Open access • Journal Article • DOI:10.1002/AELM.201900198

Stochastic Memory Devices for Security and Computing — Source link

Roberto Carboni, Daniele Ielmini

Institutions: Polytechnic University of Milan

Published on: 01 Sep 2019 - Advanced electronic materials (John Wiley & Sons, Ltd)

Topics: Stochastic computing and Phase-change memory

Stochastic memory devices for security and computing

*Roberto Carboni and Daniele Ielmini**

R. Carboni, Prof. D. Ielmini
Dipartimento di Elettronica, Informazione e Bioingegneria and Italian Universities
Nanoelectronics Team, Politecnico di Milano, Milan 20133, Italy
E-mail: daniele.ielmini@polimi.it

With the widespread use of ubiquitous mobile computing and internet of things (IoT), secured communication and chip authentication become key requirements. Hardware-based security concepts generally provide the best performance in terms of good security standard, low power consumption, and large area density. In these concepts, the stochastic properties of nanoscale devices, such as the physical and geometrical variations of the process, are harnessed for true-random number generators (TRNGs) and physical unclonable functions (PUFs). Emerging memory devices, such as resistive switching memory (RRAM), phase change memory (PCM) and spin-transfer torque magnetic memory (STT-MRAM), rely on a unique combination of physical mechanisms for transport and switching, thus appearing as the ideal source of entropy for TRNG and PUFs.

This work provides an overview of stochastic phenomena in memory devices and their use for developing security and computing primitives. First, we provide a broad classification of methods to generate true random numbers via the stochastic properties of nanoscale devices. Then, we show practical implementations of stochastic TRNG, such as hardware security and stochastic computing. The future challenges of stochastic memory concepts are finally discussed.

Keywords: phase-change memory (PCM), resistive switching memory (RRAM), spin-transfer torque memory (STTRAM), true-random number generator (TRNG), physical unclonable function (PUF), statistical variations, device scaling, memristor, neuromorphic computing, hardware authentication, security primitive.

1. Introduction

Secret communication and secure information storage have been critical throughout the history of mankind, with the first examples of cryptography dating back to 1900 B.C. in Ancient Egypt [1]. From the ancient Sparta *scytale* and Julius Caesar cypher to the polyalphabetic cypher of Leon Battista Alberti and the Enigma machine in World War II, cryptography was mostly required in military scenarios [1, 2]. On the other hand, from the second half of the twentieth century, the increasing diffusion of information and communication technologies (ICT) spurred the exchange and processing of data to be performed in a secure and verified way. As a result, from the beginning of the digital era, the field of information security has been the subject of an increasing interest [3].

Information security has been a topic of intense research since the mid 1970s, when the main purpose was to guarantee the confidentiality and integrity of data within mainframe computers [4]. As mobile computers, internet of things (IoT) and cloud computing are becoming ubiquitous, there is an ever-increasing need for secure communication [5]. Portable devices such as smartphones and tablets can now enable financial transactions and act as the primary authentication token for the user. As a result, there is a strong need for electronic chips that can (i) securely authenticate and be authenticated by other parties, (ii) securely handle private/sensitive information, and (iii) operate in an untrusted environment where the adversary might have physical access to the system [6]. These tasks must be implemented in mobile devices at the level of the integrated circuit (IC), featuring at the same time both low power consumption and small area occupation. For application in large scale IoT [7] and cyber-physical systems (CPS) [8], security methodologies must also feature high speed, low cost and robustness to physical and side-channel attacks [9].

Hardware-intrinsic security primitives such as true random number generators (TRNG) and physical unclonable functions (PUF) are gaining interest toward low-cost and high-performance security tools [10]. On the one hand, TRNG can conveniently and efficiently generate the random bitstreams required by most cryptographic and security applications [11], [12]. On the other hand, PUF can securely store a secret key in the random characteristics of an IC, by e.g. exploiting the random process fluctuations, and enable fast and low-cost authentication and secure key storage [6]. Nano-devices are currently considered as the most promising approach for TRNG and PUF thanks to the small area, the low power consumption, the scalability, the 3D integration, and the ability to offer intrinsic stochastic phenomena via the inherent physical transport and switching mechanisms. These properties are all extremely beneficial for portable and IoT applications. Nanoelectronics can provide scalable device concepts via either the well-established complementary metal oxide semiconductor (CMOS) technology, or via alternative memory concepts based on resistive, phase change, magnetic and ferroelectric materials [13], sometimes referred to as memristive devices [14]. CMOS-based TRNG [11] and PUF [15] were first introduced thanks to the strong integration capabilities and technological maturity. Nevertheless, they soon demonstrated a limited entropy quality and the need for increased area and power overhead to improve randomness [16]. On the other hand, memory devices are currently gaining increasing interest for hardware security thanks to their intrinsic stochastic behavior that can be harnessed for high-performance, low cost and low energy on chip entropy sources.

In addition to enabling data/hardware security, a simple and reliable TRNG source is becoming more and more attractive also for emerging computing paradigms, such as stochastic [17], [18] and brain-inspired computing [19], [20], which require large amounts of random bit sequences [21]. Stochastic memory devices can constitute the core for ultra-small and low-power security and computing primitives, which can perform a broad range of tasks directly within the memory [13]. This allows to avoid massive data movement between the CPU and the memory, which constitutes the so-called von Neumann bottleneck [22].

In this work, the proposed concepts of stochastic memory devices for security and computing are reviewed with emphasis on the fundamental physical mechanisms and the final applications in computing. Section II introduces the hardware primitives for security and computing applications, underlining the pivotal role of TRNG as entropy source. After an overview of CMOS-based TRNG solutions, in Section III the different stochastic memory technologies are introduced, with details on the device structure and physics. Section IV provides a classification and comparison of the stochastic phenomena at the basis of various TRNG concepts. The next Sections discuss the applications of the TRNG in security, such as the PUF concept (Section V), and in computing, like analogue multiplication performed by stochastic computing (Section VI) and stochastic neurons in cognitive computing hardware (Section VII). Finally, Section VIII concludes the Review, providing an outlook about the open challenges and active trends of the research.

2. Hardware primitives for security and computing

The widespread diffusion of interconnected devices for the IoT raises a strong need for secured communication between them. Smartphones have become the central hub for critical over-the-internet tasks, playing the role of personal authentication token, enabling financial transactions, storing private information and controlling home and car functionalities. In all these applications, guaranteeing secure data-transmission is of paramount importance. Security in internet-based communications usually requires the generation of random keys [11], [12], via on-chip random number generators (RNGs). For implementation in resource constrained IoT systems, RNGs should be compact and reliable, while featuring high-quality entropy, high throughput and low power consumption [23]. On-board RNGs are the entropy sources that constitute the backbone of hardware primitives for both security and computing applications.

For example, RNG-based security applications include physical unclonable functions (PUFs) [24], [25], which enable various schemes of unique identification of a chip for hardware authentication and IC counterfeiting prevention. Several emerging computing paradigms, such as stochastic [17], [18], [26] and brain-inspired computing [19], [20], also rely on large bitstreams of random analog/digital signals for their operation, thus requiring on-chip entropy sources. The widespread scenario of security and computing applications is summarized in **Figure 1**. In particular, the possible implementations of RNG are schematically reported in Figure 1a, while the centrality of RNG in enabling security and computing hardware primitives is underlined in Figure 1b.

A classical method for generating random bits is the pseudo-random number generator (PRNG), which generates random-looking bit streams with a deterministic algorithm initialized by a seed. The seed provides the required entropy for random number generation, since it is derived from random events such as interrupts, kernel calls, incoming TCP/IP requests, etc. [11], [27]. For example, a PRNG could be implemented by a linear-feedback shift register (LFSR), namely a digital circuit that can produce a deterministic, seed-dependent sequence of pseudo-random numbers [28], [29]. Although simple in principle, the LFSR is a finite-state machine [30], thus its output is periodic and not random over a sufficiently long observation time. Also, since the seed is derived from the user activity, it can be easily manipulated, or the knowledge of internal state and feedback tap structure of the LFSR could allow for operation monitoring [11]. These limitations of PRNG randomness are fundamentally linked with the deterministic, arithmetical generation algorithm, which cannot be the source of random numbers, as already recognized by J. von Neumann [31]. All these critical issues make the PRNG output easily susceptible to cryptoanalysis [11], [32], therefore ruling out PRNG as a good solution for security requirements in IoT devices [33].

Many of the PRNG drawbacks are solved by the true random number generator (TRNG), which generates an output bitstream based on an inherently stochastic physical process [25],

[34], [35]. The high unpredictability of TRNG provides superior entropy quality, leading to an enhanced reliability with respect to software-based RNGs [34], [25]. Various physical entropy sources were proposed and experimentally demonstrated for TRNG, like the random telegraph noise (RTN) in dielectrics [36], [37], stochastic quantum processes [35], stochastic spintronic phenomena [38], [39] and random fluctuations of transport and switching in memory devices [40], [41], [42].

Thanks to its maturity and ease of integration, the CMOS technology has been the technological platform for most hardware RNG to date [11], [36]. Various CMOS-based TRNGs were demonstrated by exploiting the noise in scaled MOSFET [36], the metastability at turn-on of the cross-coupled inverter pair constituting the core of static random-access memory (SRAM) [11], or the increased noise of dual drain MOSFET driving a voltage-controlled oscillator (VCO) [43]. However, CMOS-type TRNGs might suffer from various drawbacks: for instance, the colored noise spectrum due to capture/emission events in 1/f noise results in a biased output bitstream, requiring considerable post-processing and consequent circuit overhead. Noise in CMOS devices also critically depends on environmental/process fluctuations, whose impact can be minimized only with entropy-tracking feedback loops [11], thus resulting in additional power consumption, circuit area and added complexity.

On the other hand, alternative concepts based on emerging memory devices might enable ultra-small entropy source with high quality randomness, which makes these technologies very promising for TRNG.

3. Stochastic memristive devices

The scenario of memory devices has seen a large diversification in the last 2 decades, as alternative storage concepts have been proposed and scrutinized for possible use in high-density memory circuits. Emerging memory concepts generally depend on material-based storage, which relies on the physics of the constituent active materials [13] as opposed to the charge-based principles in flash memory [44] and conventional random access memory (RAM) like static RAM and dynamic RAM. Thanks to the specific physics and architecture, emerging memory devices show a better scalability with respect to flash memories [44], while featuring a performance which is closer to the CPU, thus enabling data-centric computing applications [45], [13]. Emerging memories thus appear promising as storage-class memory (SCM) concepts [45], [46], aiming at reducing the performance gap between memory and logic, also known as memory wall [47].

Figure 2 illustrates the most mature memory concepts developed so far, including resistive switching random access memory (RRAM) [48], [49], [50], [51], phase change memory (PCM) [52], magneto-resistive random access memory (MRAM) [53], and ferroelectric random access memory (FeRAM) [54]. Each of them is based on its peculiar transport and switching mechanisms, although sharing the 2-terminal structure, where the application of suitable voltage pulses can change one or more properties of the active material. Emerging memories are sometimes related to the memristor concept, which was originally postulated by L. Chua [55] as the fourth missing circuit element, and later experimentally linked to the resistive switching devices [56].

