An introduction to the paper by Rajendran, Karri, Wendt, Potkonjak, McDonald, Rose, and Wysocki

Nano Meets Security: Exploring Nanoelectronic Devices for Security Applications

By Amitava Dutta-Roy

The survey paper on hardware for computer security by Rajendran *et al.* that immediately follows this prolog touches upon—albeit not explicitly—on two important trends in the evolution of electronic and networking technologies. The first of them is the rapidly dwindling dimensions of electronic components. The second is the utmost urgency with which hardware *locks* are needed for ensuring computer security and the strategic roles they play in national defense and our financial safety. It may thus be worthwhile to spend few moments on these two trends before we hand over the reader to the survey of the work that are being done on several types of such hardware components. Among the components that can fortify computer security there is a relatively new entrant known as the memristor (contraction of memory resistor). It can however be realized on a nanometeric scale (10^{-9} of a meter) and it exhibits properties not encountered in other hardware elements hitherto. The combination of its small size and the special properties makes it a highly desirable candidate for use in security hardware. To comprehend the physical measurements of a memristor made to a scale of a billionth of a meter (a nanometer) we have to think of the comparable dimensions of a miniscule DNA strand in our own bodies. The tiny new player, theoretically conceived in 1971, 40 plus years later created at Hewlett Packard Research Labs, Palo Alto, Calif. and first publicly announced in 2008 in *Nature*, may soon change the established circuit design not only for security but for many other electronic circuits as well. It would thus be proper to briefly describe the concept and the operation of the memristor in this prolog.

Among the components that can fortify computer security there is a relatively new entrant known as the memristor (contraction of memory resistor).

I. Steady Downsizing of Electronics

After an uninterrupted innings of three decades ruled by diodes, triodes, and pentodes etc. a new era of electronics began in the middle of the last century. The invention of (*germanium*) transistor at the Bell Telephone Laboratories in Murray Hill, NJ, by William Bradford Shockley, John Bardeen, and Walter Hauser Brattain in 1947 initiated this revolution. In 1956, that invention brought a Nobel Prize for physics to the triumvirate who shared the glory and the financial reward. (Alas, there is no Nobel Prize for engineering. An equivalent award for engineering known as the Queen Elizabeth Prize for Engineering was instituted only in 2013.) The first *silicon* transistor in action was demonstrated by Morris Tanenbaum of the Bell Labs in 1954. Because of their small size, low-voltage operations, and low power consumption, transistors soon began replacing thermionic valves. Initially, palm-sized radios and quasi-unobtrusive hearing aids were probably the devices to take most advantage of the tiny transistors. But those devices relied on discrete transistors! Before long, accelerated research in semiconductor material sciences led to dozens and then thousands and later millions of transistors together with their metal interconnections to be fabricated on a single semiconductor chip. Better mass manufacturing techniques at fabrication plants (foundries or fabs as they are known in tech lingo) propelled down the prices of the chips. Dr. Gordon Moore, cofounder of Intel Corporation and a former director of research at the company, in 1965, made a prediction that the number of electronic components including transistors on an integrated circuit (IC) chip would double every year for the following ten years. In 1975, he revised his prediction and changed the

Digital Object Identifier: 10.1109/JPROC.2015.2418512

period of every year to every two years. The prediction popularly came to be known as Moore's law. Following that trend, the transistor count per chip was predicted to exceed 2.5 billion by 2013. Now an entire system or subsystem can be placed on a chip, that is, a "system-on-a-chip" or SoC. The area of a typical chip fabricated a couple of years ago measured anything between a few square mm to more than 400 mm^2. Each chip may contain nine million transistors per millimeter square of its surface area, a mind-boggling number. Compare this with the Electronic Numerical Integrator and Computer (ENIAC) that was installed at the University of Pennsylvania in 1946. It used 18 000 thermionic valves placed on 40 separate eight-foot-high racks. Scientists played with that configuration just to calculate ballistic trajectories. Today these calculations and much more can be done with an inexpensive laptop computer or a smart cell phone.

