ADVANCED INTELLIGENT SYSTEMS
Open Access
www.advintellsyst.com

Memristive, Spintronic, and 2D-Materials-Based Devices to Improve and Complement Computing Hardware

*Dovydas Joksas, AbdulAziz AlMutairi, Oscar Lee, Murat Cubukcu, Antonio Lombardo, Hidekazu Kurebayashi, Anthony J. Kenyon, and Adnan Mehonic**

In a data-driven economy, virtually all industries benefit from advances in information technology—powerful computing systems are critically important for rapid technological progress. However, this progress might be at risk of slowing down if the discrepancy between the current computing power demands and what the existing technologies can offer is not addressed. Key limitations to improving energy efficiency are the excessive growth of data transfer costs associated with the von Neumann architecture and the fundamental limits of complementary metal–oxide–semiconductor (CMOS) technologies, such as transistors. Herein, three approaches that will likely play an essential role in future computing systems are discussed: memristive electronics, spintronics, and electronics based on 2D materials. The authors present how these technologies may transform conventional digital computers and contribute to the adoption of new paradigms, like neuromorphic computing.

1. Introduction

Computers have become an integral part of the modern world. Technologies from instant messaging to searches on the Internet to smart assistants are enabled by devices that perform logical operations and store information over time. With such an explosion of uses, it is not surprising that energy costs have been increasing too—some estimate that information and communications technology could constitute from 8% to 21% of the global electricity demand by the end of the decade.[1] Of course, some applications may contribute to this more than others.

Most notably, artificial intelligence (AI) and machine learning (ML) have become indispensable in a wide range of rapidly growing data-centric technologies, including the Internet of things (IoT), transport, medicine, security, and entertainment. It is now recognized that AI might have a hardware problem[2] associated with huge computational demands, which are directly reflected in the energy consumption. This is not sustainable and is rapidly becoming a critical societal challenge. The soaring demand for computing power in ML vastly outpaces improvements made through Moore's scaling or innovative architectural solutions. From 2012 to 2020, hardware performance of state-of-the-art AI has improved by a factor of 317;[3] this is not enough to meet the growing computing demands of AI applications. The size of state-of-the-art AI models has been increasing exponentially, as have their training costs—from a few dollars in 2012 to millions of dollars in 2020.[4] A pressing need to develop novel technologies to address this issue at the fundamental level and build efficient AI systems has recently become acute. More fundamentally, there is a great need for low-energy computing elements, including those based on different physical principles than complementary metal–oxide–semiconductor (CMOS) transistors implementing Boolean logic.

This perspective article will discuss memristors, spintronics, and two-dimensional (2D) materials and devices, explaining how they can both improve current computing hardware and enable new computing paradigms. We will present the main physical principles and the promise of these technologies, as well as some materials and engineering challenges that must be addressed before full adoption. The role of these emerging technologies will be discussed in the context of both conventional computing, which is based on digital electronics and Boolean algebra, and promising new approaches like neuromorphic computing. This is by no means an exhaustive review and does not imply that other technologies and approaches are not going to play an important role; many alternatives will likely complement the systems we discuss here. Furthermore, the three approaches

D. Joksas, A. Lombardo, A. J. Kenyon, A. Mehonic
Department of Electronic and Electrical Engineering
University College London
Torrington Place, London WC1E 7JE, UK
E-mail: adnan.mehonic.09@ucl.ac.uk

A. A. AlMutairi
Department of Engineering
University of Cambridge
9 JJ Thomson Avenue, Cambridge CB3 0FA, UK

O. Lee, M. Cubukcu, A. Lombardo, H. Kurebayashi
London Centre for Nanotechnology
University College London
Torrington Place, London WC1E 7JE, UK

M. Cubukcu
National Physical Laboratory
Hampton Road, Teddington TW11 0LW, UK

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aisy.202200068.

DOI: 10.1002/aisy.202200068

we present often overlap—at the extreme, we might even have spintronic memristors partially based on 2D materials.[5]

1.1. Basic Principles

1.1.1. Memristors

Memristor was formalized as a circuit element in 1971[6]—an electrical property, called memristance, relating electric charge and magnetic flux was introduced. Memristor's existence was motivated by the fact that this relation filled a gap in fundamental symmetries observed in circuit theory. Since late 2000s, there has been a rebirth of interest in memristors, followed by various physical implementations. The landscape of memristive technologies and the underpinning physical mechanisms is vast and still rapidly expanding.[7]

Memristors, in most cases, are based on the concept of resistance switching. Resistance switching is a reversible process where a memristor changes its resistance with externally applied electrical stimuli. In most cases, resistance switching results in nonvolatile states with long retention times even after the stimuli are removed—the memristive device "memorizes" the resistance state. However, resistance switching can also be achieved by other types of stimuli (e.g. optical) and could lead to volatile switching, which benefits particular applications (e.g. neuronal spiking).

There exist many memristive technologies, but most rely on similar physical principles. Three examples of such technologies—redox-based resistive random-access memory (ReRAM), phase-change memory (PCM), and magnetoresistive random-access memory (MRAM)—are shown in **Figure 1**a–c. Memristors are typically implemented as simple two-terminal capacitor-like structures, where a switching layer is sandwiched

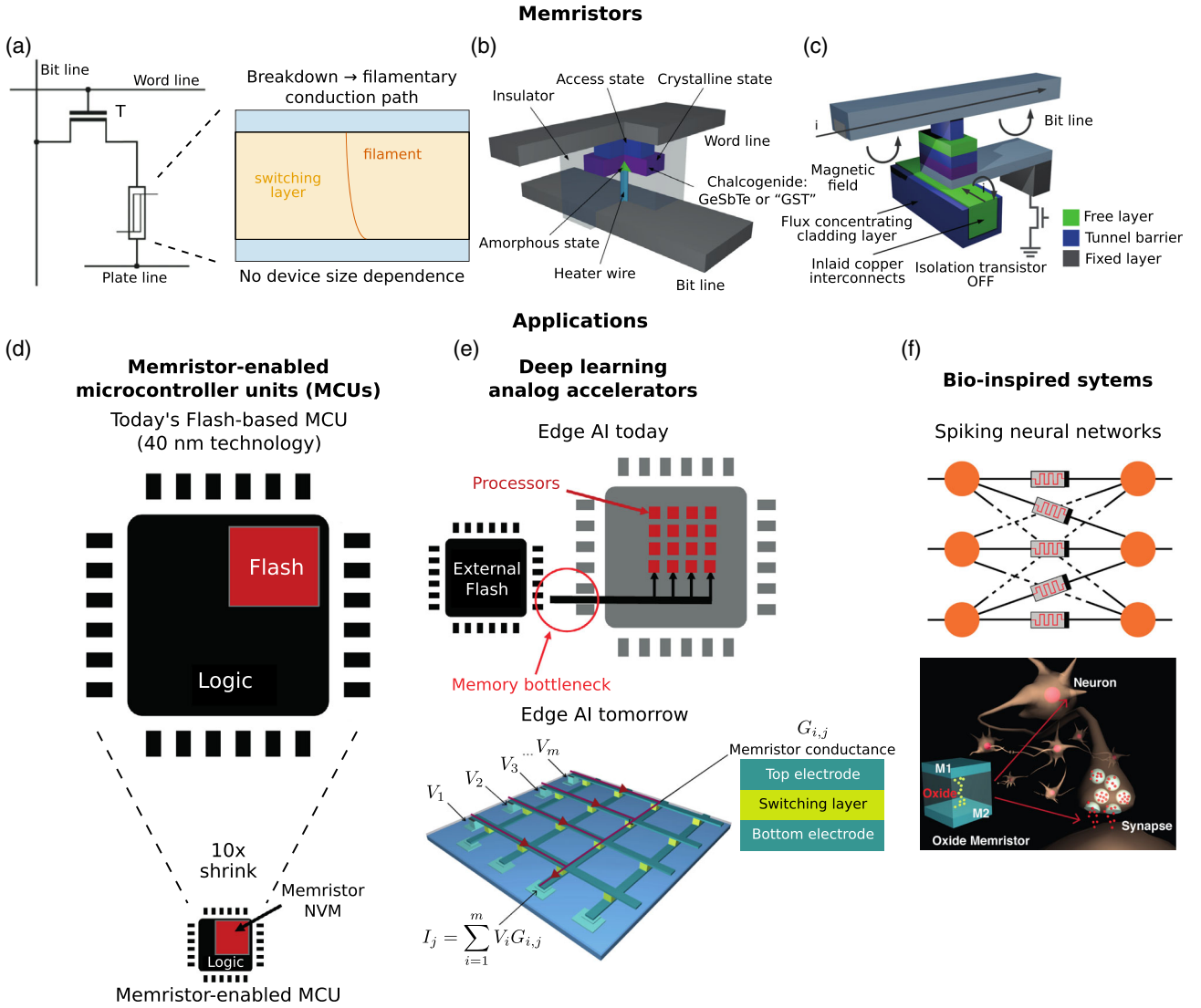


Figure 1. Overview of memristive devices and their potentials uses in computing. There exist multiple memristive technologies, including a) ReRAM, b) PCM, and c) MRAM. Common applications of memristive devices include d) embedded nonvolatile memory, e) analog deep learning accelerators based on programmable crossbars, and f) bioinspired systems implemented by memristors that emulate synapses and neurons.

between two electrodes. The resistance of the switching layer can be programmed to various resistance states with the application of voltage pulses.

Resistance switching manifests itself slightly differently in ReRAMs, PCMs, and MRAMs. In ReRAM technologies, resistance switching is based on the creation/dissolution of conductive filaments (intrinsic to the oxide layer or a result of metallic diffusion from electrodes); local nanoionic redox phenomena drive resistance switching in ReRAMs. There are different flavors of ReRAM devices, but they can be broadly divided by the type of switching: 1) intrinsic switching, which manifests itself as an intrinsic property of the switching material and 2) extrinsic switching, which is controlled by indiffusion (typically from metal electrodes) and drift of metal ions extrinsic to the fabricated switching layer.[8] Alternatively, the devices may be classified by the dominant driving forces of the switching process; this would result in electrochemical metallization cells, valence change ReRAMs, and thermochemical ReRAMs.[7] In PCMs, the switching is governed by the reversible process of crystallization and amorphization of phase-change materials. Finally, the programmable relative spin orientation of two ferromagnetic layers is the basis of MRAM operation.

It is important to note that novel devices based on different resistance switching mechanisms are still being developed. Notable examples include nanometallic memristors,[9] which rely purely on electronic effects, and Ti/ZnO/Pt structures that rely on carrier trapping/detrapping of the trap sites.[10] Such devices could provide further improvements in terms of speed, uniformity, and low-power operation.

As shown in Figure 1 d–f, a wide range of memristor applications have been suggested, including embedded digital nonvolatile memory, analog deep learning accelerators, and neuromorphic spiking systems.[11] We discuss these and other potential applications in more detail later in the text. We suggest consulting rich literature for details and descriptions of different physical mechanisms and many more types of memristive devices and technologies.[7,12,13]

1.1.2. Spintronics

Conventional electronic systems rely on electron charges—these systems use voltage levels and currents to process information. However, the electron has another intrinsic property, called "spin," making it analogous to a tiny magnet. The core concept of spintronics is to use this degree of freedom to create functional electronic devices that cannot be realized using conventional semiconductor technologies. Magnets can store digital information cheaply and reliably due to their excellent nonvolatile property; combining this with spin-dependent transport for efficient writing and read-out is a viable approach to making disruptive innovations in the electronic device market.

The quantum mechanical Pauli exclusion principle and the Coulomb interaction generate the so-called exchange coupling between spins, creating the magnetic orders of spin ensembles, with the order parameter of magnetization $\boldsymbol{M}$. The central concept of spintronics is to store information bits in local $\boldsymbol{M}$ that can be electrically written and read in an energy-efficient manner for data storage and processing.[14–16] The magnetic field $\boldsymbol{H}$ is a conventional way to control $\boldsymbol{M}$ via the Zeeman interaction ($-\boldsymbol{M} \times \boldsymbol{H}$), for example, when the two vectors are aligned in parallel, the free energy of the system becomes lower and hence stabilized. Magnetic moments are nonvolatile in general, meaning that when we switch off magnetic fields, the size and direction of the moments are unchanged. This is possible due to the presence of the aforementioned exchange interaction and magnetic anisotropies.

In a ferromagnet, where the exchange interaction aligns individual moments along the same direction, flipping one of the magnetic moments against this direction requires large energy cost, thus maintaining the total moments along the favored direction(There is an excitation state of this magnetically ordered system (called magnons) that can be realized by tilting the moments; however, this results in a slight change of the total moments). The equilibrium direction of $\boldsymbol{M}$ is determined by the magnetic free energy where—with zero external magnetic field—the magnetic anisotropy creates local minima as a function of angle, as shown in **Figure 2**a. The energy barrier between the minima characterizes the thermal stability of the moment orientation, directly relevant to the reliability for storing data in a magnetic cell. If the barrier height ΔE is too small, an accidental reversal of the magnetic moment can take place, resulting in data loss, whereas data retention of ten years is generally guaranteed when $\Delta E/(k_{\mathrm{B}}T) > 60$ in typical magnets. This mechanism is the origin of nonvolatality in magnetic materials, and optimizing parameters such as ΔE (the size of magnetic anisotropy) is one of the major topics in spintronic applications.