The RRAM device (Figure 2a) consists of a metal-insulator-metal (MIM) structure, where a dielectric layer is sandwiched between two metallic electrodes, namely the top electrode (TE) and the bottom electrode (BE). The device is initialized by a forming operation, where the application of a voltage induces the soft dielectric breakdown, leading to the formation of localized defect-rich path, or conductive filament (CF) [51]. During forming, the current is limited by a compliance system, such as a series transistor to yield the one-transistor/one-resistor (1T1R) structure [51]. Current limitation is essential to properly control the CF size and avoid the destructive breakdown of the dielectric layer. Conductivity is locally enhanced at the CF, thanks to the large density of defects such as oxygen vacancies in metal-oxide devices (Ox-RAM) [57] or metallic cations migrating from the electrodes in conductive-bridge devices (CBRAM) [58]. After the forming operation, the device shows a relatively low resistance thanks to the CF shunting the dielectric layer. This condition is referred to as low resistance state (LRS) of the RRAM. The CF can then be ruptured with the reset operation, which results in the high resistance state (HRS) of the RRAM. The LRS can be restored by a set operation, usually taking place at the opposite voltage with respect to the reset operation. The mechanisms for the reset and the set operations can be understood by the localized field-driven migration of defects resulting in the opening and reformation of the CF [59]. The set and reset processes are illustrated in Figure 2b, showing an idealized current-voltage (I-V) characteristics for a bipolar RRAM. Here, a positive voltage applied to the TE gives rise to the set transition from HRS to LRS at a characteristic voltage V_{set}, while a negative voltage applied to the TE causes the reset transition from LRS to HRS at a characteristic voltage V_{reset}. The resistance window between the LRS and the HRS is typically at least one order of magnitude, but can reach 5 orders of magnitude with the adoption of high band gap dielectrics such as SiO_x [60]. The compliance current I_C during the set operation allows to control the LRS resistance according to $R = V_C/I_C$, where V_C is a characteristic voltage generally lower than 1 V [61]. On the other hand, HRS modulation is achieved by controlling the maximum negative voltage along the reset sweep, namely the stop voltage V_{stop} [62].

In addition to the presented bipolar filamentary RRAM, other concepts of resistance switching devices were demonstrated. For example, in unipolar RRAM [14], [63], the set and reset transitions both take place at the same voltage polarity, i.e., either positive or negative voltage applied to the top electrode. This can be explained by the dominant role of the Joule heating in forming/dissolving the CF [64], [65]. Complementary switching, where the device can set and reset at both polarities have also been demonstrated [66]. Finally, non-filamentary RRAM devices were demonstrated, where the switching process is induced by the voltage-induced migration of defects which take place uniformly across the device cross section affecting the conductivity of a Schottky or tunneling barrier [67]. Thanks to the localized transport and switching, filamentary RRAM features an area-independent LRS resistance and reset current, as opposed to the area dependence of switching parameters for non-filamentary RRAM [68].

RRAM switching mechanism involves ionic migration induced by the voltage and the local Joule heating [59]. Because of the atomistic nature of the switching and the impact of the local microstructure, such as the presence of crystalline grain boundaries or grain orientation, the set and reset transitions are characterized by a significant random variation [69]. The properties of the CF can also undergo stochastic atomistic fluctuations such as defect relaxation and diffusion, which can cause significant variations of the programmed state with time [70], [71]. RRAM variations can give rise to both device-to-device (D2D) variability within a memory array or a wafer [72], and cycle-to-cycle (C2C) variability for the same device due to different defect configurations at each cycle. Such variations have an impact on all RRAM parameters, including the LRS and HRS resistance, the set voltage V_{set}, the reset voltage V_{reset} and the reset current I_{reset} [69].

RRAM devices are drawing considerable interest thanks to their ultra-fast and low current operation (below 1 ns with 15 μA pulses [73]), extremely high resistance window [60], area scalability, demonstrated in the lateral size range of 10 nm [74], and easy fabrication [62]. In addition, high endurance (10^8 cycles in SiO_x-based device [60] and 10^{12} in Ta_2O_{5-x}/TaO_{2-x}-based [75]) and high-temperature retention [62] have been demonstrated. This is stimulating an intense industrial research and development towards embedded and standalone RRAM, with several memory prototypes reported at 2x nm technology [76], [77], [78], and also embedded in microcontrollers [79].

Figure 2c schematically illustrates a PCM device, where the active switching material is a chalcogenide phase-change material, typically consisting in a ternary alloy of Ge, Sb and Te, such as $Ge_2Sb_2Te_5$ [80]. A phase-change material exists in two solid states having different structures, namely crystalline and amorphous states, with different optical and transport properties [52]. While the crystalline state is stable, the amorphous state is metastable, as it can remain unchanged over long periods of time, for example ten years at moderately high temperature, thus providing the basis for a non-volatile memory effect [13]. Figure 2d shows the programming characteristic of the PCM, namely the device resistance R measured after the application of a pulsed voltage V. Starting from the amorphous state, the application of a pulse with relatively low amplitude V induces Joule heating and consequent crystallization of the amorphous region. The application of a pulse with higher V causes the melting of the active region, followed by sudden freezing into the amorphous phase. The two phases are characterized by different band structure and resistivity. In particular, the crystalline phase shows low resistance due to the Fermi level being close to the valence band edge, thus causing a large concentration of carriers [52]. The amorphous phase is instead characterized by a relatively high resistance due to Fermi level being pinned at mid-gap [81], [82], [83].

A typical PCM cell has a mushroom shape (Figure 2c), consisting of a pillar-like bottom electrode, confining the current and Joule heating, and a hemispherical amorphous region. Cell architecture optimization showed that pore-like PCM cells are more energy efficient and more scalable with respect to mushroom cells thanks to an improved heat and current confinement [84]. Despite the bulk-type switching, as opposed to filamentary switching in RRAM, the PCM also suffers from variability effects, which make them relevant from a stochastic device viewpoint. The threshold switching phenomenon, namely the electronic transition from the off-state to the on-state in the amorphous phase, exhibits cycle-to-cycle variation in the same device, which can be characterized by either the distribution of threshold voltage for a given ramp rate, or the delay time for switching at a given applied voltage [85]. This can be explained by the dominant role of noise in triggering the threshold switching, as well as by stochastic variations in the amorphous region properties, such as the amorphous layer thickness and the position of Poole-Frenkel sites for carrier transport [86]. The amorphous phase also shows variations in the drift slope [87] and the crystallization time, in terms of both D2D and C2C variations [88]. All these variations contribute to make PCM a key enabling technology for stochastic computing primitives.

PCM looks particularly promising thanks to its technological maturity and strong performances, demonstrating sub-100 ns switching speed and more than 10^9 cycling endurance [89]. Device scaling has reached the tens of nm range, while achieving sub-10 nm cell size still requires additional research [90]. In recent years, 8 Gb prototypes were reported in the 20 nm technology [91] while the 3D crosspoint structures has been presented [90], [92].

Figure 2e shows the magnetic tunnel junction (MTJ), which is the fundamental building block for MRAM devices. The MTJ consists of a MIM stack, comprising two electrodes of ferromagnetic (FM) material, usually CoFeB, separated by a thin tunneling layer of dielectric material, usually MgO. Of the two FM layers, one is referred to as the pinned layer (PL), as its magnetic polarization is fixed, whereas the other is called free layer (FL), as the magnetic polarization can be switched in either direction. This leads to the existence of two polarization states, namely the parallel (P) state, where the two FM layers have the same magnetic polarization orientation, and the antiparallel (AP) state, where the two FM layers have opposite orientation. The two memory states correspond to two different resistive states due to the magnetoresistive effect [93], with the P- and AP-states characterized by a relatively low and high resistance, respectively. Switching between the two states takes place either by the direct application of a magnetic field in Toggle-MRAM [94] or by the spin-transfer torque (STT) effect, which was theoretically predicted [95], [96] and later experimentally demonstrated [97], [98], [99]. STT-based MRAM has become largely popular as storage class memory and possible SRAM replacement, thanks to the fast switching (below 1 ns), the extended endurance (larger than 10^{15} cycles) and the good area scaling trend, demonstrated to the 11 nm node [100]. Interest in the STT-MRAM has considerably increased after the demonstration of perpendicular spin-transfer torque (p-STT), where the magnetic polarization of the two FM layers is perpendicular to the MTJ plane, thus allowing a reduced switching current at equal retention time, hence lower power and improved scalability [101], [102] with respect to the in-plane concept (i-STT) [103].

Figure 2f shows a typical R-V curve for a STT-MRAM, indicating the two states and the corresponding switching transitions. The set transition from AP- to P-state takes place by the electrons tunneling from the PL, where their spin is polarized, to the FL, whose magnetic polarization is rotated by momentum conservation [95], [96], [104]. The reset transition, namely from P- to AP-state, takes place with the application of a current of opposite polarity. The relative change of the measured resistance is called tunnel magnetoresistance (TMR) and is typically around 200% for state-of-the-art MTJ [105]. Perpendicular STT magnetic memories are gaining considerable interest due to their fast switching [106], non-volatile states, high endurance [107], CMOS compatibility and low current operation [100]. Note that magnetization switching in MTJ depends on the interaction of the FL with random thermal fluctuations, thus STT-MRAM set/reset processes are inherently affected by stochastic variations [108]. Physical stochasticity of switching can indeed be leveraged for TRNG and other stochastic primitives based on STT-MRAM.

Endurance in STT-MRAM devices is generally limited by the dielectric breakdown of the tunnel layer, as a result of the repeated pulses inducing an electrical stress and a consequent degradation [107]. To improve the cycling reliability and achieve virtually unlimited endurance, the spin-orbit torque (SOT) MRAM has been proposed. The building block of SOT-MRAM devices is still the MTJ of Fig. 2e, although the switching between P and AP states is induced by an in-plane current flowing in a heavy metal (HM) metallization underneath the FL [109]. By decoupling the programming and read paths, the endurance can be largely improved, while maintaining all benefits of random MTJ switching for stochastic primitives [110], [111]. Recently, a 7 Mb prototype of embedded STT-MRAM with 22 nm technology has been reported [112].

Figure 2g shows a FERAM, which relies on a MIM structure including a ferroelectric (FE) layer sandwiched between 2 electrodes [113]. The polarization of the FE layer can be oriented by the application of a voltage to the MIM stack. Figure 2h shows the typical polarization-voltage (P-V) characteristic of a FERAM device, indicating that an applied positive voltage sweep results in the permanent polarization with remnant polarization $+P_r$, while a negative sweep results in a permanent remnant polarization $-P_r$. In both cases, the microscopic dipole within the FE layer can be switched only above a characteristic coercive voltage V_c. While the polarization switching can be sensed by a transient displacement current, there is generally no variation of resistance of the FERAM, which therefore cannot be used as a resistive memory similar to other memory concepts in Figure 2. Typical FE materials for implementation in FERAM technology are $PbZrTiO_3$ (PZT) [114], $BiSrTiO_3$ (BST) [115] and doped HfO_2 [116], the latter raising intense interest due to the large relevance of HfO_2 as a CMOS compatible material, extensively adopted as gate dielectric and RRAM material.

