

Review

# Memristive and CMOS Devices for Neuromorphic Computing

Valerio Milo , Gerardo Malavena, Christian Monzio Compagnoni  and Daniele Ielmini \*

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano and Italian Universities Nanoelectronics Team (IU.NET), Piazza L. da Vinci 32, 20133 Milano, Italy; valerio.milo@polimi.it (V.M.); gerardo.malavena@polimi.it (G.M.); christian.monzio@polimi.it (C.M.C.)

\* Correspondence: daniele.ielmini@polimi.it

Received: 28 November 2019; Accepted: 18 December 2019; Published: 1 January 2020



**Abstract:** Neuromorphic computing has emerged as one of the most promising paradigms to overcome the limitations of von Neumann architecture of conventional digital processors. The aim of neuromorphic computing is to faithfully reproduce the computing processes in the human brain, thus paralleling its outstanding energy efficiency and compactness. Toward this goal, however, some major challenges have to be faced. Since the brain processes information by high-density neural networks with ultra-low power consumption, novel device concepts combining high scalability, low-power operation, and advanced computing functionality must be developed. This work provides an overview of the most promising device concepts in neuromorphic computing including complementary metal-oxide semiconductor (CMOS) and memristive technologies. First, the physics and operation of CMOS-based floating-gate memory devices in artificial neural networks will be addressed. Then, several memristive concepts will be reviewed and discussed for applications in deep neural network and spiking neural network architectures. Finally, the main technology challenges and perspectives of neuromorphic computing will be discussed.

**Keywords:** neuromorphic computing; Flash memories; memristive devices; resistive switching; synaptic plasticity; artificial neural network; spiking neural network; pattern recognition

## 1. Introduction

The complementary metal-oxide semiconductor (CMOS) technology has sustained tremendous progress in communication and information processing since the 1960s. Thanks to the continuous miniaturization of the metal-oxide semiconductor (MOS) transistor according to the Moore's law [1] and Dennard scaling rules [2], the clock frequency and integration density on the chip have seen an exponential increase. In the last 15 years, however, the Moore's scaling law has been slowed down by two fundamental issues, namely the excessive subthreshold leakage currents and the increasing heat generated within the chip [3,4]. To overcome these barriers, new advances have been introduced, including the adoption of high-k materials as the gate dielectric [5], the redesign of the transistor with multigate structures [6,7], and 3D integration [8]. Besides the difficult scaling, another crucial issue of today's digital computers is the physical distinction between the central processing unit (CPU)

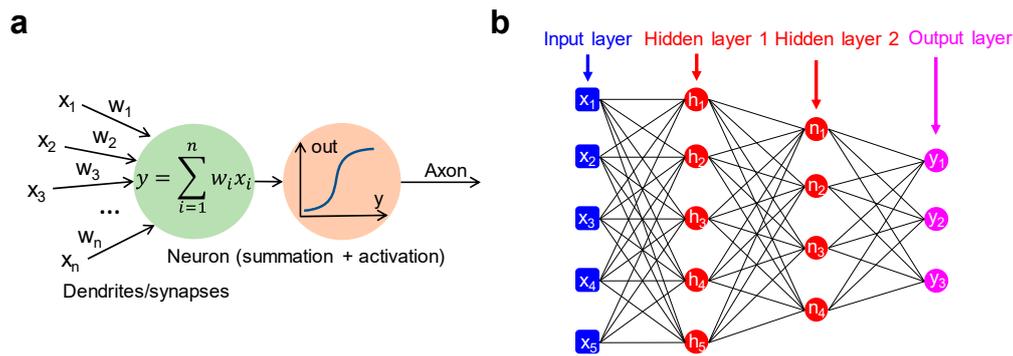
and the memory unit at the origin of extensive data movement during computation, especially for data-intensive tasks [9]. Solving the memory bottleneck requires a paradigm shift in architecture, where computation is executed in situ within the data by exploiting, e.g., the ability of memory arrays to implement matrix-vector multiplication (MVM) [10,11]. This novel architectural approach is referred to as in-memory computing, which provides the basis for several outstanding applications, such as pattern classification [12,13], analogue image processing [14], and the solution of linear systems [15,16] and of linear regression problems [17].

In this context, neuromorphic computing has been receiving increasing interest for its ability to mimic the human brain. A neuromorphic circuit consists of a network of artificial neurons and synapses capable of processing sensory information with massive parallelism and ultra-low power dissipation [18]. The realization of scalable, high density, and high-performance neuromorphic circuits generally requires the extensive adoption of memory devices serving the role of synaptic links and/or neuron elements. The device structure and operation of these memory devices may require specific optimization for neuromorphic circuits.

This work reviews the current status of neuromorphic devices, with a focus on both CMOS and memristive devices for implementation of artificial synapses and neurons in both deep neural networks (DNNs) and spiking neural networks (SNNs). The paper is organized as follows: Section 2 provides an overview of the major neuromorphic computing concepts from a historical perspective. Section 3 is an overview of the operating principles of mainstream NAND and NOR Flash technologies, and their adoption in neuromorphic networks. Section 4 describes the most important memristive concepts being considered for neuromorphic computing applications. Section 5 addresses the adoption of memristive devices in DNNs and SNNs for hardware demonstration of cognitive functions, such as pattern recognition and image/face classification. Finally, Section 6 discusses issues and future perspectives for large-scale hardware implementation of neuromorphic systems with CMOS/memristive devices.

## 2. Neuromorphic Computing Concepts

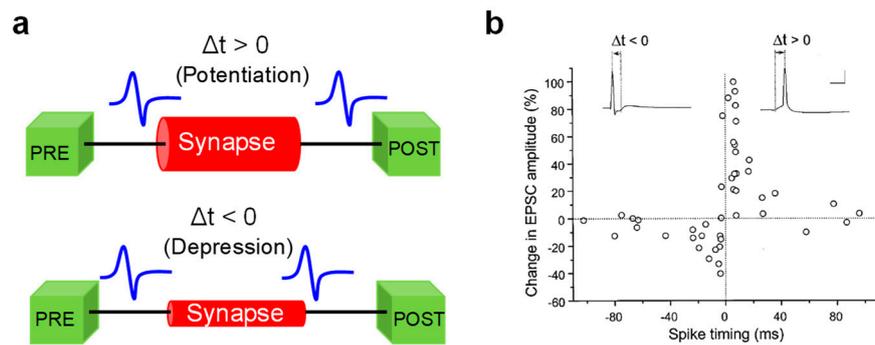
The origin of neuromorphic computing can be traced back to 1949, when McCulloch and Pitts proposed a mathematical model of the biological neuron. This is depicted in Figure 1a, where the neuron is conceived as a processing unit, operating (i) a summation of input signals ( $x_1, x_2, x_3, \dots$ ), each multiplied by a suitable synaptic weight ( $w_1, w_2, w_3, \dots$ ) and (ii) a non-linear transformation according to an activation function, e.g., a sigmoidal function [19]. A second landmark came in 1957, when Rosenblatt developed the model of a fundamental neural network called multiple-layer perceptron (MLP) [20], which is schematically illustrated in Figure 1b. The MLP consists of an input layer, one or more intermediate layers called hidden layers, and an output layer, through which the input signal is forward propagated toward the output. The MLP model constitutes the backbone for the emerging concept of DNNs. DNNs have recently shown excellent performance in tasks, such as pattern classification and speech recognition, via extensive supervised training techniques, such as the backpropagation rule [21–23]. DNNs are usually implemented in hardware with von Neumann platforms, such as the graphics processing unit (GPU) [24] and the tensor processing unit (TPU) [25], used to execute both training and inference. These hardware implementations, however, reveal all the typical limitations of the von Neumann architecture, chiefly the large energy consumption in contrast with the human brain model.



**Figure 1.** (a) Conceptual illustration of McCulloch and Pitts artificial neuron architecture, where the weighted sum of the input signals is subject to the application of a non-linear activation function yielding the output signal. (b) Schematic representation of a multilayer perceptron consisting of two hidden layers between the input and the output layer.

To significantly improve the energy efficiency of DNNs, MVM in crossbar memory arrays has emerged as a promising approach [26,27]. Memory devices also enable the implementation of learning schemes able to replicate the biological synaptic plasticity at the device level. CMOS memories, such as the static random access memory (SRAM) [28,29] and the Flash memory [30], were initially adopted to capture synaptic behaviors in hardware. In the last 10 years, novel material-based memory devices, generically referred to as memristors [31], have evidenced attractive features for the implementation of neuromorphic hardware, including non-volatile storage, low power operation, nanoscale size, and analog resistance tunability. In particular, memristive technologies, which include resistive switching random access memory (RRAM), phase change memory (PCM), and other emergent memory concepts based on ferroelectric and ferromagnetic effects, have been shown to achieve synapse and neuron functions, enabling the demonstration of fundamental cognitive primitives as pattern recognition in neuromorphic networks [32–35].

The field of neuromorphic networks includes both the DNN [36], and SNN, the latter more directly inspired by the human brain [37]. Contrary to DNNs, the learning ability in SNNs emerges via unsupervised training processes, where synapses are potentiated or depressed by bio-realistic learning rules inspired by the brain. Among these local learning rules, spike-timing-dependent plasticity (STDP) and spike-rate-dependent plasticity (SRDP) have received intense investigation for hardware implementation of brain-inspired SNNs. In STDP, which was experimentally demonstrated in hippocampal cultures by Bi and Poo in 1998 [38], the synaptic weight update depends on the relative timing between the presynaptic spike and the post-synaptic spike (Figure 2a). In particular, if the pre-synaptic neuron (PRE) spike precedes the post-synaptic neuron (POST) spike, namely the relative delay of spikes,  $\Delta t = t_{\text{post}} - t_{\text{pre}}$ , is positive, then the interaction between the two spikes causes the synapse to increase its weight, which goes under the name of synaptic potentiation. On the other hand, if the PRE spike follows the POST spike, i.e.,  $\Delta t$  is negative, then the synapse undergoes a weight decrease or synaptic depression (Figure 2b). In SRDP, instead, the rate of spikes emitted by externally stimulated neurons dictates the potentiation or depression of the synapse, with high and low frequency stimulation leading to synaptic potentiation and depression, respectively [39]. Unlike STDP relying on pairs of spikes, SRDP has been attributed to the complex combination of three spikes (triplet) or more [40–43]. In addition to the ability to learn in an unsupervised way and emulate biological processes, SNNs also offer a significant improvement in energy efficiency thanks to the ability to process data by transmission of short spikes, hence consuming power only when and where the spike occurs [18]. Therefore, CMOS and memristive concepts can offer great advantages in the implementation of both DNNs and SNNs, providing a wide portfolio of functionalities, such as non-volatile weight storage, high scalability, energy efficient in-memory computing via MVM, and online weight adaptation in response to external stimuli.

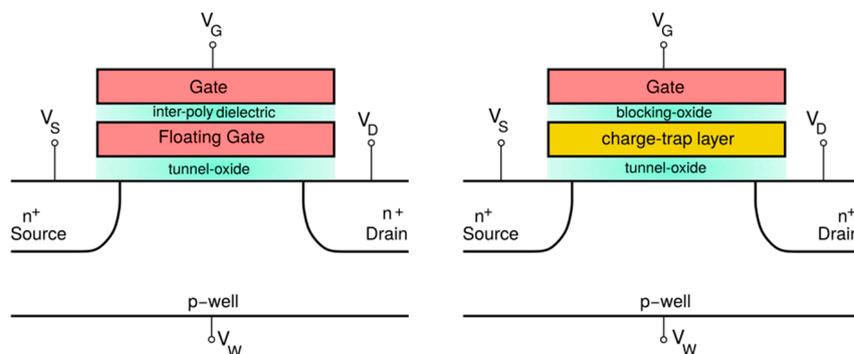


**Figure 2.** (a) Sketch of the spike-timing-dependent plasticity (STDP) learning rule. If the PRE spike arrives just before the POST spike at the synaptic terminal ( $\Delta t > 0$ ), the synapse undergoes a potentiation process, resulting in a weight (conductance) increase (**top**). Otherwise, if the PRE spike arrives just after the POST spike ( $\Delta t < 0$ ), the synapse undergoes a depression process, resulting in a weight (conductance) decrease (**bottom**). (b) Relative change of synaptic weight as a function of the relative time delay between PRE and POST spikes measured in hippocampal synapses by Bi and Poo. Reprinted with permission from [38]. Copyright 1998 Society for Neuroscience.

### 3. Mainstream Memory Technologies for Neuromorphic and Brain-Inspired Systems

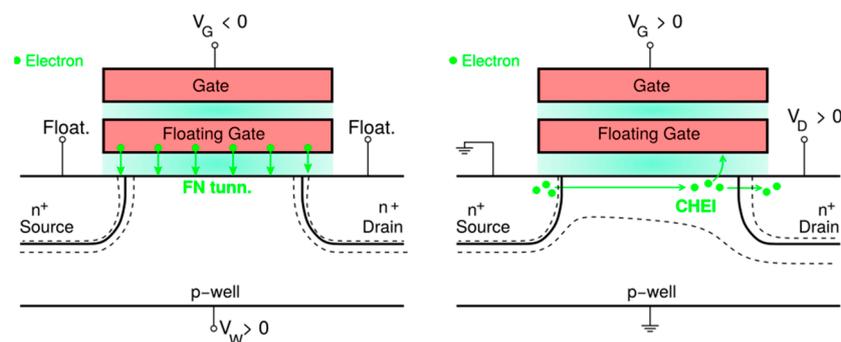
#### 3.1. Memory Transistors and Mainstream Flash Technologies

The memory transistor represents the elementary building unit at the basis of modern mainstream non-volatile storage technologies. It consists of a mainstream MOS transistor whose structure is modified to accommodate a charge-storage layer in its gate stack, allowing carriers to be confined in a well-defined region due to the resulting potential barriers. As shown in Figure 3, the most adopted solutions for such a layer are based either on highly doped polycrystalline silicon (polysilicon) or a dielectric material able to capture and release electrons and holes thanks to its peculiar high density of defects. The charge storage layer is usually referred to as floating gate in the former case, and charge-trap layer in the latter one. However, in both cases, storing a net charge in the memory transistor floating gate or charge-trap layer results in a shift of the drain current vs. gate voltage ( $I_{DS} - V_{GS}$ ) curve due to the corresponding variation of the device threshold voltage ( $V_T$ ). In particular, such variation is mainly ruled by the capacitance between the transistor gate and the charge-storage layer,  $C_{sg}$ , according to  $\Delta V_T = -Q_s / C_{sg}$ , meaning that a net positive or negative stored charge ( $Q_s$ ) is reflected in a negative or positive  $V_T$  shift ( $\Delta V_T$ ), respectively. As a consequence, a proper discretization of the stored charge in each memory transistor allows one or multiple bits of information to be stored that can be accessed through a  $V_T$  read operation.



**Figure 3.** Schematic of a memory cell exploiting (left) a highly doped polysilicon layer and (right) a dielectric layer with a high density of microscopic defects for charge storage.

In order to reliably accomplish the tuning of the stored charge and, consequently, the modification of the information content through the program (making the stored charge more negative) and erase (making the stored charge more positive) operations, suitable physical mechanisms must be selected. As schematically depicted in Figure 4, the most widely adopted physical mechanisms are the Fowler–Nordheim (FN) tunneling, for both program and erase operations, and the channel hot electron injection (CHEI), for program operation only. In the former case, the bias voltages applied to the memory transistor contacts are chosen to generate large vertical electric fields that activate carrier exchange between the substrate and the storage layer by the quantum mechanical current through the energy barrier separating them. In the latter case, instead, CHEI is achieved by accelerating the transistor on-state current electrons by applying a positive drain-to-source voltage drop ( $V_{DS}$ ). If  $V_{DS}$  is large enough, the energy acquired by the channel electrons is sufficient for them to overcome the tunnel-oxide energy barrier and to be redirected to the charge-storage layer due to the positive  $V_{GS}$ . Moreover, it is worth mentioning that, for a target  $\Delta V_T$  to be achieved over comparable time scales, CHEI requires much lower voltages to be applied with respect to FN tunneling. On the contrary, its injection efficiency is of the order of  $10^{-5}$  only, much smaller than that of FN tunneling (very close to one). A final but important remark is that for both CHEI and FN tunneling, the maximum number of program/erase cycles that can be performed on the devices is usually smaller than  $10^5$ ; in fact, for larger cycling doses, the number of defects generated in the tunnel oxide by the program/erase operations severely undermines the transistor reliability.



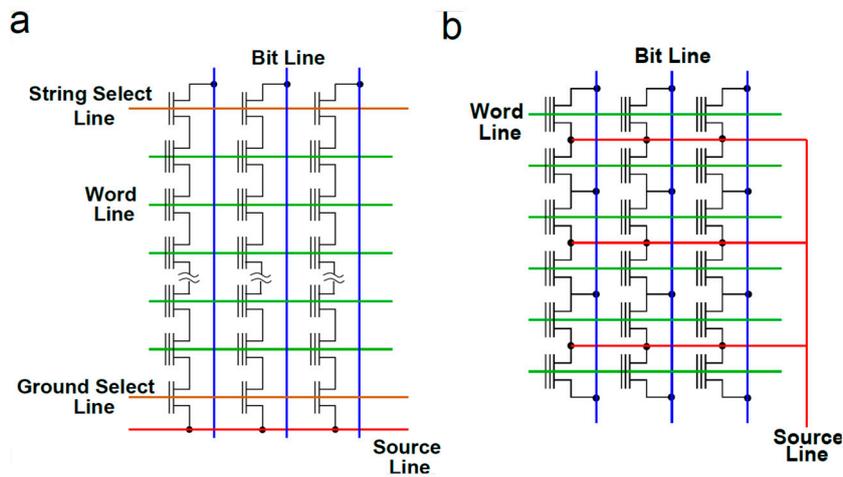
**Figure 4.** Physical mechanisms and corresponding voltage schemes exploited to change the amount of charge in the cell storage layer, consisting of (left) Fowler–Nordheim (FN) and (right) channel hot-electron injection (CHEI).