How far could we go down in miniaturizing the chips by increasing the number of tiny transistors on each of them? There were speculations that we were about to hit the limits of Moore's law because at this rapid rate of downsizing we would soon be staring at atomic dimensions. Many visionaries predicted that the Moore's law would meet its death in ten years. We will see later that the utilization of the new element memristor in real-life security hardware as highlighted by the survey paper would cast doubt on many naysayers' speculations. In other words, there is still scope for miniaturization of electronic circuits and the development of the memristor and the excitement over its use confirm that view.

II. CANDIES OR CANONS?

Sure, the Internet has brought us easy means of communications we could not even imagine even 20 years ago. In spite of many antisocial phenomena inadvertently spawned in the labyrinths of the Internet, it has made our lives simpler. It has generated thousands, if not millions, of jobs worldwide. However, the list of miseries it has heaped upon us is also long. Hacking, spread of virus and malware, cyber warfare between nations, spying (both political and industrial), and theft of financial data and health records are irksome, dangerous and ruinous even for the most ardent adherents of new technologies. The number of miscreants practicing those abominable cyber acts is growing steadily, their techniques bolder and more sophisticated by the day. It seems that they are always a step ahead of the law enforcement authorities. Most of us tend to think that all crimes against computers, networks and consequently against the society and our lives can be fixed by software alone. Admittedly, it is possible to do wonders with software. Consider, for example, the Defense Advanced Research Projects Agency (DARPA), itself the initiator of the Internet revolution, is now experimenting with Memex, a software that can sniff cyber criminal activities and practically on a fly stop some of them from spreading afar.

Indeed, at the consumer level many of us feel smug after subscribing to the service of a reputable antivirus software provider. In reality, we cannot do much beyond taking this instinctive approach. In a corporate or government environment, however, the situation is much more complex. The number of cyber attacks on those big machines and data coffers is humongous compared with the onslaughts on consumer computers. As the past loots have shown, a single such breach can mean loss of personal data—names, addresses, social security numbers and credit card numbers—of millions at one go. The attacks during the past couple of years on the financial institutions and big department stores in the United States and more recently the theft of information from the Indianapolis-based Anthem, one of the large health insurers in the United States demonstrated the villainous tendency of cyber attacks. Compared with attacks on personal computers bounty gained from a single corporate break-in is bigger by many orders. That is precisely the reason the masters of cyber crime play havoc with corporate data centers. Their invisible robotic arms attack the targets repeatedly, each time slightly altering a parameter or two or changing their strategies in the hope that they will finally be able to get into the target system. Often they succeed. Look at the recent news item published by the *New York Times* (February 14, 2015) and presumably by other news media as well about a report by Kaspersky Labs of Russia on *innovative* cyber crimes. Kaspersky is a firm specialized in detecting cyber crimes and it has discovered some hitherto unimaginable foul play. The thieves monitored mundane activities of employees of some banks, how they handled customers' monies and how they transferred funds to other banks nationally and internationally. The strategies of their attacks varied. For example, without any warning whatsoever, an ATM machine of a victim bank would, seemingly at random, start spewing out huge sums of money. (Obviously there would be a *recipient* waiting nearby to collect the proceeds of those ATMS.) The criminals even managed to see the screenshots of the bank's computers as well as videos (perhaps by hacking into the control system of the surveillance cameras) of movements of the bank personnel. Often after collecting the necessary information on the modus operandi in the banks they would patiently wait, sometimes for months before launching their attacks. Time was on their side and that minimized any suspicion. They would, for instance, suddenly inflate the balance in a target account by a certain amount and later transfer that extra money through the bank's legitimate channels to an undisclosed account with another bank. In the reported news item usually the individual amounts robbed tended to be around $10 million to avoid any auditory inspection though sums as high as $300 million were also transferred. Kaspersky reported that some 100 banks (including some top institutions in the United States) in 30 countries had been affected.

