

COVER

How We Found the Missing Memristor

**The memristor--the functional equivalent of a synapse--
could revolutionize circuit design**

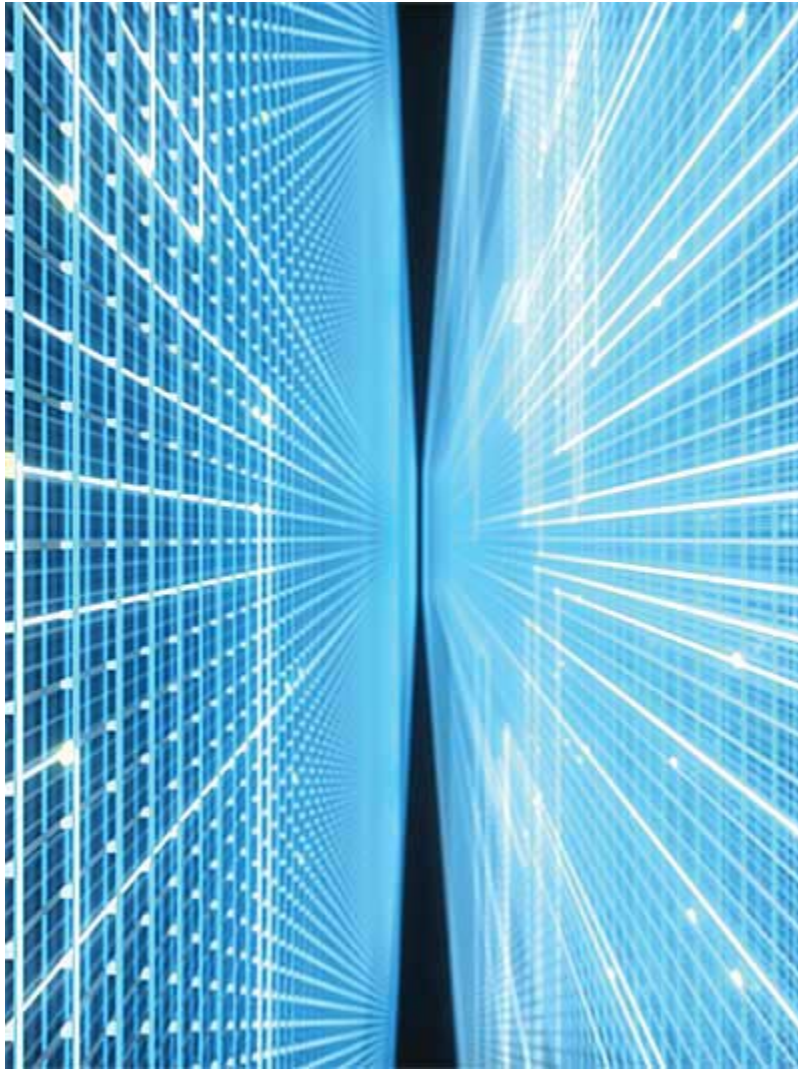


Image: Bryan Christie Design

THINKING MACHINE

This artist's conception of a memristor shows a stack of multiple crossbar arrays, the fundamental structure of R. Stanley Williams's device. Because memristors behave functionally

like synapses, replacing a few transistors in a circuit with memristors could lead to analog circuits that can think like a human brain.

BY R. Stanley Williams // December 2008

It's time to stop shrinking. Moore's Law, the semiconductor industry's obsession with the shrinking of transistors and their commensurate steady doubling on a chip about every two years, has been the source of a 50-year technical and economic revolution. Whether this scaling paradigm lasts for five more years or 15, it will eventually come to an end. The emphasis in electronics design will have to shift to devices that are not just increasingly infinitesimal but increasingly capable.

Earlier this year, I and my colleagues at Hewlett-Packard Labs, in Palo Alto, Calif., surprised the electronics community with a fascinating candidate for such a device: the memristor. It had been theorized nearly 40 years ago, but because no one had managed to build one, it had long since become an esoteric curiosity. That all changed on 1 May, when my group published the details of the memristor in *Nature*.

Combined with transistors in a hybrid chip, memristors could radically improve the performance of digital circuits without shrinking transistors. Using transistors more efficiently could in turn give us another decade, at least, of Moore's Law performance improvement, without requiring the costly and increasingly difficult doublings of transistor density on chips. In the end, memristors might even become the cornerstone of new analog circuits that compute using an architecture much like that of the brain.