The ferroelectric field-effect transistor (FEFET) is a MOS transistor where the gate dielectric is replaced by a FE layer [117]. A change in the polarization of the FE layer thus results in a shift of the threshold voltage for channel inversion, where a positive gate voltage sweep induces a decrease of the threshold voltage, whereas a negative voltage sweep causes a threshold voltage increase. Note that this threshold voltage variation is opposite to the typical effect induced by electron trapping and detrapping at interface and border traps in n-type transistors. Although the FEFET structure has three terminals, it has the advantage that a polarization change can be directly sensed as a change in channel conductance, thus enabling readout by non-destructive current sensing. The FEFET structure is also suitable for high density NAND memories with 3D structure [118], integration with 28 nm CMOS technology [119], and multilevel programming using individual FE domains [120].

Although the physical mechanisms of switching and transport in RRAM, PCM, MRAM and FERAM are quite different, they all display random physical phenomena resulting in intrinsic stochastic variations of resistance level and switching parameters. While such variations are critical issues for memory applications [72], stochastic phenomena can serve as the physical source of entropy that is needed for hardware security and stochastic computing primitives. **Table 1** provides an overview and a comparison of the figures of merit for the different stochastic memory devices, such as cell size, multi-bit capability, cycling endurance, write/read timings and power consumption.

4. TRNG concepts

The generation of random bits using memory devices must rely on a stochastic physical mechanism of transport and/or switching. The most characteristic entropy source of conventional CMOS-based TRNG circuits is thermal noise, which can be amplified and digitalized to generate true random bits [34]. Other sources include the clock jitter in an oscillator ring [121], [122] and the threshold voltage mismatch in a digital latch [11], [123]. While these TRNG circuits can take advantage of the easy integration with CMOS technology, the circuit size is generally large, involving e.g., oscillators with several stages, and digital circuits with several transistors. For example, a TRNG based on the noise of scaled MOSFET [36], is characterized by 9000 μm^2 area and a 0.25 nJ/bit energy consumption, while the best-in-class TRNG CMOS solution proposed by Intel features a 4004 μm^2 at 2.9 pJ/bit energy consumption. The circuit size and power consumption can be strongly reduced by adopting memory devices, where the physics is inherently stochastic for what concern transport and set/reset switching, the latter resulting in variations in switching time, switching voltage, and even in the parameters of the device final state after switching.

4.1. Stochastic current fluctuations

Stochastic fluctuations in bistable defects within a resistive memory, such as a RRAM in either LRS or HRS, can lead to a large current fluctuation between two random current levels, called random telegraph noise (RTN) [124]. **Figure 3a** shows the measured current for a RRAM in LRS, indicating RTN current fluctuations. RTN can be attributed to the modification of the charge state of a bistable defect close to the CF, *e.g.* due to electron trapping/detrapping combined with a structural relaxation of the defect [70]. The negatively-charged site can thus modify the carrier concentration close to the defect, causing a depletion of conduction electrons and a consequent macroscopic change of the measured current [70]. This situation is shown in Figure 3b, where a CF with simplified cylindrical geometry is locally depleted from carriers by a surface electron trap fluctuating between a neutral and a negatively charged state. The relative impact of a single defect on the measured current increases as the filamentary path of the LRS becomes smaller, leading to an increased relative amplitude Δ*I*/*I*, where *ΔI* is the difference between the high and low current levels and *I* is their average value [124]. Figure 3c shows electrostatic simulations of the electron carrier density within the CF in the presence of a negatively charged defect at the surface of the CF. Coulombic repulsion can result in a partial depletion for a large CF size (ϕ = 7 nm), or in a full depletion for a small CF size (ϕ = 1.6 nm), corresponding to a small or large relative RTN amplitude *ΔI/I*, respectively [70]. A similar effect is the RTN affecting the channel current in a MOSFET, arising from the fluctuation of a charge state of an oxide defect [125].

To understand the impact of RTN on device behavior, Figure 3a shows the experimental current voltage (I-V) characteristics for a RRAM device with HfO_x switching layer. The current trace was measured during a negative sweep with the device in the LRS, showing clear RTN discrete transitions. The rate of RTN transitions increases with V_{read}, suggesting a voltage-accelerated bistable switching process. This phenomenon is presented in Figure 3d with constant-voltage measurements of current as a function of time. Here, the average switching time between the two current levels (*i.e.* the two RTN states) increases with the read voltage for V_{read} = 50, 200 and 350 mV. Conversely, the times for the current to switch from the high to the low value (Δt_{ON}) or vice versa (Δt_{OFF}) decrease with V_{read}. Figure 3d shows finite-element method (FEM) numerical simulations of RTN, indicating the same voltage-dependent behavior as experimental data [70]. The voltage dependence of RTN can be explained by the effect of Joule heating in accelerating RTN fluctuation, which is a thermally-activated transition between two metastable states of the defect. Similarly, RTN can be accelerated at high ambient temperature [70].

While RTN is associated to the stochastic fluctuations of an individual defect, the activation of more defects leads to $1/f^{\beta}$ noise spectrum of the RRAM read current. This type of noise can thus be explained by multi-trap capture and emission events in defects, such as oxygen vacancies, within the CF of the LRS or within the localized conduction path in the HRS [71]. Figure 3e shows the measured read current I_{read} of a RRAM cell in the LRS with average resistance R = 10 kΩ measured with a biasing voltage V_{read} = 10 mV. Current fluctuations due to the $1/f$ noise result in an increasing relative standard deviation $\sigma_I/\langle I_{read}\rangle$, where $\langle I_{read}\rangle$ is the average value of I_{read} at any time t, while σ_I is the corresponding standard deviation (Figure 3e). The relative standard deviation σ_I/I increases with time due to the increasing contribution of low frequency noise typical of $1/f$ noise [71]. Figure 3f shows the power spectral density (PSD) of the current S_I, for both measured and calculated current traces, evidencing a $1/f$ behavior.

Due to the prominence of RTN and 1/f noise in RRAM devices, TRNGs based on noise entropy source were explored in the recent years, e.g., adopting RTN as entropy source for RRAM-based TRNG. **Figure 4a** shows the typical circuit architecture of a TRNG which relies on RTN in a RRAM device [37]. A contact-resistive random-access memory (CRRAM), integrated on top of the drain contact of a MOS transistor, is used as the stochastic device for the TRNG. The resulting 1T1R circuit is biased at a voltage V_R, thus each RTN fluctuation across the CRRAM device causes a fluctuation of the drain voltage V_D because of the voltage divider circuit. A comparator compares V_D to a reference voltage V_{ref} leading to a binary random output, as shown in Figure 4b. Sampling the comparator output at discrete times with a clock frequency f_{CLK} leads to a random digital bitstream, provided that $f_{CLK} << f_{RTN}$, where f_{RTN} is the average rate of RTN fluctuations.

Although the scheme in Figure 4 is relatively simple, it also has few practical issues related to both circuit and the entropy harvesting concept. First, the circuit has a relatively large area, namely 2400 F^2 in 65 nm technology, i.e. 10 μm^2 [37]. Most importantly, the RTN amplitude, rate and uniformity in the RRAM device is typically difficult to control and predict. In general, a good RNG should provide an unbiased output, which has equal 50% probability of generating either a "0" or a "1". Considering the described entropy extraction technique based on RTN phenomena, an unbiased RNG is attained only if the measured RTN signal remains at the high value for 50% of the observation time, and at the low value for the other 50%. Also, as previously described, RTN is affected by temperature and voltage, which might make the entropy source unstable in time. In addition, the amplitude of the RTN signal should be large enough to be distinguished by the integrated comparator, while the reference level V_{ref} needs to be precisely adjusted to the specific RTN signal, which depends on the resistance value and on the fluctuations level. An unbalanced output bitstream, *i.e.* a non-uniformity of the "0"/"1" probability, can be compensated with digital post-processing, such as the von Neumann algorithm [31] or cascaded XOR gates [21]. However, such post-processing circuits generally occupy additional area and consume extra energy [25].

To overcome these difficulties, **Figure 5a** shows a circuit to harvest $1/f$ noise for TRNG [40]. In this concept, the noisy current trace is sampled at t and t+Δt, then the difference of the two sampled currents $\Delta I = I(t+\Delta t) - I(t)$ is compared to 0. Finally, the random bit value is assigned to 0 or 1 depending on ΔI being either positive or negative, respectively. With respect to the RTN scheme of TRNG, this differential scheme features both a reduced area of 0.256 μm^2 (or 160 F^2 in 40 nm technology) and a low bias in the probability of extracting a "0" or a "1" bit, as the differential current ΔI shows a Gaussian distribution with average $\Delta I = 0$. The circuit design (Figure 5b) allows for a precise extraction of the current value using a timing sense amplifier (TSA) and a resistance-to-time converter (RTC) [126], while the parallel configuration of multiple devices enables up to 32 Mbps operation, with a 0.04 nJ/bit energy efficiency. Test results in Figure 5c show a minimum entropy higher than 0.999 over a broad range of temperature ($– 40 < T[°C] < 120$) and supply voltage V_{DD}. The good performance of this differential scheme is further demonstrated by the P-value, namely a figure-of-merit (FOM) for randomness of the random bit stream, for 1000 groups of 1 Mb bitstream for the bit frequency SP-800-22 statistical test specified by the National Institute for Standards and Technology (NIST) [127] (Figure 5d).

4.2. Stochastic delay time

Noise-based entropy sources are fundamentally limited by the amplitude and frequency unpredictability. To improve the controllability of TRNG, the stochastic properties of switching, such as delay time and voltage, can be used as entropy source. **Figure 6a** is a conceptual sketch for stochastic delay time in RRAM, where a square voltage pulse is applied to a device in the HRS. If the applied voltage is comparable to V_{set}, then a set transition from the HRS to the LRS takes place after a certain delay time t_D, marked by the read current increasing from I_{HRS} to I_{LRS}. Also, t_D decreases for increasing applied voltage [59]. Repeated experiments on the same device indicate that t_D is subject to a large statistical variation from cycle to cycle, as schematically shown in Figure 6a. The stochastic behavior of t_D is a consequence of the ion migration in RRAM being dependent on the local microstructure and atomistic migration of ions [71]. The delay time t_D typically shows a Poissonian distribution $P(t_D) = 1/\tau \exp(-t_D/\tau)$, where τ is the characteristic time constant for the set transition. This can be explained by the set transition being a thermally-activated process to overcome an energy barrier E_A which controls the time constant according to the Arrhenius law $\tau = \tau_0 \exp(E_A/kT)$, where τ_0 is a constant, k is the Boltzmann constant and T is the local temperature [128]. This behavior is demonstrated by distributions in Figure 6 for increasing voltages 2.6 V (b), 3.2 V (c) and 3.6 V (d) applied to a device in the HRS [129]. The time constant τ can be obtained by fitting the data to an exponential function and is shown in Figure 6e as a function of the applied voltage. The exponentially decreasing τ reflects the voltage-induced lowering of the effective energy barrier E_A, [130], [59]. These data highlight that, although the single switching event is stochastic, the switching delay time distribution can be predicted and controlled by the applied voltage [131], [129].

