methods to their systems. In the U.S., MCI WorldCom (now part of Verizon) investigated artificial ants not only for managing the company's telephone network but also for other tasks such as customer billing. The ultimate application, though, may be on the Internet, where traffic is particularly unpredictable.

To handle the demanding conditions of the Net, Dorigo and his colleague Gianni Di Caro, now at the Dalle Molle Institute for Artificial Intelligence in Lugano, Switzerland, increased the sophistication of the ant agents by taking into account several other factors, including the overall time it takes information to get from its origin to its destination. (The approach for phone networks considers just the time it takes to go from one node to another, and the traffic in the reverse direction is assumed to be the same.) Simulation results indicate that Dorigo and Di Caro's system outperforms all other routing methods in terms of both maximizing throughput and minimizing delays. In fact, extensive tests

intelligence" found in nature provides a body of principles that may lead to efficient solutions.

For instance, "pheromone logic" can enable groups of tiny robots to navigate complex environments. Instead of laying and following chemical traces, as ants do, robots lay and follow traces of light that are displayed by a video projector. Initially robots randomly explore a network in search of a goal area. They detect light traces left by other robots and preferentially follow the brighter traces, adding their own light to reinforce the existing trail. Over time, the robots select the shortest path between the starting area and the goal area.

In another example, robots mimic weaver ants that cling to one another to form a living structure (*top right*). This self-assembly allows the colony to perform tasks that none of its members could achieve alone. The European Swarm-bots project has designed autonomous robots, called s-bots, based on this principle. By forming a chain, the s-bots can retrieve heavy objects and bridge gaps that are larger than a single robot (*middle right*). *—E.B. and G.T.*

GUY THÉRAULAZ (*ants*); FRANCESCO MONDADA (*s-bots*)


SIMON GARNIER

# From Cemeteries to Databases

In some ant species, such as *Messor sancta,* workers pile up their colony's dead to clean their nests. The illustration at the right shows the dynamics of such cemetery organization. If the corpses are randomly distributed at the beginning of the experiment, the workers will form clusters within a few hours.

We have recently shown that individual ants pick up and drop corpses as a function of the density of corpses they detect in their neighborhood. The greater the size of a pile of dead ants, the less likely it is that a live ant will remove a corpse from that pile, and the more likely that a live ant will drop a dead ant on the pile. Therefore, a positive feedback results from the combination of enhancement of the dropping behavior and inhibition of the picking-up behavior.

Another phenomenon can be explained in a similar way. The workers of the ant *Leptothorax unifasciatus* sort the colony's brood systematically. Eggs and microlarvae are placed at the center of an area, the largest larvae at the periphery, and pupae and prepupae in between. One explanation of this behavior is that ants pick up and drop items according to the number of similar surrounding objects. For example, if an ant finds a large larva surrounded by eggs, it will most likely pick up the larval "misfit." And that ant will probably deposit its load in a region containing other large larvae.

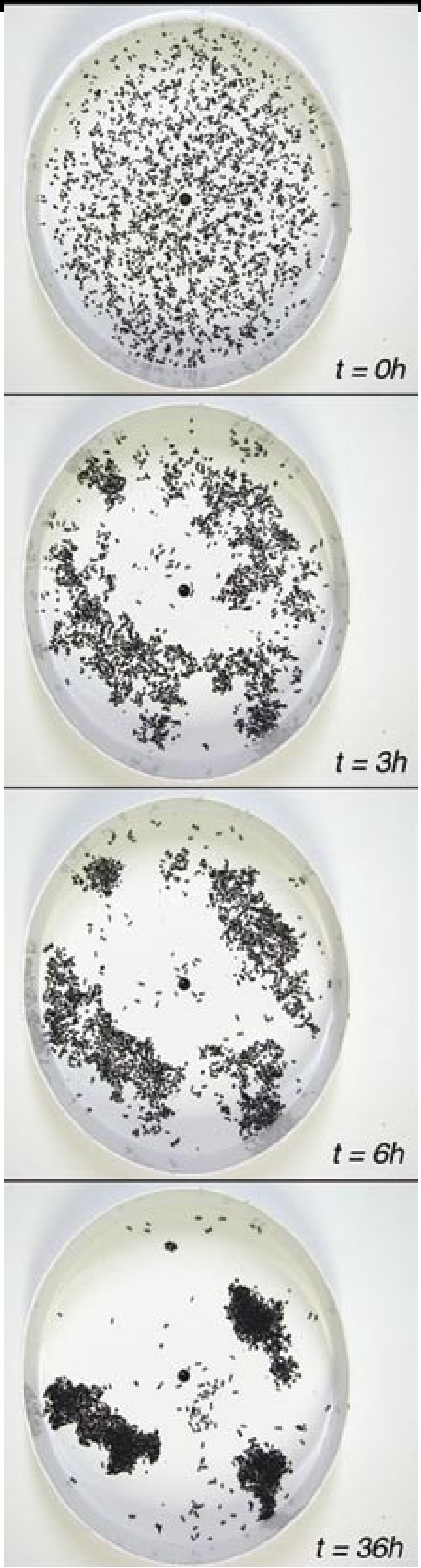
By studying such brood sorting, Erik Lumer, then at University College London, and Baldo Faieta, then at Interval Research in Palo Alto, Calif., developed a method for exploring a large database. Imagine that a bank wants to determine which of its customers is most likely to repay a loan. The problem is that many of the customers have never borrowed money from any financial institution.

But the bank has a large database of customer profiles with attributes such as age, gender, marital status, residential status, banking services used by the customer, and so on. If the bank had a way to visualize clusters of people with similar characteristics, loan officers might be able to predict more accurately whether a particular person would repay a loan. If, for example, a mortgage applicant belonged to a group dominated by defaulters, that person might not be a good credit risk.

Because clusters are generally visualized best in two dimensions (higher dimensions make the data difficult for humans to interpret), Lumer and Faieta represent each customer as a point in a plane. So each client is like a brood item, and software ants can move the clients around, picking them up and depositing them according to the surrounding items. The distance between two customers indicates how similar they are. For the single attribute of age, for instance, shorter distances depict smaller age differences. The artificial ants make their sorting decisions by considering all the different customer characteristics simultaneously. And depending on the bank's objectives, the software could mathematically weigh some of the attributes more heavily than others.

Through this kind of analysis, one cluster might contain people who are about 20 years old and single, most of them living with their parents and whose most popular banking service is interest checking. Another grouping may consist of people who are about 57, female, married or widowed, and homeowners with no mortgage.

Of course, banks and insurance companies have already used similar types of cluster analyses. But the ant-based approach enables the data to be visualized easily, and it boasts one intriguing feature: the number of clusters emerges automatically from the data, whereas conventional methods usually assume a predefined number of groups into which the data are then fit. Thus, antlike sorting has been effective in discovering interesting commonalities that might otherwise have remained hidden. —*E.B. and G.T.*




WORKER ANTS cluster their dead to clean their nest. At the outset of this experiment, 1,500 corpses are located randomly (*top*). After 36 hours, the workers have formed three piles (*bottom*). This behavior and the way in which ants sort their larvae have led to a new type of computer program for analyzing banking data.

GUY THÉRAULAZ

suggest that the ant-based method is superior to Open Shortest Path First, the protocol that the Internet currently uses, in which nodes must continually inform one another of the status of the links to which they are connected.

## A Swarm of Applications

OTHER BEHAVIORS of social insects have inspired a variety of research efforts. Computer scientists are studying insect swarms to devise different techniques for controlling a group of robots. Several applications being investigated are inspired by the coordination of traffic along pheromone trails or the formation of self-assembled chains in ant colonies [*see box on pages 44 and 45*]. Using such approaches, engineers could design relatively simple and cheap robots that would work together to perform increasingly sophisticated tasks. In another project, a model that was initially introduced to explain how ants cluster their dead and sort their larvae has become the basis of a new approach for analyzing financial data [*see box on opposite page*]. And research investigating the flexible way in which honeybees assign tasks could lead to a more efficient method for scheduling jobs in a factory [*see box at right*].

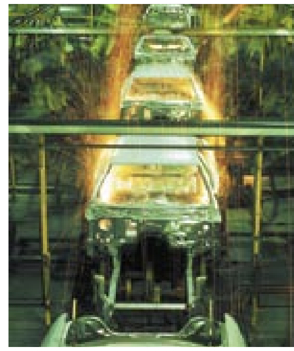
Additional examples abound. Applying knowledge of how wasps construct their nests, Dan Petrovich, then at the Air Force Institute of Technology in Dayton, Ohio, designed a swarm of tiny mobile satellites that would assemble themselves into a larger, predefined structure. H. Van Dyke Parunak of NewVectors in Ann Arbor, Mich., deploys a variety of insectlike software agents to solve manufacturing problems—for example, scheduling a complex network of suppliers to a factory. Paul B. Kantor of Rutgers University developed a swarm-intelligence approach for finding information over the World Wide Web and in other large networks. Web surfers looking for interesting sites can, if they belong to a "colony" of users, access information in the form of digital pheromones (essentially, ratings) left by fellow members in previous searches.

Indeed, the potential of swarm intelligence is enormous. It offers an alternative way of designing systems that have traditionally required centralized control and extensive preprogramming. It instead boasts autonomy and self-sufficiency, relying on direct or indirect interactions among simple individual agents. Such operations could lead to systems that can adapt quickly to rapidly fluctuating conditions.

But the field is in its infancy. Because researchers lack a detailed understanding of the inner workings of insect swarms, identifying the rules by which individuals in those swarms interact has been a huge challenge, and without such information computer scientists have had trouble developing the appropriate software. In addition, although swarm-intelligence approaches have been effective at performing a number of optimization and control tasks, the systems developed have been inherently reactive and lack the necessary overview to solve problems that require in-depth reasoning techniques. Furthermore, one criticism of the field is that the use of autonomous insectlike agents will lead to unpredictable behavior in the computers they inhabit. This characteristic may actually turn out to be a strength, though, in that it could allow such systems to adapt to solve new, unforeseen problems—a flexibility that traditional software typically lacks.

Many futurists predict that chips will soon be embedded into thousands of mundane objects, from envelopes to trash cans to heads of lettuce. Enabling all these pieces of silicon to communicate with one another in a meaningful way will require novel approaches. As high-technology author Kevin Kelly puts it, "Dumb parts, properly connected into a swarm, yield smart results." The trick, of course, is in the proper connection of all the parts. SA

## Busy as a Bee



HONEYBEES (*top*) perform tasks based on the hive's needs. By studying the way these jobs are assigned, scientists hope to develop better ways to program equipment in an automated factory (*bottom*).

In a honeybee colony, individuals specialize in certain tasks, depending on their age. Older bees, for example, tend to be the foragers for the hive. But the allocation of tasks is not rigid: when food is scarce, younger nurse bees will forage, too.

Using such a biological system as a model, we worked with Michael Campos, now a postdoctoral fellow at the California Institute of Technology, to devise a technique for scheduling paint booths in a truck factory. The booths must paint trucks coming out of an assembly line, and each booth is like an artificial bee specializing in one color. The booths can change their colors if needed, but doing so is time-consuming and costly.

Because scientists have yet to understand exactly how honeybees regulate their division of labor, we made the following assumption: an individual performs the tasks for which it is specialized unless it perceives an important need to perform another function. Thus, a booth with red paint will continue to handle orders of that color unless an urgent job requires a white truck and the other booths, particularly those specializing in white, have much longer queues.

Although this basic rule sounds simplistic, in practice it is very effective. In fact, a honeybeelike system enables the paint booths to determine their own schedules with higher efficiency—specifically, fewer color changes—than a centralized computer can provide. And the method is adept at responding to changes in consumer demand. If the number of trucks that need to be painted blue surges unexpectedly, other booths can quickly forgo their specialty colors to accommodate the unassigned vehicles. Furthermore, the system copes easily with glitches. When a paint booth breaks down, other stations compensate swiftly by immediately divvying up the additional load. *—E.B. and G.T.*

HANS PFLETSCHINGER *Peter Arnold, Inc.* (*top*); PAUL A. SOUDERS *Corbis* (*bottom*)

### MORE TO EXPLORE

**Swarm Intelligence: From Natural to Artificial Systems.** Eric Bonabeau, Marco Dorigo and Guy Théraulaz. Oxford University Press, 1999.

**Self-Organization in Biological Systems.** S. Camazine et al. Princeton University Press, 2001.

**Swarm Robotics.** Edited by Erol Sahin, William M. Spears and Alan F. T. Winfield. Springer, 2007.

# Go Forth and Replicate

## Birds do it, bees do it, but could machines do it? Computer simulations suggest that the answer is yes

**Apples beget apples**, but can machines beget machines? Today it takes an elaborate manufacturing apparatus to build even a simple machine. Could we endow an artificial device with the ability to multiply on its own? Self-replication has long been considered one of the fundamental properties separating the living from the nonliving. Historically our limited understanding of how biological reproduction works has given it an aura of mystery and made it seem unlikely that it would ever be done by a man-made object. It is reported that when René Descartes averred to Queen Christina of Sweden that animals were just another form of mechanical automata, Her Majesty pointed to a clock and said, "See to it that it produces offspring."

The problem of machine self-replication moved from philosophy into the realm of science and engineering in the late 1940s with the work of eminent mathematician and physicist John von Neumann. Some researchers have actually constructed physical replicators. Almost 50 years ago, for example, geneticist Lionel Penrose and his son, Roger (the famous physicist), built small assemblies of plywood that exhibited a simple form of self-replication. But self-replication has proved to be so difficult that most researchers study it with the conceptual tool that von Neumann developed: two-dimensional cellular automata.

Implemented on a computer, cellular automata can simulate a huge variety of self-replicators in what amount to austere universes with different laws of physics from our own. Such models free researchers from having to worry about logistical issues such as energy and physical construction so that they can focus on the fundamental questions of information flow. How is a living being able to replicate unaided, whereas mechanical objects must be constructed by humans? How does replication at the level of an organism emerge from the numerous interactions in tissues, cells and molecules? How did Darwinian evolution give rise to self-replicating organisms?

The emerging answers have inspired the development of self-repairing silicon chips [*see box on pages 54 and 55*] and autocatalyzing molecules. And this may be just the beginning. Researchers in the field of nanotechnology have long proposed that self-replication will be crucial to manufacturing molecular-scale machines, and proponents of space exploration see a macroscopic version of the process as a way to

By Moshe Sipper and James A. Reggia

Photoillustrations by David Emmite

colonize planets using in situ materials. Recent advances have given credence to these futuristic-sounding ideas. As with other scientific disciplines, including genetics, nuclear energy and chemistry, those of us who study self-replication face the twofold challenge of creating replicating machines and avoiding dystopian predictions of devices running amok. The knowledge we gain will help us separate good technologies from destructive ones.

## Playing Life

SCIENCE-FICTION STORIES often depict cybernetic self-replication as a natural development of current technology, but they gloss over the profound problem it poses: how to avoid an infinite regress. A system might try to build a clone using a blueprint—that is, a self-description. Yet the self-description is part of the machine, is it not? If so, what describes the description? And what describes the description of the description? Self-replication in this case would be like asking an architect to make a perfect blueprint of his or her own studio. The blueprint would have to contain a miniature version of the blueprint, which would contain a miniature version of the blueprint, and so on. Without this information, a construction crew would be unable to re-create the studio fully; there would be a blank space where the blueprint had been.

Von Neumann's great insight was an explanation of how to break out of the infinite regress. He realized that the self-description could be used in two distinct ways: first, as the instructions whose interpretation leads to the construction of an identical copy of the device; next, as data to be copied, uninterpreted, and attached to the newly created child so that it, too, possesses the ability to self-replicate. With this two-step process, the self-description need not contain a description of itself. In the architectural analogy, the blueprint would include a plan for building a photocopy machine. Once the new studio and the photocopier were built, the construction crew would simply run off a copy of the blueprint and put it into the new studio.

Living cells use their self-description, which biologists call the genotype, in exactly these two ways: transcription (DNA is copied mostly uninterpreted to form mRNA) and translation (mRNA is interpreted to build proteins). Von Neumann made this transcription-translation distinction several years before molecular biologists did, and his work has been crucial in understanding self-replication in nature.

To prove these ideas, von Neumann and mathematician Stanislaw M. Ulam came up with the idea of cellular automata. A cellular-automata simulation involves a chessboardlike grid of squares, or cells, each of which is either empty or occupied by one of several possible components. At discrete intervals of time, each cell looks at itself and its neighbors and decides whether to metamorphose into a different component. In making this decision, the cell follows relatively simple rules, which are the same for all cells. These rules constitute the basic physics of the cellular-automata world. All decisions and actions take place locally; cells do not know directly what is happening outside their immediate neighborhood.

The apparent simplicity of cellular automata is deceptive; it does not imply ease of design or poverty of behavior. The most famous automaton, John Horton Conway's Game of Life, produces amazingly intricate patterns. Many questions about the dynamic behavior of cellular automata are formally unsolvable. To see how a pattern will unfold, you need to simulate it fully. In its own way, a cellular-automata model can be just as complex as the real world.

> Her Majesty of Sweden pointed to a clock and said, "See to it that it produces offspring."

## Copy Machines

WITHIN CELLULAR automata, self-replication occurs when a group of components—a "machine"—goes through a sequence of steps to construct a nearby duplicate of itself. Von Neumann's machine was based on a universal constructor, a machine that, given the appropriate instructions, could create any pattern. The constructor consisted of numerous types of components spread over tens of thousands of cells and required a book-length manuscript to be specified. It has still not been simulated in its entirety, let alone actually built, on account of its complexity. A constructor would be even more complicated in the Game of Life because the functions performed by single cells in von Neumann's model—such as transmission of signals and generation of new components—have to be performed by composite structures in Life.

Going to the other extreme, it is easy to find trivial examples of self-replication. For example, suppose that a cellular automaton has only one type of component, labeled +, and that each cell fol-

THE AUTHORS

*MOSHE SIPPER* and *JAMES A. REGGIA* share a long-standing interest in how complex systems can self-organize. Sipper is an associate professor in the department of computer science at Ben-Gurion University in Israel. He is interested mainly in evolutionary computation, primarily as applied to games and bioinformatics. Reggia is a professor of computer science and neurology, working in the Institute for Advanced Computer Studies at the University of Maryland. In addition to studying self-replication, he conducts research on computational models of the brain and its disorders, such as stroke.

lows only a single rule: if exactly one of the four neighboring cells contains a +, then the cell becomes a +; otherwise it becomes vacant. With this rule, a single + grows into four more +'s, each of which grows likewise, and so forth.

Such weedlike proliferation does not shed much light on the principles of replication, because there is no significant machine. Of course, that invites the question of how you would tell a "significant" machine from a trivially prolific automaton. No one has yet devised a satisfactory answer. What is clear, however, is that the replicating structure must in some sense be complex. For example, it must consist of multiple, diverse components whose interactions collectively bring about replication—the proverbial "whole must be greater than the sum of the parts." The existence of multiple, distinct components permits a self-description to be stored within the replicating structure.

In the years since von Neumann's seminal work, many researchers have probed the domain between the complex and the trivial, developing replicators that require fewer components, less space or simpler rules. A major step forward was taken in 1984 when Christopher G. Langton, then at the University of Michigan, observed that looplike storage devices—which had formed modules of earlier self-replicating machines—could be programmed to replicate on their own. These devices typically consist of two pieces: the loop itself, which is a string of components that circulate around a rectangle, and a construction arm, which protrudes from a corner of the rectangle into the surrounding space. The circulating components constitute a recipe for the loop—for example, "go three squares ahead, then turn left." When this recipe reaches the construction arm, the automata rules make a copy of it. One copy continues around the loop; the other goes down the arm, where it is interpreted as instructions.