For nearly 150 years, the known fundamental passive circuit elements were limited to the capacitor (discovered in 1745), the resistor (1827), and the inductor (1831). Then, in a brilliant but underappreciated 1971 paper, Leon Chua, a professor of electrical engineering at the University of California, Berkeley, predicted the existence of a fourth fundamental device, which he called a memristor. He proved that memristor behavior could not be duplicated by any circuit built using only the other three elements, which is why the memristor is truly fundamental.

Memristor is a contraction of "memory resistor," because that is exactly its function: to remember its history. A memristor is a two-terminal device whose resistance depends on the magnitude and polarity of the voltage applied to it and the length of time that voltage has been applied. When you turn off the voltage, the memristor remembers its most recent resistance until the next time you turn it on, whether that happens a day later or a year later.

Think of a resistor as a pipe through which water flows. The water is electric charge. The resistor's obstruction of the flow of charge is comparable to the diameter of the pipe: the narrower the pipe, the greater the resistance. For the history of circuit design, resistors have had a fixed pipe diameter. But a memristor is a pipe that changes diameter with the amount and direction of water that flows through it. If water flows through this pipe in one direction, it expands (becoming less resistive). But send the water in the opposite direction and the pipe shrinks (becoming more resistive). Further, the memristor remembers its diameter when water

last went through. Turn off the flow and the diameter of the pipe "freezes" until the water is turned back on.

That freezing property suits memristors brilliantly for computer memory. The ability to indefinitely store resistance values means that a memristor can be used as a nonvolatile memory. That might not sound like very much, but go ahead and pop the battery out of your laptop, right now—no saving, no quitting, nothing. You'd lose your work, of course. But if your laptop were built using a memory based on memristors, when you popped the battery back in, your screen would return to life with everything exactly as you left it: no lengthy reboot, no half-dozen auto-recovered files.

But the memristor's potential goes far beyond instant-on computers to embrace one of the grandest technology challenges: mimicking the functions of a brain. Within a decade, memristors could let us emulate, instead of merely simulate, networks of neurons and synapses. Many research groups have been working toward a brain in silico: IBM's Blue Brain project, Howard Hughes Medical Institute's Janelia Farm, and Harvard's Center for Brain Science are just three. However, even a mouse brain simulation in real time involves solving an astronomical number of coupled partial differential equations. A digital computer capable of coping with this staggering workload would need to be the size of a small city, and powering it would require several dedicated nuclear power plants.

Memristors can be made extremely small, and they function like synapses. Using them, we will be able to build analog electronic circuits that could fit in a shoebox and function according to the same physical principles as a brain.

A hybrid circuit—containing many connected memristors and transistors—could help us research actual brain function and disorders. Such a circuit might even lead to machines that can recognize patterns the way humans can, in those critical ways computers can't—for example, picking a particular face out of a crowd even if it has changed significantly since our last memory of it.



Photo: Paul Sakuma/AP Photo

PICTURING MEMRISTANCE

HP Labs senior fellow R. Stanley Williams [left] and research physicist Duncan Stewart [right] explain the fourth fundamental circuit element. Williams worked with nearly 100 scientists and engineers to find the memristor.

The story of the memristor is truly one for the history books. When Leon Chua, now an IEEE Fellow, wrote his seminal paper predicting the memristor, he was a newly minted and rapidly rising professor at UC Berkeley. Chua had been fighting for years against what he considered the arbitrary restriction of electronic circuit theory to linear systems. He was convinced that nonlinear electronics had much more potential than the linear circuits that dominate electronics technology to this day.

Chua discovered a missing link in the pairwise mathematical equations that relate the four circuit quantities—charge, current, voltage, and magnetic flux—to one another. These can be related in six ways. Two are connected through the basic physical laws of electricity and magnetism, and three are related by the known circuit elements: resistors connect voltage and current, inductors connect flux and current, and capacitors connect voltage and charge. But one equation is missing from this group: the relationship between charge moving through a circuit and the magnetic flux surrounded by that circuit—or more subtly, a mathematical doppelgänger defined by Faraday’s Law as the time integral of the voltage across the circuit. This distinction is the crux of a raging Internet debate about the legitimacy of our memristor [see sidebar, [”Resistance to Memristance”](#)].