Figure 7a shows a TRNG circuit where the entropy source is the stochastic switching time in a RRAM device with Ag TE and Ag-doped SiO_2 dielectric layer [132]. In this type of devices, the Ag migration from the TE results in the formation of an unstable CF, which decays soon after the set transition with a retention time ranging from few μs to few ms [133], [134], [135], [136]. The volatile behavior is the result of the large diffusivity of Ag combined with the mechanical compressive stress in the dielectric layer [137] and the tendency to minimize the surface to volume ratio of the CF [134], [138]. In the circuit of Figure 7a, the volatile RRAM device is connected with a series resistance in a voltage divider configuration. Figure 7b shows a TRNG cycle, where the application of a voltage pulse V_1 (1) causes a set transition in the RRAM device after a stochastic delay time t_D, which causes the voltage V_2 across the series resistance to increase above the reference voltage V_{ref} (2) and the comparator output to switch to a high voltage V_3 (3). The stochastic time t_D is then compared with the clock period T_{CLK} (4) by a AND gate, yielding the pulse train V_5 (5), which is in turn measured by the counter in units of the clock period T_{CLK} (6). Based on the stochastic t_D being either an even or odd number of clock periods, a random bit 0 or 1 is assigned.

A nonvolatile RRAM could be used in this concept as well, although a reset pulse would be needed to re-initialize the device to the HRS at the beginning of a new cycle, thus increasing power consumption and complexity with respect to the volatile RRAM in Fig. 7. On the other hand, the pulse voltage V_1 should be carefully tuned to match the time window $T_{CLK} < t_D < t_P$, namely, the delay time should be significantly larger than the clock period and significantly shorter than the pulse width of V_1. This requirement usually introduces complicated probability tracking techniques [139].

To overcome the difficult device biasing scheme, devices with stochastic switching irrespective of the external biasing should be developed. An example of bias-independent random switching device is the superparamagnetic tunnel junction, sketched in **Figure 8a**. The device structure is similar to the MTJ stack of Fig. 2e, as it displays two stable magnetic states, P and AP, corresponding to low and high resistance, respectively. These two stable states are usually interpreted within the energy potential profile of Figure 8b, where P and AP states are metastable states separated by an energy barrier of amplitude ΔE dictated by the magnetic anisotropy. In a non-volatile MRAM, the FL geometry and its interface with the MgO are engineered in order to guarantee high perpendicular magnetic anisotropy (PMA), resulting in a large $\Delta E >> kT$ [100]. As a result, the FL magnetic state is stable at room temperature, and can switch to the opposite magnetization only under an external perturbation, such as a current pulse for STT-MRAM [21], [100]. On the other hand, the FL lateral dimension in a superparamagnetic tunnel junction is relatively small, thus resulting in the energy barrier being of the order of kT [21]. As a result, the FL magnetization spontaneously switches between the two stable states due to the enhanced sensitivity to thermal fluctuations [140], [141].

Figure 8c shows the time evolution of the MTJ resistance under a sensing current of 10 μA, which induces a negligible perturbation on the FL magnetization [108], while maximizing the junction lifetime thanks to the reduced stress on the dielectric [107]. The resistance changes with time by random abrupt transitions between the two resistance levels, corresponding to the two stable magnetic states of the FL. Figure 8c also shows a digitalized version of the same signal, which is obtained with a comparator. Figure 8d reports the histograms for the distribution of the transition times from AP state to P state and vice versa. The two distributions can be fitted by an exponential law, suggesting that a Poissonian statistics governs the spontaneous switching process. Such random behavior can be used as entropy source for a TRNG, where random bits can be assigned by sampling the voltage across the device at constant frequency [21].

Note that in order to obtain an unbiased random bitstream with 50% “0”/”1” probability, the distributions of the transition times should be exactly the same. However, Figure 8d indicates that the device spends more time in the P state, which was explained by the stray field induced by the PL on the FL, thus making the P-state more energy favorable [142]. Post-processing of the random bit stream can be used for eliminating the bias, although with an increased area and energy consumption.

More in general, extracting entropy from the stochastic switching time can be difficult due to its sensitivity to device parameters and process variations, requiring a probability tracking of the applied voltage for every TRNG on the same chip, or in separated chips [143].

4.3. Stochastic switching voltage

The stochastic switching voltage can also be used as fundamental entropy source in TRNG. In this concept, instead of measuring the delay time for a switching transition, one can monitor the device for a fixed amount of time, with the switching probability becoming the entropy source. This approach is based on the cycle-to-cycle variability of resistive switching devices, which demonstrate a distribution of switching voltages during repeated current-voltage measurements. For instance, considering the distribution of the set transition voltage, the application of an average set voltage $<V_{set}>$ to the device in the HRS induces a set transition with 50% probability. As a result, the device resistance measured after the application of a random set pulse of voltage $<V_{set}>$ shows a bimodal distribution, where the two sub-distributions refer to LRS and HRS. The random bitstream can thus be generated by assigning a bit value “0” or “1” to the LRS or HRS, respectively [41].

The concept of voltage-based TRNG is illustrated in **Figure 9a**, showing the measured I-V characteristics for a bipolar RRAM device during six set/reset consecutive cycles [41]. All the switching parameters, such as LRS and HRS resistance values and the set/reset switching voltages show a significant cycle-to-cycle variability, explained by the random process of formation and disruption of the CF at the microscopic level [69].

The characterization of the random set transition is performed with the sequence of triangular pulses in Figure 9b, including: (1) a positive set pulse to deterministically initialize the device in LRS, (2) a negative reset pulse with a stop voltage V_{stop} to induce transition to the HRS, (3) a positive random set pulse, with amplitude $V_A = <V_{set}>$ to stochastically induce a set event, and finally (4) a read pulse to measure the cell resistance after the random set. Figure 9c shows the resulting resistance distribution for a random set experiment with $V_A = <V_{set}> =$ 1.6 V. The data indicate a bimodal distribution, with sub-distributions corresponding to LRS around R ≈ 12 kΩ and to HRS above 100 kΩ. The origin of the bimodal distribution is clarified by Figure 9d, showing three possible I-V characteristics behaviors during the random set pulse, corresponding to state A, B and C in Figure 9c. Case A corresponds to a cycle where the V_{set} was higher than the applied V_A, thus no set transition took place and the measured resistance was still in the HRS sub-distribution of Figure 9c. Case C corresponds instead to a cycle where the V_{set} was lower with respect to V_A, thus allowing a set transition controlled by the compliance current $I_C = 50$ μA fixed by the series MOS transistor. Therefore, the device was found in the LRS sub-distribution in Figure 9c. Finally, case B corresponds to an applied V_A very close to V_{set}, where the device was able to start the set transition, but not to complete it within the pulse time. This condition results in an intermediate state between LRS and HRS, associated to the small transition region of intermediate resistance between LRS and HRS in Figure 9c. From a practical point of view, the occurrence of intermediate case B, *i.e.* the population of the flat region in Figure 9c, can be reduced by either reducing the pulse-width of the stochastic set pulse or by using a specifically engineered shape (for example, a saw-tooth shape with an abrupt drop after reaching V_A) [41].

The absence of memory effect in the entropy harvesting process is a key requirement in evaluating a specific TRNG implementation. To test the absence of memory effects and the time uniformity of the entropy source, **Figure 10a** shows the measured resistance for 500 random set cycles, evidencing a stable bimodal behavior [41]. Figure 10b shows the correlation plot of the device resistance R after cycle i+1 as a function of the resistance after cycle i for all the cycles in Figure 10a, indicating four distinct regions for the cell being in the same state (LRS and HRS, for "00" and "11" respectively) or in different state during two consecutive cycles. Figure 10c shows the corresponding histogram representation for the four populations in Fig. 10b, evidencing balanced distributions around 25%. This supports the absence of correlation between two consecutive cycles and the true physical randomness of the random bit stream.

To guarantee a proper TRNG operation, a digitalization of the analogue output values is usually desired. This can be achieved by means of a positive-feedback regeneration circuit, such as a comparator [37]. Given the large resistance window between HRS and LRS in RRAM devices, a simple CMOS inverter can be used for digitalization instead of a larger analogue comparator, which would be instead needed for digitalization in RTN-based TRNG [37], [41]. To demonstrate the operation of the complete system, comprising a 1T1R RRAM-cell as entropy source and a CMOS inverter as a second stage, Figure 10d shows the distributions of measured and calculated resistance, while Figure 10e shows the distribution of output voltage of the digitalization stage. Despite the circuit simplicity, the high uniformity of the generated random bits can be obtained only if the applied voltage exactly matches $<V_{set}>$. This requires a preliminary probability tracking procedure resulting in a significant overhead in terms of algorithm complexity, latency, circuit area and power consumption [143].

Voltage-based TRNG can also be implemented with other memory devices showing stochastic switching properties, such as the STT-MRAM in **Figure 11** [143], [144]. The TRNG concept is illustrated in Figure 11a: first, the device is initialized in the AP state by a reset pulse, then the memory cell is perturbed by applying a voltage signal with amplitude and pulse width corresponding to the 50% switching probability, which can be tracked by accurate device characterization as shown in the contour plot of the switching probability in Figure 11b. After the perturbation, the device can randomly fall in either P or AP state with equal 50% probability. Figure 11c shows the shape of the applied pulse, including initial reset, perturb and final read phase [145]. A modification of this algorithm introduced a conditional perturb scheme which applies either a positive or a negative voltage perturb pulse depending on the previous MTJ state being AP or P, respectively [143]. Avoiding the initialization step, this concept features an improved MTJ endurance and power consumption per generated bit [143]. However, the perturbation concept shares the same drawback of the RRAM-based TRNG in Figs. 9 and 10, namely, the need for a probability tracking scheme to precisely tune the perturb signal and guarantee a 50% switching probability.