Starting from the schematic structure shown in Figure 3, the arrangement of memory transistors to build memory arrays and their working conditions are strictly related to the specific targeted application. In particular, two solutions that have ruled the non-volatile memory market since their very first introduction are the NAND Flash [44] and NOR Flash [45] architectures (Figure 5). Although they share the important peculiarity that the erase operation, exploiting FN tunneling to reduce the amount of the stored negative charge, involves a large number of cells at the same time (a block of cell), some relevant differences can be mentioned.

NAND Flash technology is the main solution for the storage of large amounts of data, therefore achieving large bit storage density, i.e., the ratio between the chip capacity and its area is a mandatory requirement. For this purpose, NAND Flash memory transistors are deeply scaled (up to a feature size as small as 15 nm) and arranged in series connection, making the memory cells belonging to each string accessible only through the contacts at their top and bottom ends (Figure 5a). In such a way, the area occupancy of each cell is minimized; on the other hand, the attempt to minimize the array fragmentation and to reduce the area occupancy of the control circuitry makes the random access time to the cells quite long (tens of  $\mu\text{s}$ ), due to the consequent delays of the signals propagating over the long WLs and BLs. For this reason, programming schemes taking advantage of the low current and

high injection efficiency of FN tunneling were developed to program many memory transistors at the same time, allowing extremely high throughputs (tens or even hundreds of Mbytes/s) to be achieved.

The NOR Flash technology, on the other hand, is mainly intended for code storage, making the storage and retrieval of small packets of data (a few bytes) as fast as possible a mandatory requirement. As a consequence, in order to make each memory cell directly accessible through dedicated contacts, the memory transistors are connected in parallel, as shown in Figure 5b. Thanks to this architecture, a fast and single-cell selective program operation can be easily achieved exploiting CHEI. From the cell design standpoint, this results in a limited channel scalability, due to the need for the cell to withstand relatively high  $V_{DS}$  during its operation. Even though these features determine a larger cell footprint and, in turn, a higher cost of NOR Flash with respect to NAND Flash technologies, they allow NOR Flash arrays to guarantee a superior array reliability, being an important requirement for code storage applications.



**Figure 5.** Schematic of memory arrays based on (a) NAND Flash and (b) NOR Flash architecture.

### 3.2. Memory Transistors as Synaptic Devices in Artificial Neural Networks

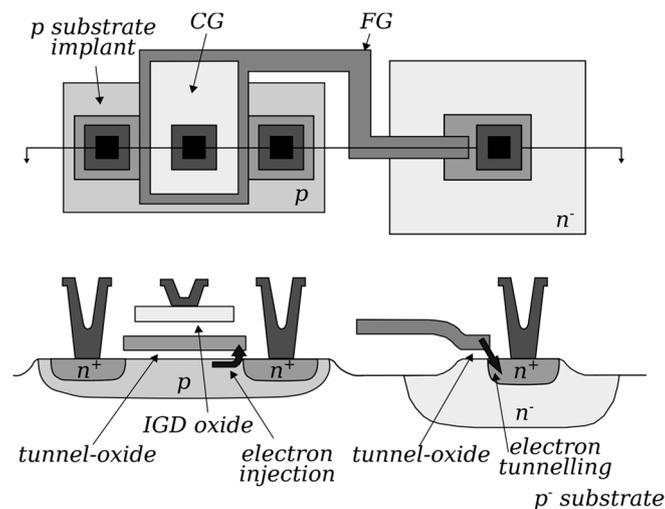
The first proposal of exploiting memory transistors as artificial synapses in artificial neural networks (ANNs) and brain-inspired neural networks dates back to the 1990s directly from the pioneering work presented in ref. [46]. The basic idea proposed there is to take advantage of the subthreshold characteristic  $I_{DS} - V_{GS}$  of an n-channel floating-gate memory transistor to reproduce the biologically observed synaptic behavior and to exploit it to build large-scale neuromorphic systems. In fact, when operated in a subthreshold regime, a memory transistor exhibits an  $I_{DS} - V_{GS}$  relation that can be expressed as:

$$I_{DS} = I_0 \cdot \exp\left[\frac{q\alpha_G(V_{GS} - V_T^{ref})}{mkT}\right] \cdot \exp\left[\frac{-q\alpha_G\Delta V_T}{mkT}\right], \quad (1)$$

where  $I_0$  is the current pre-factor,  $q$  is the elementary charge,  $m$  is the subthreshold slope ideality factor,  $kT$  is the thermal energy,  $\alpha_G$  is the gate-to-floating-gate capacitive coupling ratio, and  $\Delta V_T$  is the floating-gate transistor  $V_T$  shift from an arbitrary chosen  $V_T^{ref}$ .

With reference to the previous equation,  $I_{DS}$  can be decomposed in the product of two contributions. The first factor,  $I_0 \cdot \exp\left[\frac{q\alpha_G(V_{GS} - V_T^{ref})}{mkT}\right]$ , is a function of  $V_{GS}$  only, and represents the input presynaptic signal; the remaining scaling factor,  $W = \exp\left[\frac{-q\alpha_G\Delta V_T}{mkT}\right]$ , instead, depending on  $\Delta V_T$  but not on  $V_{GS}$ , can be thought of as the synaptic weight.

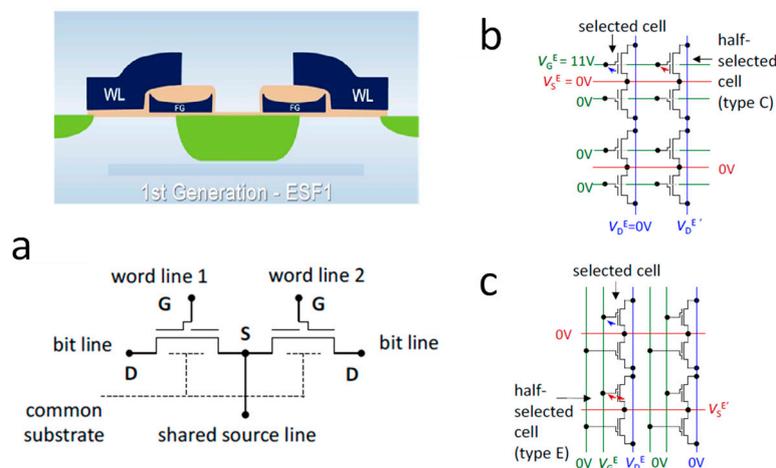
When compared with other modern approaches based on emerging memory technologies, this solution presents the clear advantages of (i) limited power consumption, thanks to the reduced currents peculiar of transistors operated below the threshold; (ii) fine weight granularity, coming to the virtually analog and bidirectional  $V_T$  tuning; and (iii) a mature and well-established CMOS fabrication technology. In particular, the relevance of the last point can be easily understood by considering the possibility of arranging a large number of floating-gate transistors in very dense and reliable memory arrays, normally employed for storage purposes. However, when exploited as synaptic arrays in neuromorphic applications, such memory arrays must meet the mandatory condition of single-cell selectivity during both program and erase operations, meaning that both the positive and negative tuning of the  $V_T$  (weight) of each memory cell (synapse) must be guaranteed. Even if this consideration makes a NOR-type array inherently more suitable to be used in these fields because of its architecture that allows direct access to each cell by dedicated contacts, its standard block-erase scheme must still be overcome. For this reason, since its very first proposal, the synaptic transistor introduced in refs. [46–48], and tested on LTD and LTP based on the STDP learning rule in refs. [30,48], includes an additional contact with respect to standard n-channel floating-gate transistors (Figure 6) to be connected to signal lines running orthogonal to the WLs [46]. While keeping CHEI for the program, the erase operation takes place by removing stored electrons by FN tunneling when a sufficiently high electric field is developed between the tunneling contact and the transistor floating gate that, as shown in Figure 3, is properly extended in close proximity of such a contact. Note that this erase scheme is indeed single-cell selective because the substrate contact, common to all the array cells, is kept to the ground.



**Figure 6.** Top view (**up**) and side view (**down**) of the synaptic transistor. Physical mechanisms exploited for program (electron injection) and erase (electron tunneling) are highlighted too. Adapted with permission from [48]. Copyright 1997, IEEE.

Although, recently, some more effort was devoted to build new custom synaptic devices and test them in SNNs [49–51], a more convincing proof of the feasibility of the floating-gate transistor to build large-scale neuromorphic systems comes from a different approach. The basic idea consists in slightly modifying the routing of commercially available NOR Flash memory arrays to enable a single-cell selective erase operation while keeping the cell structure unchanged. For this purpose, NOR memory arrays developed with a 180 nm technology by Silicon Storage Technology, Inc. (SST) [52] are chosen in refs. [53–56]. The basic memory cell, as depicted in Figure 7a, features a highly asymmetric structure presenting a floating gate only near the source side, with the gate stack at the drain side made only of the tunneling oxide. In spite of this structure, the program operation can still be performed by CHEI at

the source side; as for the erase operation, instead, a positive voltage is applied between the gate and source, resulting in the emission of stored electrons toward the gate by FN tunneling.



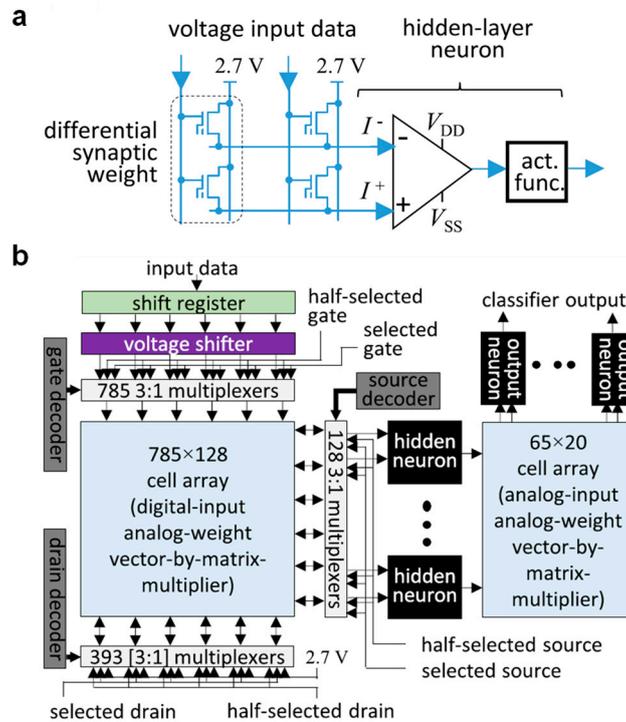
**Figure 7.** (a) Schematic cross-section of the Silicon Storage Technology (SST) cell structure (**top**) and its equivalent circuit (**bottom**) and NOR array with (b) classic and (c) modified routing together with the respective erase protocol. Reprinted with permission from [53]. Copyright 2015, IEEE.

The arrangement of such SST cells to make a NOR array is shown in Figure 7b, where the erase voltages are highlighted too. Since both WLs and SLs run parallel to each other and orthogonal to the BLs, the erase protocol involves all the cells in a row at the same time. For this reason, in refs. [54], a modification to the array routing as reported in Figure 7c is proposed, with the WLs now running parallel to the BLs. In this way, single-cell selectivity is achieved during both the program (involving WL, BL, and SL) and erase (involving WL and SL only).

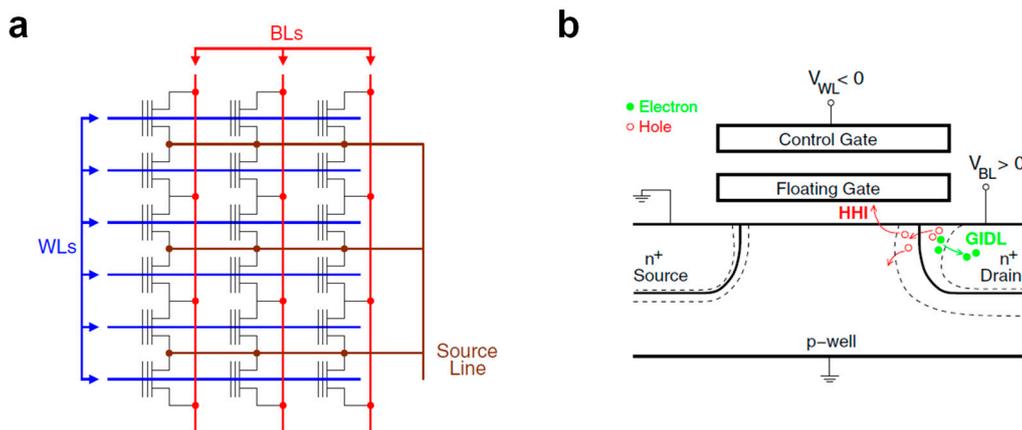
In refs. [54,55], two SST NOR arrays, re-routed as explained before, are employed to build and test a fully integrated three-layer ( $784 \times 64 \times 10$ ) ANN, trained offline on the Modified National Institute of Standards and Technology (MNIST) database for handwritten digit recognition via the backpropagation algorithm [21–23]. In particular, in order to enable the implementation of negative weights, and also to reduce random drifts and temperature sensitivity, a differential solution is adopted. As shown in Figure 8a, following this approach, each couple of adjacent memory cells implements a synaptic weight, with the resulting BL currents summed and read by CMOS artificial neurons built exploiting a differential current operational amplifier. The whole one-chip integrated network, whose schematic structure, including two synaptic arrays together with two neuron layers and some additional circuitry, is reported in Figure 8b, has shown a 94.7% classification fidelity with one-pattern classification time and energy equal to  $1 \mu s$  and less than 20 nJ, respectively. Moreover, a reduction of the total chip active area, amounting to  $1 \text{ mm}^2$  in the discussed work, is expected together with an increase of its performance when moving to the next 55 nm SST technology. In this regard, some preliminary results about MVM were already presented in ref. [56].

Although this solution based on re-routing commercially available NOR arrays appears promising, it comes together with its main drawback consisting in the increased area occupancy (the single-cell area in the modified array is 2.3 times larger than the original one). A different approach aiming at avoiding this disadvantage is proposed in [57–59]. Here, the authors suggest a modified working scheme for a mainstream double-polysilicon common-ground NOR Flash arrays developed in a 40 nm embedded technology by STMicroelectronics (Figure 9a) without any change needed in the cell or array design. While keeping CHEI as the physical mechanism for the program, single-cell selectivity during the erase is achieved by employing hot-hole injection (HHI) in the cell floating gate. In particular, by keeping the source and substrate contacts to the ground while applying a positive and negative voltage to the drain and to the gate, respectively, the developed electric field triggers the generation of holes by

band-to-band tunneling at the drain side and accelerates them (Figure 9b); if the applied voltages are high enough, the energy acquired by the holes allows them to overcome the energetic barrier of the tunnel oxide and to be redirected toward the floating gate thanks to the negative gate voltage.



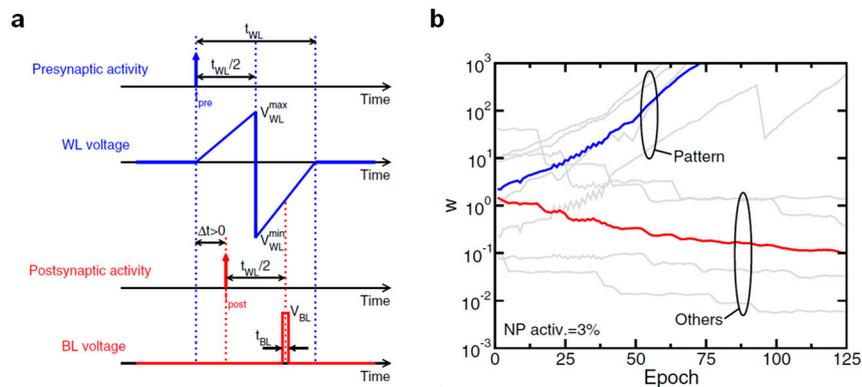
**Figure 8.** (a) Differential implementation of a synaptic connection followed by a hidden-layer neuron, consisting of a differential summing operational amplifier and an activation-function block. (b) High-level architecture of the artificial neural network and needed additional circuitry. Reprinted with permission from [55]. Copyright 2018, IEEE.



**Figure 9.** (a) Schematic for a mainstream common-ground NOR Flash array and (b) proposed physical mechanism exploited for the erase operations. Reprinted with permission from [57]. Copyright 2018, IEEE.

To validate this program/erase scheme in a brain-inspired neural network, the authors demonstrated long-term potentiation/depression through the design of the presynaptic and postsynaptic waveforms as shown in Figure 10a. The short rectangular pulse applied to the BL as a consequence of a postsynaptic fire event overlaps with a positive or negative WL voltage according to the time distance between the presynaptic and postsynaptic spike,  $\Delta t$ . In particular,  $\Delta t > 0$  leads to long-term

potentiation by HHI and  $\Delta t < 0$  leads to long-term depression by CHEI. To further confirm the validity of this protocol, a prototype two layers  $8 \times 1$  SNN was tested on pattern recognition, producing encouraging results as shown in Figure 10b; in fact, as expected, while the synapses corresponding to the input pattern are quickly potentiated, the remaining ones are gradually depressed.



**Figure 10.** (a) Pulse scheme proposed to implement the spike-timing-dependent plasticity (STDP) waveform exploiting the erase mechanism shown in Figure 9b and (b) evolution of the weights of the implemented NOR Flash-based spiking neural network during the learning phase. Reprinted with permission from [57]. Copyright 2018, IEEE.

A final remark, being of great relevance especially in DNN inference, is the finite tuning precision of the cells array,  $V_T$ , and its stability after the offline training phase. In the case of ANN based on NOR Flash memory arrays, two of the most relevant physical mechanisms causing reliability issues of this kind are program noise (PN), determining an inherent uncertainty during the program phase due to the statistical nature of electron injection in the floating gate, and random telegraph noise (RTN), inducing  $V_T$  instabilities arising from the capture and release of charge carriers in tunnel-oxide defects. In ref. [60], the authors assess the impact of both PN and RTN on a neuromorphic digit classifier through parametric Monte-Carlo simulations. The main result, relevant in terms of projection of the previously discussed results on future technological nodes, is that such non-idealities play a non-negligible role, setting a stringent requirement both on the maximum scalability of the array cell and on the adopted program/erase schemes.

