REVIEW SUMMARY

DEVICE TECHNOLOGY

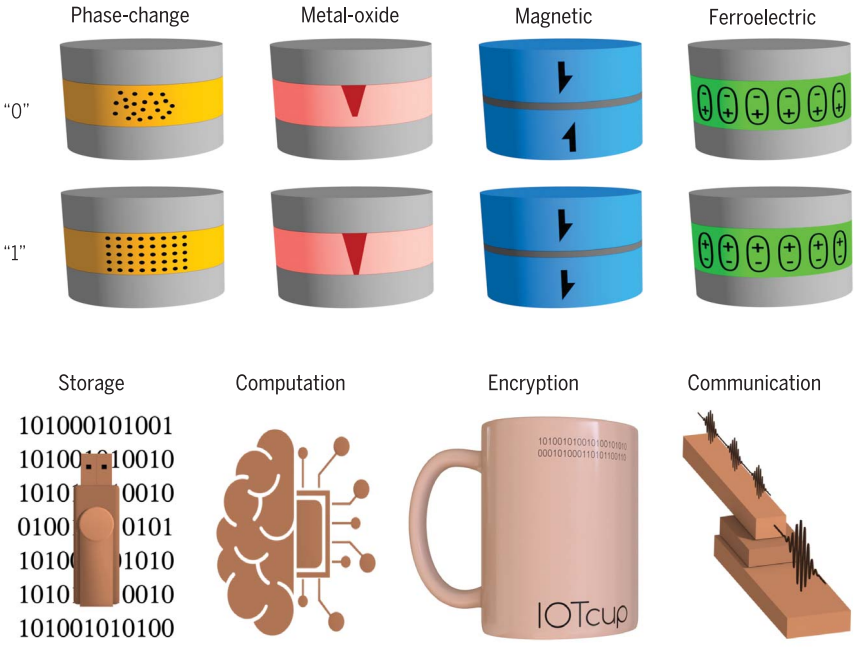
Memristive technologies for data storage, computation, encryption, and radio-frequency communication

Mario Lanza*, Abu Sebastian, Wei D. Lu, Manuel Le Gallo, Meng-Fan Chang, Deji Akinwande, Francesco M. Puglisi, Husam N. Alshareef, Ming Liu, Juan B. Roldan

BACKGROUND: Memristive devices exhibit an electrical resistance that can be adjusted to two or more nonvolatile levels by applying electrical stresses. The core of the most advanced memristive devices is a metal/insulator/metal nanocell made of phase-change, metal-oxide, magnetic, or ferroelectric materials, which is often placed in series with other circuit elements (resistor, selector, transistor) to enhance their performance in array configurations (i.e., avoid damage during state transition, minimize intercell disturbance). The memristive effect was discovered in 1969 and the first commercial product appeared in 2006, consisting of a 4-megabit nonvolatile memory based on magnetic materials. In the past few years, the switching endurance, data retention time, energy consumption, switching time, integration density, and price of memristive nonvolatile memories has been remarkably improved (depending on the materials used, values up to ~10^{15} cycles, >10 years, ~0.1 pJ, ~10 ns, 256 gigabits per die, and ≤$0.30 per gigabit have been achieved).

ADVANCES: As of 2021, memristive memories are being used as standalone memory and are also embedded in application-specific integrated circuits for the Internet of Things (smart watches and glasses, medical equipment, computers), and their market value exceeds $621 million. Recent studies have shown that memristive devices may also be exploited for advanced computation, data security, and mobile communication. Advanced computation refers to the hardware implementation of artificial neural networks by exploiting memristive attributes such as progressive conductance increase and decrease, vector matrix multiplication (in crossbar arrays), and spike timing–dependent plasticity; state-of-the-art developments have achieved >10 trillion operations per second per watt. Data encryption can be realized by exploiting the stochasticity inherent in the memristive effect, which manifests as random fluctuations (within a given range) of the switching voltages/times and state currents. For example, true random number generator and physical unclonable functions produce random codes when exposing a population of memristive devices to an electrical stress at 50% of switching probability (it is impossible to predict which devices will switch because that depends on their atomic structure). Mobile communication can also benefit from memristive devices because they could be employed as 5G and terahertz switches with low energy consumption owing to the nonvolatile nature of the resistive states; the current commercial technology is based on silicon transistors, but they are volatile and consume data both during switching and when idle. State-of-the-art developments have achieved cutoff frequencies of >100 THz with excellent insertion loss and isolation.

OUTLOOK: Consolidating memristive memories in the market and creating new commercial memristive technologies requires further enhancement of their performance, integration density, and cost, which may be achieved via materials and structure engineering. Market forecasts expect the memristive memories market to grow up to ~$5.6 billion by 2026, which will represent ~2% of the nearly $280 billion memory market. Phase-change and metal-oxide memristive memories should improve switching endurance and reduce energy consumption and variability, and the magnetic ones should offer improved integration density. Ferroelectric memristive memories still suffer low switching endurance, which is hindering commercialization. The figures of merit of memristive devices for advanced computation highly depend on the application, but maximizing endurance, retention, and conductance range while minimizing temporal conductance fluctuations are general goals. Memristive devices for data encryption and mobile communication require higher switching endurance, and two-dimensional materials prototypes are being investigated. ■



Fundamental memristive effects and their applications. Memristive devices, in which electrical resistance can be adjusted to two or more nonvolatile levels, can be fabricated using different materials (top row). This allows adjusting their performance to fulfill the requirements of different technologies. Memristive memories are a reality, and important progress is being achieved in advanced computation, security systems, and mobile communication (bottom row).

Part of *Science*'s coverage of the 75th anniversary of the discovery of the transistor.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: mario.lanza@kaust.edu.sa
Cite this article as M. Lanza *et al.*, *Science* **376**, eabj9979 (2022). DOI: 10.1126/science.abj9979

REVIEW

DEVICE TECHNOLOGY

Memristive technologies for data storage, computation, encryption, and radio-frequency communication

Mario Lanza[1]*, Abu Sebastian[2], Wei D. Lu[3], Manuel Le Gallo[2], Meng-Fan Chang[4,5], Deji Akinwande[6], Francesco M. Puglisi[7], Husam N. Alshareef[1], Ming Liu[8], Juan B. Roldan[9]

Memristive devices, which combine a resistor with memory functions such that voltage pulses can change their resistance (and hence their memory state) in a nonvolatile manner, are beginning to be implemented in integrated circuits for memory applications. However, memristive devices could have applications in many other technologies, such as non–von Neumann in-memory computing in crossbar arrays, random number generation for data security, and radio-frequency switches for mobile communications. Progress toward the integration of memristive devices in commercial solid-state electronic circuits and other potential applications will depend on performance and reliability challenges that still need to be addressed, as described here.

Each individual electronic device in an integrated circuit (IC)—such as resistors, capacitors, inductors, transistors, or diodes—controls the transport of charge carriers (electrons and holes) in a specific manner. Circuits based on these devices enable complex operations such as filtering, amplification, multiplexing, encoding, and storage (*1*). Early IC technology focused entirely on computation, with transistors performing switching functions and memory located off the chip. Modern ICs avoid the delays of remote memory for many functions with the addition of computing devices on-memory or near memory, such as floating-gate transistor nonvolatile (Flash) memories or charge-based volatile dynamic random-access memories (DRAMs) and static RAMs (SRAMs) that combine a capacitor and a transistor (Fig. 1, A to C) (*2*).

Memristive devices (often referred to as memristors or resistive switching devices) combine a resistor with memory functions (*3*, *4*). Several scientists think that the memristor, as initially defined by Chua in 1971 (*5*), has still never been realized (*3*). The use of resistive switching is more widely accepted, but it cannot fully capture the nonvolatile memory effect (i.e., the channel of a transistor also shows resistive switching when bias is applied, but it has no memory). We consider "memristive device" to be the most appropriate term for a device that behaves like a resistor with memory. Voltage pulses can change the resistance of the device, and the states are preserved without applying power (*4*). Such a programmable memory effect may help to enhance the performance of modern ICs, especially at the intersection of NAND Flash (used mainly in solid-state drives) and DRAM (used by microprocessors of computers to store data and program code when running software). When used in computers and phones, it would eliminate boot-up, reduce power consumption, and avoid loss of data when power fails.

Several physical effects in a variety of materials platforms have been explored for the implementation of memristive devices, including phase-change materials, metal oxides, magnetic materials, ferroelectric materials, carbon nanotubes, two-dimensional (2D) layered materials, polymeric and biological systems, and self-directing channels. State-of-the-art implementations have been based on metal/insulator/metal (MIM) nanocells, each of them with a lateral size as small as 10 nm × 10 nm (*6*), and have been intended mainly for use as memory in complementary metal-oxide semiconductor (CMOS) circuits. For example, in 2006, Freescale started to commercialize the first memristive product, a 4-megabit (Mb) nonvolatile memory (NVM) based on magnetic materials (*7*); in 2012, Panasonic launched a microcontroller with embedded memristive NVM made of metal-oxide materials (*8*); and in 2015, Intel and Micron started to commercialize a memristive persistent memory (a kind of memory placed in the memory bus for enhanced speed) based on phase-change MIM nanocells (*9*, *10*). Expansion of the segment of the memory market occupied by memristive devices is still limited by their cost, and more research is necessary to make them competitive alternatives.

The other main opportunity area for memristive devices stems from a non–von Neumann computing approach, in-memory computing (IMC), in which two or more programmable memory states are used. Typically implemented as crossbar arrays of vertical MIM nanocells, memristive devices can perform logical operations or complex tasks such as matrix multiplication, where multiple inputs (such as a set of numbers representing a vector) simultaneously are transformed into an output vector (*11*, *12*). The latter can also be exploited in deep neural networks (DNNs) and can execute computational tasks such as image and character recognition (i.e., artificial intelligence, or AI). The IMC approach can expend much less power than digital computation of the same operations and could have applications in areas such as robotics and Internet of Things (IoT). Other applications of memristive devices include data encryption (*13*, *14*) and radio-frequency (RF) operations for mobile communications (*15*, *16*). We review the recent progress of the most relevant memristive technologies and describe the main prospects and challenges to overcome if memristive devices are to be implemented in commercial ICs and in new device platforms.

Fundamental memristive effects

Memristive effects have been observed in devices with different structure and materials composition. Among them, two-terminal MIM nanocells (Fig. 1E) have attracted the most interest because of their good performance, simple fabrication, and high integration density in 2D or 3D crossbar arrays. In MIM nanocells, the electrical resistance of the insulator can be adjusted to two or more states by applying electrical stresses between the metallic electrodes (*4*). Memristive effects have been reported in MIM nanocells made of many different materials (*17*–*20*) and often result from atomic rearrangements induced by electrical fields or thermal effects that create conductive regions in an insulator or semiconductor, or contrariwise, return these regions to the original state (*21*).

In most studies, the quality of the memristive effect has been evaluated and compared by measuring the figures of merit of electronic memories, which include switching voltages, times, and energy as well as switching endurance and memory-state retention time (*17*, *22*).

[1]Materials Science and Engineering Program, Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. [2]IBM Research–Zurich, Rüschlikon, Switzerland. [3]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA. [4]Taiwan Semiconductor Manufacturing Company (TSMC), Hsinchu, Taiwan. [5]Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan. [6]Microelectronics Research Center, University of Texas, Austin, TX, USA. [7]Dipartimento di Ingegneria "Enzo Ferrari," Università di Modena e Reggio Emilia, 41125 Modena, Italy. [8]Key Laboratory of Microelectronic Devices and Integrated Technology, Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China. [9]Departamento de Electrónica y Tecnología de Computadores, Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain.
*Corresponding author. Email: mario.lanza@kaust.edu.sa