Another key ingredient for spintronic devices is that transport parameters (e.g., resistivity) can be controlled by $\boldsymbol{M}$. In ferromagnets, the density of states at the Fermi level for up and down spin electrons is different due to energy splitting by the exchange coupling (see Figure 2b). Magnetic tunnel junctions (MTJs) exploit this property as tunnel magnetoresistance (TMR) by having two magnetic layers with a tunnel barrier (Figure 2b), in which the tunneling probability depends on the spin polarization of electrons at the Fermi level for each electrode.[17,18] TMR devices exhibit larger resistance changes than giant magnetoresistance (GMR),[19–21] in particular TMR devices with MgO barrier.[22,23] A high TMR value is critical for reliability of read-out of spintronic devices using MTJs as well as for reducing the read-out time since it realizes a faster rate of voltage changes during reading.

Normally we switch $\boldsymbol{M}$ by applying $\boldsymbol{H}$ greater than magnet's anisotropy field, as shown in Figure 2a. However, this writing method is not scalable with downsizing since $\boldsymbol{H}$ produced by an electric current is proportional to the absolute value of the electric current, not current density. As a scalable magnetization switching mechanism, the concept of spin transfer torque (STT) was independently proposed by Slonczewski[24] and Berger.[25] In this scheme (Figure 2c), spin-polarized currents injected into a magnetic layer can exert torques via angular momentum transfer between the conduction and localized electrons.[26] An electric current through an MTJ can switch magnetization of one layer when the current size is sufficiently large. The size of this switching current density is directly relevant to the power consumption of spintronic memories, like MRAM, which stores and processes digital information by flipping $\boldsymbol{M}$ in an array of MTJs.

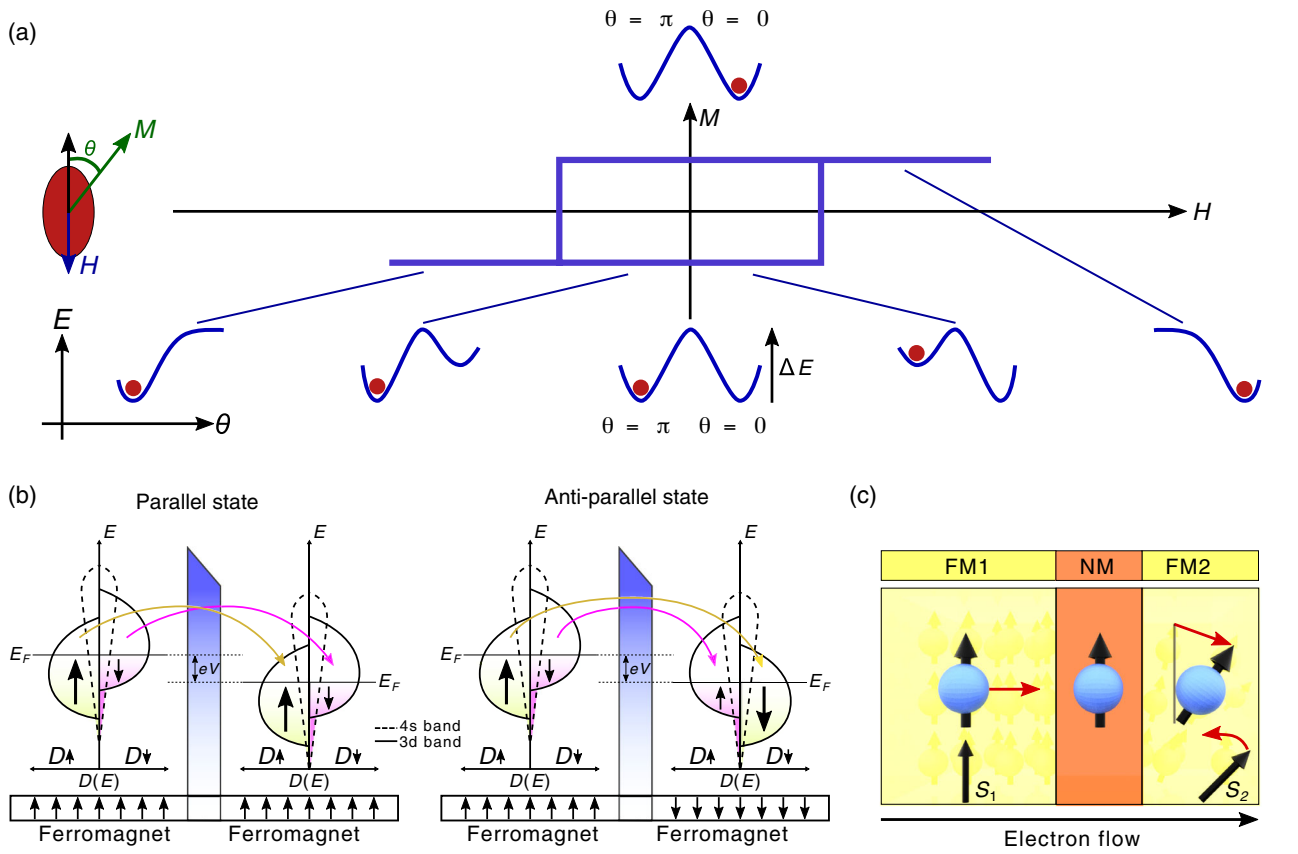


Figure 2. Basic principles of spintronics. a) Magnetic switching in a magnet with uniaxial anisotropy. When we apply magnetic fields along the easy axis of the uniaxial anisotropy, the magnetic free energy as a function of angle θ changes. At the point where the energy barrier is removed by the Zeeman energy, the magnetization switching occurs as a jump of M. b) Schematic of tunnel magnetoresistance with the density of states $D(E)$ for two magnetic electrodes for MTJ. For the parallel configuration (left image), the large $D(E)$ of the up-spin electrons at the Fermi level can produce large tunneling probability proportional to D_1^2. For the antiparallel case, the tunneling probability is smaller due to the size of $D_1 \cdot D_2$. c) Schematic of the STT mechanism. A spin-polarized electron is generated in FM1 and enters into FM2. The polarization angle of the conduction electron is tilted in FM2 as a result of angular momentum transfer into S_2. This produces a magnetic torque on S_2.

Furthermore, it is also an important parameter for footprint (density) of spintronic arrays as each MRAM cell is powered by a CMOS transistor underneath, and this element is so far the limiting factor of downsizing of MRAM. As high current requires a large CMOS transistor, a high-density MRAM can be achieved when the writing current is small. Other emerging magnetization control mechanisms include spin–orbit torques (SOTs) and voltage-controlled magnetic anisotropy (VCMA), for which readers are invited to read other studies[27–30] for more details.

1.1.3. 2D Materials

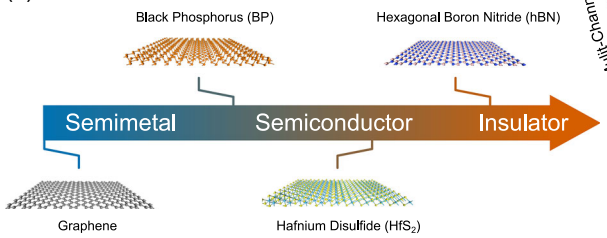
Adopting different computation variables (such as spin) and architectures (such as neuromorphic) leads to a demand for novel materials capable of supporting such technologies. In this perspective, we also explore 2D layered materials, often simply referred to as 2D materials. We believe that these materials are among the most promising candidates for future computing due to a large variety of properties they offer, the possibility of being easily combined into functional structures, and the ease of integration with existing semiconductors and fabrication lines. 2D materials are a large class of materials consisting of stacks of individual layers held together by, typically, van der Waals forces. Each layer is formed by covalently bonded atoms and exhibits fully saturated surface bonds, resulting in crystals that are stable even in the form of a single layer and hence the name "2D" materials.

Figure 3a shows a list of the most common 2D materials grouped according to their structure. X-enes are materials consisting of a single element such as graphene and silicene, whereas X-anes and fluoro-X-enes are their chemical derivatives, for example, graphane refers to hydrogenated graphene and fluorographene—to fluorinated graphene. Transition metal dichalcogenides (TMDs) are a class of compounds formed by a transition metal element (M) mainly from IV, V, or VI group and a chalcogen (X), with a generalized formula MX_2.[31] These materials form layered structures of the form XMX, with the chalcogen atoms in two hexagonal planes separated by a plane of metal atoms.[31] Semimetal chalcogenides (SMCs) are similar to TMDs; they are formed by a semimetal and a chalcogen,

(a)

Groups	Examples
X-enes	Graphene Silicene Germanene Phosphorene
X-anes	Graphane Silicane Germanane Stanane
Fluoro-X-enes	Fluorographene Fluorosilicene Fluorogermanene
TMDs	MoS_2 WS_2 $MoSe_2$ WSe_2 HfS_2 $HfSe_2$
SMCs	GaS GaSe InS InSe
MX-enes	Ti_2C Ti_3CN V_4C_3 Mo_4VC_4
2D insulators	hBN Mica

(b)



(c)

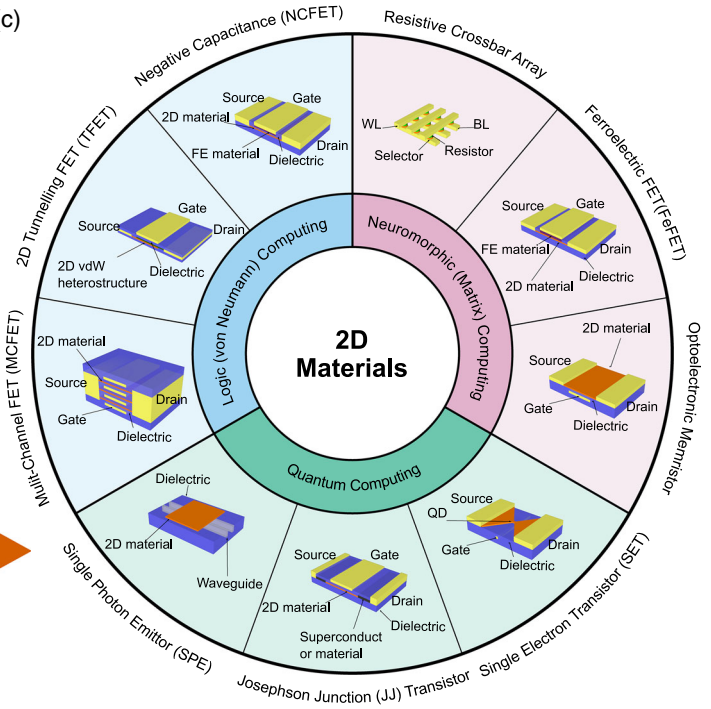


Figure 3. Overview of 2D materials and their applications. a) List of the most common 2D materials. b) The range of 2D materials' electrical properties from zero-bandgap semimetals, such as graphene, to wide-bandgap insulators, such as hexagonal boron nitride. c) Devices based on 2D materials for applications in logic, neuromorphic, and quantum computing.

usually occurring in M_2X_2 stoichiometry. MX-enes are ternary layered materials having occurred in the formula $M_{n+1}AX_n$, where M is an early transition metal, A is an element from group 13 or 14, X is either carbon or nitrogen, and n is an integer between 1 and 3. Finally, the 2D "library" also includes insulators, such as hexagonal boron nitride (hBN), an isomorph of graphene consisting of boron and nitrogen atoms.

Despite sharing a similar structure, the properties of 2D materials are incredibly diverse—the "family" of 2D materials includes semimetals, direct- and indirect-bandgap semiconductors, insulators, metals, superconductors, topological, and ferromagnetic insulators, as illustrated in Figure 3b. The lack of dangling bonds on the surface enables deterministic stacking of different 2D materials to form heterostructures without lattice matching constrains, usually referred to as van der Waals (vdW) heterostructures.[32] Such structures have atomically precise control of the thicknesses of the different layers with abrupt interfaces, leading to unprecedented flexibility in terms of materials and properties available. Moreover, by controlling the angle between the layers, it is possible to define a Moiré superlattice which provides a further degree of freedom, leading to new phenomena (such as superconductivity in twisted bilayer graphene [BLG][33]) and enabling a novel approach to electronics referred to as "twistronics".[34]

With tens of materials experimentally available and over 2000 theoretically predicted,[35] 2D materials represent one of the most promising material systems for future computing. From a manufacturing point of view, 2D materials also have significant advantages. Indeed, these materials are (sub)nanoscopic only in terms of thickness, whereas their lateral dimensions can be macroscopic, leading to a significant technological advantage over other nanomaterials because they can be processed using "conventional" semiconductor planar technology.[36] Combined with the ease of transferring them from one substrate to another, 2D materials can be easily integrated with existing technologies, particularly at the back-end of line in CMOS production lines.[37] 2D materials are strong candidates for present and future computing paradigms, including logic and neuromorphic computing, as shown in Figure 3c. Despite being beyond the scope of this perspective, it is worth noting that 2D materials, including BLG quantum dot (QD),[38] Josephson junctions,[39] and hBN single-photon emitters (SPEs),[40–42] have also been used in the field of quantum computing. Nevertheless, applications of 2D materials in the field of electronic devices goes beyond what is shown in Figure 3c. Here, we will provide a prospective overview on how 2D materials can be used as an enabling platform for the technologies discussed. The reader is invited to read other studies[43–46] for in-depth reviews on the recent progress in the field of 2D electronics.