By giving up the requirement of universal construction, which was central to von Neumann's approach, Langton showed that a replicator could be constructed from just seven unique components occupying only 86 cells. Even smaller and simpler self-replicating loops have been devised by one of us (Reggia) and our colleagues [*see box on next page*]. Because they have multiple interacting components and include a self-description, they are not trivial. Intriguingly, asymmetry plays an unexpected role: the rules governing replication are often simpler when the components are not rotationally symmetric than when they are.


## Emergent Replication

ALL THESE self-replicating structures have been designed through ingenuity and much trial and error. This process is arduous and often frustrating; a small change to one of the rules results in an entirely different global behavior, most likely the disintegration of the structure in question. But recent work has gone beyond the direct-design approach. Instead of tailoring the rules to suit a particular type of structure, researchers have experimented with various sets of rules, filled the cellular-automata grid with a "primordial soup" of randomly selected components and checked whether self-replicators emerged spontaneously.

In 1997 Hui-Hsien Chou, now at Iowa State University, and Reggia noticed that as long as the initial density of the free-floating components was above a certain threshold, small self-replicating loops reliably appeared. Loops that collided underwent annihilation, so there was an ongoing process of death as well as birth. Over time, loops proliferated, grew in size and evolved through mutations triggered by debris from past collisions. Although the automata rules were deterministic, these mutations were effectively random, because the system was complex and the components started in random locations.

Such loops are intended as abstract

machines and not as simulacra of anything biological, but it is interesting to compare them with biomolecular structures. A loop loosely resembles circular DNA in bacteria, and the construction arm acts as the enzyme that catalyzes DNA replication. More important, replicating loops illustrate how complex global behaviors can arise from simple local interactions. For example, components move around a loop even though the rules say nothing about movement; what is actually happening is that individual cells are coming alive, dying or metamorphosing in such a way that a pattern is eliminated from one position and reconstructed elsewhere—a process that we perceive as motion. In short, cellular automata act locally but appear to think globally. Much the same is true of molecular biology.

In a recent computational experiment, Jason Lohn, now at the NASA Ames Research Center, and Reggia experimented not with different structures but with different sets of rules. Starting with an arbitrary block of four components, they found they could determine a set of rules that made the block self-replicate. They discovered these rules via a genetic algorithm, an automated process that simulates Darwinian evolution.

The most challenging aspect of this work was the definition of the so-called fitness function—the criteria by which

## Build Your Own Replicator

**SIMULATING A SMALL** self-replicating loop using an ordinary chess set is a good way to get an intuitive sense of how these systems work. This particular cellular-automata model has four different types of components: pawns, knights, bishops and rooks. The machine initially comprises four pawns, a knight and a bishop. It has two parts: the loop itself, which consists of a two-by-two square, and a construction arm, which sticks out to the right.

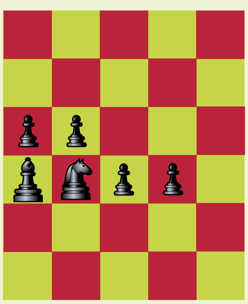
The knight and bishop represent the self-description: the knight, whose orientation is significant, determines which direction to grow, while the bishop tags along and determines how long the side of the loop should be. The pawns are fillers that define the rest of the shape of the loop, and the rook is a transient signal to guide the growth of a new construction arm.

As time progresses, the knight and bishop circulate counterclockwise around the loop. Whenever they encounter the arm, one copy goes out the arm while the original continues around the loop.

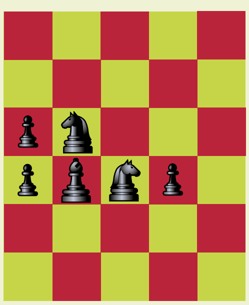
**HOW TO PLAY:** You will need two chessboards: one to represent the current configuration, the other to show the next configuration. For each round, look at each square of the current configuration, consult the rules and place the appropriate piece in the corresponding square on the other board. Each piece metamorphoses depending on its identity and that of the four squares immediately to the left, to the right, above and below. When you have reviewed each square and set up the next configuration, the round is over. Clear the first board and repeat. Because the rules are complicated, it takes a bit of patience at first. You can also view the simulation at www.cs.bgu.ac.il/~sipper/chessrep/src/chess.html

The direction in which a knight faces is significant. In the drawings here, we use standard chess conventions to indicate the orientation of the knight: the horse's muzzle points forward. If no rule explicitly applies, the contents of the square stay the same. Squares on the edge should be treated as if they have adjacent empty squares off the board. *—M.S. and J.A.R.*

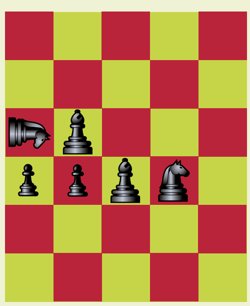
### *STAGES OF REPLICATION*



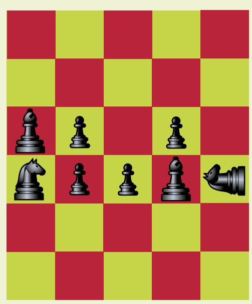
**INITIALLY,** the self-description, or "genome"—a knight followed by a bishop—is poised at the start of the construction arm.



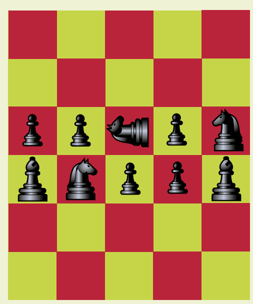
**1** The knight and bishop move counterclockwise around the loop. A clone of the knight heads out the arm.



**2** The original knight-bishop pair continues to circulate. The bishop is cloned and follows the new knight out the arm.



**3** The knight triggers the formation of two corners of the child loop. The bishop tags along, completing the gene transfer.



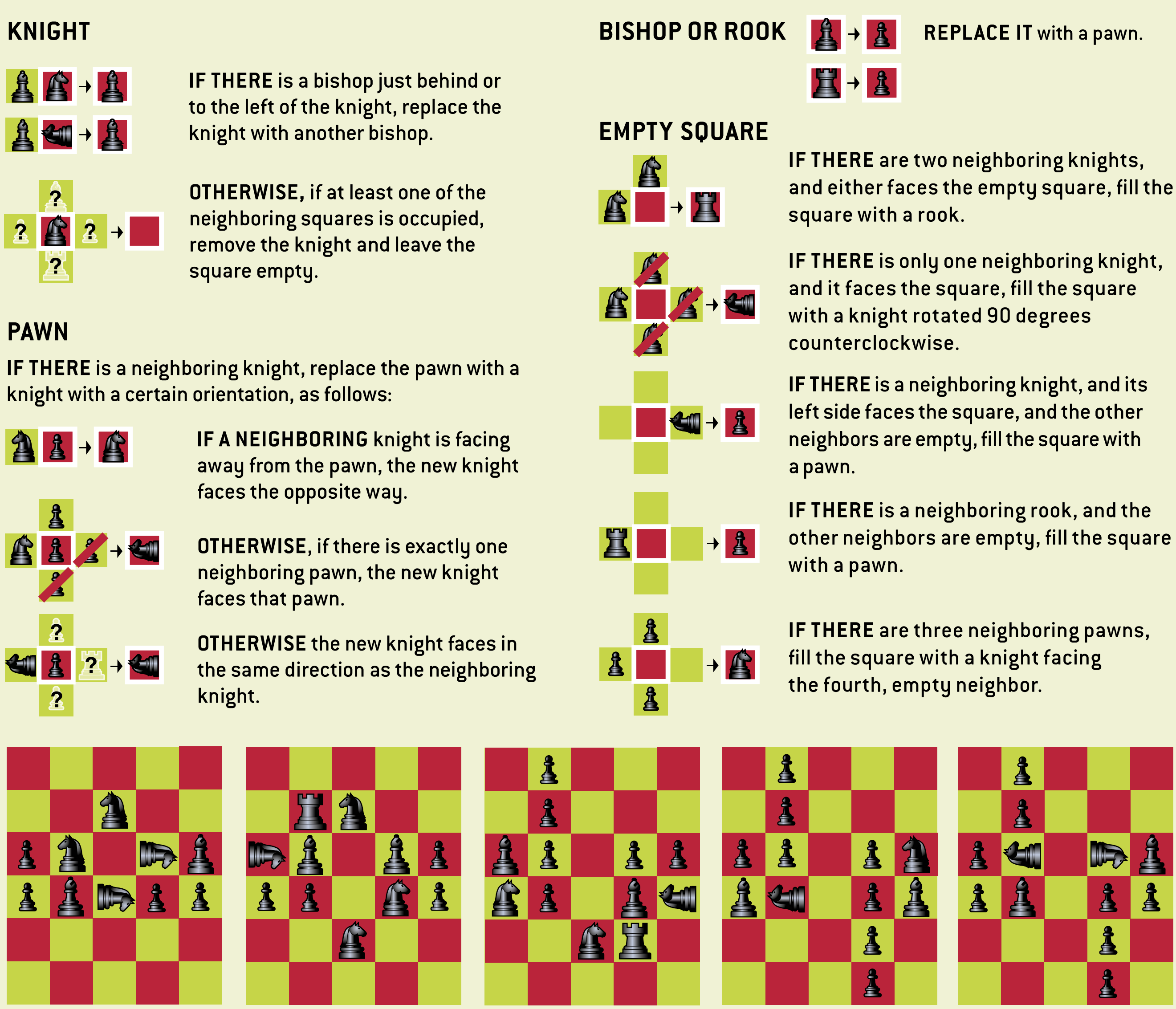
**4** The knight forges the remaining corner of the child loop. The loops are connected by the construction arm and a knight-errant.

GEORGE RETSECK

sets of rules were judged, thus separating good solutions from bad ones and driving the evolutionary process toward rule sets that facilitated replication. You cannot simply assign high fitness to those sets of rules that cause a structure to replicate, because none of the initial rule sets is likely to allow for replication. The solution was to devise a fitness function composed of a weighted sum of three measures: a growth measure (the extent to which each component type generates an increasing supply of that component), a relative position measure (the extent to which neighboring components stay together) and a replicant measure (a function of the number of actual replicators present). With the right fitness function, evolution can turn rule sets that are sterile into ones that are fecund; the process usually takes 150 or so generations.

Self-replicating structures discovered in this fashion work in a fundamentally different way than self-replicating loops do. For example, they move and deposit copies along the way—unlike replicating loops, which are essentially static. And although these newly discovered replicators consist of multiple, locally interacting components, they do not have an identifiable self-description—there is no obvious genome. The ability to replicate without a self-description may be relevant to questions about

*Continued on page 57*




**5** The knight-errant moves up to endow the parent with a new arm. A similar process, one step delayed, begins for the child loop.

**6** The knight-errant, together with the original knight-bishop pair, conjures up a rook. Meanwhile the old arm is erased.

**7** The rook kills the knight and generates the new, upward arm. Another rook prepares to do the same for the child.

**8** At last the two loops are separate and whole. The self-descriptions continue to circulate, but otherwise all is calm.

**9** The parent prepares to give birth again. In the following step, the child, too, will begin to replicate.

# Robot, Heal Thyself By George Musser

## Computers that fix themselves are the first application of artificial self-replication

Lausanne, Switzerland—Not many researchers encourage the wanton destruction of equipment in their labs. Daniel Mange, however, likes it when visitors walk up to one of his inventions and press the button marked KILL. The lights on the panel go out; a small box full of circuitry is toast. Early in May 2001 his team unveiled its latest contraption at a science festival here—a wall-size digital clock whose components you can zap at will—and told the public: Give it your best shot. See if you can crash the system.

The goal of Mange and his team is to instill electronic circuits with the ability to take a lickin' and keep on tickin'—just like living things. Flesh-and-blood creatures might not be so good at calculating $\pi$ to the millionth digit, but they can get through the day without someone pressing Ctrl-Alt-Del. Combining the precision of digital hardware with the resilience of biological wetware is a leading challenge for modern electronics.

Electronics engineers have been working on fault-tolerant circuits ever since there were electronics engineers. Computer modems would still be dribbling data at 1200 baud if it weren't for error detection and correction. In many applications, simple quality-control checks, such as extra data bits, suffice. More complex systems provide entire backup computers. The space shuttle, for example, has five processors. Four of them perform the same calculations; the fifth checks whether they agree and pulls the plug on any dissenter. The problem with these systems, though, is that they rely on centralized control. What if that control unit goes bad?

Nature has solved that problem through radical decentralization. Cells in the body are all basically identical; each takes on a specialized task, performs it autonomously and, in the event of infection or failure, commits hara-kiri so that its tasks can be taken up by new cells. These are the attributes that Mange, a professor at the Swiss Federal Institute of Technology here, and others have sought since 1993 to emulate in circuitry, as part of the "Embryonics" (embryonic electronics) project.

One of their earlier inventions, the MICTREE (microinstruction tree) artificial cell, consisted of a simple processor and four bits of data storage. The cell is contained in a plastic box roughly the size of a pack of Post-its. Electrical contacts run along the sides so that cells can be snapped together like Legos. As in cellular automata, the models used to study the theory of self-replication, the MICTREE cells are connected only to their immediate neighbors. The communication burden on each cell is thus independent of the total number of cells. The system, in other words, is easily scalable—unlike many parallel-computing architectures.

Cells follow the instructions in their "genome," a program written in a subset of the Pascal computer language. Like their biological antecedents, the cells all contain the exact same genome and execute part of it based on their position within the array, which each cell calculates relative to its neighbors.

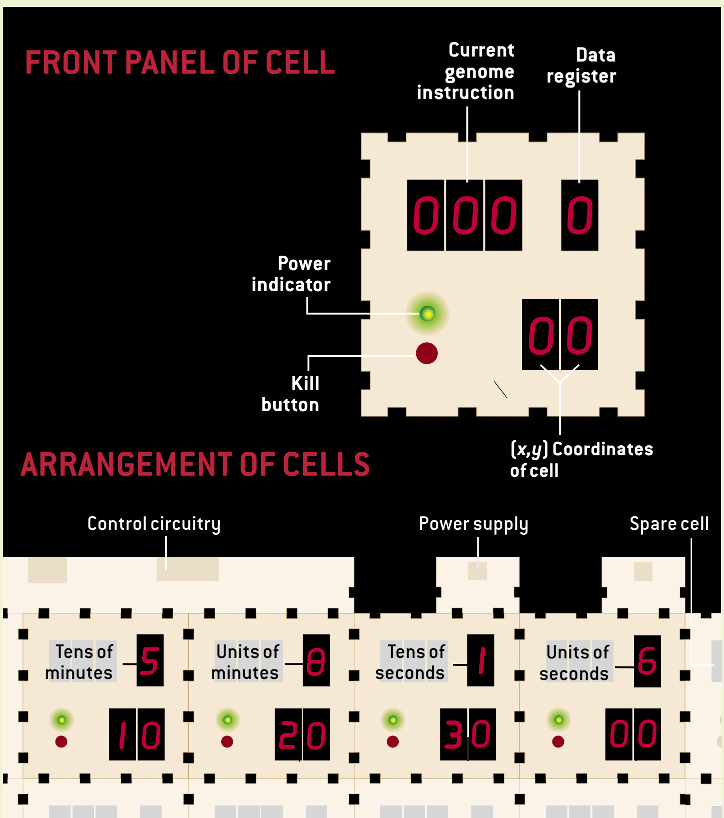
CRASH-PROOF COMPUTER is a two-dimensional array of artificial cells, each one a simple processor. In this application, four cells work together as a stopwatch, one cell per digit. Each cell counts up to either five or nine, depending on its coordinates within the array. The rest of the cells in the array are spares that take over if a cell fails or is killed. The Biodule 601 cells shown here are based on the MICTREE architecture described in the text.

© 1999 DELPHINE AURES *Eurelios*

Wasteful though it may seem, this redundancy allows the array to withstand the loss of any cell. Whenever someone presses the KILL button on a cell, that cell shuts down, and its left and right neighbors become directly connected. The right neighbor recalculates its position and starts executing the deceased's program. Its tasks, in turn, are taken up by the next cell to the right, and so on, until a cell designated as a spare is pressed into service.

Writing programs for any parallel processor is tricky, but the MICTREE array requires an especially unconventional approach. Instead of giving explicit instructions, the programmer must devise simple rules out of which the desired function will emerge. Being Swiss, Mange demonstrates by building a superreliable stopwatch. Displaying minutes and seconds requires four cells in a row, one for each digit. The genome allows for two cell types: a counter from zero to nine and a counter from zero to five. An oscillator feeds one pulse per second into the rightmost cell. After 10 pulses, this cell cycles back to zero and sends a pulse to the cell on its left, and so on down the line. The watch takes up part of an array of 12 cells; when you kill one, the clock transplants itself one cell over and carries on. Obviously, though, there is a limit to its resilience: the whole thing will fail after, at most, eight kills.

The prototype MICTREE cells are hardwired, so their processing power cannot be tailored to a specific application. In a finished product, cells would instead be implemented on a field-programmable gate array, a grid of electronic components that can be reconfigured on the fly. Mange's team custom-designed a gate array, known as MUXTREE (multiplexer tree), that is optimized for artificial cells. In the biological metaphor, the components of this array are the "molecules" that constitute a cell. Each consists of a logic gate, a data bit and a string of configuration bits that determines the function of this gate.




Building a cell out of such molecules offers not only flexibility but also extra endurance. Each molecule contains two copies of the gate and three of the storage bit. If the two gates ever give different results, the molecule kills itself for the greater good of the cell. As a last gasp, the molecule sends its data bit (preserved by the triplicate storage) and configuration to its right neighbor, which does the same, and the process continues until the rightmost molecule transfers its data to a spare. This second level of fault tolerance prevents a single error from wiping out an entire cell.

A total of 2,000 molecules, divided into four 20-by-25 cells, make up the BioWall—the giant digital clock that Mange's team put on display in 2001. Each molecule is enclosed in a small box and includes a KILL button and an LED display. Some molecules are configured to perform computations; others serve as pixels in the clock display. Making liberal use of the KILL buttons, I did my utmost to crash the system, something I'm usually quite good at. But the plucky clock just wouldn't submit. The clock display did start to look funny—numerals bent over as their pixels shifted to the right—but at least it was still legible, unlike most faulty electronic signs.

That said, the system did suffer from display glitches, which Mange attributed mainly to timing problems. Although the processing power is decentralized, the cells still rely on a central oscillator to coordinate their communications; sometimes they fall out of sync. Another Embryonics team, led by Andy Tyrrell of the University of York in England, has been studying making the cells asynchronous, like their biological counterparts. Cells would generate handshaking signals to orchestrate data transfers. The present system is also unable to catch certain types of error, including damaged configuration strings. Tyrrell's team has proposed adding watchdog molecules—an immune system—that would monitor the configurations (and one another) for defects.

Although these systems demand an awful lot of overhead, so do other fault-tolerance technologies. "While Embryonics appears to be heavy on redundancy, it actually is not that bad when compared to other systems," Tyrrell argues. Moreover, MUXTREE should be easier to scale down to the nano level; the "molecules" are simple enough to really be molecules. Says Mange, "We are preparing for the situation where electronics will be at the same scale as biology."

On a philosophical level, Embryonics comes very close to the dream of building a self-replicating machine. It may not be quite as dramatic as a robot that can go down to RadioShack, pull parts off the racks, and take them home to resolder a connection or build a loving mate. But the effect is much the same. Letting machines determine their own destiny—whether reconfiguring themselves on a silicon chip or reprogramming themselves using a neural network or genetic algorithm—sounds scary, but perhaps we should be gratified that machines are becoming more like us: imperfect, fallible but stubbornly resourceful.