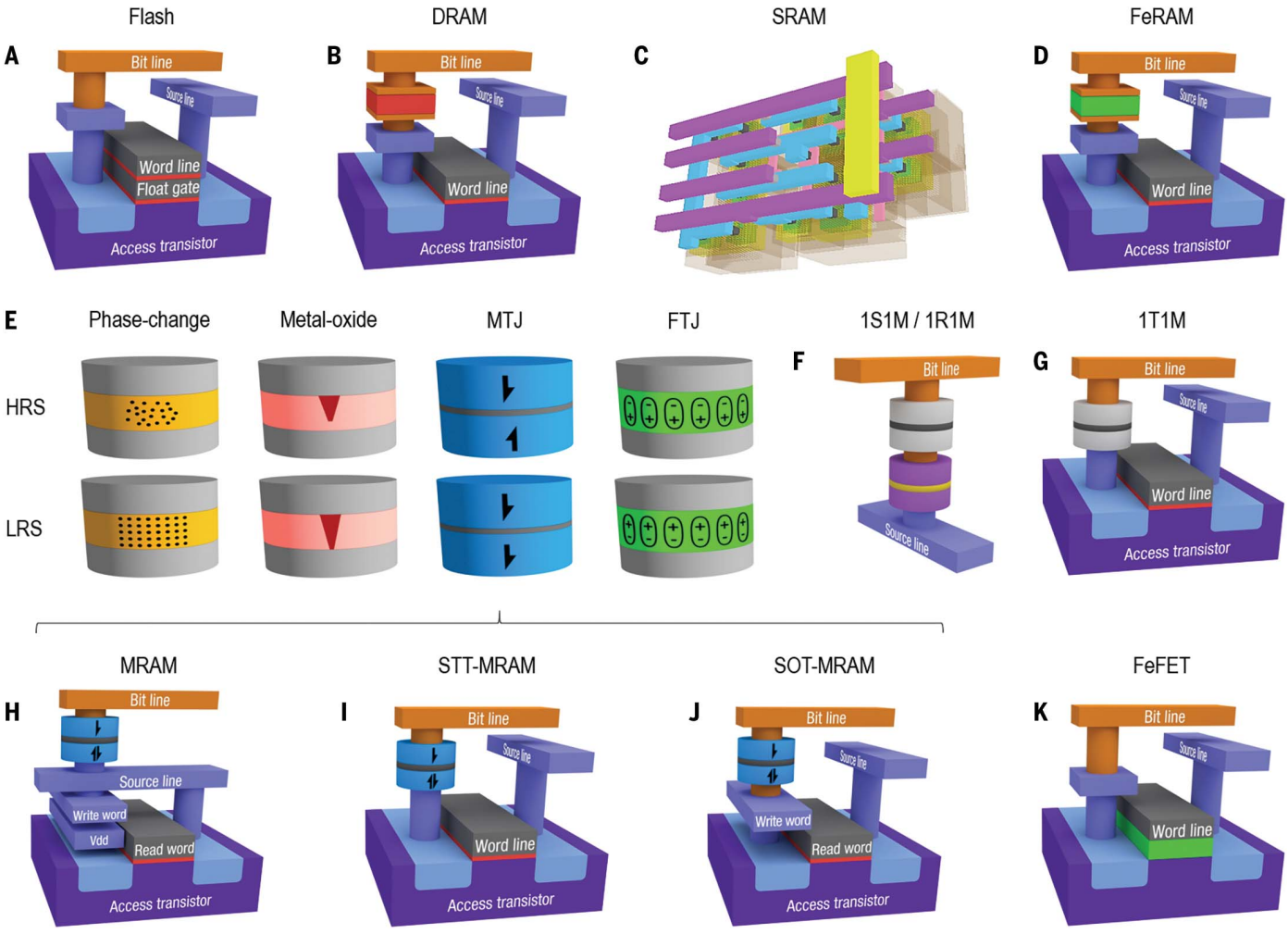


Fig. 1. Cell structure of the mainstream memories compared to memristive devices. (**A** to **D**) Three-dimensional schematics of the main non-memristive memories. Shown in (C) is the real layout of a six-transistor SRAM cell designed using Electric VLSI Design System software, and it is reproduced from (*126*). (**E**) Main memristive MIM nanocells and their working principles. (**F** and **G**) Resulting cell when adding one selector/resistor or one transistor in series to the MIM nanocell (represented as a light/dark/light gray cylinder). Such configurations are referred to as one-selector-one-memristor (1S1M), one-resistor-one-memristor (1R1M), and one-transistor-one-memristor (1T1M). Note that several works in the literature use the term 1T1R to refer to one-transistor-one-RRAM, i.e., the MIM nanocell (made of metal oxide and named RRAM) is referred to with the letter "R"; we did not use this notation here to avoid confusion. (**H** to **J**) Main memory cells derived from the magnetic tunnel junction. (**K**) Ferroelectric FET, showing that the ferroelectric material is integrated directly on the conductive channel.

According to these metrics, MIM-like memristive devices made of phase-change materials, metal oxides, magnetic materials, and ferroelectric materials have exhibited the best performance. However, in other memristive technology applications, different figures of merit are more important and other materials have shown even better performance. For example, the memristive devices made of 2D materials (such as hexagonal boron nitride, or h-BN) can process terahertz signals for RF devices (*23*).

Chalcogen-rich alloys such as $Ge_2Sb_2Te_5$ and $Ag_5In_5Sb_{60}Te_{30}$ can undergo a phase transition from a crystalline (low-resistance) to an amorphous (insulating) state (*18*). This memristive effect is exploited in phase-change memories (PCMs) that use rapid resistive (Joule) heating from high write currents followed by cooling to change the conductance state. Memristive devices made from metal oxides such as TaO_x and HfO_2, often referred to as resistive random-access memories (RRAMs), can change their electrical resistance in two different ways (*4*). If the electrodes are made of metals with a high diffusivity (such as Cu or Ag), the electrical field can move metallic ions from the electrodes into the insulator, which changes the overall resistance of the MIM cell. However, if the electrodes are metals with low diffusivity (such as Pt or W), the electrical field can only move the O ions within the metal-oxide insulator, leaving behind metallic atoms with dangling bonds that can enable electron flow.

In memristive devices made of magnetic materials (magnetoresistive random-access memory, or MRAM), the external electrical stress produces a change in the polarization of a magnetic tunnel junction (MTJ). In a MTJ, two magnetic layers, which could be Fe, Co, or CoFeB, are separated for a few nanometers by an insulator (such as MgO or Al_2O_3). One of the magnetic layers has a pinned magnetic state (the spin of the electrons cannot change), and when electrical stresses of different polarities are applied the magnetic state of the other (free) magnetic layer, it can change its direction (parallel or antiparallel with respect to the pinned one). This produces a net change of the out-of-plane resistance of the MTJ because electrons are more likely to tunnel across the insulator when both magnetic states have parallel orientation (*19*). This effect was first implemented in spin-transfer torque MRAM (STT-MRAM), which uses electrons with aligned

spins to torque the magnetic domains (consuming less power), and later in spin-orbit torque MRAM (SOT-MRAM), which is similar to STT-MRAM but with an additional adjacent metal line used to write the memory levels of the device (*19*). The MTJ in SOT-MRAM is not exposed to the write current, which provides almost unlimited endurance, but as it is a three-terminal device, its integration into crossbar arrays is more complex (Fig. 1, H to J).

Memristive devices can also be fabricated by placing a ferroelectric insulator a few nanometers thick [such as $Pb(Zr,Ti)O_3$ or $BaTiO_3$] between two electrodes (such as Pt, Co, or $La_{0.7}Sr_{0.3}MnO_3$) (*20*). The ferroelectric insulator is formed by crystalline unit cells that act as dipoles, and its orientation can be controlled depending on the external bias applied. For each dipole orientation the transmission coefficient across the insulator is different, and therefore the out-of-plane resistance can be tuned (*4*). This type of cell, known as ferroelectric tunnel junction (FTJ), is especially attractive because the quantum tunneling current used to write the device is very low, reducing the overall power consumption. FTJ should not be confused with ferroelectric random access memory (FeRAM), a device that uses a ferroelectric capacitor for nonvolatile data storage (Fig. 1D) and in which any readout mechanism would be more complex and destructive—hence, FeRAM is not a memristive device. The ferroelectric material can also be integrated between the semiconductor channel and the gate electrode of a field-effect transistor to form a three-terminal memristive device (often referred to as ferroelectric field-effect transistor or FeFET), which enables modulation of the channel conductivity by changing the orientation and magnitude of the ferroelectric polarization, leading to a large number of possible channel conductance levels (Fig. 1K).

Controlling the current across two-terminal memristive MIM nanocells to fulfill the specifications of commercial ICs can be very challenging. For individual devices, avalanche currents could appear during write operation (often referred to as overshoot) caused by self-accelerated thermal effects associated with the flow of current, which may introduce irreversible atomic rearrangements in the metal or insulating films (or both) that trigger the failure of the device. At an integration level, the presence of numerous memristive MIM nanocells in crossbar arrays can create interference between them. This effect, often referred to as sneak path current, can result in unintended state writing and reading in a nearby device when another is being addressed.

For these reasons, commercial ICs exploiting memristive devices often integrate an additional element in series to the memristive MIM nanocell (*4*), such as one transistor, one selector, or one resistor. Although the selector and the resistor are also two-terminal devices that can be integrated on top of the memristive MIM nanocell so that no additional area is consumed on a chip (see Fig. 1F for a simple example of vertical integration), the integration of a three-terminal transistor (Fig. 1G) remarkably increases the complexity and reduces the integration density of the entire circuit. Even though transistors can be very small, the real lateral size of the smallest transistors (in the 5-nm node) is actually ~20 nm (*24*), which is larger than the minimum size of state-of-the-art memristive MIM nanocells with acceptable performance: 10 nm (*6*). Academic groups have published observations of the memristive effect in even smaller structures with lateral sizes down to 2 nm (*25*) and even in one single atom (*26*), but in such reports the endurance was always very limited (<100 cycles), the yield was low (<50%), and the variability was high (not quantified).

Moreover, the transistors used to enable and disable MIM-like memristive nanocells in a crossbar array must drive high write currents (often >1 μA), which implies making them much bigger than the MIM nanocell itself. Some designs allow placing the crossbar array of memristive MIM nanocells directly above other necessary peripheral hardware to maximize integration density. Some memristive devices consist of a MIM nanocell with a third electrode adjacent to the insulator to provide an additional degree of control over the flow of electrons in a more compact manner (*27*). Nonetheless, for some memristive technologies, such as data encryption and mobile communication, ultrahigh integration density may not be required. Therefore, rather than limiting this review to any specific device structure, we highlight the studies that achieved the highest performance without regard to which memristive structure was used. For each memristive technology, we spotlight the performance of commercial devices (if any) and discard those studies in which performance claims have not been supported by sufficient data.

Two-state memristive memories

Memristive devices exhibiting two stable resistive states, a high resistive state (HRS) and a low resistive state (LRS), can be used to emulate the ones and zeros of the binary code, and therefore can be used to build NVMs. However, commercial NVMs must fulfill very stringent requirements for integration with current ICs. Among these requirements are integration densities up to 1 gigabyte (GB)/mm^2, writing voltages <3 V, switching energy <10 pJ, switching time <10 ns, writing endurance >10^{10} cycles, HRS/LRS resistance ratio >10, and small resistance fluctuations over time if no bias is applied (<10% for >10 years are preferred) (*22*). Some memristive devices have fulfilled such stringent criteria, but their manufacturing costs are orders of magnitude higher than that of the mainstream NVMs, such as the NAND Flash. Although the structure of a memristive MIM nanocell is simple, the manufacturing cost increases resulting from the need of additional elements (series transistor, selector, or resistor) and, even more importantly, because of the custom back-end-of-line (BEOL) interconnections outside the standard processing needed for CMOS transistors. Thus, the market segment occupied by memristive NVMs is still very small. As of 2021, they represent 0.5% of the ~$127 billion memory market (*28*). Table 1 presents a comparison of the performance of the mainstream versus memristive NVMs.

PCM is well understood in terms of device physics and manufacturability. In 2020, 90% of memristive NVMs commercialized were PCM (*28*). The main assets of PCM are high scalability (<10 nm) and low programming voltage (<3 V). The 3D XPoint technology developed by Intel/Micron (*9*, *10*) connects the PCM with an amorphous selector. For a minimum lateral feature size F, it is the only technology that has achieved a $4F^2$ cell size, and layer-by-layer stacking can further boost density. Dual in-line memory modules (DIMMs) with up to 512 GB of storage are offered. This maximum value doubles the density of current DRAM-based DIMM with lower cost.

The write speed of commercial PCMs (between 50 and 100 ns) is much longer than in other memristive NVMs because of the long crystallization process, and PCM shows limited endurance of 10^7 cycles (*29*). Nonetheless, PCM can expand the memory capacity of a system and reduce the amount of DRAM while maintaining high bandwidth, and can potentially reduce energy consumption and overall cost. As of 2020, Intel/Micron's PCMs are being used by multiple companies as persistent memory, which is placed in the memory bus for enhanced speed. ST-Microelectronics is qualifying PCMs for automotive applications such as microcontrollers for driver assistance systems, secure gateways, powertrain systems, and vehicle electrification (*30*).

PCMs require a high write current to produce sufficient Joule heating to melt the chalcogen-rich alloy. Although the programming current scales with device dimension, ~10 pJ is still required for switching a device with a lateral size of ~400 nm^2, which is two orders of magnitude higher than other memristive NVMs and three orders higher than DRAM (*31*). PCM shows inherent variability of switching voltages, times, and energies, as well as state resistances from one cycle to another and from one device to another, because the programming is based on atomic rearrangements.

Table 1. Comparison of the best performances of commercial stand-alone memories in 2021. Only stand-alone products are considered because estimating the density, performance, and cost of memristive devices embedded in other circuits may be challenging and inaccurate (such data are often intellectual property and are therefore not disclosed). Prototype chips with better values may have been demonstrated elsewhere but are not being commercialized as stand-alone products (because they are embedded or because they may still not fulfill industrial reliability requirements of stand-alone memories). "L" is the number of layers in a three-dimensional configuration; "F" is the minimum lithography feature size. The data are extracted from (*28*). The array energy is a relative estimation of the energy cost compared to the other types of memories. The applications list is nonexhaustive.