Chua’s memristor was a purely mathematical construct that had more than one physical realization. What does that mean? Consider a battery and a transformer. Both provide identical voltages—for example, 12 volts of direct current—but they do so by entirely different mechanisms: the battery by a chemical reaction going on inside the cell and the transformer by taking a 110 V ac input, stepping that down to 12 V ac, and then transforming that into 12 V dc. The end result is mathematically identical—both will run an electric shaver or a cellphone, but the physical source of that 12 V is completely different.

Conceptually, it was easy to grasp how electric charge could couple to magnetic flux, but there was no obvious physical interaction between charge and the integral over the voltage.

Chua demonstrated mathematically that his hypothetical device would provide a relationship between flux and charge similar to what a nonlinear resistor provides between voltage and current. In practice, that would mean the device’s resistance would vary according to the amount of charge that passed through it. And it would remember that resistance value even after the current was turned off.

He also noticed something else—that this behavior reminded him of the way synapses function in a brain.

Even before Chua had his eureka moment, however, many researchers were reporting what they called "anomalous" current-voltage behavior in the micrometer-scale devices they had built out of unconventional materials, like polymers and metal oxides. But the idiosyncrasies were usually ascribed to some mystery electrochemical reaction, electrical breakdown, or other spurious phenomenon attributed to the high voltages that researchers were applying to their devices.

As it turns out, a great many of these reports were unrecognized examples of memristance. After Chua theorized the memristor out of the mathematical ether, it took another 35 years for us to intentionally build the device at HP Labs, and we only really understood the device about two years ago. So what took us so long?

It's all about scale. We now know that memristance is an intrinsic property of any electronic circuit. Its existence could have been deduced by Gustav Kirchhoff or by James Clerk Maxwell, if either had considered nonlinear circuits in the 1800s. But the scales at which electronic devices have been built for most of the past two centuries have prevented experimental observation of the effect. It turns out that the influence of memristance obeys an inverse square law: memristance is a million times as important at the nanometer scale as it is at the micrometer scale, and it's essentially unobservable at the millimeter scale and larger. As we build smaller and smaller devices, memristance is becoming more noticeable and in some cases dominant. That's what accounts for all those strange results researchers have described. Memristance has been hidden in plain sight all along. But in spite of all the clues, our finding the memristor was completely serendipitous.

In 1995, I was recruited to HP Labs to start up a fundamental research group that had been proposed by David Packard. He decided that the company had become large enough to dedicate a research group to long-term projects that would be protected from the immediate needs of the business units. Packard had an altruistic vision that HP should "return knowledge to the well of fundamental science from which HP had been withdrawing for so long." At the same time, he understood that long-term research could be the strategic basis for technologies and inventions that would directly benefit HP in the future. HP gave me a budget and four researchers. But beyond the comment that "molecular-scale electronics" would be interesting and that we should try to have something useful in about 10 years, I was given carte blanche to pursue any topic we wanted. We decided to take on Moore's Law.

At the time, the dot-com bubble was still rapidly inflating its way toward a resounding pop, and the existing semiconductor road map didn't extend past 2010. The critical feature size for the transistors on an integrated circuit was 350 nanometers; we had a long way to go before atomic sizes would become a limitation. And yet, the eventual end of Moore's Law was obvious. Someday semiconductor researchers would have to confront physics-based limits to their relentless descent into the infinitesimal, if for no other reason than that a transistor cannot be smaller than an atom. (Today the smallest components of transistors on integrated circuits are roughly 45 nm wide, or about 220 silicon atoms.)

That's when we started to hang out with Phil Kuekes, the creative force behind the Teramac (tera-operation-per-second multiarchitecture computer)—an experimental supercomputer built at HP Labs primarily from defective parts, just to show it could be done. He gave us the idea to

build an architecture that would work even if a substantial number of the individual devices in the circuit were dead on arrival. We didn't know what those devices would be, but our goal was electronics that would keep improving even after the devices got so small that defective ones would become common. We ate a lot of pizza washed down with appropriate amounts of beer and speculated about what this mystery nanodevice would be.