---

*George Musser is an imperfect but resourceful staff editor and writer.*

*Continued from page 53*
how the earliest biological replicators originated. In a sense, researchers are seeing a continuum between nonliving and living structures.

Many researchers have tried other computational models besides the traditional cellular automata. In asynchronous cellular automata, cells are not updated in concert; in nonuniform cellular automata, the rules can vary from cell to cell. Another approach altogether is Core War and its successors, such as ecologist Thomas S. Ray's Tierra system. In these simulations the "organisms" are computer programs that vie for processor time and memory. Ray has observed the emergence of "parasites" that co-opt the self-replication code of other organisms.

## Getting Real

SO WHAT GOOD are these machines? Von Neumann's universal constructor can compute in addition to replicating, but it is an impractical beast. A major advance has been the development of simple yet useful replicators. In 1995 Gianluca Tempesti of the Swiss Federal Institute of Technology in Lausanne simplified the loop self-description so it could be interlaced with a small program—in this case, one that would spell the acronym of his lab, "LSL." His insight was to create automata rules that allow loops to replicate in two stages. First the loop, like Langton's loop, makes a copy of itself. Once finished, the daughter loop sends a signal back to its parent, at which point the parent sends the instructions for writing out the letters.

Drawing letters was just a demonstration. The following year Jean-Yves Perrier, Jacques Zahnd and one of us (Sipper) designed a self-replicating loop with universal computational capabilities—that is, with the computational power of a universal Turing machine, a highly simplified but fully capable computer. This loop has two "tapes," or long strings of components, one for the program and the other for data. The loops can execute an arbitrary program in addition to self-replicating. In a sense, they are as complex as the computer that simulates them. Their main limitation is that the program is copied unchanged from parent to child, so all loops carry out the same set of instructions.

In 1998 Chou and Reggia swept away this limitation. They showed how self-replicating loops carrying distinct information, rather than a cloned program, can be used to solve a problem known as satisfiability. The loops can be used to determine whether the variables in a logical expression can be assigned values such that the entire expression evaluates to "true." This problem is NP-complete—in other words, it belongs to the family of nasty puzzles, including the famous traveling salesman problem, for which there is no known efficient solution. In Chou and Reggia's cellular-automata universe, each replicator received a different partial solution. During replication, the solutions mutated, and replicators with promising solutions were allowed to proliferate while those with failed solutions died out.

Although various teams have created cellular automata in electronic hardware, such systems are probably too wasteful for practical applications; automata were never really intended to be implemented directly. Their purpose is to illuminate the underlying principles of replication and, by doing so, to inspire more concrete efforts. The loops provide a new paradigm for designing a parallel computer from either transistors or chemicals.

In 1980 a NASA team led by Robert Freitas, Jr., proposed planting a factory on the moon that would replicate itself, using local lunar materials, to populate a large area exponentially. Indeed, a similar probe could colonize the entire galaxy, as physicist Frank J. Tipler of Tulane University has argued. In the nearer term, computer scientists and engineers have experimented with the automated design of robots. Although these systems are not truly self-replicating—the offspring are much simpler than the parent—they are a first step toward fulfilling the queen of Sweden's request.

Should physical self-replicating machines become practical, they and related technologies will raise difficult issues, including the *Terminator* film scenario in which artificial creatures outcompete natural ones. We prefer the more optimistic, and more probable, scenario that replicators will be harnessed to the benefit of humanity. The key will be taking the advice of 14th-century English philosopher William of Ockham: *entia non sunt multiplicanda praeter necessitatem*—entities are not to be multiplied beyond necessity. SA

In a sense, researchers are seeing a continuum between nonliving and living structures.

### MORE TO EXPLORE

**Simple Systems That Exhibit Self-Directed Replication.** J. Reggia, S. Armentrout, H. Chou and Y. Peng in *Science,* Vol. 259, pages 1282–1287; February 26, 1993.

**Emergence of Self-Replicating Structures in a Cellular Automata Space.** H. Chou and J. Reggia in *Physica D,* Vol. 110, Nos. 3–4, pages 252–276; December 15, 1997.

**Special Issue: Von Neumann's Legacy: On Self-Replication.** Edited by M. Sipper, G. Tempesti, D. Mange and E. Sanchez in *Artificial Life,* Vol. 4, No. 3; Summer 1998.

**Towards Robust Integrated Circuits: The Embryonics Approach.** D. Mange, M. Sipper, A. Stauffer and G. Tempesti in *Proceedings of the IEEE,* Vol. 88, No. 4, pages 516–543; April 2000.

Moshe Sipper's Web page on artificial self-replication is at **www.cs.bgu.ac.il/~sipper/selfrep**

Animations of self-replicating loops can be found at **www.necsi.org/postdocs/sayama/sdsr/java**

For John von Neumann's universal constructor, see **www.sq3.org.uk/Evolution/JvN**

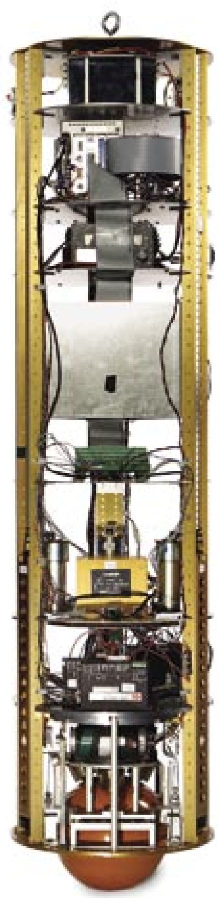
A new mode of locomotion will enable mobile robots to stand tall and move gracefully through busy everyday environments

# BALLBOTS

By Ralph Hollis

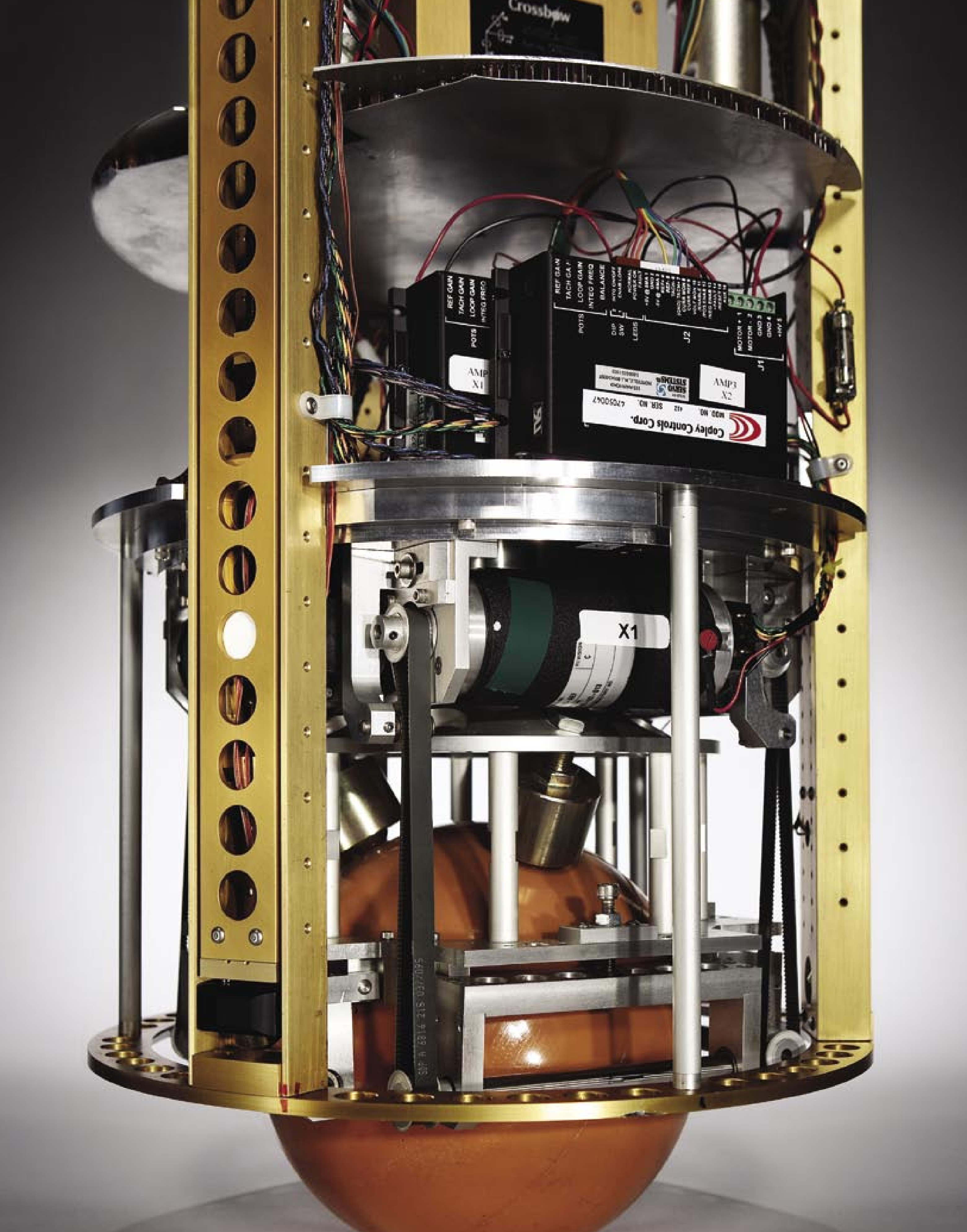
The dream of intelligent mobile robots that assist people during their day-to-day activities in homes, offices and nursing facilities is a compelling one. Although a favorite subject of science-fiction writers and robotics researchers, the goal seems always to lie well off in the future, however. Engineers have yet to solve fundamental problems involving robotic perception and world modeling, automated reasoning, manipulation of objects, and locomotion.

Researchers have produced robots that, while falling far short of the ideal, can do some remarkable things. In 2002 one group dropped off a robot at the entrance to the annual meeting of the American Association for Artificial Intelligence in Edmonton, Alberta. The clever machine soon found its way to the registration booth, signed up for the conference, was assigned a lecture room, proceeded to that location and finally presented a brief talk about itself at the appointed hour. Some robots have in the meantime served effectively as interactive museum tour guides, whereas others show promise as nursing home assistants. Computer scientists and engineers have also equipped mobile systems with arms and hands for manipulating objects. All these experimental devices travel about on bases supported by three or four wheels. Designers call this configuration



MOBILE ROBOTICS takes a different path with the ballbot's unique single, spherical drive-wheel design.

BRIAN MARANAN PINEDA


Crossbow
REF GAIN
TACH GAIN
LOOP GAIN
INTEG FREQ
BALANCE
POTS
AMP3
X2
Copley Controls Corp.
X1

"statically stable" because it keeps the robots upright even at rest.

Robots tall enough to interact effectively in human environments have a high center of gravity and must accelerate and decelerate slowly, as well as avoid steep ramps, to keep from falling over. To counter this problem, statically stable robots tend to have broad bodies on wide wheelbases, which greatly restricts their mobility through doorways and around furniture or people.

Several years ago I decided to sidestep the need for large wheelbases by designing and building a tall, skinny and agile robot that balances on, and is propelled by, a single spherical wheel. Such a simple machine, with its high center of gravity, would be able to move quickly in any direction. The system would rely on active balancing and thus be "dynamically stable"—that is, it would remain erect only if it made continual corrections to its body attitude. I realized this design would constitute a hitherto unstudied class of wheeled mobile robots. For lack of anything better, I called it a ballbot.

My students and I have operated our ballbot now for several years, studying its stability properties and suitability for operating in human environments. During that time, many visitors to our laboratory have found its uncanny ability to balance and roam about on a single spherical wheel to be quite remarkable.

## Maintaining Balance

WE HUMANS KEEP BALANCE with help from the vestibular senses in our inner ears. This information is combined with input from other senses, such as vision, to control muscles in our legs and feet to enable us to stand upright without falling down. A ballbot maintains equilibrium in a somewhat analogous fashion. First, the machine must have some goal to achieve, such as to remain in one place or to move in a straight line between two locations. Second, it must always know the direction of gravity's pull and be able to measure the orientation of its body with respect to this vertical reference. Third, it must have means to rotate the ball in any direction and to measure its travel along the floor. Finally, the ballbot must have a method, or control policy, that processes the sensor data it measures to generate commands for ball rotation that attempt to satisfy the goals.

**Many visitors find its uncanny ability to balance and roam about on a single spherical wheel to be quite remarkable.**

Solving the "problem of the vertical" has proved to be a challenging exercise throughout history [*see box on page 62*]. Our solution takes advantage of tremendous recent advances in computing, fiber optics and microelectromechanical systems (MEMS) that have enabled the production of low-cost devices that emulate the function of the traditional spinning gyroscope.

We use a system that features three fiber-optic gyroscopes mounted orthogonally (at right angles to one another) in a box that is rigidly attached to the ballbot body [*see box on opposite page*]. These gyroscopes contain no rotating masses. Each gyroscope features a light source, a detector and a coil of optical fiber. Light waves travel around the coil in opposite directions and interfere with one another at the detector. During operation, the ballbot body, with its three gyroscopic, angular-motion sensors, rotates in various directions, but the light waves inside them travel at a fixed speed regardless of any movement. Accordingly, a small path difference between the clockwise- and counterclockwise-propagating waves results in each sensor. In each case, the path difference causes the interference fringes at the detector to shift, producing an output that is proportional to angular velocity, an effect noted as far back as 1913 by French physicist Georges Sagnac. A small computer integrates the three angular velocities to produce pitch (forward/backward tilt), roll (left/right tilt) and yaw (rotation around the vertical) angles taken by the robot's body.

To report the correct vertical orientation, all gyroscopes must take into account the earth's rotation. They are also subject to numerous other small effects that cause errors and drift over time. Our system incorporates three MEMS accelerometers, set orthogonally in the same box alongside the gyroscopes. As the ballbot moves around, these sensors report the resulting instantaneous acceleration values for each orientation, which the computer then combines to yield an overall acceleration direction and magnitude that can be averaged over time. (The accelerometers' readings cannot be used directly for balancing.) The outcome is a reliable long-term indicator of the direction of gravity that the system uses to correct the drift of the fiber-optic gyroscopes.

## Moving with the Ball

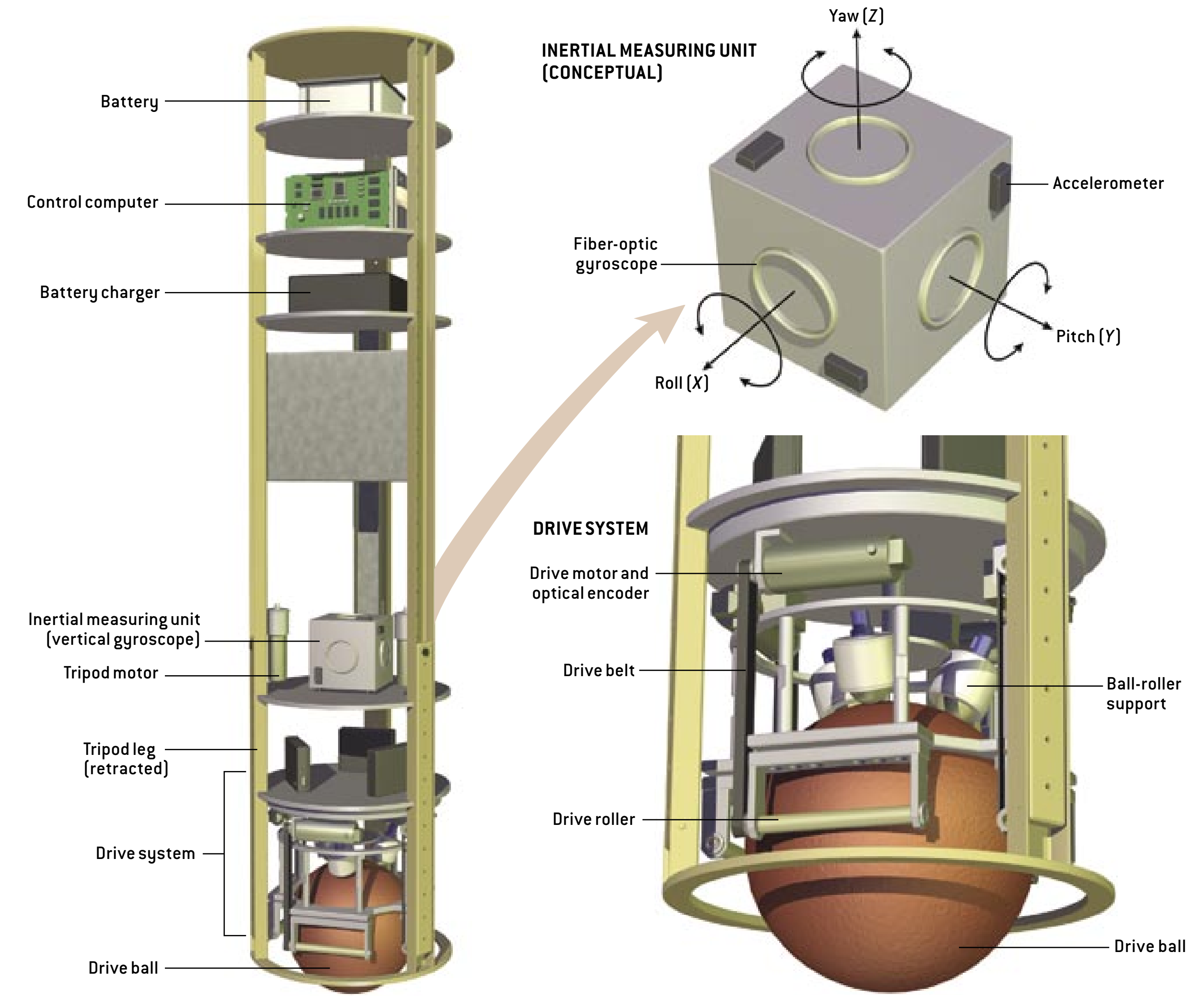
SEVERAL METHODS EXIST for driving a ball in various directions using motors. We strove for simplicity in our design for the ballbot's drive mechanism. When one moves a mechanical computer mouse about on the desktop, the rubber-coated ball on the underside causes a

## Overview/*Mobile Robots*

- To interact with people in their everyday environments, intelligent mobile robots will need to stand tall, as well as to move surely and gracefully.
- Most current experimental mobile robots feature wide wheelbases, which hinder their movements through cramped, chaotic human settings.
- A ballbot—a tall, thin robot that travels about on a ball-shaped wheel that enables it to move rapidly in any direction—may provide the flexible locomotive capabilities that future robots will need to aid people in their daily lives.

## BALLBOT ARCHITECTURE

In some ways, a ballbot (*left*) resembles a ballpoint pen that is five feet tall. The fiber-optic gyroscopes and accelerometers (*top right*), which are mounted at right angles to one another to sense motion in the pitch, roll and yaw directions, generate the vertical orientation data the computer control policy needs to determine how to maintain balance [*see box on next page for explanation of underlying principles*]. The drive-ball mechanism (*bottom right*), which operates something like an inverse computer mouse, provides the ballbot's motive force. Motorized drive rollers turn the ball, and optical encoders measure the ballbot's travel. To stay upright when shut down, the machine deploys its tripod legs.




GEORGE RETSECK

pair of orthogonally mounted rollers to turn. The measured rotation of the rollers provides input to the computer to traverse the cursor across the screen. Just the opposite happens in the ballbot: output from the ballbot's computer commands a set of motors to turn rollers that rotate the ball, thus causing the robot to travel in any direction along the floor. It is essentially an "inverse mouse ball" drive. Currently motors actuate the ball in the pitch and roll directions. An additional motor (not yet installed) will rotate the body in yaw, which will allow the ballbot to face in any direction.