For such big jobs the thieves (or spies) do not depend on the tedious process of unearthing of passwords. Instead, they themselves or their accomplices create or know of the backdoors that many user entities—corporations or governments—may not even be aware of. Imagine a computer chip or a SoC guarding the entry port of a computer or a network. It acts like a guard at the entrance of an office building who is responsible for its safety. Before employing any security guard we check the detailed background (friends, family, habits and hobbies etc.) of the applicant and only when we are completely satisfied we can call the person "trusted." In the context of a computer chip do we know where it was designed and manufactured, and the path it precisely took to reach the hands of an integrator so that we could decide if it could be trusted or not?

It is not difficult to imagine that somewhere in the globalized manufacturing and supply chain of computer components a mole may intentionally be placed on a chip intended to guard an entry point. That mole literally can open flood gates to other venomous digital bugs that may devastate a government's defense system or the nation's communications network or even the country's vital power grid. Today most computers that are crucial to our safety and freedom are interconnected through webs of networks and once the wave of Internet of Things hit us they will be more so. Hence, once they gain access those bugs or agents may replicate themselves in the connected computers. With so much insurgency all over the world it is a horrendous scenario. To combat this crisis we have to be ever vigilant and that increases security budgets by leaps and bounds.

III. COMPLEXITY IN DESIGN AND FABRICATION OF CHIPS

Design and fabrication of modern computer chips have turned out to be a complex mixture of fundamental sciences and high tech. Integration of the chips in systems is not easy either. Furthermore, unlike in the yesteryears today there is not one single entity that we can call up to order a component or a subsystem that is made under its own roof. The supply chain is globalized. Chip design, for example, is mostly done with the help of specialty software that may be developed anywhere in the world. The designs contain thousands, if not millions, of codes. Blocks of software codes are constantly traded between contractors and their subcontractors. It is cheaper for a contractor to purchase blocks of relevant codes from a third party than to develop them from scratch in its own premises There is so much that goes at the design stage that it is quite possible to imperceptibly put in a few lines of undesirable codes among millions without anybody raising an eyebrow. Those codes may appear innocuous and stay idle for an indeterminate amount of time after which may carry out a preprogrammed malicious act. The hostile codes may also be activated by a command from outside, even wirelessly. They dig hidden holes that can endanger our digital society and the possibilities for them to commit such malevolent acts are limitless. Also, remember that a networked computer is not even essential for hostile agents to cause harm. They can act even if placed in stand-alone systems, for example, those found on a fighter jet or an antiballistic missile.

The philosophy of fabrication of chips has also changed during the past decade and a half. In the past the US government used to be the biggest buyer of electronics. Consequently, those days it could dictate the terms and conditions of procurement, manufacture and integration of chips. But not anymore! Compared to global demand for electronics, purchases by the government is minimal. The government can no longer leverage its requirements to acquire electronic products by insisting on specific suppliers that could be trusted. It is a safe bet therefore to assume that commercial products, in very large numbers, enter into the government systems.

As to the manufacture of chips, it needs heavy investment—estimated to be more than $10 billion and up—to build a decent foundry that only a few world-class companies can afford. Essential upgrade of the fabs may also cost tons of money, perhaps as much as $5 billion. Thus, system integrators must depend on chips made by large foundries flush with money. Most of them are situated in countries outside the United States. It is not implied that any of these multibillion dollar and reputed companies from abroad would plant moles on the chips at the design or fabrication stage. However, the large fabs themselves depend on third-party contractors for highly specialized items or codes of software. Hence, should a disaster happen who is to be held accountable?

Fearing the hidden backdoor implants in chips the Department of Defense (DoD) computers, in 2004, initiated a Trust Foundries Program. Under this program the DoD inspectors would certify some foundries as *trusted*. Note, however, that the inspectors could label only the processes as trusted and not the individual chips. (To read further about the DoD program see the IEEE Spectrum May 1, 2008 posting, "The Hunt for the Kill Switch" by Sally Adee).