	NAND Flash	NOR Flash	DRAM	FeRAM	PCM	RRAM	STT-MRAM
Cell area	4/176L F^2	6 to 30 F^2	6 to 8 F^2	6 to 30 F^2	4/4L F^2	6 to 30 F^2	6 to 30 F^2
Bits per die	1 Tb	2 Gb	16 Gb	8 Mb	256 Gb	8 Mb	1 Gb
Retention	>10 years	>10 years	50 ms	>10 years	>10 years	>10 years	>10 years
Endurance	~10^4 cycles	~10^5 cycles	~10^{15} cycles	~10^{15} cycles	~10^7 cycles	~10^6 cycles	~10^{15} cycles
Read time	~100 µs	~100 µs	~10 ns	10 to 100 ns	10 to 100 ns	~100 ns	~10 ns
Write time	~10 µs	10 to 100 ns	~10 ns	10 to 100 ns	10 to 100 ns	~100 ns	~10 ns
Cell energy	~10 fJ	~100 pJ	~10 fJ	~0.1 pJ	~10 pJ	~0.1 pJ	~0.1 pJ
Array energy	High	High	High	Low	Medium	Medium	Medium/low
2021 price	$0.014/Gb	~$10/Gb	$0.50/Gb	>$1000/Gb	≤$0.30/Gb	~$1000/Gb	$40 to 70/Gb
Main application	Mass storage (USB, SSD)	Code execution, data storage	Computer and data memory (run software)	IoT, robotics, computing	Persistent memory and DIMM	Low-power, low-density IoT/ASICs	FPGA controllers, medical tools

Another concern is device reliability, including the spontaneous resistance drift over time arising from the structural relaxation of the amorphous phase and thermal cross-talk between adjacent cells when scaled to smaller, denser device arrays.

RRAM has demonstrated fast speed (<10 ns), large HRS/LRS resistance ratios (>100), low switching energy (<0.1 pJ), and high scalability. Fujitsu commercializes low-power 8-Mb stand-alone RRAM chips that can operate at 1.6 V with an average read current of 0.15 mA, which makes them suitable for different applications within the IoT, such as smart watches and glasses and hearing aids (*32*). Sandisk/Toshiba reported the development of stand-alone RRAM memory chips with a much higher integration density up to 32 GB (using 24-nm-node technology), where the CMOS-compatible RRAM fabrication process allows controllers and selectors to be located directly under the crossbar arrays (*33*). However, such high-density RRAM developments are still at the prototype stage and have not been commercialized.

Embedded RRAM solutions targeting system-on-chip applications have been provided by several companies. Intel has produced TaO_x-based 7.2-MB embedded RRAM with 22-nm-node FinFET (a field-effect transistor in which a double gate wraps around a fin-shaped source-drain contact) and low-power technology (leakage current of <1 pA per cell when used to build a six-transistor SRAM), showing 10 years of retention at 85°C and endurance of 10^4 cycles (*34*). TSMC has developed embedded RRAM at 22- and 40-nm nodes (*35*, *36*). Three-dimensional RRAM structures have been investigated as a way to further increase storage density. For example, an eight-layer TiN/HfO_2/TaO_x/Ti/TiN/W RRAM cell was fabricated by stacking eight electrode/insulation layers, etching vertical holes through the layers, and covering the holes with switching medium and electrode (*37*). RRAMs can be switched with write energy down to ~0.1 pJ (*5*, *38*).

During switching, the formation and rupture of a conductive nanofilament across the MIM nanocell are typically thermally assisted, which exponentially accelerates the migration of the active species (*39*). This process in turn can cause mechanical damage to the film in the form of plastic deformation, which results in unwanted and uncontrollable variations of switching voltages and state resistances. Such variations may be aggravated as operation proceeds, resulting in device failure (*40*). Some studies reported high switching endurances as high as >10^{10} cycles, although the real endurance of RRAM (as well as PCM) is a matter of controversy because many studies used unreliable characterization methods that present very few data points (*40*). Maximum endurances between 10^6 and 10^7 cycles are the most repeated (by different groups), but these are still insufficient and are hindering the use of RRAM as mainstream NVM. For a commercial single-level cell RRAM device, the reported endurance is >10,000 cycles and the resistance ratio ranges from 2 to 10 (*41*) for megabit-level array dimension. During the retention-after-cycling test, the experimental read window is ~7 µA after 10,000 cycles of set and reset operations (*42*). Approaches that can guide the active ionic species during filament growth, such as local doping, nanopore formation, geometry optimization, and defect engineering (among others), need to be actively investigated to minimize the stochastic behavior (*43*, *44*).

MRAM has attracted intense interest since the early 2000s, and products that target a wide range of applications are being commercialized, such as microcontrollers and watches. In general, trade-offs between write speed, endurance, and retention can be tuned to satisfy different application requirements. For NOR Flash-like applications, better retention (>10 years) is desired while energy consumption can be relaxed to ~100 pJ per transition. In this case, a material with a higher energy barrier can be used to enhance robustness to thermal disturbance, at a cost of higher writing energy. STT-MRAM, which offers better scalability and endurance (~10^{12}) than NOR Flash, has been demonstrated by several companies as embedded NVM at 22- or 28-nm nodes (*45*, *46*). It also features smaller cell size and nonvolatility relative to SRAM, although the speed and endurance are slightly worse (~10 ns and ~10^{15} cycles for MRAM versus ~1 ns and ~10^{16} cycles for SRAM).

Stand-alone memory chips have been produced to replace battery-backed SRAM or DRAM; they do not need to periodically refresh, so they can consume much less energy. STT-MRAM has the potential to replace SRAM in applications where performance can be relaxed for lower cost and lower energy consumption, such as mobile devices or IoT. Scenarios such as the last-level cache (a type of ultrafast memory between the RAM and the central processing unit that serves as a synchronizing buffer) have also been proposed with optimized materials that can achieve write speed as fast as 4 ns for the 14-nm node (*47*). The low HRS/LRS resistance ratio in MRAM (~2) also complicates the design of sensing circuitry. Finally, MRAM typically

involves complex stacks with 20 to 30 different metal and insulating layers, where the deposition and etching of this stack must be precisely controlled to ensure functionality and performance.

The development of FTJs and FeFETs is relatively recent and has yet to be commercialized. Initial studies have focused on single devices or small arrays, and array-scale characterization is still lacking (*48*). FTJ shows promising properties for applications that require low energy consumption, including low write energy (~0.1 pJ) and long retention (~10 years). Challenges facing FTJ include low switching speed and endurance relative to other NVMs. In an effort to increase the switching speed, an optimized Ag/$BaTiO_3$/Nb:$SrTiO_3$ stack was investigated as a means of achieving electric field–driven polarization reversal in the ferroelectric layer, leading to a switching time of 0.6 ns and lower switching energy (estimated to be 500 aJ per bit if the memristive MIM nanocell were 50 nm wide) (*49*). The highest endurance of FTJs (~10^7 cycles) was reported in HfZrO-based FTJ (*50*). Global Foundries have recently demonstrated FeFET using a 28-nm CMOS technology (*51*) as well as a 22-nm fully depleted silicon-on-insulator process (*52*), but the endurance was only ~10^5 cycles. New ferroelectric nanomaterials, such as 2D layered ferroelectric materials, are being investigated for NVM applications (*53*), but such activities are still being conducted exclusively by academics and are in a very early stage. Non-memristive FeRAM started to be commercialized by Samsung in 1996, and state-of-the-art devices offer very high endurance (~10^{15} cycles), high switching speed (~10 ns), long data retention (>10 years), and low power consumption (~0.1 pJ). However, its scalability is limited to a maximum storage capacity of few (~8) MB per die, which has limited its market size to ~$315 million in 2020 (<0.5% among all stand-alone memories) (*28*).

Some memristive devices exhibiting more than two stable resistive states have been proposed for multilevel NVMs, which would remarkably enhance the integration density because each MIM cell could store multiple data bits. PCM and RRAM possess higher HRS/LRS resistance ratios (>100) than MRAM (~2) and FTJ/FeFET (<100); therefore, they might support multilevel storage through write-and-verify schemes. However, the high variability of the state currents from one programming cycle to another and from one device to another make it very difficult to reliably identify each state.

Advanced computation with memristive devices

By exploiting the physical attributes of memristive devices and their array-level organization, it is also possible to perform certain computational tasks in the memory itself without the need to shuttle data between the memory and processing units. This IMC computational paradigm is finding a range of applications including scientific computing and deep learning (*11*, *12*). Memristive devices exhibiting two or more stable states can perform in-memory arithmetic operations such as matrix-vector multiplication (MVM). For example, to perform the operation $\mathbf{Ax} = \mathbf{b}$, the elements of matrix $\mathbf{A}$ are mapped linearly to the conductance values of memristive devices organized in a crossbar configuration. The values of the input vector $\mathbf{x}$ are mapped linearly to the amplitudes (durations) of read voltages and are applied to the crossbar along the rows. The resulting current (charge) measured along the columns of the array will be proportional to the result of the computation, $\mathbf{b}$. Yet another attribute exploited for computation is accumulative behavior, whereby the device conductance progressively increases or decreases with the successive application of programming pulses, which enables tuning of the synaptic weights of a machine learning model during training.

As shown in Fig. 2A, an IMC engine would ideally comprise a network of IMC cores, each of which would perform a MVM primitive along with some light digital postprocessing. Each IMC core comprises a crossbar array of memristive devices along with the bit-line drivers, analog-to-digital (ADC) converters, modest custom digital compute units to postprocess the raw ADC outputs, local controllers, transceivers, and receivers. Figure 2B presents the evolution of silicon-verified memristive IMC cores published in recent years.

In a DNN implemented with a standard von Neumann (CMOS) architecture, millions of synaptic weights are shuttled between memory and processor during deep learning inference and training, which consumes considerable energy and time. Recent studies have suggested that a DNN can be mapped onto multiple IMC cores that communicate with each other (*54*). The MVM operation corresponding to the realization of each DNN layer is performed in-memory, as described earlier. The results are then passed through a nonlinear activation function and input to the next layer. The nonlinear activation function is typically implemented at the core periphery, using analog or digital circuits, although recent studies proposed that memristive devices exhibiting highly nonlinear volatile switching could also perform that task (*55*).

Chips targeting DNN inference with IMC using memristive devices have been fabricated using RRAM (*41*, *56*), PCM (*57*, *58*), and MRAM (*59*, *60*). Usually, at least two devices per weight in a differential configuration are used to implement signed weights. The state-of-the-art experimental demonstrations of DNN inference based on IMC have reported a competitive energy efficiency of more than 10 trillion operations per second per watt (TOPS/W) for MVMs (see Table 2). However, a critical aspect for IMC implementations is the custom offline training and/or on-chip retraining of the network needed to mitigate the effects of stuck-at devices, device noise, and circuit level nonideality on network accuracy (*61*). It could also be possible to train the network entirely on-chip such that all the hardware nonidealities would be included as constraints during training. However, device-related challenges to performing precise weight updates need to be identified and overcome to obtain software-equivalent training accuracy with this approach (*62*). Another important research topic is the design of efficient intra- and interlayer pipelines such that all the cores on the chip are always active during inference, together with flexible core-to-core communication and control (*63*).

Another application domain for IMC in deep learning is spiking neural networks (SNNs). SNNs are neural networks that exhibit spatiotemporally sparse communication via spikes and are thus more biologically plausible than analog-valued communications. Moreover, neurons and synapses could have additional internal dynamics. SNNs offer great opportunities for local on-chip learning, exploiting temporal codes, and working with new types of event-based sensors. Memristive devices such as PCM (*64*) and RRAM (*65*) have been proposed to be integrated as part of the synapse and neuron circuits in a hardware SNN. Most of the early efforts have focused on implementing unsupervised learning with local learning rules with these devices. For example, spike timing–dependent plasticity (STDP), which adjusts a synaptic weight according to the relative timing between its output and input neuron spikes, can be implemented by applying multiple overlapping programming pulses to the devices (*64*, *65*). However, it is generally difficult to use STDP learning rules to reach the accuracy of conventional DNNs trained with back-propagation (*66*). Therefore, recent efforts have instead relied on converting a previously trained nonspiking DNN to an SNN (*66*), which is then implemented on memristive IMC hardware for inference (*67*). With the incorporation of additional bioinspired neuronal and synaptic dynamics, SNNs could potentially outperform conventional deep learning in certain application domains, and memristive devices could be exploited to natively implement such dynamics (*68*).