2. Conventional Computing Hardware

Digital computers are the basis of our information and communication technologies. Logic gates, such as NAND or NOR, implement Boolean algebra, which is used for all digital information processing. Field-effect transistors (FETs), fundamental building blocks of digital circuits, have followed Moore's scaling law for more than 50 years. We are still managing to scale

transistors; however, the scaling rate has slowed down over the last years.[47] There is a tremendous motivation to investigate post-CMOS technologies, starting from innovations in and understanding of materials and basic nanoscale devices. ReRAM, spintronic, and 2D-based devices could *all* potentially offer better scaling prospects, as well as improved energy efficiency and speed. These emerging technologies could serve as improved realizations of digital memory and logic, which are used in all conventional, general-purpose computers.

2.1. Memory

ReRAM, PCM, and MRAM devices can all be operated as binary memory with two well-defined nonvolatile memory states. Both ReRAM and MRAM devices compare favorably against currently used Flash technology, beating it in most performance metrics.[48,49] Microcontroller units (MCUs) are the first and most attractive applications for these emerging nonvolatile memory technologies. Today's MCUs use embedded NOR Flash, which cannot be easily scaled beyond 28 nm node size; this represents a critical bottleneck, especially considering that more applications are becoming data intensive (e.g., automotive MCU needs to operate on a significant amount of data collected by numerous sensors found in modern cars). Both ReRAM and MRAM present an attractive opportunity to replace NOR Flash in embedded memory applications offering better scaling (down to most aggressive nodes, <10 nm) and faster programming/reading speeds (<5 ns). Beyond embedded memory, ReRAM and MRAM are also considered as data storage and thus as a replacement for NAND Flash. NAND Flash is scalable to most aggressive nodes; however, ReRAM and MRAM offer better reading speed and lower energy. Another attractive potential application could replace or augment static random-access memory (SRAM) in edge AI applications,[50,51] where ReRAM/MRAM offers similar reading speeds but better scalability and energy efficiency.

In general, ReRAM—when used as nonvolatile digital memory—offers 1) excellent scalability (e.g., 10×10 nm[52] and likely below[53]), which is highly competitive with current memory technologies, like SRAM and Flash; 2) large resistance ratio (>10 and much more) critical for fast sensing and reading speeds; 3) fast programming (typically <100 ns, although there are reports of 100 ps programming[54]); 4) excellent endurance (10^{12} switching cycles have been reported[55]); and 5) small operational energy (e.g., sub-pJ bit^{-1}[56]).

In terms of commercialization of ReRAM, in 2013, Panasonic released the first MCU with embedded ReRAM.[57] Many other companies are currently developing ReRAM technologies, including Rambus, 4DS, Dialog Semiconductor, Crossbar, Intrinsic Semiconductor Technologies, Weebit Nano, eMemory, and global foundries such as Taiwan Semiconductor Manufacturing Company (TSMC).

MRAM consists of an array of MTJs connected with read and write lines for its memory operation. **Figure 4** displays three different types of individual MRAM cells with different writing mechanisms. In particular, the STT writing method has become ripe for industrial applications, and two magnetic layers are magnetized along the perpendicular to the junction plane to minimize footprint. Nonvolatility offers significant advantages in energy saving against volatile memories, such as dynamic random-access memory (DRAM), which requires constant power to maintain their stored information as energy loss.

Major electronics companies have been focusing on MRAM development. Samsung and the partnership between Everspin and Global Foundaries announced their release of a 1 GB-embedded MRAM on their 28/22 nm technology nodes.[58,59] The write speed of their technology is orders of magnitude faster than eFlash (200 ns vs tens of μs), with comparable read speeds, providing a power advantage over eFlash in many applications. Intel announced that they are embedding STT-MRAM into devices using its 22 nm FinFET process, with a bit yield rate of greater than 99.9%.[60]

STT-MRAMs are believed to be more suitable to embedded memory applications for industrial-grade MCUs, autonomous vehicles, and various IoT devices.[61] Using its high-speed nature, STT-MRAM has been considered as an alternative to SRAM applications[62] as well as L3/L4 cache replacement, which requires high performance in terms of density, write efficiency, bandwidth, and endurance.[63] We point curious readers to more detailed review papers[64,65] as there is an excellent summary table of STT-MRAM specs against other memory applications.

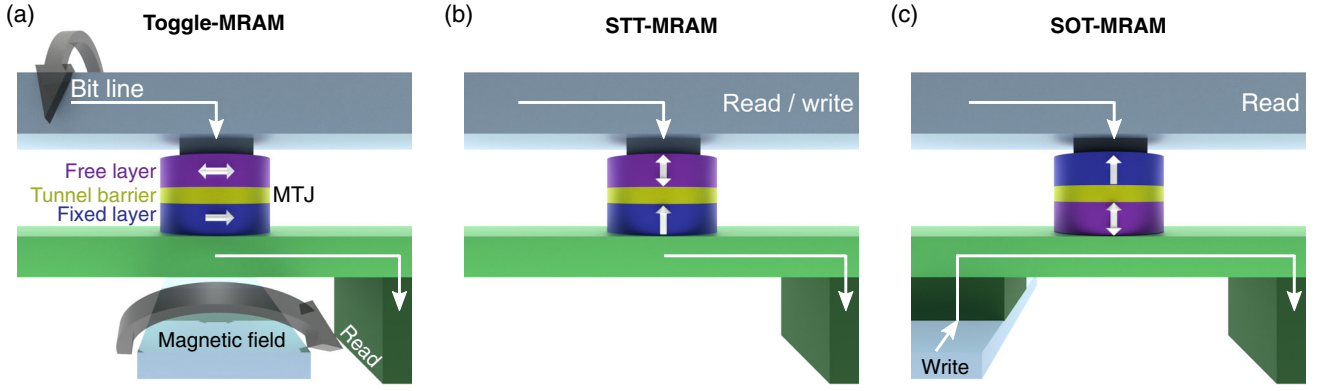


Figure 4. Schematics of different MRAM architectures. a) Toggle-MRAM uses magnetic fields to switch magnetization in an MTJ. b) STT-MRAM directly passes an electric current through an MTJ to write their cells. c) In SOT-MRAM, an electric current flows through the write line, which generates magnetic torques on the layer above.

Emerging writing mechanisms of MRAM cells, such as SOT and VCMA, have been extensively studied for the next generation of MRAM.[28,58,64,65] Wafer-scale SOT-MRAMs compatible with CMOS technologies have been demonstrated,[66] together with fast switching demonstration (less than 400 ps) in a perpendicularly magnetized SOT-MRAM cell,[67] that show high-speed switching, as well as improved endurance for both standalone-memory and processing-in-memory (PIM) applications.[68] PIM refers to performing computational tasks within the memory units where the memory units within these applications need to have high endurance and fast writing/reading since data are more rapidly accessed inside.[69] Combining SOT and STT writing mechanisms is expected to reduce the writing current down to a range of 10–100 fJ bit^{-1}[58,70,71]

2.2. Logic

2.2.1. Field-Effect transistors

Since the groundbreaking work of Geim and Novoselov that experimentally unveiled the electronic properties of graphene in 2004,[72] significant attention has been put into its use for transistors. That is due to graphene's atomic thickness, extremely high room-temperature mobility, saturation velocity and thermal conductivity, and the ambipolarity of its field effect. Because of the lack of bandgap, however, graphene field-effect transistors (GFETs) cannot be switched off. As a result, GFETs exhibit only a modest ON/OFF ratio of ≈10, which is not suitable for transistor logic applications, where current ratios in excess of 10^4 are required.[73] Nevertheless, GFETs have been used in analog RF electronics, where switching off is not essential, achieving cutoff frequencies in excess of 400 GHz,[74] and in applications directly benefiting of the ambipolarity of the field effect, such as high-frequency mixers.[75]

The possibility of isolating individual atomically thin crystals demonstrated by graphene paved the way to the exploration of other 2D materials, in particular TMDs. Molybdenum- and tungsten-based TMDs, such as MoS_2 WS_2 and WS_2, are of particular interest for future transistor logic applications as they are atomically thin semiconductors, which can enable reduction of the characteristic length of FETs beyond the limit faced by silicon.[76] Scaling of body thickness by adopting ultrathin-body on insulator and fin field-effect transistor (FinFET) structures has indeed been key to reduce short-channel effects and extend Moore's law.[77] However, the reduction of body thickness in bulk semiconductor below 5 nm is accompanied by a rapid decrease of charge carrier mobility due to thickness variation, dangling bonds, and roughness, resulting in a limit to further scaling.[78] Conversely, 2D semiconductors have thickness <1 nm (e.g., single-layer $MoS_2 \approx 0.65$ nm) and mobility in excess of 100cm^2 Vs^{-1}, significantly higher than sub-5 nm silicon.[76] Moreover, in 3D semiconductors, there is usually a tradeoff between bandgap and effective mass and therefore mobility. Materials with higher bandgap normally show larger effective mass and lower mobility, imposing a compromise between performance and power consumption. This is not the case in 2D semiconductors, where mobility is determined by phonon scattering,[79] thus enabling materials combining large bandgap and high mobility. Saturation velocity also plays a very important role in ultrascale devices, where the in-plane field can easily exceed 1 kV cm^{-1}; however, the data available for TMDs are scattered and would require a more thorough investigation. TMDs are extremely interesting candidates for future multichannel field-effect transistor (MCFET) to reduce the scaling length of FETs beyond the limits imposed by silicon.

2.2.2. Tunneling FETs

One of the main figures of merit when assessing CMOS efficiency is the energy-delay product of its metal–oxide–semiconductor field-effect transistors (MOSFETs). One of the main factors governing the EDP is the SS, which is a measurement of the gate voltage required to change the drain current by a factor of ten. SS in MOSFETs, regardless of the channel material, is thermodynamically limited by the Boltzmann limit. In MOSFETs

$$SS = k_B T \ln(10)\left(1 + \frac{C_s}{C_{ox}}\right) \quad (1)$$

where C_s and C_{ox} are the semiconductor capacitance (or depletion layer capacitance) and the gate dielectric capacitance, respectively. It is clear that even if $C_{ox} \gg |C_s$, SS will never drop below $k_B T \ln(10)$ (≈ 60 mV dec^{-1} at room temperature).

An alternative to thermionic injections over an energy barrier is tunneling field-effect transistors (TFETs). They rely on band-to-band tunneling (BTBT), resulting in SS not limited to 60 mV/dec. However, to achieve steep SS beyond the thermal limit, the energy window for tunneling needs to be sharp, which can only be attained with very abrupt interface. This has proven to be challenging in conventional planar homojunction TFETs because controlling the doping profile to the atomic level is extremely difficult. Bulk heterojunction TFETs, on the other hand, have been demonstrated to outperform their homojunction counterpart. Nevertheless, the fabrication of such sharp interfaces is still challenging.

2D materials, owing to their inherently atomically flat surfaces, are well suited for such applications, as they can form a sharp interface ideal for tunneling. Different material combinations have been explored, such as graphene/boron nitride/graphene,[80] graphene/WS_2/graphene,[81] MoS_2/WSe_2,[82] black phosphorus/SnS_{22},[83] and SnS_2/WSe_2.[84] More interestingly, heterostructures between 2D materials and a 3D conventional one can bring the best of both worlds. In particular, MoS_2/germanium TFETs have been reported to achieve "record" SS of 3.9 mV dec^{-1} at room temperature, combined with higher current density compared with other subthermionic transistors.[85]

2.2.3. Negative Capacitance Field-Effect Transistors

Steep SS can also be attained by modifying the gating mechanism in MOSFETs. In these devices, the gate controls the channel through direct capacitive approach. Negative capacitance field-effect transistor (NCFET) utilizes ferroelectric (FE) materials, which exhibit metastable spontaneous polarization, which can be triggered through an external field from a low state to high state. NCFETs employ this abrupt change to switch the device

from low (OFF) state to high (ON) state. However, it is important to note that an appropriate dielectric material (DE) needs to be connected in series with the FE layer to stabilize the negative capacitance state and reduce hysteresis.[86,87] The aforementioned SS formula needs be changed to include the FE layer effect. Hence

$$\begin{aligned} SS &= k_B T \ln(10)\left(1+\frac{C_s}{C_{FE}+C_{ox}}\right) \\ &= k_B T \ln(10)\left(1-\frac{C_s}{|C_{FE}|-C_{ox}}\right) \end{aligned} \quad (2)$$

where C_{FE} is the capacitance of the FE layer.[88]

It is clear that to achieve sub-60 mV/dec SS, C_{ox} must be larger than $|C_{FE}|$, which adds another criterion for choosing the suitable dielectric. As in MOSFETs, NCFETs benefits from improved gate control that 2D materials exhibit due to their thinness. Hence, SS as low as 25 mV dec^{-1} has been achieved in MoS_2 NCFET with $Hf_{0.5}Zr_{0.5}O_2$ FE with low hysteresis (≈28 mV).[89] In addition, based on the industrial direction for MOSFETs, we expect that an all-2D-stacked negative-capacitance gate-all-around field-effect transistor (GAAFET) that can combine steep SS and high ON current would be one of the most promising structures for future logic devices.