We could now ask how would we test *individual* chips so that they fell in one of the two bins: one for the trusted and another for those not worthy of our trust? One way would be to test the input–output characteristics of a chip not only by checking the digital codes but also its power consumption, temperature, RF emission and time of response, etc. The assumption was that if any of the parameters did not correspond to the expected value it could be rejected as not trustworthy. Another way of looking at a chip would be to etch it away layer by layer and look into it with an electron microscope. But the process would destroy the chip even if no mole is found.

Researchers are also working on nondestructive creation of 3-D images of a chip by impinging it with X-rays

from several angles and then combining them. However, all such methods are time consuming and even if a chip, out of perhaps millions of similar chips, is found to be benign, there would be no guarantee that other chips in an entire batch could be trusted. Examination of just a few samples of batch-processed chips would not meet with the stringent security requirements.

What if we could insert an agent on the entry port of a chip that could throw challenges at an intruder in a fashion similar to a *friend or foe* query? Would it be alright to let the intruder in if its responses to the challenges are correct? What if the creator of the intruder knew about the nature of challenges in advance and designed the responses accordingly? These are valid questions and for this reason many high-level research groups are working to find an acceptable solution. It is very likely that soon multiple solutions will be found and then we have to choose the one that promises to be pilfer-proof, stable and affordable. Once such a solution is encountered for government machines large corporations may also accept that idea though it is likely that the miscreants will leap forward with something more malicious. So, a long-time security solution is our priority.

In 2011, the problem of testing chips prompted the DoD's Intelligence Advanced Research Projects Activity (IARPA) to initiate yet another program, that of *split* manufacturing. This meant that different part of a chip would be made by different fabs where the workers will not even know the function of each individual part. Hence, a potential miscreant would not be able to guess where to place the mole in a sea of transistors. These seemingly disjointed parts would finally be sent to a trusted integrator for integrating them by further processing or by digital coding. That trusted integrator may not need an ultra modern fab and thus it might be easier to find one.

In spite of the suggestion for split manufacturing the fundamental question remains: how to test each chip? The cost for doing such an inspection on a wide scale would be prohibitive. Work is in progress on placing a checker on the chip itself that can automatically detect a mole placed by its side. If that research succeeds we will have found a better way to secure our computers and networks without incurring the extra costs for inspection. (The IEEE Spectrum paper of February 2015 "The Trojan-Proof Chip" by Subhasish Mitra *et al.* explains the topic from the point of view of a group of researchers in quest for an ideal security trap.)

IV. MEMRISTOR, THE WONDER ELEMENT

In our EE classes, we were taught about the existence of three two-terminal passive electrical components: resistor, capacitor, and inductor. We also memorized the equations linking these components with four electric circuit variables, i.e., voltage, current, charge, and magnetic flux. These three elements combined with active components such as transistors were all we needed to design electric circuits. Indeed, even the supercomputers have been designed with this notion in mind. But a great surprise was lurking behind our backs.

Symmetry happens to be a natural order in physical sciences. Some forty plus years ago Prof. Leon Chua of the University of California at Berkeley and a Fellow of IEEE, while studying the fundamental equations governing the four variables of an electric circuit, noticed that an equation was missing to fulfill their symmetry conditions. That missing equation would link electrical charge with magnetic flux. In September of 1971, he published a paper in the IEEE Transactions of Circuit Theory in which he theorized that besides the three elements we knew so well there must be a fourth the properties of which would satisfy that missing equation and the symmetry. That element would be nonlinear and would exhibit hysteretic properties (somewhat similar to that found in magnets). Interestingly, Chua's fourth element, being hysteretic, would *remember* its state at which it was the last time an electric current was passed through it. Chua named the element memristor (short for memory plus resistor). No combination of the three passive elements could behave as a memristor but Chua was able to construct a circuit combining passive components with active devices (i.e., transistors) that would mimic a memristor. But memristor itself as a single component did not exist in real life.