Recently, IMC has also been used to realize associative memory, which is an essential component of several machine learning algorithms. An associative memory compares input

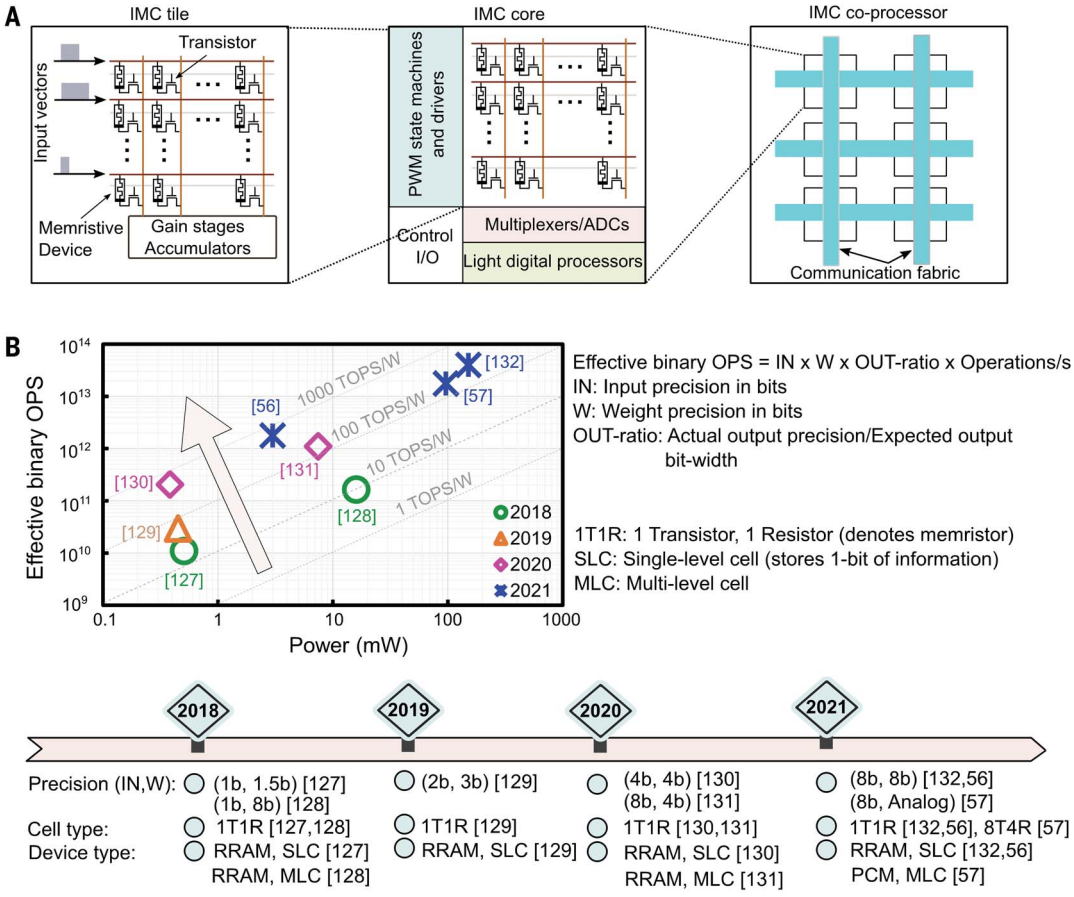


Fig. 2. Memristive cores for in-memory computing. (**A**) An IMC coprocessor typically comprises a network of IMC cores. Each IMC core has one or more crossbar arrays of unit cells comprising memristive devices along with the bit-line drivers, analog-to-digital (ADC) converters, modest custom digital compute units to post-process the raw ADC outputs, local controllers, and input/output interfaces. Several such IMC cores along with memory buffers, additional digital processing units, and global control units are interconnected by a communication fabric to realize a full-fledged IMC coprocessor. (**B**) An illustration of the evolution of post-silicon validated (fabricated and measured) memristive IMC cores published in recent years. There is a steady increase in the compute efficiency (effective binary operations per second per watt), represented by the diagonal lines in the graph. Note that the actual output precision in these cores is often less than the expected output bit width, and hence this has to be taken into consideration as well. The unit cells comprise one or more field-effect transistors and memristive devices storing binary or analog information. A noteworthy innovation that led to higher IMC core energy efficiency is the use of hybrid analog-digital readout circuits for MVM digitization, instead of purely analog readout schemes for which the signal margin decreases significantly with an increase in IN-W precision and the number of accumulations. Hybrid readout was first used in 2020 (*130*) and later in the 2021 cores, which led to a notable increase in energy efficiency.

search data with the data stored in it and finds the data entry with the closest match to the input data (*69*). This function can be realized by a content-addressable memory (CAM), which can be implemented with in-memory operations on memristive devices to reduce area and power consumption relative to traditional digital CAMs (*69*). However, although a conventional CAM finds an exact match between the input and stored data, it cannot compute the degree of match for each data entry with high precision (*70*). This limitation can be avoided by encoding the stored data directly in a crossbar array and computing, in parallel with IMC, the Hamming distances of each stored data vector with the input search data vector through in-memory dot products (*71*). This soft-reading type of associative memory search capability is used in several learning frameworks, such as hyperdimensional computing (*71*) and memory-augmented neural networks (*72*). Other applications that can leverage the lookup-table aspect of associative memory include tree-based models, finite-state machines, and pattern matching for genome sequencing (*69*, *73*).

Memristive IMC has also found applications in scientific computing. A prominent example is solving systems of linear equations (*74*), which can be used in a wide range of applications such as regression (*75*) and solving partial differential equations (*76*). One way to realize an accurate linear solver is to use the fast but imprecise MVM through IMC in an iterative linear solver, obtain an approximate solution, and then refine this solution based on the residual error calculated precisely through digital computing (*74*). At the system level, energy savings of as much as a factor of 6.8 were demonstrated with this method on large linear systems (>10,000 equations) relative to digital-only solutions. Nonetheless, the precision of the MVM performed with IMC ultimately prevents its application to ill-conditioned problems (when a small change in the input leads to a large change in the answer, which makes the solution to the equation hard to find) (*74*). Research avenues to increase MVM precision through bit slicing could enlarge the application space of IMC to also cover applications in scientific computing where high computational accuracy is required (*77*).

Another promising application of IMC is for combinatorial optimization problems, such as the traveling salesman problem, graph partitioning, Boolean satisfiability, and integer linear programming. Boltzmann machines and Hopfield networks have been proposed to address these computationally intensive, typically nondeterministic polynomial-time hard problems (*78*, *79*). IMC can be used to efficiently compute the inner products associated with these networks. Moreover, to achieve proper convergence, the native device noise injected in those inner products can be exploited as an explicit source of noise to force the network to continuously explore the solution space (*80*, *81*). An alternate approach to efficient search for solutions of combinatorial optimization problems includes the use of dynamics of networks of coupled nonlinear analog oscillators realized using memristive devices (*82*).

Table 2. State-of-the-art inference demonstrations with in-memory computing. Chip-level experimental demonstrations of neural network inference based on in-memory computing and comparison with one chip of a digital CMOS accelerator. Target values of these systems are application-specific; in general, maximization of the memory size, precision, energy, and area efficiency are desired.

Device	PCM	PCM	RRAM	MRAM	SRAM	SRAM	Digital CMOS
CMOS technology	14 nm	40 nm	22 nm	22 nm	16 nm	28 nm	16 nm
Memory size	65.5K cells	2M cells	4 Mb	128 kb	4.5 Mb	1 Mb	5 Mb
Input/weight/ output precision	8b/analog/8b	8b/8b/19b	8b/8b/14b	1b/1b/4b	4b/4b/8b	8b/8b/22b	8b/8b/8b
Network	MLP/ResNet-9	ResNet-20	ResNet-20	6-layer CNN	VGG	ResNet-20	ResNet-50
Dataset	MNIST CIFAR-10	CIFAR-10 CIFAR-100	CIFAR-10 CIFAR-100	CIFAR-10	CIFAR-10	CIFAR-10 CIFAR-100	ImageNet
Accuracy	98.3% 85.6%	91.89% 67.53%	92.01% 67.17%	90.1%	91.5%	92.08% 67.81%	No accuracy loss
Energy efficiency with max precision	10.5 TOPS/W	20.5 TOPS/W	15.6 TOPS/W	5.1 TOPS/W	121 TOPS/W	27.75 TOPS/W	0.96 TOPS/W
Energy efficiency (normalized to 1bIN-1bW)	336 TOPS/W	1312 TOPS/W	998.4 TOPS/W	5.1 TOPS/W	1936 TOPS/W	1776 TOPS/W	61.44 TOPS/W
Area efficiency with max precision	1.59 TOPS/ mm^2	0.026 TOPS/ mm^2	0.005 TOPS/ mm^2	0.758 TOPS/ mm^2	2.67 TOPS/ mm^2	0.1 TOPS/ mm^2	1.29 TOPS/ mm^2
Area efficiency (normalized to 1bIN-1bW)	50.88 TOPS/ mm^2	1.664 TOPS/ mm^2	0.32 TOPS/ mm^2	0.758 TOPS/ mm^2	42.72 TOPS/ mm^2	6.4 TOPS/ mm^2	82.56 TOPS/ mm^2
Reference	(*57*)	(*133*)	(*41*)	(*59*)	(*86*)	(*134*)	(*135*)

Memristive devices could also be used to implement a physical reservoir for reservoir computing, where the collective dynamics of an ensemble of such devices is used to perform certain machine learning tasks. For example, in (*83*), the use of a collection of tungsten oxide memristive devices was proposed to classify spoken digits. In (*84*), the authors used a reservoir of 1 million PCM devices and exploited their crystallization dynamics to classify binary random processes into correlated and uncorrelated classes. A reservoir of perovskite halide-based dynamic memristive devices was used to analyze neural signals in real time (*85*).

Besides memristive devices, IMC can also be realized with SRAM-based compute elements, recent demonstrations of which have shown impressive energy efficiency (*86*). However, with the potential for substantially smaller areal footprint and even 3D integration, memristive IMC is expected to have a substantial density advantage even at very advanced CMOS nodes (*87*). Additionally, the nonvolatility of memristive devices enables power cycling without reloading the operands from an external memory. There is also an emerging trend of information processing increasingly transitioning to the edge (as opposed to the data centers) and even to the end device (mobile devices and home assistants), driven by the cost of transmitting data and by privacy concerns. "Always on" computing systems, which operate at very low energy per area footprint, are ideal for these applications. Memristive devices may also play a key role in this space both for memory and computing applications (*88*).

Memristive devices for security applications

The intrinsic variability of the switching voltages, times, and energies of memristive devices from one cycle to another, as well as the fluctuations of their state resistance over time, could be used to generate unpredictable strings of bits, which are essential in user cryptographic systems (Fig. 3). One example is true random number generators (TRNGs), which are used daily by most electronic systems to generate unpredictable codes, such as the one-time passwords sent by banks to our phones when making online payments (Fig. 3A). Another example is physical unclonable functions (PUFs), a type of circuit that generates a unique string of bits that serves as a fingerprint for device identification, which could be integrated in any object if the power consumption is very low or if the circuit is self-powered (Fig. 3B). Unpredictable strings of bits are not generated through software because they could be easily attacked (*89*), and this limitation represents a huge opportunity and market for memristive devices.

State-of-the-art TRNGs and PUFs rely on the intrinsic stochasticity of some physical quantities available in electron devices or circuits. In modern TRNGs, a frequently adopted entropy source is thermal noise, which can be harvested from either a large resistor, or jitter in ring oscillators, or the modulations it induces in analog-to-digital converters (*90*–*92*). The main advantages of these TRNGs are high throughput (>1 Mb/s), low-voltage operation (≤0.5 V), high randomness, small size, and good scalability. However, in these systems, the electrical power needed to harvest the noise is too high, and they are vulnerable to noise and cryogenic attacks (*93*). Alternative approaches rely on the metastability of flip-flop circuits or on the time evolution of chaotic systems. Still, they typically exhibit large power consumption (>1 mW) (*94*), although low-power systems have been recently proposed (*95*). In addition, when randomness is harvested from an ensemble of devices as in this case and in some digital systems, careful minimization of process variations must be adopted (*96*).