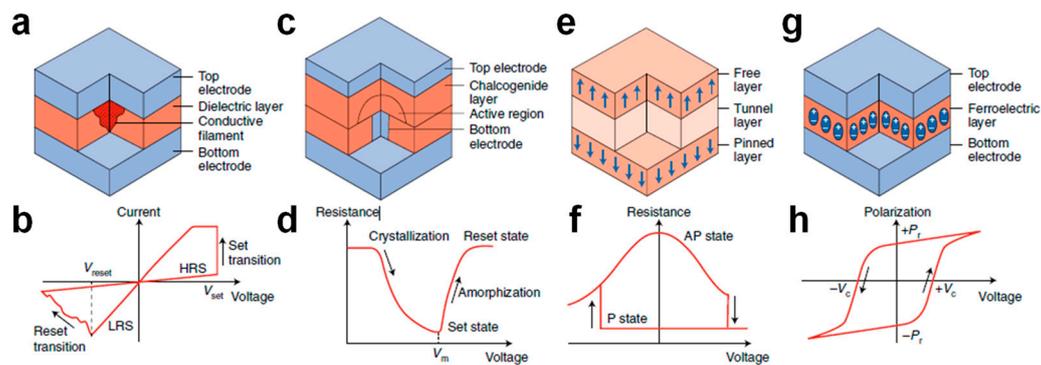
#### 4. Memristive Technologies

To replicate neural networks in hardware, memristive devices have been recently investigated for the realization of compact circuits capable of emulating neuron and synapse functionalities. Increasing interest toward these novel device concepts first results from their ability to store information at the nanoscale in an analogue and non-volatile way. Also, they allow the memory to be combined with the computing function, enabling in-situ data processing, also referred to as in-memory computing [11], which is currently the major approach toward the achievement of energy-efficient computing paradigms beyond the von Neumann bottleneck. In detail, the landscape of memristive technologies can be divided into the classes of memristors with two or three terminals, which are explained in the following subsections.

##### 4.1. Memristive Devices with 2-Terminal Structure

As shown in Figure 11, the class of memristive devices with a two-terminal structure covers various physical concepts, such as resistive switching random access memory (RRAM), phase change memory (PCM), spin-transfer torque magnetic random access memory (STT-MRAM), and ferroelectric random access memory (FeRAM), which share a very simple structure consisting of a metal-insulator-metal (MIM) stack, where an insulating layer is sandwiched between two metallic electrodes called the top

electrode (TE) and bottom electrode (BE), respectively. As a voltage pulse is applied, these devices undergo a change of physical properties of the material used as the switching layer, which results in a change of the resistance for RRAM and PCM, magnetic polarization for STT-MRAM, and electrical polarization for FeRAM. Importantly, all these memristive elements offer the opportunity to read, write, and erase the information in memory states by electrical operations on the device, thus making them potentially more attractive in terms of scalability than other memory concepts, such as the Flash memories based on charge storage.



**Figure 11.** Sketch of the most promising two-terminal memristive devices used in neuromorphic computing applications. (a) Structure of resistive switching random access memory (RRAM) device where the insulating switching layer is sandwiched between two metal electrodes. (b) Current-voltage characteristics of RRAM displaying that the application of a positive voltage causes an abrupt resistance transition, called set, leading the device from the high resistance state (HRS) to the low resistance state (LRS) while the application of a negative voltage causes a more gradual resistance transition, called reset, leading the device from LRS to HRS. (c) Structure of phase change memory (PCM) device where a chalcogenide active layer is sandwiched between two metal electrodes. (d) Resistance-voltage characteristics of PCM displaying that the crystallization process in the active layer gradually leading the PCM from HRS to LRS is achieved at voltages below the melting voltage,  $V_m$ , while the amorphization process gradually leading the PCM from LRS to HRS is achieved at voltages above  $V_m$ . (e) Structure of spin-transfer torque magnetic random access memory (STT-MRAM) device, where a tunnel layer is sandwiched between two ferromagnetic metal electrodes. (f) Resistance-voltage characteristics of STT-MRAM displaying two binary resistance transitions leading the device from the anti-parallel (AP) to the parallel (P) state (set) at positive voltage and from P to AP (reset) at negative voltage. (g) Structure of ferroelectric random access memory (FeRAM) device, where a ferroelectric layer is sandwiched between two metal electrodes. (h) Polarization-voltage characteristics displaying binary operation between two states with a positive residual polarization,  $+P_r$ , and a negative residual polarization,  $-P_r$ , achieved by application of a positive and negative voltage, respectively. Reprinted with permission from [11]. Copyright 2018, Springer Nature.

Figure 11a shows the MIM stack of the RRAM device, where an insulating oxide material serves as the switching layer [61–63]. To initiate the device, a preliminary electrical operation called forming is performed by application of a positive voltage at TE by causing a soft breakdown process, leading to the creation of a high conductivity path containing oxygen vacancies and/or metallic impurities, also known as a conductive filament (CF), within the oxide layer. This results in the change of the resistance of the device from the initial high resistance state (HRS) to the low resistance state (LRS). After forming, in the case of bipolar RRAM devices, the application of negative/positive voltage pulses at TE leads the device to experience reset and set transitions, respectively. The application of a negative pulse causes the rupture of CF (reset process), leading to the opening of a depleted gap via drift/diffusion migration of ion defects from BE to TE, hence to the HRS. On the other hand, the application of a positive pulse allows the gap to be filled via field-driven migration of ion defects from TE to BE, thus leading the device back to LRS (set process) [64,65]. Two resistance transitions can be noted by the current-voltage

characteristics shown in Figure 11b, which evidence both the abrupt nature of the set process due to the positive feedback loop involving the two driving forces for ion migration, namely the electric field and temperature, and the more gradual dynamics of the reset process due to the negative feedback occurring within the device as a negative pulse is applied [66]. Similar to the bipolar RRAM described in Figure 11b, which typically relies on switching layers, including  $\text{HfO}_x$  [67],  $\text{TaO}_x$  [68],  $\text{TiO}_x$  [69],  $\text{SiO}_x$  [70], and  $\text{WO}_x$  [71], the conductive-bridge random access memory (CBRAM), where metallic CFs are created/disrupted between active Cu/Ag electrodes, has also received strong interest in recent years [72]. In addition to bipolar RRAM concepts, another type of filamentary RRAM called unipolar RRAM, typically based on NiO [73–75], has been widely investigated, evidencing that pulses with the same polarity can induce both set and reset processes as a result of the key role played by Joule heating for the creation/disruption of CF [73,75]. Moreover, the RRAM concept also includes non-filamentary devices referred to as uniform RRAM, exhibiting an interface resistive switching due to the uniform change of a Schottky or tunneling barrier on the whole cell area [76]. One of the fundamental features making RRAM suitable for in-memory computing is the opportunity to modulate its resistance in an analog way, thus enabling multilevel operation via the storage of at least 3 bit [77–81]. In addition to multilevel operation, it also combines high scalability up to 10 nm in size [82] and the opportunity to achieve 3D integration [83].

Figure 11c shows the schematic structure of a PCM device, which relies on a chalcogenide material, such as  $\text{Ge}_2\text{Sb}_2\text{Te}_5$  (GST) [84], as the switching layer. Here, resistance variation arises from an atomic configuration change within the active layer from the crystalline to the amorphous phase and vice-versa via application of unipolar voltage pulses at TE [85–87]. As a voltage higher than the voltage,  $V_m$ , needed to induce the melting process within the active layer is applied across the cell, local melting takes place within the chalcogenide material, leading the device to HRS as a result of the pinning of the Fermi level at the midgap. Otherwise, if the applied voltage is below  $V_m$ , a gradual crystallization process is triggered via local Joule heating, leading PCM to LRS [88]. These physical processes can be better visualized by the resistance-voltage characteristics in Figure 11d, where the set transition displays a gradual behavior due to the gradual crystallization process induced by Joule heating while the reset transition displays faster dynamics than the set transition. Compared to RRAM, where the HRS/LRS ratio is about 10, PCM offers a higher resistance window, ranging from 100 to 1000, which makes PCM very attractive for multilevel operation as reported in [89], where a 3 bits/cell PCM device was demonstrated. Moreover, in addition to classic GST, other materials, such as GeSb [90], doped In-Ge-Te [91], and Ge-rich GST [92], have been investigated, receiving strong interest since they offer higher crystallization temperatures for enhanced retention performances.

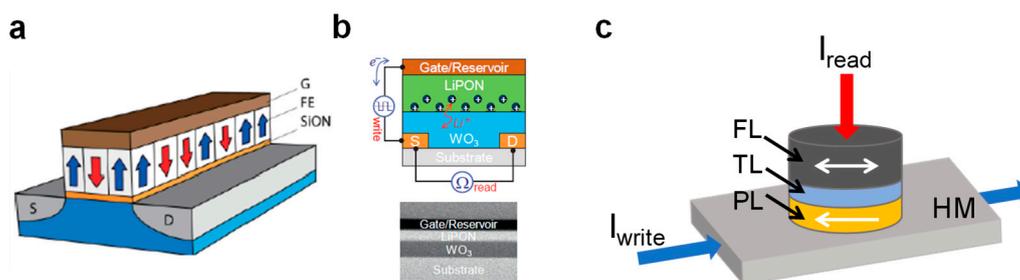
Figure 11e shows the schematic structure of an STT-MRAM device based on an MIM stack called magnetic tunnel junction (MTJ), including an ultrathin tunneling layer (TL), typically in  $\text{MgO}$ , interposed between two ferromagnetic (FM) metal electrodes, typically in CoFeB, called the pinned layer (PL) and free layer (FL), respectively [93–95]. Unlike RRAM and PCM enabling multilevel operation, STT-MRAM allows only two states to be stored, with a very small resistance window of the order of a factor 2 [94] because of the tunnel magneto-resistance (TMR) effect [96]. The two states are encoded in the relative orientation between PL magnetic polarization, which is fixed, and FL magnetic polarization, which is instead free to change via the spin-transfer torque physical mechanism discovered by Slonczewski [97] and Berger [98] in 1996. As a positive voltage is applied at TE, a current of electrons with the same spin-polarization of the fixed layer is transmitted through the tunneling layer, causing the transition of the polarization orientation from anti-parallel (AP) to parallel (P), which leads the device to LRS. In contrast, as a negative bias is applied, the reflection back of electrons entering from the free layer with the opposite magnetization takes place, thus causing the transition from the P to AP state, hence from LRS to HRS. Figure 11f shows the resistance response of the STT-MRAM device as a function of the applied voltage, evidencing that the application of positive/negative voltage pulse induces set/reset transition with very abrupt dynamics, which further supports the incompatibility of

STT-MRAM with multilevel applications. However, STT-MRAM has shown high potential in scalability, as reported in ref. [99], fast switching speed [100], and almost unlimited cycling endurance [101,102].

Figure 11g shows the MIM stack of FeRAM, where an insulating layer based on a ferroelectric (FE) material, typically in doped  $\text{HfO}_2$  [103] or perovskite materials [104,105], is sandwiched between two metal electrodes. Its operation principle relies on the polarization switching within the FE layer due to the rotation of electrical dipoles under an external bias [106]. As shown by the polarization-voltage characteristics in Figure 11h, a positive voltage above the coercive voltage,  $+V_c$ , at TE induces the set transition, leading the device to exhibit a positive residual polarization,  $+P_r$ , whereas a voltage more negative than  $-V_c$  leads the device to exhibit a negative residual polarization,  $-P_r$ . Importantly, note that the FE switching process does not impact on the device resistance, which makes FeRAM unusable as resistive memory.

#### 4.2. Memristive Devices with Three-Terminal Structure

In addition to the two-terminal devices, memristive concepts also include the class of three-terminal devices whose main examples are those depicted in Figure 12, namely (a) the ferroelectric field-effect transistor (FeFET) [107], (b) the electro-chemical random access memory (ECRAM) [108], and (c) the spin-orbit torque magnetic random access memory (SOT-MRAM) [109]. Other interesting three-terminal concepts that have been recently investigated for neuromorphic computing applications are the 2D semiconductor-based mem-transistors [110,111] and the domain-wall-based magnetic memories [112,113].



**Figure 12.** Sketch of three fundamental examples of three-terminal memristive devices. (a) Schematic structure of ferroelectric field-effect transistor (FeFET) device, where the ferroelectric switching phenomenon allows the transistor threshold voltage to be modulated, thus gradually changing the channel conductivity. (b) Schematic structure of electro-chemical random access memory (ECRAM) device, where the channel conductivity is controlled by the migration of ion species, e.g.,  $\text{Li}^+$  ions, into an electrolyte material being induced by the voltage applied at the gate terminal. (c) Schematic structure of spin-orbit torque magnetic random access memory (SOT-MRAM), where the current flow in a heavy metal (HM) line causes a polarization switching in the MTJ-free layer, resulting in a device conductance change. Reprinted with permission from [107,108]. Copyright 2017, IEEE. Copyright 2018, IEEE.

Figure 12a shows the structure of the FeFET consisting of an MOS transistor with an FE material, such as doped- $\text{HfO}_2$  [103], and perovskites [106], serving as the gate dielectric. Here, the application of external pulses at the gate terminal induces a non-volatile polarization switching within the FE dielectric, leading to a change of the transistor threshold, hence of the channel conductivity, which can be probed simply by reading the current at the drain terminal. As a result, the FeFET concept allows significant issues due to transient read currents and destructive read operation limiting FeRAM operation to be overcome. This three-terminal device has recently been operated into memory arrays with 28 nm CMOS technology [114] and exhibits a strong potential for the development of 3D structures [115]. Also, it has been operated to replicate synapse [116] and neuron [117,118] functions, which, combined with 3D integration opportunity, makes it a strong candidate for neuromorphic computing applications.

Figure 12b illustrates the device structure of the ECRAM consisting of an MOS transistor where a solid-state electrolyte based on inorganic materials, such as lithium phosphorous oxynitride (LiPON) [108,119], or organic materials, such as poly (3, 4-ethylenedioxythiophene):polystyrene sulfonate (PEDOT:PSS) [120], is used as the gate dielectric. Its operation relies on the intercalation/de-intercalation of ions in a channel layer to tune the device conductance. As reported in [108], the intercalation of  $\text{Li}^+$  ions into the  $\text{WO}_3$  layer by application of a positive voltage at the gate terminal leads the device to experience a conductance increase whereas the de-intercalation of  $\text{Li}^+$  ions under negative bias leads the device to experience a conductance decrease. The linear conductance change is achievable in ECRAM thanks to the decoupling of read/write paths, which makes this device concept very attractive for synaptic applications, mainly for hardware implementation of synaptic weights in ANNs, where analog and symmetric weight updates play a crucial role. Also, the device investigated in [108] provides fast operation at the nanosecond timescale, thus opening the way toward a significant acceleration of the training process in hardware ANNs.

Figure 12c shows the device structure of the SOT-MRAM, where a heavy metal (HM) line, typically in Pt [121] or Ta [122], is located under an MTJ. This three-terminal device is programmed by the flow of a horizontal current through the HM line, which induces a spin accumulation as a result of the spin Hall or the Rashba effects [123,124], leading to the switching of magnetic polarization in the MTJ FL. Unlike the program operation, the read operation can be performed by measuring the vertical current flowing in MTJ as a result of the TMR effect, which means that the three-terminal structure of SOT-MRAM offers the opportunity to decouple read/write current paths and consequently improve the endurance performance compared with STT-MRAM. Regarding device applications, SOT-MRAM was used to implement neuromorphic computing in ANNs, by exhibiting the synapse function [125], the neuron function [126], and the associative memory operation [127].

## 5. Memristive Neuromorphic Networks

Thanks to their rich physics and nanoscale size, memristive concepts are believed to be promising candidates to achieve the huge density and behavior of real synapses and neurons, thus enabling brain-like cognitive capabilities in hardware neural networks. Based on this appealing approach, many hardware or mixed hardware/simulation implementations of the neural networks currently dominating the neuromorphic computing scenario, namely the DNNs and the SNNs, have been proposed.

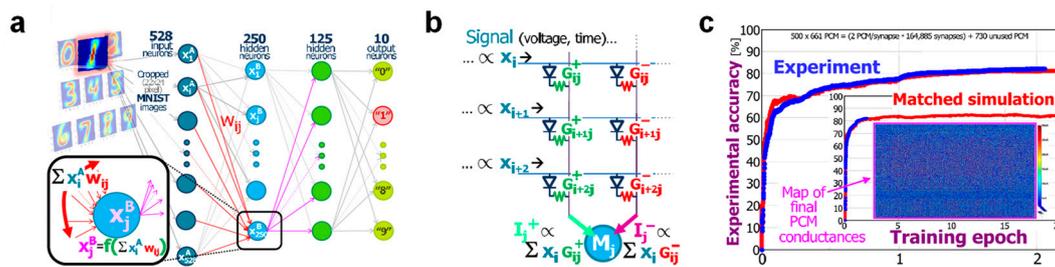
### 5.1. DNNs with Memristive Synapses

DNNs encompass various ANN architectures, such as feedforward MLP and convolutional neural network (CNN) [36], that have attracted wide interest in the neuromorphic computing scenario thanks to the excellent performance achieved in machine learning tasks, such as image classification [128], face verification [129], and speech recognition [130]. Because of the very high complexity of the CNN architecture, which consists of a deep hierarchy of convolutional layers followed by some fully connected layers, and processing strategy, which is based on the extraction of the most significant features of submitted images via the application of large sets of filters, hardware implementation of DNN tasks with memory devices has mostly been focused on feedforward MLP networks. In this type of ANN, the training phase is based on a supervised learning algorithm called backpropagation [21–23] and consists of three sub-procedures called forward propagation, backward propagation, and weight update [36]. Note that although the backpropagation algorithm is chiefly considered lacking in biological plausibility [131], recent works have questioned this aspect [132]. During training, upon any input presentation from a training database containing images of objects, digits, or faces, the input signal propagates in the forward direction from the input to output layer, passing through the multiplication by synaptic weights of each layer and the summation at the input of each hidden/output neuron. Forward propagation yields an output signal, which is compared with the target response of the network, namely the label of the submitted image, thus leading to the calculation of the corresponding error signal. At this point, the calculated error signal is propagated in the backward direction from the output

to the input layer and is used to update all the synaptic weights, hence the name backpropagation. Repeating this scheme for every image of the training database for a certain number of presentation cycles or epochs, the optimization of synaptic weights is achieved, leading the network to specialize on the training database. After, the training phase is followed by the test phase, namely the phase where the classification ability of DNN is evaluated by submitting another database, called the test dataset, only once, via forward propagation of the signal encoded in all the test examples [36].