We were designing something that wouldn't even be relevant for another 10 to 15 years. It was possible that by then devices would have shrunk down to the molecular scale envisioned by David Packard or perhaps even *be* molecules. We could think of no better way to anticipate this than by mimicking the Teramac at the nanoscale. We decided that the simplest abstraction of the Teramac architecture was the crossbar, which has since become the de facto standard for nanoscale circuits because of its simplicity, adaptability, and redundancy.

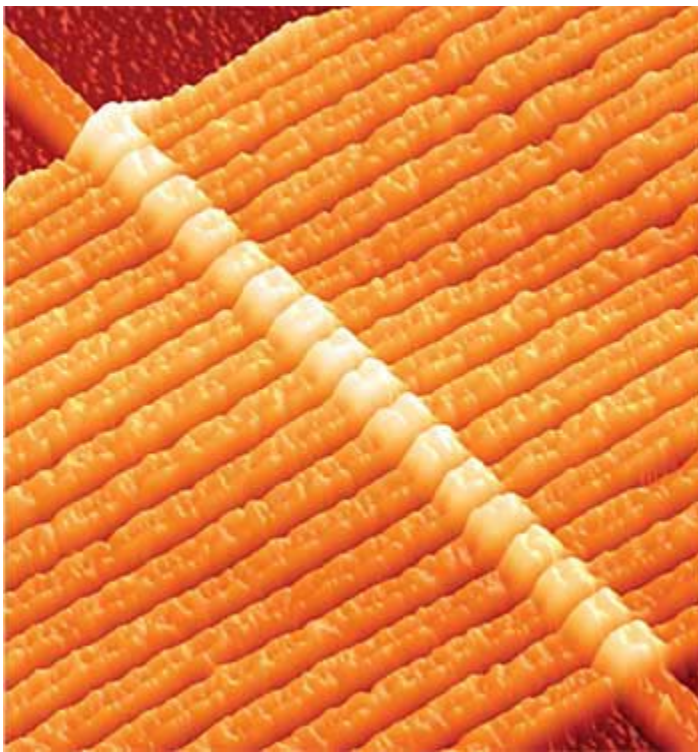


Image: R. STANLEY WILLIAMS/HP LABS

CROSSBAR ARCHITECTURE

A memristor's structure, shown here in a scanning tunneling microscope image, will enable dense, stable computer memories.

The crossbar is an array of perpendicular wires. Anywhere two wires cross, they are connected by a switch. To connect a horizontal wire to a vertical wire at any point on the grid, you must close the switch between them. Our idea was to open and close these switches by applying voltages to the ends of the wires. Note that a crossbar array is basically a storage system, with an

open switch representing a zero and a closed switch representing a one. You read the data by probing the switch with a small voltage.

Like everything else at the nanoscale, the switches and wires of a crossbar are bound to be plagued by at least some nonfunctional components. These components will be only a few atoms wide, and the second law of thermodynamics ensures that we will not be able to completely specify the position of every atom. However, a crossbar architecture builds in redundancy by allowing you to route around any parts of the circuit that don't work. Because of their simplicity, crossbar arrays have a much higher density of switches than a comparable integrated circuit based on transistors.

But implementing such a storage system was easier said than done. Many research groups were working on such a cross-point memory—and had been since the 1950s. Even after 40 years of research, they had no product on the market. Still, that didn't stop them from trying. That's because the potential for a truly nanoscale crossbar memory is staggering; picture carrying around the entire Library of Congress on a thumb drive.

One of the major impediments for prior crossbar memory research was the small off-to-on resistance ratio of the switches (40 years of research had never produced anything surpassing a factor of 2 or 3). By comparison, modern transistors have an off-to-on resistance ratio of 10 000 to 1. We calculated that to get a high-performance memory, we had to make switches with a resistance ratio of at least 1000 to 1. In other words, in its off state, a switch had to be 1000 times as resistive to the flow of current as it was in its on state. What mechanism could possibly give a nanometer-scale device a three-orders-of-magnitude resistance ratio?

We found the answer in scanning tunneling microscopy (STM), an area of research I had been pursuing for a decade. A tunneling microscope generates atomic-resolution images by scanning a very sharp needle across a surface and measuring the electric current that flows between the atoms at the tip of the needle and the surface the needle is probing. The general rule of thumb in STM is that moving that tip 0.1 nm closer to a surface increases the tunneling current by one order of magnitude.

