SHADES OF
THE
SINGULARITY

RAY KURZWEIL'S
FUTURE IMPERFECT
BY JOHN RENNIE

THINKING LIKE
A HUMAN-WITH
MEMRISTORS
BY MASSIMILIANO VERSACE
& BEN CHANDLER

0 74470 64849 9

1 2>
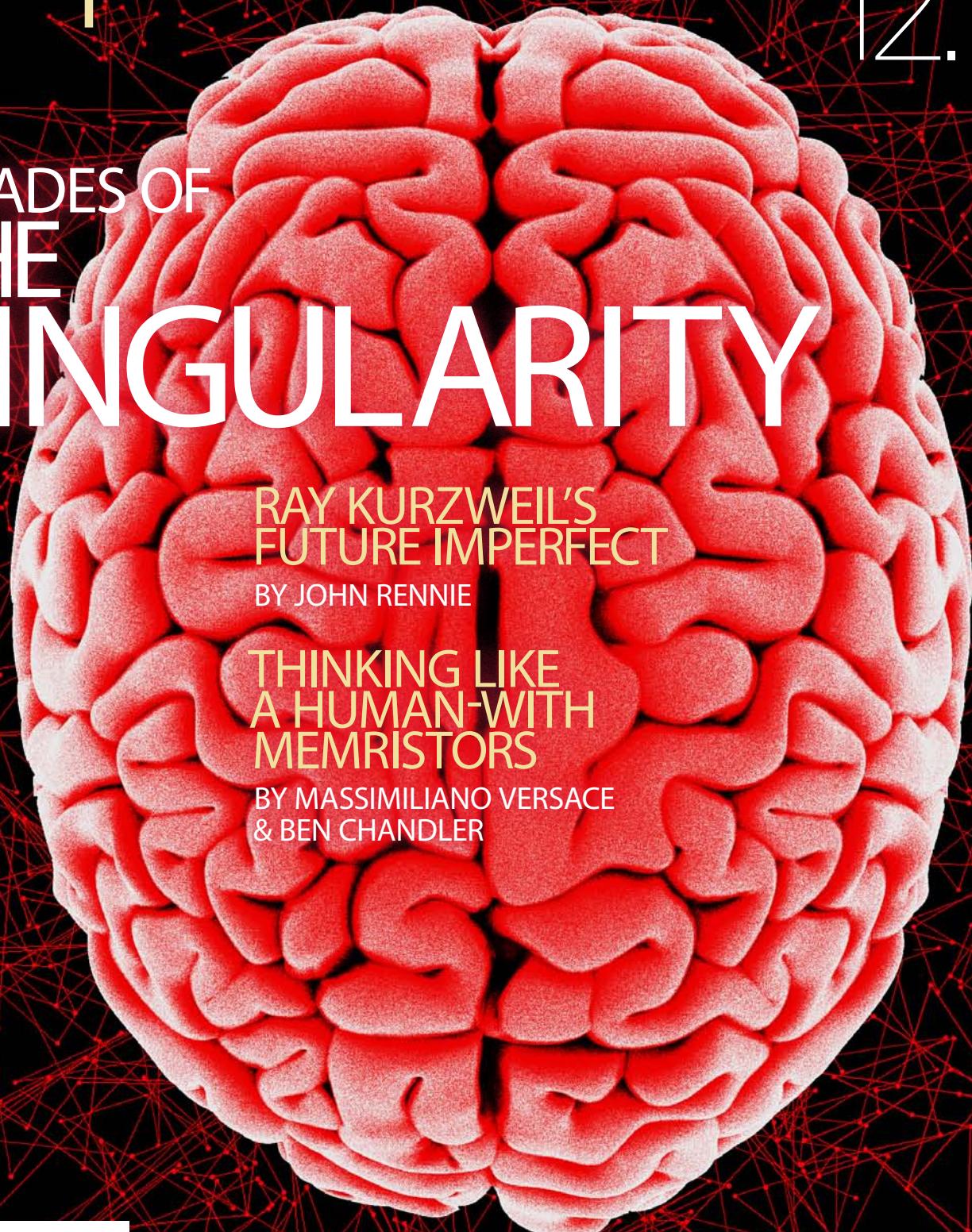
◆ IEEE

Meet MoNETA—
the **brain-inspired
chip** that will
outsmart us all

The Brain
of a New
Machine

BY MASSIMILIANO VERSACE & BEN CHANDLER
ILLUSTRATIONS BY CHAD HAGEN
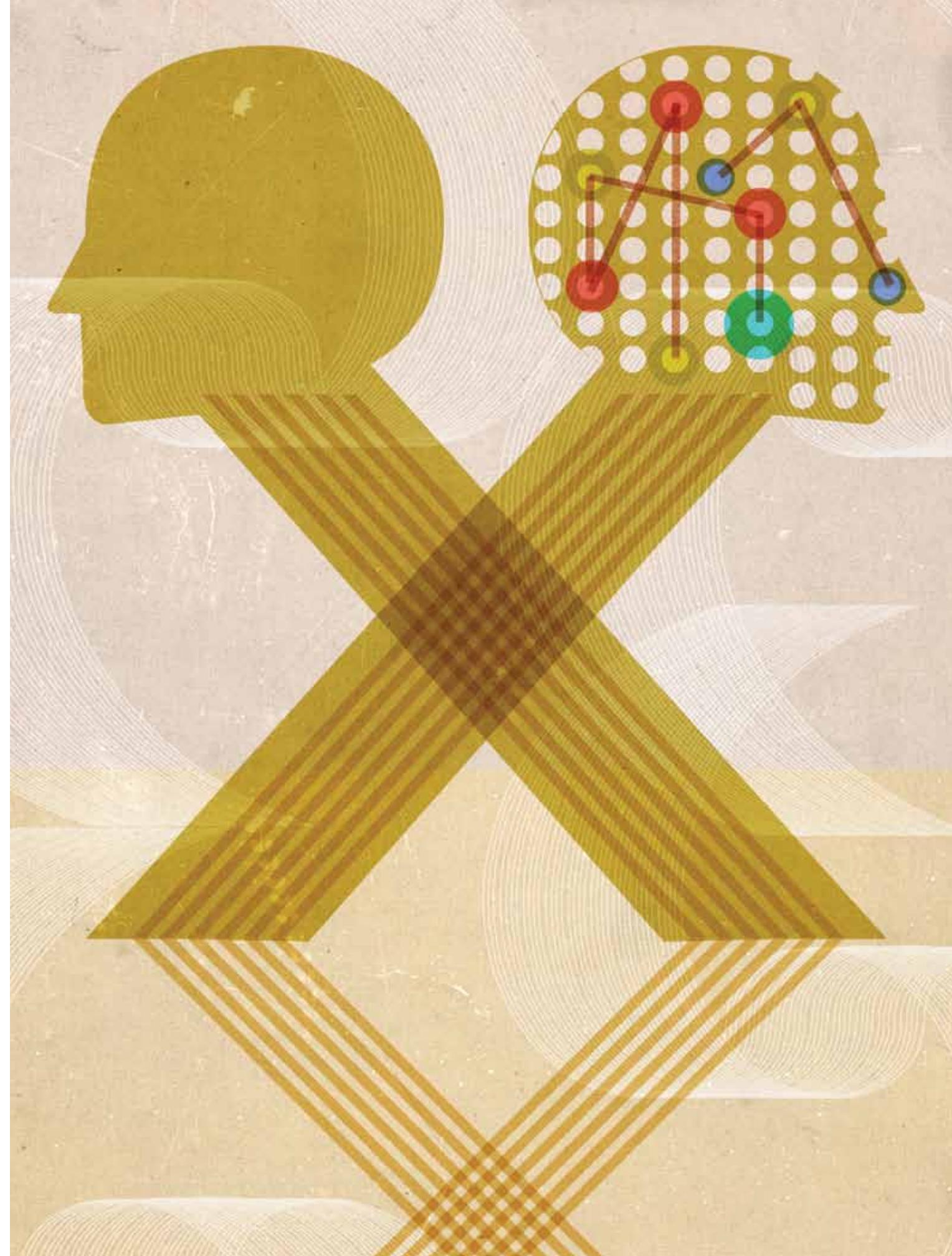
STOP US IF YOU'VE HEARD THIS ONE BEFORE: In the near future, we'll be able to build machines that learn, reason, and even emote their way to solving problems, the way people do.

If you've ever been interested in artificial intelligence, you've seen that promise broken countless times. Way back in the 1960s, the relatively breathless invention of the transistor prompted breathless predictions that machines would outsmart their human handlers within 20 years. Now, 50 years later, it seems the best we can do is automated tech support, intoned with a preternatural calm that may or may not send callers into a murderous rage.

So why should you believe us when we say we finally have the technology that will lead to a true artificial intelligence? Because of MoNETA, the brain on a chip. MoNETA (Modular Neural Exploring Traveling Agent) is the software we're designing at Boston University's department of cognitive and neural systems, which will run on a brain-inspired microprocessor under development at HP Labs in California. It will function according to the principles that distinguish us mammals most profoundly from our fast but witless machines. MoNETA (the goddess of memory—cute, huh?) will do things no computer ever has. It will perceive its surroundings, decide which information is useful, integrate that information into the emerging structure of its reality, and in some applications, formulate plans that will ensure its survival. In other words, MoNETA will be motivated by the same drives that motivate cockroaches, cats, and humans.

