Article

Multiscale Modeling for Application-Oriented Optimization of Resistive Random-Access Memory

Paolo La Torraca [1,*], Francesco Maria Puglisi [2], Andrea Padovani [3] and Luca Larcher [1,3]

[1] Dipartimento di Scienze e Metodi dell'Ingegneria, Università di Modena e Reggio Emilia, Via Amendola 2, 42122 Reggio Emilia, Italy; Luca_Larcher@amat.com
[2] Dipartimento di Ingegneria "Enzo Ferrari", Università di Modena e Reggio Emilia, Via P. Vivarelli 10/1, 41125 Modena, Italy; francescomaria.puglisi@unimore.it
[3] Applied Materials, Via Sicilia 32, 42122 Reggio Emilia, Italy; Andrea_Padovani@amat.com
* Correspondence: paolo.latorraca@unimore.it

Received: 31 July 2019; Accepted: 20 October 2019; Published: 23 October 2019

Abstract: Memristor-based neuromorphic systems have been proposed as a promising alternative to von Neumann computing architectures, which are currently challenged by the ever-increasing computational power required by modern artificial intelligence (AI) algorithms. The design and optimization of memristive devices for specific AI applications is thus of paramount importance, but still extremely complex, as many different physical mechanisms and their interactions have to be accounted for, which are, in many cases, not fully understood. The high complexity of the physical mechanisms involved and their partial comprehension are currently hampering the development of memristive devices and preventing their optimization. In this work, we tackle the application-oriented optimization of Resistive Random-Access Memory (RRAM) devices using a multiscale modeling platform. The considered platform includes all the involved physical mechanisms (i.e., charge transport and trapping, and ion generation, diffusion, and recombination) and accounts for the 3D electric and temperature field in the device. Thanks to its multiscale nature, the modeling platform allows RRAM devices to be simulated and the microscopic physical mechanisms involved to be investigated, the device performance to be connected to the material's microscopic properties and geometries, the device electrical characteristics to be predicted, the effect of the forming conditions (i.e., temperature, compliance current, and voltage stress) on the device's performance and variability to be evaluated, the analog resistance switching to be optimized, and the device's reliability and failure causes to be investigated. The discussion of the presented simulation results provides useful insights for supporting the application-oriented optimization of RRAM technology according to specific AI applications, for the implementation of either non-volatile memories, deep neural networks, or spiking neural networks.

Keywords: AI; neuromorphic computing; multiscale modeling; memristor; optimization; RRAM; simulation

---

1. Introduction

Artificial Neural Networks (ANNs) are possibly the most prominent computational model used in modern artificial intelligence (AI) applications. The computational scheme of ANN, loosely based on biological neural networks, comprises a collection of interconnected processing elements, referred to as artificial neurons, often organized in layers [1,2].

An ever-increasing effort has been devoted to the experimentation of different ANN architectures, which has led to the development of ANNs constituted by an extremely high number of neurons and layers [1,2], known as deep neural networks (DNNs). Thanks to their high flexibility and

application-agnostic nature, DNNs have proved extremely effective in a variety of applications, such as image recognition [3,4] and speech recognition [5], easily outperforming other ANN architectures and machine learning techniques (i.e., Support Vector Machines, K-Nearest Neighbors).

However, the exceptional performance shown by the most advanced DNNs requires a heavy investment in terms of energy and hardware resources, during both the training of the network and the inference [6–8]. In fact, in most practical applications, DNNs are implemented in software and executed on von Neumann computer architectures that have to provide (i) sufficient computational power for fast training and inference, and (ii) a sufficiently large and fast memory for storing all the artificial neuron weights and any partial result. Nonetheless, the energy efficiency of training and inference must also be considered.

In order to achieve fast and efficient DNNs, computation has been progressively shifted from central processing units (CPUs) to optimized coprocessors, referred to as AI accelerators or Neural Processing Units (NPUs), that enable high parallelization by exploiting the DNN layered topology. Graphical Processing Units (GPUs) have readily been used as NPUs for their inherent capability to conduct efficient vector and matrix operations [9], followed by the development of the first Tensor Processing Units (TPUs), capable of efficient tensor operations and better data reuse [10]. Application-specific integrated circuits (ASICs) have recently been developed as special-purpose NPUs, co-designed with the desired DNN, exchanging a low flexibility for a higher energy efficiency, an optimized memory size and use, and a low area occupation [11–14].

Despite their steady improvements, even ASIC NPUs are now facing technological limits (i.e., the imminent end of Moore's law and the already broken Dennard scaling) and the intrinsic limits of the von Neumann architecture, mainly the so-called von Neumann bottleneck (i.e., the limited data transfer speed between the processor and memory, as well as the energy required for the data transfer itself). This is especially true in applications with strict volume, weight, and power constraints, such as automotive, battery-powered vehicles (i.e., drones); mobile devices; and Internet of Things (IoT) devices. In this context, memristor-based neuromorphic systems can potentially overcome the limitations posed by von Neumann architectures, not only to DNNs, but to AI in general.

Memristors are electronic passive devices characterized by a pinched hysteresis I-V curve, thus exhibiting a time-varying and non-volatile electrical resistance [15,16]. Ideally, the conductivity of a memristor can be arbitrarily modified by applying a proper electrical stimulus, indefinitely retaining the resistance state in the absence of external stimuli.

The peculiar memristor characteristics, combined with their extremely small size, have made them very appealing for the development of new non-volatile memories (NVMs), characterized by a low power, high density, and possibly multilevel data storage [17,18].

Even more importantly for AI applications, memristive crossbar arrays have been proposed as DNN accelerators, natively implementing an in-memory (typically analog) vector-matrix multiplication [18–20], which is the fundamental operation for the computation of a layer output in DNNs. By storing the weights of a DNN layer in the memristive array as conductance values, matrix-vector multiplication can be executed at once by only exploiting Ohm's and Kirchhoff's laws. The computation of a single DNN layer requires a single execution step and, moreover, it is performed at the data location, preventing the massive data transfer of weight values between the memory and processor required in von Neumann architectures.

Memristors are also finding applications in more biologically-plausible neural networks, such as Spiking Neural Networks (SNNs), in which the computation is performed by mimicking the actual operation of biological neurons (i.e., spatio-temporal coding of the information, synaptic plasticity, cell membrane V-I relationship, generation of action potentials, and intrinsically stochastic behavior) [18,19,21]. Different memristor-based artificial synapse implementations have been proposed, and have successfully exhibited both deterministic and stochastic spike-time-dependent plasticity (STDP) [21,22]. Their application to SNNs has highlighted the potential for supervised and unsupervised learning, and adaptation to input stimuli. A recent analysis on the Hodgkin–Huxley neuron model [23],

known for being the most biologically-sound model of the neuron action potential, suggested that memristors are key components for its implementation [24]. Moreover, the intrinsic stochasticity of real memristors can be exploited for mimicking the probabilistic and noisy behavior of biological neurons [22], enabling more biologically-plausible artificial neuron implementations and overcoming the limits of the strictly deterministic CMOS implementations. Memristor-based SNNs thus have a huge potential, providing a new architectural and computational paradigm approaching a brain-like computation and circumventing the von Neumann bottleneck at once.

Since the fabrication of the first TiO_2 memristor in 2008 [25], many memristor technologies have been proposed [19,26]. In Resistive Random-Access Memories (RRAMs), the switching is induced by the formation of a conductive path in a dielectric material, controlled by an ion-based mechanism [27–31]. In Phase-Change Memories (PCM), the modulation of a chalcogenide material phase (i.e., amorphous and crystalline) through localized Joule heating allows the resistance switching [32–34]. In a Ferroelectric Tunnel Junction (FTJ), the tunneling electroresistance of a ferroelectric material is modulated by setting its internal polarization [35–37]. In a Magnetic Tunnel Junction (MTJ), the tunneling electroresistance of a thin insulator enclosed between two ferromagnetic layers is modulated by their magnetic polarization [38].

Although the proposed memristor technologies are characterized by a simple structure and are easy to fabricate, the physical principles underlying their analog resistance switching and, in general, the implications of the atomic material properties (i.e., defects, phase, morphology) on the electrical performances, are extremely complex and are still not comprehensively understood. This lack of knowledge is currently hampering the development of memristor-based systems.

In fact, all the memristor applications previously discussed ask for devices with different performance metrics, as summarized in Table 1. Memristors must thus be appropriately chosen, designed, and optimized to satisfy the performance requirements for each specific application. This, in turn, requires a deep knowledge of the physical mechanisms responsible for the resistance switching phenomena, their interplay, and how they are affected by the device materials and geometry. However, the partial comprehension of the physical mechanisms underlying memristor operation prevents the application-oriented design and tuning of the devices.

Table 1. Summary of desired performance metrics for memristors.

Metric	NVM [1] [26]	DNN [2] [39]	SNN [3] [40]
Feature size	<12 nm	<10 nm	-
Number of levels	≥2 (1 bit)	>100 (6.45 bits)	≥64 (6 bits)
Dynamic range (on/off ratio)	-	≥100	≥00
State retention	>1 year	>10 years	>10 years
Device endurance	$>10^3$ cycles	$>10^9$ cycles	$>10^9$ cycles
Energy consumption	<100 pJ/write	<10 fJ/programming pulse	<10 fJ/spike
Linearity	-	Yes	Yes
Symmetry	-	Yes	-
Switching time	<100 µS	<100 ns	-

[1] System level performance for replacing a NAND flash memory. [2] Performance for memristors in a deep neural network (DNN) accelerator with a crossbar memory array architecture. [3] Performance for memristors as a Spiking Neural Network (SNN) artificial synapse.

Physical multiscale modeling and simulation provide a powerful tool for investigating the physical mechanisms responsible for analog switching in memristors and to highlight the effects of the material properties (including defects) on the device performance. This information can then be used to further develop memristive technology, optimizing the properties of the devices (i.e., geometry and materials) to match the specifications required for the desired application (i.e., electric properties, data retention, variability, and noise), and strongly reducing its time-to-market.

In this paper, we use a multiscale modeling platform to simulate RRAM devices and investigate the physical mechanisms underlying their operation. The simulations are designed to highlight the effects

of the device geometry, materials, forming conditions (i.e., temperature, current compliance, voltage stress mode), and programming. The discussion on the simulation results shows the relevance of the obtained insights for the different AI applications (i.e., non-volatile memories, deep neural networks, or spiking neural networks) and provides useful design principles for RRAM application-oriented design.

2. Materials and Methods

As highlighted in Section 1, the design of memristive devices requires complete knowledge of all the involved resistance switching phenomena, including the effects of the device geometry and materials.

In RRAMs, the fundamental mechanisms include (i) the interaction between the electronic and ionic transport, (ii) the effects induced by the applied electric field, and (iii) the influence of the microscopic material properties [41].

Due to the complexity of those mechanisms, exacerbated by their interplay, the physical modeling of RRAMs presented in this work is extremely advantageous, as it easily allows the following: (i) prediction of the device performance (i.e., switching time, endurance, and retention) from the material properties (which is one of the key novel aspects of this work, which allows the materials and the process conditions to be directly screened when targeting specific applications); (ii) investigation of the trade-off between the device scaling and variability; (iii) evaluation of the process effects on the device and its materials through the interpretation of electrical characterization data; and (iv) co-design of the device materials and geometries for satisfying the specific application requirements (see Table 1).

In this section, we first review the physical mechanism underlying the RRAM operation, and then present a multiscale modeling platform that includes all the presented effects and mechanisms.