2.2.4. Memristor-Based Logic

There are several ways of using memristors for digital logic. For instance, memristors have been considered as programmable switches for field-programmable gate arrays (FPGAs) in the past.[90,91] Although, currently, these switches are implemented using SRAM, memristor-based switches could lead to significantly improved energy efficiency, for example, reducing cell area by 40% and energy-delay-product, by 28%.[92] Alternatively, memristors could be used to implement IMPLY (implication $p \Rightarrow q$ is false only when p is true and q is false) logic gates.[93] The interest comes from the fact that an IMPLY gate with the FALSE operation (FALSE operation always yields a logical zero) comprises a complete logic structure. Memristive implementation of this fundamental logic element could lead to memristor-based logic circuits. More details and performance comparisons involving this approach can be found in the study by Kvatinsky et al.[94]

3. Future Computing Hardware

While existing compute infrastructure based on Boolean algebra offers many advantages, new hardware paradigms can 1) improve the efficiency of existing computing tasks and 2) implement functionality that would be infeasible to realize using conventional computers.

One example is neuromorphic computing, which mimics the structure and/or operation of the brain.[95] Neuromorphic computing can encompass efficient implementations of both well-established concepts, like artificial neural networks (ANNs), and exotic approaches to information processing, like spiking neural networks (SNNs) and reservoir computing. This paradigm aims to perform complex tasks, including recognition and classification, with little energy.[96–98] Multiple emerging technologies hold promise of making these new approaches to computing hardware a reality.

3.1. Artificial Neural Networks on Crossbar Arrays

ANNs are implemented on digital computers, but they are very resource-intensive because of 1) large amounts of data being processed and 2) the nature of conventional computer architectures. Modern neural networks can often have billions of parameters,[99] and von Neumann architecture, which most computers are built around, is not well suited to handle such large models. Time and energy is mostly spent *not* on performing computations, but on repeatedly moving data between memory and computing units.[100]

Resistive crossbars—one of the simplest examples of neuromorphic hardware—may offer a solution to this problem. In these structures, resistive elements are arranged in an array, as shown in **Figure 5**a. Ohm's law achieves multiplication of

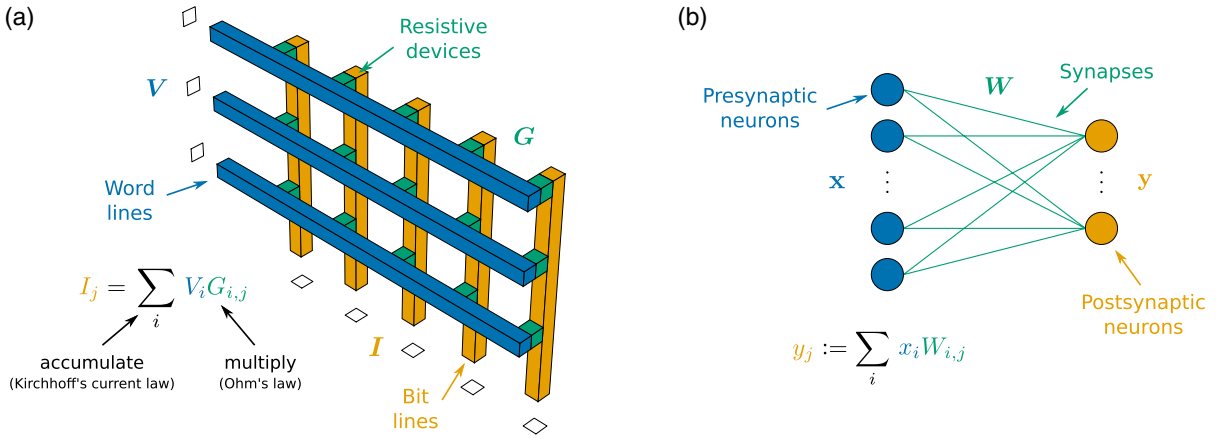


Figure 5. The computing principles behind crossbar-array-based dot-product engines and fully connected synaptic layers. a) Using resistive devices in each of the vertical (bit) lines, crossbar arrays can compute dot products of voltages and conductances. When multiple of these bit lines are combined, one can compute products of voltage vectors ***V*** and conductance matrices ***G***. b) Synapses in neural networks scale the incoming signals. Before nonlinear transformations, these scaled signals are added together by the postsynaptic neurons.

voltages and conductances, while Kirchhoff's current law achieves addition of currents. With the crossbar structure, these are combined, producing multiply-accumulate operations or multiplication of voltage vectors and conductance matrices. Using pairs of devices,[101] the principle can be easily extended to handle negative numbers, thus achieving in-memory multiplication of *arbitrary* vectors and matrices. Such crossbar are usually referred to as dot-product engines (DPEs).

Hardware acceleration of linear algebra operations is easily applicable to ML and ANNs in specific. Fully connected neural networks heavily rely on vector-matrix multiplication to compute outputs of the synaptic layers; this is demonstrated in Figure 5b. During training, optimal weights $\boldsymbol{W}$ are determined; this is typically done using gradient descent.[102] After that, during a process called inference, only the inputs $\boldsymbol{x}$ change—with each new example, outputs $\boldsymbol{y}$ are either used for prediction directly or are passed along to the next synaptic layer. The fact that weights do not change during inference is one of the primary reasons as to why crossbars are an appealing candidate for their physical implementation. Inference can be accelerated by encoding weights into conductances and inputs—into voltages. The ability of DPEs to compute vector–matrix products means that, this way, the synaptic layers of ANNs can be implemented in memory, that is, there is no need to transfer the weights during computation, only the inputs have to be applied in the form of voltage vectors.

Easily programmable resistive devices are perfect candidates for DPE implementations. Memristors are one example of such devices—one may encode matrix values into the conductances of memristors embedded in the crossbar array. Such programming can be done using voltage pulses, which require very little energy.[103] Examples of such devices include Ta/HfO_2[101,104] and SiO_x memristors. Spintronic devices can also be used to emulate synaptic behavior—MTJs can act as a local nonvolatile digital memory or as a continuously varying resistance.[105–107] For example, the conductance of a three-terminal MTJ device can be encoded by controlling the magnitude and the direction of the current flowing through the underlying heavy-metal layer.[106]

Several neuromorphic proof-of-concept devices have also been realized using 2D materials. That includes atomically thin MoS_2 memristors having switching ratio $> 10^4$ and stable operation up to 50 GHz,[108] memristors consisting of multilayer MoS_2 encapsulated between graphene layers capable of high-temperature (> 300°C) operation,[109] lithium-ion intercalated few-layer metal dichalcogenides, and phosphorus trichalcogenides.[110] Different switching mechanisms have been identified in 2D materials, including formation of conductive filaments,[111] grain boundary migration,[112] phase transition,[113] oxygen migration,[109] and graphene which have been showed to improve the I_{ON}/I_{OFF} ratio in tetrahedral amorphous carbon-resistive metal–insulator–metal (MIM) devices.[114] In addition, three-terminal memristors based on 2D materials have shown great promise due to the additional tunability and functionality provided though the additional gate terminal. An example of three-terminal memristors is synaptic transistors, which utilize wide range of mechanisms, such as floating gate flash memory[115] and gate-controlled charge trapping in gate dielectric.[116] On the other hand, ferroelectric field-effect transistors (FeFETs) utilize a FE layer in place of the gate dielectric. As a result, nonvolatile states can be written to the device with gate control.[117] Finally, memtransistors operate similarly to its two-terminal counterparts (memristors) with the exception that the resistance of the device is gate controlled. In fact, several mechanisms governing resistive switching in memtransistors have been demonstrated, such as grain boundary migration,[112] FE switching,[118] and gate-controlled vdW heterojunctions.[119]

Of course, with any of these technologies, due to the analog nature of computations, the idealized vector-matrix computation in Figure 5a is often difficult to achieve. First, it may be challenging to set devices to the desired values of conductances $G_{i,j}$. As an example, devices like memristors may get stuck in a certain conductance state[120] or even fail to electroform (i.e., become conductive),[121] experience random telegraph noise (RTN)[122,123] or programming variability,[124] or have their conductance state drift over time.[125] Even more difficult to tackle are nonidealities that result in deviations from the linear (with respect to conductance and/or voltage) behavior, which DPEs rely on; such nonidealities include I–V nonlinearity[126,127] and line resistance.[128–130]

There are multiple ways of utilizing DPEs for the implementation of ANNs. The most obvious one has been alluded to earlier—neural network weights may be mapped onto crossbar conductances after they have been trained on digital computers. However, it may also be possible to train ANNs directly on crossbar arrays thus saving time, energy, and even preventing unnecessary greenhouse gas emissions. That is attractive because training a large ANN on a conventional digital architecture may emit as much CO_2 as five cars throughout their lifetimes.[131]

Ex situ training is the most straightforward way of learning the weights of neural networks that are later implemented physically. Such ANNs can utilize a training process that is no different from the one used to train conventional networks. Training on a digital computer is the simplest approach, but it obviously has disadvantages due to the mismatch between well-behaved conventional electronic systems and crossbar arrays consisting of analog devices.

If one does not take nonidealities into account, networks trained ex situ may perform considerably worse on crossbar arrays, compared to their digital counterparts. For example, small number of achievable states, limited dynamic range, device-to-device (D2D) variability, and I–V nonlinearities may all contribute to higher error rate.[132] In addition, system-level issues, including the aforementioned line resistance,[101,133] may disturb the distribution of currents and increase the error further.

This may be partly addressed by modifying ex situ training so that the nonidealities are considered *before* deploying ANNs onto DPEs. It is possible to model the behavior of analog devices, like memristors, and adjust the expected outputs of the hardware neural network accordingly. Even for stochastic nonidealities, the nature of the stochasticity may inform the training process and make ANNs more robust. That is not unique to crossbar-based neural networks as noise can make even conventional ANNs more robust.[134]

There are multiple ways of taking nonidealities into account during training. For example, the cost function (which quantifies

how close ANN outputs are to the expected ones) may be modified to incorporate the randomness associated with device behavior.[135] Alternatively, network weights can be disturbed to represent nonidealities, like read and write noise.[136] Where the effects of nonidealities cannot be represented by injecting noise into the weights, their behavior can be redefined to reflect, for example, I–V nonlinearities.[127]

Although ex situ training can significantly improve the performance, it is important to consider that it relies on a number of assumptions. If the modeling of nonidealities is inaccurate, that will be reflected in the training on a digital computer and may result in deviations from intended behavior when ANNs are implemented physically. However, this may be partly hedged against by including randomness in the modeling. Randomness may represent the uncertainty in not only the device behavior, but also the designers' understanding of how the devices behave. Therefore, it can improve the performance when the modeling is not perfectly accurate or even when different nonidealities manifest themselves.[127]

Finally, one may employ in situ training, which can refer to either full or partial training directly on crossbar arrays. Performing ANN training on real devices can help networks adapt to specific instantiations of nonideal behavior—no two analog devices are the same, but in situ, unlike ex situ, training can take individual variations into account without the need to model the behavior. In situ approach makes networks more robust to nonidealities, like faulty devices and D2D variability.[137] One may even combine the two paradigms—conventional ex situ training can be used to produce ANN weights, after which in situ retraining is used to recover from defects, like stuck devices.[138]

Unfortunately, training networks in situ is challenging. Because conventional ML methods rely on incremental adjustments of synaptic weights, analog devices may often be too unreliable for the task. For example, the training process can be negatively affected by the asymmetry and nonlinearity of conductance changes,[137] both of which are common in, for example, memristive devices. Approaches for dealing with this include adjusting the fabrication process[139,140] and using digital electronic devices in conjunction with the analog ones.[141]

3.2. Spiking Neural Networks

Although ANNs are loosely inspired by the brain, they are highly inefficient compared to biological systems. This is due to the fact that there are fundamental differences between the two systems. The adopted models of brain learning involve dynamic adjustment of synaptic strengths by the neuronal spiking activity. In comparison, learning in ANNs is based on gradient descent methods, which adjust weights in order to optimize an objective function.

There is significant research interest in developing SNNs as it is believed they could yield much better energy efficiency. The fundamental difference is that in SNNs, time is used directly to encode and process information—it is encoded in the time of arrival of binary events ("spikes"). Two main functional units needed for the implementation of SNNs are neurons and synapses. Neurons are typically implemented as simple leaky integrate-and-fire neurons, which are capable of integrating signals over time and producing spikes when a certain threshold is reached. In terms of the synaptic functionalities, apart from adjustable strength, it is necessary to implement different local learning rules, such as spike-time-dependent plasticity, spike-rate-dependent plasticity, short-term plasticity, long-term potentiation, and long-term depression.