For decades Chua's memristor remained shelved as an academic curiosity. Nobody associated the modern integrated circuits with the memristor though much research continued with various material for further downsizing of electronics, especially the transistors.

The HP Research Lab was one of the leading labs conducting such research on new materials under the guidance of Dr. R. Stanley Williams, a HP Fellow and also a Fellow of the IEEE. They had their early share of frustrations familiar to most researchers. However, in August 2006, the Williams team serendipitously found a structure made with two layers of TiO_2 sandwiched between two platinum electrodes that behaved like Chua's memristor. Williams reported his finding to the prestigious journal Nature that published it in its Letters section in May 2008. Williams himself wrote an eloquent story "How we found the missing memristor" on this development in the IEEE Spectrum of December 2008.

A flurry of praise, criticisms, and denials followed. It was reported that Chua himself said: "I am delighted because I never thought this would happen in my lifetime." But history of science is full of such stories about the time lag between theoretical predictions and their realization. They include Galileo's prediction of the rotation of earth and the recent discovery of Higg's Boson.

A simplistic illustration of the structure of a real memristor is shown in Fig. 1. It consists of two layers of TiO_2, one on top of the other. The bottom layer is a perfect

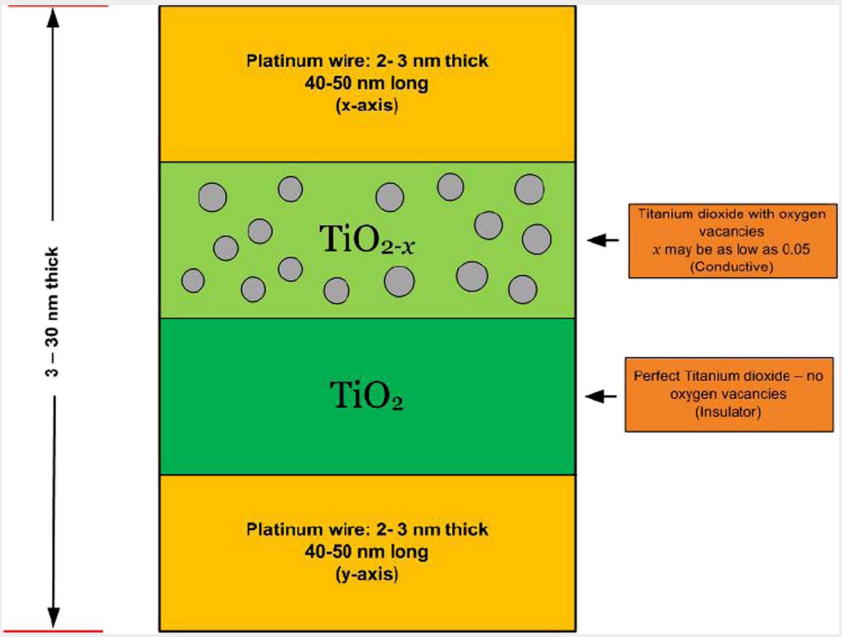

Fig. 1. *A simplistic illustration of a memristor structure (Not to scale). A tiny fraction of oxygen deficient TiO_2 molecules make the upper layer conductive (Source: IEEE Spectrum).*

titanium dioxide and behaves as an insulator. The upper layer, however, consists of TiO_{2-x} that has oxygen deficiencies and thus behaves like a conductive material. The oxygen deficiencies move around inside the material like bubbles (remember the movement of holes and electrons near a semiconductor junction?). Two platinum strips, one in x and the other in y direction, are placed on top and bottom of the composite titanium structure. When a positive voltage is applied to the top platinum wire the oxygen deficiencies are repelled and pushed down. Thus, the thickness of the conductive TiO_{2-x} increases bringing down its resistance. If, on the other hand, we apply a negative voltage to the top platinum wire the oxygen deficiencies will be attracted pulling to boundary between TiO_2 and TiO_{2-x} upwards increasing the resistance of the slab.