2.1. RRAM Devices

In RRAMs, the analog resistance switching is controlled by a reversible, voltage-driven, and ion-based mechanism that allows the geometry of a conductive path within a dielectric layer to be modulated. Different switching mechanisms have been proposed for the implementation of RRAM [26,42,43], with each one requiring the comprehensive knowledge of different physical processes for being thoroughly understood.

In CBRAMs [44–46], a conductive filament (CF) is created (and dissolved) in a solid electrolyte by means of redox reactions. By applying a positive voltage to the “active” electrode (the anode), it releases metallic cations into the electrolyte that migrate towards the “inert” electrode (the cathode) pushed by the electric field. Once they reach the cathode, the cations are reduced, contributing to the formation of the CF and eventually connecting the two electrodes. Through reversing the redox process by applying a negative voltage to the anode, the CF is progressively dissolved.

Optimizing CBRAMs requires an understanding of the kinetics and interplay between the redox processes and the transport of metallic cations in solid electrolytes (i.e., drift and diffusion). Moreover, all the effects related to the material interactions and geometry must be considered.

In OxRAMs [47–51], a conductive path constructed of oxygen vacancies is formed in a high-k oxide layer (e.g., TiO_2, HfO_2, TaO_5), breaking the oxide atomic bonds (Figure 1a).

When applying a voltage to the oxide layer (Figure 1b), its lattice bonds are stretched. With a sufficiently high voltage, the bonds eventually break, generating a negatively charged oxygen ion and a positively charged oxygen vacancy (Frenkel pair). If no recombination of the pair occurs, the oxygen ion and vacancies migrate in opposite directions under the action of the electric field. However, due to their relatively high diffusion energy barrier [52], the oxygen vacancies experience little to no motion, locally increasing the electrical conduction and power dissipation in the material. The resulting temperature increment supports the creation of new oxygen ions/vacancies, leading to the rapid formation of a highly conductive path of oxygen vacancies between the two electrodes (Figure 1c). The conductive path can be broken by recombining the oxygen vacancies with the oxygen ions, i.e.,

bringing the oxygen ions near the oxygen vacancies by applying a negative voltage to the oxide layer (Figure 1d).

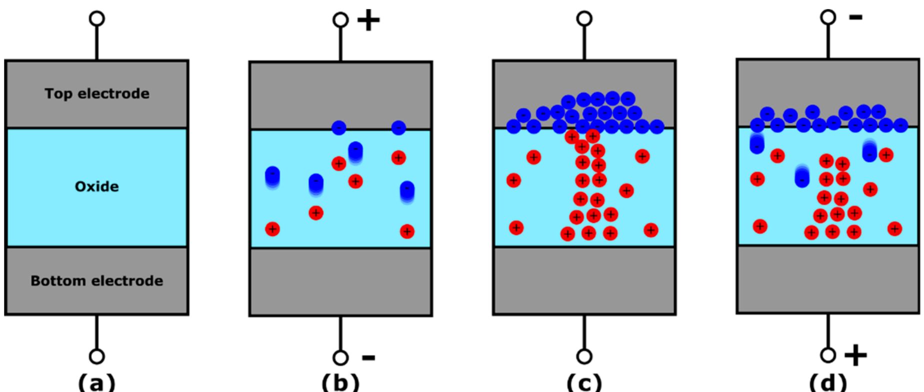

Figure 1. Structure and operation of a filamentary OxRAM device. (**a**) Device in pristine conditions. (**b**) Forming operation: a positive voltage is applied to the top electrode, generating oxygen ion/vacancy pairs. The ions (blue spheres) migrate under the effect of the electric field, leaving the oxygen vacancies (red spheres) behind. (**c**) Formed device with a conductive filament (CF) made of oxygen vacancies connecting the two electrodes. (**d**) Reset operation: a negative voltage is applied to the top electrode, bringing the oxygen ions near to the oxygen vacancies. The resulting recombination leads to partial dissolution of the CF and the formation of a dielectric barrier.

The analog resistance switching in OxRAMs relies on the modulation of a conductive path created during a forming process, consisting of a current-controlled breakdown of the dielectric. After the formation, only a thin portion of the formed conductive path is affected by the oxygen ion/vacancies generation and recombination processes. The conductance switching thus requires precise control of the geometrical properties of a thin insulating barrier.

Depending on the forming conditions (i.e., temperature, current compliance, voltage stress mode), the conductive path can assume either a uniform or filamentary shape. In the former [51], the dielectric conduction is uniformly modulated across its section and the resistance is thus controlled through the insulating barrier thickness only. In the latter [47–50], a CF is formed within the dielectric, enabling control of the resistance by both the barrier thickness and the CF diameter.

The design of OxRAMs requires an understanding of the complex processes of oxygen ions/vacancies generation, diffusion, and recombination under the action of an external electric field; their interplay with the charge transport in dielectrics; and the effects of the material properties and geometries on all those processes.

In this work, we focus on the multiscale simulation of OxRAM devices, and investigate the interplay between the physical mechanisms involved and their impact on the device performance, reliability, and variability. Specifically, we mainly consider OxRAM devices with a one-oxide-layer stack made of $TiN/HfO_x/TiO_y/TiN$ (where TiO_y is a parasitic layer). For comparison, we also consider a two-oxide-layer stack made of $TiN/Ta_2O_x/TiO_y/TiN$, showing the different properties enabled by such structural and material combination.

2.2. Multiscale Modeling of RRAMs

The multiscale modeling platform sketched in Figure 2 allows the RRAM devices to be thoroughly investigated. Starting from the key material properties, calculated using ab-initio methods [53–55], and the other device-specific properties (e.g., geometry and materials), the platform models the electrical device response considering all the complex physical mechanisms involved in the different RRAM operations, while accounting for all the resulting changes in the 3D electric and temperature fields, and in the material structure.

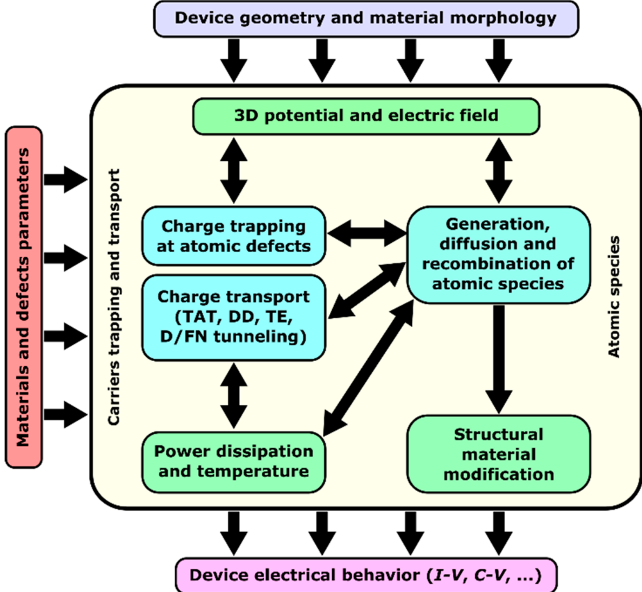

Figure 2. Multiscale modeling platform for RRAM devices. The device-level simulation engine includes three main modules: one for the simulation of charge transport; one for charge trapping; and one for the simulation of atomic species generation, diffusion, and recombination. The simulation requires the material parameters, calculated using ab-initio methods, and a definition of the device geometry. The modeling platform considers the 3D potential and electric field and how they are affected by localized trapped charge and power dissipation at defects. The results are the various characteristic curves describing the device electrical response.

The multiscale modeling platform comprises three main modules, addressing charge transport, charge trapping, and the generation/diffusion/recombination of atomic species (i.e., oxygen vacancies and interstitial ions), respectively.

For a comprehensive simulation of charge transport, many mechanisms are considered, including trap-assisted tunneling (TAT), thermo-ionic emission (TE), the Poole–Frenkel effect (PF), and Drift-Diffusion (DD) through conduction and valence bands, and through sub-bands originating from metal-rich regions formed within the oxide. Interestingly, in most RRAM devices, the TAT and DD phenomena dominate the charge transport dynamics, the first at low defect densities (i.e., the conduction in the oxide is mainly defect-assisted, typical of binary and ternary oxides [56–62]), and the latter for sufficiently high defect densities (i.e., a conductive path is present within the oxide). The TE and the PF phenomena are typically negligible in materials with a medium/high bandgap, such as transition metal oxides [63].

The TAT conduction is described using a multiphonon TAT model, inherently considering the electron-phonon coupling oxides by accounting for the atomic lattice rearrangement in the vicinity of the defect due to the presence of a trapped charge [56–58,64]. The TAT transport considered is fully described in [65]. This model requires knowledge of the relaxation energy, E_{REL}, associated with the atomic lattice relaxation process; the defect thermal ionization energy, E_T; and the defect density in the material, N_T. Both E_{REL} and E_T are calculated by means of ab initio methods (i.e., Density Functional Theory and Molecular Dynamics simulations). The DD conduction is best described by adopting the Landauer approach [66], which accounts for delocalized electron flow in the conductive path.

The defect-assisted charge transport naturally results in localized power dissipation at the defect sites, affecting the temperature distribution in the device. The power dissipation is self-consistently computed across the entire device volume by including the charge carrier's energy released at both the defects (at every charge trapping event) and the lattice (due to inelastic scattering mechanisms, i.e., optical and acoustic phonons). The temperature distribution in the device volume is calculated from the power dissipation by solving the Fourier's Law for heat conduction.

The charge transport dynamics are strongly coupled to the different phenomena related to atomic species in the oxide (i.e., generation, diffusion, recombination). Therefore, they must be consistently calculated to effectively model the structural material modifications occurring during RRAM operations (i.e., forming, setting, and resetting).

The generation of atomic species is described by a thermochemical bond breakage model [67], consisting of compact effective-energy formulas accounting for the microscopic material properties and the two main generation mechanisms: (i) the breakage of atomic bonds, enhanced by the local electric and temperature field profiles and locally favored by the possible presence of precursors [68] and other defects, and (ii) the redox reactions that occur at the interfaces, as well-favored by the local electric and temperature fields. The resulting generation rate is

$$G(x, y, z) = G_{0,G} \exp\left[-\frac{E_{A,G} - b \cdot F(x, y, z)}{k_B T(x, y, z)}\right], \quad (1)$$

where $b = p_0 \frac{2+k}{3}$ is the bond polarization factor related to the molecular dipole moment, p_0 is the bond breakage activation energy, k is the material dielectric constant, $G_{0,G}$ is the effective bond vibration frequency, $E_{A,G}$ is the bond breakage activation energy, $F(x, y, z)$ is the 3-D electric field, k_B is the Boltzmann's constant, and $T(x, y, z,)$ is the 3-D temperature field. The material-related parameters (i.e., the effective bond vibration frequency $G_{0,G}$, the molecular bond polarizability p_0, and the bond breakage activation energy $E_{A,G}$) are calculated using ab-initio methods [54,69].

The transport of atomic species is dominated by a DD mechanism, driven by the electric field and strongly accelerated by the local temperature, described by the equation

$$G_D(x, y, z) = G_{0,D} \exp\left[-\frac{E_{A,D} - \gamma \cdot E(x, y, z)}{k_B T(x, y, z)}\right], \quad (2)$$

where $G_{0,D}$ is the effective bond vibration frequency, $E_{A,D}$ is the diffusion activation energy, γ is the field acceleration factor, $F(x, y, z)$ is the 3-D electric field, k_B is the Boltzmann's constant, and $T(x, y, z,)$ is the 3-D temperature field.

However, both the electric and temperature fields are in turn affected by the presence of atomic species, implying a strong and complex coupling between the transport mechanism and the fields in the device. For correctly modeling the DD transport, the internal device conditions (e.g., current, trapped charge distribution, electric and temperature fields) are updated every time an individual defect is generated, recombined, or moved.