4.4. Differential schemes

To overcome the need for probability tracking in time- or voltage-based TRNG, differential schemes offer a promising solution that can be implemented with either RRAM [42] and STT-MRAM [139]. In these TRNG concepts, high quality entropy is obtained by the comparison of two resistive switching devices [42] or the same device in two consecutive cycles [139]. **Figure 12a** shows a parallel-set TRNG circuit, which comprises two RRAM devices P and Q in parallel configuration, both connected to a common select transistor where the drain terminal is connected to the input node of a digital comparator [42]. Figure 12b shows the applied waveform sequence for a TRNG cycle, including 1) an independent reset of P and Q, 2) a stochastic set pulse where the same voltage pulse is applied to P and Q in parallel, and 3) a differential read performed by the application of a voltage 2*V_{read} across the two devices, while the transistor is in the off state (*i.e.* V_{gate} = 0 V). The application of a positive voltage across the one-transistor/two-resistor (1T2R) structure in Figure 12a causes the set transition to take place randomly in one of the two RRAM only. In fact, as the first cell starts the transition from the HRS to the LRS, the voltage across both RRAMs decreases due to the voltage divider with the series transistor, thus preventing any set transition to happen in the second RRAM device. In this TRNG scheme, the cycle-to-cycle variability of V_{set} plays the role of entropy source. To demonstrate the operation of this concept, Figure 12c shows the cycle-to-cycle output values of V_{out} and V_{out2}, namely the transistor drain voltage and the comparator output, respectively, while Figure 12d shows their corresponding probability distributions.

Similar to the parallel set, where the parameter V_{set} is compared in the differential pair, other configurations are also possible, such as parallel reset and series reset configuration [42]. In general, the parallel set transition appears as the most promising approach, given the abrupt dynamics of the set transition as opposed to the more gradual reset event. The abrupt set transition is explained by the physical positive feedback characterizing the filament growth process, where the first initiation of the CF induces an increase of the local Joule heating, hereby accelerating the further growth of the filament [69]. This highlights the direct impact of the physics of the entropy-generating process on the performance of the specific TRNG.

A typical drawback of the differential pair approach is the assumption that cycle-to-cycle variation dominates over the cell-to-cell variation, which is not valid in general [146]. In presence of a large mismatch between the two cells in the differential pairs, e.g., where one cell systematically displays a lower V_{set} than the other cell, the TRNG might show deviations from the uniform behavior, leading to a biased output. Although this might be acceptable for PUF applications, where the random unique key has to be generated only once in the lifetime of the device, it might cause excessive non-uniformity in TRNG [42].

To prevent cell-to-cell mismatch in the differential TRNG concept, the random bit can be obtained by comparing two consequent switching cycles in the same device, instead of the same cycle in two distinct devices [139]. This is highlighted in **Figure 13**, which shows the applied voltage waveform and the corresponding device current during two consecutive set/reset cycles [139]. The waveform was applied to a perpendicular STT-MRAM device with CoFeB FM layers and MgO tunnel barrier [147]. Positive and negative triangular pulses are applied to induce stochastic set and reset events, respectively. Both pulses have a pulse duration of 1 μs, although the concept can be scaled to shorter pulse-widths, thanks to the fast switching of STT-MRAM. The stochastic switching is evidenced by the different set and reset voltages in cycles n–1 and n, resulting in different current waveforms between the two cycles. The same concept was also demonstrated with applied rectangular voltage pulses, where the stochastic switching is evidenced by the different switching delay time between the two different cycles. As a result, the same TRNG scheme can adopt either the stochastic distribution of switching voltage or the stochastic distribution of switching time as entropy source [139]. Figure 13b shows the probability distribution of the integrated current $Q_n = \int idt$, with i being the device current response, while Figure 13c shows the corresponding distribution of the difference $\Delta Q_n = Q_n - Q_{n-1}$. Given the highly symmetric distribution of ΔQ_n, the latter is chosen as the statistical variable for random bit generation, where a random bit value 0 or 1 is assigned for $\Delta Q_n < 0$ or $\Delta Q_n > 0$, respectively [139].

In general, TRNG schemes require whitening methodologies, like the Von Neumann algorithm [143] or the XOR operation [21], to achieve an unbiased bitstream. On the other hand, the concept described in Figure 13 demonstrated good performance against the standard statistical test of the National Institute for Standards and Technology (NIST) even without any whitening process to improve the data uniformity [139]. The triangular pulse in Figure 13 leads to improved results compared to the rectangular pulse, as the latter needs to be tuned to a relatively narrow range of voltage [139]. This behavior can be understood by considering the applied voltage dependence of the switching parameters t_D and V_{set} (or their equivalent parameters for the reset transition) for rectangular and triangular pulses. Considering the temperature- and voltage-activated Poissonian process for STT switching, in the case of a rectangular pulse, the delay time for the set transition t_D can be written as [148]:

$$t_D = \tau_0 \exp(\Delta(1 - V/V_0)), \tag{1}$$

where V_0 and τ_0 are constant, V is the applied voltage, and Δ is the thermal stability factor. Given the exponential dependence in Equation (1), only a narrow window of voltages corresponds to a switching time comparable to the applied pulse width t_P. On the other hand, the set voltage under a triangular pulse, where the applied voltage is ramped according to $V(t) = 2{*}V_A{*}t/t_P$, can be estimated from the integrated switching probability reaching 1, namely $\int_0^{t_{set}} t_D^{-1}\, dt$, where $t_{set} = t_P V_{set}/(2V_A)$ is the time for the set transition and t_D is defined by Equation (1). Thus, the set voltage V_{set} under a triangular pulse is given by [149], [128]:

$$V_{set} \approx V_0 \left(\frac{t_0 V_A}{t_P V_0}\right), \tag{2}$$

indicating a logarithmic dependence of V_{set} on the maximum applied voltage V_A. Owing to the different voltage dependence in Equations (1) and (2), a time-based approach (*i.e.* the rectangular pulse in this concept) requires probability tracking for optimal performance, while the triangular pulse approach displays a reduced sensitivity to external bias. As a result, the triangular pulse approach is also more robust against security threats such as the application of an external magnetic field or a temperature variation (*e.g.* induced by an external laser [23]) which would affect the switching voltage.

Table 2 summarizes the TRNG concepts presented in Section 4, describing the different technologies and the corresponding performances in terms of probability tracking and post-processing requirements. Specifically, there are obvious advantages in terms of area, energy consumption and absence of whitening/post-processing algorithm in most concepts are clear with respect to the previously described CMOS-based TRNG solutions.

5. Physical unclonable function (PUF) for chip authentication

Secret information transmission with classical mathematical cryptography has relied on sufficiently hard-to-break algorithms (i.e. the "lock") and secret keys since its inception [150]. While secret key generation can be efficiently performed by TRNG, secret key storage typically involves a relatively large area occupation and power consumption, since it usually relies on nonvolatile electrically erasable read-only memory (EEPROM) or a battery-backed static random-access memory (SRAM). In resource-constrained environments, such as low-power IoT devices, securely storing secret keys with low energy consumption is becoming an increasingly difficult task [151]. In addition, emerging attack techniques like data tampering and side-channel attacks constitute severe security threats [152], [23]. This has led to an intense research interest for hardware-intrinsic security primitives where secret key storage is not in the digital memory.

In this scenario, stochastic memory devices enabling physical unclonable function (PUF) are a promising solution. A PUF is a physical system that statistically maps an input digital word to an output one through a secret key depending on an intrinsically stochastic property of the chip, *e.g.* the silicon process variation or the inherent physical variability of device parameters [6]. Note that in general the physical mapping implemented by the hardware is extremely difficult and prohibitively expensive to characterize and reproduce, guaranteeing the high security of PUF systems. These properties make PUF an excellent scheme to uniquely identify either a component or a circuit, thus acting as an electronic fingerprint and enabling hardware authentication and preventing IC counterfeiting [15].

The operation of a PUF system is schematically shown in **Figure 14**, where PUF is represented as a black box that for each input challenge c returns an output response r = *f*(c), with *f* describing the unique internal physical characteristics of the PUF. A specific PUF instance is defined by the set of possible challenge-response pairs (CRP), which are sampled and recorded by a central verification authority to be used later. Typically, the device P where a PUF is installed is delivered through a trusted supply chain or environment to prevent counterfeiting. Once deployed to the final user of P, the central authority can verify the authenticity of the device P by simply checking whether the response r'_i to a specific challenge c_i corresponds to the previously recorded response r_i (Figure 14). This hardware authentication technique can also prevent any unauthorized overproduction of a particular device from the manufacturer [15].

The classification of PUF systems is based on the number of unique CRPs, which defines two main categories, namely (i) the weak PUF, which only supports a limited set of CRPs, and (ii) the strong PUF, with a large set of CRPs [6]. The number of available CRPs increases linearly or polynomially with the number of basic cells, *i.e.* with the number of building blocks forming the PUF systems [24], while it increases exponentially in a strong PUF [15]. The weak PUF is also referred to as physically obfuscated key (POK), since it is mostly used for the generation and storage of cryptographic keys [153], [154].

Digital static random access memory (SRAM) offers one of the most common implementation of the PUF circuit exploiting the metastable states of cross-coupled inverters [123]. In fact, the response bit from each inverter pair addressed by the challenge is determined by which of the two nominally equal-sized inverters reaches the tri-state point faster. Memory-based PUFs are relatively easy to design even with low area overhead, although they are typically weak PUFs (POKs), since their CRP space is limited by the available memory capacity [155]. As a result, their CRP set can be completely explored within a polynomial measurement time, compromising their use for authentication. On the other hand, strong PUFs are practically immune to brute-force attacks, thanks to their large CRP set [15].

Although there is no general metric to certify a PUF system in terms of security properties, the following characteristics can be considered the best figures of merit (FOM) for PUF [15]:

- Reliability: A PUF should always give the same response to a given challenge over a wide range of operating conditions (voltage, temperature etc.)
- Unpredictability: The PUF response to an arbitrary challenge should not be predicted based on either the CRPs of another PUF or from the previous CRPs of the same PUF.
- Unclonability: The CRP mapping of a PUF cannot be physically or mathematically cloned, even for the original manufacturer of the PUF.
- Physical Unbreakability: Any physical attempt to maliciously modify the PUF should result in a malfunction or a permanent damage of the chip.

Even though PUF systems are extremely promising for low-cost chip authentication, they should be strong enough against emerging attacks aiming at building a model of the PUF. Such kind of attacks try to break the PUF security by developing a model of the specific PUF instance based on the observation of a subset of input/output pairs. For instance, machine-learning attacks have been shown to be relatively successful in breaking PUFs [156], [157].