The downside of the outstanding results achieved running DNNs in software on high-performance digital computers, such as GPU and TPU, or very large servers is given by the excessive power consumption and latency due to the von Neumann architecture. To overcome this issue, memristive devices, in particular RRAM and PCM, have been intensively investigated to accelerate artificial intelligence (AI) applications in hardware thanks to their ability to execute in-memory computing with extremely high energy efficiency and speed by exploiting basic physical laws, such as the Ohm's law and Kirchhoff's law [11]. However, hardware implementation of a real in-situ weight update for DNN training has been challenged by critical non-idealities affecting the conductance response of the majority of memristive devices, mainly RRAM and PCM, during set (potentiation) and reset (depression) processes, such as the non-linearity, the asymmetry, and the stochasticity [34,133,134]. Motivated by these significant limitations, a wide range of alternative materials and technologies have been intensively investigated, leading to the recent emergence of novel concepts, such as ECRAM [108] and the ionic floating gate [135], thanks to their highly linear, symmetric, and analog conductance behavior.

In the last 10 years, great advances in crossbar-based demonstrations of DNNs for pattern classification have been achieved using RRAM and PCM devices [12,13,136–138]. In ref. [12], a medium-scale crossbar array containing 165,000 PCM devices with a one-transistor-one-resistor (1T1R) structure was used to demonstrate an image classification task by hardware implementation of the three-layer DNN schematically shown in Figure 13a. This network is based on an input layer with 528 input neurons, a first hidden layer with 250 neurons, a second hidden layer with 125 neurons, and an output layer with 10 neurons, and was operated on a cropped version ( $22 \times 24$  pixels) of handwritten digit images from the MNIST database for training and test operations. To implement positive and negative synaptic weights of the network, Burr et al. proposed a differential configuration based on pairs of 1T1R PCM cells with conductance,  $G^+$  and  $G^-$ , respectively, as shown in Figure 13b. According to this structure, each weight can be potentiated or depressed by increasing  $G^+$  with fixed  $G^-$  or increasing  $G^-$  with fixed  $G^+$ , respectively. Also, the network was implemented with software neurons, providing the conversion of the sum of input currents into an output voltage by application of the tanh non-linear function. After the training process, which was carried out on 5000 MNIST images by using a complex pulse overlap scheme, the network's classification ability was evaluated, leading to a best performance of only 83% due to the asymmetry and non-linearity of the PCM  $G^-$  response (Figure 13c). To tackle this limitation, a novel artificial synapse combining the 1T1R differential pair with a three-transistor/one-capacitor (3T1C) analog device was presented in ref. [138]. This led the PCM-based DNNs with improved hardware synapses to match the software performance on both the MNIST and CIFAR databases [139]. Later, other DNN implementations in small-scale 1T1R RRAM crossbar arrays were demonstrated, enabling MNIST classification with 92% test performance [137] and gray-scale face classification on the Yale face database with 91.5% performance [136], thanks to the RRAM conductance responses displaying high linearity and symmetry in both update directions. Moreover, an alternative approach aiming at combining high performance with high energy efficiency was proposed in ref. [140]. Here, after an off-line training resulting in the optimization of synaptic weights in the software, the floating-point accuracy of synaptic weights was reduced only to five levels, which were stored in a hardware 4 kbit HfO<sub>2</sub> RRAM array using a novel multilevel programming scheme. The following execution of the inference phase with the experimental conductances stored into the 4 kbit RRAM array led to a maximum classification accuracy of 83%. A simulation-based study showed that the implementation of synaptic weights using more conductance levels can move performance beyond 90% with larger arrays.

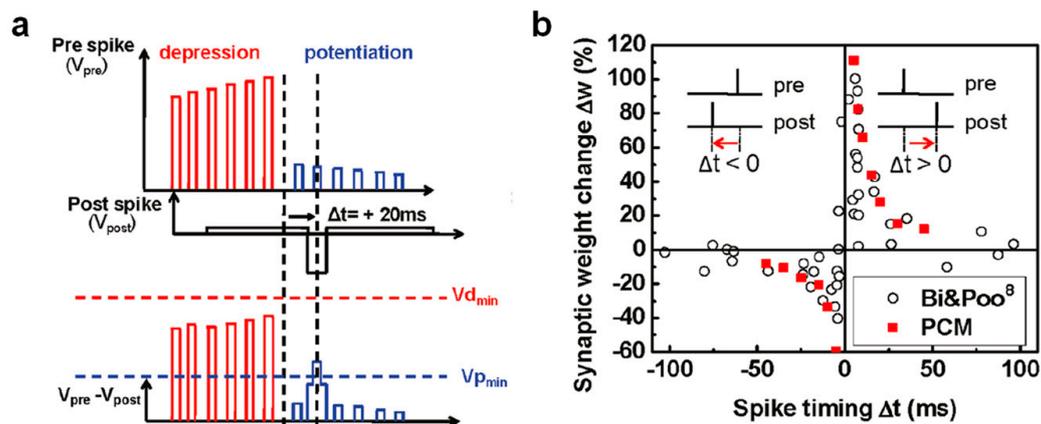


**Figure 13.** (a) Schematic representation of a three-layer DNN operated on the MNIST database for an image classification task. (b) Weight implementation in DNN by differential pairs of 1T1R PCM cells with conductances  $G_{ij}^+$  and  $G_{ij}^-$ , which provide a positive current and a negative current, respectively. (c) Experimental classification accuracy achieved by three-layer DNN during the inference phase. Reprinted with permission from [12]. Copyright 2014, IEEE. Deep neural networks, DNNs; Modified National Institute of Standards and Technology, MNIST; one-transistor-one-resistor, 1T1R.

## 5.2. SNNs with Memristive Synapses

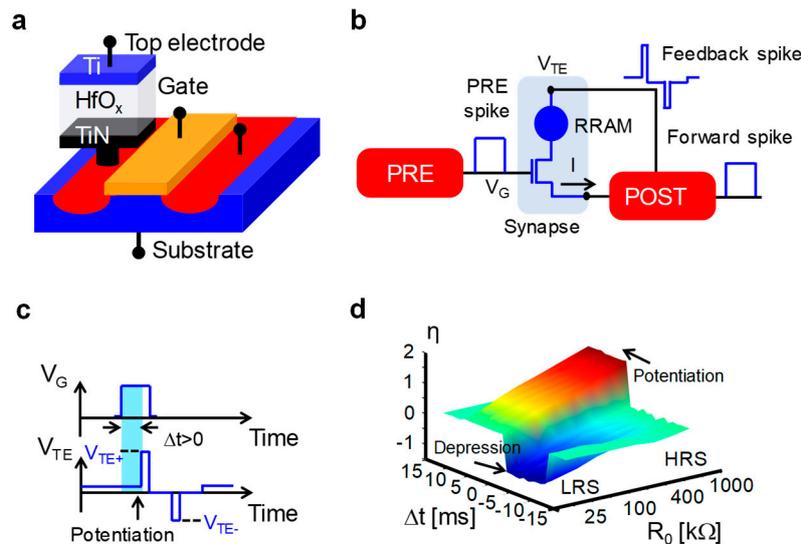
Although DNNs have shown to be capable of excellent performance in fundamental cognitive functions, exceeding the human ability in some cases [128,141], the interest in SNNs is rapidly increasing thanks to their attempt to replicate structure and operation principles of the most efficient computing machine found in nature, which is the biological brain. The brain can efficiently learn, recognize, and infer in an unsupervised way thanks to the plasticity of biological synapses controlled by local rules, such as STDP, which has recently inspired many hardware implementations of synaptic plasticity at the device and network level exploiting the attractive physical properties of memristive devices.

One of the earliest STDP demonstrations at the memristive device level was performed by Jo and coauthors in ref. [142] by using an Ag/Si-based CBRAM device as the synapse and a time-division multiplexing approach based on synchronous time frames which was designed to achieve STDP characteristics thanks to the conversion of the time delay into the amplitude of the pulse to be applied across the synaptic device. After this precursor implementation, another scheme based on voltage overlap at the terminals of memristive synapses was experimentally demonstrated in both RRAM [143] and PCM [144]. Both works demonstrate potentiation and depression characteristics very close to biological STDP, exploiting the analog modulation of device conductance achieved via the superposition of voltage spikes with suitably tailored waveforms. Specifically, Kuzum et al. proposed the voltage waveforms shown in Figure 14a as PRE and POST spikes for achieving potentiation in PCM devices [144]. As the relative delay is positive, in this case  $\Delta t = 20$  ms, the overlap of the PRE spike, which consists of a sequence of high positive pulses with increasing amplitudes followed by another sequence of small positive pulse with decreasing amplitudes, with the POST spike, which consists of a single 8 ms long negative pulse, leads the total voltage across the PCM cell,  $V_{pre} - V_{post}$ , to only cross the minimum threshold for potentiation,  $v_p$ , thus leading the synapse to undergo potentiation via a set process within PCM. Changing the sign of  $\Delta t$ , depression was also demonstrated, thus allowing the STDP characteristics shown in Figure 14b to be achieved, which exhibit a very nice agreement with the Bi and Poo measurements [38]. Moreover, note that this scheme offers the opportunity to finely tune the shape of STDP characteristics, by suitably designing the PRE spike waveform [144]. Taking inspiration from this approach based on overlapping spikes across the memristive device, more recently, other significant STDP demonstrations were achieved in individual two-terminal memristive devices, thus enabling unsupervised learning in small-scale memristive SNNs [145–149]. However, the synapse implementation using individual two-terminal memristive devices might suffer from serious issues, such as (i) the requirement to control the current during set transition in RRAM devices to avoid an uncontrollable CF growth [64], which would reduce the synapse reliability during potentiation; (ii) the sneak paths challenging the operation of crossbar arrays; and (iii) the high energy consumption.



**Figure 14.** (a) PRE and POST spike waveforms applied at terminals of a PCM-based synaptic device to change its weight via an overlap-based STDP scheme. The application of a positive time delay of 20 ms leads to a conductance increase (potentiation) in the PCM synapse since the spike overlap leads the effective voltage across the PCM to cross the potentiation threshold whereas the higher depression threshold is not hit. (b) Measured weight change as a function of the spike timing achieved using a PCM synapse against experimental data collected by Bi and Poo in biological synapses. Reprinted with permission from [144]. Copyright 2012, American Chemical Society.

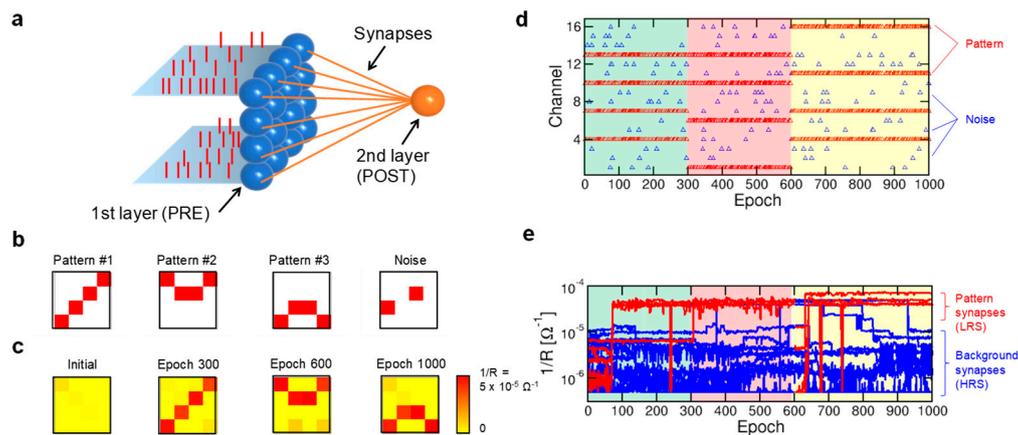
To overcome these drawbacks, a novel hybrid CMOS/memristive STDP synapse using the 1T1R structure was proposed in refs. [150,151]. Figure 15a shows the schematic structure of the 1T1R device presented in ref. [151], where a Ti/HfO<sub>x</sub>/TiN RRAM is serially connected to the drain of an MOS transistor acting as selector and current limiter. As schematically shown in Figure 15b, the ability of the 1T1R cell to operate as a synapse capable of STDP was validated in the hardware [152]. The 1T1R synapse operation can be explained as follows. The application of a pulse designed as a PRE spike at the gate terminal of the transistor combined with the low voltage bias applied at the TE of the RRAM device activates a current flowing toward the BE. At this point, the current enters in an integrate-and-fire circuit implementing POST where it is integrated, causing an increase of the POST internal potential,  $V_{int}$ . As a sequence of PRE spikes leads the POST to cross its internal threshold, the POST emits both a forward spike toward the next neuron layer and a suitably designed spike, including a positive pulse followed by a negative pulse, being delivered at TE, thus creating the conditions for synaptic weight update according to STDP [151]. As shown in Figure 15c, if the PRE spike anticipates the POST spikes ( $\Delta t > 0$ ), only the positive pulse of the POST spike with amplitude  $V_{TE+}$  ( $V_{TE+} > V_{set}$ ) overlaps with the PRE spike, thus inducing a set transition within the RRAM device, leading RRAM to LRS, and, therefore, the synapse to be potentiated. Otherwise, if the PRE spike follows the POST spike ( $\Delta t < 0$ ), only the negative pulse with amplitude  $V_{TE-}$  ( $|V_{TE-}| > |V_{reset}|$ ) overlaps with the PRE spike, thus inducing a reset transition within the RRAM device, leading RRAM to HRS, and, therefore, the synapse to be depressed (not shown). Thanks to this operation principle, the 1T1R synapse was shown to capture STDP functionality implementing the 3D characteristics shown in Figure 15d, where the relative change in conductance,  $\eta = \log_{10}(R_0/R)$ , is plotted as a function of the initial resistance state,  $R_0$ , and relative delay,  $\Delta t$ . They support potentiation/depression at positive/negative  $\Delta t$ , evidencing that maximum potentiation is obtained for  $R_0 = \text{HRS}$ , whereas maximum depression is obtained for  $R_0 = \text{LRS}$ . If the 1T1R synapse is initially in LRS/HRS, no potentiation/depression occurs because it cannot overcome the boundary conductance values set by LRS and HRS [151–153]. Importantly, note that the weight change in the 1T1R synapse can be induced only via spike overlap, hence only for delays in the range  $-10 \text{ ms} < \Delta t < 10 \text{ ms}$  in this experiment [152].



**Figure 15.** (a) Schematic structure of the 1T1R RRAM structure. (b) Schematic representation of the 1T1R structure as a synapse to achieve STDP in hardware via overlapping PRE and POST voltage spikes applied at the gate terminal and RRAM top electrode, respectively. (c) Schematic sketch of PRE and POST overlapping spikes leading to synapse potentiation via the activation of a set process in the RRAM cell. (d) STDP characteristics experimentally demonstrated in the 1T1R RRAM synapse. Adapted with permission from [151,152]. Copyright 2016, IEEE.

Although the STDP characteristics achieved in the 1T1R RRAM synapse [151,152] display a squared shape due to binary operation of the RRAM cell instead of the exponentially decaying behavior observed in biological experiments, the plasticity of the 1T1R synapse was exploited in many SNN implementations enabling neuromorphic tasks, such as unsupervised learning of space/spatiotemporal patterns [151,152,154,155], the extraction of auditory/visual patterns [156,157], pattern classification [158–160], and associative memory [161–163], in both simulation and hardware.

Figure 16a shows the schematic representation of the RRAM-based SNN used in ref. [152] to demonstrate unsupervised learning of visual patterns in hardware. This perceptron SNN consists of 16 PREs connected to a single POST via individual synapses with the 1T1R RRAM structure of Figure 15a. Pattern learning experiment is based on three sequential phases where only one  $4 \times 4$  visual pattern among Pattern #1, Pattern #2, and Pattern #3 shown in Figure 16b is submitted to the input layer, and was conducted using a stochastic approach according to which the probability to submit the pattern image or a random noise image similar to the last  $4 \times 4$  pattern in Figure 16b at every epoch is 50%. Using this training approach, Figure 16c shows that the submission of three patterns alternated with noise resulted in the on-line adaptation of SNN synapses to the presented pattern in all three phases, evidencing a selective potentiation of synapses within the submitted pattern due to the correlated spiking activity of corresponding PREs and the depression of synapses outside the pattern, typically called background synapses, due to the uncorrelated nature of noise inducing POST spike-PRE spike depression sequences for the background with a high probability [151,152]. Note that the frequency and amount of submitted noise has to be carefully designed to prevent learning dynamics from becoming unstable [164]. To further support the unsupervised pattern learning ability of SNN with 1T1R RRAM synapses, Figure 16d shows the raster plot of spikes generated by PREs during the whole experiment, leading to the time evolution of synaptic conductance evidenced in Figure 16e, where the pattern/background synaptic conductance converges to LRS/HRS at the end of each training phase. Note that the stochastic approach used in this experiment also allowed for the implementation of multiple pattern learning by a winner-take-all scheme [165] based on the use of software inhibitory synapses between 2 POSTs, and unsupervised learning of gray-scale images [152].