Surprisingly, if voltage of either polarity is withdrawn deficiencies are found to be *frozen* in space. Hence, if we probe the sandwich with a tiny voltage (compared to the voltage V applied earlier) we will find the resistance high or low depending on the value and polarity of V before it was withdrawn. In other words, the memristor remembers its last state and that memory is nonvolatile. This is unlike the behavior of a DRAM which loses its memory once the power is switched off. If we plot an i–v curve for the memristor we will find a pattern similar to the one shown in Fig. 2. Prof Chua called the curve "pinched hysteresis loops" and Williams named it "bow ties."

Because of its swings in resistance from high to low with a large ratio between them caused by the polarity and magnitude of the applied voltages at its terminals the memristor behaves like an on-off switch. Thus, a nonvolatile random access memory may be created with a large number of memristors placed on a simple mesh of crossbars of thin vertical and horizontal metallic wires. For fabricating such a crossbar, Dr. Williams of HP used about 2–3-nm-thick platinum wires (Fig. 3). The simplicity of a crossbar architecture is not the only advantage the

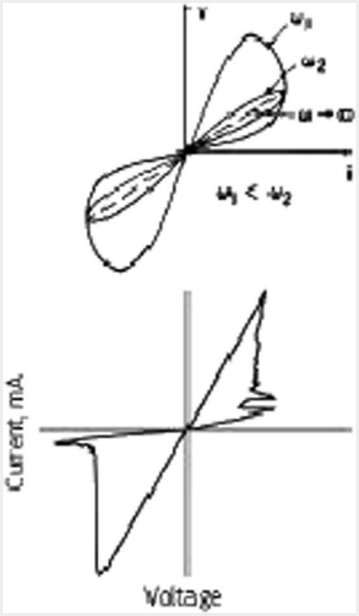

Fig. 2. *I–V characteristic of a memristor. "Pinched-hysteresis loops7"–Chua "Bow ties"–Williams (Source: IEEE Spectrum).*

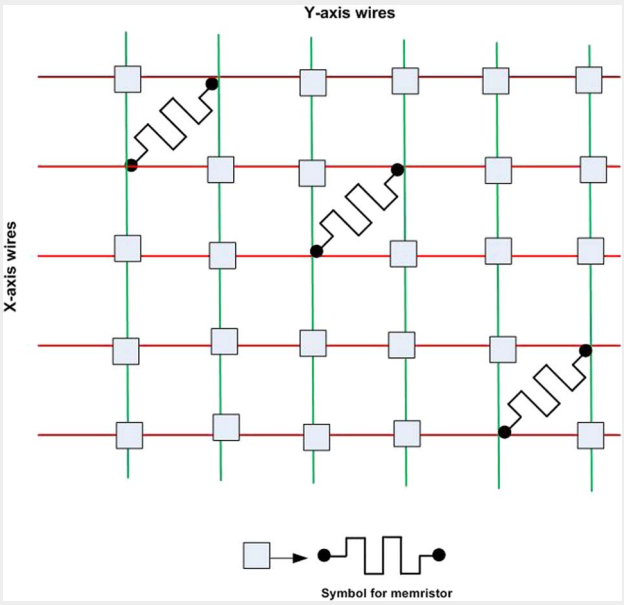

Fig. 3. *Crossbar architecture using memristors.*

memristor can offer. It has many characteristics the combination of which is not found in any other device. Some of these characteristics are listed below.