Much as a circus clown might perch atop a ball, the ballbot's body stands atop the ball wheel. The ball is a hollow aluminum sphere covered with a thick layer of polyurethane rubber. Such a drive scheme exhibits frictional and damping behavior because sliding always occurs between the ball and rollers, for which compensation must be made. Three ball bearings between the ball and body support the body's weight.

To infer ball rotation and hence travel distance, we used optical encoders that are fitted to each of the drive motors. Each encoder has a fixed light source opposite a light detector. A transparent, rotating mask (with many fine opaque stripes) attached to the motor shaft sits between them. As the motor turns, the mask rotates, causing the striped pattern to alternately block and transmit the light beam. The ballbot's main computer counts these events to measure ball rotation and thus distance traveled.

## The Problem of the Vertical

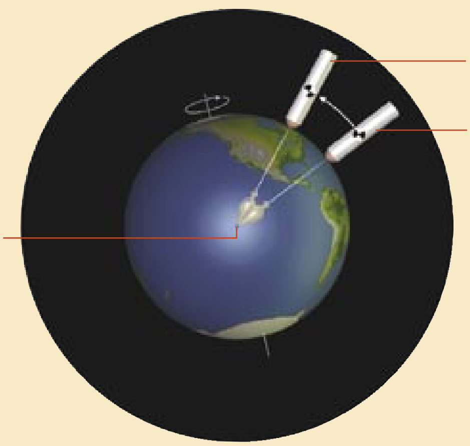
Finding the up/down orientation, what early aviators called the problem of the vertical, continues to be difficult even today. A plumb bob hanging from a string reveals the vertical, but a ballbot equipped with such a pendulum reference would become confused because motion (say, from position A to B, below) would cause the bob to swing to and fro.

Alternatively, the ballbot could rely on a gyroscope. The gyro's wheel would be supported by gimbals, which would allow its axis to point arbitrarily. By driving the wheel with a motor, it could be spun rapidly with its axis aligned vertically before the ballbot began to operate. The inertia of such a gyro would keep it pointing in the same direction regardless of movement. Equipping the gimbals with angle sensors would allow measurement of the body's forward/backward (pitch) and its left/right (roll) attitudes. This approach has problems, however. The gyro's axis would remain fixed in space while the earth rotates and hence would depart from the vertical.

German engineer Maximilian Schuler first formulated a solution to this problem in 1923 by imagining a pendulum string long enough to reach the center of the earth. Such a long string would always point downward regardless of motion. This pendulum would, in fact, have a period of about 84.4 minutes, the so-called Schuler period. He showed how small torques exerted on a gyroscope could increase the period of a short, practical pendulum to 84.4 minutes (and thus make it behave like a Schuler pendulum), which would keep it oriented along the direction of gravity.

The ballbot could, in theory, use such a gyro with a short pendulum. As the ballbot moves, the directions of the pendulum's swing could be measured over time and averaged to yield a value that faithfully represented the vertical (because the lateral accelerations would cancel out over time, leaving gravity dominant). The result could be used to exert torques on the gyro to make it stay vertical.

We opted for another solution. Our ballbot uses fiber-optic gyroscopes and microelectromechanical accelerometers that together emulate the functions of a mechanical gyro and pendulum that behaves like a Schuler pendulum. The result is a gravity-seeking, or "vertical," gyro that serves as a reference for balancing. *—R.H.*




THE AUTHOR

*RALPH HOLLIS* is a research professor at Carnegie Mellon University's Robotics Institute, with an appointment in the department of electrical and computer engineering. He received his Ph.D. in solid-state physics from the University of Colorado at Boulder in 1975. Before joining Carnegie Mellon in 1993, Hollis worked at North American Aviation and the IBM Thomas J. Watson Research Center. His current research focuses on agent-based microassembly of electromechanical products, human-computer interaction through the sense of touch, and dynamically stable mobile robots.

## Ball Control

SIMPLY STATED, the ballbot uses its knowledge of the vertical to determine how to rotate its ball to balance and move about. Fortunately, the ballbot is fundamentally an inverted pendulum, a mechanism that physicists have studied extensively. We use the techniques of optimal control theory to find a strategy or policy for driving the ballbot to its goal while simultaneously minimizing the effort it takes to get there. The ballbot has eight internal states that the policy must take into account: four for its forward/backward motion and four for its left/right motion. For each of these directions, the system measures or infers (from the onboard sensors) the robot's position and speed, as well as the tilt and tilt rate of the body.

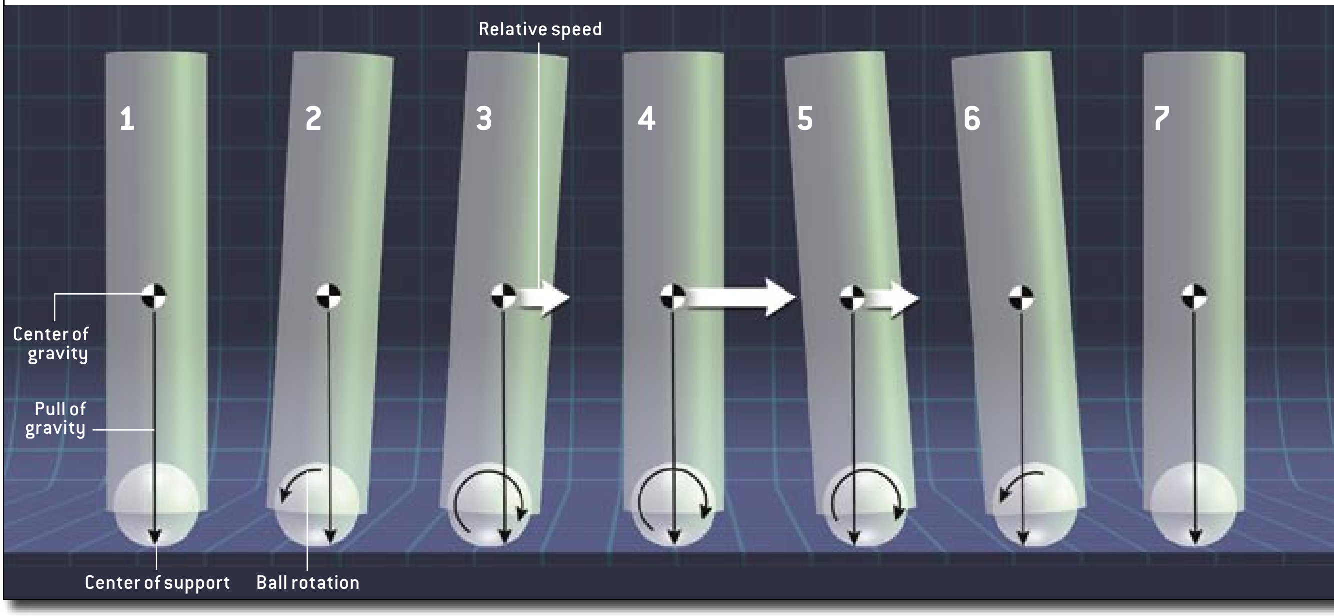
We employ a simplified linear mathematical model to describe the ballbot's dynamics. Rudolf Kalman, a Hungarian-American mathematical system theorist, invented in 1960 an elegant method for deriving control policies for such systems, which he called the linear quadratic regulator. This approach considers the measurements of the system's internal states to be proportional to the values of the states themselves. Further, it assumes that the states change over time at a rate proportional to the values of the states plus a proportional contribution of any control actions that might occur, such as motor torques. Kalman's technique cleverly minimizes an integral function over time that includes a quadratic measure of the states plus a quadratic measure of the control actions. Its solution yields a final set of constants, which, when multiplied by each of the internal states, gives a recommended, or optimal, control action for the ballbot to take at each moment in time. These calculations run several hundred times a second in the ballbot's main computer.

When the ballbot's goal is to stand still, its control policy tries to simultaneously drive the body's position and speed, as well as its tilt and tilt rate, to zero in each direction, while minimizing the actions needed to do so. When its objective is to go from one place to another, the control policy automatically

GEORGE RETSECK

## TRAVELING FROM HERE TO THERE

To maintain balance when still, the ballbot must keep its center of gravity directly over its center of support (*1*). Orientation sensors determine the vertical direction, which the machine then compares with its current attitude. During movement, the ballbot manipulates its center of gravity to best effect. To go from one point to another on level ground, for example, the drive ball first rotates slightly in the direction opposite to the intended direction of travel (*2*), which tilts the body forward a bit to initiate the move. Next, the ball spins in the direction of motion to accelerate ahead (*3*). While the ballbot is at constant velocity, the body must remain nearly vertical (*4*). The opposite actions must occur to decelerate the machine (*5*) and then prepare it to halt (*6*), which together bring it to a stop (*7*). When traversing inclines, the body must lean into slopes to keep its equilibrium.




institutes a retrograde ball rotation to establish a body tilt, allowing it to accelerate forward. As the goal position is approached, the ball automatically speeds up to reverse the tilt and bring the ballbot to rest [*see box above*].

## Moving Ahead

WE HAVE BEGUN to experiment with the ballbot, interacting with it over a wireless radio link. We plan to add a pair of arms, as well as a head that pans and tilts, with a binocular vision system and many other sensors, in an effort to develop the machine into a capable robot with a significant degree of autonomy. Our goals are to understand how well such robots can perform around people in everyday settings and to compare quantitatively its performance, safety and navigation abilities with those of traditional, statically stable robots. Our hypothesis is that the latter may turn out to be an evolutionary dead end when it comes to operating in such environments.

We are not alone in betting on the notion of dynamically stable robots. Other research groups have produced two-wheeled robots that are dynamically stable in the pitch direction but statically stable in the roll orientation. Although these robots are not omnidirectional like a ballbot is, they show promise for agile mobility—especially outdoors.

It may turn out that dynamically stable biped robots, perhaps in humanoid form, will have the long-term edge—particularly for their ability to deal with stairways. Research teams worldwide are working intensively to develop these complex and often expensive machines. Meanwhile it would seem that ballbots will serve as interesting and effective platforms for studying how mobile robots can interact dynamically and gracefully with humans in the places where people live. SA

GEORGE RETSECK

### MORE TO EXPLORE

For a discussion of gyroscopic principles, see **Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance.** Reprint edition. Donald MacKenzie. MIT Press, 1993.

**A Dynamically Stable Single-Wheeled Mobile Robot with Inverse Mouse-Ball Drive.** T. B. Lauwers, G. A. Kantor and R. L. Hollis in *Proceedings of the 2006 IEEE International Conference on Robotics and Automation (ICRA '06),* May 2006.

**One Is Enough!** Tom Lauwers, George Kantor and Ralph Hollis in *Robotics Research: Results of the 12th International Symposium ISRR.* Springer Tracts in Advanced Robotics, Vol. 28, pages 327–336; 2007.

**Socially Distributed Perception: GRACE Plays Social Tag at AAAI 2005.** M. P. Michalowski, S. Šabanović, C. DiSalvo, D. Busquets, L. M. Hiatt, N. A. Melchior and R. Simmons in *Autonomous Robots,* Vol. 22, No. 4, pages 385–397; May 2007.

Ballbot information (including demonstration videos): **www.msl.ri.cmu.edu/projects/ballbot**

Information on the linear quadratic regulator: **http://en.wikipedia.org/wiki/Linear-quadratic_regulator**

Information on accelerometer principles: **www.efunda.com/formulae/vibrations/sdof_eg_accelerometer.cfm**

ENGINEERS AIM to develop electroactive polymers that can lengthen and contract like human muscles. Eventually these creations may even surpass our capabilities.

# Artificial Muscles

Novel motion-producing devices—actuators, motors, generators—based on polymers that change shape when stimulated electrically are reaching commercialization

By Steven Ashley

It's only a $100 toy—an aquarium of swimming robotic fish developed by the Eamex Corporation in Osaka, Japan. What makes it remarkable is that the brightly colored plastic fish propelling themselves through the water in a fair imitation of life do not contain mechanical parts: no motors, no driveshafts, no gears, not even a battery. Instead the fish swim because their plastic innards flex back and forth, seemingly of their own volition. They are the first commercial products based on a new generation of improved electroactive polymers (EAPs), plastics that move in response to electricity.

For decades, engineers who build actuators, or motion-generating devices, have sought an artificial equivalent of muscle. Simply by changing their length in response to nerve stimulation, muscles can exert controlled amounts of force sufficient to blink an eyelid or hoist a barbell. Muscles also exhibit the property of scale invariance: their mechanism works equally efficiently at all sizes, which is why fundamentally the same muscle tissue powers both insects and elephants. Something like muscle might therefore be useful in driving devices for which building tiny electric motors is not easily accomplished.

EAPs hold promise for becoming the artificial muscles of the future. Investigators are already ambitiously working on EAP-based alternatives to many of today's technologies. And they aren't afraid to pit their creations against nature's. A few years ago several individuals, including Yoseph Bar-Cohen, a senior research scientist at the Jet Propulsion Laboratory (JPL) in Pasadena, Calif., posted a challenge to the electroactive polymer research community to drum up interest in the field: a race to build the first EAP-driven robotic arm that could beat a human arm wrestler one on one. Later, they began searching for sponsors to subsidize a cash prize for the winner. The first such contest was held in March 2005, and the outcome was disappointing for robot designers: a 17-year-old girl easily defeated her three mechanized opponents, each demonstrating a different type of artificial muscle.

Research continued despite this result, and perhaps the most promising of the current EAP efforts is being conducted by SRI International, a nonprofit contract-research laboratory based in Menlo Park, Calif. Another pioneer in the field of EAPs is Micromuscle AB, a company based in Linköping, Sweden, that focuses on medical device applications in the areas of cardiovascular treatment and drug delivery.

In 2003 SRI launched a spin-off company, Artificial Muscle, Inc. (AMI), to commercialize the EAP technology it had patented. AMI now manufactures actuators and transducers (touch sensors) that employ its electroactive polymer artificial-muscle technology.

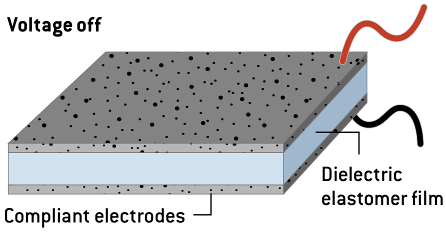
## HOW ELECTRICITY MAKES A PLASTIC EXPAND

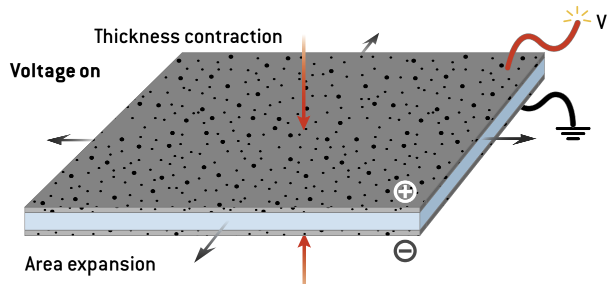
The fundamental mechanism underlying new artificial-muscle products is relatively simple. When exposed to high-voltage electric fields, dielectric elastomers—such as silicones and acrylics—contract in the direction of the electric field lines and expand perpendicularly to them, a phenomenon physicists term Maxwell stress. The new devices are basically rubbery capacitors—two charged parallel plates sandwiching a dielectric material. When the power is on, plus and minus charges accumulate on the opposite electrodes. They attract each other and squeeze down on the polymer insulator, which responds by expanding in area.

Engineers laminate thin films of dielectric elastomers (typically 30 to 60 microns thick) on the front and back with conductive carbon particles suspended in a soft polymer matrix. When connected by wires to a power source, the carbon layers serve as flexible electrodes that expand in area along with the material sandwiched in the middle. This layered plastic sheet serves as the basis for a wide range of novel actuation, sensory and energy-generating devices.

Dielectric elastomers, which can grow by as much as 400 percent of their nonactivated size, are by no means the only types of electroactive materials or devices, although they represent some of the more effective examples.

The graph at the right compares the performance of various classes of actuation materials and devices. These include well-established motion-generating products driven by electric current as well as applied electrostatic and electromagnetic fields. Strain refers to the amount of displacement or travel per unit length the




These solid-state devices are intended for use in audio speakers, power generators, motors, pumps, valves, sensors and actuators. The company's Universal Muscle Actuator is the first high-production-volume platform that can serve as a fundamental building block for advanced linear actuator designs. AMI recently introduced, for example, the DLP-95 autofocus lens positioner, a compact device that adjusts lenses for focusing and zooming.

The firm's long-term goal? Only to replace a substantial number of the myriad electric motors we use regularly, not to mention many other common motion-generating mechanisms, with smaller, lighter, cheaper products using SRI's novel actuators. "We believe this technology has a good chance to revolutionize the field of mechanical actuation," states Philip von Guggenberg, the lab's director of business development. "We'd like to make the technology ubiquitous, the kind of thing you could pick up in hardware stores."

### Materials That Move

BAR-COHEN HAS SERVED as the unofficial coordinator for the highly diverse community of international EAP researchers since the mid-1990s. Back during the field's infancy, "the electroactive polymer materials I read about in scientific papers didn't work as advertised," he recalls, chuckling slyly. "And as I already had obtained NASA funding to study the technology, I was forced to look around to find who was working in this area to find something that did." Within a few years Bar-Cohen had learned enough to help establish the first scientific conference on the topic, start publishing an EAP newsletter, post an EAP Web site and edit two books on the nascent technology.

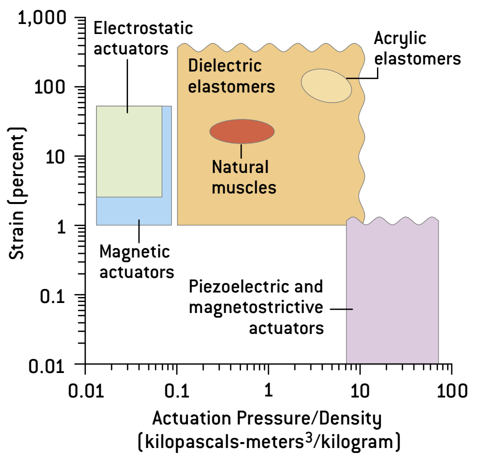
Sitting among arrays of lab tables strewn with prototype actuation devices and test apparatuses in a low-slung research building on the JPL campus, Bar-Cohen reviews the history of the field he has come to know so well. "For a long time," he begins, "people have been working on ways to move objects without electric motors, which can be too heavy and bulky for many applications. Until the development of EAPs, the standard replacement technology for motors were piezoelectric ceramics, which have been around for some time."

In piezoelectric materials, mechani-

## Overview/*Electroactive Polymers*

- Physicists and chemists have long sought to develop lightweight materials that grow or shrink significantly in length or volume when subjected to electric stimulation. Such substances could serve as the drivers for novel motion-generating devices (generally called actuators)—possible replacements for the ubiquitous electric motor, which is often too bulky and heavy for smaller-scale applications.
- A new generation of electroactive polymer materials displays sufficient physical response to electrical excitation to power new classes of actuators as well as innovative sensors and energy generators. Products based on this "artificial muscle" technology are just starting to hit the market.

JAMIE KRIPKE (*preceding pages*); SARA CHEN (*above and opposite page*) devices can create. Actuation pressure/density is a measure of the force they produce. Dielectric elastomers can generate more strain and force than many of the competing technologies. Their properties in this regard are similar to those of natural animal muscle—hence the moniker "artificial muscles."




cal stress causes crystals to electrically polarize, and vice versa. Hit them with electric current, and they deform; deform them, and they generate electricity.