The energy efficiency argument relies on hopes of developing dedicated hardware platforms[142] because current von Neumann architectures are not best suited for the implementation of SNN algorithms. Although there exist many CMOS-based implementations of SNN hardware accelerators,[143–152] these systems are still lacking in terms of the energy efficiency of biological counterparts. It is believed that emergent technologies will be able to directly implement critical functionalities using voltages and currents much lower than CMOS equivalents.[153]

Memristive technology has been used to implement multiple elements of the SNN paradigm. Synaptic functionalities were implemented by incorporating temporal plasticity as well as particular local learning rules.[154–156] PCM memristors,[157] ReRAMs,[158,159] and Mott-based memristors[160,161] have all been used for emulating neuronal activity. For more details and a much more comprehensive overview of using memristors for SNNs, we refer readers to another study.[153]

Spintronic devices, too, may be used for physical implementations of SNNs. The oscillatory behavior of biological neurons can be emulated using spin-torque nano-oscillators (STNOs);[164,165] the required power may be achieved when assisted by a microwatt nanosecond laser pulse.[166] When the system is configured toward the limit of superparamagnetism, the random spiking of biological neurons can be emulated to perform population coding and probabilistic computing.[163,167] **Figure 6**a shows a schematic of probabilistic computing with probabilistic-bits (p-bits), where the structural design of the MTJs benefits from the low-energy barrier of the superparamagnetic tunnel junctions. The analog input voltage, I, to some junction can cause a nonlinear response to the digital output voltage, m, (Figure 6b) and form random fluctuations analogous to 0's and 1's of a stochastic neuron at room temperature. Nevertheless, other systems such as memristors or nanoarrays or exploiting nonlinear dynamics in variant forms of magnetic spin textures like domain walls or skyrmions can also be engineered to facilitate such properties,[168–172] demonstrating the potential of spintronic devices as artificial neuromorphic components.

Photonic circuits represent another possible approach to neuromorphic computing and SNNs in particular.[173] For example, black phosphorus has been used to emulate excitatory and inhibitory action potentials using oxidation-related defects.[174] Also, WSe_2/hBN heterostructures have been used as 7-bit nonvolatile optoelectronic memories[175] and for colored and mixed-color pattern recognition.[176] Further, the developments in the field of optoelectronic memristive devices[177] could provide further flexibility and extended functionality, such as in-sensory computing.[178] In many cases, the operation of these devices requires both electronic and optical stimulation;[179] however, fully optically operable memristors can be realized[180] with favorable properties for neuromorphic computing.

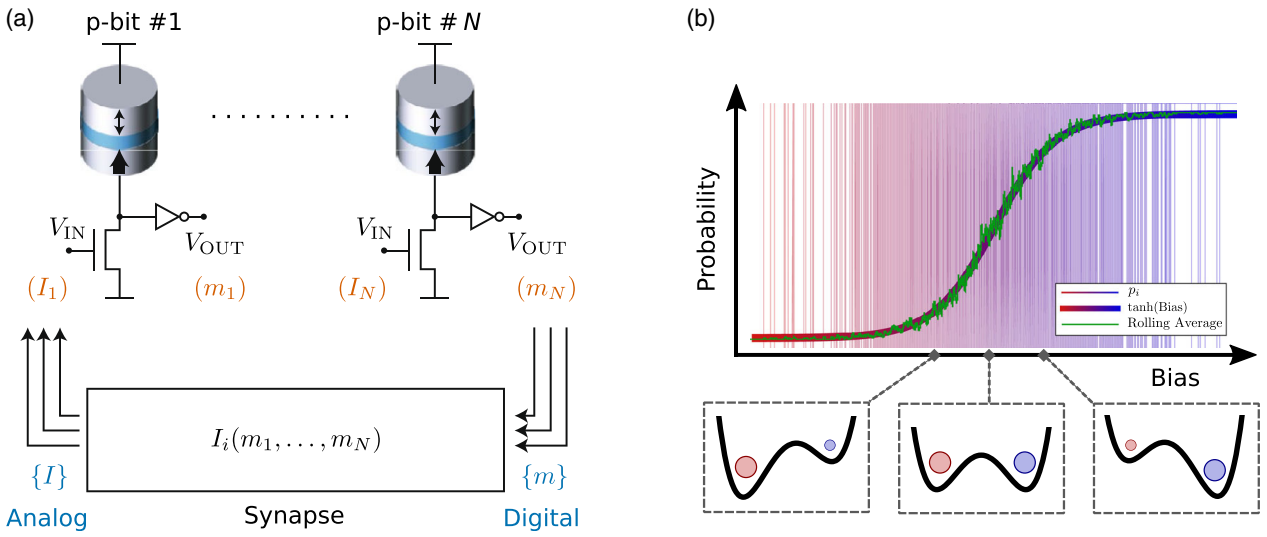


Figure 6. Spintronic approaches to spiking neural networks. a) Schematic representation of p-bit computing scheme. Superparamagnetic tunnel junctions offer extremely low-energy barriers, which can be exploited to solve complex problems. The analog input voltage to some junction, *I*, can cause a nonlinear response to the digital output voltages as shown in (b) and form random fluctuations analogous to 0's and 1's of a stochastic neuron at room temperature. Adapted with permission.[162] 2020, Springer Nature. b) The control of bias voltages changes the relative energies of two states. Adapted with permission.[163] 2019, AIP Publishing.

3.3. Reservoir Computing

In addition to the aforementioned fully connected ANNs, there also exist recurrent neural networks (RNNs). These networks contain recurrent connections and can be incredibly useful when dealing with time series data.[181] However, RNNs can suffer from vanishing and exploding gradients, which makes their training especially difficult.[182]

Given the challenges of RNNs, reservoir computing has been suggested as an alternative.[183] It relies on systems that exhibit rich dynamic behaviors to do the computations "for free." Like activation functions in conventional ANNs may introduce nonlinearities, physical "reservoirs," which are complex, nonlinear, and have short-term memory properties, are able to map inputs to the nonlinear dynamics of a high-dimensional system. This enables to perform training only on the last synaptic—and usually linear—layer. The principles behind reservoir computing are visualized in **Figure 7**a.

Many kinds of memristors hold promise as potential mediums of reservoir computing. One of the factors enabling this is the fact that many memristors exhibit short-term memory properties. In the case of some memristors, repeatedly applying voltage pulses may gradually increase the response, while the absence of the pulses will make the devices decay toward their original resting state.[186] In addition, nonlinear *I*–*V* characteristics of memristive devices can be incredibly useful for reservoir computing applications.[184]

One may also use spintronic devices in reservoir computing applications. Figure 7b shows an experimental demonstration of using a single STNO facilitated with an MTJ as a reservoir. It

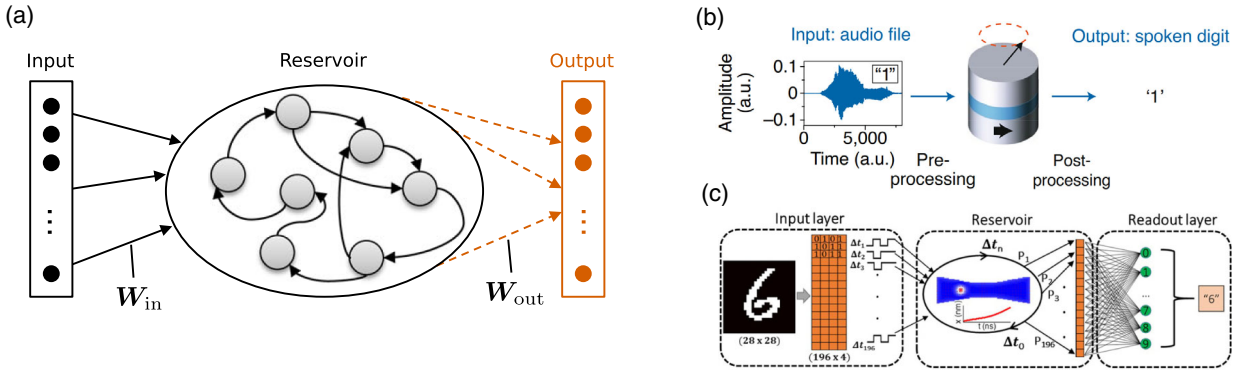


Figure 7. Operating principles and example implementations of reservoir computing. a) Inputs and the interconnected nonlinear units of the reservoir are connected through a set of weights, $\boldsymbol{W}_{in}$. Those reservoir nodes and the outputs are connected through another set of weights, $\boldsymbol{W}_{out}$; during training, only $\boldsymbol{W}_{out}$ need to be learned. Adapted under the terms of the CC-BY license.[184] Copyright 2021, The Authors. Published by Springer Nature. b) Experimental implementation of physical reservoir computing using spin–torque nano-oscillator for spoken digit recognition. Adapted with permission.[162] 2020, Springer Nature. c) Numerical demonstration of the physical reservoir computing scheme using skyrmion positions for classification of handwritten digits. Adapted with permission.[185] 2019, AIP Publishing.

exploited time multiplexing to emulate up to 400 neurons by tuning the state of each neuron at periodic intervals. The relationships between the input current and the oscillation frequency can bring a nonlinear response, and the motion of spins in the free layer showed history dependence as a response to the amplitudes of analog audio signals. Another example has been demonstrated by exploiting spinwaves in 3D space using small-sized metal electrodes to apply and detect the input and output voltages (currents).[187] The system was configured as a stacked device consisting of a thin yttrium iron garnet layer between the conductive substrate and magnetoelectric coupling layer. The nonlinear effects and the history-dependent motion of the spinwaves were achieved by controlling the stability of the precession of the spins by reducing the applied bias DC magnetic field, allowing the device to satisfy the reservoir computation criteria. Yet another proposed medium for reservoir computing has been magnetic skyrmions due to their stability and controllable history-dependent nonlinear effects. In an example design in the study by Jiang[185] (shown in Figure 7c), handwritten digits were converted into an input sequence of current pulses, which was fed into a magnetic skyrmion memristor. The nonlinear relationship between the positions of the magnetic skyrmions allowed the system to be configured as a physical reservoir. In addition to this approach, a wide range of different systems have been proposed and investigated, including the manipulation of skyrmion fabrics, skyrmion position, and interaction of multiple skyrmions.[188–190]

4. Outlook and Conclusion

Here we discuss the basics of three emerging nanoscale technologies with great potential to improve and extend the infrastructure of compute hardware. One plausible scenario that addresses the growing diversity and complexity of computational problems includes a synergy between more conventional, digital systems, and new paradigms of computing hardware. General-purpose computing will likely remain best implemented on digital systems, which use Boolean logic and higher-precision computing. However, some applications, like ANNs, which are currently realized on these digital systems, could benefit from speed and power efficiency that neuromorphic hardware accelerators offer. Further, computing approaches like SNNs, which are even less fit for conventional computers, could be implemented using devices that exhibit more exotic behaviors, such as synaptic plasticity or neuronal spiking. Finally, there are paradigms of computing that are feasible or possible only with devices that exhibit certain physical behavior; an example of this is reservoir computing. Memristive, spintronic, and 2D-materials-based devices will likely play a role in both the improvement of digital hardware and the adoption of more novel approaches.

Many systems would benefit from fast low-power memristive hardware, but, at the same time, some are constrained by additional requirements. For example, memristive ANNs could in theory be used by autonomous driving companies; however, these companies often utilize driving data to improve their ML models and deploy the updated models continuously.[191,192] Even if ANNs are trained ex situ and identical versions are deployed onto memristive systems, each physical instantiation will be at least slightly different. This could affect not only the behavior of individual vehicles, but also the ML pipeline, that is, data that are collected and then used to improve the models[193] that are deployed to *all* cars. In general, we can identify multiple challenges of memristive systems that need to be addressed before widescale deployment in the real world: 1) nonidentical behavior of identically designed systems[194] 2) stochasticity, including possibly changing behavior over time[195,196] 3) difficulty of reprogramming once deployed in the real world 4) difficulty of identifying hardware faults.[197,198]

Where safety and behavior reproducibility are key, special attention currently needs to be paid to the treatment of device stochasticity, variability, and reliability. This is especially true when memristors are used unconventionally (i.e., not for digital nonvolatile memory, but as analog memory and neuromorphic computational primitives). Similarly, applications where hardware needs to be constantly reconfigured (e.g., updating ML models in autonomous vehicles) would be challenging—even in controlled environments, programming memristive devices remains difficult.[124,199] In addition, cycling endurance might need to be improved to match the endurance of volatile memory (e.g., 10^{16} cycles in SRAM).