Researchers have suspected for decades that real artificial intelligence can't be done on traditional hardware, with its rigid adherence to Boolean logic and vast separation between memory and processing. But that knowledge was of little use until about two years ago, when HP built a new class of electronic device called a memristor. Before the memristor, it would have been impossible to create something with the form factor of a brain, the low power requirements, and the instantaneous internal communications. Turns out that those three things are key to making anything that resembles the brain and thus can be trained and coaxed to behave like a brain. In this case, form is function, or more accurately, function is hopeless without form.

Basically, memristors are small enough, cheap enough, and efficient enough to fill the bill. Perhaps most important, they have key characteristics that resemble those of synapses. That's why they will be a crucial enabler of an artificial intelligence worthy of the term.

The entity bankrolling the research that will yield this new artificial intelligence is the U.S. Defense Advanced Research Projects Agency (DARPA). When work on the brain-inspired microprocessor is complete, MoNETA's first starring role will likely be in the U.S. military, standing in for irreplaceable humans in scout vehicles searching for roadside bombs or navigating hostile terrain. But we don't expect it to spend much time confined to a niche. Within five years, powerful, brainlike systems will run on cheap and widely available hardware.

How brainlike? We're not sure. But we expect that the changes MoNETA will foment in the electronics industry over the next couple of decades will be astounding.

ARTIFICIAL INTELLIGENCE hasn't stood still over the past half century, even if we never got the humanlike assistants that some thought we'd have by now. Computers diagnose patients over the Internet. High-end cars help keep you from straying out of your lane. Gmail's Priority Inbox does a pretty decent job of prioritizing your e-mails.

But even the most helpful AI must be programmed explicitly to carry out its one specific task. What we want is a general-purpose intelligence that can be set loose on any problem; one that can adapt to a new environment without having to be retrained constantly; one that

can tease the single significant morsel out of a gluttonous banquet of information the way we humans have evolved to do over millions of years.

Think about that MoNETA-enabled military scout vehicle for a moment. It will be able to go into a mission with partially known objectives that change suddenly. It will be able to negotiate unfamiliar terrain, recognize a pattern that indicates hostile activity, make a new plan, and hightail it out of the hostile area. If the road is blocked, it will be able to make a spur-of-the-moment decision and go off-road to get home. Intuition, pattern recognition, improvisation, and the ability to negotiate ambiguity: All of these things are done really well by mammalian brains—and absolutely abysmally by today's microprocessors and software.

Consider Deep Blue, IBM's 1.4-ton supercomputer, which in 1997 faced then world chess champion Garry Kasparov. In prior years, Kasparov had defeated the computer's predecessors five times. After a taut series comprising one win apiece and three draws, Deep Blue finally trounced Kasparov in game six. Nevertheless, Deep Blue was not intelligent. To beat Kasparov, its special-purpose hardware used a brute-force strategy of simply calculating the value of 200 million possible chess moves each second. In the same amount of time, Kasparov could plan roughly two chess positions.

Over the next 10 years, computing capabilities skyrocketed: By 2007 the processing power of that 1.4-ton supercomputer had been contained within a Cell microprocessor roughly the size of a thumbnail. In the decade between them, transistor counts had jumped from 7.5 million on an Intel Pentium II to 234 million on the Cell. But that explosion of computing power did not bring artificial intelligence the slightest bit closer, as DARPA's Grand Challenge has amply demonstrated.

DARPA had launched the Grand Challenge to create autonomous vehicles that could drive themselves without human intervention. AI had been credited (again) with a major victory, when Stanley, Stanford's Volkswagen Touareg, drove itself 212 kilometers (132 miles) across California's Mojave desert to claim the US $2 million prize. One giant leap for AI!

Not really. The next phase of DARPA's challenge upped the ante, demanding AI-controlled cars whose intelligence could conquer not just the wide-open desert but busy city streets. For eight days in 2007, DARPA set research teams loose on George Air Force Base, a desolate speck in Victorville, Calif. This time, the cars had to navigate basic traffic conditions according to California law, merging, passing, parking, negotiating intersections—the stuff most American teenagers can do by age 16.

The results were sobering. Cars tricked out with state-of-the-art sensors, positioning systems, and in one case, 14 blade servers, were utterly undone by obstacles as common as a breadbox-size rock. Within a few hours, almost half the teams had been removed from the race for such infractions as running amok in a parking lot or smashing into each other while trying to share a single lane on a road.