To develop a strong PUF for hardware-security, TRNG such as those discussed in Section 4 can enable a small circuit area, low power consumption, and high performance in terms of uniqueness and reliability. For instance, TRNG based on stochastic RRAM are suitable for the implementation of reconfigurable PUF, although the memory implementation results in a limited number of CRPs, hence not suitable for a strong PUF [158]. **Figure 15a** shows a strong PUF concept based on a cross-point array of RRAM devices [155]. Here, the entropy source is provided by the broad distribution of RRAM resistance in LRS and HRS shown in Figure 15b, while the current sneak paths in the array allow to enlarge the set of CRPs. A sneak path is a parasitic current flowing in non-selected cells with low resistance, which is detrimental for cell read-out margin in pure memory applications [159], [160], [161]. In this implementation, instead, sneak paths provide the unclonable function which enables the exponential scaling of the CRP set needed in the strong PUF. In the NxN cross-point PUF of Figure 15a, the challenge consists in a N-bit vector applied to the N rows, where an input bit value of 1 corresponds to an applied voltage equal to V_{DD}, while the row is left floating for a bit value of 0. The current from the N columns is then read and converted to an N-bit response by a sense amplifier. Theoretically, the maximum number of CRPs is 2^N, since each row may be either floating or biased. The actual number of CRPs is reduced since 50% of the rows are required to be biased in order to generate a comparable range of column currents for different challenges [155]. Thus, it is estimated that CRP set is around 5 x 10^{75} for an array of 256 x 256 bits. RRAM devices in the array are randomly initialized at the beginning of the PUF operation, resulting in large stochastic current distribution (Figure 15b). A similar PUF concept exploited the variability of self-heating PCM cells [162], [163] in cross-point array, demonstrating the wide applicability of stochastic memory devices to PUF systems.

6. Stochastic computing

Various unconventional methods of computation are currently under scrutiny to possibly overcome the von Neumann bottleneck of standard computing in data-intensive computing tasks [22]. Among the emerging computing concepts, stochastic computing is drawing considerable interest thanks to its massive parallelism, soft error tolerant operation, reduced area and low power [164]. In this computing paradigm, first conceived in the 1960s with the pioneering works of Gaines [165], [166] and Poppelbaum *et al.* [167], data are treated as probabilities, *i.e.* each bit of a stochastic number N is randomly assigned a "1" value with probability p_N [164]. This radically differs from the usual binary representation, used in conventional computers [168].

Figure 16a shows a stochastic number generator (SNG) [164], where the output of a multibit TRNG is compared to a reference voltage V_{ref} at the input of an analogue comparator. The output will thus be a "0" if the generated random bit R is below V_{ref}, otherwise the output is "1". As a consequence, the probability of obtaining a "1" in the bitstream will be a stochastic codification of the analogue value V_{ref}/V_{DD}, with V_{DD} being the analogue voltage value of bit

"1". From a mathematical point of view, the output stochastic spiking signal can be modeled as a Bernoulli process, where each bit is independent of the previous ones [169].

To better understand the methodology of stochastic computing, we consider the case of the multiplication of two numerical data. While multiplication in a conventional computer [170] involves an extremely large number of transistors, only a simple AND gate is needed in stochastic computing [165]. This is shown in Figure 16b, where the AND output yields a "1" only if both input signals A e B (*i.e.* in the two input stochastic numbers) are equal to "1". The probability p_{out} for the output to be "1" is thus given by:

$$p_{out} = p_A*p_B, \quad (3)$$

where p_A and p_B are the probabilities for A and B to be "1", respectively. The accuracy of the stochastic multiplication increases with the length of the stochastic bitstream, where doubling the number of bits for each stochastic number can add a bit of extra precision to the multiplication [164]. As a result, stochastic computing tends to be preferred in low-precision applications, hence relatively short bit-streams [164]. Another concern is that A and B should be statistically independent in Equation (3).

The difficulties in producing truly uncorrelated random bitstream by PRNGs hindered the development of this computing paradigm in the early days of its conception. However, the availability of TRNG in stochastic memory devices provides the necessary stochastic bit stream within a small area and low power consumption. For instance, Gaba *et al.* [131], [171] demonstrated stochastic computing with random input spiking signals obtained by the stochastic switching of RRAM devices. Lv and Wang [110] showed that a single SOT-MRAM is capable of stochastic addition and multiplication with extremely low power and area occupation. A similar MRAM device showed a current-controlled Poissonian spike generation with the aid of simple CMOS circuitry [172]. The device structure, shown in Figure 16c, comprises a CoFeB/MgO/Ta MTJ built on top of a 10 nm-thick Ta layer. Here, the current injected in the Ta layer will produce a spin current orthogonal to the MTJ, inducing the magnetization reversal of the CoFeB free layer thanks to the SOT effect [109]. Note that the MTJ switching is inherently stochastic due to the thermal fluctuations affecting the free-layer, which is described by the switching probability plot as a function of the heavy metal current shown in Figure 16c.

A possible application of such controllable stochastic switching element is the probabilistic inference from real-time data, which is currently gaining interest for enabling low-cost cognitive intelligence. Figure 16d shows a Bayesian network (BN), namely a graphical model to describe conditional independence between variables [173], [174], [175]. Here, each node represents random variables, while each link describes the direct dependence among variables [173]. The BN in the figure include 4 variables, namely the probability of being "*cloudy* (C)", "*rainy* (R)", having the "*sprinkler on* (S)" or the grass being "*wet* (W)" [172]. The dependence between each variable is quantified with the conditional probabilities of a transition to a particular node from its parent node, as described by the conditional probability table (CPT) [172]. This BN can be used in a probabilistic inference engine to estimate the probability of hidden causes from a given observation [172]. For instance, the observation could be wet grass in Figure 16d, with the objective of estimating the cause, whether the sprinkler is on, or it is raining. To this purpose, Bayes' Rule allows for the calculation of the posterior probability P(A|B) that B is caused by A, given by:

$$P(A|B) = P(B|A)P(A)/P(B), \tag{4}$$

where P(A) is the probability of A, P(B) is the probability of B, and P(B|A) is the probability that A is caused by B. A Bayesian inference engine must clearly implement Bayes' rule of Equation 4, which can be achieved by various classical computing approaches such as a field programmable gate array (FPGA) [176], stochastic digital circuits [177], [172], or analog based probabilistic hardware [178]. With respect to these CMOS-based solutions, stochastic memory devices can provide huge advantages in terms of area occupation [179]. For instance, Figure 16e shows a circuit to calculate the probability for having the "sprinkler on" in the situation of "wet grass is true”, based on the Bayes’ rule $P(S = 1|W = 1) = P(S = 1 \cap W = 1)/P(W = 1)$. The intersection operation can be implemented by multiplying the two spiking signals with an AND gate, while the division operation can be performed by a dedicated circuit measuring the matching of the spiking rates of the two signals [180], [172]. In this circuit, probabilities are encoded within spiking signals generated by the SNG of Figure 16c, thus showing compact design, low power consumption, and high resilience with respect to variations and noise in the computational units [172].

7. Neuromorphic computing

Taking the inspiration from the extremely high area and power efficiency of neuro-biological systems, neuromorphic computing appears as a promising candidate for solving complicated tasks in a dynamic environment requiring adaptation [181]. An essential building block in neuromorphic engineering is the neuron, consisting in the active computing element generating spikes in response to the stimulation by other neurons or sensors [181]. Classical neuron models are the Hodgkin-Huxley model [182] and other threshold-based models [183], which however are quite demanding in terms of area occupation in CMOS technology, due to, e.g., either the large capacitors needed to emulate membrane potential dynamics, or the large number of transistors [184], [185]. Also, CMOS circuits are generally not suitable for generating stochastic behaviors which play a key role in spiking signal encoding and transmission [186]. On the other hand, stochastic memory devices can straightforwardly emulate the neuronal behavior in the nanometer scale.

Figure 17a shows a schematic representation of an artificial neuron, receiving weighted spiking signals from various pre-synaptic sources [187]. These signals are summed, *e.g.* via the Kirchhoff’s law, and integrated within the neuron, where the consequent increase of the membrane potential over a threshold generates an output spike. The integration function can be realized by a PCM device [187], where the crystalline fraction in the active material can play the role of the biological membrane potential [188]. Figure 17b shows, in fact, that the application of electrical spikes to the PCM leads to the growth of the crystalline phase, similar to the evolution of the membrane potential in the biological neuron. An increasing crystalline fraction leads to a conductance variation in response to incoming spikes, as also illustrated by Figure 17c for various spike amplitudes and pulse widths. As the PCM conductance reaches a threshold values of 2 μS, the PCM is reset to the initial state, and a firing event is generated. Figure 17c indicates that the average firing frequency increases for increasing pulse amplitude and duration. The PCM neuron can be scaled down to few tens of nanometers, while reproducing the typical stochastic dynamics of the biological neuron directly at the device level [189].

Recent investigations suggest that the intrinsic resilience of the brain against variation and noise is thanks to the redundancy of the neuron and synapses. For instance, a population of multiple binary resistive memories has been shown to lead to neuromorphic synapses with analogue tunable weight [190]. Neuron redundancy has been shown to be ubiquitous in the retina [191], in the visual cortex [192] and in the motor cortex [193], where information encoding and processing is performed by populations of neurons instead of single units. For instance, **Figure 18a** illustrates the mechanism for detecting the visual angle of an incoming stimulus via a population of neurons. Here, each neuron is associated to a specific range of visual angles, which is then detected by the weighted sum of neuron outputs. Each neuron of a specific population is characterized by a tuning curve, meaning that its response has an increased firing rate only for a narrow window of input values [194]. Population coding can be straightforwardly achieved by stochastic memory devices where the spiking frequency can be controlled by either the applied voltage or current. For instance, superparamagnetic tunnel junctions display current-controlled random resistive switching between the P- and AP- states, as shown in Figure 18b-d for increasing applied current [21], [194]. This stochastic switching displays the same Poissonian statistics of the neuron firing activity [195]. Figure 18e shows the spiking rate as a function of the applied current, namely, the device tuning curve, which can be explained by the classical spin-transfer torque theory [149], where the average spiking rate r is given by:

$$r = r_0/\cosh(\Delta E*I/k_BT*I_c), \tag{5}$$

where k_BT is the thermal energy, I is the biasing current, I_c is a critical current, and $r_0 = \varphi_0*\exp(-\Delta E/k_BT)$ is the natural frequency of the transition across an energy barrier ΔE with an attempt frequency $\varphi_0 \approx 1$ GHz. Calculations by Equation 5 in Figure 18e indicate a Gaussian function, thus allowing the construction of a basis set for computing by simply considering different peak positions [196]. Figure 18f shows a population ensemble, where junctions in each population are tuned to a different input current range. This can be achieved with a set of identical junctions by the application of a different bias current I_{bias} to each junction [194].

Physical randomness in memory devices also enables the description of biological neurons that operate at the edge of chaos [197]. In this regime, neural networks show improved computational capabilities [198], which can be used to solve optimization problems [199]. Some of these problems belong to the class of NP (non-deterministic polynomial)-hard problems, whose solution is usually an extremely intensive computing task. Nevertheless, this class of problems is of great interest, with applications to airline traffic control, CPU processes scheduling and gene sequencing. Recently, the intrinsically noisy behavior of memristive systems was applied to solve a spin glass problem with simulated annealing [200] and to the maximum cut problem [201], thus underlining the opportunities offered by devices showing chaotic behavior. Chaotic switching can be displayed by the niobium dioxide (NbO_2) Mott insulator, where the I-V curve displays two threshold-switching events as shown in **Figure 19a** [202]. Mott insulators show a temperature-dependent metal-insulator transition, which is assumed to be responsible for the second negative differential resistance (NDR-2) at relatively high current in Figure 19a. On the other hand, the first negative differential resistance (NDR-1) at relatively low current is attributed to the non-linear current transport [202].