We believe that memristors can be the most useful where computing needs to be fast, low power and/or local (i.e., not in the cloud). The last possibility flows from the first two—data-intensive applications like ANNs consume a lot of power, thus the computing often takes place remotely; however, memristive technologies—due to their speed and power efficiency—can enable to perform the computations locally.[200,201] We therefore believe that these devices are very well suited for applications like the IoT where potential violations of privacy remain a significant issue.[202] Memristive implementations of data-intensive tasks would not only eliminate the need to send data to the server, but also ensure low-power operation and high speed.

Spintronics is another promising approach that can advance the state of the art in multiple paradigms of computing. Spintronic memory and logic circuits are expected to open a novel route to manipulate information more efficiently and their prototypes have been actively proposed.[28,58,64,65] In the coming decade, we predict an increased dominance of hybrid CMOS-spintronic computing architectures based on MRAM techniques such as STT, SOT, and VCMA. Moreover, the desired progress in speed, energy, and scaling will also require the use of advanced materials such as antiferromagnets,[203] 2D materials,[204,205] and topological insulators.[206] Spintronic devices are also being employed in a new class of computer architecture such as all spin logic (ASL)[207] and logic-in-memory (LIM).[208] LIM structures are hybrid in nature, combining contemporary spintronics components, such as MTJs, with current CMOS devices. Advancement in fabrication technology (e.g., 3D back-end process) enabled the growth of MTJs on the silicon layer without compromising the functionality of the circuit.[209] Circuits developed using LIM hold advantages over the conventional CMOS technologies due to their lower power dissipation, nonvolatility, high density, fast reading capability, infinite endurance, and 3D fabrication adaptability.[210]

The properties of spintronic devices (e.g., high-speed dynamics from GHz to potentially THz ranges, nonvolatality,

plasticity and nonlinearity) offer ample room for accessing numerous building blocks that can mimic the key features of biological synapses and neurons.[105,107,168–172,211] In spintronic devices, the processing/transfer of information can be achieved via spin currents, spin waves, microwave signals, or magnetic spin textures such as domain walls and skyrmions. Such properties can potentially find their unique positions in the electronics market by offering more compact and energy-efficient approaches, exploiting the spin degree of freedom.

While proof-of-concept spintronics-based neuromorphic computing implementations have been demonstrated,[105,162,211,212] there remain a number of key challenges. Although many creative and exciting ideas have been proposed, it is important to consider the viability of mass production and scalability when it comes to spintronics-based neuromorphic computing. Likewise, traditional algorithms used on CMOS technology require enhanced tuning to harness the maximum potential of such spintronic neuromorphic chips. Similar to von Neumann architecture for conventional computing, a dedicated architecture is a prerequisite for wide-scale implementation of neuromorphic computing.[213] Furthermore, additional research is required to increase the capability of the proposed devices for example, enhancing the coupling efficiency between the MTJ layers and the relatively low ratio of maximum-to-minimum resistance of the existing devices.[162]

2D materials are yet another key enabler for future computing technologies. Taken individually, or in combination to form heterostructures with tailored properties, they offer an unprecedented playground for both conventional and emerging forms of computing. However, there are a number of challenges to overcome before their full potential is realized.

The first is doping because the ion implantation processes commonly used in semiconductor industry are not applicable to 2D materials due to their atomic thickness.[214] Instead of replacing atoms in the crystal lattice (as in substitutional doping used for 3D semiconductors), doping in 2D materials is normally achieved either by physisorption, covalent bonding of impurities (chemical doping), or proximity with compounds, which modifies the dielectric environment and leads to local gating effect (sometimes referred to as solid-state doping).[215] Unfortunately, to date, none of these methods fully satisfy the stringent requirement of ultrascaled devices and more research effort should be devoted to identifying an industry-compatible, precise, stable, and reproducible doping method.

The second challenge to overcome is related to the deposition of high-κ dielectrics. Indeed, the lack of dangling bonds in 2D materials' surfaces complicates the growth of thin, uniform insulating layers by atomic layer deposition and, often, "seed" layers are required to facilitate the growth. Dielectrics are not only important for the functionality of devices (e.g., as gate dielectric in MOSFETs) but also to encapsulate 2D materials, as their properties are often significantly degraded by substrate, contamination, roughness, and charged impurities. A promising alternative is represented by 2D dielectrics, which form atomically sharp interfaces with other 2D materials. HbN is by far the most explored 2D dielectric, which enabled experimental investigation of transport phenomena and proof-of-concept devices.[216,217] However, low dielectric constant ($\approx$3) and difficulty in scalable production of multilayer hBN limit its applicability in high-performance computing. A more promising option is represented by the possibility of oxidizing hafnium and zirconium-based multilayer TMDs to form high-κ dielectrics HfO_2 and ZrO_2.[218,219] This approach is of particular interest as it is the equivalent to the oxidation of silicon and results in almost-perfect interfaces between the pristine semiconducting part and the oxidized surface.

The third challenge is represented by contacts. Contact resistance is usually high and cannot be reduced by ion implantation as in 3D semiconductors. Moreover, due to the Schottky junction formed when depositing metals on 2D semiconductor, contact resistance is also modified by applied gate voltage, introducing additional delays, and complicating the analysis of devices.[214] Theoretical and experimental effort should be devoted toward this essential but often disregarded aspect of computing. Finally, scalable production of 2D materials should be optimized, in particular for what concerns reproducibility and control over defects and contaminations. Chemical vapor deposition (CVD) growth has made impressive progress in the last ten years; however, some fundamental challenges remain, such as the lack of an industrially scalable, clean transfer of graphene. Our view is that 2D materials do not represent a replacement, but rather a complement to current bulk semiconductor technology. The relative ease of integrability of such materials into established semiconductor production lines will indeed be key for a synergy between the two technologies and enable new, high-performance computing.

Memristors, spintronics, and 2D materials are rapidly developing and changing fields. New developments span materials, devices, circuit/system design, and algorithmic approaches. This perspective article provides a basic introduction to central ideas, explores potential advantages over conventional CMOS technologies, and lists some pressing challenges that still need to be addressed. Memristors, spintronics, and 2D-based electronics are among the most promising candidates for supporting future computing systems. There is a strong possibility that they will coexist and complement other emerging technologies and approaches, as well as conventional electronics systems.

Acknowledgements

A.M. acknowledges funding from the Royal Academy of Engineering under the Research Fellowship scheme, A.J.K. acknowledges funding from the Engineering and Physical Sciences Research Council (EP/P013503/1) and the Leverhulme Trust (RPG-2016-135), D.J. acknowledges studentship funding from the Engineering and Physical Sciences Research Council (ref. 2094654), and A.A. acknowledges funding from the Saudi Ministry of Education.

Conflict of Interest

Adnan Mehonic and Anthony J. Kenyon are co-founders of Intrinsic, a company developing memristor technology.

Keywords

machine learning, memristors, spintronics, neuromorphic computing, 2D materials

Received: March 10, 2022
Revised: June 1, 2022
Published online:

[1] N. Jones, *Nature* **2018**, *561*, 163.
[2] *Nat. Electron.* **2018**, *1*, 205.
[3] A. Mehonic, A. J. Kenyon, **2022**, *604*, 255.
[4] ARK Invest. Big Ideas, **2021**, https://web.archive.org/web/20220227204750/https://research.ark-invest.com/hubfs/1_Download_Files_ARK-Invest/White_Papers/ARK%E2%80%93Invest_BigIdeas_2021.pdf (accessed: Febraury 2022).
[5] Q. Shao, et al. Spintronic Memristors for Computing, arXiv: 2112.02879, **2021**.
[6] L. Chua, *IEEE Trans. Circuit Theory* **1971**, *18*, 507.
[7] D. Ielmini, R. Waser, *Resistive Switching: From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications*, John Wiley & Sons, Hoboken, NJ **2015**.
[8] A. Mehonic, A. J. Kenyon, in *Defects at Oxide Surfaces* (Eds: J. Jupille, G. Thornton, Springer, New York **2015**, pp. 401–428.
[9] Y. Lu, A. Alvarez, C.-H. Kao, J.-S. Bow, S.-Y. Chen, I.-W. Chen, *Nat. Electron.* **2019**, *2*, 66.
[10] R. Pan, J. Li, F. Zhuge, L. Zhu, L. Liang, H. Zhang, J. Gao, H. Cao, B. Fu, K. Li, *Appl. Phys. Lett.* **2016**, *108*, 013504.
[11] A. Mehonic, A. Sebastian, B. Rajendran, O. Simeone, E. Vasilaki, A. J. Kenyon, *Adv. Intell. Syst.* **2020**, *2*, 2000085.
[12] S. Raoux, *Chem. Rev.* **2010**, *110*, 240.
[13] J. Grollier, D. Querlioz, M. D. Stiles, *Proc. IEEE*, **2016**, *104*, 2024.
[14] A. Hirohata, et al., *J. Magn. Magn. Mater.* **2020**, *509*, 166711.
[15] I. Zuti´c, J. Fabian, S. Sarma, *Rev. Mod. Phys.* **2004**, *76*, 323.
[16] S. A. Wolf, D. D. Awschalomr, A. Buhrmanj, M. Daughtons, V. Molnárm, L. Roukesa, Y. Chtchelkanova, D. M. Treger, *Science* **2001**, *294*, 1488.
[17] M. Julliere, *Phys. Lett. A* **1975**, *54*, 225.
[18] S. Yuasa, D. D. Djayaprawira, *J. Phys. D Appl. Phys.* **2007**, *40*, R337.
[19] M. N. Baibich, J. M. Broto, A. Fert, F. Nguyen Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, J. Chazelas, *Phys. Rev. Lett.* **1988**, *61*, 2472.
[20] G. Binasch, *Phys. Rev. B* **1989**, *39*, 4828.
[21] E. Tsymbal, D. Pettifor, *Solid State Phys.* **2001**, *56*, 113.
[22] S. S. P. Parkin, C. Kaiser, A. Panchula, P. M. Rice, B. Hughes, M. Samant, S.-H. Yang, *Nat. Mater.* **2004**, *3*, 862.
[23] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, K. Ando, *Nat. Mater.* **2004**, *3*, 868.
[24] J. Slonczewski, *J. Magn. Magn. Mater.* **1996**, *159*, L1.
[25] L. Berger, *Phys. Rev. B* **1996**, *54*, 9353.
[26] D. Ralph, M. Stiles, *J. Magn. Magn. Mater.* **2008**, *320*, 1190.
[27] A. Manchon, J. Železný, I. M. Miron, T. Jungwirth, J. Sinova, A. Thiaville, K. Garello, P. Gambardella. *Rev. Mod. Phys.* **2019**, *91*, 035004.
[28] Q. Shao, et al., *IEEE Tran. Magnet.* **2021**, *57*, 1.
[29] F. Matsukura, Y. Tokura, H. Ohno, *Nat. Nanotechnol.* **2015**, *10*, 209.
[30] T. Nozaki, T. Yamamoto, S. Miwa, M. Tsujikawa, M. Shirai, S. Yuasa, Y. Suzuki, *Micromachines* **2019**, *10*, 327.
[31] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, M. S. Strano, *Nat. Nanotechnol.* **2012**, *7*, 699.
[32] K. S. Novoselov, A. Mishchenko, A. Carvalho, A. H. C. Neto, *Science* **2016**, *353*.
[33] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, P. Jarillo-Herrero, *Nature* **2018**, *556*, 43.
[34] S. Carr, D. Massatt, S. Fang, P. Cazeaux, M. Luskin, E. Kaxiras, *Phys. Rev. B* **2017**, *95*, 075420.
[35] N. Mounet, M. Gibertini, P. Schwaller, D. Campi, A. Merkys, A. Marrazzo, T. Sohier, I. E. Castelli, A. Cepellotti, G. Pizzi, N. Marzari, *Nat. Nanotechnol.* **2018**, *13*, 246.
[36] W. Kong, H. Kum, S.-H. Bae, J. Shim, H. Kim, L. Kong, Y. Meng, K. Wang, C. Kim, J. Kim, *Nat. Nanotechnol.* **2019**, *14*, 927.
[37] D. Neumaier, S. Pindl, M. C. Lemme, *Nat. Mater.* **2019**, *18*, 525.
[38] L. Banszerus, S. Möller, E. Icking, K. Watanabe, T. Taniguchi, C. Volk, C. Stampfer, *Nano Lett.* **2020**, *20*, 2005.
[39] D. Rodan-Legrain, Y. Cao, J. M. Park, S. C. de la Barrera, M. T. Randeria, K. Watanabe, T. Taniguchi, P. Jarillo-Herrero, *Nat. Nanotechnol.* **2021**, *16*, 769.
[40] J. C. Stewart, Y. Fan, J. S. H. Danial, A. Goetz, A. S. Prasad, O. J. Burton, J. A. Alexander-Webber, S. F. Lee, S. M. Skoff, V. Babenko, S. Hofmann, *ACS Nano* **2021**, *15*, 13591.
[41] T. T. Tran, K. Bray, M. J. Ford, M. Toth, I. Aharonovich, *Nat. Nanotechnol.* **2016**, *11*, 37.
[42] N. Mendelson, et al., *Nat. Mater.* **2021**, *20*, 321.
[43] C. Liu, H. Chen, S. Wang, Q. Liu, Y.-G. Jiang, D. W. Zhang, M. Liu, P. Zhou, *Nat. Nanotechnol.* **2020**, *15*, 545.
[44] W. Huh, D. Lee, C.-H. Lee, *Adv. Mater.* **2020**, *32*, 2002092.
[45] Q. Zhao, Z. Xie,Y.-P. Peng, K. Wang, H. Wang, X. Li, H. Wang, J. Chen, H. Zhang, X. Yan, *Mater. Horiz.* **2020**, *7*, 1495.
[46] S. Batool, M. Idrees, S. R. Zhang, S. T. Han, Y. Zhou, *Nanoscale Horiz.* **2022**, *7*, 480
[47] M. M. Waldrop, *Nat. News* **2016**, *530*, 144.
[48] S. Bhatti, et al., *Mater. Today* **2017**, *20*, 530.
[49] F. Zahoor, T. Z. A. Zulkifli, F. A. Khanday, *Nanoscale Res. Lett.* **2020**, *15*, 1.
[50] S. Sakhare, M. Perumkunnil, T. H. Bao, S. Rao, W. Kim, D. Crotti, F. Yasin, S. Couet, J. Swerts, S. Kundu, D. Yakimets, R. Baert, H. R. Oh, A. Spessot, A. Mocuta, G. S. Kar, A. Furnemont, in *IEEE Int. Electron Devices Meeting*, IEEE, Piscataway, NJ **2018**, p. 18–3.
[51] M. Rios, F. Ponzina, G. Ansaloni, A. Levisse, D. Atienza, in *2021 Design, Automation & Test in Europe Conf. & Exhibition (DATE)*, IEEE, Grenoble, France **2021**,pp. 1881–1886.
[52] B. Govoreanu, G. S. Kar, Y.-Y. Chen, V. Paraschiv, S. Kubicek, A. Fantini, I. P. Radu, L. Goux, S. Clima, R. Degraeve, N. Jossart, O. Richard, T. Vandeweyer, K. Seo, P. Hendrickx, G. Pourtois, H. Bender, L. Altimime, D. J. Wouters, J. A. Kittl, M. Jurczak, in *Int. Electron Devices Meeting*, IEEE, Washington, DC, USA **2011**, p. 31–6.
[53] V. V. Zhirnov, R. Meade, R. K. Cavin, G. Sandhu, *Nanotechnology* **2011**, *22*, 254027.
[54] A. C. Torrezan, J. P. Strachan, G. Medeiros-Ribeiro, R. S. Williams, *Nanotechnology* **2011**, *22*, 485203.
[55] M.-J. Lee, C. B. Lee, D. Lee, S. R. Lee, M. Chang, J.-H. Hur, Y.-B. Kim, C.-J. Kim, D. H. Seo, S. Seo, U.-I. Chung, I.-K. Yoo, K. Kim, *Nat. Mater.* **2011**, *10*, 625
[56] F. Miao, J. P. Strachan, J. J. Yang, M.-X. Zhang, I. Goldfarb, A. C. Torrezan, P. Eschbach, R. D. Kelley, G. Medeiros-Ribeiro, R. S. Williams, *Adv. Mater.* **2011**, *23*, 5633.
[57] Panasonic Industry. 8bit Ultra Low Power MN101L, **2022**, https://web. archive. org /web/20220306155110/https://industrial.panasonic.com/ww/products/pt/mn101l (accessed: March 2022).
[58] S. Newsroom Samsung Electronics Starts Commercial Shipment of eMRAM Product Based on 28nm FD-SOI Process, https://news.samsung.com/global/samsung-electronicsstarts-commercial-shipment-of-emram-product-based-on-28nm-fd-soi-process, (accessed: February, 2022).
[59] GlobalFoundries, Making New Memories: 22nm eMRAM is Ready to Displace eFlash, https: //gf.com/blog/making-new-memories-22nm-emram-ready-displace-eflash. (accessed: February, 2022).
[60] D. McGrath, Intel Says FinFET-Based Embedded MRAM is Production-Ready, https://www.eetimes.com/intel-says-finfet-based-embedded-mram-is-production-ready/, (accessed: February, 2022).