NOW CONSIDER THE HUMBLE RAT. Its biological intelligence uses general-purpose "wetware"—the biochemical hardware and software puree that is the brain—to solve tasks like those of the Grand Challenge cars, with much better results. First, a hungry rat will explore creatively for food. It might follow familiar, memorized routes that it has learned are safe, but at the same time it must integrate signals from different senses as it encounters various objects in the environment. The rat can recognize dangerous objects such as a mousetrap and will often avoid them even though it may never have seen the object at that particular angle before. After eating, the rat can quickly disengage its current plan and switch to its next priority. All these simultaneous challenges, with all their varied complexities, are impractical for a machine, because you can't fit a computer that size into a vehicle smaller than a semi. And yet they are negotiated by a brain whose networks of millions of neurons and billions of synapses are distributed across many brain areas—a brain that weighs no more than 2 grams and can operate on the power budget of a Christmas-tree bulb.

Why is the rat brain so superior? In a word, architecture. The brain of an adult rat is composed of 21 million nerve cells called neurons (the human brain has about 100 billion). Neurons talk to each other by way of dendrites and axons. You can think of these tendrils as the in-boxes (dendrites) and out-boxes (axons) of the individual neuron, transmitting electrical impulses from one neuron to another. Most of the processing performed in the nervous system happens in the junctions between neurons. Such a junction, between one neuron's dendrite and a neighboring neuron's axon, is a space called a synapse.

Computational neuroscience has focused largely on building software that can simulate or replicate a mammal's brain in the classic von Neumann computer architecture. This architecture separates the place where data is processed from the place where it is stored, and it has been the staple of computer architectures since the 1960s [see sidebar, "The Great Brain Race"]. Researchers figured that, given enough powerful CPUs, creating programs that emulate the "software" of the brain is a logical outcome.

But that's a little like saying that given enough words, creating a novel is the logical outcome. Architecture is key here. To understand why, compare the path of a hypothetical bit of data inside a conventional microprocessor with its path inside a brain.

Recall that on a standard computer, the memory and processor are separated by a data channel, or bus, between the area where the data's stored and where it's worked on [see illustration, "Hardware vs. Wetware"]. That channel's fixed capacity means that only limited amounts of data can be "checked out" and worked on at any given instant. The processor reserves a small number of slots, called registers, for storing data during computation. After doing all the necessary computation, the processor writes the result back to memory—again, using the data bus. Usually, this routine doesn't pose much of a problem: To minimize the amount of traffic flowing on the fixed-capacity bus, most modern proces-

sors augment the registers with a cache memory that provides temporary storage very close to the point of computation. If an often-repeated computation demands multiple pieces of data, the processor will keep them in that cache, which the computational unit can then access much more quickly and more efficiently than it can the main memory.

However, that caching scheme won't work for the sort of computational challenges you'd encounter trying to simulate a brain. Even relatively simple brains have tens of millions of neurons connected by billions of synapses, so any attempt to simulate such a vast interconnection would gobble up a cache as big as the computer's main memory—which would render the machine immediately useless.

Why? The vast majority of the computing and power budget of such a brain-simulating system—computer scientists call it a neuromorphic architecture—goes to mimicking the sort of signal processing that happens inside the brain's synapses. Indeed, modeling just one individual synapse requires the following to happen in the machinery: The synapse's state—how likely it is to pass on a signal-like input from a neuron, which is the major factor in how strong the association is between any two neurons—is in a location in main memory. To change that state, the processor must package an electronic signal for transfer over the main bus. That signal must travel between 2 and 10 centimeters to reach the physical memory and then must be unpackaged to actually access the desired memory location.

Now multiply that sequence by up to 8000 synapses—as much as a single rat neuron might have. Then multiply that by the number of neurons in the brain you're emulating—billions. Congratulations! You've just modeled an entire *millisecond* of brain activity.

A biological brain is able to quickly execute this massive simultaneous information orgy—and do it in a small package—because it has evolved a number of stupendous shortcuts. Here's what happens in a brain: Neuron 1 spits out an impulse, and the resultant information is sent down the axon to the synapse of its target, Neuron 2. The synapse of Neuron 2, having stored its own state locally, evaluates the importance of the information coming from Neuron 1 by integrat-

Hardware vs. Wetware To understand the difference between the architecture of the brain and a standard computer, compare the path of a hypothetical bit of data in a brain with that in a brain simulation.