The stochastic nature of the mechanisms involved in RRAM is successfully accounted for by using a kinetic Monte Carlo approach, which allows a consideration of phenomena like the intrinsic variability of the forming, set, and reset processes, and the occurrence of Random Telegraph Noise [70–74], together with their dependence on the material properties and geometry, providing insights for their optimization.

The presented modeling platform successfully reproduces the RRAM electrical responses to arbitrary voltage and current inputs, allowing for the extraction of various characteristic curves of the device (I-V, C-V, and G-V). The parameters used in all the simulations are reported in Table 2.

Table 2. Simulation parameters.

Symbol	Quantity	Material		
		HfO_2	Ta_2O_5	TiO_2
	Material Parameters			
E_G	Band-gap (eV)	5.8	3.6	1.3
χ	Electron affinity (eV)	2.4	3.4	4.3
k	Relative dielectric permittivity	21	25	95
m_e*	Electron tunneling effective mass	0.25 m_0	0.3 m_0	0.2 m_0
WF	TiN work function (eV)	4.57	4.57	4.57
	Defect Parameters			
E_T	Defect thermal ionization energy (eV)	1.7–2.7	0.8–1.2	0.1–0.5
E_{REL}	Defect relaxation energy (eV)	1.19	0.88	0.7
N_T	Defect density (cm^{-3})	5×10^{19}	5×10^{19}	5×10^{19}
	Metal-Oxygen Bond Breakage Parameters			
p_0	Polarizability (eÅ)	5.2	1.8	4
$E_{A,G}$	Activation energy (eV)	2.1	1.0	5.3
$G_{0,G}$	Effective bond vibration frequency (Hz)	4.5×10^{13}	4.5×10^{13}	4.5×10^{13}
	Oxygen Ion Diffusion Parameters			
Γ	Field acceleration factor (eÅ)	0.3	0.2	0.4
$E_{A,D}$	Activation energy (eV)—in x/y/z direction	0.8/0.8/0.7	1.2/1.2/1.0	1.0/1.0/0.75
$G_{0,D}$	Effective bond vibration frequency (Hz)	4.5×10^{13}	4.5×10^{13}	4.5×10^{13}

3. Results

The multiscale modeling platform presented in Section 2 combines all the most relevant physical mechanisms involved in RRAM operation, directly connecting the electrical performance of RRAM devices to their geometries and to the microscopic properties of the employed materials. Noticeably, thanks to the multiscale nature of the modeling platform, it can be used to gain insights into RRAM devices at different levels.

At the physical level, the modeling platform can be used to investigate the included physical mechanisms and the effects of their interplay on the generation of the conductive path. At the device level, it allows the whole conductance switching cycle (i.e., forming, set, and reset operations) to be simulated, and can therefore be used to predict the performance of RRAM devices. Moreover, its inherently stochastic implementation allows the device variability and reliability (retention and endurance) properties to be effectively investigated. Multiscale modeling thus provides a powerful tool for accelerating the further development and optimization of RRAM technology, focusing on the specific application (i.e., NVM, ANN, or SNN).

In this section, we use the presented multiscale modeling platform to perform multiscale simulations of RRAM devices. The different simulations are specifically designed to highlight the effects of the device geometry, materials, forming conditions (i.e., temperature, current compliance, voltage stress mode), and programming. The parameters used for the simulations are summarized in Table 2.

3.1. Conductance Switching Cycle

The presented multiscale modeling platform allows the whole conductance switching cycle to be simulated. Starting from a pristine device with specified geometries and materials, it is possible to investigate the device formation and the following conductance switching operations (SET and RESET) under different conditions. The electric and temperature fields and the location of atomic species in the device can also be monitored during simulations to gain insights into the involved switching mechanisms and their interplay.

As an example, Figure 3a shows the I-V characteristic curves of a simulated device made of a TiN/5 nm HfO_x/TiO_y/TiN stack RRAM device, including the forming (solid red), the reset (dotted green), and set (dashed blue) operations. The simulations are performed by applying ramped voltages and a compliance current of 10 µA.

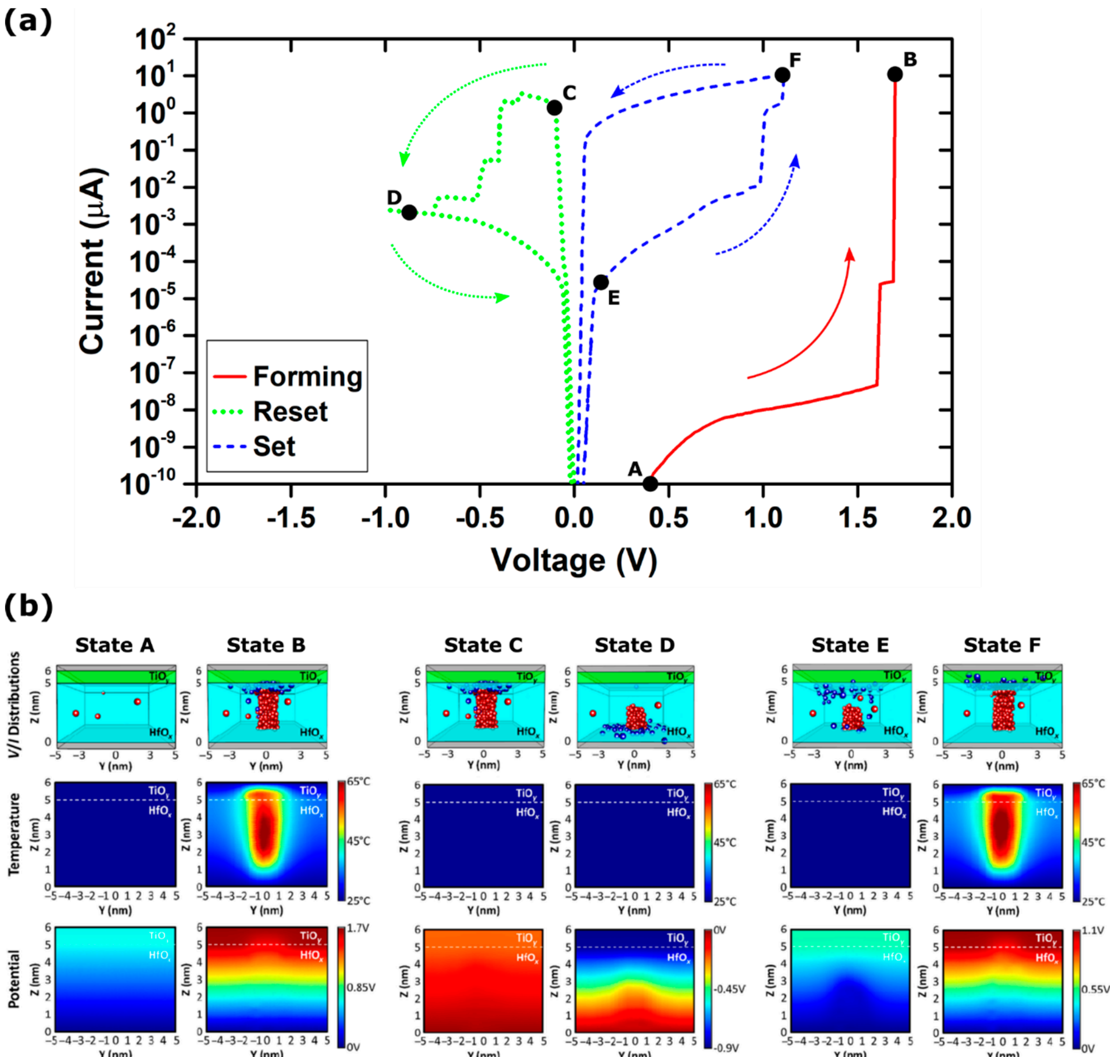

Figure 3. Simulation results of a TiN/5 nm HfO_x/TiO_y/TiN stack RRAM device: (**a**) Current–Voltage characteristic during forming (solid red), reset (dotted green), and set (dashed blue) operations with a compliance current of 10 µA; (**b**) oxygen vacancy (red spheres) and ion (blue spheres) distribution, temperature profile, and potential profile of the simulated device at different operation stages (labeled A, B, C, D, E, and F in Figure 3a).

In the pristine state (state A), the conduction is dominated by the TAT through the relatively few preexisting defects in the material (i.e., oxygen vacancies accumulated at grain boundaries), which determines the very high initial resistance of the device. At a low voltage ($V < V_{FORM}$ = 1.7 V in Figure 3), both the electric field and the power dissipation are too small to result in strong localized defect generation, leading to a very modest and uniform generation of defects across the whole volume of the device. Once the applied voltage exceeds V_{FORM}, the oxide bonds start to break under the effect of the increasing electric field, creating a significant number of atomic defects. Due to the different diffusion energy barriers, the oxygen ions generated drift towards the top electrode under the effect of the electric field, while the oxygen vacancies mostly remain in place. Noticeably, the oxygen ions accumulate at the TiO_x layer, creating the so-called “oxygen reservoir”. The newly created oxygen

vacancies support the TAT, locally increasing the current flow, power dissipation, and temperature, in turn assisting with the generation of new defects. A thermally-driven positive feedback process is thus established, leading to the rapid formation of a CF. After the formation (state B), the device is in a low resistance state. The conduction is dominated by DD in the vacancy-rich regions constituting the CF, while the oxygen ion counterparts are gathered at the top electrode.

The reset operation is simulated by the device configuration resulting from the described forming operation. This approach is advantageous since, in comparison to other approaches proposed in the literature [66,75], no a-priori assumptions of the CF structure and characteristics are required. Upon the application of a negative voltage ramp (state C), the oxygen ions gather in the oxygen reservoir during the forming drift towards the bottom electrode under the effect of the electric field. During their motion, the oxygen ions can recombine with the oxygen vacancy defects in the oxide, leading to the progressive formation of a thin dielectric barrier within the CF, causing resistive switching to the high resistance state (state D). Interestingly, this process is not associated with a large temperature increment in the device, suggesting that it is mostly driven by the electric field.

Lastly, the set operation is simulated by the device configuration obtained at the end of the reset operation. Upon the application of a positive voltage ramp (state E), the device experiences a process similar to that of forming, but confined to the thin dielectric barrier created during the reset process. Since the applied voltage drops almost completely across the thin dielectric barrier, the electric field required for the breaking of oxide bonds and the generation of new defects is easily exceeded at relatively low voltages ($V < V_{FORM}$). The restoration of the CF is thus initiated at lower voltages compared to the forming process. The same thermally-driven positive feedback described for the forming process is established, resulting in a quick restoration of the CF and the switch to a low resistance state (state F).

3.2. Effects of Forming Conditions

The presented multiscale modeling platform allows the effects of the forming condition (i.e., temperature, current compliance, voltage stress mode) on the performance and properties exhibited by the device after the forming process to be investigated.

The beneficial effects of a high-temperature forming process have been thoroughly reported in the literature, and have been associated with lower forming voltages and variability of the low resistance state, while improving the stability and reliability of the device [76–78]. The external temperature affects the forming process assisting the defect generation and promoting the oxygen ion diffusion in the device, leading to a lower density of oxygen ions near the conductive path after the forming process. The subsequent oxygen ion/vacancy recombination is strongly reduced, resulting in a higher stability and lower variability of the conductive path.

This is evidenced in Figure 4a, which shows the low state resistance distribution exhibited by TiN/5 nm HfO_x/TiO_y/TiN RRAM stacks after the forming process at two different external temperatures (i.e., 25 and 125 °C), using a ramped voltage and a 1 μA compliance current. In accordance with the literature [76], the experimental data (marked by the symbols in Figure 4a) show that the higher forming temperature leads to a tighter resistance distribution.