Bar-Cohen lifts a small grayish disk off one of the lab benches, saying, "This one's made of PZT—lead zirconate titanate." He explains that electric current makes the piezoelectric PZT shrink and expand by a fraction of a percent of its total length. Not much motion but useful nonetheless.

In an adjoining room, Bar-Cohen shows off foot-long impact drills driven by PZT disks that he is building with his JPL colleagues and Cybersonics, Inc., in Erie, Pa. "Inside this cylinder is a stack of piezoelectric disks," he states. "When activated with alternating current, the stack beats ultrasonically on a mass that hops up and down at a high rate, driving a bit into solid rock." To one side sit piles of stone blocks into which drill bits have cut deep holes.

As a demonstration of how effectively piezoceramics can perform as actuators, it is impressive. But many applications would demand electroactive materials that grow by more than just a fraction of a percent.

## Plastics That React

POLYMERS THAT change shape in response to electricity, according to Bar-Cohen, can be sorted into two groups: ionic and electronic types, each with complementary advantages and disadvantages.

Ionic EAPs (which include ionic polymer gels, ionomeric polymer-metal composites, conductive polymers and carbon nanotubes) work on the basis of electrochemistry—the mobility or diffusion of charged ions. They can run directly off batteries because even single-digit voltages will make them bend significantly. The catch is that they generally need to be wet and so must be sealed within flexible coatings. The other major shortcoming of many ionic EAPs (especially the ionomeric polymer-metal composites) is that "as long as the electricity is on, the material will keep moving," Bar-Cohen notes, adding: "If the voltage is above a certain level, electrolysis takes place, which causes irreversible damage to the material."

In contrast, electronic EAPs (such as ferroelectric polymers, electrets, dielectric elastomers and electrostrictive graft elastomers) are driven by electric fields. They require relatively high voltages, which can cause uncomfortable electric shocks. But in return, electronic EAPs can react quickly and deliver strong mechanical forces. They do not need a protective coating and require almost no current to hold a position.

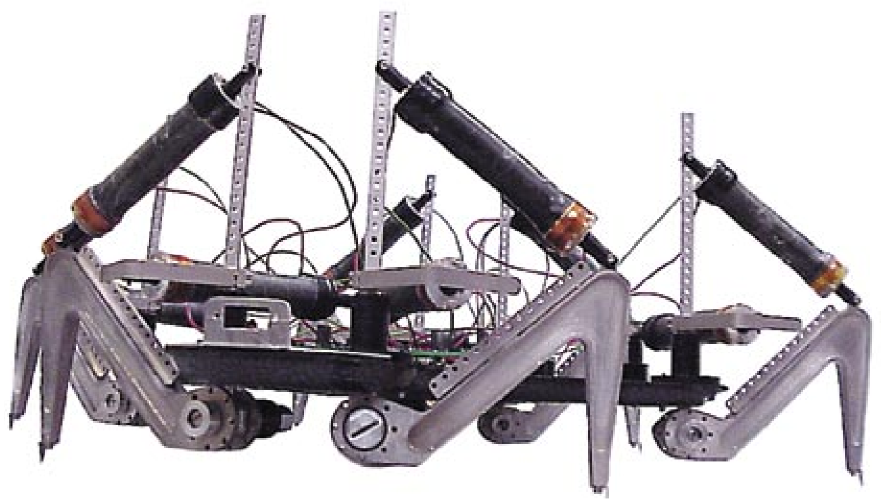
SRI's artificial-muscle material falls into the electronic EAP classification. The long, bumpy and sometimes serendipitous road to its successful development is a classic example of the vagaries of technological innovation.

## Electrifying Rubber

"SRI INTERNATIONAL began work on artificial muscles in 1992 under contract to the Japanese micromachine program," says Ron Pelrine, the physicist-turned-mechanical engineer who leads the SRI team. Japanese officials were looking for a new kind of microactuator technology. A few SRI scientists started searching for a motion-generating material that resembled natural muscle in terms of force, stroke (linear displacement) and strain (displacement per unit length or area).

"We looked at a whole bunch of possible actuation technologies," Pelrine recalls. Eventually, however, they considered electrostrictive polymers, a class of materials then being investigated by Jerry Scheinbeim of Rutgers University. The hydrocarbon molecules in those polymers are arranged in semicrystalline arrays featuring piezoelectriclike properties.

When exposed to an electric field, all insulating plastics, such as polyurethane, contract in the direction of the field lines and expand perpendicularly to them. This phenomenon, which differs from electrostriction, is called Maxwell stress. "It had been known for a long time but was regarded gen-



SRI INTERNATIONAL

INSECTLIKE ROBOT (named Flex) walks on legs powered by extension roll-type artificial muscles.

erally as a nuisance effect," Pelrine says.

He recognized that polymers softer than polyurethane would squash more under electrostatic attraction and thus would provide greater mechanical strains. Working with soft silicones, the SRI scientists soon demonstrated quite acceptable strains of 10 to 15 percent. With further research those numbers rose to 20 to 30 percent. To distinguish the new actuator materials, silicones and other softer plastics were christened dielectric elastomers (they are also called electric-field-actuated polymers).

Having identified several promising polymer materials, the group focused for much of the remainder of the 1990s on the nuts and bolts of building devices for specific applications. Much of the SRI team's new external funding support and research direction came at the time from the Defense Advanced Research Projects Agency (DARPA) and the Office of Naval Research, whose directors were primarily interested in using the technology for military purposes, including small reconnaissance robots and lightweight power generators.

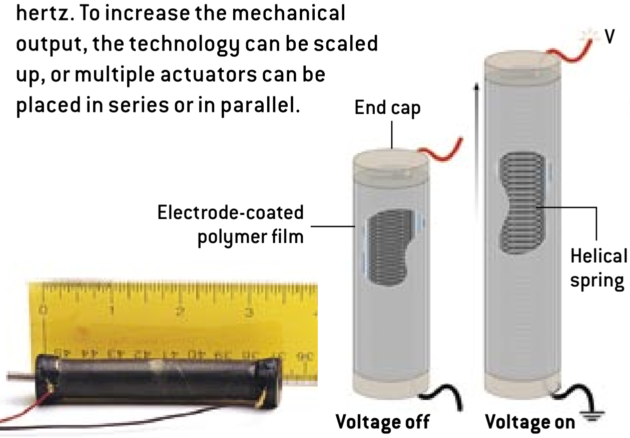
As the elastomers began to exhibit much larger strains, the engineers realized that the electrodes would have to become expandable as well. Ordinary metal electrodes cannot stretch without breaking. "Previously, people didn't have to worry about this issue, because they were working with materials that provided strains of 1 percent or so," Pelrine notes. Eventually the team developed compliant electrodes based on carbon particles in an elastomeric matrix. "Because the electrodes expand along with the plastic," he points out, "they can maintain the electric field between them across the entire active region." SRI International patented this concept, one of the keys to subsequent artificial-muscle technology.

Eager to demonstrate, Pelrine holds out what looks like a six-inch-square picture frame with plastic sandwich wrap stretched tautly across it. "See, this polymer material is very stretchy," he says, pushing a finger into the transparent film. "It's actually a double-sided adhesive tape that's sold at low cost in large rolls." On both sides of the middle of the sheet are the black, nickel-size compliant electrodes, trailing wires.

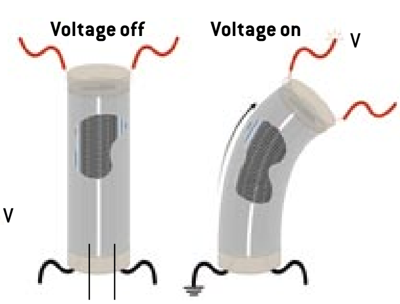
Pelrine turns a control knob on the electric power supply. Instantly, the dark circle of the paired electrodes grows to the diameter of a quarter. When he returns the knob to its original position, the disk shrinks back immediately. He flashes a grin and repeats the sequence a few times, explaining: "Fundamentally, our devices are capacitors—two charged parallel plates sandwiching a dielectric material. When the power is on, plus and minus charges accumulate on the opposite electrodes. They attract each other and squeeze down on the polymer

## SPRING ROLLS, SNAKES AND ROBOT ARMS

Polymers that expand in response to electricity make possible roll actuators that extend—or bend—on command. Engineers first roll up two layers of dielectric elastomer sheet (laminated on both sides with flexible electrodes) into compact cylinders. Often the film materials are wrapped around a compressed helical spring that holds a high circumferential prestrain on the films, thereby enhancing device performance. So-called spring rolls can serve in many applications, such as actuators for robotic and prosthetic mechanisms, valves and pumps, and wherever simple linear motion is required. To date, roll actuators have produced up to 30 newtons of force (about 6.6 pounds), linear displacements (strokes) up to about two centimeters and cyclic speeds of more than 50 hertz. To increase the mechanical output, the technology can be scaled up, or multiple actuators can be placed in series or in parallel.




Relatively simple modifications can yield devices that bend on command. Researchers spray specially patterned electrodes onto the dielectric elastomer film in such a fashion that the roll incorporates two independently energized actuators on either side (split lengthwise). If only the left half receives voltage, the right one inhibits the resulting motion and causes the device to bend toward the right (*below*). If only the right half is activated, the roll bends left. If both halves are energized, the roll extends.




More complicated arrays of independent electrodes can create more complex motion. Applications for bending rolls include snakelike robots and manipulators, steerable catheters and endoscopes, legged robots, and pointing mechanisms for antennas.

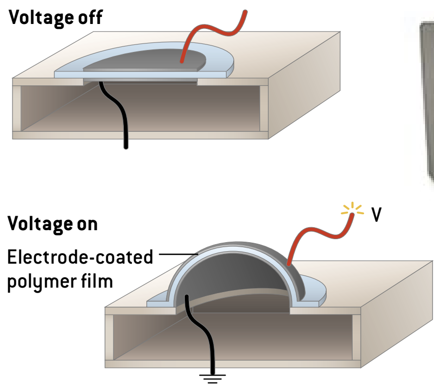
SARA CHEN (*illustrations*); JAMIE KRIPKE (*photographs*) insulator, which responds by expanding in area."

Although several promising materials had been identified, achieving acceptable performance in practical devices proved to be a challenge. A couple of breakthroughs in 1999 drew considerable interest from government and industry, however. One arose from the observation that stretching the polymers before electrically activating them somehow vastly improved their performance. "We started to notice that there seemed to be a sweet spot at which you get optimum performance," remembers engineer Roy Kornbluh, another team member. "No one was sure exactly why, but prestretching the polymers increased breakdown strengths [resistance to the passage of current between electrodes] by as much as 100 times." Actuation strains improved to a similar degree. Although the reason is still unclear, former SRI chemist Qibing Pei believes that "prestretching orients the molecular chains along the plane of expansion and also makes it stiffer in that direction." To achieve the prestraining effect, SRI's actuator devices incorporate an external bracing structure.

The second key discovery came about primarily because the researchers "were testing every stretchy material we could find—what we call an Edisonian approach," Pelrine says with amusement. (Thomas Edison systematically tried all kinds of materials for suitability as lightbulb filaments.) "At my home, we had placed a polymeric door lock on the refrigerator to keep my toddler from getting in. As he got older, we didn't need the lock anymore, so I removed it. But since it was made of a stretchy material, I decided to test its strain properties. Surprisingly, it gave very good performance." Tracking down the material and determining its composition took no small effort, but in the end the mystery polymer "turned out to be an acrylic elastomer that could provide tremendous strains and energy output—as much as 380 percent linear strain. These two developments allowed us to start applying the dielectric elastomers to real-world actuator devices," the researcher says.

## PUMP UP THE MEMBRANE

**Diaphragm actuators are made by stretching dielectric elastomer films over an opening in a rigid frame. Typically the diaphragm is biased, or pushed up or down, by a spring, light air pressure, foam or other means. Biasing makes the diaphragm actuate in one direction (up or down) rather than simply wrinkling randomly when voltage is applied. Diaphragm actuators can displace volume, making them suitable for pumps or loudspeakers. Alternative drive technologies such as piezoelectric materials have long been in use, but dielectric elastomer diaphragms offer larger displacements. Some designs can deflect from an initial flat position to a convex shape (*below*).**





SARA CHEN (*illustrations*); JAMIE KRIPKE (*photograph*)

## Making It Real

THE SRI TEAM'S general approach is flexible, encompassing many designs and even different polymers. As Pei says, "This is a device, not a material." According to Pelrine, the team can produce the actuation effect using various polymers, including acrylics and silicones. Even natural rubber works to some extent. In the extreme temperatures of outer space, for example, artificial muscles might best be made of silicone plastics, which have been demonstrated in a vacuum at –100 degrees Celsius. Uses that require larger output forces might involve more polymer or ganging up several devices in series or in parallel.

"Because the dielectric elastomers can be purchased off the shelf and we'd use at most only a few square feet of material in each device, the actuators would be very low cost, particularly in volume production," SRI's von Guggenberg estimates.

The voltages required to activate dielectric elastomer actuators are relatively high—typically one to five kilovolts—so the devices can operate at a very low current (generally, high voltage means low current). They also use thinner, less expensive wiring and keep fairly cool. "Up to the point at which the electric field breaks down and current flows across the gap [between the electrodes], more voltage gives you greater expansion and greater force," Pelrine says.

"High voltage can be a concern," Kornbluh comments, "but it's not necessarily dangerous. After all, fluorescent lights and cathode-ray tubes are high-voltage devices, but nobody worries about them. It's more of an issue for mobile devices because batteries are usually low voltage, and thus additional electric conversion circuits would be needed." Moreover, at Pennsylvania State University, Qiming Zhang and his research group have managed to lower the activation voltages of certain electrostrictive polymers by combining them with other substances to create composites.

When asked about the durability of SRI's dielectric elastomer actuators, von Guggenberg acknowledges a need for more study but attests to a "reasonable indication" that they continue to work sufficiently long for commercial use: "For example, we ran a device for one client that produces moderate, 5 to 10 percent strains for 10 million cycles."

Another generated 50 percent area strains for a million cycles.

Although artificial-muscle technology can weigh significantly less than comparable electric motors—the polymers themselves have the density of water—efforts are ongoing at SRI to cut their mass by reducing the need for the external structure that prestrains the polymers. Pei, for instance, is experimenting with chemical processing to eliminate the need for the relatively heavy frame.

## Building Products

HAVING DEVELOPED a basic mechanism, the SRI team soon began work on a flood of application concepts:

**Linear actuators.** To make what they call spring rolls, the engineers wrap several layers of prestrained laminated dielectric elastomer sheet around a helical spring. The tension spring supports the circumferential prestrain, whereas the lengthwise prestrain of the film holds the spring compressed [*see box on page 68*]. Voltage makes the film squeeze in thickness and relax lengthwise so that the device extends. The spring rolls can therefore generate high force and stroke in a compact package. Kornbluh reports that automakers are interested in these mechanisms as replacements for the many small electric motors found in cars, such as in motorized seat-position controls and in the valve controls of high-efficiency camless engines.

**Bending rolls.** Taking the same basic spring roll, engineers can connect electrodes to create two or more distinct, individually addressed sections around the circumference. Electrically activating that section makes its side of the roll extend, so the entire roll bends away from that side [*see box on page 68*]. Mechanisms based on this design could engage in complicated motions that would be difficult to accomplish using conventional motors, gears and linkages. Possible uses would be in steerable medical catheters and in so-called snake robots.

**Push-pull actuators.** Pairs of dielectric elastomer films or of spring rolls can be arranged in a "push-pull" configuration so that they work against each other and thus respond in a more linear ("one input yields one output") fashion. Shuttling voltage from one device to the other can shift the position of the whole assembly back and forth; activating both devices makes the assembly rigid at a neutral point. In this way, the actuators act like the opposing bicep and tricep muscles that control movements of the human arm.

**Loudspeakers.** Stretch a dielectric elastomer film over a frame that has an aperture in it. Expanding and contracting rapidly according to the applied voltage signal, the diaphragm will then emit sound. This configuration can yield a lightweight, inexpensive flat-panel speaker whose vibrating medium is both the driver and sound-generating panel. Current designs offer good performance in the mid- and high-frequency ranges. The speaker configuration has not yet been optimized as a woofer, although no obstacle prevents it from operating well at low frequencies [*see box on preceding page*].

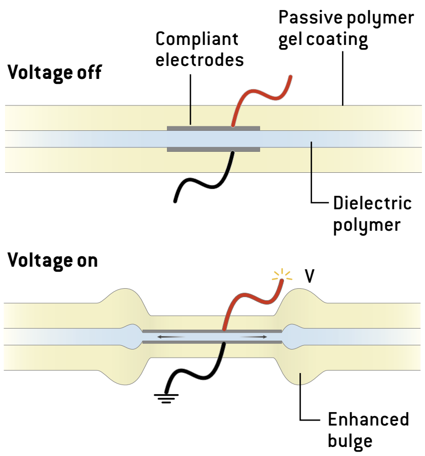
**Pumps.** The design of a dielectric elastomer diaphragm pump is analogous to that of a low-frequency loudspeaker to which engineers have added a fluid chamber and two one-way check valves to control the flow of liquid. Artificial muscles are well suited to powering microfluidic pumps, for example, on the lab-on-a-chip devices prized by medicine and industry.

**Sensors.** Because of their nature, all SRI's dielectric elastomer devices exhibit a change in capacitance when they are bent or stretched. Thus, it is possible to make a sensor that is compliant and operates at low voltage. According to Kornbluh, the team came close to getting an automaker to adopt the technology as a sensor for measuring the tension of a seat belt. Such sensors could similarly be incorporated in fabrics and other materi-

### CONTROLLABLE SURFACE TEXTURES

Changing the texture of a surface can be desirable in a variety of applications, such as "active" military camouflage materials that can alter their reflectance. Surface texturing can also help control air or water flow over the surfaces of airplanes or ships. Touch-based, or haptic, displays could be based on changes in texture.

Most dielectric elastomer actuators take advantage of large-scale deformations in the plane of the film. Alterations in the thickness of the film, on the other hand, are barely perceptible. By coating the thin-film sheets and patterned compliant electrodes with a much thicker and softer layer of polymeric gel, however, thickness changes can be greatly amplified so that they are readily apparent. As the film grows in the plane, the gel spreads out along with the expanding film and bunches up at the points at which the film compresses.



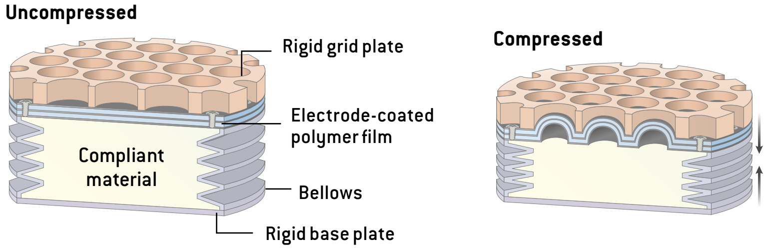

SARA CHEN (*illustrations*); JAMIE KRIPKE (*photograph*)

## ELECTRIC BOUNCE IN EVERY STEP

Dielectric elastomers can produce electric power. In generator mode, a voltage is applied across the dielectric elastomer, which is deformed by external force. As the shape of the elastomer changes, the effective capacitance of the device also alters and, with the appropriate electronics, electrical energy is generated. The energy density of these materials when used as a generator is high, which means that they can be made lighter than other technologies.

Dielectric elastomers are well suited to applications in which electrical power comes from relatively large motions, such as those produced by wind energy, waves and human activity. Capturing the compression energy of a shoe heel when it strikes the ground during walking or running is a good way to generate portable electrical power. This energy is free in the sense that it does not place an additional burden on the wearer. The heel-strike generator effectively couples the compression of the heel to the deformation of an array of multilayer diaphragms.