BRAIN In the mammalian brain, storage and computation happen at the same time and in the same place. Neuron 1 sends a signal down the axon to Neuron 2. The synapse of Neuron 2 evaluates the importance of the information coming from Neuron 1 by contrasting it with its own previous state and the strength of its connection to Neuron 1. Then, these two pieces of information—the information from Neuron 1 and the state of Neuron 2's synapse—flow toward the body of Neuron 2 over the dendrites. By the time that information reaches the body of Neuron 2, there is only a single value—all computation has already taken place during the information transfer.
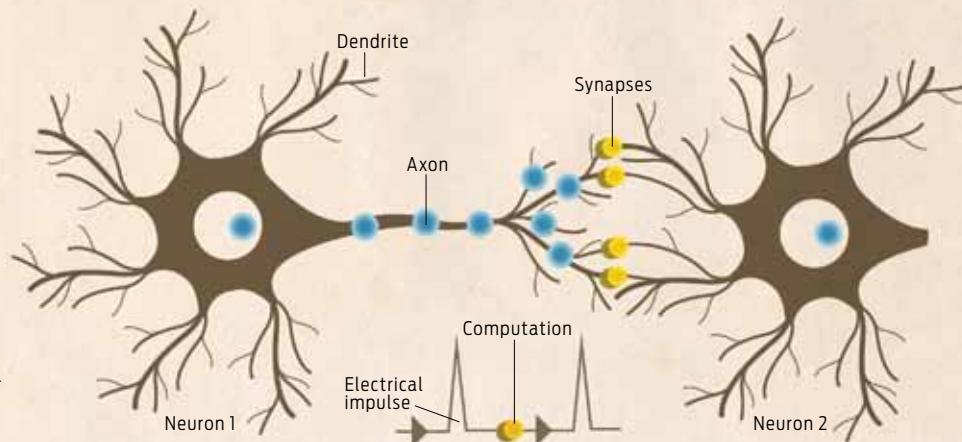
COMPUTER On a computer, the memory and processor are physically separated—a significant physical distance separates the areas where the data is stored from the areas where it is manipulated. Modeling just a single synapse requires the following to happen in the machinery: The synapse's state is in a location in main memory. To change that state, a signal must originate somewhere on the processor, travel to the edge of the processor, be packaged for transfer over the main bus, travel between 2 and 10 centimeters to reach the physical memory, and then be unpackaged to actually access the desired memory location. Multiplying that sequence by up to 8000 synapses—as many as in a single rat neuron—and then again by the brain's billions of neurons yields a single millisecond of brain activity.
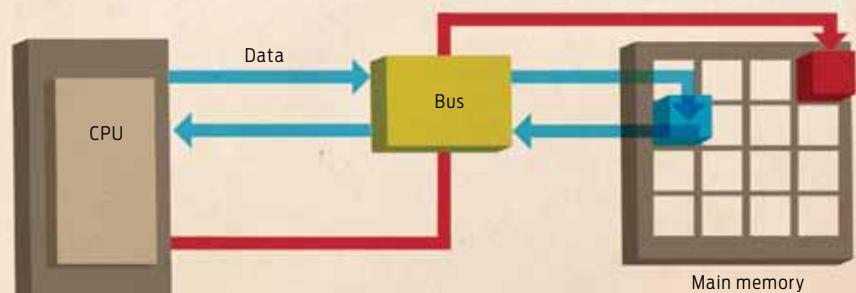
Labels on diagram: Dendrite, Synapses, Axon, Computation, Electrical impulse, Neuron 1, Neuron 2, Data, CPU, Bus, Main memory

---

ing it with its own previous state and the strength of its connection to Neuron 1. Then, these two pieces of information—the information from Neuron 1 and the state of Neuron 2's synapse—flow toward the body of Neuron 2 over the dendrites. And here is the important part: By the time that information reaches the body of Neuron 2, there is only a single value—*all processing has already taken place during the information transfer*. There is never any need for the brain to take information out of one neuron, spend time processing it, and then return it to a different set of neurons. Instead, in the mammalian brain, storage and processing happen at the same time and in the same place.

That difference is the main reason the human brain can run on the same power budget as a 20-watt lightbulb. But reproducing the brain's functionality on even the most advanced supercomputers would require a dedicated power plant. To be sure, locality isn't the only difference. The brain has some brilliantly efficient components that we just can't reproduce yet. Most crucially, brains can operate at around 100 millivolts. Complementary metal-oxide-semiconductor logic circuits, however, require a much higher voltage to function properly (close to 1 volt), and the higher operating voltage means that more power is expended in transmitting the signal over wires.