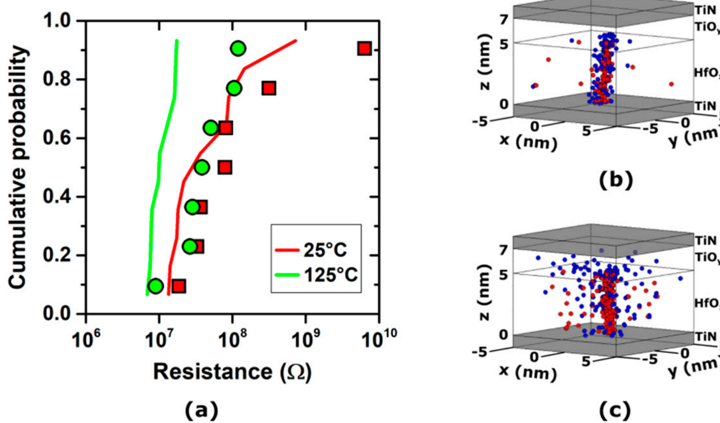

Figure 4. (**a**) Experimental (symbols) and simulated (lines) cumulative distributions of TiN/5 nm HfO_x/TiO_y/TiN RRAM stacks' low state resistances after forming at 25 and 125 °C. (**b**) Resulting oxygen ions/vacancies distribution in the device after forming at 25 °C. (**c**) Resulting oxygen ions/vacancies distribution in the device after forming at 125 °C. The forming process was performed using a ramped voltage and a 1 μA compliance current.

The effects of the forming temperature on the oxygen ions/vacancies generation and diffusion can be effectively investigated using the presented multiscale modeling platform. The simulations of the forming processes (shown by the lines in Figure 4), whilst not perfectly matching the measured samples, accurately reproduce the trend and order of magnitude of the experimental data, showing a tighter resistance distribution in the higher temperature forming case. The discrepancy between the experimental results and those of the simulation can be ascribed to several effects not considered in these simulations, mainly, the oxide thickness and area variations due to the fabrication process tolerances, and process-dependent interface effects between the oxide and the electrodes.

The microscopic differences caused by the higher forming temperature can be better appreciated in Figure 4b,c, which shows the simulated oxygen ions/vacancies distribution in the device after the forming process. At a low temperature (25 °C), the conductive path (in this case, a CF) made of oxygen vacancies is tightly surrounded by oxygen ions affecting its resistance and stability. Conversely, at a high temperature (125 °C), the oxygen ions are scattered in the device volume, affecting the CF to a lesser extent and leading to a tighter resistance distribution.

The platform can therefore be used to explore the best strategies to control the device-to-device variability of RRAMs for specific applications: for instance, high-temperature forming could reduce the variability for high-accuracy DNNs, while precisely controlling the temperature during the forming process could be used to obtain the desired variability level, optimizing the performance of stochastic learning networks.

The forming compliance current has a very strong impact on the morphology of the conductive path as it allows the defect generation processes in the device to be controlled by limiting the maximum current flow and power dissipation in the device. For example, in filamentary RRAMs, it has been observed that the magnitude of the compliance current allows the diameter of the CF to be controlled [57]. Moreover, it also greatly affects the low resistance state magnitude and variability: a sufficiently high compliance current allows the formation of a dense population of defects, forming low-resistance CFs with a similar morphology, while a low compliance current leads to a sparse population of defects, forming weak CF characterized by a larger variability. This is evidenced in Figure 5, which shows the low state resistance distribution exhibited by TiN/5 nm HfO_x/TiO_y/TiN RRAM stacks after the forming process at three different compliance currents (i.e., 1, 5, and 10 μA), using a ramped voltage and an external temperature of 25 °C.

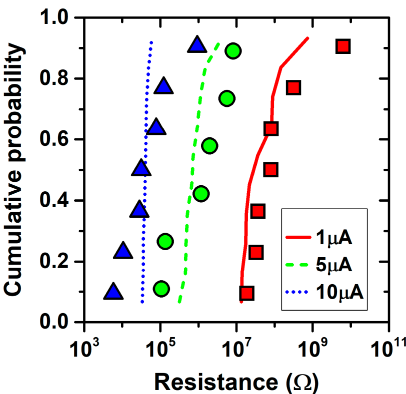

Figure 5. Experimental (symbols) and simulated (lines) cumulative distributions of TiN/5 nm HfO_x/TiO_y/TiN RRAM stacks' low state resistances after forming at 1, 5, and 10 µA. The forming process was performed using a ramped voltage and 25 °C external temperature.

The effects of the compliance current on the defect generation processes and the resulting device resistance magnitude and variability can be effectively investigated using the presented multiscale modeling platform. The simulations of the forming processes, shown in Figure 5, reproduce the trend found in the experiments, highlighting that the lower variability of the low resistance state is in fact related to the higher density of oxygen vacancies generated at higher compliance currents. The platform confirms its effectiveness for investigating the best strategies to control the device-to-device variability for specific applications, as previously mentioned.

In filamentary RRAM, the morphology of the CF is also strongly affected by the forming voltage stress mode (i.e., the time-varying profile of the forming voltage), as it determines different distributions of oxygen ions and vacancies in the device volume. Evaluating the effect of a specific voltage stress mode is highly desirable, since it allows the best strategy for a specific application to be identified. However, this task is extremely complex, as it requires the interplay between the (possibly) time-varying forming voltage, the defect generation processes driving the growth of the CF, and the oxygen ions diffusion to be considered. The presented multiscale modeling platform provides a powerful tool for such investigation.

Figure 6 shows the simulation of the oxygen ions/vacancies distribution in TiN/5 nm HfO_x/TiO_y/TiN RRAM stacks after the forming process using three different voltage stress modes (i.e., constant voltage, ramped voltage, and pulsed voltage), a compliance current of 1 µA, and an external temperature of 25 °C.

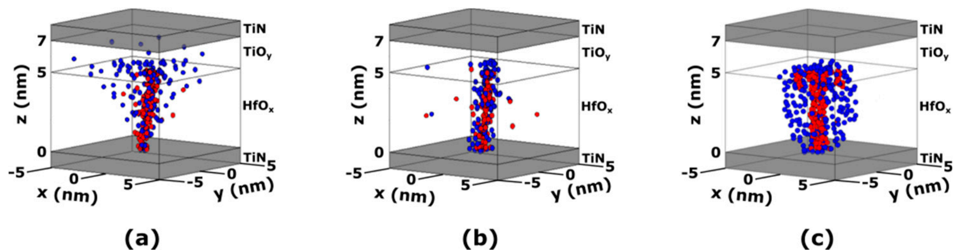

Figure 6. Simulated distributions of the oxygen ions (blue) and vacancies (red) in TiN/5 nm HfO_x/TiO_y/TiN RRAM stacks after forming with a (**a**) constant voltage, (**b**) ramped voltage, and (**c**) pulsed voltage. The forming process was performed using a 1 µA compliance current and 25 °C external temperature.

Noticeably, the oxygen ions distribution is significantly affected by the voltage stress mode. At constant voltage forming, the oxygen ions are mostly accumulated near the top electrode, while ramped

and pulsed voltage forming results in a more uniform distribution. The pulsed voltage forming, however, leads to a higher radial diffusion of the oxygen ions compared to the constant voltage forming.

Comparing the oxygen ions and vacancies distribution shown in Figures 4 and 6, it can be noticed that the distribution corresponding to a low-temperature forming process (Figure 4b) is similar to the one of a ramped voltage forming process (Figure 6b). Conversely, the distribution corresponding to a high-temperature forming process (Figure 4c) is closer to that of a pulsed voltage forming process (Figure 6c). The similarity of the oxygen ions and vacancies suggests a similar behavior of the device. This is confirmed by experimental results [79], showing that pulsed voltage forming is associated with a tighter resistance distribution with respect to constant or ramped voltage forming.

The voltage stress mode can thus be exploited for optimization of the oxygen ions distribution in the formed device for achieving a tighter resistance distribution, similar to the previously discussed forming temperature. The platform can support such optimization, allowing for the fine tuning of RRAM device-to-device variability according to the specific application requirements.

3.3. Analog Resistance Switching Optimization

As previously discussed, the analog resistance switching in RRAM devices is driven by two different mechanisms. The switching from a high resistance state to a low resistance state is dominated by thermal positive feedback triggered by an electric field, leading to a mostly uncontrolled generation of oxygen vacancies and resulting in the abrupt formation of a conductive path. Conversely, the switching from a low resistance state to a high resistance state is determined by the recombination of oxygen ions/vacancies driven by the electric field, resulting in the new formation of a high-resistance dielectric barrier. The dielectric barrier formation, being unassisted by the temperature, is typically slower than the formation of the conductive path.

The differences between the two switching mechanisms cause strong asymmetry in the device characteristics, as evidenced in Figure 3, which has been recognized as detrimental for many applications (i.e., ANNs and SNNs). Moreover, the extremely fast formation of the conductive path prevents the fine control of analog resistance switching, potentially limiting both the number of distinguishable resistance levels (thus the number of bits per cell in NVM) and the device linearity.

A well-established method for mitigating the non-idealities exhibited by RRAM devices is pulsed programming, i.e., controlling the resistance switching by applying a sequence of voltage pulses to the device. Ideally, each set (reset) voltage pulse should decrease (increase) the device resistance by a small amount, according to the device's electrical characteristics, enabling fine tuning of the resistance [80].

In the simplest approach, the applied pulses are all identical in amplitude and width (with opposite signs in set and reset operations), allowing for better control of the conductive path formation. However, this is often not sufficient for the full compensation of RRAM non-idealities, requiring sequences of non-identical pulses and possibly a combination of positive and negative pulses for each set or reset operation [80]. Fine tuning of the pulse sequence shape, amplitude, and width is thus of paramount importance for exploiting the full potential of RRAM devices.

A combined design of the device geometries and materials can be beneficial for controlling the RRAM non-idealities and achieving linear resistance analog switching without using complex voltage pulse schemes [81]. A simple solution consists of using a two-layer RRAM stack made of two different dielectric materials, comprising a thin (~1–2 nm) low-k (LK) material and a thicker high-k (HK) material. The resulting distribution of the electric field across the device allows the switching mechanism to be controlled and gradual modulation of the device resistance to be produced. In such a configuration, the structural material changes responsible for the resistance switching (i.e., the conductive path and dielectric barrier formation) are confined to the LK layer, where the electric field is the highest, regardless of the layers' thickness. The confinement of these phenomena in the LK layer is the key factor that leads to a gradual change of the electrical resistance in the device, which is crucial to achieving precise and linear analog switching. Moreover, this two-layer structure allows for better control of the oxygen ions diffusion. The low electric field in the HK layer hinders the diffusion of

oxygen ions coming from the LK layer, leading to the formation of an ion reservoir at the LK–HK interface. Having the oxygen ions reservoir near the conductive path contributes to smoothening the whole switching process and results in gradual modulation of the electrical resistance, with obvious benefits for applications requiring a linear and gradual conductance change, e.g., DNNs.

The presented multiscale modeling platform allows the devices' non-idealities and their effects on the device performance (i.e., symmetry, linearity, number of levels) to be evaluated. The retrieved information can then be exploited, supporting optimization of the device structure and materials, and the programming pulse sequence for achieving linear analog resistance switching.

Starting from a one-layer device, Figure 7 shows the experimental and simulated conductance evolution of TiN/6 nm HfO_x/TiO_y/TiN RRAM stacks under the application of set pulse trains with an identical width (1 ms) and different voltage amplitudes (0.7, 0.8, 0.9, and 1 V). After each single set pulse, the conductance was read with a read pulse with 0.1 V amplitude and 1 ms width, which did not significantly influence the device resistance.