Figure 19b shows the measured current while the device was biased with a constant voltage comparable to the threshold voltage in Figure 19a. The current shows large chaotic current oscillations which can be explained by local temperature fluctuations, while serving as a valuable computational resource, e.g. in Hopfield networks [203]. Hopfield networks can solve constraint satisfaction problems, such as the travelling salesman problem or the Ising optimization [204]. However, deterministic Hopfield networks must generally face the problem of getting stuck at a local minimum of the energy landscape, as shown in Figure 19c [202]. This condition can be overcome with a stochastic Hopfield network, which is sometimes referred to as Boltzmann machine [205], implemented with Mott insulators for describing chaotic oscillations and TaO_x memory devices for storing the states. Similar to the simulated annealing approach, chaotic oscillations allow for reaching the global minimum as shown in Figure 19d. Figure 19e shows the simulation results for a stochastic Hopfield network during 1000 random instances of the travelling salesman problem, demonstrating the role of random fluctuations in reaching the global minimum. This work underlines that the integration of a source of controllable chaos looks more and more important for future brain-inspired computation.

8. Conclusions and future outlook

The paradigm for data collection and processing has dramatically evolved in the last decades, going from a centralized model to the distributed IoT. This scenario involves intense internet-based communication among IoT devices, thus requiring enhanced security for data storage/transmission and hardware authentication. On the other hand, conventional hardware security primitives are not suitable for the resource constrained environment of IoT devices, where low area occupation and energy efficiency are of paramount importance. This challenge is currently spurring a strong research effort toward high-performance/low-power devices for hardware security functionalities. Originally conceived as next-generation memories, various memory technologies are currently gaining interest as ultra-small entropy sources, enabling hardware intrinsic security primitives like true random number generators (TRNGs). This paper reviews the physical processes leading to stochastic fluctuations in resistive switching memory and the corresponding circuital solutions enabling their exploitations, underlining the opportunities and challenges in using nanodevices as entropy sources for TRNG.

Although the current state-of-the-art for stochastic memory-based security primitives is already encouraging, many challenges need to be addressed for a practical implementation in IoT and integrated systems. In particular, a device optimization needs to be aimed at high-frequency operation (larger than 1 Gbit/s), low-energy per bit (tens of fJ range), aggressive area scalability (below 20 nm) and high endurance (at least 10^{16} cycles). Most importantly, a CMOS-compatible technological process is of paramount importance for an easy integration capability. Moreover, the single device should be engineered to maximize its stochastic behavior, which is generally unwanted in memory applications. From the circuital point of view, research should be focused on minimizing area, energy and complexity overhead. For instance, TRNG schemes which need neither post-processing algorithm nor entropy-tracking feedback loop allow for a more area/energy efficient design. In general, a device/system co-optimization methodology will prove to be more and more important.

In addition to typical TRNG application to cryptography, stochastic memory devices are the workhorse for other emerging applications, ranging from hardware authentication with PUFs to unconventional non-von Neumann computing. In this regard, an application-specific optimization of the device geometry, materials, and operation algorithms will be needed for reaching optimal performances. The proposed computing paradigms have been discussed, underlining their need for hardware TRNG which makes stochastic memory devices a good candidate for stochastic and neuromorphic computing hardware primitives with low power consumption and scalability to high density.

Finally, stochastic memories enable the hardware reconfigurability, where simple building blocks assembled in the same fundamental structure (*e.g.* the RRAM devices in a crosspoint array) can be used for various purposes, such as memory, computing or security. The multipurpose reconfigurability makes emerging memories a key enabling technology for IoT technology, paving the way for small area, low power and smart systems capable of performing a wide range of tasks (*e.g.* pattern recognition and classification, fast/low-power analog computation, authentication, etc.) within a single hardware chip.

Acknowledgements

The authors would like to thank E. Vernocchi for her critical reading of the manuscript. This article received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 648635).

Received: ((will be filled in by the editorial staff))
Revised: ((will be filled in by the editorial staff))
Published online: ((will be filled in by the editorial staff))

References

[1] G. Dieter, Computer Security, UK: John Wiley & Sons, 2005.

[2] B. A. Forouzan, Cryptography and Network Security, New York, NY, USA: McGraw-Hill, 2007.

[3] W. Stallings, Cryptography and Network Security, Pearson Education India, 2006.

[4] J. Rajendran, R. Karri, J. B. Wendt, M. Potkonjak, N. R. McDonald, G. S. Rose and B. T. Wysocki, "Nanoelectronic solutions for hardware security," *IACR Cryptology ePrint Archive,* no. 2012:575, 2012.

[5] C. Stergiou, K. E. Psannis, B.-G. Kim and B. Gupta, "Secure integration of iot and cloud computing," *Future Generation Computer Systems,* vol. 78, p. 964–975, 2018.

[6] C. Herder, M.-D. Yu, F. Koushanfar and S. Devadas, "Physical unclonable functions and applications: A tutorial," *Proceedings of the IEEE,* vol. 102, no. 8, p. 1126–1141, 2014.

[7] M.-W. Ryu, S.-S. L. Jaeho Kim and M.-H. Song, "Survey on internet of things," *SmartCR,* vol. 2(3), pp. 195-202, 2012.

[8] K.-K. R. Choo, M. M. Kermani, R. A.-. akhsh and M. Govindarasu, "Emerging embedded and cyber physical system security challenges and innovations," *IEEE Transactions on Dependable and Secure Computing,* vol. 3, p. 235–236, 2017.

[9] F. Tehranipoor, "Towards implementation of robust and low-cost security primitives for resource-constrained IoT devices," *arXiv preprint arXiv:1806.05332,* 2018.

[10] H. Nili, G. C. Adam, B. Hoskins, J. K. Mirko Prezioso, M. R. Mahmoodi, F. M. Bayat, O. Kavehei and D. B. Strukov, "Hardware-intrinsic security primitives enabled by analogue state and nonlinear conductance variations in integrated memristors," *Nature Electronics,* vol. 1(3):197, 2018.

[11] S. K. Mathew, S. Srinivasan, M. A. Anders, H. Kaul, S. K. Hsu, F. Sheikh, A. Agarwal, S. Satpathy and R. K. Krishnamurthy, "2.4 gbps, 7 mw all-digital pvt-variation tolerant true random number generator for 45 nm cmos high-performance microprocessors," *IEEE Journal of Solid-State Circuits,* vol. 47, no. 11, p. 2807–2821, 2012.

[12] J. Katz, A. J. Menezes, P. C. V. Oorschot and S. A. Vanstone, Handbook of applied cryptography, CRC Press, 1996.

[13] D. Ielmini and H.-S. P. W. Wong, "In-memory computing with resistive switching devices," *Nature Electronics,* vol. 1, pp. 333-343, 2018.

[14] J. J. Yang, D. B. Strukov and D. R. Stewart, "Memristive devices for computing," *Nature nanotechnology,* vol. 8(1):13, 2013.

[15] C.-H. Chang, Y. Zheng and L. Zhang, "A retrospective and a look forward: Fifteen years of physical unclonable function advancement," *IEEE Circuits and Systems Magazine,* vol. 17(3):32–62, 2017.

[16] G. S. Rose, "Security meets nanoelectronics for internet of things applications," *Proceedings of the 26th edition on Great Lakes Symposium on VLSI, ACM,* pp. 181-183, 2016.

[17] A. Alaghi and J. Hayes, "Survey of stochastic computing," *ACM Trans. Embed. Comput. Syst.,* vol. 12, no. 2S, p. 92:1–92:19, 2013.

[18] J. S. Friedman, L. E. Calvet, P. Bessière, J. Droulez and D. Querlioz, "Bayesian inference with Müller C-elements," *IEEE Trans. Circuits Syst. I, Reg. Papers,* vol. 63, no. 6, p. 895–904, 2016.

[19] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo and Y. Nakamura, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science,* vol. 347, no. 6197, p. 668–673, 2014.

[20] G. Pedretti, V. Milo, S. Ambrogio, R. Carboni, S. Bianchi, A. Calderoni, N. Ramaswamy, A. S. Spinelli and D. Ielmini, "Stochastic learning in neuromorphic hardware via spike timing dependent plasticity with RRAM synapses," *IEEE J. Emerg. Sel. Topics Circuits Syst.,* vol. 8, no. 1, pp. 77-85, 2018.

[21] D. Vodenicarevic, N. Locatelli, A. Mizrahi, J. S. Friedman, A. F. Vincent, M. Romera, A. Fukushima, K. Yakushiji, H. Kubota and S. Yuasa, "Low-energy truly random number generation with superparamagnetic tunnel junctions for unconventional computing," *Phys. Rev. Appl.,* vol. 8, no. 5, p. 054045, 2017.

[22] J. Backus, "Can programming be liberated from the von Neumann style?: a functional style and its algebra of programs," *Communications of the ACM,* vol. 21.8, p. 613–641, 1978.

[23] S. Ghosh, "Spintronics and security: Prospects, vulnerabilities, attack models, and preventions," *Proceedings of the IEEE,* vol. 104, no. 10, p. 1864–1893, 2016.

[24] R. Pappu, B. Recht, J. Taylor and N. Gershenfeld, "Physical One-Way Functions," *Science,* vol. 297, no. 5589, pp. 2026-2030 , 2002.

[25] S. Sahay and M. Suri, "Recent trends in hardware security exploiting hybrid CMOS-resistive memory circuits," *Semiconductor Science and Technology,* vol. 32, no. 12, 2017.

[26] W. Maass, "Noise as a resource for computation and learning in networks of spiking neurons," *Proceedings of the IEEE,* vol. 102, no. 5, p. 860–880, 2014.

[27] G. Alvarez and S. Li, " Some basic cryptographic requirements for chaos-based cryptosystems," *International journal of bifurcation and chaos,* vol. 16, no. 8, p. 2129–2151, 2006.

[28] Maxim Integrated, "Pseudo random number generation using linear feed-back shift registers," 2010. [Online]. Available: http://www.maximintegrated.com/an4400..

[29] Xilinx, "Efficient Shift Registers, LFSR Counters, and Long Pseudo- Random Sequence Generators," 7 July 1996. [Online]. Available: https://www.xilinx.com/support/documentation/application_notes/xapp052.pdf.

[30] J. Belzer, A. G. Holzman and A. Kent, Encyclopedia of Computer Science and Technology, CRC Press, 1975.

[31] J. V. Neumann, "Various techniques used in connection with random digits," *Appl. Math Ser.,* vol. 12, no. 5, pp. 36-38, 1951.

[32] J. Kelsey, B. Schneier, D. Wagner and C. Hall, "Cryptanalytic attacks on pseudorandom number generators," *In International Workshop on Fast Software Encryption, Springer,* p. 168–188, 1998.