Now, replicating the structure we've described above is not totally impossible with today's silicon technology. A true artificial intelligence *could* hypothetically run on conventional hardware, but it

would be fantastically inefficient. Inefficient hardware won't stop us from running neuromorphic algorithms (such as machine vision), but we would need an entire massive cluster of high-performance graphics processing units (GPUs) to handle the parallel computations, which would also come with the power requirements of a midwestern college town.

So how do you build something that has an architecture like the brain's? Here's DARPA's gambit: Change your architecture to merge memory and computation. The memristor is the best technology out there for the task. That's because the memristor is the first memory technology with enough power efficiency and density to rival biological computation. With these devices, we are confident we can build an AI that can approximate the size and power requirements of a mammal's brain.

PARTLY TO AVOID THE FOLLY of trying to coax intelligence from fundamentally dumb hardware, DARPA launched a program called SyNAPSE (Systems of Neuromorphic Adaptive Plastic Scalable Electronics) in 2008. The timing was good. That year, HP Labs had created a functioning memristor, a device hailed as the fourth fundamental electronic

component, after the resistor, capacitor, and inductor. The concept wasn't new. In 1971, professor Leon Chua of the University of California, Berkeley, reasoned that a memristor would behave like a resistor with a conductance that changed as a function of its internal state and the voltage applied. In other words, because a memristor could remember how much current had gone through it, it could work as an essentially nonvolatile memory. And sure enough, Korean dynamic RAM giant Hynix Semiconductor made a splash recently when it chose the device as a possible foundation for its next-generation memory. But because memristors can remember their past state without using any power, their biggest potential all along has been as a realistic analogue to synapses in brains.

Here's why. A memristor is a two-terminal device whose resistance changes depending on the amount, direction, and duration of voltage that's applied to it. But here's the really interesting thing about a memristor: Whatever its past state, or resistance, it freezes that state until another voltage is applied to change it. Maintaining that state requires no power. That's different from a dynamic RAM cell, which requires regular charge to maintain its state. The upshot is that thousands of memristors could substitute for massive banks of power-hogging memory. Just to be clear, the memristor is not magic—its memristive state does decay over time. That decay can take hours or centuries depending on the material, and stability must often be traded for energy requirements—which is one of the major research reasons memristors aren't flooding the market yet.

Physically, a memristor is just an oxide junction between two perpendicular metal wires. The generic memristor can be thought of as a nano-size sandwich—the bread is the intersection of the two crossing wires. Between the "bread" slices is an oxide; charge-carrying bubbles of oxygen move through that oxide and can be pushed up and down through the material to determine the state—the last resistance—across the memristor. This resistance state is what freezes when the power is cut. Recent DARPA-sponsored work at HP has yielded more complex memristors, so this description is necessarily a bit generic. The important thing to recall is that the memristor's "state" can be considered analogous to the state of the synapse that we mentioned earlier: The state of the synapse depends how closely any two neurons are linked, which is a key part of the mammalian ability to learn new information.

The architecture of the brain-inspired microprocessor under development at HP Labs can be thought of as a kind of memristor-based multicore chip [see illustration, "MoNETA: A Mind Made of Memristors"]. Nowadays, high-end microprocessors all have multiple cores, or processing units. But instead of the eight or so cores typical of such a microprocessor, the HP hardware will contain hundreds of simple, garden-variety silicon processing cores, and each of these will have its own ultradense thicket of memristor lattices.

Each silicon core is directly connected to its own immediately accessible megacache made up of millions of memristors, meaning that every single core has its own private massive bank of memory. Memristors are incredibly tiny, even by the standards of today's semiconductor transistors: HP senior fellow Stan Williams claims that with advances in fabrication processes for stacking many crossbars on a single chip, within a couple of decades it will be possible to build a nonvolatile memristor-based memory with a petabit (a quadrillion bits) per square centimeter.

Though memristors are dense, cheap, and tiny, they also have a high failure rate at present, characteristics that bear an intriguing resemblance to the brain's synapses. It means that the architecture must by definition tolerate defects in individual circuitry, much the way brains gracefully degrade their performance as synapses are lost, without sudden system failure.

Basically, memristors bring data close to computation, the way biological systems do, and they use very little power to store that information, just as the brain does. For a comparable function, the new hardware will use two to three orders of magnitude less power than Nvidia's Fermi-class GPU. For the first time we will begin to bridge the main divide between biological computation and traditional computation. The use of the memristor addresses the basic hardware challenges of neuromorphic computing: the need to simultaneously move and manipulate data, thereby drastically cutting power consumption and space. You might think that to achieve processing that's more like thinking than computation would require more than just new hardware—it would also require new software. You'd be wrong, but in a way that might surprise you.

Basically, without this paradigm shift in hardware architecture, you couldn't even think about building MoNETA.