Noticeably, only pulse trains with amplitudes of 0.8 V and above induce a variation of the device conductance, mostly during the first few pulses (<5) of the sequence. The conductance quickly saturates, highlighting both the nonlinear characteristic of the device and the abrupt formation of the conductive path. These trends are accurately reproduced by the simulations. The fluctuation in conductance exhibited by the experimental results can be ascribed to the random nature of oxygen migration caused by the voltage pulses [82]. In fact, even after the conductance saturation, each pulse causes a small random motion of the oxygen ions in the RRAM stack, resulting in fluctuations of the conductance around a mean value.

Taking advantage of the RRAM stack's nonlinear and saturating behavior, it is possible to associate a set conductance value with a specific pulse voltage. Further simulations could support the optimization of pulse amplitudes for the best discrimination of conductance levels.

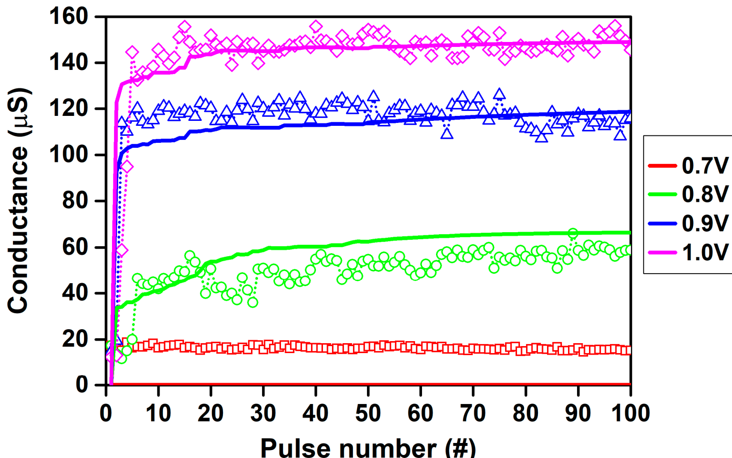

Figure 7. Experimental (symbols) and simulated (lines) conductance evolution of TiN/6 nm HfO_x/TiO_y/TiN RRAM stacks under the application of set pulse trains with a 1 ms width and variable amplitude: (red) 0.7 V; (green) 0.8 V; (blue) 0.9 V; (magenta) 1.0 V. Initial forming was performed with a 100 µA compliance current.

Similar results are obtained by modulating the pulse width, as shown in Figure 8a, which depicts the experimental and simulated conductance evolution of the same device under the application of set pulse trains with an identical voltage amplitude (0.9 V) and different widths (10 µS, 100 µS, and 1 ms). As shown in Figure 8b, it is possible to take advantage of the nonlinear and saturating behavior exhibited by the considered RRAM stack to associate a set conductance value with a specific pulse width, which achieves six well-separated and recognizable resistance levels with an approximately linear characteristic. Finely tuning the programming pulse train for both set and reset processes allows

a linear resistance update to be achieved for any specific RRAM device. The best combination of pulse amplitude, width, and sequence can be effectively investigated by the presented multiscale modeling platform, in order to optimize the synaptic behavior of the device. On the other hand, the platform could be useful for implementing design-circuit co-design strategies to enhance the device performance for multi-level memory applications, increasing the robustness and possibly reducing the bit error rate.

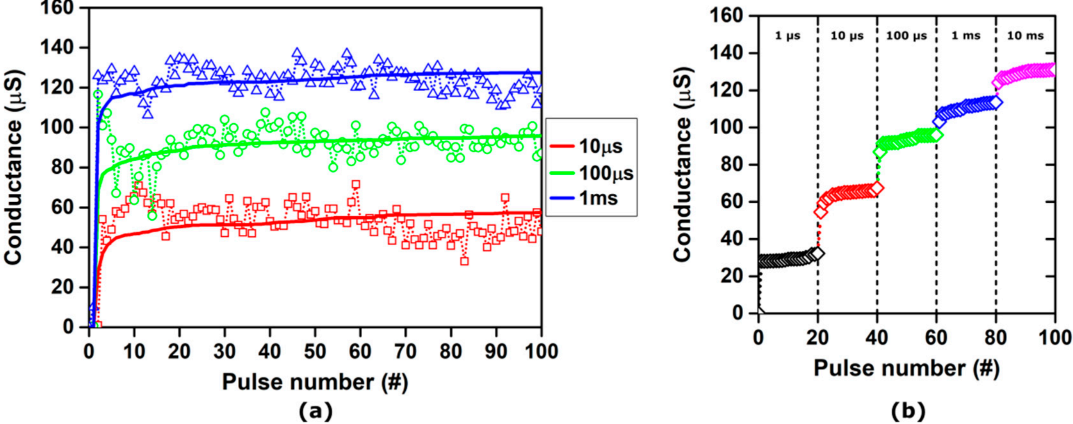

Figure 8. (**a**) Experimental (symbols) and simulated (lines) conductance evolution of TiN/6 nm HfO_x/TiO_y/TiN RRAM stacks under the application of set pulse trains with a 0.9 V amplitude and different widths: (red) 10 µS; (green) 100 µS; (blue) 1 ms. Initial forming was performed with a 100 µA compliance current. (**b**) Simulated conductance modulation obtained using set pulse trains with an increasing width: (black) 1 µS; (red) 10 µS; (green) 100 µS; (blue) 1 ms; (magenta) 10 ms.

Using a two-layer RRAM stack made of a thin LK layer (i.e., Ta_2O_5) and a thick HK layer (i.e., TiO_2), a linear conductance characteristic can be obtained. Figure 9 shows the experimental and simulated conductance evolution of TiN/2 nm Ta_2O_x/35-nm TiO_y/TiN RRAM stacks during both set and reset operations, under the application of pulse trains with an identical width (100 µS) and different voltage amplitudes. After each single pulse, the conductance was read with a read pulse with a 0.1 V amplitude and 1 ms width, which did not significantly influence the device resistance.

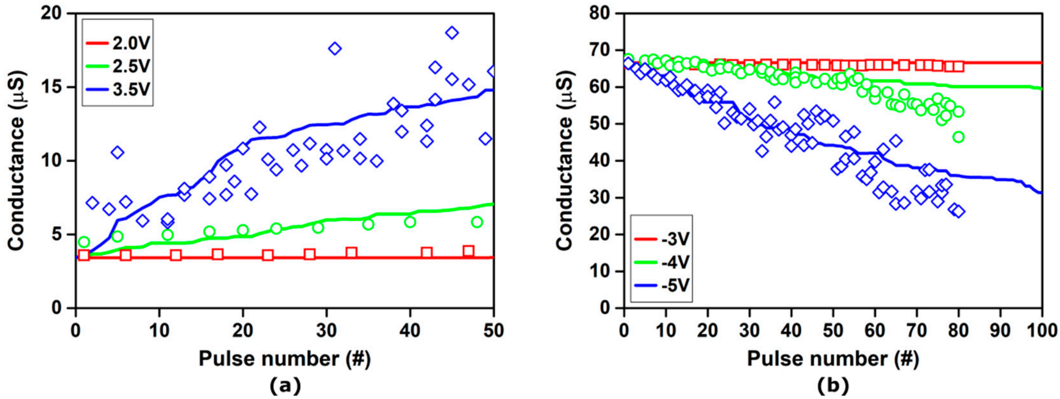

Figure 9. Experimental (symbols) and simulated (lines) conductance evolution of TiN/2 nm Ta_2O_x/35-nm TiO_y/TiN RRAM stacks under the application of set pulse trains with a 100 µS width and variable amplitude. (**a**) Set operation: (red) 2 V; (green) 2.5 V; (blue) −3.5 V. (**b**) Reset operation: (red) −3 V; (green) −4 V; (blue) −5 V.

Noticeably, the considered device exhibits an extremely linear behavior during both the set and reset operations, especially if compared with the characteristics of the one-layer device (Figures 7 and 8). However, the pulse voltage required for the switching to occur is considerably larger, starting at 2.5 V

for the set process and requiring up to −4 V for the reset process. All these trends are accurately reproduced by the simulations.

While the two-layer stack structure allows for the fabrication of RRAM devices with an intrinsically linear characteristic, the lack of symmetry still prevents their use for ANN and SNN applications and could be detrimental in NVMs. The symmetry of the electrical characteristic can be pursued by further investigating the novel two-layer stack, using the presented modeling platform to optimize the materials and geometries combination, or by using specifically designed pulse sequences. Interestingly, the latter solution is relatively easy to implement: thanks to the intrinsic linearity of the set and reset characteristics, symmetry can be obtained by using different pulses (i.e., with different amplitudes) in the two operations.

3.4. Switching Reliability Optimization

The fundamental mechanism enabling analog resistance switching in RRAM devices is the diffusion of oxygen ions in the oxide layer. This process is greatly affected by the interaction of ions with the surrounding lattice, thus by the material's properties, and in turns affects many of the device's electrical properties. For example, as previously discussed, precise control of oxygen ion diffusion is key to achieving well-separated resistance levels. Even more importantly, the properties of the ion diffusion process are of paramount importance for evaluating the variability and endurance of RRAM devices [83]. In fact, precise switching between different resistance levels requires the dielectric barrier to be consistently modulated for the full lifespan of the device, with little to no variations between the cycles. Conversely, the newly proposed stochastic learning algorithms for SNNs can take advantage of device variability and non-uniformity. In both cases, the oxygen ion diffusion process must be thoroughly investigated for specific optimization of the devices. The presented multiscale modeling platform, fully accounting for the oxygen ion kinetics, can be used to investigate its effects on the reliability and variability of RRAMs.

The simulations of a TiN/5 nm HfO_x/TiO_y/TiN stack RRAM device, shown in Figure 10, reveal the high sensitivity of the device reset operation to the oxygen ion diffusion kinetics properties. Starting from a correctly formed device (Figure 10a), the reset operation was simulated considering slow and anisotropic (motion predominantly in the vertical direction) oxygen diffusion (Figure 10b), corresponding to well-performed reset operation. Then, starting from the same post-formed conditions, the simulation was performed for fast and anisotropic oxygen diffusion (Figure 10c), corresponding to an excessively high voltage reset process, and for slow and isotropic oxygen diffusion (Figure 10d). Interestingly, this last condition describes the diffusive motion of oxygen ions due to the sole temperature field.

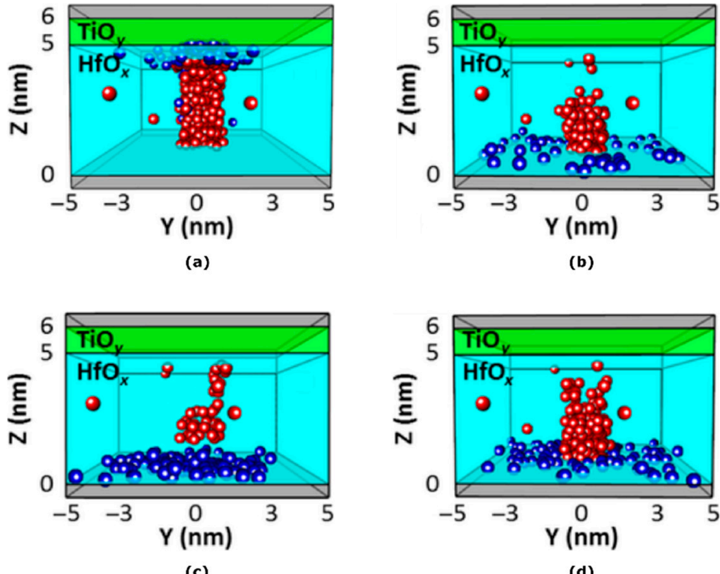

Figure 10. Simulated distributions of oxygen ions (blue) and vacancies (red) in TiN/5 nm HfO_x/TiO_y/TiN RRAM stacks: (**a**) after forming; (**b**) after the reset operation with slow and anisotropic oxygen diffusion, showing an efficient reset process; (**c**) after the reset operation with excessively fast and anisotropic oxygen diffusion, showing an inefficient reset process with the formation of a dielectric barrier near the bottom electrode; (**d**) after the reset operation with slow and isotropic oxygen diffusion (with significant motion in the radial direction), showing an inefficient reset process due to the excessive motion of the oxygen ions away from the CF.