SRI engineers expect that, with further development, a device will be able to generate about a watt during normal walking. A unit in each shoe should provide enough electricity to power a cellular phone, for example. Such a device is being developed for the U.S. military to supply power to soldiers in the field, but the technology has civilian uses as well.




als as fibers, strips or coatings, he says.

**Surface texturing and smart surfaces.** If the polymers are imprinted with patterns of electrodes, arrays of dots or shapes can be raised on a surface on demand. This technology might find use as an active camouflage fabric that can change its reflectance as desired or as a mechanism for making "riblets" that improve the aerodynamic drag characteristics of airplane wings [*see box on opposite page*].

**Power generators.** Again, because these materials act as soft capacitors, variable-capacitance power generators and energy harvesters can be built from them. DARPA and the U.S. Army funded development of a heel-strike generator, a portable energy source that soldiers and others in the field could use to power electronic devices in place of batteries. An average-size person taking a step each second can produce about a watt of power using a device now under development [*see box above*]. Von Guggenberg says this concept has caught the interest of footwear companies. The devices could similarly be attached to backpack straps or car-suspension components. In principle, this approach could also be applied to wave generators or wind-power devices.

SRI researchers have tested a more radical concept—"polymer engines." Propane fuel was burned inside a chamber, and the pressure from the resulting combustion products distorted a dielectric elastomer diaphragm, generating electricity. Such designs might eventually lead to efficient, extremely small generators in the centimeter-or-less size range.

But truly marketable products are still to come. "At this point we're building turnkey devices that we can place in the hands of engineers so they can play with them and get comfortable with the technology," von Guggenberg notes. "We hope it's just a matter of time before every engineer will consider this technology as they design new products." SA

*Steven Ashley is a staff writer and editor.*

## MORE TO EXPLORE


**High-Speed Electrically Actuated Elastomers with Over 100% Strain.** Ron Pelrine, Roy Kornbluh, Qibing Pei and Jose Joseph in *Science,* Vol. 287, pages 836–839; February 4, 2000.

**Electroactive Polymer (EAP) Actuators as Artificial Muscles: Reality, Potential, and Challenges.** Edited by Yoseph Bar-Cohen. SPIE Press Monograph, Vol. PM98; 2001.

**Polymer Actuators.** Peter Sommer-Larsen and Roy Kornbluh in *Proceedings of Actuator 2002.* 8th International Conference on New Actuators, Bremen, Germany, June 2002.

**An All-Organic Composite Actuator Material with High Dielectric Constant.** Q. M. Zhang et al. in *Nature,* Vol. 419, pages 284–287; September 19, 2002.

**Engineering a Muscle: An Approach to Artificial Muscle Based on Field-Activated Electroactive Polymers.** Roy Kornbluh et al. in *Neurotechnology for Biomimetic Robots.* Edited by Joseph Ayers, Joel Davis and Alan Rudolph. MIT Press, 2002.

**Smart Structures and Materials 2003: Electroactive Polymer Actuators and Devices (EAPAD).** Edited by Yoseph Bar-Cohen. *Proceedings of the SPIE,* Vol. 5051; 2003.

Instructions for making electroactive polymers are available online at **http://ndeaa.jpl.nasa.gov/nasa-nde/lommas/eap/EAP-recipe.htm**

The WorldWide ElectroActive Polymer Newsletter is at **http://ndeaa.jpl.nasa.gov/nasa-nde/lommas/eap/WW-EAP-Newsletter.html**


SARA CHEN (*illustrations*); JAMIE KRIPKE (*photograph*)

# Controlling Robots with the Mind

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

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

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

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

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

If we had done everything correctly, the two robot arms

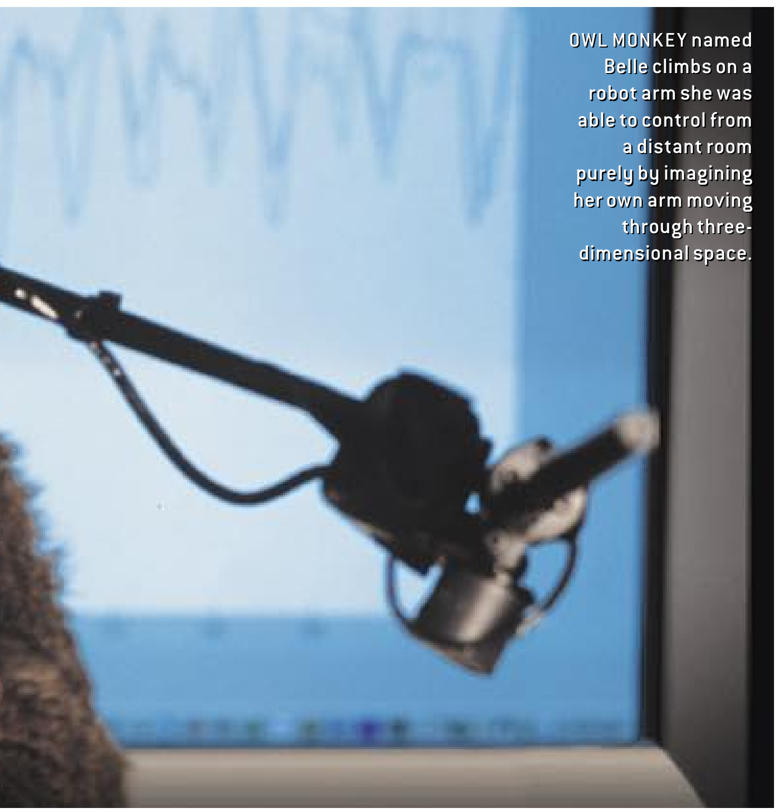
would behave as Belle's arm did, at exactly the same time. We would have to translate her neuronal activity into robot commands in just 300 milliseconds—the natural delay between the time Belle's motor cortex planned how she should move her limb and the moment it sent the instructions to her muscles.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

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

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

The big question is, of course, whether we can make a practical, reliable system. Doctors have no means by which to repair spinal cord breaks or damaged brains. In the distant future, neuroscientists may be able to regenerate injured neurons or program stem cells (those capable of differentiating into various cell types) to take their place. But in the near future, brain-machine interfaces (BMIs), or neuroprostheses, are a more viable option for restoring motor function. Success in 2002 with macaque monkeys that completed different tasks than those we asked of Belle has gotten us even closer to this goal.

JIM WALLACE *Duke University Photography*



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

## From Theory to Practice

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

There were caveats, however. Georgopoulos had recorded the activity of single neurons one at a time and from only one motor area. This approach left unproved the underlying hypothesis that some kind of coding scheme emerges from the simultaneous activity of many neurons distributed across multiple cortical areas. Scientists knew that the frontal and parietal lobes—in the forward and rear parts of the brain, respectively—interacted to plan and generate motor commands. But technological bottlenecks prevented neurophysiologists from making widespread recordings at once. Furthermore, most scientists believed that by cataloguing the properties of neurons one at a time, they could build a comprehensive map of how the brain works—as if charting the properties of individual trees could unveil the ecological structure of an entire forest!

Fortunately, not everyone agreed. When the two of us met 19 years ago at Hahnemann University, we discussed the challenge of simultaneously recording many single neurons. By 1993 technological breakthroughs we had made allowed us to record 48 neurons spread across five structures that form a rat's sensorimotor system—the brain regions that perceive and use sensory information to direct movements.

Crucial to our success back then—and since—were new electrode arrays containing Teflon-coated stainless-steel microwires that could be implanted in an animal's brain. Neurophysiologists had used standard electrodes that resemble rigid needles to record single neurons. These classic electrodes worked well but only for a few hours, because cellular compounds collected around the electrodes' tips and eventually insulated them from the current. Furthermore, as the subject's brain moved slightly during normal activity, the stiff pins damaged neurons. The microwires we devised in our lab (later produced by NB Labs in Denison, Tex.) had blunter tips, about 50 microns in diameter, and were much more flexible. Cellular substances did not seal off the ends, and the flexibility greatly reduced neuron damage. These properties enabled us to produce recordings for months on end, and having tools for reliable recording allowed us to begin developing systems for translating brain signals into commands that could control a mechanical device.

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

## A Rat's Brain Controls a Lever

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

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

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

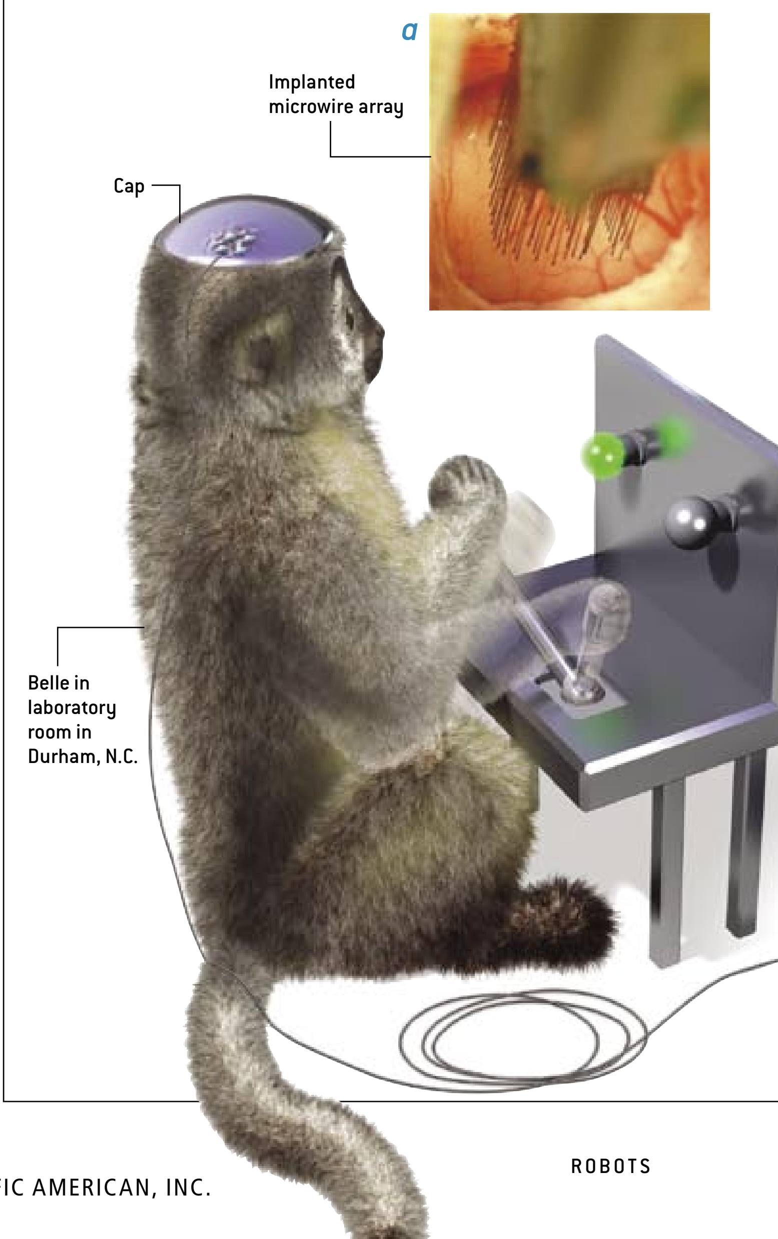
After several hours the rat realized it no longer needed to

## Overview/*Brain Interfaces*

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

## BELLE'S 600-MILE REACH

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



press the bar. If it just looked at the bar and imagined its forelimb pressing it, its neurons could still express the firing pattern that our brain-machine interface would interpret as motor commands to move the lever. Over time, four of six rats succeeded in this task. They learned that they had to "think through" the motion of pressing the bar. This is not as mystical as it might sound; right now you can imagine reaching out to grasp an object near you—without doing so. In similar fashion, a person with an injured or severed limb might learn to control a robot arm joined to a shoulder.

## A Monkey's Brain Controls a Robot Arm

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

At this time, one of us (Nicolelis) moved to Duke and established a neurophysiology laboratory there. Together we built an interface to simultaneously monitor close to 100 neu-

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

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


BRYAN CHRISTIE DESIGN

rons, distributed across the frontal and parietal lobes. We proceeded to try it with several owl monkeys. We chose owl monkeys because their motor cortical areas are located on the surface of their smooth brain, a configuration that minimizes the surgical difficulty of implanting microwire arrays. The microwire arrays allowed us to record the action potentials in each creature's brain for several months.

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

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

As our scientific work proceeded, we acquired a more advanced Harvey box from Plexon. Using it and some custom, real-time algorithms, our computer sampled and integrated the action potentials every 50 to 100 milliseconds. Software translated the output into instructions that could direct the actions of a robot arm in three-dimensional space. Only then did we try to use a BMI to control a robotic device. As we watched our multijointed robot arm accurately mimic Belle's arm movements on that inspiring afternoon in 2000, it was difficult not to ponder the implausibility of it all. Only 50 to 100 neurons randomly sampled from tens of millions were doing the needed work.

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

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

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

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

THE AUTHORS

*MIGUEL A. L. NICOLELIS* and *JOHN K. CHAPIN* have collaborated for more than 17 years. Nicolelis, a native of Brazil, received his M.D. and Ph.D. in neurophysiology from the University of São Paulo. After postdoctoral work at Hahnemann University, he joined Duke University, where he now co-directs the Center for Neuroengineering and is Anne W. Deane Professor of neuroscience, biomedical engineering, and psychological and brain sciences. Chapin received his Ph.D. in neurophysiology from the University of Rochester and has held faculty positions at the University of Texas and the MCP Hahnemann University School of Medicine (now Drexel University College of Medicine). He is currently professor of physiology and pharmacology at the State University of New York Downstate Medical Center.

## Exploiting Sensory Feedback

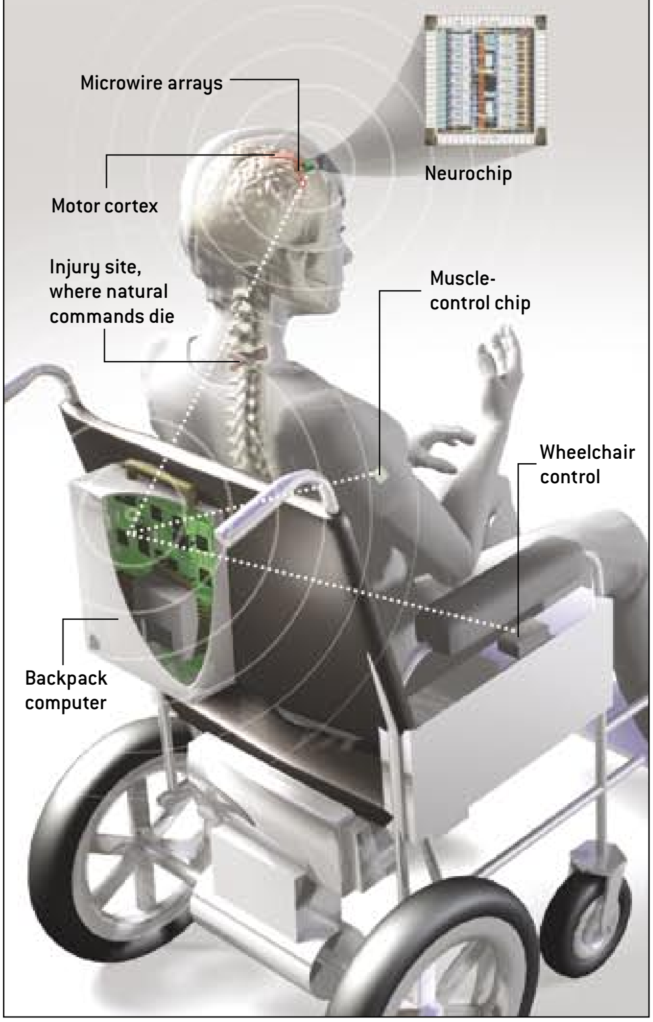
BELLE PROVED THAT A BMI can work for a primate brain. But could we adapt the interface to more complex brains? In May 2001 we began studies with three macaque monkeys at Duke. Their brains contain deep furrows and convolutions that resemble those of the human brain.

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

## A VISION OF THE FUTURE

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

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




BRYAN CHRISTIE DESIGN

had to position the cursor quickly inside the target—within 0.5 second—by rapidly manipulating the joystick.

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

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

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

We observed another encouraging result over the past seven years. In two owl monkeys, we continued to record large numbers of single neurons daily for more than five years after the implantation surgery. Even in macaque monkeys such as Aurora, we have now been able to maintain viable recordings for more than one year. In all these studies, the monkeys continued to live normal lives, performing all typical primate behaviors without exhibiting any adverse side effects. These findings support the notion that, in the near future, it may be possible to develop implants that will last for many years in human patients who suffer from severe paralysis.

During the past five years there have been many important developments in the field of brain-machine interfaces. Aurora and her "colleague" Ivy extended the frontier of the field by becoming the first primates to use a BMI for both reaching and grasping movements. Through long-term training, both Aurora and Ivy learned to use a BMI to control a robot arm without actually moving their own arms. Analysis of the data collected during these experiments by Misha Lebedev, a senior research associate in our laboratory at Duke, revealed that a significant percentage of cortical neurons—located in multiple areas of Aurora's and Ivy's brains—incorporated the dynamic properties of the robot arm as if it were part of their own bodies. Surprisingly, this neuronal incorporation of an artificial arm did not affect the ability of these neurons to continue to participate in the controlling of the monkeys' biological arms. Indeed, some cortical neurons contribute to the operation of both biological and artificial arms.

Since 2003 we have been translating the methods and technologies developed at the Duke University Center for Neuroengineering (DUCN) to the clinical arena. We have used our BMI approach to develop new methods to help neurosurgeons perform a vital surgical procedure in patients severely affected by Parkinson's disease. During these surgeries, we have been able to test whether our BMI approach would work in humans. In 2004 we reported what became the first intraoperative demonstration that an invasive BMI, based on multielectrode recordings of subcortical structures (yielding a maximum of 50 recorded neurons per patient), could reproduce simple hand movements in human subjects. Four years and almost 30 patients later, we have collected enough data for our first clinical trials in which a BMI will be used to restore upper limb mobility in severely paralyzed patients. These clinical trials, which will be carried out through a collaboration between the DUCN and the Sírio-Libanês Hospital in São Paulo, Brazil, are scheduled to start at the end of 2008.

We have created even more challenging experiments to test the limits of our BMI apparatus as well. Using technology developed in a collaboration between our two laboratories, Nathan Fitzsimmons, a graduate student in the department of neurobiology at Duke, has shown that the implants we used to record brain activity in primates can also be used to deliver brief electrical messages to the primate brain. Using this technique, called multichannel cortical microstimulation, Fitzsimmons was able to use spatiotemporal patterns of electrical stimulation, delivered through multiple electrodes implanted in the somatosensory cortex of an owl monkey, to inform the animal which of two identical boxes contained a food pellet. Because no visual cue was provided, the monkey had to learn to decode the electrical messages delivered directly into its brain to solve the behavioral task. After a few weeks of practice, two owl monkeys learned to use these abstract electrical cues to find their rewards.

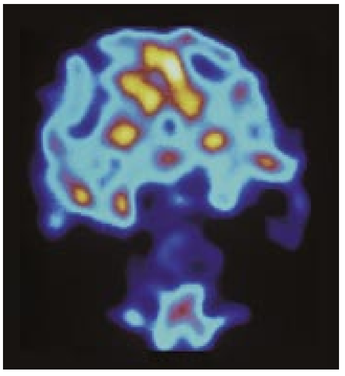
Using the same technology, Joseph O'Doherty, a graduate student in the department of biomedical engineering at Duke, demonstrated for the first time that proficiency in the operation of BMIs can be achieved by direct interactions with brain tissue rather than by relying exclusively on sensory channels such as vision or touch. In his apparatus, multichannel cortical microstimulation is used to enable an animal to choose between two identical targets on a computer screen. Once the animal decodes the electrical message delivered to its brain, it has to use its neuronal activity alone to move the computer cursor to the correct target. These experiments support the hypothesis that, in the future, a sensor-equipped prosthetic device will be able to send "feedback" information directly to the human brain, enabling the patient to generate the appropriate motor output to control the movements of the device. We call this type of device a brain-machine-brain interface, or BMBI.