[61] Z. Guo, Z. et al. *Proc. IEEE* **2021**, *109*, 1398).
[62] J. Alzate, et al. in *IEEE Int. Electron Devices Meeting (IEDM)*, IEEE, Piscataway, NJ **2019**, pp. 2.4.1–2.4.4.
[63] Sakhare, S. et al. in *IEEE Int. Electron Devices Meeting (IEDM)*, IEEE, Piscataway, NJ **2018**, pp. 18.3.1–18.3.4.
[64] G. Finocchio, M. Di Ventra, K. Y. Camsari, K. Everschor-Sitte, P. K. Amirie, Z. Zeng, *J. Magn. Magn. Mater.* **2021**, *521*, 167506.
[65] B. Dieny, et al., *Nat. Electron.* **2020**, *3*, 446.
[66] K. Garello, et al. in *Symp. on VLSI Technology*, **2019**, pp. T194–T195.
[67] M. Cubukcu, et al., *IEEE Trans. Magnet.* **2018**, *54*, 1.
[68] Z. He, S. Angizi, F. Parveen, D. Fan, in *IEEE/ACM Int. Symp. on Nanoscale Architectures*, IEEE, Piscataway, NJ **2017**, 97–102.
[69] A. Sebastian, M. Le Gallo, R. Khaddam-Aljameh, E. Eleftheriou, *Nat. Nanotechnol.* **2020**, *15*, 529.
[70] A. Van den Brink, S. Cosemans, S. Cornelissen, M. Manfrini, A. Vaysset, W. Van Roy, T. Min, H. J. M. Swagten, B. Koopmans, *Appl. Phys. Lett.* **2014**, *104*, 012403.
[71] M. Wang, et al., *Nat. Electron.* **2018**, *1*, 582.
[72] K. S. Novoselov, et al., *Science* **2004**, *306*, 666.
[73] F. Schwierz, *Nat. Nanotechnol.* **2010**, *5*, 487.
[74] R. Cheng, et al., *Proc. Natl. Acad. Sci.* **2012**, *109*, 11588.
[75] H. Wang, A. L. Hsu, T. Palacios, *IEEE Microwave Magazine* **2012**, *13*, 114.
[76] Y. Liu, X. Duan, H.-J. Shin, S. Park, Y. Huang, X. Duan, *Nature* **2021**, *591*, 43.
[77] IEEE, *International Roadmap for Devices and Systems (IRDS)*, 2020 ed., IEEE, Piscataway, NJ **2020.**
[78] M. Poljak, V. Jovanovic, D. Grgec, T. Suligoj, *IEEE Trans. Electron Dev.* **2012**, *59*, 1636.
[79] L. Cheng, Y. Liu, *J. Am. Chem. Soc.* **2018**, *140*, 17895.
[80] L. Britnell, R. V. Gorbachevr, J. B. D. Bellef, S. Mishchenkot, G. I. Katsnelsonl, E. V. Morozovn, M. R. P. Leista, K. Geim, K. S. Novoselov, L. A. Ponomarenko, *Science* **2012**, *335*, 947.
[81] T. Georgiou, et al., *Nat. Nanotechnol.* **2013**, *8*, 100.
[82] T. Roy, et al., *ACS Nano* **2015**, *9*, 2071.
[83] R. Yan, et al., *Nano Lett.* **2015**, *15*, 5791.
[84] T. Roy, M. Tosun, M. Hettick, G. Ho Ahn, C. Hu, A. Javey, *Appl. Phys. Lett.* **2016**, *108*, 083111.
[85] D. Sarkar, et al., *Nature* **2015**, *526*, 91.
[86] Cho, H. W. et al. *NPJ 2D Mater. Appl.* **2021**, *5*.
[87] M. A. Alam, M. Si, P. D. Ye, *Appl. Phys. Lett.* **2019**, *114*, 090401.
[88] L. Tu, et al., *Nat Commun* **2020**, *11*, 101.
[89] P. Pujar, et al. *Adv. Funct. Mat.* **2021**, *31*, 2170224.
[90] D. B. Strukov, K. K. Likharev, *Nanotechnology* **2005**, *16*, 888.
[91] G. S. Snider, R. S. Williams, *Nanotechnology* **2007**, *18*, 035204.
[92] T. Sakamoto, S. S. Wong, Y. Y. Liauw, in *Resistive Switching: From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications*, John Wiley & Sons, Ltd **2016**, p. 695.
[93] J. Borghetti, G. S. Snider, P. J. Kuekes, J. J. Yang, D. R. Stewart, R. S. Williams, *Nature* **2010**, *464*, 873.
[94] S. Kvatinsky, et al., *IEEE Trans. Very Large Scale Integr. VLSI Syst.* **2013**, *22*, 2054.
[95] S. Furber, *J. Neural Eng.* **2016**, *13*, 051001.
[96] K. Roy, A. Jaiswal, P. Panda, *Nature* **2019**, *575*, 607.
[97] I. Chakraborty, A. Jaiswal, A. K. Saha, S. K. Gupta, K. Roy, *Appl. Phys. Rev.* **2020**, *7*, 021308.
[98] Q. Wei, J. Tang, B. Gao, X. Li, H. Qian, H. Wu, in *Neuromorphic Devices for Brain-Inspired Computing*, **2021**, p. 173.
[99] N. Shazeer, et al. Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer, arXiv: 1701.06538 [cs.LG], **2017**.
[100] M. A. Zidan, J. P. Strachan, W. D. Lu, *Nat. Electron.* **2018**, *1*, 22.
[101] C. Li, et al., *Nat. Electron.* **2018**, *1*, 52.
[102] S. Ruder, An Overview of Gradient Descent Optimization Algorithms, arXiv: 1609.04747 [cs.LG] **2016**.
[103] K. E. Nikiruy, A. V. Emelyanov, V. A. Demin, V. V. Rylkov, A. V. Sitnikov & P. K. Kashkarov, *Tech. Phys. Lett.* **2018**, *44*, 416.
[104] A. Mehonic, et al., *Adv. Mater.* **2018**, *30*, 1801187.
[105] M. Romera, P. Talatchian, S. Tsunegi, F. Abreu Araujo, V. Cros, P. Bortolotti, J. Trastoy, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. Ernoult, D. Vodenicarevic, T. Hirtzlin, N. Locatelli, D. Querlioz, J. Grollier, *Nature* **2018**, *563*, 230.
[106] A. Sengupta, K. Roy, *Appl. Phys. Express* **2018**, *11*, 030101.
[107] K. M. Song, J.-S. Jeong, B. Pan, X. Zhang, J. Xia, S. Cha, T.-E. Park, K. Kim, S. Finizio, J. Raabe, J. Chang, Y. Zhou, W. Zhao, W. Kang, H. Ju, S. Woo, *Nat. Electron.* **2020**, *3*, 148.
[108] M. Kim, et al. *Nat. Commun.* **2018**, *9*.
[109] M. Wang, S. Cai, C. Pan, C. Wang, X. Lian, Y. Zhuo, K. Xu, T. Cao, X. Pan, B. Wang, S.-J. Liang, J. J. Yang, P. Wang, F. Miao, *Nat. Electron.* **2018**, *1*, 130.
[110] J. Zhu, et al., *Adv. Mater.* **2018**, *30*, 1800195.
[111] L. F. Sun, et al. *Nat. Commun.* **2019**, *10*.
[112] V. K. Sangwan, et al., *Nat. Nanotechnol.* **2015**, *10*, 403.
[113] F. Zhang, et al. *Nat. Mater.* **2019**, *18*, 55.
[114] A. K. Ott, et al., *2D Mater.* **2018**, *5*, 045028.
[115] S. Bertolazzi, D. Krasnozhon, A. Kis, *ACS Nano* **2013**, *7*. 3246.
[116] H. Tian, et al., *Nano Lett.* **2015**, *15*, 8013.
[117] Z.-D. Luo, et al. *ACS Nano* **2022**, *16*, 3362.
[118] K. Liu, B. Dang, T. Zhang, Z. Yang, L. Bao, L. Xu, C. Cheng, R. Huang, Y. Yang, *Adv. Mater.* 2108826.
[119] Y. Zhai, et al., *Adv. Funct. Mater.* **2022**, *32*, 2108440
[120] B. Zhang, N. Uysal, D. Fan, R. Ewetz, in *Proc. of the 24th Asia and South Pacific Design Automation Conf.*, Association for Computing Machinery, New York, NY, USA **2019**, pp. 438–443.
[121] J. J. Yang, et al., *Nanotechnology* **2009**, *20*, 215201.
[122] A. V. Yakimov, et al., *Appl. Phys. Lett.* **2019**, *114*, 253506.
[123] J. Park, et al., *Chaos Solitons Fract.* **2021**, *152*, 111388.
[124] K. M. Kim, et al., *Sci. Rep.* **2016**, *6*, 1.
[125] D. J. Kim, et al., *Appl. Phys. Lett.* **2013**, *103*, 142908.
[126] C. Sung, et al., *Nanotechnology* **2018**, *29*, 115203.
[127] D. Joksas, et al., *Adv. Sci.* **2022**, 2105784.
[128] A. Chen, *IEEE Trans. Electron Dev.* **2013**, *60*, 1318.
[129] A. Serb, W. Redman-White, C. Papavassiliou, T. Prodromakis, *IEEE Trans. Circuits Syst. I* **2015**, *63*, 827.
[130] D. Joksas, A. Mehonic, *SoftwareX* **2020**, *12*, 100617.
[131] E. Strubell, A. Ganesh, A. McCallum, Energy and Policy Considerations for Deep Learning in NLP. arXiv: 1906.02243 [cs.CL], **2019**.
[132] A. Mehonic, D. Joksas, W. H. Ng, M. Buckwell, A. J. Kenyon, *Front. Neurosci.* **2019**, *13*, 593.
[133] D. Joksas, P. Freitas, Z. Chai, W. H. Ng, M. Buckwell, C. Li, W. D. Zhang, Q. Xia, A. J. Kenyon, A. Mehonic, *Nat. Commun.* **2020**, *11*.
[134] X. Liu, S. Si, Q. Cao, S. Kumar, C. J. Hsieh, Neural SDE: Stabilizing Neural ODE Networks with Stochastic Noise. arXiv: 1906.02355 [cs.LG], **2019**.
[135] Y. Zhu, et al. in Design, Automation & Test in Europe Conf. & Exhibition (DATE), **2020**, pp. 1590–1593.
[136] V. Joshi, et al., *Nature Commun.* **2020**, *11*, 1.
[137] G. W. Burr, et al., *IEEE Trans. Electron Dev.* **2015**, *62*, 3498.
[138] C. Liu, M. Hu, J. P. Strachan, H. Li in *54th ACM/EDAC/IEEE Design Automation Conf. (DAC)*, Association for Computing Machinery, New York, NY, USA **2017**, 1–6.
[139] J. Woo, et al., *IEEE Electron Dev. Lett.* **2016**, *37*, 994.
[140] W. Wu, et al., *Technology* **2018**.
[141] S. Ambrogio, et al., *Nature* **2018**, *558*, 60.
[142] D. V. Christensen, et al. *Neuromorph. Comput. Eng.* **2022**, https://doi.org/10.1088/2634-4386/ac4a83.
[143] B. V. Benjamin,et al. *Proc. IEEE* **2014**, *102*, 699.