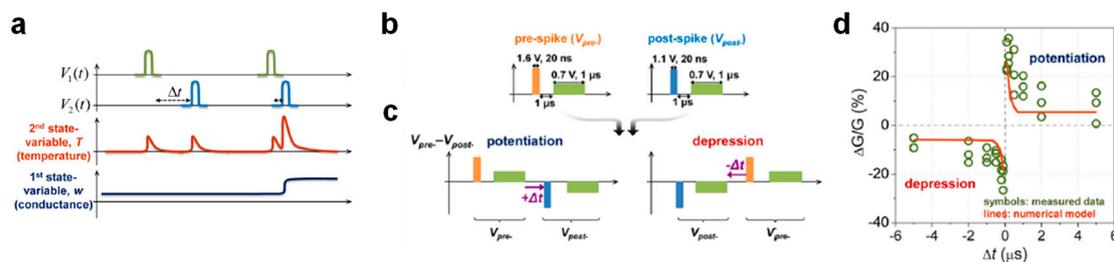


**Figure 16.** (a) Schematic sketch of a single-layer perceptron network where a  $4 \times 4$  input layer is fully connected to a single POST. (b) Sequence of three visual patterns (Pattern #1, Pattern #2, and Pattern #3) submitted to the neural network during training process and an example of a random noise image, which is alternatively applied to patterns according to a stochastic approach. (c) Conductance/weight color plots measured at epochs 0, 300, 600, and 1000 evidencing the ability of the synaptic weights to adapt to submitted patterns thanks to selective potentiation of pattern synapses and noise-induced depression of background synapses. (d) Raster plot of PRE spikes applied to pattern and background input channels during the learning experiment. (e) Time evolution of the measured synaptic conductance during three phases of the unsupervised learning experiment showing convergence of pattern/background synapses to LRS/HRS. Reprinted from [152].

The main drawbacks generally limiting the implementation of synaptic plasticity in overlap-based synaptic concepts, such as the 1T1R synapse, are the pulse duration and energy efficiency. Overlap-based implementations first require a pulse width of the order of time delays to allow for conductance change within the device, which results in pulses with a long duration causing a high power consumption. In addition to this, the need for long pulses to program overlap-based memristive devices also causes too slow signal processing in large neuromorphic networks, which leads to low throughput performance [166].

An alternative approach to achieve synaptic plasticity overcoming the limitations affecting overlap-based memristive devices consists of the adoption of non-overlap memristive devices, such as the second-order memristor [167,168]. Unlike first-order memristors, such as RRAM and PCM, where device conductance can change only if overlapping voltage pulses are applied at device terminals, resistive switching in second-order memristors can take place by sequential application of two spikes with a certain  $\Delta t$  at device terminals as a result of short-term memory effects encoded in the time evolution of second-order variables, e.g., the internal temperature. As shown in Figure 17a, if  $\Delta t$  is long, two sequential spikes applied at terminals of a second-order memristor induce small independent changes in temperature, which results in no conductance change. On the contrary, if  $\Delta t$  is short, the superposition of the effects of applied spikes results in a large change in temperature thanks to a limited thermal constant of about 500 ns, thus leading to a long-term conductance variation in the device as a result of short-term memory effects. Importantly, short memory effects observed in second-order memristors have recently attracted great interest because they can allow for the emulation in hardware of a fundamental biological process playing a key role in the real synapse response as the  $\text{Ca}^{2+}$  ion dynamics [169,170] and to finely replicate biological STDP and SRDP [168,171]. An interesting STDP demonstration by a second-order memristor is reported in ref. [168]. Here, a Pt/Ta<sub>2</sub>O<sub>5-x</sub>/TaO<sub>y</sub>/Pd RRAM device was operated as a non-overlap synapse to achieve STDP via sequential application of PRE and POST voltages. As shown in Figure 17b, the PRE spike consists of a positive pulse with amplitude of 1.6 V and duration of 20 ns followed after 1  $\mu\text{s}$  by a longer positive pulse with amplitude of 0.7 V and duration of 1  $\mu\text{s}$  whereas the POST spike includes a positive pulse with amplitude of 1.1 V and duration of 20 ns followed after 1  $\mu\text{s}$  by a longer positive pulse with amplitude of 0.7 V and 1  $\mu\text{s}$

width. Note that both the first pulse, called the programming element, and the second pulse, called the heating element, within PRE and POST spikes cannot cause independently a conductance change in the RRAM device. The application of the PRE/POST spike at TE/BE of the RRAM device results in an effective voltage drop across the device, evidencing a PRE–POST spike sequence for positive  $\Delta t$  and POST–PRE spike sequence for negative  $\Delta t$ , as shown in Figure 17c. In the case of the PRE–POST spike sequence ( $\Delta t > 0$ ), the heating effect of the PRE spike affects the POST spike, making the positive change in conductance due to the negative programming pulse in the POST higher than the negative change in conductance due to the positive programming pulse in the PRE, hence causing the non-overlap RRAM synapse to undergo potentiation. On the other hand, in the case of the POST–PRE sequence ( $\Delta t < 0$ ), the opposite occurrence order of spikes results in an effective negative conductance change in the Pt/Ta<sub>2</sub>O<sub>5-x</sub>/TaO<sub>y</sub>/Pd RRAM device, resulting in the depression of the non-overlap synapse. Figure 17d shows the STDP characteristics experimentally measured in the Pt/Ta<sub>2</sub>O<sub>5-x</sub>/TaO<sub>y</sub>/Pd RRAM device for variable  $\Delta t$  in the range  $-6 \mu\text{s} - 6 \mu\text{s}$ , which exhibit strong similarity with biological data and a good agreement with simulation results achieved by a numerical model of the second-order memristor.



**Figure 17.** (a) Schematic representation of a non-overlap scheme enabling STDP in second-order memristors. Short-term memory effects observed in second-order physical variables, e.g., internal temperature, allow for the implementation of potentiation/depression for short/long delays. (b) PRE and POST spike waveforms applied at top electrode (TE) and bottom electrode (BE) to implement non-overlap STDP. (c) Effective voltage across a second-order memristor to induce potentiation (left) and depression (right). (d) STDP characteristics measured in a second-order memristor against calculated curves achieved by numerical modeling. Reprinted with permission from [168]. Copyright 2015, American Chemical Society.

Similar to the second-order memristor device, other memristive concepts also allowed bio-realistic synaptic plasticity to be demonstrated using non-overlap schemes. In ref. [172], an atomic switch RRAM, whose stack includes a silver BE, an Ag<sub>2</sub>S-based solid electrolyte, and a metal TE separated from the Ag<sub>2</sub>S layer by a nanogap, was proposed as an artificial synapse thanks to the short-term memory effects controlling its physical processes. In fact, the application of voltage pulses at TE induces the gradual creation of an Ag atomic bridge within the nanogap leading to a short-term potentiation process after a few pulses, resulting in an incomplete atomic bridge, which is followed by a long-term potentiation process achieved after many pulses resulting in the formation of a complete atomic bridge. In addition to short-term plasticity due to the spontaneous relaxation process of the atomic bridge, this non-overlap device also offers the opportunity to capture SRDP potentiation and depression depending on whether the frequency of the applied pulses is high or low. Thanks to this functionality, the sequential learning of visual patterns was demonstrated in a  $7 \times 7$  array of Ag<sub>2</sub>S inorganic synaptic devices.

Another memristive concept to implement non-overlap synapses in hardware was recently presented in ref. [171]. Here, a hybrid device based on the serial configuration of a volatile RRAM with a SiO<sub>x</sub>N<sub>y</sub>:Ag stack serving as the select device and a non-volatile RRAM serving as the resistive device, also known as a one-selector-one-resistor (1S1R) structure, was designed to demonstrate non-overlap synaptic plasticity for neuromorphic computing. Exploiting spontaneous relaxation of CF similar to the one taking place in atomic switches, the introduction of a volatile RRAM or diffusive memristor

in series to a non-volatile RRAM, where conductance change can only be induced by the electric field, enabled 1S1R synapses capable of both SRDP and STDP depending on the rate or occurrence timing of PRE and POST spikes applied in sequence at TE. Note that the strong potential of 1S1R synapses for neuromorphic computing applications was also investigated in simulation in [173,174]. Moreover, diffusive memristors developed in ref. [171] were used as neurons to build in hardware a fully memristive neural network, which was shown to achieve outstanding performance in a pattern classification task by the implementation of unsupervised learning [175].

## 6. Discussion

While neuromorphic networks have recently demonstrated an excellent ability in fundamental cognitive computing applications, such as image classification and speech recognition, their large-scale hardware implementation is still a major challenge. Achieving such a goal primarily requires nanoscale, energy-efficient, and fast devices capable of emulating faithfully high-density, ultra-low power operation and low latency of biological synapses and neurons. Moreover, depending on the architecture (DNN or SNN) and the application of neuromorphic networks, such devices should also fulfill other significant requirements, such as high retention, high linearity in conductance response, and long endurance [35]. In Table 1, the CMOS-based and memristive emerging memory devices investigated for neuromorphic computing we discussed in Sections 3 and 4 are compared in terms of performance, reliability, and suitability for DNN, with the distinction between training and inference phases, and SNN applications; however, it is evidenced that no emerging memory device can currently optimize all the metrics for any network architecture and application.

**Table 1.** Comparison of key features exhibited by CMOS mainstream memory devices and memristive emerging memory devices under investigation to implement neuromorphic computing in hardware. Adapted from [35].

Technology	CMOS Mainstream Memories		Memristive Emerging Memories						
	NOR Flash	NAND Flash	RRAM	PCM	STT-MRAM	FeRAM	FeFET	SOT-MRAM	Li-ion
ON/OFF Ratio	10 <sup>4</sup>	10 <sup>4</sup>	10–10 <sup>2</sup>	10 <sup>2</sup> –10 <sup>4</sup>	1.5–2	10 <sup>2</sup> –10 <sup>3</sup>	5–50	1.5–2	40–10 <sup>3</sup>
Multilevel operation	2 bit	4 bit	2 bit	2 bit	1 bit	1 bit	5 bit	1 bit	10 bit
Write voltage	<10 V	>10 V	<3V	<3V	<1.5 V	<3 V	<5 V	<1.5 V	<1 V
Write time	1–10 $\mu$ s	0.1–1 ms	<10 ns	~50 ns	<10 ns	~30 ns	~10 ns	<10 ns	<10 ns
Read time	~50 ns	~10 $\mu$ s	<10 ns	<10 ns	<10 ns	<10 ns	~10 ns	<10 ns	<10 ns
Stand-by power	Low	Low	Low	Low	Low	Low	Low	Low	Low
Write energy (J/bit)	~100 pJ	~10 fJ	0.1–1 pJ	10 pJ	~100 fJ	~100 fJ	<1 fJ	<100 fJ	~100 fJ
Linearity	Low	Low	Low	Low	None	None	Low	None	High
Drift	No	No	Weak	Yes	No	No	No	No	No
Integration density	High	Very High	High	High	High	Low	High	High	Low
Retention	Long	Long	Medium	Long	Medium	Long	Long	Medium	-
Endurance	10 <sup>5</sup>	10 <sup>4</sup>	10 <sup>5</sup> –10 <sup>8</sup>	10 <sup>6</sup> –10 <sup>9</sup>	10 <sup>15</sup>	10 <sup>10</sup>	>10 <sup>5</sup>	>10 <sup>15</sup>	>10 <sup>5</sup>
Suitability for DNN training	No	No	No	No	No	No	Moderate	No	Yes
Suitability for DNN inference	Yes	Yes	Moderate	Yes	No	No	Yes	No	Yes
Suitability for SNN applications	Yes	No	Yes	Yes	Moderate	Yes	Yes	Moderate	Moderate

To efficiently execute DNN online training in hardware, high speed and low energy consumption are two essential features of synaptic devices to maximize the network throughput, namely the rate of trained patterns, and enable DNNs in embedded systems, respectively. In addition to these features,

high accuracy in weight update operation imposes the use of devices exhibiting a conductance response with a high degree of linearity. This functionality makes almost all the emerging devices unsuitable as synaptic devices for online training. The only exception is represented by novel Li-ion devices, which appear to be very promising, with a simulated performance of around 98% [119], even though the necessary technology maturity and high-density integration have not been reached yet. Alternatively, more complex structures, including multiple pair of memristive devices, such as PCM and RRAM, could mitigate the need for high linearity, but at the expense of a lower integration density [176].

Differently from DNN online training consisting of forward propagation, backpropagation, and weight update operations, DNN inference only relies on forward propagation, which means that the high linearity needed to accurately update the weights is not an essential feature of synaptic devices for this task. Specifically, hardware suitable for optimizing the inference process should primarily exhibit low latency to accelerate the classification of each test pattern and low-power consumption to enable DNN inference at the edge. In addition to these features, high retention of analogue states is also essential to prevent charge fluctuations in CMOS devices [177], stochastic noise in RRAM [178], and resistance drift in PCM [179] from degrading the weights programmed in one shot after the off-line training procedure. These requirements can be fulfilled not only by Li-ion devices, as in the case of DNN training, but also by CMOS floating gate memory [55], RRAM [137], and PCM [148] devices thanks to their ability to finely tune the conductance with analog precision to encode the stored weights.

On the other hand, hardware implementation of brain-inspired SNNs for sensors or embedded systems primarily requires high energy efficiency to enable sensory information processing for long times even in limited-energy environments. The high endurance of synaptic and neuron devices is also strongly required in that SNN operation relies on a learning approach based on continuous synaptic updates and continuous reset operations of integrate-and-fire neurons upon fire events. In addition to these features, a high resistance window could be useful for accurate continual learning although multilevel weight storage could be not strictly needed, as shown by significant applications using binary stochastic memory devices, such as STT-MRAM. Therefore, both NOR Flash memory [57], despite higher operating voltages, and all the memristive emerging devices show a strong potential for hardware implementation of SNNs emulating the efficiency and 3D architecture of the biological brain.

Although some limitations currently hinder the large-scale industrialization of memory-centric neuromorphic technology, the rich physics of memory devices can also offer additional biologically inspired functionalities and more. For instance, besides synaptic implementation, integrate-and-fire neuron functionality has been recently demonstrated in various types of memristive devices, including RRAM [180], volatile RRAM [175], Mott memristor [181], PCM [182], STT-MRAM [183,184], SOT-MRAM [126], and paramagnetic MTJs [185], thus opening the way for hardware implementation of high-density fully memristive neural networks with a high area and energy efficiency. Also, thanks to the short-term memory effects observed in some materials, a more realistic implementation of biological synaptic behavior taking into account the impact of spatiotemporal patterns has been achieved [171–173]. Moving from the standpoint of the device to that of the system, in-memory computing with memristive devices is opening the way to the exploration of new learning algorithms exhibiting strong similarity with human experience, such as reinforcement learning [186], which has already been shown to enable complex tasks [187].

Finally, memristive devices are receiving increasing interest for the development of other computing concepts by neuromorphic networks with high computational power, such as the Hopfield recurrent neural network [188]. Although high acceleration performance has been achieved for the solution of hard constraint-satisfaction problems (CSPs), such as the Sudoku puzzle, via CMOS-based circuits [189], FPGA [190], and quantum computing circuits [191], the use of memristive devices in crossbar-based neural networks can further speed up computation by the introduction of a key resource as the noise [192] without the requirement of additional sources [193]. Moreover, very recent studies have also evidenced the strong potential of memristive devices for the execution of complex algebraic tasks, including the solution of linear systems and differential equations, such as the Schrödinger

and Fourier equations, in crossbar arrays in only one computational step [16], thus overcoming the latency of iterative approaches [15]. Therefore, these achievements suggest CMOS/memristive devices as enablers of novel high-efficiency computing paradigms capable of revolutionizing many fields of our society.

## 7. Conclusions

This work provides an overview of the most promising devices for neuromorphic computing covering both CMOS and memristive device concepts. Physical MVM in memristive/CMOS crossbar arrays implementing DNNs and SNNs has enabled both fundamental cognitive applications, such as image and speech recognition, and the solution of algebraic and constraint-satisfaction problems in hardware. These milestones can thus pave the way to highly powerful and energy-efficient neuromorphic hardware based on CMOS/memristive technologies, making AI increasingly pervasive in future society.