During the past two years we have shown that the same principles used to reproduce upper limb movements can also be applied to the design of a BMI aimed at restoring bipedal locomotion patterns. By simultaneously recording the electrical activity of large populations of somatosensory and motor cortical neurons in monkeys that learned to walk bipedally in a hydraulic treadmill, we were able to predict in real time the kinematic parameters needed to reproduce the animals' steps under a variety of conditions—such as different treadmill speeds and forward and backward walking. To test the reliability of this approach, we carried out a fun experiment in collaboration with our colleagues Gordon Cheng and Mitsuo Kawato from the Robotics Laboratory at ATR in Kyoto, Japan. In this experiment, brain signals recorded while a monkey walked in a treadmill at Duke fed a series of linear models like those used in the experiments with Aurora. The outputs of these models were then sent in real time to ATR and used to control the bipedal locomotion of a sophisticated humanoid robot. That allowed our monkey in North America to control the steps of a robot in Japan, literally expanding that primate's brain reach to the other side of the earth. Feed-

## STOPPING SEIZURES

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

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



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

TIM BEDDON *SPL/Photo Researchers, Inc.*

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

back signals derived from sensors spread across the robot's body were then sent back to North Carolina so that the first primate ever to operate a real-time, transcontinental BMI could have plenty to brag about for years to come!

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

## Reality Check

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

Surgical implantation of electrode arrays will always be of medical concern, for instance. Investigators need to evaluate whether highly dense microwire arrays can provide viable recordings without causing tissue damage or infection in humans. Progress toward dense arrays is already under way. Duke electronics technician Gary Lehew has designed ways to increase significantly the number of microwires mounted in an array that is light and easy to implant. We can now implant multiple arrays, each of which has up to 160 microwires and measures five by eight millimeters, smaller than a pinkie fingernail. Using this approach, we can now record almost 500 neurons simultaneously.

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

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

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

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

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

## MORE TO EXPLORE


**Ensemble Recordings of Human Subcortical Neurons as a Source of Motor Control Signals for a Brain-Machine Interface.** P. G. Patil, J. M. Carmena, M.A.L. Nicolelis and D. A. Turner in *Neurosurgery,* Vol. 55, No. 1, pages 27–38; July 2004.

**Cortical Ensemble Adaptation to Represent Velocity of an Artificial Actuator Controlled by a Brain-Machine Interface.** M. A. Lebedev, J. M. Carmena, J. E. O'Doherty, M. Zacksenhouse, C. S. Henriquez, J. C. Principe and M.A.L. Nicolelis in *Journal of Neuroscience,* Vol. 25, No. 19, pages 4681–4693; May 11, 2005.

**Brain Machine Interfaces: Past, Present and Future.** M. A. Lebedev and M.A.L. Nicolelis in *Trends in Neurosciences,* Vol. 29, No. 9, pages 536–546; September 2006.

**Seeking the Neural Code.** Miguel A. L. Nicolelis and Sidarta Ribeiro in *Scientific American,* Vol. 295, No. 6, pages 70–77; December 2006.

**Primate Reaching Cued by Multichannel Spatiotemporal Cortical Microstimulation.** N. A. Fitzsimmons, W. Drake, T. L. Hanson, M. A. Lebedev and M.A.L. Nicolelis in *Journal of Neuroscience,* Vol. 27, No. 21, pages 5593–5602; May 23, 2007.

# INNOVATIONS FROM A ROBOT RALLY

By W. Wayt Gibbs

**The Grand Challenge competition spurred advances in laser sensing, computer vision and autonomous navigation—not to mention a thrilling race for the $2-million prize**

THE MOST VALUABLE AND COMPLEX component in a modern vehicle typically is also the most unreliable part of the system. Driving accidents usually have both a human cause and a human victim. To certain engineers—especially those who build robots—that is a problem with an obvious solution: replace the easily distracted, readily fatigued driver with an ever attentive, never tiring machine.

The U.S. military, which has been losing soldiers to roadside bombs in Iraq for several years, is particularly keen on this idea. But by 2002 more than a decade of military-funded research on autonomous ground vehicles had produced only a few slow and clumsy prototypes.

So that year the Pentagon authorized its Defense Advanced Research Projects Agency (DARPA) to take an unconventional approach: a public competition with a $1-million prize.


MACHINE ON A MISSION: Sandstorm swivels its laser-scanning "eye" (*inside silver dome*) to peer around a tight turn as it negotiates Beer Bottle Pass in the 2005 Grand Challenge race, followed by a DARPA chase vehicle. The autonomous Humvee drove the 132-mile course at an average speed of 18.6 miles an hour but was bested by a slightly faster robot.

## An Endurance Race for Robots

Enthusiasm ran high among the 550-odd engineers from seven nations and 42 U.S. states who gathered in Pasadena, Calif., in August 2004 to hear DARPA officials lay down the rules for the 2005 Grand Challenge race. Many had already set aside day jobs and invested their own savings to begin work on a self-navigating ground vehicle, in hopes of earning a shot at the $2-million prize in October 2005. Few seemed discouraged by the results of the first Grand Challenge, held on March 13, 2004, when only 13 teams were able to field machines for the 142-mile course and none cleared the first mountain crossing.

Sandstorm, constructed by the Red Team at Carnegie Mellon University, had traveled fastest and farthest in the 2004 event, driving at up to 36 miles an hour before straying off the edge of a narrow hairpin turn 7.4 miles into the route. But even as it fell far short of the goal, Sandstorm's performance set new records in off-road robotics and ignited the imagination of many of the roboticists, students and backyard mechanics here.

Ron Kurjanowicz, the DARPA program manager for the 2005 Grand Challenge, spelled out the rules. Any kind of traction-propelled vehicle could enter, but officials would disqualify any robot that interfered with another, damaged the environment or communicated with humans in any way during the race. The course, delineated by a computer-readable list of GPS waypoints, would be held secret until 4 A.M. on race day. "This year you should be prepared to drive 175 miles in 10 hours or less," Kurjanowicz said. The robots will have to negotiate many obstacles, DARPA director Anthony J. Tether warned. "There are gullies, washouts, stopped vehicles, underpasses, utility towers. And on the morning of the race we're going to place several tank traps on the road," he said, showing a photograph of a scary-looking obstruction built from three crisscrossed iron girders.

"Our job is to look for crazy people with crazy ideas," Tether said, only half in jest, "and then to bring those ideas as quickly as possible from the 'far side' of technology to the near side. Looking at the crowd here today, I'd say we've done that."


MENAGERIE OF ROBOTS in the 2005 Grand Challenge included several that, like Team Cornell's Spider (*top*), were based on military vehicles. Many entrants adapted pickup trucks or SUVs in order to focus their efforts on inventing new software and sensors, such as Team DAD's 64-laser terrain scanner (*middle*). And some vehicles were designed from scratch, including Team Jefferson's Tommy (*bottom*).

DARPA (*preceding pages*); VAUGHN YOUTZ *Zuma/Corbis* (*this page, top*); DIGITAL AUTO DRIVE (*middle*); DARPA (*bottom*)

The next February DARPA director Anthony J. Tether announced that the Grand Challenge—the first long-distance race for driverless vehicles—would be held in the Mojave Desert in March 2004. When no robot completed that course, DARPA doubled the prize and scheduled a second running, through a different part of the desert, for October 2005.

The point of the Grand Challenge was not to produce a robot that the military could move directly to mass production, Tether says. The aim was to energize the engineering community to tackle the many problems that must be solved before vehicles can pilot themselves safely at high speed over unfamiliar terrain. "Our job is to take the technical excuse off the table, so people can no longer say it can't be done," Tether explained at the qualifying event held 10 days before the October 8 race.

Clearly, it can be done—and done in more than one way. This time five autonomous vehicles crossed the finish line, four of them navigating the 132-mile course in well under the 10 hours required to be eligible for the cash prize.

More important than the race itself are the innovations that have been developed by Grand Challenge teams, including some whose robots failed to finish or even to qualify for the race. These inventions provide building blocks for a qualitatively new class of ground vehicles that can carry goods, plow fields, dig mines, haul dirt, explore distant worlds—and, yes, fight battles—with little or no human intervention.

"The potential here is enormous," insists Sebastian Thrun, director of Stanford University's Artificial Intelligence Laboratory and also head of its robot

## Overview/*The Grand Challenge 2005*

- Five out of 23 competing robots successfully navigated a 132-mile course through the Mojave Desert in October 2005 as part of the DARPA Grand Challenge race. To qualify for the $2-million prize, the driverless vehicles had to finish in less than 10 hours. Four turned in elapsed times under 7.5 hours.
- The race inspired innovations in location tracking, road and obstacle perception, and high-speed path planning. A successor competition in 2007, the Urban Challenge, showcased robots capable of driving themselves safely through light traffic.
- These technologies may appear in future military, agricultural, industrial and even consumer vehicles. Some are already being commercialized.

racing team. “Autonomous vehicles will be as important as the Internet.”

## From Here to There

IF ROBOTICS IS EVER to fulfill Thrun’s bold prediction, it will have to leap technical hurdles somewhat taller than those posed by DARPA’s competition. The Grand Challenge did define many of the right problems, however. To succeed in such a race, vehicles first have to plot a fast and feasible route for the long journey ahead. Next, the robots need to track their location precisely and find the road (if there is one), as well as any obstacles in their way. Finally, the machines must plan and maneuver over a path that avoids obstructions yet stay on the trail, especially at high speed and on slippery terrain.

Two hours before the event began, DARPA officials unveiled the course by handing out a computer file listing 2,935 GPS waypoints—a virtual trail of bread crumbs, one placed every 237 feet on average, for the robots to follow—plus speed limits and corridor widths. Many teams simply copied this file to their robots unchanged. But some used custom-built software to try to rapidly tailor a route within the allowed corridor that could win the race.

The Red Team, based at Carnegie Mellon University, raised this mission-planning task to a military level of sophistication. In a mobile office set up near the starting chutes, 13 route editors, three speed setters, three managers, a statistician and a strategist waited for the DARPA CD. Within minutes of its arrival, a “preplanning” system that the team had built with help from Science Applications International Corporation, a major defense contractor, began overlaying the race area with imagery drawn from a 1.8-terabyte database containing three-foot-resolution satellite and aerial photographs, digital-elevation models and laser-scanned road profiles gathered during nearly 3,000 miles of reconnaissance driving in the Mojave.

The system automatically created initial routes for Sandstorm and H1ghlander, the team’s two racers, by converting every vertex to a curve, calculating a safe speed around each curve, and knocking the highest allowable speeds down to limits derived from months of desert trials at the Nevada Automotive Testing Center. The software then divided the course and the initial route into segments, and the manager assigned one segment to each race editor.

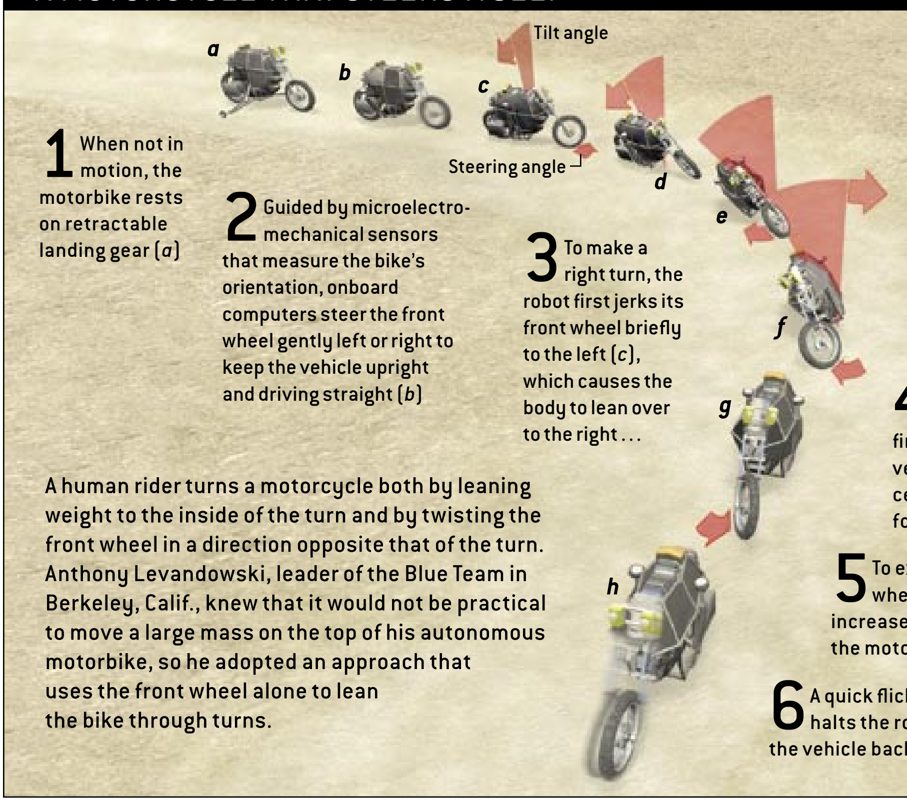
Flipping among imagery, topographic maps and reconnaissance scans, the editors tweaked the route to take tight turns the way a race driver would and to shy away from cliff edges. They marked “slow” any sections near gates, washouts and underpasses; segments on paved roads and dry lake beds were assigned “warp speed.”

The managers repeatedly reassigned segments so that at least four pairs of eyes reviewed each part of the route. Meanwhile, in a back room, team leaders pored over histograms of projected speeds and estimates of elapsed time. Team leader William “Red” Whittaker ordered completion times of 6.3 hours for H1ghlander and 7.0 hours for Sandstorm, and the system adjusted the commanded speeds to make it so.

## Hitting the Road

ROADS CHANGE—desert roads more than most—so no map is ever entirely up-

### A MOTORCYCLE THAT STEERS ITSELF




A human rider turns a motorcycle both by leaning weight to the inside of the turn and by twisting the front wheel in a direction opposite that of the turn. Anthony Levandowski, leader of the Blue Team in Berkeley, Calif., knew that it would not be practical to move a large mass on the top of his autonomous motorbike, so he adopted an approach that uses the front wheel alone to lean the bike through turns.

GHOSTRIDER ran unassisted for 20 miles at a time in desert testing.

JOHN KOCON (*illustration*); ANTHONY LEVANDOWSKI (*photograph*)

## The Race to the Starting Line

Forty-three robots rolled onto the infield of the California Speedway in Fontana, Calif., on September 28, 2005, for the Grand Challenge semifinals. Over the next eight days, each robot would get at least four chances to run a speed trial through the roughly two-mile course, which officials had cluttered with fence gates, parked cars, stacked tires and a tunnel that blocked GPS reception. Chris Urmson of the Red Team and Sebastian Thrun of the Stanford Racing Team surveyed the competition from the top of the grandstands. As they gazed at the gamut of robots ranging from a 275-pound minibike to a 15-ton military truck, Urmson broke into a grin. "This may be the coolest thing I have ever seen," he said.

The vehicles on display had been sifted from a much larger crop. DARPA accepted applications from 195 groups, including three high school teams, 35 university squads and all 15 finalists from the 2004 Grand Challenge. Many high-powered universities that sat out the first race—including Stanford, Cornell, Princeton, the University of California, Los Angeles, and the Massachusetts Institute of Technology—had entered in the second.

Only 118 teams made the first cut, based on a technical summary and a video of the vehicle in action. In May, DARPA officials visited each team for an on-site inspection and demonstration on a 220-yard (200-meter) zigzag course. The inspectors timed each robot and placed garbage cans in its path to test its obstacle-dodging abilities over three runs.

The more advanced teams sent their robots on a longer fourth run to show off the machines' driving skills. In Cedar Rapids, Iowa, Team Terramax's Oshkosh truck was able to back up and realign its eight-foot-wide body to squeeze around traffic cones with just inches to spare. At an old steel mill site in Pittsburgh, the Red Team's H1ghlander Hummer sped at 25 miles an hour over rubble-strewn road and shot through a railroad underpass.

From the first day of the qualifiers, it was clear that the technology had taken giant strides in the past 18 months. Eleven of the 43 contestants completed the obstacle course on their first try, and 25 robots had done so by the end of the trial—some hitting speeds over 40 miles an hour. Two of the finishers had crashed badly on some of their runs and were eliminated. DARPA sent the remaining 23 on to Primm, Nev., to take their shot at the $2-million prize.

CUT: M.I.T.'s Manticore (*top*) failed its on-site demonstration. IRV, built by Indy Robot Racing, flamed out on the straw bales in Fontana (*middle*). Team ENSCO was a contender for the prize until its Dexter robot (*bottom*) struck a boulder in the road at mile 81 of the race.

GRANT JORDAN (*top*); STEVEN WALLACE *Advanced Network Management Lab, Indiana University* (*middle*); SARAH WILKIN (*bottom*)

to-date. And even the perfect route is of no value unless the robot always knows where it is and where it needs to go next. Every vehicle in the Grand Challenge was equipped with differential GPS receivers. They are generally accurate to better than three feet, but overpasses and canyons block the GPS signal, and it sometimes shifts unpredictably.

Most teams thus added other tracking systems to their robots, typically inertial navigation systems that contain microelectromechanical accelerometers or fiber-optic gyroscopes. But two of the competitors created technologies that promise to be more accurate or less expensive, or both.

A team of high school students from Palos Verdes, Calif., found inspiration in the optical mouse used with desktop computers. They installed a bright lamp in their Doom Buggy robot and directed the white light onto the ground through optical tubing. A camera aimed at the bright spot picks up motion in any horizontal direction, acting as a two-dimensional odometer accurate to one millimeter. "We call it the GroundMouse," says team member Ashton Larson.

The Intelligent Vehicle Safety Technologies (IVST) team, staffed by professional engineers from Ford, Honeywell, Delphi and Perceptek, used a similar technique on its autonomous pickup truck. A radar aimed at the ground senses Doppler shifts in the frequency of the reflected beam, from which the robot then calculates relative motion with high precision. Whenever the vehicle loses the GPS fix on its position, it can fall back on dead-reckoning navigation from its radar odometer.

In the desert, even human drivers sometimes have difficulty picking out a dirt trail. It takes very clever software indeed to discriminate terrain that is probably road from terrain that is probably not. Such software, Tether says, "is a big part of what I call the 'secret sauce' that makes this technology work."

The experience of the Grand Challenge suggests that for robots, laser scanners provide the best view for this task. By rapidly sweeping an infrared laser beam across a swath of the world in front of the machine, a scanner creates a three-dimensional "point cloud" of the environment. A single laser beam cannot cover both distant objects and nearby road with sufficient fidelity, however, so a robot typically uses several in concert.

More lasers are not necessarily better. IRV, the Indy Robot Racing Team's autonomous Jeep, sported 11. But when the vehicle's sensors were knocked out of alignment, it ran over hay bales, caught fire and was eliminated during the qualification round. Without accurate calibration, laser scanners place obstacles in the wrong spot on the robot's internal map, drawing the vehicle into the very objects that it is trying to avoid.

David Hall of Team DAD, a two-man operation from Morgan Hill, Calif., created a novel laser sensor that addresses the calibration problem by fixing 64 lasers inside a motorized circular platform that whirls 10 times a second [*see box below*]. A bank of fast digital signal processors, programmed in the low-level Assembly language, handles the flood of data. In prerace trials, the sensor was able to pick out obstacles the size of a person from up to 500 feet away.