TO BUILD A BRAIN, you need to throw away the conceit of separate hardware and software because the brain doesn't work that way. In the brain it's all just wetware. If you really wanted to replicate a mammalian brain, software and hardware would need to be inextricable. We have no idea how to build such a system at the moment, but the memristor has allowed us to take a big step closer by approximating the biological form factor: hardware that can be both small and ultralow power.

Where HP is taking care of the hardware component of the neuromorphic processor, we are building the software—the brain models that will populate the hardware. Our biological algorithms will create this entity: MoNETA. Think of MoNETA as the application software that does the recognizing, reasoning, and learning. HP chose our team at Boston University to build it because of our experience at the Center of Excellence for Learning in Education, Science, and Technology (CELEST), funded by the National Science Foundation. At CELEST, computational modelers, neuroscientists, psychologists, and engineers collaborate with researchers from Harvard, MIT, Brandeis, and BU's own department of cognitive and neural systems. CELEST was established to study basic principles of how the brain plans, organizes, communicates, and remembers.

To allow the brain models and the neuromorphic hardware to interact, HP built a kind of special-purpose operating system called Cog Ex Machina. Cog, built by HP principal investigator Greg Snider, lets system designers interact with the underlying hardware to do neuromorphic computation. Neuromorphic computation means computation that can be divided up between hardware that processes like the body of a neuron and hardware that processes the way dendrites and axons do.

The two kinds of cores deal with processing in fundamentally different ways. A "neuron-type" CPU architecture makes this core flexible, letting it handle any operation you throw at it. In that way, its

characteristics resemble those of the neuron. But the trade-off is that the core sucks up a lot of power, so like neurons, these elements should make up only a small percentage of the system.

A "dendritic" core works more like a GPU, an inexpensive and high-performance microprocessor. Like a dendrite, a GPU has a rigid architecture that is optimized for only a specific kind of computation—in this case, the complicated linear algebra operations that approximate what happens inside a dendrite. Because GPUs are optimized for parallel computation, we can use them to approximate the distributed computation that dendrites carry out. But there's a cost to using these, too: GPU cores perform only a limited set of operations. The dendrite cores in the final DARPA hardware will be much less flexible than neuron cores, but they will store extraordinary amounts of state information in their massive memristor-based memory banks, and like the tendrils of neurons, they will make up the vast bulk of the system's computational elements. Memristors, finally, will act as the synapses that mediate the information transfer between the dendrites and axons of different neurons. For a programmer, taking full advantage of a machine like this—with its two different core types and complicated memory-storage overlay—is tremendously challenging, because the problems need to be properly partitioned across those two radically different types of processors. Thanks to Cog, we computational neuroscientists can forget about the hardware and focus on developing the soul inside the machine.

MoNETA WILL BE A general-purpose mammalian-type intelligence, an artificial, generic creature known as an animat. With the DARPA hardware, we think we will be able to fit this level of intelligence into a shoebox.

The key feature distinguishing MoNETA from other AIs is that it won't have to be explicitly programmed. We are engineering MoNETA to be as adaptable and efficient as a mammal's brain. We intend to set it loose on a variety of situations, and it will learn dynamically.

Biological intelligence is the result of the coordinated action of many highly interconnected and plastic brain areas. Most prior research has focused on modeling those individual parts of the brain. The results, while impressive in some cases, have been a piecemeal assortment of experiments, theories, and models that each nicely describes the architecture and function of a single brain area and its contribution to perception, emotion, and action. But if you tried to stitch those findings together, you would more likely end up with a nonfunctioning Frankenstein's monster than anything like a mammalian intelligence.

Truly general-purpose intelligence can emerge only when everything happens all at once: In intelligent creatures like our humble rat, all perception (including auditory and visual inputs, or the brain areas responsible for the generation of fine finger movements), emotion, actions, and reactions combine and interact to guide behavior. Perceiving without action, emotion, higher reasoning, and learning would not only fail to lead to a general purpose AI, it wouldn't even pass a commonsense Turing test.

Creating this grail-like unified architecture has been precluded by several practical limitations. The most important is the lack of a unified theory of the brain. But the creation of large centers such as CELEST has advanced our understanding of what key aspects of biological intelligence might be applicable to our task of building a general-purpose AI.

SYNAPSES

NEURONS

BRAIN

HUMAN

MoNETA: A Mind Made of Memristors

DARPA's neuromorphic chip is still a long way from reaching biological efficiency. Even with optimistic assumptions about how much information single artificial synapses can store, the hardware still has one-hundredth the efficiency of biology. Still, that's 2000 times as power efficient as today's best supercomputers.