Recent studies used the cycling variability, the intrinsic stochastic nature of the write and erase processes, or both in memristive devices (Fig. 3A) to design TRNGs (*13*, *14*), which showed promising performance (tens of megabits per second throughput at few picojoules of energy per bit) and the potential to achieve ultrafast (>1 GHz) and low-energy (few to tens of femtojoules per bit) operation. Non-memristive FeRAM implementation was demonstrated with off-the-shelf components (*97*) but its energy efficiency can hardly be reduced below 1 pJ per bit, whereas memristive FeFETs are currently limited by their insufficient endurance (*98*). Improvements are obtained with PCM (*99*) and MRAM (*100*), with the limiting factor being the programming times for PCM and the high writing current for MRAM. In all cases, complex peripheral circuitry is required to finely tune the applied

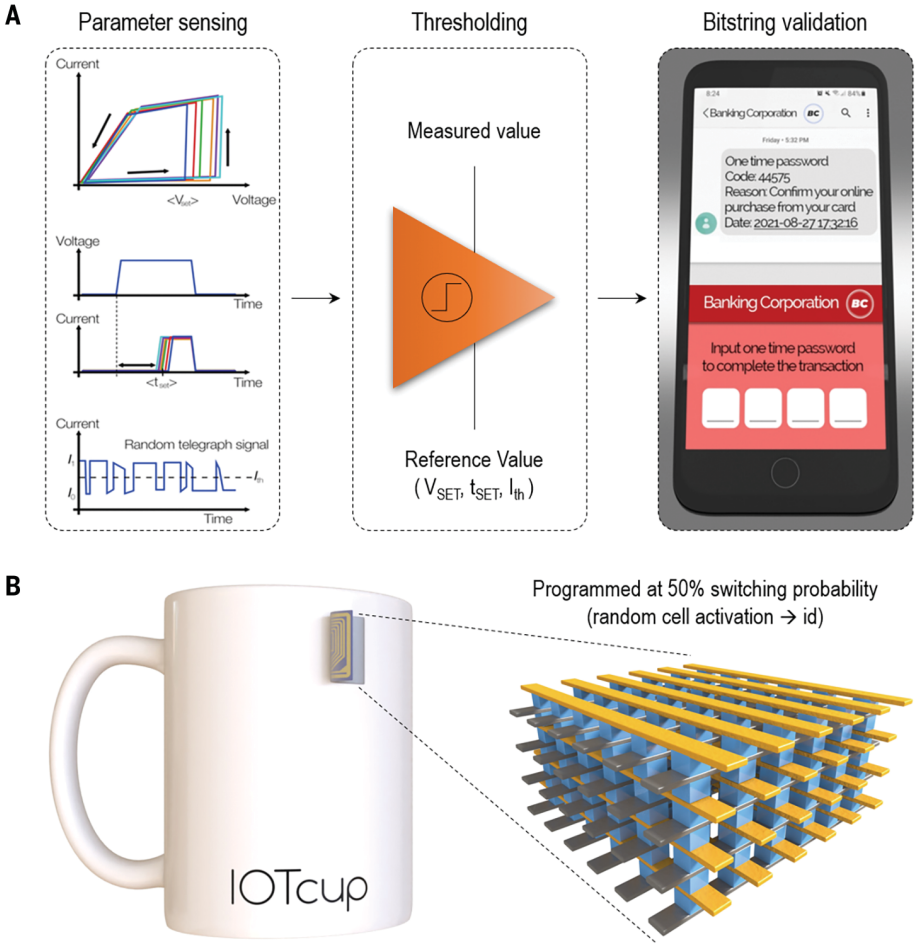


Fig. 3. Application of memristive devices for TRNG and PUF for encryption systems. (**A**) Block diagram of a memristive TRNG system for one-time password generation used for online payments. The polarization of memristive devices generates random fluctuations of some of its figures of merit (e.g., set voltage, set time, state current), which can be compared with a number to produce a string of zeros (e.g., lower) and ones (i.e., higher), with which random passwords can be generated. (**B**) Illustration of an application of a memristive PUF. When a population of memristors are exposed to a specific stress near its switching threshold (i.e., a voltage close to the average set voltage), some of these memristors will switch to a LRS, and others will not. Predicting which ones will switch depends on the atomic structure of each device, and therefore prediction is impossible. This can be employed to generate a digital fingerprint that can be used to identify objects.

voltage, track the time evolution of the switching statistics, and mitigate thermal effects, although promising results have been recently demonstrated using MRAM (*101*). In addition, the need of a full switching cycle to generate each random bit reduces the overall endurance of the TRNG circuit.

Memristive devices exhibiting volatile resistive switching could be a good alternative way to increase endurance because the atomic rearrangements produced by the stress are volatile (*102*), and also because they require simpler processing circuitry. In addition, volatile resistive switching has been observed at much lower current ranges, which reduces the power consumption (less than 1 μW per bit). However, the switching delay is difficult to reduce to below a few tens of microseconds, which in turn limits both the throughput to a few tens of kilobits per second and the energy efficiency to a few picojoules per bit. Nonetheless, further device engineering aimed at reducing the switching delay may make this option appealing for self-powered devices within IoT by bringing the throughput and the energy efficiency into the megabit-per-second and femtojoule-per-bit range, respectively. The possible detrimental role of endurance limitation and temperature variations still needs to be elucidated.

Another possibility is to exploit random telegraph noise (RTN), a quantum phenomenon related to the trapping and detrapping of electrons in the insulating film of memristive MIM nanocells, which appears as the abrupt switch of the measured current between two or more values at random times (*103*–*105*). A key advantage of RTN-based TRNGs is that voltages as low as few millivolts are sufficient to generate RTN at low currents (<100 nA), which enables ultralow power levels (~1 nW, excluding the peripheral electronics). Moreover, the dominant role of RTN over thermal noise improves immunity to cryptographic attacks, and device endurance and stability are expected to be high because no atomic rearrangements are needed to produce the switching. However, the switching times of TRNGs based on RTN signals observed in memristors can be large (micro- to milliseconds), limiting the throughput to a few tens of kilobits per second. Recent studies have demonstrated how these memristive TRNGs based on RTN can effectively be used to drive pseudo-RNGs (high-throughput deterministic RNG circuits) to realize energy-efficient, fast, hybrid TRNGs (*103*).

For PUFs implementations, state-of-the-art solutions typically exploit small random-delay differences that result from manufacturing variations on symmetrical electrical paths on a chip (such as arbiter PUF) or in multiple ring oscillators (*106*). However, silicon-based PUFs require a very large number of devices to guarantee secure operation, and they are relatively large both in physical implementation and energy consumption. Moreover, they are susceptible to side attacks—attempts to gain information from a system's operation (such as changes in power consumption)—and are not the ideal choice for exposed IoT systems. Memory-based PUFs based on SRAM or Flash technology do not require as many devices, are faster, and consume less energy. However, SRAM-based designs are also vulnerable to side attacks because the amount of thermal energy released when the cell settles to a stable state (randomly chosen between logic 0 or 1) upon power-up depends on the final state and can be measured from the outside (*107*).

Flash-based designs are slow, energy-intensive, and require large voltages (up to 15 V), which makes them unsuitable for IoT applications. Conversely, memristive devices offer a CMOS-compatible, fast, low-power alternative to Flash-based PUFs but keep the same fingerprint generation scheme. Memristive PUFs are normally implemented using a crossbar array of memristive MIM nanocells. Then, a voltage pulse is applied with amplitude and width that correspond to a 50% switching probability, thereby generating a random pattern of written and erased cells (Fig. 3B). Alternatively, the leakage current through each erased cell is directly compared against a threshold to generate a random bit. This approach has been validated on MRAM arrays (*108*), which guarantee high speed, as well as on ferroelectric devices (*109*, *110*) that inherently show better energy efficiency, with

intermediate performance achieved by using PCM (*111*). All realizations showed good resilience to cryptographic attacks, decent temperature stability, and resistance against supply voltage variations. Finally, in (*112*), the authors demonstrated a PUF implementation using a crossbar array of TaO_x-based RRAM devices and combined the inherent variability of the set voltage with a robust ternary and complementary programming paradigm. This approach guarantees resilience to side attacks by means of complementary programming, requiring low voltages (≤2 V), low programming (few picojoules per bit), and low reading energy (<1 pJ/bit) at high speed (few nanoseconds of cycling) and low error rate (≤1 part per million).

Memristive devices for 5G and terahertz switches

Memristive devices exhibiting two stable resistive states may also be used as RF switches—that is, passive critical components needed to route or reconfigure high-frequency signals through communication channels in wireless systems (*15*, *22*, *113*). Modern wireless systems contain a massive number of communication channels over a wide range of frequencies into the terahertz (THz) regime in order to transmit multimedia data at rates of up to 10 Gb/s for 5G networks and 100 Gb/s for the evolving 6G standard (*23*, *114*). Silicon transistors, operating in the ON and OFF electronic states, are currently the main technology used for RF switches, primarily because of their advantages of chip integration and cost (*113*). However, transistors are volatile devices that consume energy both during switching and when idle, and thus offer poor energy efficiency that can substantially reduce the battery life of mobile devices (*15*, *115*). The former is unavoidable as it represents work, but the latter is wasted energy that simply maintains the ON or OFF states.

In addition, future wireless systems such as 6G and beyond will require switches operating at frequencies exceeding 100 GHz. This requirement is a challenge for conventional transistor devices because the ON and OFF states, characterized primarily by ON-resistance (R_{ON}) and OFF-capacitance (C_{OFF}), are directly coupled. The former determines the extent of the signal loss in the ON state (insertion loss), and the latter is responsible for preventing the signal from transmitting in the OFF state (isolation) (Fig. 4A). In high-speed logic devices, dimensional scaling is used to reduce device capacitances in order to achieve higher speeds and frequencies, among other parameters. However, dimensional downscaling, as implemented conventionally, conversely increases the channel resistance. This trade-off constrains transistor high-frequency prospects, defined in terms of the cutoff frequency (F_C) figure-of-merit, $F_C = (2\pi R_{ON} C_{OFF})^{-1}$, a metric particularly useful for benchmarking switch technologies.

For these reasons, there is growing interest in emerging materials that can produce nonvolatile RF switches (*15*, *16*). The energy efficiency of nonvolatile switches can be benchmarked either by direct measurement of the switching energy or indirectly in terms of the recently proposed energy figure of merit, $E_{FOM} = V_{SET} I_{ON} \tau$, where V_{SET}, I_{ON}, and τ are the set voltage, ON current, and switching time, respectively (*23*). Both RRAM and PCM are under investigation for this noncomputing application (*15*, *116*), as they can both transfer the RF signal in LRS (i.e., $R_{LRS} = R_{ON}$) and produce a capacitance effect that blocks it in HRS ($C_{OFF} = C_{HRS}$). However, for RRAM, R_{ON} is typically in the range of kilohms, in part the result of progress in device optimization for low-power storage and computing applications. This ON-state resistance range is too large to meet the <10-ohm requirement for RF switches and would result in undesirably high insertion loss (*15*). Further research is required on metal-oxide RRAM devices to achieve R_{ON} values of <10 ohms.

In contrast, three-terminal PCM RF switches with an integrated heater (Fig. 4B) offer low R_{ON}, modest C_{OFF}, high signal power handling, and endurance in the billions of cycles (*117*), so these could be a practical alternative to CMOS transistors for so-called sub–6 GHz 5G systems. Indeed, GeTe-based lateral PCM RF switches have now been integrated into a BEOL 200-mm (12-inch) wafer foundry manufacturing process (*116*, *117*) and should become increasingly available in integrated chips. Lateral PCM RF switches incur two principal challenges. One is the need for an integrated microheater to trigger the material transition between the amorphous and crystalline phases. The microheater complicates the device BEOL integration (*116*, *118*), and also results in low

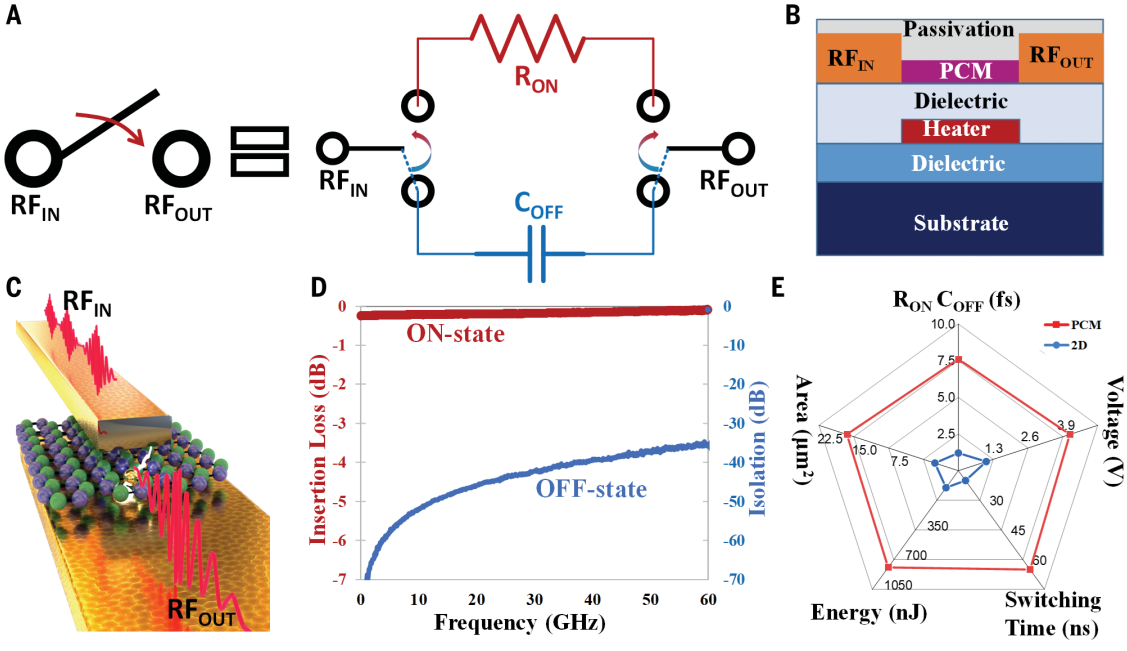


Fig. 4. Memristive radio-frequency switches. (**A**) Schematic symbol of an RF switch, which can be represented by the circuit schematic consisting of R_{ON} and C_{OFF} when the switch is in the ON and OFF states, respectively. (**B** and **C**) Simplified device structures of 3-terminal PCM, and 2D h-BN RF switches. (**D**) Scattering parameters of a high-performance 2D switch based on monolayer h-BN featuring R_{ON} = 2.8 ohm, C_{OFF} = 0.44 fF, F_C = 129 THz. [Adapted from (*23*)] (**E**) A radar chart comparing high-performance contemporary PCM and 2D RF switches along five performance metrics. In this chart, smaller numbers/pentagons represent superior performance. Two additional parameters of great interest are power handling and endurance, with higher values indicative of superior reliability.

switching speeds, typically in the hundreds of nanosecond to microsecond range (*15*, *118*). The other challenge is the direct coupling of R_{ON} and C_{OFF}, an issue that also affects planar transistor switches, hence limiting the operational frequencies to the tens of gigahertz (*15*, *117*). Slow switching speeds and limited-frequency bandwidth present a challenge in using PCM RF switches for millimeter-wave 5G or next-generation 6G systems.