**Author Contributions:** Writing—original draft preparation, V.M. and G.M.; writing—review and editing, V.M., G.M., D.I. and C.M.C.; visualization, V.M. and G.M.; supervision, D.I. and C.M.C.; funding acquisition, D.I. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 648635).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Moore, G.E. Cramming more components onto integrated circuits. *Electronics* **1965**, *38*, 114–117. [[CrossRef](#)]
2. Dennard, R.H.; Gaensslen, F.H.; Yu, H.-N.; Rideout, V.L.; Bassous, E.; LeBlanc, A.R. Design of ion-implanted MOSFET’s with very small physical dimensions. *IEEE J. Solid State Circuits* **1974**, *9*, 256–268. [[CrossRef](#)]
3. Horowitz, M. Computing’s energy problem (and what we can do about it). In Proceedings of the 2014 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, USA, 9–13 February 2014; pp. 10–14. [[CrossRef](#)]
4. Waldrop, M.M. The chips are down for Moore’s law. *Nature* **2016**, *530*, 144–147. [[CrossRef](#)] [[PubMed](#)]
5. Robertson, J. High dielectric constant oxides. *Eur. Phys. J. Appl. Phys.* **2004**, *28*, 265–291. [[CrossRef](#)]
6. Ferain, I.; Colinge, C.A.; Colinge, J.-P. Multigate transistor as the future of classical metal-oxide-semiconductor field-effect transistors. *Nature* **2011**, *479*, 310–316. [[CrossRef](#)]
7. Kuhn, K. Considerations for ultimate CMOS scaling. *IEEE Trans. Electron Devices* **2012**, *59*, 1813–1828. [[CrossRef](#)]
8. Shulaker, M.M.; Hills, G.; Park, R.S.; Howe, R.T.; Saraswat, K.; Wong, H.-S.P.; Mitra, S. Three-dimensional integration of nanotechnologies for computing and data storage on a single chip. *Nature* **2017**, *547*, 74–78. [[CrossRef](#)]
9. Wong, H.-S.P.; Salahuddin, S. Memory leads the way to better computing. *Nat. Nanotechnol.* **2015**, *10*, 191–194. [[CrossRef](#)]
10. Truong, S.N.; Min, K.-S. New memristor-based crossbar array architecture with 50-% area reduction and 48-% power saving for matrix-vector multiplication of analog neuromorphic computing. *J. Semicond. Technol. Sci.* **2014**, *14*, 356–363. [[CrossRef](#)]
11. Ielmini, D.; Wong, H.-S.P. In-memory computing with resistive switching devices. *Nat. Electron.* **2018**, *1*, 333–343. [[CrossRef](#)]
12. Burr, G.W.; Shelby, R.M.; di Nolfo, C.; Jang, J.W.; Shenoy, R.S.; Narayanan, P.; Virwani, K.; Giacometti, E.U.; Kurdi, B.; Hwang, H. Experimental demonstration and tolerancing of a large-scale neural network (165,000 synapses) using phase-change memory as the synaptic weight element. In Proceedings of the 2014 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 15–17 December 2014; pp. 697–700. [[CrossRef](#)]
13. Prezioso, M.; Merrih-Bayat, F.; Hoskins, B.D.; Adam, G.C.; Likharev, K.K.; Strukov, D.B. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **2015**, *521*, 61–64. [[CrossRef](#)] [[PubMed](#)]

14. Li, C.; Hu, M.; Li, Y.; Jiang, H.; Ge, N.; Montgomery, E.; Zhang, J.; Song, W.; Dávila, N.; Graves, C.E.; et al. Analogue signal and image processing with large memristor crossbars. *Nat. Electron.* **2018**, *1*, 52–59. [[CrossRef](#)]
15. Le Gallo, M.; Sebastian, A.; Mathis, R.; Manica, M.; Giefers, H.; Tuma, T.; Bekas, C.; Curioni, A.; Eleftheriou, E. Mixed-precision in-memory computing. *Nat. Electron.* **2018**, *1*, 246–253. [[CrossRef](#)]
16. Sun, Z.; Pedretti, G.; Ambrosi, E.; Bricalli, A.; Wang, W.; Ielmini, D. Solving matrix equations in one step with crosspoint resistive arrays. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4123–4128. [[CrossRef](#)]
17. Sun, Z.; Pedretti, G.; Bricalli, A.; Ielmini, D. One-step regression and classification with crosspoint resistive memory arrays. *Sci. Adv.* **2019**, in press.
18. Indiveri, G.; Liu, S.-C. Memory and information processing in neuromorphic systems. *Proc. IEEE* **2015**, *103*, 1379–1397. [[CrossRef](#)]
19. McCulloch, W.S.; Pitts, W.A. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [[CrossRef](#)]
20. Rosenblatt, F. *The Perceptron: A Perceiving and Recognizing Automaton Project Para*; Report 85-460-1; Cornell Aeronautical Laboratory: Buffalo, NY, USA, 1957.
21. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representation by backpropagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
22. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
23. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
24. Coates, A.; Huval, B.; Wang, T.; Wu, D.; Ng, A.Y.; Catanzaro, B.C. Deep learning with COTS HPC systems. In Proceedings of the 30th International Conference on Machine Learning (ICML), Atlanta, GA, USA, 16–21 June 2013; Volume 28, pp. 1337–1345.
25. Jouppi, N.P.; Young, C.; Patil, N.; Patterson, D.; Agrawal, G.; Bajwa, R.; Bates, S.; Bathia, S.; Boden, N.; Borchers, A.; et al. In-Datcenter performance analysis of a Tensor Processing Unit™. In Proceedings of the 44th International Symposium on Computer Architecture (ISCA), Toronto, ON, Canada, 24–28 June 2017; pp. 1–12. [[CrossRef](#)]
26. Hu, M.; Graves, C.E.; Li, C.; Li, Y.; Ge, N.; Montgomery, E.; Davila, N.; Jiang, H.; Williams, R.S.; Yang, J.J.; et al. Memristor-based analog computation and neural network classification with a dot product engine. *Adv. Mater.* **2018**, *30*, 1705914. [[CrossRef](#)] [[PubMed](#)]
27. Xia, Q.; Yang, J.J. Memristive crossbar arrays for brain-inspired computing. *Nat. Mater.* **2019**, *18*, 309–323. [[CrossRef](#)] [[PubMed](#)]
28. Merolla, P.A.; Arthur, J.V.; Alvarez-Icaza, R.; Cassidy, A.S.; Sawada, J.; Akopyan, F.; Jackson, B.L.; Imam, N.; Guo, C.; Nakamura, Y.; et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **2014**, *345*, 668–673. [[CrossRef](#)] [[PubMed](#)]
29. Moradi, S.; Qiao, N.; Stefanini, F.; Indiveri, G. A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors (DYNAPs). *IEEE Trans. Biomed. Circuits Syst.* **2017**, *12*, 106–122. [[CrossRef](#)] [[PubMed](#)]
30. Ramakrishnan, S.; Hasler, P.; Gordon, C. Floating gate synapses with spike-time-dependent plasticity. *IEEE Trans. Biomed. Circuits Syst.* **2011**, *5*, 244–252. [[CrossRef](#)] [[PubMed](#)]
31. Strukov, D.B.; Snider, G.S.; Stewart, D.R.; Williams, R.S. The missing memristor found. *Nature* **2008**, *453*, 80–83. [[CrossRef](#)]
32. Kuzum, D.; Yu, S.; Wong, H.-S.P. Synaptic electronics: Materials, devices and applications. *Nanotechnology* **2013**, *24*, 382001. [[CrossRef](#)]
33. Burr, G.W.; Shelby, R.M.; Sebastian, A.; Kim, S.; Kim, S.; Sidler, S.; Virwani, K.; Ishii, M.; Narayanan, P.; Fumarola, A.; et al. Neuromorphic computing using non-volatile memory. *Adv. Phys. X* **2017**, *2*, 89–124. [[CrossRef](#)]
34. Yu, S. Neuro-inspired computing with emerging nonvolatile memory. *Proc. IEEE* **2018**, *106*, 260–285. [[CrossRef](#)]
35. Ielmini, D.; Ambrogio, S. Emerging neuromorphic devices. *Nanotechnology* **2020**, *31*, 092001. [[CrossRef](#)]
36. Sze, V.; Chen, Y.H.; Yang, T.-J.; Emer, J.S. Efficient processing of Deep Neural Networks: A tutorial and survey. *Proc. IEEE* **2017**, *105*, 2295–2329. [[CrossRef](#)]

37. Maass, W. Networks of spiking neurons: The third generation of neural network models. *Neural Netw.* **1997**, *10*, 1659–1671. [[CrossRef](#)]
38. Bi, G.-Q.; Poo, M.-M. Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **1998**, *18*, 10464–10472. [[CrossRef](#)] [[PubMed](#)]
39. Sjostrom, P.J.; Turrigiano, G.G.; Nelson, S.B. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* **2001**, *32*, 1149–1164. [[CrossRef](#)]
40. Gjorgjieva, J.; Clopath, C.; Audet, J.; Pfister, J.P. A triplet spike timing dependent plasticity model generalizes the Bienenstock-Cooper-Munro rule to higher-order spatiotemporal correlations. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 19383–19388. [[CrossRef](#)] [[PubMed](#)]
41. Pfister, J.-P.; Gerstner, W. Triplets of spikes in a model of spike timing-dependent plasticity. *J. Neurosci.* **2006**, *26*, 9673–9682. [[CrossRef](#)]
42. Rachmuth, G.; Shouval, H.-Z.; Bear, M.F.; Poon, C.-S. A biophysically-based neuromorphic model of spike rate- and timing-dependent plasticity. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, E1266–E1274. [[CrossRef](#)]
43. Milo, V.; Pedretti, G.; Carboni, R.; Calderoni, A.; Ramaswamy, N.; Ambrogio, S.; Ielmini, D. A 4-Transistors/1-Resistor hybrid synapse based on resistive switching memory (RRAM) capable of spike-rate-dependent plasticity (SRDP). *IEEE Trans. Very Large Scale Integrat. (VLSI) Syst.* **2018**, *26*, 2806–2815. [[CrossRef](#)]
44. Monzio Compagnoni, C.; Goda, A.; Spinelli, A.S.; Feeley, P.; Lacaita, A.L.; Visconti, A. Reviewing the evolution of the NAND Flash technology. *Proc. IEEE* **2017**, *105*, 1609–1633. [[CrossRef](#)]
45. Bez, R.; Camerlenghi, E.; Modelli, A.; Visconti, A. Introduction to Flash memory. *Proc. IEEE* **2003**, *91*, 489–502. [[CrossRef](#)]
46. Hasler, P.; Diorio, C.; Minch, B.A.; Mead, C. Single transistor learning synapses. In Proceedings of the 7th International Conference on Neural Information Processing Systems (NIPS), Denver, CO, USA, 28 November–1 December 1994; pp. 817–824.
47. Diorio, C.; Hasler, P.; Minch, B.A.; Mead, C.A. A single-transistor silicon synapse. *IEEE Trans. Electron Devices* **1996**, *43*, 1972–1980. [[CrossRef](#)]
48. Diorio, C.; Hasler, P.; Minch, B.A.; Mead, C.A. A floating-gate MOS learning array with locally computed weight updates. *IEEE Trans. Electron Devices* **1997**, *44*, 2281–2289. [[CrossRef](#)]
49. Kim, H.; Park, J.; Kwon, M.-W.; Lee, J.-H.; Park, B.-G. Silicon-based floating-body synaptic transistor with frequency-dependent short- and long-term memories. *IEEE Electron Device Lett.* **2016**, *37*, 249–252. [[CrossRef](#)]
50. Kim, H.; Hwang, S.; Park, J.; Yun, S.; Lee, J.-H.; Park, B.-G. Spiking neural network using synaptic transistors and neuron circuits for pattern recognition with noisy images. *IEEE Electron Device Lett.* **2018**, *39*, 630–633. [[CrossRef](#)]
51. Kim, C.-H.; Lee, S.; Woo, S.Y.; Kang, W.-M.; Lim, S.; Bae, J.-H.; Kim, J.; Lee, J.-H. Demonstration of unsupervised learning with spike-timing-dependent plasticity using a TFT-type NOR Flash memory array. *IEEE Trans. Electron Devices* **2018**, *65*, 1774–1780. [[CrossRef](#)]
52. Technology is driving the latest automotive designs. Available online: <http://www.sst.com> (accessed on 20 December 2019).
53. Merrikh Bayat, F.; Guo, X.; Om’mani, H.A.; Do, N.; Likharev, K.K.; Strukov, D.B. Redesigning commercial floating-gate memory for analog computing applications. In Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS), Lisbon, Portugal, 24–27 May 2015; pp. 1921–1924. [[CrossRef](#)]
54. Guo, X.; Merrikh Bayat, F.; Bavandpour, M.; Klachko, M.; Mahmoodi, M.R.; Prezioso, M.; Likharev, K.K.; Strukov, D.B. Fast, energy-efficient, robust, and reproducible mixed-signal neuromorphic classifier based on embedded NOR flash memory technology. In Proceedings of the 2017 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2–6 December 2017; pp. 151–154. [[CrossRef](#)]
55. Merrikh Bayat, F.; Guo, X.; Klachko, M.; Prezioso, M.; Likharev, K.K.; Strukov, D.B. High-performance mixed-signal neurocomputing with nanoscale floating-gate memory cell arrays. *IEEE Trans. Neural Netw. Learning Syst.* **2018**, *29*, 4782–4790. [[CrossRef](#)] [[PubMed](#)]
56. Guo, X.; Merrikh Bayat, F.; Prezioso, M.; Chen, Y.; Nguyen, B.; Do, N.; Strukov, D.B. Temperature-insensitive analog vector-by-matrix multiplier based on 55 nm NOR flash memory cells. In Proceedings of the IEEE Custom Integrated Circuits Conference (CICC), Austin, TX, USA, 30 April–3 May 2017; pp. 1–4. [[CrossRef](#)]

57. Malavena, G.; Spinelli, A.S.; Monzio Compagnoni, C. Implementing spike-timing-dependent plasticity and unsupervised learning in a mainstream NOR Flash memory array. In Proceedings of the 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 1–5 December 2018; pp. 35–38. [[CrossRef](#)]
58. Malavena, G.; Filippi, M.; Spinelli, A.S.; Monzio Compagnoni, C. Unsupervised learning by spike-timing-dependent plasticity in a mainstream NOR Flash memory array: Part I—Cell operation. *IEEE Trans. Electron Devices* **2019**, *66*, 4727–4732. [[CrossRef](#)]
59. Malavena, G.; Filippi, M.; Spinelli, A.S.; Monzio Compagnoni, C. Unsupervised learning by spike-timing-dependent plasticity in a mainstream NOR Flash memory array: Part II—Array learning. *IEEE Trans. Electron Devices* **2019**, *66*, 4733–4738. [[CrossRef](#)]
60. Malavena, G.; Petrò, S.; Spinelli, A.S.; Monzio Compagnoni, C. Impact of program accuracy and random telegraph noise on the performance of NOR Flash-based neuromorphic classifier. In Proceedings of the 49th European Solid-State Device Research Conference (ESSDERC), Cracow, Poland, 23–26 September 2019; pp. 122–125. [[CrossRef](#)]
61. Waser, R.; Aono, M. Nanoionics-based resistive switching memories. *Nat. Mater.* **2007**, *6*, 833–840. [[CrossRef](#)] [[PubMed](#)]
62. Wong, H.-S.P.; Lee, H.-Y.; Yu, S.; Chen, Y.-S.; Wu, Y.; Chen, P.-S.; Lee, B.; Chen, F.T.; Tsai, M.-J. Metal oxide RRAM. *Proc. IEEE* **2012**, *100*, 1951–1970. [[CrossRef](#)]
63. Ielmini, D. Resistive switching memories based on metal oxides: Mechanisms, reliability and scaling. *Semicond. Sci. Technol.* **2016**, *31*, 063002. [[CrossRef](#)]
64. Ielmini, D. Modeling the universal set/reset characteristics of bipolar RRAM by field- and temperature-driven filament growth. *IEEE Trans. Electron Devices* **2011**, *58*, 4309–4317. [[CrossRef](#)]
65. Larentis, S.; Nardi, F.; Balatti, S.; Gilmer, D.C.; Ielmini, D. Resistive switching by voltage-driven ion migration in bipolar RRAM—Part II: Modeling. *IEEE Trans. Electron Devices* **2012**, *59*, 2468–2475. [[CrossRef](#)]
66. Ambrogio, S.; Balatti, S.; Gilmer, D.C.; Ielmini, D. Analytical modeling of oxide-based bipolar resistive memories and complementary resistive switches. *IEEE Trans. Electron Devices* **2014**, *61*, 2378–2386. [[CrossRef](#)]
67. Lee, H.Y.; Chen, P.S.; Wu, T.Y.; Chen, Y.S.; Wang, C.C.; Tzeng, P.J.; Lin, C.H.; Chen, F.; Lien, C.H.; Tsai, M.-J. Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO<sub>2</sub> based RRAM. In Proceedings of the 2008 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 15–17 December 2008; pp. 1–4. [[CrossRef](#)]
68. Lee, M.-J.; Lee, C.B.; Lee, D.; Lee, S.R.; Chang, M.; Hur, J.H.; Kim, Y.-B.; Kim, C.-J.; Seo, D.H.; Seo, S.; et al. A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta<sub>2</sub>O<sub>5-x</sub>/TaO<sub>2-x</sub> bilayer structures. *Nat. Mater.* **2011**, *10*, 625–630. [[CrossRef](#)] [[PubMed](#)]
69. Park, S.-G.; Magyari-Köpe, B.; Nishi, Y. Impact of oxygen vacancy ordering on the formation of a conductive filament in TiO<sub>2</sub> for resistive switching memory. *IEEE Electron Device Lett.* **2011**, *32*, 197–199. [[CrossRef](#)]
70. Bricalli, A.; Ambrosi, E.; Laudato, M.; Maestro, M.; Rodriguez, R.; Ielmini, D. Resistive switching device technology based on silicon oxide for improved on-off ratio—Part I: Memory devices. *IEEE Trans. Electron Devices* **2018**, *65*, 115–121. [[CrossRef](#)]
71. Chien, W.C.; Chen, Y.C.; Lai, E.K.; Lee, F.M.; Lin, Y.Y.; Chuang, A.T.H.; Chang, K.P.; Yao, Y.D.; Chou, T.H.; Lin, H.M.; et al. A study of switching mechanism and electrode material of fully CMOS compatible tungsten oxide ReRAM. *Appl. Phys. A* **2011**, *102*, 901–907. [[CrossRef](#)]
72. Kozicki, M.N.; Barnaby, H.J. Conductive bridge random access memory—materials, devices and applications. *Semicond. Sci. Technol.* **2016**, *31*, 113001. [[CrossRef](#)]
73. Russo, U.; Ielmini, D.; Cagli, C.; Lacaita, A.L. Filament conduction and reset mechanism in NiO-based resistive-switching memory (RRAM) devices. *IEEE Trans. Electron Devices* **2009**, *56*, 186–192. [[CrossRef](#)]
74. Lee, H.D.; Magyari-Köpe, B.; Nishi, Y. Model of metallic filament formation and rupture in NiO for unipolar switching. *Phys. Rev. B* **2010**, *81*, 193202. [[CrossRef](#)]
75. Ielmini, D.; Bruchhaus, R.; Waser, R. Thermochemical resistive switching: Materials, mechanisms and scaling projections. *Phase Transit.* **2011**, *84*, 570–602. [[CrossRef](#)]
76. Sawa, A. Resistive switching in transition metal oxides. *Mater. Today* **2008**, *11*, 28–36. [[CrossRef](#)]
77. Russo, U.; Kamalanathan, D.; Ielmini, D.; Lacaita, A.L.; Kozicki, M.N. Study of multilevel programming in Programmable Metallization Cell (PMC) memory. *IEEE Trans. Electron Devices* **2009**, *56*, 1040–1047. [[CrossRef](#)]