We needed some similar mechanism by which we could change the effective spacing between two wires in our crossbar by 0.3 nm. If we could do that, we would have the 1000:1 electrical switching ratio we needed.

Our constraints were getting ridiculous. Where would we find a material that could change its physical dimensions like that? That is how we found ourselves in the realm of molecular electronics.

Conceptually, our device was like a tiny sandwich. Two platinum electrodes (the intersecting wires of the crossbar junction) functioned as the "bread" on either end of the device. We oxidized the surface of the bottom platinum wire to make an extremely thin layer of platinum dioxide, which is highly conducting. Next, we assembled a dense film, only one molecule thick, of specially designed switching molecules. Over this "monolayer" we deposited a 2- to 3-nm

layer of titanium metal, which bonds strongly to the molecules and was intended to glue them together. The final layer was the top platinum electrode.

The molecules were supposed to be the actual switches. We built an enormous number of these devices, experimenting with a wide variety of exotic molecules and configurations, including rotaxanes, special switching molecules designed by James Heath and Fraser Stoddart at the University of California, Los Angeles. The rotaxane is like a bead on a string, and with the right voltage, the bead slides from one end of the string to the other, causing the electrical resistance of the molecule to rise or fall, depending on the direction it moves. Heath and Stoddart's devices used silicon electrodes, and they worked, but not well enough for technological applications: the off-to-on resistance ratio was only a factor of 10, the switching was slow, and the devices tended to switch themselves off after 15 minutes.

Our platinum devices yielded results that were nothing less than frustrating. When a switch worked, it was spectacular: our off-to-on resistance ratios shot past the 1000 mark, the devices switched too fast for us to even measure, and having switched, the device's resistance state remained stable for years (we still have some early devices we test every now and then, and we have never seen a significant change in resistance). But our fantastic results were inconsistent. Worse yet, the success or failure of a device never seemed to depend on the same thing.

We had no physical model for how these devices worked. Instead of rational engineering, we were reduced to performing huge numbers of Edisonian experiments, varying one parameter at a time and attempting to hold all the rest constant. Even our switching molecules were betraying us; it seemed like we could use anything at all. In our desperation, we even turned to long-chain fatty acids—essentially soap—as the molecules in our devices. There's nothing in soap that should switch, and yet some of the soap devices switched phenomenally. We also made control devices with no molecule monolayers at all. None of them switched.

We were frustrated and burned out. Here we were, in late 2002, six years into our research. We had something that worked, but we couldn't figure out why, we couldn't model it, and we sure couldn't engineer it. That's when Greg Snider, who had worked with Kuekes on the Teramac, brought me the Chua memristor paper from the September 1971 *IEEE Transactions on Circuits Theory*. "I don't know what you guys are building," he told me, "but this is what I want."

To this day, I have no idea how Greg happened to come across that paper. Few people had read it, fewer had understood it, and fewer still had cited it. At that point, the paper was 31 years old and apparently headed for the proverbial dustbin of history. I wish I could say I took one look and yelled, "Eureka!" But in fact, the paper sat on my desk for months before I even tried to read it. When I did study it, I found the concepts and the equations unfamiliar and hard to follow. But I kept at it because something had caught my eye, as it had Greg's: Chua had included a graph that looked suspiciously similar to the experimental data we were collecting.

The graph described the current-voltage (I-V) characteristics that Chua had plotted for his memristor. Chua had called them "pinched-hysteresis loops"; we called our I-V characteristics "bow ties." A pinched hysteresis loop looks like a diagonal infinity symbol with the center at the zero axis, when plotted on a graph of current against voltage. The voltage is first increased from

zero to a positive maximum value, then decreased to a minimum negative value and finally returned to zero. The bow ties on our graphs were nearly identical [see graphic, "[Bow Ties](#)"].

That's not all. The total change in the resistance we had measured in our devices also depended on how *long* we applied the voltage: the longer we applied a positive voltage, the lower the resistance until it reached a minimum value. And the longer we applied a negative voltage, the higher the resistance became until it reached a maximum limiting value. When we stopped applying the voltage, whatever resistance characterized the device was frozen in place, until we reset it by once again applying a voltage. The loop in the I-V curve is called hysteresis, and this behavior is startlingly similar to how synapses operate: synaptic connections between neurons can be made stronger or weaker depending on the polarity, strength, and length of a chemical or electrical signal. That's not the kind of behavior you find in today's circuits.