The Red Team took a different but equally innovative approach with its two robots. Each carries a single long-range laser that can do the job of many, because it swivels, rolls and nods on top of an articulated arm called a gimbal. Protected by a dome and windshield that look like a giant eyeball on top of the robot, the laser can tilt up or down when the vehicle climbs or descends. As the robot approaches a turn, the gimbal swivels left or right, keeping its eye trained on the road.

Red Team engineers also mounted fiber-optic gyroscopes to each of the gimbal's three axes and linked them via a feedback system to actuators that stabilize the laser so that it holds steady even as the vehicle jumps underneath it. The team failed to integrate that stabilization capability with the robots' other systems in time to use it for the race, however.

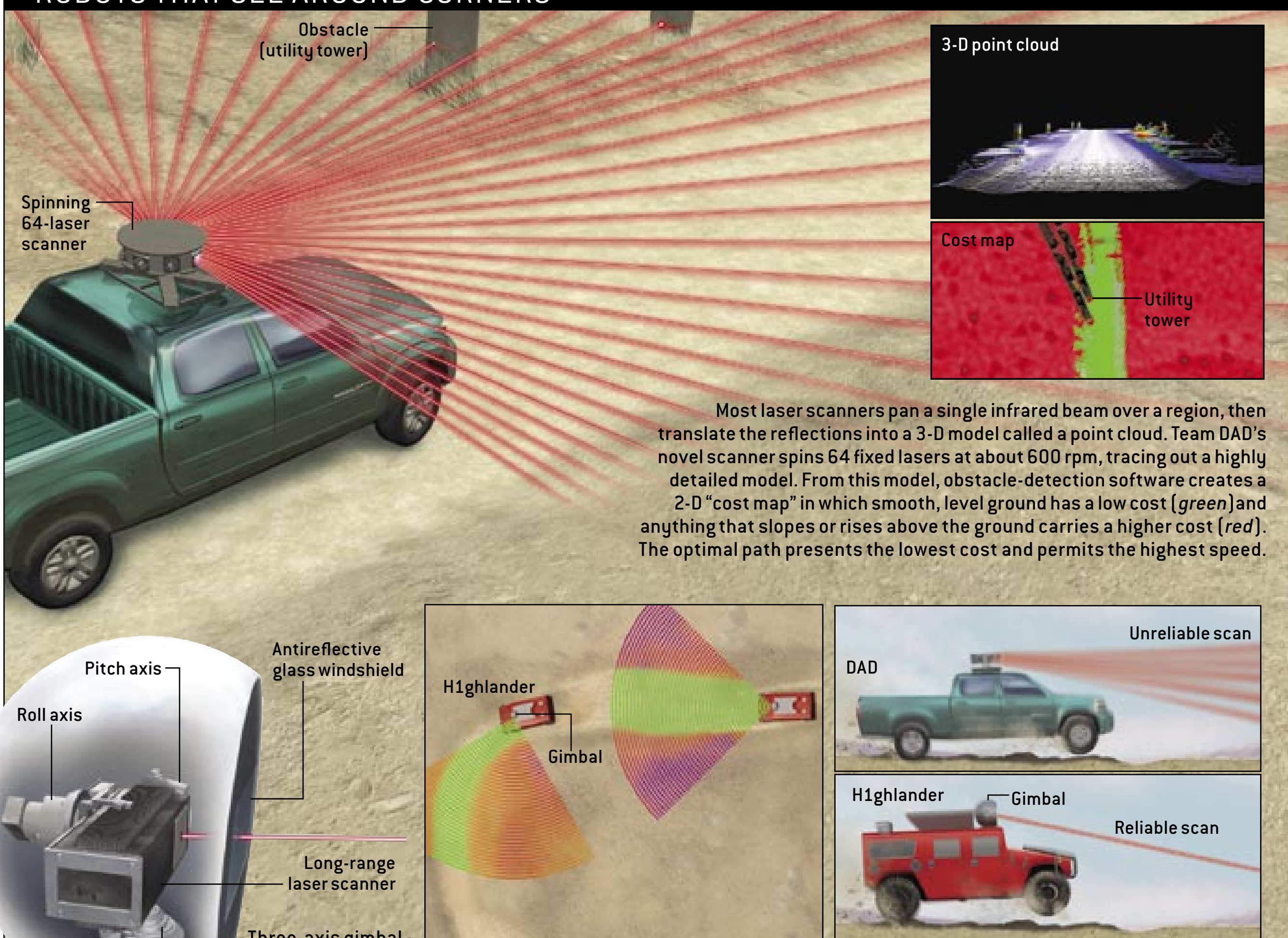
## A Path to the Future

INDISPENSABLE AS lasers seem to be, they have their drawbacks. At $25,000 to more than $100,000 each, the price of long-range laser scanners is formidable. Other kinds of sensors, such as video cameras and radars, can see farther and cost less. Yet these have their own weaknesses, and they produce torrents of data that are infamously hard to interpret.

Many teams equipped their robots with a combination of sensors. But only a few succeeded in building systems that could integrate the disparate perspectives to deduce a safe and fast path ahead—and do so many times a second.

Team Terramax's 15-ton robotic Oshkosh truck completed the course thanks

### ROBOTS THAT SEE AROUND CORNERS




Most laser scanners pan a single infrared beam over a region, then translate the reflections into a 3-D model called a point cloud. Team DAD's novel scanner spins 64 fixed lasers at about 600 rpm, tracing out a highly detailed model. From this model, obstacle-detection software creates a 2-D "cost map" in which smooth, level ground has a low cost (*green*) and anything that slopes or rises above the ground carries a higher cost (*red*). The optimal path presents the lowest cost and permits the highest speed.

Three-axis gimbal on the Red Team's H1ghlander and Sandstorm robots is able to pitch, roll and yaw to point the laser scanner in any forward direction.

Hairpin turns are typically blind turns for robots. But DAD, H1ghlander and Sandstorm can often see the other side of a tight curve before they get there. DAD has a full 180-degree field of view; the Red Team robots swivel their gimbals.

Rough roads jostle fixed lasers, creating gaps in the 3-D model. The Red Team gimbal senses such jolts with fiber-optic gyroscopes, then uses its actuators to cancel out the motion. The result is more reliable perception, especially when looking far ahead.

JOHN KOCON (*illustrations*); JOSHUA ANHALT AND TUGRUL GALATALI, RED TEAM (*point cloud*); LUCY READING-IKKANDA AND RED TEAM (*cost map*) in part to a novel "trinocular" vision system designed by Alberto Broggi's group at the University of Parma in Italy. The program selects from among three possible pairs of cameras to get an accurate stereo view of the near, medium or distant terrain. The higher its speed, the farther out the robot peers.

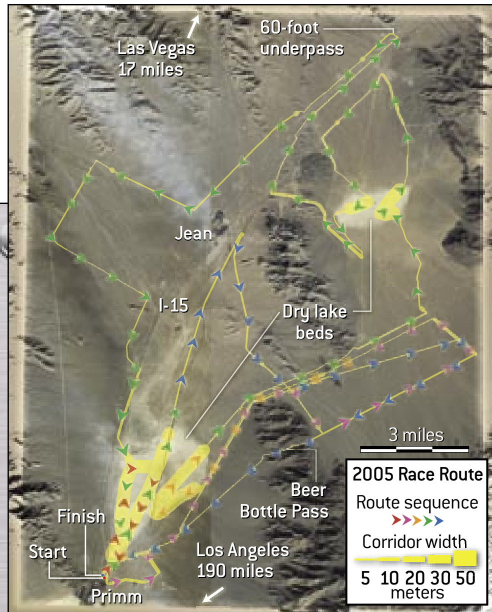
After the competition, Thrun reflected that one of the key advantages of his Stanford team's Stanley robot, which won the race and the $2 million, was its vision-based speed switch. Stanley uses a simple but powerful form of machine learning to hit the gas whenever it spots a smooth road extending into the distance [*see box on page 88*].

Some of the innovations with the greatest reach, however, appeared on robots that never reached the finish line. The IVST team, for example, devoted desert trials to discovering the optimum sensor configurations for its Desert Tortoise in a variety of "contexts"—such as washboard trail, paved highway or interstate underpass. As the robot drives, explains team leader William Klarquist, "the vehicle chooses an appropriate context that switches off some sensors, switches on others, and reassigns the confidence that it places in each one." This technique should allow a robot to move from desert to, say, farmland and still perform well by loading a new set of contexts.

In IRV, the Indy Robot Racing Team demonstrated a "plug and play" system for sensors, a feature that is probably a prerequisite for the creation of an autonomous-vehicle industry. The far-flung team of more than 100 engineers needed a way to swap sensors and software modules in and out of the robot easily as the participants tested and refined the system. So they invented a network protocol (analogous to the hypertext transfer protocol on which the Web runs) for autonomous driving.

Each sensor on IRV plugs into a dedicated computer, which boils the raw data down to a set of obstacle coordinates and sizes and then translates that into the network protocol. Every sensor computer broadcasts its obstacle list to all other sensors and to the robot's central path-planning computer. The standard makes removing a malfunctioning

## And the Winner Is ...




GRAND CHALLENGE course began and ended in Primm, Nev. Over its 131.7 miles, the route (*sequence of colored arrows*) included two long underpasses, several railroad crossings and a mountain pass.

"Last year, the night before the race, I just kept thinking, 'Don't screw up, Sandstorm,'" recalls Chris Urmson, one of the team's technical leaders. "This year it's more a feeling of anticipation—like Santa Claus is coming."

At 4 A.M. plus 90 seconds, the course comes up on screens in the Red Team's route-planning trailer. "Hmm, this is exciting," says Alexander Gutierrez as he scans the convoluted route.

Michael Montemerlo, lead programmer for the Stanford team, is looking at a similar display of the course on his laptop inside Stanley. "What the heck? There's all kinds of overlap—it keeps going in and out. There: there are the mountains, right at the end." Sebastian Thrun, the team leader, looks over his shoulder. "It's short," Thrun says. "That's sad."

H1ghlander is the first to launch into the rising dawn. If it sticks to its schedule—and in months of desert testing it always has—the vehicle will finish at 1 P.M. after a 6.3-hour run. Stanley starts five minutes later, followed by Sandstorm and the remainder of the 23 robots at five- to 10-minute intervals.


VICTORIOUS Stanley claimed the $2-million prize for Stanford with an elapsed time of 6.9 hours. Four other robots completed the course: Sandstorm in 7.1 hours, H1ghlander in 7.2, KAT-5 in 7.5, and Terramax in 12.9.

By 8:35 A.M. Team DAD's pickup with the spinning laser has passed IVST's truck and is gaining on Sandstorm. An hour later H1ghlander rolls through a 40-mile-an-hour dust storm, having widened its lead on Stanley by seven minutes. Stanley is meanwhile pulling farther ahead of Sandstorm, which the Red Team commanded to drive at a conservative 7.0-hour pace as part of a hare-and-tortoise strategy. Team ENSCO's Dexter, which started in the middle of the pack, is making great time.

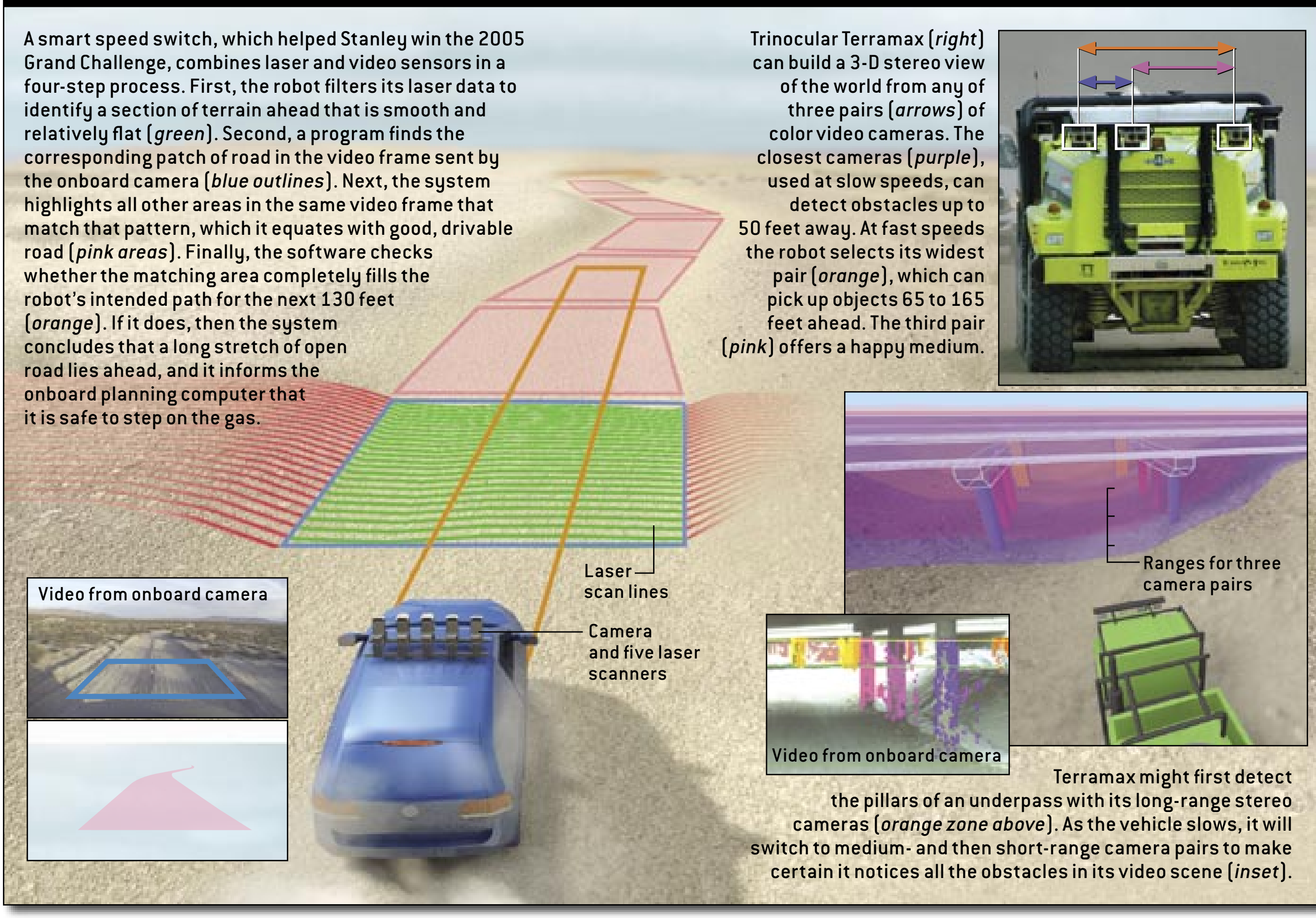
As H1ghlander crosses the railroad and hits rolling terrain, it stops midway up a hill and slips back to the bottom, then climbs and falls again. On the third attempt it crests the hill, but clearly the robot's engine is flagging. Stanley catches up, and shortly past noon whoops erupt from the large crowd of Stanford and Volkswagen spectators as Stanley takes the lead.

At 1:51 P.M. Stanley appears at the finish line, soon followed by H1ghlander and Sandstorm. Team Gray's KAT-5 arrives at sunset, as officials pause Terramax, which spends the night idling in the desert and completes its mission the next morning. After checking the robots' time logs, DARPA director Anthony J. Tether pronounces Stanley the Grand Challenge victor, by a margin of 11 minutes.

GENE BLEVINS *L.A. Daily News/Corbis* (*Stanley*); W. WAYT GIBBS (*route map*) AND IKONOS (*satellite image*)

Page Intentionally Blank

## VISION LINKED TO SPEED




A smart speed switch, which helped Stanley win the 2005 Grand Challenge, combines laser and video sensors in a four-step process. First, the robot filters its laser data to identify a section of terrain ahead that is smooth and relatively flat (*green*). Second, a program finds the corresponding patch of road in the video frame sent by the onboard camera (*blue outlines*). Next, the system highlights all other areas in the same video frame that match that pattern, which it equates with good, drivable road (*pink areas*). Finally, the software checks whether the matching area completely fills the robot's intended path for the next 130 feet (*orange*). If it does, then the system concludes that a long stretch of open road lies ahead, and it informs the onboard planning computer that it is safe to step on the gas.

Trinocular Terramax (*right*) can build a 3-D stereo view of the world from any of three pairs (*arrows*) of color video cameras. The closest cameras (*purple*), used at slow speeds, can detect obstacles up to 50 feet away. At fast speeds the robot selects its widest pair (*orange*), which can pick up objects 65 to 165 feet ahead. The third pair (*pink*) offers a happy medium.

Terramax might first detect the pillars of an underpass with its long-range stereo cameras (*orange zone above*). As the vehicle slows, it will switch to medium- and then short-range camera pairs to make certain it notices all the obstacles in its video scene (*inset*).

radar or upgrading a buggy vision algorithm as simple as a changing a tire.

Soon after the dust had settled from the Grand Challenge, DARPA announced it would hold a new robot competition, called the Urban Challenge, in November 2007. This time the course was much shorter—just 60 miles—and followed paved roads with low speed limits in a suburban setting on an abandoned U.S. Air Force base. But the difficulty level was raised by a new kind of obstacle: human-driven cars driving among the robots and through four-way intersections in the course. Many of the 11 robots that qualified for the final event were built by teams that had done well in the Grand Challenge, including Stanford, Carnegie Mellon and Oshkosh. Seven of the robotic competitors sported the spinning-laser system invented by Team DAD, now sold by Velodyne as a commercial product, and Oshkosh retained trinocular vision on its Terramax truck. None of the Urban Challengers used a stabilized gimbal, however, and Stanford's autonomous Passat used little of the software that guided Stanley to a win in 2005.

The automatons were required to obey California traffic laws, and judges deducted points for every moving violation, such as when the Massachusetts Institute of Technology's Talos (which ultimately placed fourth) collided with Cornell University's Skynet robot. Terramax was disqualified when it nearly slammed into a building. Carnegie Mellon's Boss, built from a Chevy Tahoe, turned in the fastest, cleanest performance and took the $2-million grand prize. Stanford's Junior robot won second place and $1 million, whereas Virginia Tech's Victor Tango vehicle earned $500,000 for third.

Asked whether the government would sponsor yet another robot race, DARPA's Tether said no. But the military is clearly interested in roboticizing its supply convoys, a lead that Oshkosh is actively pursuing. Whittaker and others are hoping to persuade NASA to send an autonomous vehicle to the moon. And commercial investors have been circling. So whatever else happens, these robots will keep moving. SA

---


*W. Wayt Gibbs, a contributing editor at* Scientific American, *is executive editor of Intellectual Ventures in Bellevue, Wash.*


### MORE TO EXPLORE


**High Speed Navigation of Unrehearsed Terrain: Red Team Technology for Grand Challenge 2004.** Chris Urmson et al. Carnegie Mellon University Technical Report CMU-RI-TR-04-37; June 2004.

**Adaptive Road Following Using Self-Supervised Learning and Reverse Optical Flow.** David Lieb, Andrew Lookingbill and Sebastian Thrun in *Proceedings of Robotics: Science and Systems I;* June 2005. Available online at **www.roboticsproceedings.org**

DARPA Grand Challenge Web site: **www.darpa.mil/grandchallenge**



JOHN KOCON (*illustrations*); SEBASTIAN THRUN *Stanford University (video frames)*; V. ROSS JOHNSON AND SHAHN TORONTOW (*photograph, top right*); ALBERTO BROGGI *University of Parma (inset at bottom right)*