Recently, nonvolatile RF switches based on monolayer crystalline nonmetallic 2D materials such as MoS_2 or h-BN (Fig. 4C) have exhibited high F_C (>100 THz), fast switching time (<1 ns), and low switching voltage (≤1 V) with outstanding insertion loss and isolation (Fig. 4D) (*23*, *115*, *119*) and commensurately high data rates. A defect-assisted virtual conductive point process has been shown to decouple R_{ON} from C_{OFF}, which enables straightforward scaling to higher frequencies while maintaining low insertion loss (*23*). The atomic length scale (<1 nm) of conduction in 2D RF switches affords a relatively flat insertion loss (Fig. 4D), unlike larger micrometer-scale RF switches characterized by noticeable frequency dependence (also known as dispersion) caused by device inductive effects at high frequencies. With regard to energy consumption, 2D RF switches are more than two orders of magnitude more energy-efficient than emerging switches in terms of the energy figure of merit (*23*). These metrics are superior to those of other emerging RF switches based on VO_2 and microelectromechanical systems. Contemporary PCM and 2D devices, the two leading emerging RF switch technologies, are compared in Fig. 4E. An important challenge of 2D materials–based RF switches is their integration on silicon wafers, which often results in large amounts of defects that reduce yield and increase device variability (relative to PCM- and RRAM-based switches), as well as their endurance (hundreds of cycles).

Challenges and prospects

As solid-state memory, PCM is an appealing candidate to be incorporated into the memory bus if its endurance, switching time, and energy are improved from ~10^7 cycles, ~50 to 100 ns, and ~10 pJ to 10^9 cycles, 10 ns, and 1 pJ, respectively. These improvements could be achieved through device engineering such as the use of superlattice chalcogens (e.g., $[(GeTe)_x/(Sb_2Te_3)_y]_N$), substrates with low thermal conductivity and confined geometries (*120*). RRAM may show faster switching speeds and lower energy consumption than PCM, but endurance (~10^6 cycles) and device-to-device and cycle-to-cycle variability are still major obstacles limiting its use. STT-MRAM could replace SRAM or embedded DRAM if read and write challenges can be overcome, such as by increasing switching endurance up to ~10^7 cycles and ensuring reliable state identification. Potential solutions could involve materials that have high energy barriers or otherwise enhance the HRS/LRS resistance ratio, optimization of the sensing amplifier to achieve accurate state distinction, or both (*121*).

Impediments to FTJ commercialization are low CMOS compatibility and poor endurance (~10^6 cycles). The first of these might be mitigated using orthorhombic HfO_2, a material readily used in microelectronics, but this material results in lower HRS/LRS resistance ratios (~2). A FeFET based on HfO_2 has shown higher HRS/LRS resistance ratios of ~10^5 (*122*), but the integration of the ferroelectric material at the gate of the transistor is more complex. A feasible alternative would be the use of CMOS-compatible van der Waals ferroelectric materials such as $CuInP_2S_6$, which may enable this value to increase to >10^7 (*123*). However, when using this material in exploratory studies, it is important to avoid mechanical exfoliation when synthesizing the 2D materials to ensure good scalability. In any case, the maximum endurance demonstrated using this approach is ~10^4 cycles, and further improvements are necessary. The optimization of the metal-insulator interfaces to avoid the presence of traps (which reduce device reliability) is one of the most important factors to consider.

The specific requirements that memristive devices need to fulfill when used for IMC depend highly on the application. However, attributes such as HRS/LRS resistance ratio, endurance, retention, and intrinsic variability are important for most computing applications. It is also beneficial to have a LRS resistance high enough to limit the impact of the voltage drop in the lines of the crossbar array during writing and readout. To make memristive IMC highly competitive against custom digital accelerators and SRAM-based IMC, further improvements in compute density (in excess of 10 TOPS/mm^2) and compute precision (equivalent to four- to five-bit fixed-point arithmetic) are required. To improve the compute density, besides scaling both the memristive devices and the associated access devices, high-density memristive arrays need to be integrated at the back end of a CMOS wafer. However, recent advances in heterogeneous integration such as through-silicon-via or hybrid bonding could open new possibilities whereby the memristive array fabrication could be decoupled from the design of advanced CMOS peripheral circuitry. To improve the compute precision, it is essential to minimize the temporal conductance fluctuations (such as noise and conductance drift), and new device concepts such as projected memory are being explored (*124*).

For memristive devices used in data encryption, the main challenge is to fabricate highly energy-efficient memristive devices capable of few-femtojoule, low-voltage, subnanosecond switching with high switching randomness that also shows extended endurance. A specific challenge for RTN-based solutions is instead related to the stability and magnitude of RTN signals; these could be improved by clever process and device design approaches, such as obtaining confinement of defects in 2D materials (*103*). In memristive PUFs, one challenge is to reduce the large sensitivity to voltage fluctuations and noise that could enable an attacker to hijack the device and force-program unintended malicious fingerprints. Another is to develop complex peripheral circuits that could correctly compensate for the temperature dependence of the switching statistics and thus avoid introducing severe bias in fingerprint generation.

Regarding memristive devices for mobile communication, 2D materials seem to provide good performance, but yield (<75%) and endurance (hundreds of cycles) fall well short of the >99.99% yield and >10^9 cycles needed in practical systems. These problems might be mitigated using multilayer 2D materials, which can have high yield (~98%) and endurance (~80,000 cycles) (*125*). The issue of endurance calls for deeper understanding of the underlying phenomena responsible for the memristive effect in atomically thin crystalline materials, as well as further research into the aging effect of the phenomena. If 2D memristive devices continue to improve, then 2D RF switches with satisfactory endurance could be considered for millimeter-wave and 6G integrated wireless systems. Moreover, in these systems integration density is not a problem, so the use of series transistors to avoid overshoot should also result in better endurance.

Memristive devices began to be commercialized as nonvolatile memories in 2006, but their share of the memory market has only now started to increase rapidly. The commercialization of other memristive products beyond nonvolatile memories may still take a few years. Nonetheless, memristive devices are a reality and we will start to see them more and more in the electronic products that we use daily.

REFERENCES AND NOTES

1. W. Jung, Ed., *Op Amp Applications Handbook* (Elsevier, 2005). doi: 10.1016/B978-0-7506-7844-5.X5109-1
2. S. Mittal, G. Verma, B. Kaushik, F. A. Khanday, A survey of SRAM-based in-memory computing techniques and applications. *J. Systems Archit.* **119**, 102276 (2021). doi: 10.1016/j.sysarc.2021.102276
3. J. Kim, Y. V. Pershin, M. Yin, T. Datta, M. Di Ventra, An Experimental Proof that Resistance-Switching Memory Cells Are Not Memristors. *Adv. Electron. Mater.* **6**, 2000010 (2020). doi: 10.1002/aelm.202000010
4. D. Ielmini, R. Waser, *Resistive Switching: From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications* (Wiley, 2015).
5. L. O. Chua, Memristor—The missing circuit element. *IEEE Trans. Circuit Theory* **CT-18**, 50 (1971).