Looking at Chua's graphs was maddening. We now had a big clue that memristance had something to do with our switches. But how? Why should our molecular junctions have anything to do with the relationship between charge and magnetic flux? I couldn't make the connection.

Two years went by. Every once in a while I would idly pick up Chua's paper, read it, and each time I understood the concepts a little more. But our experiments were still pretty much trial and error. The best we could do was to make a lot of devices and find the ones that worked.

But our frustration wasn't for nothing: by 2004, we had figured out how to do a little surgery on our little sandwiches. We built a gadget that ripped the tiny devices open so that we could peer inside them and do some forensics. When we pried them apart, the little sandwiches separated at their weakest point: the molecule layer. For the first time, we could get a good look at what was going on inside. We were in for a shock.

What we had was not what we had built. Recall that we had built a sandwich with two platinum electrodes as the bread and filled with three layers: the platinum dioxide, the monolayer film of switching molecules, and the film of titanium.

But that's not what we found. Under the molecular layer, instead of platinum dioxide, there was only pure platinum. Above the molecular layer, instead of titanium, we found an unexpected and unusual layer of titanium dioxide. The titanium had sucked the oxygen right out of the platinum dioxide! The oxygen atoms had somehow migrated through the molecules and been consumed by the titanium. This was especially surprising because the switching molecules had not been significantly perturbed by this event—they were intact and well ordered, which convinced us that they must be doing something important in the device.

The chemical structure of our devices was not at all what we had thought it was. The titanium dioxide—a stable compound found in sunscreen and white paint—was not just regular titanium dioxide. It had split itself up into two chemically different layers. Adjacent to the molecules, the oxide was stoichiometric TiO_2 , meaning the ratio of oxygen to titanium was perfect, exactly 2 to 1. But closer to the top platinum electrode, the titanium dioxide was missing a tiny amount of its oxygen, between 2 and 3 percent. We called this oxygen-deficient titanium dioxide TiO_{2-x} , where x is about 0.05.

Because of this misunderstanding, we had been performing the experiment backward. Every time I had tried to create a switching model, I had reversed the switching polarity. In other words, I had predicted that a positive voltage would switch the device off and a negative voltage would switch it on. In fact, exactly the opposite was true.

It was time to get to know titanium dioxide a lot better. They say three weeks in the lab will save you a day in the library every time. In August of 2006 I did a literature search and found about 300 relevant papers on titanium dioxide. I saw that each of the many different communities researching titanium dioxide had its own way of describing the compound. By the end of the month, the pieces had fallen into place. I finally knew how our device worked. I knew why we had a memristor.

The exotic molecule monolayer in the middle of our sandwich had nothing to do with the actual switching. Instead, what it did was control the flow of oxygen from the platinum dioxide into the titanium to produce the fairly uniform layers of TiO_2 and TiO_{2-x} . The key to the switching was this bilayer of the two different titanium dioxide species [see diagram, "[How Memristance Works](#)"]. The TiO_2 is electrically insulating (actually a semiconductor), but the TiO_{2-x} is conductive, because its oxygen vacancies are donors of electrons, which makes the vacancies themselves positively charged. The vacancies can be thought of like bubbles in a glass of beer, except that they don't pop—they can be pushed up and down at will in the titanium dioxide material because they are electrically charged.

Now I was able to predict the switching polarity of the device. If a positive voltage is applied to the top electrode of the device, it will repel the (also positive) oxygen vacancies in the TiO_{2-x} layer down into the pure TiO_2 layer. That turns the TiO_2 layer into TiO_{2-x} and makes it conductive, thus turning the device on. A negative voltage has the opposite effect: the vacancies are attracted upward and back out of the TiO_2 , and thus the thickness of the TiO_2 layer increases and the device turns off. This switching polarity is what we had been seeing for years but had been unable to explain.

On 20 August 2006, I solved the two most important equations of my career—one equation detailing the relationship between current and voltage for this equivalent circuit, and another equation describing how the application of the voltage causes the vacancies to move—thereby writing down, for the first time, an equation for memristance in terms of the physical properties of a material. This provided a unique insight. Memristance arises in a semiconductor when both electrons and charged dopants are forced to move simultaneously by applying a voltage to the system. The memristance did not actually involve magnetism in this case; the integral over the voltage reflected how far the dopants had moved and thus how much the resistance of the device had changed.