[144] S. Schmitt, J. Klähn, G. Bellec, A. Grübl, M. Güttler, A. Hartel, S. Hartmann, D. Husmann, K. Husmann, S. Jeltsch, V. Karasenko, M. Kleider, C. Koke, A. Kononov, C. Mauch, E. Müller, P. Müller, J. Partzsch, M. A. Petrovici, S. Schiefer, S. Scholze, V. Thanasoulis, B. Vogginger, R. Legenstein, W. Maass, C. Mayr, R. Schüffny, J. Schemmel, K. Meier, *Int. Joint Conf. on Neural Networks*, **2017**, pp. 2227–2234.
[145] P. Lichtsteiner, C. Posch, T. A. Delbruck, *IEEE J. Solid-State Circuits* **2008**, *43*, 566.
[146] S. Moradi, N. Qiao, F. Stefanini, G. Indiveri, *IEEE Trans. Biomed. Circuits Syst.* **2017**, *12*, 106.
[147] R. Wang, et al., *IEEE Trans. Biomed. Circuits Syst.* **2017**, *11*, 574.
[148] P. U. Diehl, M. Cook, *Front. Comput. Neurosci.* **2015**, *9*, 99.
[149] S. B. Furber, F. Galluppi, S. Temple, L. A. Plana, *Proc. IEEE* **2014**, *102*, 652.
[150] M. Davies, et al., *IEEE Micro* **2018**, *38*, 82.
[151] J. Pei, L. Deng, S. Song, M. Zhao, Y. Zhang, S. Wu, G. Wang, Z. Zou, Z. Wu, W. He, F. Chen, N. Deng, S. Wu, Y. Wang, Y. Wu, Z. Yang, C. Ma, G. Li, W. Han, H. Li, H. Wu, R. Zhao, Y. Xie, L. Shi, *Nature* **2019**, *572*, 106.
[152] G. K. Chen, R. Kumar, H. E. Sumbul, P. C. Knag, R. K. A. Krishnamurthy, *IEEE J. of Solid-State Circuits* **2019**, *54*, 992.
[153] *Memristive Devices for Brain-Inspired Computing: From Materials, Devices, and Circuits to Applications-Computational Memory, Deep Learning, and Spiking Neural Networks* (eds S. Spiga, A. Sebastian, D. Querlioz, B. Rajendran), Woodhead Publishing, Swaston, Cambridge **2020**.
[154] S. H. Jo, et al., *Nano Lett.* **2010**, *10*, 1297.
[155] Z. Wang, S. Joshi, S. E. Savel'ev, H. Jiang, R. Midya, P. Lin, M. Hu, N. Ge, J. P. Strachan, Z. Li, Q. Wu, M. Barnell, G.-L. Li, H. L. Xin, R. S. Williams, Q. Xia, J. J. Yang, *Nat. Mater.* **2017**, *16*, 101.
[156] K. Zarudnyi, et al., *Front. Neurosci.* **2018**, *12*, 57.
[157] T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, E. Eleftheriou, *Nat. Nanotechnol.* **2016**, *11*, 693.
[158] S. Diaz-Pier, M. Naveau, M. Butz-Ostendorf, A. Morrison, *Front. Neuroanatomy* **2016**, *10*, 57.
[159] I. Gupta, et al., *Nature Communications* **2016**, *7*, 1.
[160] M. D. Pickett, G. Medeiros-Ribeiro, R. S. Williams, *Nat. Mater.* **2013**, *12*, 114.
[161] S. Kumar, R. S. Williams, Z. Wang, *Nature* **2020**, *585*, 518.
[162] J. Grollier, et al., *Nature Electron.* **2020**, *3*, 360.
[163] K. Y. Camsari, B. M. Sutton, S. Datta, *Appl. Phys. Rev.* **2019**, *6*, 011305.
[164] K. Yogendra, D. Fan, K. Roy, *IEEE Trans. Magnet.* **2015**, *51*, 1.
[165] K. Yogendra, D. Fan, B. Jung, K. Roy, *IEEE Trans. on Electron Dev.* **2016**, *63*, 1674.
[166] H. Farkhani, et al. in *14th Int. Conf. on Design Technology of Integrated Systems In Nanoscale Era (DTIS)*, **2019**), pp. 1–5.
[167] A. Mizrahi, et al. *Nat. Commun.* **2018**, *9*.
[168] Y. Huang, W. Kang, X. Zhang, Y. Zhou, W. Zhao *Nanotechnology* **2017**, *28*.
[169] S. Li, et al. *Nanotechnology* **2017**, *28*.
[170] J. Grollier, D. Querlioz, M. D. Stiles, *Proc. IEEE* **2016**, *104*, 2024.
[171] X. Chen, et al., *Nanoscale* **2018**, *10*, 6139.
[172] M.-C. Chen, A. Sengupta, K. Roy, *IEEE Trans. Magnet.* **2018**, *54*, 1.
[173] B. J. Shastri, et al., *Nat. Photonics* **2021**, *15*, 102.
[174] T. Ahmed, et al., *Small* **2019**, *15*, 1900966.
[175] D. Xiang, et al. *Nat. Commun.* **2018**, *9*.
[176] S. Seo, et al. *Nat. Commun.* **2018**, *9*.
[177] W. Xue, W. Ci, X.-H. Xu, G. Liu, *Chin. Phys. B* **2020**, *29*, 048401.
[178] T.-Y. Wang, et al., *Nano Energy* **2021**, *89*, 106291.
[179] A. Mehonic, T. Gerard, A. J. Kenyon, *Appl. Phys. Lett.* **2017**, *111*, 233502.
[180] L. Hu, et al., *Adv. Funct. Mater.* **2021**, *31*, 2005582.
[181] J. T. Connor, R. D. Martin, L. E. Atlas, *IEEE Trans. Neural Networks* **1994**, *5*, 240.
[182] R. Pascanu, T. Mikolov, Y. Bengio, in *Int. Conf. on Machine Learning*, PMLR, Atlanta, Georgia, USA **2013**, pp. 1310–1318.
[183] H. Jaeger, *GMD Tech. Rep.* **2001**, *148*, 13.
[184] Y. Zhong, et al. *Nat. Commun.* **2021**, *12*, 1.
[185] W. Jiang, et al., *Appl. Phys. Lett.* **2019**, *115*, 192403.
[186] C. Du, et al., *Nat. Commun.* **2017**, *8*, 1.
[187] R. Nakane, G. Tanaka, A. Hirose, *IEEE Access* **2018**, *6*, 4462.
[188] G. Bourianoff, D. Pinna, M. Sitte, K. Everschor-Sitte, *AIP Adv.* **2018**, *8*, 055602.
[189] D. Prychynenko, et al., *Phys. Rev. Appl.* **2018**, *9*.
[190] D. Pinna, G. Bourianoff, K. Everschor-Sitte, *Phys. Rev. Appl.* **2020**, *14*.
[191] comma.ai, openpilot GitHub page, **2022**. https://web.archive.org/web/20220212070443/https://github.com/commaai/openpilot (accessed: February, 2022).
[192] Tesla, Artificial Intelligence & Autopilot, **2022**. https://web.archive.org/web/20220223164345/https://www.tesla.com/en_GB/AI, (accessed: February, 2022).
[193] Y. Zhou, Y. Yu B. Ding, in *Int. Conf. on Artificial Intelligence and Computer Engineering,* **2020**, pp. 494–500.
[194] D. Chabi, W. Zhao, D. Querlioz, J. O. Klein, in *IEEE/ACM Int. Symp. on Nanoscale Architectures*, IEEE, Piscataway, NJ **2011**, pp. 137–143.
[195] B. Tian, L. Liu, M. Yan, J. Wang, Q. Zhao, N. Zhong, P. Xiang, L. Sun, H. Peng, H. Shen, T. Lin, B. Dkhil, X. Meng, J. Chu, X. Tang, C. Duan, *Adv. Electron. Mater.* **2019**, *5*, 1800600.
[196] Y. Li, et al., *Nat. Electron.* **2021**, *4*, 348.
[197] Z. Dong, C. Li, D. Qi, L. Luo, S. Duan, *IEEE Access* **2016**, *4*, 2604.
[198] C.-Y. Chen, K. Chakrabarty, in *Design, Automation & Test in Europe Conf. & Exhibition (DATE)*, IEEE, Grenoble, France **2021**, pp. 1074–1077.
[199] A. P. James, L. O. Chua, *IEEE Trans. Circuits Syst. I.* **2021**.
[200] K. V. Pham, S. B. Tran, T. V. Nguyen, K.-S. Min, *Micromachines* **2019**, *10*, 141.
[201] O. Krestinskaya, A. P. James, L. O. Chua, *IEEE Trans, Neural Networks Learn. Syst.* **2019**, *31*, 4.
[202] C. Lee, G. Ahmed, *Int. J. Technol. Innovat. Manag.* **2021**, *1*, 18.
[203] T. Jungwirth, X. Marti, P. Wadley, J. Wunderlich, *Nat. Nanotechnol.* **2016**, *11*, 231.
[204] X. Lin, W. Yang, K. L. Wang, W. Zhao, *Nat. Electron.* **2019**, *2*, 274.
[205] H. Kurebayashi, J. H. Garcia, S. Khan, J. Sinova, S. Roche, *Nat. Rev. Phys.* **2022**, *4*, 150.
[206] D. Pesin, A. H. MacDonald, *Nat. Mater.* **2012**, *11*, 409.
[207] J. Kim, et al. *Proc. IEEE* **2014**, *103*, pp. 106.
[208] S. Matsunaga, et al. in *Design, Automation & Test in Europe Conf. & Exhibition*, **2009**, pp. 433–435.
[209] S. Tehrani, et al., *IEEE Trans. Magnet.* **1999**, *35*, 2814.
[210] S. Verma, A. A. Kulkarni, B. K. Kaushik, *IEEE Nanotechnol. Mag.* **2016**, *10*, 13.
[211] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, J. Grollier, *Nature* **2017**, *547*, 428.
[212] M. Romera, et al., *Nat. Commun.* **2022**, *13*, 1.
[213] Christensen, D. V. et al. *Neuromorph. Comput. Eng.* **2022**, https://doi.org/10.1088/2634-4386/ac4a83.
[214] M. Chhowalla, D. Jena, H. Zhang, *Nat. Rev. Mater.* **2016**, *1*, 16052.
[215] C. J. Benjamin, S. K. Zhang, Z. H. Chen, *Nanoscale* **2018**, *10*, 5148.
[216] C. R. Dean, et al., *Nature* **2013**, *497*, 598.
[217] C. R. Dean, et al., *Nat. Nanotechnol.* **2010**, *5*, 722.
[218] Y. Y. Wang, et al., *J. Appl. Phys.* **2020**, *127*, 214303.
[219] T. Jin, et al., *ACS Appl. Mater. Interfaces* **2021**, *13*, 10639.