6. B. Govoreanu *et al*., 10×10nm^2 Hf/HfO$_x$ crossbar resistive RAM with excellent performance, reliability and low-energy operation. In *2011 IEEE International Electron Devices Meeting* (2011). doi: 10.1109/IEDM.2011.6131652
7. D. Lammers, "MRAM debut cues memory transition." *EE Times* (7 October 2006); www.eetimes.com/mram-debut-cues-memory-transition/.
8. "The new microcontrollers with on-chip non-volatile memory ReRAM" [press release]. Panasonic (15 May 2012).
9. D. C. Kau *et al*., A stackable cross point phase change memory. In *2009 IEEE International Electron Devices Meeting (IEDM)* (2009). doi: 10.1109/IEDM.2009.5424263
10. "Intel and Micron Produce Breakthrough Memory Technology" (28 July 2015); https://newsroom.intel.com/news-releases/intel-and-micron-produce-breakthrough-memory-technology/#gs.vtoa8u.
11. D. Ielmini, H.-S. P. Wong, In-memory computing with resistive switching devices. *Nat. Electron.* **1**, 333–343 (2018). doi: 10.1038/s41928-018-0092-2
12. A. Sebastian, M. Le Gallo, R. Khaddam-Aljameh, E. Eleftheriou, Memory devices and applications for in-memory computing. *Nat. Nanotechnol.* **15**, 529–544 (2020). doi: 10.1038/s41565-020-0655-z; pmid: 32231270
13. B. Lin *et al*., A High-Speed and High-Reliability TRNG Based on Analog RRAM for IoT Security Application. In *2019 IEEE International Electron Devices Meeting (IEDM)* (2019). doi: 10.1109/IEDM19573.2019.8993486
14. S. Balatti *et al*., Physical unbiased generation of random numbers with coupled resistive switching devices. *IEEE Trans. Electron Dev.* **63**, 2029–2035 (2016). doi: 10.1109/TED.2016.2537792
15. N. Wainstein, G. Adam, E. Yalon, S. Kvatinsky, RadioFrequency Switches Based on Emerging Resistive Memory Technologies: A Survey. *Proc. IEEE* **109**, 77–95 (2021). doi: 10.1109/JPROC.2020.3011953
16. S. Pi, M. Ghadiri-Sadrabadi, J. C. Bardin, Q. Xia, Nanoscale memristive radiofrequency switches. *Nat. Commun.* **6**, 7519 (2015). doi: 10.1038/ncomms8519; pmid: 26108890
17. M. Lanza *et al*., Recommended Methods to Study Resistive Switching Devices. *Adv. Electron. Mater.* **5**, 1800143 (2018). doi: 10.1002/aelm.201800143
18. C. H. Sie, thesis, Iowa State University (1969). doi: 10.31274/rtd-180813-1655
19. S. Bhatti *et al*., Spintronics based random access memory: A review. *Mater. Today* **20**, 530–548 (2017). doi: 10.1016/j.mattod.2017.07.007
20. V. Garcia, M. Bibes, Ferroelectric tunnel junctions for information storage and processing. *Nat. Commun.* **5**, 4289 (2014). doi: 10.1038/ncomms5289; pmid: 25056141
21. Y. Yang *et al*., Probing electrochemistry at the nanoscale: In situ TEM and STM characterizations of conducting filaments in memristive devices. *J. Electroceram.* **39**, 73–93 (2017). doi: 10.1007/s10832-017-0069-y
22. International Roadmap for Devices and Systems (2020); https://irds.ieee.org/.
23. M. Kim *et al*., Analogue switches made from boron nitride monolayers for application in 5G and terahertz communication systems. *Nat. Electron.* **3**, 479–485 (2020). doi: 10.1038/s41928-020-0416-x
24. B. Pangrle, A node by any other name. *Semiconductor Engineering* (2014); https://semiengineering.com/a-node-by-any-other-name/.
25. S. Pi *et al*., Memristor crossbar arrays with 6-nm half-pitch and 2-nm critical dimension. *Nat. Nanotechnol.* **14**, 35–39 (2019). doi: 10.1038/s41565-018-0302-0; pmid: 30420759
26. S. M. Hus *et al*., Observation of single-defect memristor in an MoS$_2$ atomic sheet. *Nat. Nanotechnol.* **16**, 58–62 (2021). doi: 10.1038/s41565-020-00789-w
27. V. K. Sangwan *et al*., Multi-Terminal Memtransistors from Polycrystalline Monolayer MoS$_2$. *Nature* **554**, 500–504 (2018). doi: 10.1038/nature25747; pmid: 29469093
28. "Emerging non-volatile memory" [market analysis report]. Yole Development (2021); www.i-micronews.com/products/emerging-non-volatile-memory-2021/.
29. K. Vättö, I. Cutress, R. Smith, "Analyzing Intel-Micron 3D XPoint: The next generation non-volatile memory." Anandtech (31 July 2015).
30. "STMicroelectronics Now Sampling Embedded PCM for Automotive Microcontrollers" (10 December 2018).
31. S. Yu, P.-Y. Chen, Emerging memory technologies: Recent trends and prospects. *IEEE Solid-State Circuits Mag.* **8**, 43–56 (2016). doi: 10.1109/MSSC.2016.2546199
32. "Non-volatile Memory with very small operating current: ReRAM"; www.fujitsu.com/jp/group/fsm/en/products/reram/.
33. T.-y. Liu *et al*., A 130.7-mm^2 2-Layer 32-Gb ReRAM Memory Device in 24-nm Technology. *IEEE J. Solid-State Circuits* **49**, 140–153 (2013). doi: 10.1109/JSSC.2013.2280296
34. O. Golonzka *et al*., Non-Volatile RRAM Embedded into 22FFL FinFET Technology. In *2019 Symposium on VLSI Technology* (2019). doi: 10.23919/VLSIT.2019.8776570
35. Y.-C. Chiu *et al*., A 40nm 2Mb ReRAM Macro with 85% Reduction in FORMING Time and 99% Reduction in Page-Write Time Using Auto-FORMING and Auto-Write Schemes. In *2019 Symposium on VLSI Technology* (2019). doi: 10.23919/VLSIT.2019.8776540
36. Taiwan Semiconductor Manufacturing Company, Memory research portal, RRAM section; https://research.tsmc.com/english/research/memory/rram/publish-time-1.html.
37. Q. Luo *et al*., 8-Layers 3D vertical RRAM with excellent scalability towards storage class memory applications. In *2017 IEEE International Electron Devices Meeting (IEDM)* (2017). doi: 10.1109/IEDM.2017.8268315
38. A. C. Torrezan, J. P. Strachan, G. Medeiros-Ribeiro, R. S. Williams, Sub-nanosecond switching of a tantalum oxide memristor. *Nanotechnology* **22**, 485203 (2011). doi: 10.1088/0957-4484/22/48/485203; pmid: 22071289
39. S. H. Lee *et al*., Quantitative, Dynamic TaO$_x$ Memristor/Resistive Random Access Memory Model. *ACS Appl. Electron. Mater.* **2**, 701–709 (2020). doi: 10.1021/acsaelm.9b00792
40. M. Lanza *et al*., Standards for the Characterization of Endurance in Resistive Switching Devices. *ACS Nano* **15**, 17214–17231 (2021). doi: 10.1021/acsnano.1c06980
41. J.-M. Hung *et al*., A four-megabit compute-in-memory macro with eight-bit precision based on CMOS and resistive random-access memory for AI edge devices. *Nat. Electron.* **4**, 921–930 (2021). doi: 10.1038/s41928-021-00676-9
42. C.-C. Chou *et al*., A 22nm 96KX144 RRAM Macro with a Self-Tracking Reference and a Low Ripple Charge Pump to Achieve a Configurable Read Window and a Wide Operating Voltage Range. In *2020 IEEE Symposium on VLSI Circuits* (2020).
43. Q. Liu *et al*., Controllable growth of nanoscale conductive filaments in solid-electrolyte-based ReRAM by using a metal nanocrystal covered bottom electrode. *ACS Nano* **4**, 6162–6168 (2010). doi: 10.1021/nn1017582; pmid: 20853865
44. S. Choi *et al*., SiGe epitaxial memory for neuromorphic computing with reproducible high performance based on engineered dislocations. *Nat. Mater.* **17**, 335–340 (2018). doi: 10.1038/s41563-017-0001-5; pmid: 29358642
45. W. J. Gallagher *et al*., 22nm STT-MRAM for Reflow and Automotive Uses with High Yield, Reliability, and Magnetic Immunity and with Performance and Shielding Options. In *2019 IEEE International Electron Devices Meeting (IEDM)* (2019). doi: 10.1109/IEDM19573.2019.8993469
46. K. Lee *et al*., 1Gbit High Density Embedded STT-MRAM in 28nm FDSOI Technology. In *2019 IEEE International Electron Devices Meeting (IEDM)* (2019). doi: 10.1109/IEDM19573.2019.8993551
47. D. Edelstein *et al*., A 14 nm Embedded STT-MRAM CMOS Technology. In *2020 IEEE International Electron Devices Meeting (IEDM)* (2020). doi: 10.1109/IEDM13553.2020.9371922
48. L. Chen *et al*., Ultra-low power Hf$_{0.5}$Zr$_{0.5}$O$_2$ based ferroelectric tunnel junction synapses for hardware neural network applications. *Nanoscale* **10**, 15826–15833 (2018). doi: 10.1039/C8NR04734K; pmid: 30105324
49. C. Ma *et al*., Sub-nanosecond memristor based on ferroelectric tunnel junction. *Nat. Commun.* **11**, 1439 (2020). doi: 10.1038/s41467-020-15249-1; pmid: 32188861
50. Y. Goh, S. Jeon, The effect of the bottom electrode on ferroelectric tunnel junctions based on CMOS-compatible HfO$_2$. *Nanotechnology* **29**, 335201 (2018). doi: 10.1088/1361-6528/aac6b3; pmid: 29786620
51. M. Trentzsch *et al*., A 28 nm HKMG super low power embedded NVM technology based on ferroelectric FETs. In *IEDM Tech. Dig.* (December 2016). doi: 10.1109/IEDM.2016.7838397
52. E. T. Breyer, H. Mulaosmanovic, T. Mikolajick, S. Slesazeck, Reconfigurable NAND/NOR logic gates in 28 nm HKMG and 22 nm FD-SOI FeFET technology. In *IEDM Tech. Dig.* (December 2017). doi: 10.1109/IEDM.2017.8268471
53. K. Zhu *et al*., The development of integrated circuits based on two-dimensional materials. *Nat. Electron.* **4**, 775–785 (2021). doi: 10.1038/s41928-021-00672-z
54. A. Shafiee *et al*., ISAAC: A convolutional neural network accelerator with in-situ analog arithmetic in crossbars. *ACM SIGARCH Comput. Architect. News* **44**, 14–26 (2016). doi: 10.1145/3007787.3001139
55. S. Oh *et al*., Energy-efficient Mott activation neuron for full-hardware implementation of neural networks. *Nat. Nanotechnol.* **16**, 680–687 (2021). doi: 10.1038/s41565-021-00874-8; pmid: 33737724
56. C.-X. Xue *et al*., A 22nm 4Mb 8b-Precision ReRAM Computing-in-Memory Macro with 11.91-195.7 TOPS/W for Tiny AI Edge Devices. In *IEEE International Solid-State Circuits Conference (ISSCC)* (2021). doi: 10.1109/ISSCC42613.2021.9365769
57. R. Khaddam-Aljameh *et al*., HERMES Core – A 14nm CMOS and PCM-based In-Memory Compute Core using an array of 300ps/LSB Linearized CCO-based ADCs and local digital processing. In *IEEE Symposium on VLSI Technology* (2021).
58. P. Narayanan *et al*., Fully on-chip MAC at 14nm enabled by accurate row-wise programming of PCM-based weights and parallel vector-transport in duration-format. In *IEEE Symposium on VLSI Technology* (2021).
59. P. Deaville *et al*., A maximally row-parallel MRAM in-memory-computing macro addressing readout circuit sensitivity and area. In *European Solid-state Devices and Circuits Conference* (2021). doi: 10.1109/ESSCIRC53450.2021.9567807
60. S. Jung *et al*., A crossbar array of magnetoresistive memory devices for in-memory computing. *Nature* **601**, 211–216 (2022). doi: 10.1038/s41586-021-04196-6; pmid: 35022590
61. V. Joshi *et al*., Accurate deep neural network inference using computational phase-change memory. *Nat. Commun.* **11**, 2473 (2020). doi: 10.1038/s41467-020-16108-9; pmid: 32424184
62. S. Yu, Neuro-inspired computing with emerging nonvolatile memory. *Proc. IEEE* **106**, 260–285 (2018). doi: 10.1109/JPROC.2018.2790840
63. M. Dazzi *et al*., Efficient pipelined execution of CNNs based on in-memory computing and graph homomorphism verification. *IEEE Trans. Comput.* **70**, 922–935 (2021). doi: 10.1109/TC.2021.3073255
64. M. Ishii *et al*., On-Chip Trainable 1.4M 6T2R PCM Synaptic Array with 1.6K Stochastic LIF Neurons for Spiking RBM. In *2019 IEEE International Electron Devices Meeting* (2019).
65. A. Serb *et al*., Unsupervised learning in probabilistic neural networks with multi-state metal-oxide memristive synapses. *Nat. Commun.* **7**, 12611 (2016). doi: 10.1038/ncomms12611; pmid: 27681181
66. P. U. Diehl *et al*., Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. In *International Joint Conference on Neural Networks (IJCNN)* (2015). doi: 10.1109/IJCNN.2015.7280696
67. A. Valentian *et al*., Fully integrated spiking neural network with analog neurons and RRAM synapses. In *2019 IEEE International Electron Devices Meeting (IEDM)* (2019). doi: 10.1109/IEDM19573.2019.8993431
68. M. Pfeiffer, T. Pfeil, Deep learning with spiking neurons: Opportunities and challenges. *Front. Neurosci.* **12**, 774 (2018). doi: 10.3389/fnins.2018.00774; pmid: 30410432
69. C. Li *et al*., Analog content-addressable memories with memristors. *Nat. Commun.* **11**, 1638 (2020). doi: 10.1038/s41467-020-15254-4; pmid: 32242006
70. G. Karunaratne *et al*., Robust high-dimensional memory-augmented neural networks. *Nat. Commun.* **12**, 2468 (2021). doi: 10.1038/s41467-021-22364-0; pmid: 33927202
71. G. Karunaratne *et al*., In-memory hyperdimensional computing. *Nat. Electron.* **3**, 327–337 (2020). doi: 10.1038/s41928-020-0410-3
72. K. Ni *et al*., Ferroelectric ternary content-addressable memory for one-shot learning. *Nat. Electron.* **2**, 521–529 (2019). doi: 10.1038/s41928-019-0321-3
73. C. E. Graves *et al*., In-memory computing with memristor content addressable memories for pattern matching. *Adv. Mater.* **32**, e2003437 (2020). doi: 10.1002/adma.202003437; pmid: 32761709
74. M. Le Gallo *et al*., Mixed-precision in-memory computing. *Nat. Electron.* **1**, 246–253 (2018). doi: 10.1038/s41928-018-0054-8
75. Z. Sun, G. Pedretti, A. Bricalli, D. Ielmini, One-step regression and classification with cross-point resistive memory arrays. *Sci. Adv.* **6**, eaay2378 (2020). doi: 10.1126/sciadv.aay2378; pmid: 32064342