We finally had a model we could use to engineer our switches, which we had by now positively identified as memristors. Now we could use all the theoretical machinery Chua had created to help us design new circuits with our devices.

Triumphantly, I showed the group my results and immediately declared that we had to take the molecule monolayers out of our devices. Skeptical after years of false starts and failed

hypotheses, my team reminded me that we had run control samples without molecule layers for every device we had ever made and that those devices had never switched. And getting the recipe right turned out to be tricky indeed. We needed to find the exact amounts of titanium and oxygen to get the two layers to do their respective jobs. By that point we were all getting impatient. In fact, it took so long to get the first working device that in my discouragement I nearly decided to put the molecule layers back in.

A month later, it worked. We not only had working devices, but we were also able to improve and change their characteristics at will.

But here is the real triumph. The resistance of these devices stayed constant whether we turned off the voltage or just read their states (interrogating them with a voltage so small it left the resistance unchanged). The oxygen vacancies didn't roam around; they remained absolutely immobile until we again applied a positive or negative voltage. That's memristance: the devices remembered their current history. We had coaxed Chua's mythical memristor off the page and into being.

Emulating the behavior of a single memristor, Chua showed, requires a circuit with at least 15 transistors and other passive elements. The implications are extraordinary: just imagine how many kinds of circuits could be supercharged by replacing a handful of transistors with one single memristor.

The most obvious benefit is to memories. In its initial state, a crossbar memory has only open switches, and no information is stored. But once you start closing switches, you can store vast amounts of information compactly and efficiently. Because memristors remember their state, they can store data indefinitely, using energy only when you toggle or read the state of a switch, unlike the capacitors in conventional DRAM, which will lose their stored charge if the power to the chip is turned off. Furthermore, the wires and switches can be made very small: we should eventually get down to a width of around 4 nm, and then multiple crossbars could be stacked on top of each other to create a ridiculously high density of stored bits.

Greg Snider and I published a paper last year showing that memristors could vastly improve one type of processing circuit, called a field-programmable gate array, or FPGA. By replacing several specific transistors with a crossbar of memristors, we showed that the circuit could be shrunk by nearly a factor of 10 in area and improved in terms of its speed relative to power-consumption performance. Right now, we are testing a prototype of this circuit in our lab.

And memristors are by no means hard to fabricate. The titanium dioxide structure can be made in any semiconductor fab currently in existence. (In fact, our hybrid circuit was built in an HP fab used for making inkjet cartridges.) The primary limitation to manufacturing hybrid chips with memristors is that today only a small number of people on Earth have any idea of how to design circuits containing memristors. I must emphasize here that memristors will never eliminate the need for transistors: passive devices and circuits require active devices like transistors to supply energy.

The potential of the memristor goes far beyond juicing a few FPGAs. I have referred several times to the similarity of memristor behavior to that of synapses. Right now, Greg is designing new circuits that mimic aspects of the brain. The neurons are implemented with transistors, the axons are the nanowires in the crossbar, and the synapses are the memristors at the cross points. A circuit like this could perform real-time data analysis for multiple sensors. Think about it: an intelligent physical infrastructure that could provide structural assessment monitoring for bridges. How much money—and how many lives—could be saved?

I'm convinced that eventually the memristor will change circuit design in the 21st century as radically as the transistor changed it in the 20th. Don't forget that the transistor was lounging around as a mainly academic curiosity for a decade until 1956, when a killer app—the hearing aid—brought it into the marketplace. My guess is that the real killer app for memristors will be invented by a curious student who is now just deciding what EE courses to take next year.

About the Author

R. STANLEY WILLIAMS, a senior fellow at Hewlett-Packard Labs, wrote this month's cover story, "How We Found the Missing Memristor" [p. 28]. Earlier this year, he and his colleagues shook up the electrical engineering community by introducing a fourth fundamental circuit design element. The existence of this element, the memristor, was first predicted in 1971 by IEEE Fellow Leon Chua, of the University of California, Berkeley, but it took Williams 12 years to build an actual device.