MEMRISTOR

GPU-type processor (Artificial dendrite)

GPU-type cores

CPU-type cores (Artificial neurons)

CPU processors (Artificial neurons)

CPU processors

PROCESSORS

DARPA SyNAPSE Hardware Goals
- 10^6 "neurons" (neuron cores) per square centimeter
- 10^{10} "synapses" per square centimeter (memristors)
- About 100 milliwatts per square centimeter
- Total power consumption: 1 kilowatt

MICROCHIP

Human Cortex
- About 10^9 neurons per square centimeter
- About 10^{10} synapses per square centimeter
- About 2 milliwatts per square centimeter
- Total power consumption: 20 watts

ROBOT

Cortical-Scale Hardware System
- 10 000 neuromorphic chips
- 10^{10} "neurons"
- 10^{14} "synapses"
- Total power consumption: 1 kilowatt

HOW WILL WE KNOW we've succeeded? How will we know that all this effort and new hardware and new software have yielded what we want—an artificial intelligence? We'll know we have successfully built an animat when we are able to motivate MoNETA to run, swim, and find food dynamically, without being programmed explicitly to do so.

It should learn throughout its lifetime without needing constant reprogramming or needing to be told a priori what is good for it, and what is bad. This is a true challenge for traditional AI: It is not possible to preprogram a lifetime of knowledge into a virtual or robotic animat. Such wisdom has to be learned from the interaction between a brain—with its large (but not infinite) number of synapses that store memories—and an environment that is constantly changing and dense with information.

The animat will learn about objects in its environment, navigate to reach its goals, and avoid dangers without the need for us to program specific objects or behaviors. Such an ability comes standard-issue in mammals, because our brains are plastic throughout our lives. We learn to recognize new people and places, and we acquire new skills without being told to do so. MoNETA will need to do the same.

We will test our animat in a classic trial called the Morris water navigation task. In this experiment, neuroscientists teach a rat to swim through a water maze, using visual cues, to a submerged platform that the rat can't see. That task might seem simple, but it's anything but. To get to the platform, the rat must use many stupendously sophisticated brain areas that synchronize vision, touch, spatial navigation, emotions, intentions, planning, and motor commands. Neuroscientists have studied the water maze task at great length, so we know a great deal about how a rat's anatomy and physiology react to the task. If we can train the animat to negotiate this maze, we'll be confident that we have taken an important first step toward simulating a mammalian intelligence.

By the middle of next year, our researchers will be working with thousands of candidate animats at once, all with slight variations in their brain architectures. Playing intelligent designers, we'll cull the best ones from the bunch and keep tweaking them until they unquestionably master tasks like the water maze and other, progressively harder experiments. We'll watch each of these simulated animats interacting with its environment and evolving like a natural organism. We expect to eventually find the "cocktail" of brain areas and connections that achieves autonomous intelligent behavior. We will then incorporate those elements into a memristor-based neural-processing chip. Once that chip is manufactured, we will build it into robotic platforms that venture into the real world. Robot companions for the elderly, robots to be sent to Mars to forage autonomously, and unmanned aerial vehicles will be just the beginning.

Will these chips "experience" vision and emotions by simulating and appropriately connecting the brain areas known to be involved in the subjective experience associated with them? It's too soon to say. However, our goal is not to replicate subjective experience—consciousness—in a chip but rather to build functional machines that can behave intelligently in complex environments. In other words, the idea is to make machines that behave as if they are intelligent, emotionally biased, and motivated, *without* the constraint that they are actually aware of these feelings, thoughts, and motivations.

NEUROMORPHIC CHIPS won't just power niche AI applications. The architectural lessons we learn here will revolutionize all future CPUs. The fact is, conventional computers will just not get significantly more powerful unless they move to a more parallel and locality-driven architecture. While neuromorphic chips will first supplement today's CPUs, soon their sheer power will overwhelm that of today's computer architectures.

The semiconductor industry's relentless push to focus on smaller and smaller transistors will soon mean transistors have higher failure rates. This year, the state of the art is 22-nanometer feature sizes. By 2018, that number will have shrunk to 12 nm, at which point atomic processes will interfere with transistor function; in other words, they will become increasingly unreliable. Companies like Intel, Hynix, and of course HP are putting a lot of resources into finding ways to rely on these unreliable future devices. Neuromorphic computation will allow that to happen on both memristors and transistors.

It won't be long until all multicore chips integrate a dense, low-power memory with their CMOS cores. It's just common sense.

Our prediction? Neuromorphic chips will eventually come in as many flavors as there are brain designs in nature: fruit fly, earthworm, rat, and human. All our chips will have brains. ❑