An efficient reset operation was only obtained in the first case, where the oxygen ions are carried towards the CF in a mostly anisotropic way (some radial motion is beneficial) and are provided with enough time for recombination with the CF oxygen vacancies. In the case of fast diffusion of the oxygen ions (i.e., for an excessively high-voltage reset process), the latter condition is hindered: a significant portion of the oxygen ions are swept through the oxide, recombining along the CF length and eventually forming a dielectric barrier near the bottom electrode, as suggested in [52]. Finally, the slow and isotropic diffusion of oxygen ions leads to an inefficient reset of the device, as the excessive scattering of ions in the device volume prevents the formation of a fully dielectric barrier. The simulation results are in accordance with the experimental results showing thermally-driven oxygen ions diffusion from the "reservoir" to the CF [84].

The set operation is also affected by the diffusion of newly generated oxygen ions towards the TiO_x "reservoir" at the top electrode. Significant radial motion of the oxygen ions during the set operation would spread the ions in the "reservoir", not creating ion storage in sole proximity to the CF, leading to an increasingly ineffective reset operation and possibly to failure of the device. The cycling endurance of a device is clearly negatively affected by excessive radial motion of the oxygen ions induced by the diffusion. Therefore, convenient process recipes to optimize material properties can be implemented based on the results of these simulations to optimize the endurance and switching uniformity of RRAM devices, depending on the specific application target and its requirements.

4. Discussion

Memristors have an extremely high potential for revolutionizing both the memory and AI fields. In fact, memristors have been proposed for implementing high-density, high-speed, and low-power NVMs, and crossbar memory arrays supporting fast and efficient in-memory vector-matrix multiplication for DNN acceleration, and are finding applications in the development of SNNs, enabling the realization of artificial synapses exhibiting an STDP learning mechanism and biologically plausible artificial neurons (i.e., in accordance with the Hodgkin–Huxley model).

As illustrated in Table 1, each of the memristor applications poses different constraints for the performance metrics, demanding devices with rather different characteristics. However, the design and optimization of memristors is still difficult and, most of the time, impractical, due to the limited knowledge on the resistance switching phenomena, their interplay, and the effects of the device materials and geometry on such phenomena.

In this context, the described physical multiscale modeling and simulation provide an extremely powerful tool for the application-oriented optimization of RRAM-based memristors.

As shown in Section 3, the presented multiscale modeling platform allows the device properties, such as its macroscopic electrical properties, its resistance switching dynamics, and its microscopic properties (i.e., the oxygen ions and vacancies distribution), to be simultaneously extracted, and how such properties are affected by the device materials, geometry, and forming conditions to be investigated. Despite a few differences between the simulations and the experimental results, mainly evidenced in Figures 4 and 6, the platform proved capable of capturing the trends and fundamental relationships between the OxRAM devices' microscopic properties and their behavior.

The discrepancies shown by the simulations can be ascribed to several effects not considered by the platform, such as the oxide thickness and area variations due to the fabrication process tolerances, and process-dependent interface effects between the oxide and the electrodes. The platform can therefore be further improved by including the said effects. Moreover, it could be extended for simulating other memory devices, such as CBRAMs (i.e., including the chemical reaction phenomena at the device interfaces) and Phase Change Memories (PCMs) (i.e., accounting for the phase change in sub-regions of the device).

Nevertheless, the information obtained through the presented multiscale simulation can be effectively used to determine the best combination of OxRAM materials, geometry, forming conditions, and pulse schemes for the desired application, dramatically reducing the time required for its marked deployment.

The application of memristors to NVMs is less demanding performance-wise [26]. Noticeably, for binary NVM applications, the conductance update linearity and symmetry are not required, greatly relaxing both the device and pulse scheme design.

Even in this simple case, device optimization must account for its microscopic properties in order to satisfy the state retention and endurance constraints. In fact, as highlighted by the simulations presented in Section 3.4, the motion of oxygen ions in the device is of paramount importance for achieving an efficient and enduring resistance switching mechanism.

The multiscale simulations can effectively support the optimization of an RRAM-based memristor for NVMs. The device materials, geometry, forming condition, and pulse scheme can be optimized according to the energy consumption and switching time constraints, and the oxygen ions and vacancies distribution can be extracted at the same time. This information can then be used to discard the configurations exhibiting an instable conductive path, thus not satisfying the state retention and endurance constraints.

From the perspective of extending the number of levels of the memristive NVMs, the support of multiscale simulations is even more advantageous, as the higher number of levels requires a tighter resistance distribution of each level. As shown in Sections 3.1 and 3.2, the resistance variability of the devices can be precisely controlled by finely tuning the forming conditions (i.e., temperature, compliance current, and voltage stress), and the presented simulations allow the best forming condition to be explored.

It should be noted that a device with non-linear and/or non-symmetric resistance switching can still be used in multilevel applications if compensated for by designing a suitable pulse scheme [85–87], as shown in Section 3.3. With the support of the described multiscale simulations, the required pulse scheme can be co-designed with the device properties and optimized to achieve the required number of levels. However, linear and symmetric resistance switching would be highly desirable, as it allows for easier and more effective level control, without requiring complex pulse schemes.

As shown in Section 3.3, using a two-layer geometry RRAM allows linear resistance switching to be achieved. In this case, the interplay between the resistance switching phenomena of the two different dielectrics determines an extremely complex behavior, which can be effectively investigated through a multiscale simulation of the device. With the support of the extracted information, the two-layer structure (i.e., its materials and geometry) can be easily optimized to achieve the desired linear, and possibly symmetric, resistance switching.

The implementation of a DNN accelerator with a crossbar memory array architecture requires memristors with very different characteristics [39].

First, the number of levels is dictated by the weight precision required by the implemented DNN: a higher number of levels corresponds to a better learning capability, higher inference accuracy, and higher robustness. A precision as low as 6 bits (i.e., 64 levels) has been shown to be effective for both training and inference [88]. Moreover, the resistance switching must be absolutely linear and symmetric to avoid significant losses in the DNN accuracy [89]. This is especially true for online-trained DNNs, while for offline-trained DNNs, the non-ideal conductance update can be compensated for by suitable writing schemes.

For this application, a two-layer structure is clearly beneficial and, as stated before, the support of the described multiscale simulations allows a better understanding of the device's complex behavior and optimization of its structure to match the requirements.

Due to the high number of levels required, the dynamic range must increase in parallel with the weight precision to ensure a sufficient separation of the levels, while ensuring a sufficiently high resistance value of the low-resistance state in order to limit the inference energy consumption. The energy consumption of the training phase is instead determined by the programming energy, which must be limited. A related performance metric is the switching time, which must also be as low as possible to ensure both fast and energy-efficient training.

All of these metrics are interconnected and their relationship with the device geometry and materials, and their trade-offs cannot be easily appreciated and designed. The low-resistance state value and the variability of the resistance levels can be controlled by the forming conditions, as previously stated and as shown in Sections 3.1 and 3.2, but the effects on the required programming energy and switching time are not obvious. The multiscale simulation provides an extremely powerful tool to explore the possible solutions and possibly optimize the device materials, geometry, and forming conditions to match the requirements.

Finally, the reliability-related metrics (i.e., endurance and retention) are extremely demanding and critical: the training phase (or the weights set up in the case of offline training), requiring a large number of switching operations, can be stressful for the devices and the weight values must be retained (ideally) indefinitely. As for the NVM application, the stability of the conductive path in a specific device can be easily investigated though multiscale simulations, allowing for the recognition of unsuitable solutions.

The desired performance for memristors implementing SNN artificial synapses is extremely similar to that required for DNN accelerators [40], especially those concerning the number of levels and the reliability-related metrics (i.e., endurance and retention). This is reasonable, since SNNs also undergo a stressful training phase (or synaptic weight set up), requiring many switching operations. Moreover, the specific requirements for feature size and switching time can be reasonably assumed to be similar to those of DNN accelerators. Additionally, recent works suggest that artificial synapses with symmetric conductance updates allow for a better accuracy in SNNs [90].

Noticeably, the application-level effects of memristors' stochasticity, i.e., uniformity (or lack of thereof) and variability, are little discussed in the literature. In NVM applications, both the non-uniformity and high variability of the memory cell can be detrimental [39], posing a challenge for technology development. Neural network applications exhibit a good tolerance to device-to-device and cycle-to-cycle variations, especially if online training is used [88,89]. Interestingly, the exploitation of memristors' stochasticity has recently been proposed for implementing stochastic learning algorithms

for SNNs [22,91]. In such a context, the device's variability, non-uniformity, and noise become key components of the learning algorithm, thus requiring a precise design. All those properties can also be investigated using multiscale simulations, e.g., as previously stated, the device's variability can be controlled and optimized by tuning the forming conditions.

5. Conclusions

We have shown that multiscale modeling and simulation can effectively support the application-oriented optimization of RRAM devices.

With the support of the presented multiscale modeling platform, we simulated the microscopic behavior of RRAMs and investigated the effects of the device geometry, materials, forming conditions (i.e., temperature, current compliance, voltage stress mode), and programming on the device performance. The multiscale simulations allowed the properties of RRAMs during their whole operation, from the forming process to the subsequent set-reset cycle, to be investigated, providing information about the device linearity, symmetry, dynamic range, and reliability.

The presented multiscale simulations provide useful design principles for RRAM technology optimization according to the specific AI application, for the implementation of non-volatile memories, deep neural networks, or spiking neural networks.

Moreover, the multiscale simulation allows the effects of different implementations of the device (i.e., different geometries or materials) to be explored, and both the forming conditions and the pulse scheme (i.e., amplitude, width, sequence) to be finely tuned for achieving the desired performance.

Author Contributions: Conceptualization, P.L.T. and F.M.P.; methodology, A.P.; software, A.P.; validation, L.L.; formal analysis, A.P. and L.L.; investigation, A.P. and F.M.P.; resources, L.L.; data curation, A.P.; writing—original draft preparation, P.L.T.; writing—review and editing, F.M.P. and L.L.; visualization, P.L.T.; supervision, L.L.; project administration, L.L.; funding acquisition, L.L.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Goodfellow, I.; Bengio, Y.; Courville, A. *The Deep Learning Book*; MIT Press: Cambridge, MA, USA, 2017; Volume 521, p. 785.
2. Geron, A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media, Inc.: Newton, MA, USA, 2019; ISBN 9781492032649.
3. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
4. Kaensar, C. A Comparative Study on Handwriting Digit Recognition Classifier Using Neural Network, Support Vector Machine and K-Nearest Neighbor. *Adv. Intell. Syst. Comput.* **2013**, *209*, 155–163.
5. Imran, A.S.; Shahrebabaki, A.S.; Olfati, N.; Svendsen, T. A Study on the Performance Evaluation of Machine Learning Models for Phoneme Classification. In Proceedings of the 2019 11th International Conference on Machine Learning and Computing—ICMLC '19, Zhuhai, China, 22–24 February 2019; pp. 52–58.
6. Li, D.; Chen, X.; Becchi, M.; Zong, Z. Evaluating the Energy Efficiency of Deep Convolutional Neural Networks on CPUs and GPUs. In Proceedings of the 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom), Atlanta, GA, USA, 8–10 October 2016; pp. 477–484.
7. Canziani, A.; Paszke, A.; Culurciello, E. An Analysis of Deep Neural Network Models for Practical Applications. *arXiv* **2016**, arXiv:1605.07678.2016.
8. Strubell, E.; Ganesh, A.; McCallum, A. Energy and Policy Considerations for Deep Learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019.