76. M. A. Zidan *et al*., A general memristor-based partial differential equation solver. *Nat. Electron.* **1**, 411–420 (2018). doi: 10.1038/s41928-018-0100-6
77. B. Feinberg *et al*., Enabling scientific computing on memristive accelerators. In *2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)* (2018). doi: 10.1109/ISCA.2018.00039
78. M. N. Bojnordi, E. Ipek, Memristive Boltzmann machine: A hardware accelerator for combinatorial optimization and deep learning. In *IEEE International Symposium on High Performance Computer Architecture (HPCA)* (2016). doi: 10.1109/HPCA.2016.7446049
79. M. R. Mahmoodi, M. Prezioso, D. B. Strukov, Versatile stochastic dot product circuits based on nonvolatile memories for high performance neurocomputing and neurooptimization. *Nat. Commun.* **10**, 5113 (2019). doi: 10.1038/s41467-019-13103-7; pmid: 31704925
80. F. Cai *et al*., Power-efficient combinatorial optimization using intrinsic noise in memristor hopfield neural networks. *Nat. Electron.* **3**, 409–418 (2020). doi: 10.1038/s41928-020-0436-6
81. H. Mostafa, L. K. Müller, G. Indiveri, An event-based architecture for solving constraint satisfaction problems. *Nat. Commun.* **6**, 8941 (2015). doi: 10.1038/ncomms9941; pmid: 26642827
82. S. Kumar, J. P. Strachan, R. S. Williams, Chaotic dynamics in nanoscale NbO_2 Mott memristors for analogue computing. *Nature* **548**, 318–321 (2017). doi: 10.1038/nature23307; pmid: 28792931
83. J. Moon *et al*., Temporal data classification and forecasting using a memristor-based reservoir computing system. *Nat. Electron.* **2**, 480–487 (2019). doi: 10.1038/s41928-019-0313-3
84. A. Sebastian *et al*., Temporal correlation detection using computational phase-change memory. *Nat. Commun.* **8**, 1115 (2017). doi: 10.1038/s41467-017-01481-9; pmid: 29062022
85. X. Zhu, Q. Wang, W. D. Lu, Memristor networks for real-time neural activity analysis. *Nat. Commun.* **11**, 2439 (2020). doi: 10.1038/s41467-020-16261-1; pmid: 32415218
86. H. Jia *et al*., A programmable neural-network inference accelerator based on scalable in-memory computing. In *IEEE International Solid-State Circuits Conference (ISSCC)* (2021). doi: 10.1109/ISSCC42613.2021.9365788
87. B. Murmann, Mixed-signal computing for deep neural network inference. In *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* (2021).
88. J. Hartmann *et al*., Artificial intelligence: Why moving it to the edge? In *ESSCIRC 2021 - IEEE 47th European Solid State Circuits Conference* (2021). doi: 10.1109/ESSCIRC53450.2021.9567817
89. B. Schneier, *Applied Cryptography* (Wiley, 2015).
90. C. S. Petrie, J. A. Connelly, A noise-based IC random number generator for applications in cryptography. *IEEE Trans. Circ. Syst. I* **47**, 615–621 (2000). doi: 10.1109/81.847868
91. D. Liu, Z. Liu, L. Li, X. Zou, A Low-Cost Low-Power Ring Oscillator-Based Truly Random Number Generator for Encryption on Smart Cards. *IEEE Trans. Circ. Syst. II* **63**, 608–612 (2016). doi: 10.1109/TCSII.2016.2530800
92. S. K. Mathew *et al*., μRNG: A 300–950 mV, 323 Gbps/W All-Digital Full-Entropy True Random Number Generator in 14 nm FinFET CMOS. *IEEE J. Solid-State Circuits* **51**, 1695–1704 (2016). doi: 10.1109/JSSC.2016.2558490
93. M. Soucarros, J. Clediere, C. Dumas, P. Elbaz-Vincent, Fault analysis and evaluation of a true random number generator embedded in a processor. *J. Electron. Test.* **29**, 367–381 (2013). doi: 10.1007/s10836-013-5356-1
94. N. Nguyen, G. Kaddoum, F. Pareschi, R. Rovatti, G. Setti, A fully CMOS true random number generator based on hidden attractor hyperchaotic system. *Nonlinear Dyn.* **102**, 2887–2904 (2020). doi: 10.1007/s11071-020-06017-3
95. J.-C. Hsueh, V. H.-C. Chen, An ultra-low voltage chaos-based true random number generator for IoT applications. *Microelectronics* **87**, 55–64 (2019). doi: 10.1016/j.mejo.2019.03.013
96. S. K. Mathew *et al*., 2.4 Gbps, 7 mW all-digital PVT-variation tolerant true random number generator for 45 nm CMOS high-performance microprocessors. *IEEE J. Solid-State Circuits* **47**, 2807–2821 (2012). doi: 10.1109/JSSC.2012.2217631
97. M. I. Rashid *et al*., True Random Number Generation Using Latency Variations of FRAM. In *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* (2021). doi: 10.1109/TVLSI.2020.3018998
98. H. Mulaosmanovic, T. Mikolajick, S. Slesazeck, Random Number Generation Based on Ferroelectric Switching. *IEEE Electron Device Lett.* **39**, 135–138 (2018). doi: 10.1109/LED.2017.2771818
99. E. Piccinini, R. Brunetti, M. Rudan, Self-Heating Phase-Change Memory-Array Demonstrator for True Random Number Generation. *IEEE Trans. Electron Dev.* **64**, 2185–2192 (2017). doi: 10.1109/TED.2017.2673867
100. A. Fukushima *et al*., Spin dice: A scalable truly random number generator based on spintronics. *Appl. Phys. Express* **7**, 083001 (2014). doi: 10.7567/APEX.7.083001
101. K. Yang *et al*., A 28NM Integrated True Random Number Generator Harvesting Entropy from MRAM. In *2018 IEEE Symposium on VLSI Circuits* (2018). doi: 10.1109/VLSIC.2018.8502431
102. H. Jiang *et al*., A novel true random number generator based on a stochastic diffusive memristor. *Nat. Commun.* **8**, 882 (2017). doi: 10.1038/s41467-017-00869-x; pmid: 29026110
103. C. Wen *et al*., Advanced Data Encryption using 2D Materials. *Adv. Mater.* **33**, e2100185 (2021). doi: 10.1002/adma.202100185; pmid: 34046938
104. F. M. Puglisi, P. Pavan, Guidelines for a Reliable Analysis of Random Telegraph Noise in Electronic Devices. *IEEE Trans. Instrum. Meas.* **65**, 1435–1442 (2016). doi: 10.1109/TIM.2016.2518880
105. R. Brederlow *et al*., A low-power true random number generator using random telegraph noise of single-oxide-traps. In *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)* (2006). doi: 10.1109/ISSCC.2006.1696222
106. S. S. Mansouri, E. Dubrova, Ring oscillator physical unclonable function with multi level supply voltages. In *2012 IEEE 30th International Conference on Computer Design (ICCD)* (2012). doi: 10.1109/ICCD.2012.6378703
107. D. E. Holcomb, W. Burleson, K. Fu, Initial SRAM state as a fingerprint and source of true random numbers for RFID tags. In *Proceedings of the Conference on RFID Security* (2007).
108. A. Kumar, S. Sahay, M. Suri, Switching-Time Dependent PUF Using STT- MRAM. In *2018 31st International Conference on VLSI Design and 2018 17th International Conference on Embedded Systems (VLSID)* (2018). doi: 10.1109/VLSID.2018.103
109. J. Yu *et al*., A novel physical unclonable function (PUF) using 16 × 16 pure-HfOx ferroelectric tunnel junction array for security applications. *Nanotechnology* **32**, 485202 (2021). doi: 10.1088/1361-6528/ac1dd5
110. S. Kim *et al*., Physical Unclonable Functions Using Ferroelectric Tunnel Junctions. *IEEE Electron Device Lett.* **42**, 816–819 (2021). doi: 10.1109/LED.2021.3075427
111. N. Noor, H. Silva, Phase Change Memory for Physical Unclonable Functions. In *Applications of Emerging Memory Technology*, M. Suri, Ed. (Springer, 2020). doi: 10.1007/978-981-13-8379-3_3
112. B. Cambou, M. Orlowski, PUF designed with resistive RAM and ternary states. In *Proceedings of the 11th Annual Cyber and Information Security Research Conference* (2016). doi: 10.1145/2897795.2897808
113. B. Yu *et al*., Ultra-Wideband Low-Loss Switch Design in High-Resistivity Trap-Rich SOI With Enhanced Channel Mobility. *IEEE Trans. Microw. Theory Tech.* **65**, 3937–3949 (2017). doi: 10.1109/TMTT.2017.2696944
114. V. Petrov, T. Kurner, I. Hosako, IEEE 802.15.3d: First Standardization Efforts for Sub-Terahertz Band Communications toward 6G. *IEEE Commun. Mag.* **58**, 28–33 (2020). doi: 10.1109/MCOM.001.2000273
115. M. Kim *et al*., Zero-static power radio-frequency switches based on MoS_2 atomristors. *Nat. Commun.* **9**, 2524 (2018). doi: 10.1038/s41467-018-04934-x; pmid: 29955064
116. G. Slovin *et al*., Monolithic Integration of Phase-Change RF Switches in a Production SiGe BiCMOS Process with RF Circuit Demonstrations. In *2020 IEEE/MTT-S International Microwave Symposium (IMS)* (2020). doi: 10.1109/IMS30576.2020.9223824
117. N. El-Hinnawy, G. Slovin, J. Rose, D. Howard, A 25 THz FCO (6.3 fs $R_{ON}C_{OFF}$) Phase-Change Material RF Switch Fabricated in a High Volume Manufacturing Environment with Demonstrated Cycling > 1 Billion Times. In *2020 IEEE/MTT-S International Microwave Symposium (IMS)* (2020). doi: 10.1109/IMS30576.2020.9223973
118. N. El-Hinnawy *et al*., Experimental Demonstration of AlN Heat Spreaders for the Monolithic Integration of Inline Phase-Change Switches. *IEEE Electron. Device Lett.* **39**, 610–613 (2018). doi: 10.1109/LED.2018.2806383
119. R. Ge *et al*., Atomristor: Nonvolatile Resistance Switching in Atomic Sheets of Transition Metal Dichalcogenides. *Nano Lett.* **18**, 434–441 (2018). doi: 10.1021/acs.nanolett.7b04342; pmid: 29236504
120. A. I. Khan *et al*., Ultralow-switching current density multilevel phase-change memory on a flexible substrate. *Science* **373**, 1243–1247 (2021). doi: 10.1126/science.abj1261; pmid: 34516795
121. S. M. Alam, "STT-MRAM Fundamentals, Challenges, and Applications" (webinar, Santa Clara Valley IEEE Magnetics Society, 1 December 2020).
122. M. Saitoh *et al*., HfO_2-based FeFET and FTJ for Ferroelectric-Memory Centric 3D LSI towards Low-Power and High-Density Storage and AI Applications. In *2020 IEEE International Electron Devices Meeting (IEDM)* (2020). doi: 10.1109/LED.2017.2771818
123. J. Wu *et al*., High tunnelling electroresistance in a ferroelectric van der Waals heterojunction via giant barrier height modulation. *Nat. Electron.* **3**, 466–472 (2020). doi: 10.1038/s41928-020-0441-9
124. I. Giannopoulos *et al*., 8-bit precision in-memory multiplication with projected phase-change memory. In *2018 IEEE International Electron Devices Meeting (IEDM)* (2018). doi: 10.1109/IEDM.2018.8614558
125. S. Chen *et al*., Wafer-scale integration of 2D materials in high-density memristive crossbar arrays for artificial neural networks. *Nat. Electron.* **3**, 638–645 (2020). doi: 10.1038/s41928-020-00473-w
126. M. Binggeli, VLSI Design Layout and Simulation of a 6T SRAM Cell, Course EE 4432: Introduction to VLSI Systems, Idaho State University; https://docplayer.net/25821420-Ee-4432-vlsi-design-layout-and-simulation-of-a-6t-sram-cell.html.
127. W.-H. Chen *et al*., A 65nm 1Mb Nonvolatile Computing-in-Memory ReRAM Macro with sub-16ns Multiply-and-Accumulate for Binary DNN AI Edge Processors. In *IEEE International Solid-State Circuits Conference (ISSCC)* (2018). doi: 10.1109/ISSCC.2018.8310400
128. R. Mochida *et al*., A 4M synapses integrated analog ReRAM based 66.5 TOPS/W neural-network processor with cell current controlled writing and flexible network architecture. In *IEEE Symposium on VLSI Technology* (2018). doi: 10.1109/VLSIT.2018.8510676
129. C. Xue *et al*., A 1Mb multibit ReRAM computing-in-memory macro with 14.6 ns parallel MAC computing time for CNN-based AI edge processors. In *IEEE International Solid-State Circuits Conference (ISSCC)* (2019). doi: 10.1109/ISSCC.2019.8662395
130. C. Xue *et al*., A 22nm 2Mb ReRAM compute-in-memory macro with 121-28TOPS/W for multibit MAC computing for Tiny AI Edge Devices. In *IEEE International Solid-State Circuits Conference (ISSCC)* (2020). doi: 10.1109/ISSCC19947.2020.9063078
131. P. Yao *et al*., Fully hardware-implemented memristor convolutional neural network. *Nature* **577**, 641–646 (2020). doi: 10.1038/s41586-020-1942-4; pmid: 31996818
132. J. Yoon *et al*., A 40nm 64Kb 56.67TOPS/W Read-Disturb-Tolerant Compute-in-Memory/Digital RRAM Macro with Active-Feedback-Based Read and In-Situ Write Verification. In *IEEE International Solid-State Circuits Conference (ISSCC)* (2021). doi: 10.1109/ISSCC42613.2021.9365926
133. W.-S. Khwa *et al*., A 40-nm, 2M-Cell, 8b-Precision, Hybrid SLC-MLC PCM Computing-in-Memory Macro with 20.5 - 65.0TOPS/W for Tiny-AI Edge Devices. In *2022 IEEE International Solid-State Circuits Conference (ISSCC)* (2022).
134. P.-C. Wu *et al*., A 28nm 1Mb Time-Domain Computing-in-Memory 6T-SRAM Macro with a 6.6ns Latency, 1241GOPS and 37.01TOPS/W for 8b-MAC Operations for Edge-AI Devices. *In 2022 IEEE International Solid-State Circuits Conference (ISSCC)*(2022).
135. B. Zimmer *et al*., A 0.32–128 TOPS, scalable multi-chip-module-based deep neural network inference accelerator with ground-referenced signaling in 16 nm. *IEEE J. Solid-State Circuits* **55**, 920–932 (2020). doi: 10.1109/JSSC.2019.2960488

ACKNOWLEDGMENTS

M.L. acknowledges S. Bertolazzi from Yole Développement for advice on memory market trends, S. Pazos from the King Abdullah University of Science and Technology for useful

discussions on device and system technology, and E. Sahagun from Scixel for graphics design. **Funding:** M.L. acknowledges support from the King Abdullah University of Science and Technology. A.S. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement numbers 682675 and 966764). D.A. acknowledges funding from Office of Naval Research grant N00014-20-1-2104 and Air Force Research Laboratory award FA9550-21-1-0460. **Author contributions:** M.La., A.S., W.L., M.L.G., M.-F.C., D.A., F.M.P., H.A., and J.B.R. wrote the manuscript and prepared the figures. M.Li. revised and edited the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Not applicable.

Submitted 30 September 2021; accepted 25 April 2022
10.1126/science.abj9979

Science

Memristive technologies for data storage, computation, encryption, and radio-frequency communication

Mario LanzaAbu SebastianWei D. LuManuel Le GalloMeng-Fan ChangDeji AkinwandeFrancesco M. PuglisiHusam N. AlshareefMing LiuJuan B. Roldan

Science, 376 (6597), eabj9979. • DOI: 10.1126/science.abj9979

Putting memristors to work

Memristors, which are resistors that change conductivity and act as memories, are not only being used in commercial computing but have several application areas in computing and communications. Lanza *et al*. review how devices such as phase-change memories, resistive random-access memories, and magnetoresistive random-access memories are being integrated into silicon electronics. Memristors also are finding use in artificial intelligence when integrated in three-dimensional crossbar arrays for low-power, non–von Neuman architectures. Other applications include random-number generation for data encryption and radiofrequency switches for mobile communications. —PDS

Science (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.