78. Balatti, S.; Larentis, S.; Gilmer, D.C.; Ielmini, D. Multiple memory states in resistive switching devices through controlled size and orientation of the conductive filament. *Adv. Mater.* **2013**, *25*, 1474–1478. [[CrossRef](#)] [[PubMed](#)]
79. Yu, S.; Gao, B.; Fang, Z.; Yu, H.; Kang, J.; Wong, H.-S.P. A low energy oxide-based electronic synaptic device for neuromorphic visual systems with tolerance to device variation. *Adv. Mater.* **2013**, *25*, 1774–1779. [[CrossRef](#)]
80. Zhao, L.; Chen, H.-Y.; Wu, S.-C.; Jiang, Z.; Yu, S.; Hou, T.-H.; Wong, H.-S.P.; Nishi, Y. Multi-level control of conductive nano-filament evolution in HfO<sub>2</sub> ReRAM by pulse-train operations. *Nanoscale* **2014**, *6*, 5698–5702. [[CrossRef](#)]
81. Prakash, A.; Park, J.; Song, J.; Woo, J.; Cha, E.-J.; Hwang, H. Demonstration of low power 3-bit multilevel cell characteristics in a TaO<sub>x</sub>-based RRAM by stack engineering. *IEEE Electron Device Lett.* **2015**, *36*, 32–34. [[CrossRef](#)]
82. Govoreanu, B.; Kar, G.S.; Chen, Y.-Y.; Paraschiv, V.; Kubicek, S.; Fantini, A.; Radu, I.P.; Goux, L.; Clima, S.; Degraeve, R.; et al. 10 × 10 nm<sup>2</sup> Hf/HfO<sub>x</sub> crossbar resistive RAM with excellent performance, reliability and low-energy operation. In Proceedings of the 2011 IEEE International Electron Devices Meeting (IEDM), Washington, DC, USA, 5–7 December 2011; pp. 729–732. [[CrossRef](#)]
83. Baek, I.G.; Park, C.J.; Ju, H.; Seong, D.J.; Ahn, H.S.; Kim, J.H.; Yang, M.K.; Song, S.H.; Kim, E.M.; Park, S.O.; et al. Realization of vertical resistive memory (VRRAM) using cost effective 3D process. In Proceedings of the 2011 IEEE International Electron Devices Meeting (IEDM), Washington, DC, USA, 5–7 December 2011; pp. 737–740. [[CrossRef](#)]
84. Yamada, N.; Ohno, E.; Nishiuchi, K.; Akahira, N.; Takao, M. Rapid-phase transitions of GeTe-Sb<sub>2</sub>Te<sub>3</sub> pseudobinary amorphous thin films for an optical disk memory. *J. Appl. Phys.* **1991**, *69*, 2849–2856. [[CrossRef](#)]
85. Raoux, S.; Welnic, W.; Ielmini, D. Phase change materials and their application to non-volatile memories. *Chem. Rev.* **2010**, *110*, 240–267. [[CrossRef](#)]
86. Burr, G.W.; BrightSky, M.J.; Sebastian, A.; Cheng, H.Y.; Wu, J.Y.; Kim, S.; Sosa, N.E.; Papandreou, N.; Lung, H.-L.; Pozidis, H.; et al. Recent progress in Phase-Change Memory technology. *IEEE J. Emerg. Sel. Top. Circuits Syst. JETCAS* **2016**, *6*, 146–162. [[CrossRef](#)]
87. Fong, S.W.; Neumann, C.M.; Wong, H.-S.P. Phase-Change Memory—Towards a storage-class memory. *IEEE Trans. Electron Devices* **2017**, *64*, 4374–4385. [[CrossRef](#)]
88. Ielmini, D.; Lacaíta, A.L.; Pirovano, A.; Pellizzer, F.; Bez, R. Analysis of phase distribution in phase-change nonvolatile memories. *IEEE Electron Device Lett.* **2004**, *25*, 507–509. [[CrossRef](#)]
89. Athmanathan, A.; Stanisavljevic, M.; Papandreou, N.; Pozidis, H.; Eleftheriou, E. Multilevel-cell Phase-Change Memory: A viable technology. *IEEE J. Emerg. Sel. Top. Circuits Syst. JETCAS* **2016**, *6*, 87–100. [[CrossRef](#)]
90. Chen, Y.C.; Rettner, C.T.; Raoux, S.; Burr, G.W.; Chen, S.H.; Shelby, R.M.; Salinga, M.; Risk, W.P.; Happ, T.D.; McClelland, G.M.; et al. Ultra-thin phase-change bridge memory device using GeSb. In Proceedings of the 2006 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 11–13 December 2006; pp. 1–4. [[CrossRef](#)]
91. Morikawa, T.; Kurotsuchi, K.; Kinoshita, M.; Matsuzaki, N.; Matsui, Y.; Fujisaki, Y.; Hanzawa, S.; Kotabe, A.; Terao, M.; Moriya, H.; et al. Doped In-Ge-Te Phase Change Memory featuring stable operation and good data retention. In Proceedings of the 2007 IEEE International Electron Devices Meeting (IEDM), Washington, DC, USA, 10–12 December 2007; pp. 307–310. [[CrossRef](#)]
92. Zuliani, P.; Varesi, E.; Palumbo, E.; Borghi, M.; Tortorelli, I.; Erbetta, D.; Dalla Libera, G.; Pessina, N.; Gandolfo, A.; Prelini, C.; et al. Overcoming temperature limitations in phase change memories with optimized Ge<sub>x</sub>Sb<sub>y</sub>Te<sub>z</sub>. *IEEE Trans. Electron Devices* **2013**, *60*, 4020–4026. [[CrossRef](#)]
93. Chappert, C.; Fert, A.; Van Dau, F.N. The emergence of spin electronics in data storage. *Nat. Mater.* **2007**, *6*, 813–823. [[CrossRef](#)]
94. Kent, A.D.; Worledge, D.C. A new spin on magnetic memories. *Nat. Nanotech.* **2015**, *10*, 187–191. [[CrossRef](#)]
95. Locatelli, N.; Cros, V.; Grollier, J. Spin-torque building blocks. *Nat. Mater.* **2014**, *13*, 11–20. [[CrossRef](#)]
96. Julliere, M. Tunneling between ferromagnetic films. *Phys. Lett. A* **1975**, *54*, 225–226. [[CrossRef](#)]
97. Slonczewski, J.C. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **1996**, *159*, L1–L7. [[CrossRef](#)]

98. Berger, L. Emission of spin waves by a magnetic multilayer traversed by a current. *Phys. Rev. B* **1996**, *54*, 9353–9358. [[CrossRef](#)] [[PubMed](#)]
99. Novak, J.J. Dependence of voltage and size on write error rates in spin-transfer torque magnetic random-access memory. *IEEE Magn. Lett.* **2016**, *7*, 1–4. [[CrossRef](#)]
100. Saida, D.; Kashiwada, S.; Yakabe, M.; Daibou, T.; Hase, N.; Fukumoto, M.; Miwa, S.; Suzuki, Y.; Nuguchi, H.; Fujita, S.; et al. Sub-3 ns pulse with sub-100  $\mu$ A switching of 1x-2x nm perpendicular MTJ for high-performance embedded STT-MRAM towards sub-20 nm CMOS. In Proceedings of the 2016 IEEE Symposium on VLSI Technology, Honolulu, HI, USA, 14–16 June 2016; pp. 1–2. [[CrossRef](#)]
101. Carboni, R.; Ambrogio, S.; Chen, W.; Siddik, M.; Harms, J.; Lyle, A.; Kula, W.; Sandhu, G.; Ielmini, D. Understanding cycling endurance in perpendicular spin-transfer torque (p-STT) magnetic memory. In Proceedings of the 2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 3–7 December 2016; pp. 572–575. [[CrossRef](#)]
102. Kan, J.J.; Park, C.; Ching, C.; Ahn, J.; Xie, Y.; Pakala, M.; Kang, S.H. A study on practically unlimited endurance of STT-MRAM. *IEEE Trans. Electron Devices* **2017**, *64*, 3639–3646. [[CrossRef](#)]
103. Bösccke, T.S.; Mueller, J.; Brauhaus, D.; Schroeder, U.; Boettger, U. Ferroelectricity in hafnium oxide thin films. *Appl. Phys. Lett.* **2011**, *99*, 102903. [[CrossRef](#)]
104. Takashima, D.; Takeuchi, Y.; Miyakawa, T.; Itoh, Y.; Ogiwara, R.; Kamoshida, M.; Hoya, K.; Doumae, S.M.; Ozaki, T.; Kanaya, H.; et al. A 76-mm<sup>2</sup> 8-Mb chain ferroelectric memory. *IEEE J. Solid State Circuits* **2001**, *36*, 1713–1720. [[CrossRef](#)]
105. Sakai, S.; Takahashi, M.; Takeuchi, K.; Li, Q.H.; Horiuchi, T.; Wang, S.; Yun, K.Y.; Takamiya, M.; Sakurai, T. Highly scalable Fe(Ferroelectric)-NAND Cell with MFIS(Metal-Ferroelectric-Insulator-Semiconductor) structure for sub-10nm Tera-bit capacity NAND Flash memories. In Proceedings of the Joint Non-Volatile Semiconductor Memory Workshop and International Conference on Memory Technology and Design, Opio, France, 18–22 May 2008; pp. 103–105. [[CrossRef](#)]
106. Mikolajick, T.; Dehm, C.; Hartner, W.; Kasko, I.; Kastner, M.J.; Nagel, N.; Moert, M.; Mazure, C. FeRAM technology for high density applications. *Microelectron. Reliab.* **2001**, *41*, 947–950. [[CrossRef](#)]
107. Mulaosmanovic, H.; Ocker, J.; Müller, S.; Noack, M.; Müller, J.; Polakowski, P.; Mikolajick, T.; Slesazeck, S. Novel ferroelectric FET based synapse for neuromorphic systems. In Proceedings of the 2017 IEEE Symposium on VLSI Technology, Kyoto, Japan, 5–8 June 2017; pp. T176–T177. [[CrossRef](#)]
108. Tang, J.; Bishop, D.; Kim, S.; Copel, M.; Gokmen, T.; Todorov, T.; Shin, S.H.; Lee, K.-T.; Solomon, P.; Chan, K.; et al. ECRAM as scalable synaptic cell for high-speed, low-power neuromorphic computing. In Proceedings of the 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 1–5 December 2018; pp. 292–295. [[CrossRef](#)]
109. Cubukcu, M.; Boulle, O.; Drouard, M.; Garelo, K.; Avci, C.O.; Miron, I.M.; Langer, J.; Ocker, B.; Gambardella, P.; Gaudin, G. Spin-orbit torque magnetization switching of a three-terminal perpendicular magnetic tunnel junction. *Appl. Phys. Lett.* **2014**, *104*, 042406. [[CrossRef](#)]
110. Sangwan, V.K.; Lee, H.-S.; Bergeron, H.; Balla, I.; Beck, M.E.; Chen, K.-S.; Hersam, M.C. Multi-terminal memtransistors from polycrystalline monolayer molybdenum disulfide. *Nature* **2018**, *544*, 500–504. [[CrossRef](#)]
111. Zhu, X.; Li, D.; Liang, X.; Lu, W.D. Ionic modulation and ionic coupling effects in MoS<sub>2</sub> devices for neuromorphic computing. *Nat. Mater.* **2019**, *18*, 141–148. [[CrossRef](#)]
112. Bhowmik, D.; Saxena, U.; Dankar, A.; Verma, A.; Kaushik, D.; Chatterjee, S.; Singh, U. On-chip learning for domain wall synapse based fully connected neural network. *J. Magn. Magn. Mater.* **2019**, *489*, 165434. [[CrossRef](#)]
113. Sharad, M.; Augustine, C.; Panagopoulos, G.; Roy, K. Spin-based neuron model with domain wall magnets as synapse. *IEEE Trans. Nanotech.* **2012**, *11*, 843–853. [[CrossRef](#)]
114. Trentzsch, M.; Flachowsky, S.; Richter, R.; Paul, J.; Reimer, B.; Utes, D.; Jansen, S.; Mulaosmanovic, H.; Müller, S.; Slesazeck, S.; et al. A 28nm HKMG super low power embedded NVM technology based on ferroelectric FETs. In Proceedings of the 2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 3–7 December 2016; pp. 294–297. [[CrossRef](#)]
115. Florent, K.; Pesic, M.; Subirats, A.; Banerjee, K.; Lavizzari, S.; Arreghini, A.; Di Piazza, L.; Potoms, G.; Sebaai, F.; McMitchell, S.R.C.; et al. Vertical ferroelectric HfO<sub>2</sub> FET based on 3-D NAND architecture: Towards dense low-power memory. In Proceedings of the 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 1–5 December 2018; pp. 43–46. [[CrossRef](#)]

116. Jerry, M.; Chen, P.-Y.; Zhang, J.; Sharma, P.; Ni, K.; Yu, S.; Datta, S. Ferroelectric FET analog synapse for acceleration of deep neural network training. In Proceedings of the 2017 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2–6 December 2017; pp. 139–142. [[CrossRef](#)]
117. Mulaosmanovic, H.; Chicca, E.; Bertele, M.; Mikolajick, T.; Slesazek, S. Mimicking biological neurons with a nanoscale ferroelectric transistor. *Nanoscale* **2018**, *10*, 21755–21763. [[CrossRef](#)] [[PubMed](#)]
118. Fang, Y.; Gomez, J.; Wang, Z.; Datta, S.; Khan, A.I.; Raychowdhury, A. Neuro-mimetic dynamics of a ferroelectric FET-based spiking neuron. *IEEE Electron Device Lett.* **2019**, *40*, 1213–1216. [[CrossRef](#)]
119. Fuller, E.J.; El Gabaly, F.; Leonard, F.; Agarwal, S.; Plimpton, S.J.; Jacobs-Gedrim, R.B.; James, C.D.; Marinella, M.J.; Talin, A.A. Li-ion synaptic transistor for low power analog computing. *Adv. Mater.* **2017**, *29*, 1604310. [[CrossRef](#)] [[PubMed](#)]
120. Van de Burgt, Y.; Lubberman, E.; Fuller, E.J.; Keene, S.T.; Faria, G.C.; Agarwal, S.; Marinella, M.J.; Talin, A.A.; Salleo, A. A non-volatile organic electrochemical device as a low-voltage artificial synapse for neuromorphic computing. *Nat. Mater.* **2017**, *16*, 414–418. [[CrossRef](#)] [[PubMed](#)]
121. Garello, K.; Avci, C.O.; Miron, I.M.; Baumgartner, M.; Ghosh, A.; Auffret, S.; Boule, O.; Gaudin, G.; Gambardella, P. Ultrafast magnetization switching by spin-orbit torques. *Appl. Phys. Lett.* **2014**, *105*, 212402. [[CrossRef](#)]
122. Lo Conte, R.; Hrabec, A.; Mihai, A.P.; Schulz, T.; Noh, S.-J.; Marrows, C.H.; Moore, T.A.; Kläui, M. Spin-orbit torque-driven magnetization switching and thermal effects studied in Ta\CoFeB\MgO nanowires. *Appl. Phys. Lett.* **2014**, *105*, 122404. [[CrossRef](#)]
123. Garello, K.; Miron, I.M.; Avci, C.O.; Freimuth, F.; Mokrousov, Y.; Blügel, S.; Auffret, S.; Boule, O.; Gaudin, G.; Gambardella, P. Symmetry and magnitude of spin-orbit torques in ferromagnetic heterostructures. *Nat. Nanotechnol.* **2013**, *8*, 587–593. [[CrossRef](#)]
124. Miron, I.M.; Garello, K.; Gaudin, G.; Zermatten, P.-J.; Costache, M.V.; Auffret, S.; Bandiera, S.; Rodmacq, B.; Schuhl, A.; Gambardella, P. Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection. *Nature* **2011**, *476*, 189–193. [[CrossRef](#)]
125. Borders, W.A.; Fukami, S.; Ohno, H. Characterization of spin-orbit torque-controlled synapse device for artificial neural network applications. *Jpn. J. Appl. Phys.* **2018**, *57*, 1002B2. [[CrossRef](#)]
126. Sengupta, A.; Choday, S.H.; Kim, Y.; Roy, K. Spin orbit torque based electronic neuron. *Appl. Phys. Lett.* **2015**, *106*, 143701. [[CrossRef](#)]
127. Borders, W.A.; Akima, H.; Fukami, S.; Moriya, S.; Kurihara, S.; Horio, Y.; Sato, S.; Ohno, H. Analogue spin-orbit torque device for artificial-neural-network-based associative memory operation. *Appl. Phys. Express* **2017**, *10*, 013007. [[CrossRef](#)]
128. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1026–1034. [[CrossRef](#)]
129. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. DeepFace: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708. [[CrossRef](#)]
130. Xiong, W.; Droppo, J.; Huang, X.; Seide, F.; Seltzer, M.L.; Stolcke, A.; Yu, D.; Zweig, G. Toward human parity in conversational speech recognition. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 2410–2423. [[CrossRef](#)]
131. Grossberg, S. Competitive learning: From interactive activation to adaptive resonance. *Cogn. Sci.* **1987**, *11*, 23–63. [[CrossRef](#)]
132. Whittington, J.C.R.; Bogacz, R. Theories of error back-propagation in the brain. *Trends Cogn. Sci.* **2019**, *23*, 235–250. [[CrossRef](#)] [[PubMed](#)]
133. Burr, G.W.; Shelby, R.M.; Sidler, S.; di Nolfo, C.; Jang, J.; Boybat, I.; Shenoy, R.S.; Narayanan, P.; Virwani, K.; Giacometti, E.U.; et al. Experimental demonstration and tolerancing of a large-scale neural network (165,000 synapses) using phase-change memory as the synaptic weight element. *IEEE Trans. Electron Devices* **2015**, *62*, 3498–3507. [[CrossRef](#)]
134. Woo, J.; Yu, S. Resistive memory-based analog synapse: The pursuit for linear and symmetric weight update. *IEEE Nanotechnol. Mag.* **2018**, *12*, 36–44. [[CrossRef](#)]