9. Guzhva, A.; Dolenko, S.; Persiantsev, I. Multifold Acceleration of Neural Network Computations Using GPU. In *Computer Vision–ECCV 2012*; Springer Science and Business Media LLC: Berlin, Germany, 2009; Volume 5768, pp. 373–380.
10. Jouppi, N.P.; Borchers, A.; Boyle, R.; Cantin, P.-L.; Chao, C.; Clark, C.; Coriell, J.; Daley, M.; Dau, M.; Dean, J.; et al. In-Datacenter Performance Analysis of a Tensor Processing Unit. *ACM SIGARCH Comput. Arch. News* **2017**, *45*, 1–12. [CrossRef]
11. Chen, C.; Krishna, T.; Emer, J.; Sze, V. Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Networks "Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Accessed Terms of Use. *Deep Convol. Neural Netw.* **2017**, *52*, 127–138.
12. Moons, B.; Uytterhoeven, R.; Dehaene, W.; Verhelst, M. 14.5 Envision: A 0.26-to-10TOPS/W subword-parallel dynamic-voltage-accuracy-frequency-scalable Convolutional Neural Network processor in 28nm FDSOI. In Proceedings of the 2017 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, USA, 5–9 February 2017; Volume 60, pp. 246–247.
13. Bankman, D.; Yang, L.; Moons, B.; Verhelst, M.; Murmann, B. An always-on 3.8 μ J/86% CIFAR-10 mixed-signal binary CNN processor with all memory on chip in 28nm CMOS. *IEEE J. Solid-State Circuits* **2018**, *54*, 158–172. [CrossRef]
14. Moons, B.; Bankman, D.; Yang, L.; Murmann, B.; Verhelst, M. BinarEye: An always-on energy-accuracy-scalable binary CNN processor with all memory on chip in 28nm CMOS. In Proceedings of the 2018 IEEE Custom Integrated Circuits Conference (CICC), San Diego, CA, USA, 8–11 April 2018; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2018; pp. 1–4.
15. Chua, L. Memristor-The missing circuit element. *IEEE Trans. Circuit Theory* **1971**, *18*, 507–519. [CrossRef]
16. Chua, L. Resistance switching memories are memristors. *Appl. Phys. A* **2011**, *102*, 765–783. [CrossRef]
17. Wang, L.; Yang, C.; Wen, J.; Gai, S.; Peng, Y. Overview of emerging memristor families from resistive memristor to spintronic memristor. *J. Mater. Sci. Mater. Electron.* **2015**, *26*, 4618–4628. [CrossRef]
18. Zidan, M.A.; Chen, A.; Indiveri, G.; Lu, W.D. Memristive computing devices and applications. *J. Electroceramics* **2017**, *39*, 4–20. [CrossRef]
19. Burr, G.W.; Shelby, R.M.; Sebastian, A.; Kim, S.S.; Kim, S.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]
20. Ambrogio, S.; Narayanan, P.; Burr, G.W.; Tsai, H.-Y.; Shelby, R.M. Recent progress in analog memory-based accelerators for deep learning. *J. Phys. D Appl. Phys.* **2018**, *51*, 283001.
21. Ielmini, D. Brain-inspired computing with resistive switching memory (RRAM): Devices, synapses and neural networks. *Microelectron. Eng.* **2018**, *190*, 44–53. [CrossRef]
22. Sengupta, A.; Srinivasan, G.; Roy, D.; Roy, K. Stochastic Inference and Learning Enabled by Magnetic Tunnel Junctions. In Proceedings of the 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 1–5 December 2019.
23. Hodgkin, A.; Huxley, A. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Bull. Math. Boil.* **1990**, *52*, 25–71. [CrossRef]
24. Chua, L. Memristor, Hodgkin–Huxley, and Edge of Chaos. *Nanotechnology* **2013**, *24*, 383001. [CrossRef] [PubMed]
25. Strukov, D.B.; Snider, G.S.; Stewart, D.R.; Williams, R.S. The missing memristor found. *Nature* **2008**, *453*, 80–83. [CrossRef] [PubMed]
26. Chen, A.; Hutchby, J.; Zhirnov, V.; Bourianoff, G. *Emerging Nanoelectronic Devices*; Chen, A., Hutchby, J., Zhirnov, V., Bourianoff, G., Eds.; John Wiley & Sons Ltd.: Chichester, UK, 2014; ISBN 9781118958254.
27. Choi, S.J.; Lee, J.H.; Bae, H.J.; Yang, W.Y.; Kim, T.W.; Kim, K.H. Improvement of CBRAM Resistance Window by Scaling Down Electrode Size in Pure-GeTe Film. *IEEE Electron Device Lett.* **2009**, *30*, 120–122. [CrossRef]
28. Wang, S.Y.; Huang, C.W.; Lee, D.Y.; Tseng, T.Y.; Chang, T.C. Multilevel resistive switching in Ti/Cu_xO/Pt memory devices. *J. Appl. Phys.* **2010**, *108*, 114110. [CrossRef]
29. Rahaman, S.Z.; Maikap, S.; Das, A.; Prakash, A.; Wu, Y.H.; Lai, C.-S.; Tien, T.-C.; Chen, W.-S.; Lee, H.-Y.; Chen, F.T.; et al. Enhanced nanoscale resistive switching memory characteristics and switching mechanism using high-Ge-content Ge0.5Se0.5 solid electrolyte. *Nanoscale Res. Lett.* **2012**, *7*, 614. [CrossRef]

30. Padovani, A.; Woo, J.; Hwang, H.; Larcher, L. Understanding and Optimization of Pulsed SET Operation in HfO_x-Based RRAM Devices for Neuromorphic Computing Applications. *IEEE Electron Device Lett.* **2018**, *39*, 672–675. [CrossRef]
31. Sung, C.; Padovani, A.; Beltrando, B.; Lee, D.; Kwak, M.; Lim, S.; Larcher, L.; Della Marca, V.; Hwang, H. Investigation of I-V linearity in TaO_x-Based RRAM devices for neuromorphic applications. *IEEE J. Electron Devices Soc.* **2019**, *7*, 404–408. [CrossRef]
32. Wright, C.D.; Liu, Y.; Kohary, K.I.; Aziz, M.M.; Hicken, R.J. Arithmetic and Biologically-Inspired Computing Using Phase-Change Materials. *Adv. Mater.* **2011**, *23*, 3408–3413. [CrossRef] [PubMed]
33. 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]
34. 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. Mol. Neurosci.* **2016**, *10*, 384012. [CrossRef] [PubMed]
35. Chanthbouala, A.; Crassous, A.; Garcia, V.; Bouzehouane, K.; Fusil, S.; Moya, X.; Allibe, J.; Dlubak, B.; Grollier, J.; Xavier, S.; et al. Solid-state memories based on ferroelectric tunnel junctions. *Nat. Nanotechnol.* **2012**, *7*, 101–104. [CrossRef] [PubMed]
36. Wen, Z.; Li, C.; Wu, D.; Li, A.; Ming, N. Ferroelectric-field-effect-enhanced electroresistance in metal/ferroelectric/semiconductor tunnel junctions. *Nat. Mater.* **2013**, *12*, 617–621. [CrossRef] [PubMed]
37. Boyn, S.; Girod, S.; Garcia, V.; Fusil, S.; Xavier, S.; Deranlot, C.; Yamada, H.; Carrétéro, C.; Jacquet, E.; Bibes, M.; et al. High-performance ferroelectric memory based on fully patterned tunnel junctions. *Appl. Phys. Lett.* **2014**, *104*, 052909. [CrossRef]
38. Chappert, C.; Fert, A.; Van Dau, F.N. The emergence of spin electronics in data storage. *Nat. Mater.* **2007**, *6*, 813–823. [CrossRef]
39. Yu, S. Neuro-Inspired Computing with Emerging Nonvolatile Memory. *Proc. IEEE* **2018**, *106*, 260–285. [CrossRef]
40. Zhang, T.; Yang, K.; Xu, X.; Cai, Y.; Yang, Y.; Huang, R. Memristive Devices and Networks for Brain-Inspired Computing. *Phys. Status Solidi (RRL)-Rapid Res. Lett.* **2019**, *13*, 1–21.
41. Padovani, A.; Larcher, L.; Puglisi, F.M.; Pavan, P. Multiscale modeling of defect-related phenomena in high-k based logic and memory devices. In Proceedings of the 2017 IEEE 24th International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA), Chengdu, China, 4–7 July 2017; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2017; pp. 1–6.
42. Puglisi, F.M.; Larcher, L.; Padovani, A.; Pavan, P. Bipolar Resistive RAM Based on HfO_2: Physics, Compact Modeling, and Variability Control. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2016**, *6*, 171–184. [CrossRef]
43. Puglisi, F.M.; Padovani, A.; Pavan, P.; Larcher, L. Advanced modeling and characterization techniques for innovative memory devices: The RRAM case. In *Advances in Non-Volatile Memory and Storage Technology*; Woodhead Publishing: Sawston, UK; Cambridge, UK, 2019; pp. 103–135.
44. Kund, M.; Beitel, G.; Pinnow, C.-U.; Rohr, T.; Schumann, J.; Symanczyk, R.; Ufert, K.; Muller, G. Conductive bridging RAM (CBRAM): An emerging non-volatile memory technology scalable to sub 20nm. In Proceedings of the IEEE InternationalElectron Devices Meeting, 2005. IEDM Technical Digest, Washington, DC, USA, 5 December 2005; pp. 754–757.
45. Nail, C.; Molas, G.; Blaise, P.; Piccolboni, G.; Sklenard, B.; Cagli, C.; Bernard, M.; Roule, A.; Azzaz, M.; Vianello, E.; et al. Understanding RRAM endurance, retention and window margin trade-off using experimental results and simulations. In Proceedings of the 2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 3–7 December 2016.
46. Goux, L.; Radhakrishnan, J.; Belmonte, A.; Witters, T.; Devulder, W.; Redolfi, A.; Kundu, S.; Houssa, M.; Kar, G.S. Key material parameters driving CBRAM device performances. *Faraday Discuss.* **2019**, *213*, 67–85. [CrossRef] [PubMed]
47. Lee, B.; Wu, Y.; Yu, S.; Wong, P. Low-power $TiN/Al_2O_3/Pt$ resistive switching device with sub-20 μA switching current and gradual resistance modulation. *J. Appl. Phys.* **2011**, *110*, 94104.
48. 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]