- Memristor is a two-terminal device that does not need a control electrode. Its footprint is tiny.
- Many such devices can be stacked vertically making it ultradense memory. The miniscule memristors would save physical space of a circuit by many orders of magnitude.
- They can act as both a flash memory and a dynamic random access memory (DRAM) that stores numbers in the working memory of a computer. This property can eliminate or at least reduce communication between the processor and the DRAM in the present machines making them more efficient. Note that the memory in a DRAM is volatile whereas memristor memory is not. Hewlett Packard and its partner South Korea-based Hynix are now working on computer architectures that will increasingly use memristors.
- A memristor should be able to act both as a processor and a memory making a system work faster and consume less energy making it ideal for mobile and wearable devices.
- The memristor also acts like synapses in the brain. In future, it should be possible to employ memristors for creating better artificial intelligence and diagnosing malfunctions of the brain. Researchers are already working on these projects.
- Since the memristor processor-memory combination is so tiny it would be ideal as a hardware lock for security purposes. This is of much relevance to the survey paper.

According to a report ("Six-State Memristor Opens Door to Weird Computing," IEEE Spectrum online, November 21, 2014) researchers at Trinity College Dublin has found the evidence that a memristor is also capable of exhibiting at least six states instead of only two as in a off-on switch. The researchers there further believe that a sufficiently small memristor could exhibit as many as ten states. If that happens would we start operating 10-based computers?

It is interesting, even if it is only an academic curiosity, to note that Prof. Chua *et al.* published a paper "The First Man-made Memristor: Circa 1801" in the Proceedings of the IEEE, January 2015. In this paper, the authors state that the English scientist Sir Humphry Davy (1778–1829) inadvertently invented the memristor when he demonstrated his carbon arc lamp powered by an electric battery, presumably the first source of light without using fire. Chua's group has recreated the Davy lamp and measured its characteristics that show unmistakable memoristic behavior.

V. THE SURVEY

Now that we have a fair idea of downsizing of electronics, the urgency and strategies for the use of hardware in security locks, and a new element known as memristor we should be able to appreciate the survey paper written by Rajendran*et al.* The authors have surveyed the published literature on research done around tiny electronic locks found in a nanometric world.

The authors have surveyed six different types of nano devices that could be used in hardware security circuits. They are: memristor, resistive random access memory (RRAM), contact-resistive random access memory (CRRAM), phase change memory (PCM), spin transfer torque access memory (STT-RAM), and orthogonal spin transfer random access memory (OST-RAM).The paper illustrates the structure of each of these devices including that of the memristor.

Complementary metal oxide semiconductor devices (CMOS) are now deemed to be the standard in many computer architectures including those used for security purposes. Furthermore, the processes of manufacture of CMOS are widely accepted even for the devices that are not CMOS-like. Rajendran *et al.* therefore have done well by deciding to compare the performances of nano devices with that of CMOS. Rajendran writes that the authors have tried to show "what new things can be done with new devices that we cannot do with CMOS devices."

The list of papers the authors have cited is indeed long. It is thus evident that the task of finding a hardware-based security solution is important and urgent. The issue of cloning of devices is discussed and so is the issue of challenges and responses (friend or foe!). The authors have devoted considerable space to the formulation of challenges and responses that seem to be harder than a non-specialist can imagine.

Table 1 of the paper gives the characteristics of all devices considered by the authors. Finally, a careful examination of Table 4 of the paper leads us to the conclusion that the memristor offers the best possible performance.

Asked about the physical measurements the lead author Rajendran replied: "Given that HP has a 50 nm × 50 nm × 50 nm memristor, a rough calculation puts the figure around 100 million memristor in a chip of size 1 mm × 1 mm. This is a highly conservative number. HP or other companies can do a lot better, may be by several order of magnitude or more."

The paper introduces many terms and abbreviations that a nonspecialist may not be familiar with. Each of the devices occupies its own niche that creates a plethora of new terms. The reader should treat each section of the paper with patience until the meaning of a term and its significance to the security issue become clear. Overall, the authors have done a splendid job in bringing to the forefront an urgent issue of the digital world. ■