135. Fuller, E.J.; Keene, S.T.; Melianas, A.; Wang, Z.; Agarwal, S.; Li, Y.; Tuchman, Y.; James, C.D.; Marinella, M.J.; Yang, J.J.; et al. Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing. *Science* **2019**, *364*, 570–574. [[CrossRef](#)]
136. Yao, P.; Wu, H.; Gao, B.; Eryilmaz, S.B.; Huang, X.; Zhang, W.; Zhang, Q.; Deng, N.; Shi, L.; Wong, H.-S.P.; et al. Face classification using electronic synapses. *Nat. Commun.* **2017**, *8*, 15199. [[CrossRef](#)]
137. Li, C.; Belkin, D.; Li, Y.; Yan, P.; Hu, M.; Ge, N.; Jiang, H.; Montgomery, E.; Lin, P.; Wang, Z.; et al. Efficient and self-adaptive in-situ learning in multilayer memristor neural networks. *Nat. Commun.* **2018**, *9*, 2385. [[CrossRef](#)]
138. Ambrogio, S.; Narayanan, P.; Tsai, H.; Shelby, R.M.; Boybat, I.; di Nolfo, C.; Sidler, S.; Giordano, M.; Bodini, M.; Farinha, N.C.P.; et al. Equivalent-accuracy accelerated neural-network training using analogue memory. *Nature* **2018**, *558*, 60–67. [[CrossRef](#)] [[PubMed](#)]
139. Krizhevsky, A. Learning Multiple Layers of Features from Tiny Images. 2009. Ch. 3. Available online: <https://www.cs.toronto.edu/~kriz/cifar.html> (accessed on 20 December 2019).
140. Milo, V.; Zambelli, C.; Olivo, P.; Pérez, E.; Mahadevaiah, M.K.; Ossorio, O.G.; Wenger, C.; Ielmini, D. Multilevel HfO<sub>2</sub>-based RRAM devices for low-power neuromorphic networks. *APL Mater.* **2019**, *7*, 081120. [[CrossRef](#)]
141. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484–489. [[CrossRef](#)] [[PubMed](#)]
142. Jo, S.H.; Chang, T.; Ebong, I.; Bhadviya, B.B.; Mazumder, P.; Lu, W. Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett.* **2010**, *10*, 1297–1301. [[CrossRef](#)]
143. Yu, S.; Wu, Y.; Jeyasingh, R.; Kuzum, D.; Wong, H.-S.P. An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation. *IEEE Trans. Electron Devices* **2011**, *58*, 2729–2737. [[CrossRef](#)]
144. Kuzum, D.; Jeyasingh, R.G.D.; Lee, B.; Wong, H.-S.P. Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing. *Nano Lett.* **2012**, *12*, 2179–2186. [[CrossRef](#)]
145. Serb, A.; Bill, J.; Khiat, A.; Berdan, R.; Legenstein, R.; Prodromakis, T. Unsupervised learning in probabilistic neural networks with multi-state metal-oxide memristive synapses. *Nat. Commun.* **2016**, *7*, 12611. [[CrossRef](#)]
146. Prezioso, M.; Mahmoodi, M.R.; Merrikh-Bayat, F.; Nili, H.; Kim, H.; Vincent, A.; Strukov, D.B. Spike-timing-dependent plasticity learning of coincidence detection with passively integrated memristive circuits. *Nat. Commun.* **2018**, *9*, 5311. [[CrossRef](#)]
147. Hansen, M.; Zahari, F.; Kohlstedt, H.; Ziegler, M. Unsupervised Hebbian learning experimentally realized with analogue memristive crossbar arrays. *Sci. Rep.* **2018**, *8*, 8914. [[CrossRef](#)]
148. Boybat, I.; Le Gallo, M.; Nandakumar, S.R.; Moraitis, T.; Parnell, T.; Tuma, T.; Rajendran, B.; Leblebici, Y.; Sebastian, A.; Eleftheriou, E. Neuromorphic computing with multi-memristive synapses. *Nat. Commun.* **2018**, *9*, 2514. [[CrossRef](#)]
149. Vincent, A.F.; Larroque, J.; Locatelli, N.; Romdhane, N.B.; Bichler, O.; Gamrat, C.; Zhao, W.S.; Klein, J.-O.; Galdin-Retailleau, S.; Querlioz, D. Spin-transfer-torque magnetic memory as a stochastic memristive synapse for neuromorphic systems. *IEEE Trans. Biomed. Circ. Syst.* **2015**, *9*, 166–174. [[CrossRef](#)] [[PubMed](#)]
150. Ambrogio, S.; Ciocchini, N.; Laudato, M.; Milo, V.; Pirovano, A.; Fantini, P.; Ielmini, D. Unsupervised learning by spike timing dependent plasticity in phase change memory (PCM) synapses. *Front. Neurosci.* **2016**, *10*, 56. [[CrossRef](#)] [[PubMed](#)]
151. Ambrogio, S.; Balatti, S.; Milo, V.; Carboni, R.; Wang, Z.; Calderoni, A.; Ramaswamy, N.; Ielmini, D. Neuromorphic learning and recognition with one-transistor-one-resistor synapses and bistable metal oxide RRAM. *IEEE Trans. Electron Devices* **2016**, *63*, 1508–1515. [[CrossRef](#)]
152. Pedretti, G.; Milo, V.; Ambrogio, S.; Carboni, R.; Bianchi, S.; Calderoni, A.; Ramaswamy, N.; Spinelli, A.S.; Ielmini, D. Memristive neural network for on-line learning and tracking with brain-inspired spike timing dependent plasticity. *Sci. Rep.* **2017**, *7*, 5288. [[CrossRef](#)] [[PubMed](#)]
153. Prezioso, M.; Merrikh-Bayat, F.; Hoskins, B.; Likharev, K.; Strukov, D. Self-adaptive spike-timing-dependent plasticity of metal-oxide memristors. *Sci. Rep.* **2016**, *6*, 21331. [[CrossRef](#)] [[PubMed](#)]

154. Milo, V.; Pedretti, G.; Carboni, R.; Calderoni, A.; Ramaswamy, N.; Ambrogio, S.; Ielmini, D. Demonstration of hybrid CMOS/RRAM neural networks with spike time/rate-dependent plasticity. In Proceedings of the 2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 3–7 December 2016; pp. 440–443. [[CrossRef](#)]
155. Wang, W.; Pedretti, G.; Milo, V.; Carboni, R.; Calderoni, A.; Ramaswamy, N.; Spinelli, A.S.; Ielmini, D. Learning of spatiotemporal patterns in a spiking neural network with resistive switching synapses. *Sci. Adv.* **2018**, *4*, eaat4752. [[CrossRef](#)]
156. Suri, M.; Bichler, O.; Querlioz, D.; Palma, G.; Vianello, E.; Vuillaume, D.; Gamrat, C.; DeSalvo, B. CBRAM devices as binary synapses for low-power stochastic neuromorphic systems: Auditory (cochlea) and visual (retina) cognitive processing applications. In Proceedings of the 2012 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 10–13 December 2012; pp. 235–238. [[CrossRef](#)]
157. Suri, M.; Bichler, O.; Querlioz, D.; Cueto, O.; Perniola, L.; Sousa, V.; Vuillaume, D.; Gamrat, C.; DeSalvo, B. Phase change memory as synapse for ultra-dense neuromorphic systems: Application to complex visual pattern extraction. In Proceedings of the 2011 IEEE International Electron Devices Meeting (IEDM), Washington, DC, USA, 5–7 December 2011; pp. 79–82. [[CrossRef](#)]
158. Ambrogio, S.; Balatti, S.; Milo, V.; Carboni, R.; Wang, Z.; Calderoni, A.; Ramaswamy, N.; Ielmini, D. Novel RRAM-enabled 1T1R synapse capable of low-power STDP via burst-mode communication and real-time unsupervised machine learning. In Proceedings of the IEEE Symposium on VLSI Technology, Honolulu, HI, USA, 14–16 June 2016; pp. 1–2. [[CrossRef](#)]
159. Garbin, D.; Vianello, E.; Bichler, O.; Rafhay, Q.; Gamrat, C.; Ghibaud, G.; De Salvo, B.; Perniola, L. HfO<sub>2</sub>-based OxRAM devices as synapses for convolutional neural networks. *IEEE Trans. Electron Devices* **2015**, *62*, 2494–2501. [[CrossRef](#)]
160. Muñoz-Martín, I.; Bianchi, S.; Pedretti, G.; Melnic, O.; Ambrogio, S.; Ielmini, D. Unsupervised learning to overcome catastrophic forgetting in neural networks. *IEEE J. Exp. Solid State Comput. Devices Circuits* **2019**, *5*, 58–66. [[CrossRef](#)]
161. Eryilmaz, S.B.; Kuzum, D.; Jeyasingh, R.; Kim, S.B.; BrightSky, M.; Lam, C.; Wong, H.-S.P. Brain-like associative learning using a nanoscale non-volatile phase change synaptic device array. *Front. Neurosci.* **2014**, *8*, 205. [[CrossRef](#)]
162. Milo, V.; Ielmini, D.; Chicca, E. Attractor networks and associative memories with STDP learning in RRAM synapses. In Proceedings of the 2017 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2–6 December 2017; pp. 263–266. [[CrossRef](#)]
163. Milo, V.; Chicca, E.; Ielmini, D. Brain-inspired recurrent neural network with plastic RRAM synapses. In Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 27–30 May 2018; pp. 1–5. [[CrossRef](#)]
164. Pedretti, G.; Milo, V.; Ambrogio, S.; Carboni, R.; Bianchi, S.; Calderoni, A.; Ramaswamy, N.; Spinelli, A.S.; Ielmini, D. Stochastic learning in neuromorphic hardware via spike timing dependent plasticity with RRAM synapses. *IEEE J. Emerg. Sel. Top. Circuits Syst. JETCAS* **2018**, *8*, 77–85. [[CrossRef](#)]
165. Masquelier, T.; Guyonneau, R.; Thorpe, S.J. Competitive STDP-based spike pattern learning. *Neural Comput.* **2009**, *21*, 1259–1276. [[CrossRef](#)] [[PubMed](#)]
166. Nair, M.V.; Muller, L.K.; Indiveri, G. A differential memristive synapse circuit for on-line learning in neuromorphic computing systems. *Nano Futures* **2017**, *1*, 035003. [[CrossRef](#)]
167. Pershin, Y.V.; Di Ventra, M. Neuromorphic, digital, and quantum computation with memory circuit elements. *Proc. IEEE* **2012**, *100*, 2071–2080. [[CrossRef](#)]
168. Kim, S.; Du, C.; Sheridan, P.; Ma, W.; Choi, S.H.; Lu, W.D. Experimental demonstration of a second-order memristor and its ability to biorealistically implement synaptic plasticity. *Nano Lett.* **2015**, *15*, 2203–2211. [[CrossRef](#)]
169. Zucker, R.S.; Regehr, W.G. Short-term synaptic plasticity. *Ann. Rev. Physiol.* **2002**, *64*, 355–405. [[CrossRef](#)]
170. Markram, H.; Wang, Y.; Tsodyks, M. Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5323–5328. [[CrossRef](#)]
171. Wang, Z.; Joshi, S.; Savel'ev, S.E.; Jiang, H.; Midya, R.; Lin, P.; Hu, M.; Ge, N.; Strachan, J.P.; Li, Z.; et al. Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. *Nat. Mater.* **2017**, *16*, 101–108. [[CrossRef](#)]

172. Ohno, T.; Hasegawa, T.; Tsuruoka, T.; Terabe, K.; Gimzewski, J.K.; Aono, M. Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nat. Mater.* **2011**, *10*, 591–595. [[CrossRef](#)]
173. Wang, W.; Bricalli, A.; Laudato, M.; Ambrosi, E.; Covi, E.; Ielmini, D. Physics-based modeling of volatile resistive switching memory (RRAM) for crosspoint selector and neuromorphic computing. In Proceedings of the 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 1–5 December 2018; pp. 932–935. [[CrossRef](#)]
174. Wang, W.; Laudato, M.; Ambrosi, E.; Bricalli, A.; Covi, E.; Lin, Y.-H.; Ielmini, D. Volatile resistive switching memory based on Ag ion drift/diffusion—Part II: Compact modeling. *IEEE Trans. Electron Devices* **2019**, *66*, 3802–3808. [[CrossRef](#)]
175. Wang, Z.; Joshi, S.; Savel'ev, S.; Song, W.; Midya, R.; Li, Y.; Rao, M.; Yan, P.; Asapu, S.; Zhuo, Y.; et al. Fully memristive neural networks for pattern classification with unsupervised learning. *Nat. Electron.* **2018**, *1*, 137–145. [[CrossRef](#)]
176. Cristiano, G.; Giordano, M.; Ambrogio, S.; Romero, L.P.; Cheng, C.; Narayanan, P.; Tsai, H.; Shelby, R.M.; Burr, G.W. Perspective on training fully connected networks with resistive memories: Device requirements for multiple conductances of varying significance. *J. Appl. Phys.* **2018**, *124*, 151901. [[CrossRef](#)]
177. Nicosia, G.; Paolucci, G.M.; Monzio Compagnoni, C.; Resnati, D.; Miccoli, C.; Spinelli, A.S.; Lacaita, A.L.; Visconti, A.; Goda, A. A single-electron analysis of NAND Flash memory programming. In Proceedings of the 2015 IEEE International Electron Devices Meeting (IEDM), Washington, DC, USA, 7–9 December 2015; pp. 378–381. [[CrossRef](#)]
178. Ambrogio, S.; Balatti, S.; Cubeta, A.; Calderoni, A.; Ramaswamy, N.; Ielmini, D. Statistical fluctuations in HfO<sub>x</sub> resistive-switching memory (RRAM): Part I—Set/Reset variability. *IEEE Trans. Electron Devices* **2014**, *61*, 2912–2919. [[CrossRef](#)]
179. Ielmini, D.; Lacaita, A.L.; Mantegazza, D. Recovery and drift dynamics of resistance and threshold voltages in phase change memories. *IEEE Trans. Electron Devices* **2007**, *54*, 308–315. [[CrossRef](#)]
180. Lashkare, S.; Chouhan, S.; Chavan, T.; Bhat, A.; Kumbhare, P.; Ganguly, U. PCMO RRAM for integrate-and-fire neuron in spiking neural networks. *IEEE Electron Device Lett.* **2018**, *39*, 484–487. [[CrossRef](#)]
181. Pickett, M.D.; Medeiros-Ribeiro, G.; Williams, R.S. A scalable neuristor built with Mott memristors. *Nat. Mater.* **2013**, *12*, 114–117. [[CrossRef](#)]
182. Tuma, T.; Pantazi, A.; Le Gallo, M.; Sebastian, A.; Eleftheriou, E. Stochastic phase-change neurons. *Nat. Nanotechnol.* **2016**, *11*, 693–699. [[CrossRef](#)]
183. Torrejon, J.; Riou, M.; Araujo, F.A.; Tsunegi, S.; Khalsa, G.; Querlioz, D.; Bortolotti, P.; Cros, V.; Yakushiji, K.; Fukushima, A.; et al. Neuromorphic computing with nanoscale spintronic oscillators. *Nature* **2017**, *547*, 428–431. [[CrossRef](#)]
184. Wu, M.-H.; Hong, M.-C.; Chang, C.-C.; Sahu, P.; Wei, J.-H.; Lee, H.-Y.; Sheu, S.-S.; Hou, T.-H. Extremely compact integrate-and-fire STT-MRAM neuron: A pathway toward all-spin artificial deep neural network. In Proceedings of the IEEE Symposium on VLSI Technology, Kyoto, Japan, 9–14 June 2019; pp. T34–T35. [[CrossRef](#)]
185. Mizrahi, A.; Hirtzlin, T.; Fukushima, A.; Kubota, H.; Yuasa, S.; Grollier, J.; Querlioz, D. Neural-like computing with populations of superparamagnetic basis functions. *Nat. Commun.* **2018**, *9*, 1533. [[CrossRef](#)]
186. Wittenberg, G.M.; Sullivan, M.R.; Tsien, J.Z. Synaptic reentry reinforcement based network model for long-term memory consolidation. *Hippocampus* **2002**, *12*, 637–647. [[CrossRef](#)]
187. Wang, Z.; Li, C.; Song, W.; Rao, M.; Belkin, D.; Li, Y.; Yan, P.; Jiang, H.; Lin, P.; Hu, M.; et al. Reinforcement learning with analogue memristor arrays. *Nat. Electron.* **2019**, *2*, 115–124. [[CrossRef](#)]
188. Hopfield, J.J. Searching for memories, Sudoku, implicit check bits, and the iterative use of not-always-correct rapid neural computation. *Neural Comput.* **2008**, *20*, 1119–1164. [[CrossRef](#)] [[PubMed](#)]
189. Mostafa, H.; Müller, L.K.; Indiveri, G. An event-based architecture for solving constraint satisfaction problems. *Nat. Commun.* **2015**, *6*, 8941. [[CrossRef](#)] [[PubMed](#)]
190. Traversa, F.L.; Ramella, C.; Bonani, F.; Di Ventra, M. Memcomputing NP-complete problems in polynomial time using polynomial resources and collective states. *Sci. Adv.* **2015**, *1*, e1500031. [[CrossRef](#)] [[PubMed](#)]
191. Denchev, V.S.; Boixo, S.; Isakov, S.V.; Ding, N.; Babbush, R.; Smelyanskiy, V.; Martinis, J.; Neven, H. What is the computational value of finite-range tunneling? *Phys. Rev. X* **2016**, *6*, 031015. [[CrossRef](#)]

192. Maass, W. Noise as a resource for computation and learning in networks of spiking neurons. *Proc. IEEE* **2014**, *102*, 860–880. [[CrossRef](#)]
193. Cai, F.; Kumar, S.; Van Vaerenbergh, T.; Liu, R.; Li, C.; Yu, S.; Xia, Q.; Yang, J.J.; Beausoleil, R.; Lu, W.; et al. Harnessing intrinsic noise in memristor Hopfield neural networks for combinatorial optimization. *arXiv* **2019**, arXiv:1903.11194.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).