49. Long, B.; Li, Y.; Jha, R. Switching Characteristics of Ru/HfO_2/TiO_{2-x}/Ru RRAM Devices for Digital and Analog Nonvolatile Memory Applications. *IEEE Electron Device Lett.* **2012**, *33*, 706–708. [CrossRef]
50. Matveyev, Y.; Egorov, K.; Markeev, A.; Zenkevich, A.; Matveyev, Y. Resistive switching and synaptic properties of fully atomic layer deposition grown TiN/HfO_2/TiN devices. *J. Appl. Phys.* **2015**, *117*, 044901. [CrossRef]
51. Wang, Y.-F.; Lin, Y.-C.; Wang, I.-T.; Lin, T.-P.; Hou, T.-H. Characterization and Modeling of Nonfilamentary Ta/TaO_x/TiO_2/Ti Analog Synaptic Device. *Sci. Rep.* **2015**, *5*, 10150. [CrossRef]
52. Bersuker, G.; Gilmer, D.C.; Veksler, D.; Kirsch, P.; Vandelli, L.; Padovani, A.; Larcher, L.; McKenna, K.; Shluger, A.; Iglesias, V.; et al. Metal oxide resistive memory switching mechanism based on conductive filament properties. *J. Appl. Phys.* **2011**, *110*, 124518. [CrossRef]
53. Foster, A.S.; Gejo, F.L.; Shluger, A.L.; Nieminen, R.M. Vacancy and interstitial defects in hafnia. *Phys. Rev. B* **2002**, *65*, 174117. [CrossRef]
54. Ramo, D.M.; Gavartin, J.L.; Shluger, A.L.; Bersuker, G. Spectroscopic properties of oxygen vacancies in monoclinic HfO_2 calculated with periodic and embedded cluster density functional theory. *Phys. Rev. B* **2007**, *75*, 205336. [CrossRef]
55. Robertson, J.; Gillen, R. Defect densities inside the conductive filament of RRAMs. *Microelectron. Eng.* **2013**, *109*, 208–210. [CrossRef]
56. Vandelli, L.; Padovani, A.; Larcher, L.; Southwick, R.G.; Knowlton, W.B.; Bersuker, G. A Physical Model of the Temperature Dependence of the Current through SiO_2/HfO_2 Stacks. *IEEE Trans. Electron Devices* **2011**, *58*, 2878–2887. [CrossRef]
57. Vandelli, L.; Padovani, A.; Larcher, L.; Broglia, G.; Ori, G.; Montorsi, M.; Bersuker, G.; Pavan, P. Comprehensive physical modeling of forming and switching operations in HfO_2 RRAM devices. In Proceedings of the 2011 International Electron Devices Meeting, Washington, DC, USA, 5–7 December 2011.
58. Larcher, L.; Padovani, A.; Pirrotta, O.; Vandelli, L.; Bersuker, G. Microscopic understanding and modeling of HfO_2 RRAM device physics. In Proceedings of the 2012 International Electron Devices Meeting, San Francisco, CA, USA, 10–13 December 2012.
59. Vandelli, L.; Padovani, A.; Larcher, L.; Bersuker, G. Microscopic Modeling of Electrical Stress-Induced Breakdown in Poly-Crystalline Hafnium Oxide Dielectrics. *IEEE Trans. Electron Devices* **2013**, *60*, 1754–1762. [CrossRef]
60. Padovani, A.; Larcher, L.; Bersuker, G.; Pavan, P. Charge Transport and Degradation in HfO_2 and HfO_x Dielectrics. *IEEE Electron Device Lett.* **2013**, *34*, 680–682. [CrossRef]
61. Padovani, A.; Larcher, L.; Pirrotta, O.; Vandelli, L.; Bersuker, G. Microscopic Modeling of HfO_x RRAM Operations: From Forming to Switching. *IEEE Trans. Electron Devices* **2015**, *62*, 1998–2006. [CrossRef]
62. Larcher, L.; Padovani, A.; Di Lecce, V. Multiscale modeling of neuromorphic computing: From materials to device operations. In Proceedings of the 2017 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2–6 December 2017; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2017.
63. Larcher, L.; Padovani, A.; Puglisi, F.M.; Pavan, P. Extracting Atomic Defect Properties From Leakage Current Temperature Dependence. *IEEE Trans. Electron Devices* **2018**, *65*, 5475–5480. [CrossRef]
64. Zhang, M.; Huo, Z.; Yu, Z.; Liu, J.; Liu, M. Unification of three multiphonon trap-assisted tunneling mechanisms. *J. Appl. Phys.* **2011**, *110*, 114108. [CrossRef]
65. Larcher, L.; Padovani, A.; Pavan, P. Leakage current in HfO2 stacks: From physical to compact modeling. In Proceedings of the Workshop on Compact Modeling, San Jose, CA, USA, 18–21 June 2012.
66. Di Ventra, M. *Electrical Transport in Nanoscale Systems*; Cambridge University Press: Cambridge, UK, 2008; Volume 34, ISBN 9780511755606.
67. McPherson, J.; Kim, J.-Y.; Shanware, A.; Mogul, H. Thermochemical description of dielectric breakdown in high dielectric constant materials. *Appl. Phys. Lett.* **2003**, *82*, 2121. [CrossRef]
68. Padovani, A.; Gao, D.Z.; Shluger, A.; Larcher, L. A microscopic mechanism of dielectric breakdown in SiO_2 films: An insight from multi-scale modeling. *J. Appl. Phys.* **2017**, *121*, 155101. [CrossRef]
69. Foster, A.S.; Shluger, A.L.; Nieminen, R.M. Mechanism of Interstitial Oxygen Diffusion in Hafnia. *Phys. Rev. Lett.* **2002**, *89*, 225901. [CrossRef]
70. Puglisi, F.M.; Pavan, P.; Padovani, A.; Larcher, L.; Bersuker, G. RTS noise characterization of HfO_x RRAM in high resistive state. *Solid-State Electron.* **2013**, *84*, 160–166. [CrossRef]

71. Veksler, D.; Bersuker, G.; Vandelli, L.; Padovani, A.; Larcher, L.; Muraviev, A.; Chakrabarti, B.; Vogel, E.; Gilmer, D.C.; Kirsch, P.D. Random telegraph noise (RTN) in scaled RRAM devices. In Proceedings of the 2013 IEEE International Reliability Physics Symposium (IRPS), Anaheim, CA, USA, 14–18 April 2013; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2013.
72. Puglisi, F.M.; Pavan, P.; Vandelli, L.; Padovani, A.; Bertocchi, M.; Larcher, L. A microscopic physical description of RTN current fluctuations in HfO_x RRAM. In Proceedings of the 2015 IEEE International Reliability Physics Symposium, Monterey, CA, USA, 19–23 April 2015; IEEE: Piscataway, NJ, USA, 2015.
73. Puglisi, F.M.; Larcher, L.; Padovani, A.; Pavan, P. A Complete Statistical Investigation of RTN in HfO_2-Based RRAM in High Resistive State. *IEEE Trans. Electron Devices* **2015**, *62*, 2606–2613. [CrossRef]
74. Nminibapiel, D.M.; Veksler, D.; Shrestha, P.R.; Campbell, J.P.; Ryan, J.T.; Baumgart, H.; Cheung, K.P.; Kim, J.-H. Impact of RRAM Read Fluctuations on the Program-Verify Approach. *IEEE Electron Device Lett.* **2017**, *38*, 736–739. [CrossRef] [PubMed]
75. 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]
76. Butcher, B.; Bersuker, G.; Young-Fisher, K.G.; Gilmer, D.C.; Kalantarian, A.; Nishi, Y.; Geer, R.; Kirsch, P.D.; Jammy, R. Hot forming to improve memory window and uniformity of low-power HfO_x-based RRAMs. In Proceedings of the 2012 4th IEEE International Memory Workshop, Milan, Italy, 20–23 May 2012; pp. 1–4.
77. Butcher, B.; Bersuker, G.; Vandelli, L.; Padovani, A.; Larcher, L.; Kalantarian, A.; Geer, R.; Gilmer, D. Modeling the effects of different forming conditions on RRAM conductive filament stability. In Proceedings of the 2013 5th IEEE International Memory Workshop, Monterey, CA, USA, 26–29 May 2013; pp. 52–55.
78. Traore, B.; Xue, K.-H.; Vianello, E.; Molas, G.; Blaise, P.; De Salvo, B.; Padovani, A.; Pirrotta, O.; Larcher, L.; Fonseca, L.R.C.; et al. Investigation of the role of electrodes on the retention performance of HfO_x based RRAM cells by experiments, atomistic simulations and device physical modeling. In Proceedings of the 2013 IEEE International Reliability Physics Symposium (IRPS), Anaheim, CA, USA, 14–18 April 2013.
79. Lorenzi, P.; Rao, R.; Irrera, F. Forming Kinetics in HfO_2-Based RRAM Cells. *IEEE Trans. Electron Devices* **2013**, *60*, 438–443. [CrossRef]
80. Chen, P.-Y.; Lin, B.; Wang, I.-T.; Hou, T.-H.; Ye, J.; Vrudhula, S.; Seo, J.-S.; Cao, Y.; Yu, S. Mitigating effects of non-ideal synaptic device characteristics for on-chip learning. In Proceedings of the 2015 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), Austin, TX, USA, 2–6 November 2015; pp. 194–199.
81. Larcher, L.; Padovani, A.; Woo, J.; Hwang, H.; Pesic, M. RRAM synapse optimization: From material stack to device performance. *IEEE Trans. Electron Devices*. under review.
82. Gao, B.; Kang, J.; Zhou, Z.; Chen, Z.; Huang, P.; Liu, L. Metal oxide resistive random access memory based synaptic devices for brain-inspired computing. *Jpn. J. Appl. Phys.* **2016**, *55*, 4. [CrossRef]
83. Chen, B.; Lu, Y.; Gao, B.; Fu, Y.; Zhang, F.; Huang, P.; Chen, Y.; Liu, L.; Liu, X.; Kang, J.; et al. Physical mechanisms of endurance degradation in TMO-RRAM. In Proceedings of the 2011 International Electron Devices Meeting; Institute of Electrical and Electronics Engineers (IEEE), Washington, DC, USA, 5–7 December 2011.
84. Chen, Y.Y.; Komura, M.; Degraeve, R.; Govoreanu, B.; Goux, L.; Fantini, A.; Raghavan, N.; Clima, S.; Zhang, L.; Belmonte, A.; et al. Improvement of data retention in HfO_2/Hf 1T1R RRAM cell under low operating current. In Proceedings of the 2013 IEEE International Electron Devices Meeting, Washington, DC, USA, 9–11 December 2013.
85. Puglisi, F.M.; Wenger, C.; Pavan, P. A Novel Program-Verify Algorithm for Multi-Bit Operation in HfO_2 RRAM. *IEEE Electron Device Lett.* **2015**, *36*, 1030–1032. [CrossRef]
86. Belmonte, A.; Fantini, A.; Redolfi, A.; Houssa, M.; Jurczak, M.; Goux, L. Optimization of the write algorithm at low-current (10μA) in Cu/Al_2O_3-based conductive-bridge RAM. In Proceedings of the 2015 45th European Solid State Device Research Conference (ESSDERC), Graz, Austria, 14–18 September 2015; pp. 114–117.
87. Woo, J.; Moon, K.; Song, J.; Kwak, M.; Park, J.; Hwang, H. Optimized Programming Scheme Enabling Linear Potentiation in Filamentary HfO_2 RRAM Synapse for Neuromorphic Systems. *IEEE Trans. Electron Devices* **2016**, *63*, 5064–5067. [CrossRef]
88. Islam, R.; Li, H.; Chen, P.-Y.; Wan, W.; Chen, H.-Y.; Gao, B.; Wu, H.; Yu, S.; Saraswat, K.C.; Wong, H.-S.P.; et al. Device and materials requirements for neuromorphic computing. *J. Phys. D Appl. Phys.* **2019**, *52*, 113001. [CrossRef]

89. 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]
90. 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] [PubMed]
91. Yu, S.; Gao, B.; Fang, Z.; Yu, H.; Kang, J.; Wong, H.-S.P. Stochastic learning in oxide binary synaptic device for neuromorphic computing. *Front. Mol. Neurosci.* **2013**, *7*, 1–9. [CrossRef] [PubMed]

© 2019 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/).