[33] S. Chari, C. Jutla, J. R. Rao and P. Rohatgi, "A cautionary note regarding evaluation of aes candidates on smart-cards," *In Second Advanced Encryption Standard Candidate Conference,* p. 133–147, 1999.

[34] B. Jun and P. Kocher, "The Intel random number generator," *Cryptography Research Inc. White paper,* vol. 27, pp. 1-8, 1999.

[35] N. Gisin, G. Ribordy, W. Tittel and H. Zbinden, "Quantum cryptography," *Reviews of modern physics,* vol. 74, no. 1, p. 145, 2002.

[36] R. Brederlow, R. Prakash, C. Paulus and R. Thewes, "A low-power true random number generator using random telegraph noise of single oxide-traps," *IEEE International Solid-State Circuits Conference (ISSCC 2006). Digest of Technical Papers. ,* p. 1666–1675, 2006.

[37] C.-Y. Huang, W. C. Shen, Y.-H. Tseng, Y.-C. King and C.-J. Lin, "A contact-resistive-random-access-memory-based true-random-number generator," *IEEE Electron Device Letters,* vol. 33, no. 8, p. 1108, 2012.

[38] A. Fukushima, T. Seki, K. Yakushiji, H. Kubota, H. Imamura, S. Yuasa and K. Ando, "Spin dice: A scalable truly random number generator based on spintronics," *Applied Physics Express:,* vol. 7, no. 8, p. 083001, 2014.

[39] S. Chun, S.-B. Lee, M. Hara, W. Park and S.-J. Kim, "High-density physical random number generator using spin signals in multidomain ferromagnetic layer," *Advances in Condensed Matter Physics,* 2015.

[40] Z. Wei, Y. Katoh, S. Ogasahara, Y. Yoshimoto, K. Kawai, Y. Ikeda, K. Eriguchi, K. Ohmori and S. Yoneda, "True random number generator using current difference based on a fractional stochastic model in 40-nm embedded reram," *Electron Devices Meeting (IEDM), 2016 IEEE International,,* pp. 4-8.

[41] S. Balatti, S. Ambrogio, Z. Wang and D. Ielmini, "True random number generation by variability of resistive switching in oxide-based devices," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems,* vol. 5, no. 2, p. 214–221, 2015.

[42] S. Balatti, S. Ambrogio, R. Carboni, V. Milo, Z. Wang, A. Calderoni, N. Ramaswamy and D. Ielmini, "Physical unbiased generation of random numbers with coupled resistive switching devices," *IEEE Transactions on Electron Devices,* vol. 63, no. 5, p. 2029– 2035, 2016.

[43] S.-h. Zhou, W. Zhang and N.-J. Wu, " An ultra-low power cmos random number generator," *Solid-State Electronics,* vol. 52, no. 2, p. 233–238, 2008.

[44] C. M. Compagnoni, A. Goda, A. S. Spinelli, P. Feeley, A. L. Lacaita and A. Visconti, "Reviewing the Evolution of the NAND Flash Technology," *Proceedings of the IEEE,* 2017.

[45] H.-S. P. Wong and S. Salahuddin, "Memory leads the way to better computing," *Nature Nanotechnology,* vol. 10, no. 3, p. 191–194, 2015.

[46] R.F.Freitas and W.W.Wilcke, "Storage-classmemory:Thenextstorage system technology," *IBM Journal of Research and Development,* vol. 52, no. 4.5, p. 439–447, 2008.

[47] W. A. Wulf and S. A. McKee, "Hitting the memory wall: implications of the obvious," *ACM SIGARCH Computer Architecture News,* vol. 23, p. 20–24 , 1995.

[48] H. Akinaga and H. Shima, "Resistive random access memory (ReRAM) based on metal oxides," *Proceedings of the IEEE,* vol. 98, no. 12, p. 2237–2251, 2010.

[49] H.-S. P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F. T. Chen and M.-J. Tsai., "Metal–oxide RRAM," *Proceedings of the IEEE,* vol. 100, no. 6, p. 1951–1970, 2012.

[50] R. Waser and M. Aono, "Nanoionics-based resistive switching memories," *Nature materials,* vol. 6, no. 11, p. 833, 2007.

[51] D. Ielmini, "Resistive switching memories based on metal oxides: mech- anisms, reliability and scaling," *Semiconductor Science and Technology,* vol. 31, no. 6, p. 063002, 2016.

[52] S. Raoux, W. Welnic and D. Ielmini, "Phase change materials and their application to non-volatile memories," *Chem. Rev.,* vol. 110, p. 240–267, 2010.

[53] A. D. Kent and D. C. Worledge, "A new spin on magnetic memories," *Nat. Nanotech.,* vol. 41, no. 10, p. 187–191 , 2015.

[54] T. Mikolajick, C. Dehm, W. Hartner, I. Kasko, M. Kastner, N. Nagel, M. Moert and C. Mazure, "FeRAM technology for high density applications," *Microelectron. Reliab.,* vol. 41, p. 947–950, 2001.

[55] L. O. Chua, "Memristor-The Missing Circuit Element," *IEEE Transactions on Circuit Theory,* Vols. CT-18, no. 5, 1971.

[56] D. B. Strukov, G. S. Snider, D. R. Stewart and R. S. Williams, "The missing memristor found," *Nature,* vol. 453, p. 80–83, 2008.

[57] A. Beck, J. G. Bednorz, C. Gerber, C. Rossel and D. Widmer, "Reproducible switching effect in thin oxide films for memory applications," *Appl. Phys. Lett.,* vol. 77, no. 139, 2000.

[58] Q. Liu, J. Sun, H. Lv, S. Long, K. Yin, N. Wan, Y. Li, L. Sun and M. Liu, "Real-time observation on dynamic growth/dissolution of conductive filaments in oxide-electrolyte-based ReRAM," *Adv. Mater.,* vol. 24, no. 47, p. 1844–1849, 2012.

[59] S. Larentis, F. Nardi, S. Balatti, D. C. Gilmer and D. Ielmini., "Resistive switching by voltage-driven ion migration in bipolar RRAM—Part ii: Modeling.," *IEEE Transactions on Electron Devices,* vol. 59, no. 9, p. 2468–2475, 2012.

[60] A. Bricalli, E. Ambrosi, M. Laudato, M. Maestro, R. Rodriguez and D. Ielmini, "SiOx-based resistive switching memory (RRAM) for crossbar storage/select elements with high on/off ratio," *IEEE International Electron Devices Meeting (IEDM),* 2016 .

[61] D. Ielmini, "Modeling the universal set/reset characteristics of bipolar rram by field- and temperature-driven filament growth," *IEEE Transactions on Electron Devices,* vol. 58, no. 12, p. 4309–4317, 2011.

[62] A. Bricalli, E. Ambrosi, M. Laudato, M. Maestro, R. Rodriguez and D. Ielmini, "Resistive Switching Device Technology Based on Silicon Oxide for Improved ON–OFF Ratio—Part I: Memory Devices," *IEEE Transactions on Electron Devices,* vol. 65, no. 1, pp. 115-121.

[63] K. M. Kim, D. S. Jeong and C. S. Hwang, "Nanofilamentary resistive switching in binary oxide system: a review on the present status and outlook.," *Nanotechnology,* vol. 22, p. 254002, 2011.

[64] D. Ielmini, R. Bruchhaus and R. Waser, "Thermochemical resistive switching: Materials, mechanisms and scaling projections," *Phase Transition,* vol. 84, no. 570, 2011.

[65] U. Russo, D. Ielmini, D. Cagli and A. L. Lacaita, "Filament conduction and reset mechanism in NiO-based resistive-switching memory (RRAM) devices," *IEEE Transactions Electron Devices,* vol. 56, no. 186, 2009.

[66] F. Nardi, S. Balatti, S. Larentis, C. Gilmer and D. Ielmini, "Complementary switching in oxide-based bipolar resistive switching memory (RRAM)," *IEEE Transactions Electron Devices,* vol. 60, no. 70, 2013.

[67] A. Sawa, "Resistive switching in transition metal oxides," *Materials Today,* vol. 11, no. 6, 2008.

[68] D. Ielmini, S. Spiga, F. Nardi, C. Cagli, A. Lamperti, E. Cianci and M. Fanciulli, "Scaling analysis of submicrometer nickel-oxide-based resistive switching memory devices," *J. Appl. Phys.,* vol. 109, p. 034506, 2011.

[69] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy and D. Ielmini, "Statistical fluctuations in HfOx resistive-switching memory: Part i-set/reset variability," *IEEE Transactions on electron devices,* vol. 61, no. 8, p. 2912–2919, 2014.

[70] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy and D. Ielmini, "Statistical fluctuations in HfOx resistive-switching memory: Part ii—random telegraph noise.," *IEEE Transactions on Electron Devices,* vol. 61, no. 8, pp. 2920-2927, 2014.

[71] S. Ambrogio, S. Balatti, V. McCaffrey, D. C. Wang and 6. 2. Daniele Ielmini. Noise-induced resistance broadening in resistive switching memory—part i: Intrinsic cell behavior. IEEE Transactions on Electron Devices, "Noise-induced resistance broadening in resistive switching memory—part i: Intrinsic cell behavior.," *IEEE Transactions on Electron Devices,* vol. 62, no. 11, p. 3805–3811, 2015.

[72] S. Ambrogio, S. Balatti, V. McCaffrey, D. C. Wang and D. Ielmini., "Noise-induced resistance broadening in resistive switching memory—part ii: Array statistics.," *IEEE Transactions on Electron Devices,* vol. 62, no. 11, p. 3812–3819, 2015.

[73] B. J. Choi, A. C. Torrezan, J. P. Strachan, P. G. Kotula, A. J. Lohn, M. J. Marinella, Z. Li, R. S. Williams and J. J. Yang, "High-Speed and Low-Energy Nitride Memristors," *Adv. Funct. Mater.,* vol. 26, pp. 5290-5296, 2016.

[74] B. Govoreanu, G. Kar, Y-Y.Chen, V. Paraschiv, S. Kubicek, A.Fantini, I. P. Radu, L. Goux, S. Clima, R. Degraeve, N. Jossart, O. Richard, T. Vandeweyer, K. Seo, P. Hendrickx, G. Pourtois, H. Bender, L. Altimime, D. Wouters, J. Kittl and M. Jurczak, "10x10 nm^2 Hf/HfOx Crossbar Resistive RAM with Excellent Performance,

Reliability and Low-Energy Operation," *IEEE Int. Electron Devices Meet. (IEDM),* p. 31.6.1, 2011.

[75] M.-J. Lee, C. B. Lee, D. Lee, S. R. Lee, M. Chang, J. H. Hur, Y.-B. Kim, C.-J. Kim, D. H. Seo, S. Seo, U.-I. Chung, I.-K. Yoo and K. Kim, "A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta2O5−x/TaO2−x bilayer structures," *Nature Materials,* vol. 10, p. 625–630, 2011.

[76] T.-y. Liu, T. H. Yan, R. Scheuerlein, Y. Chen, J. K. Lee, G. Balakrishnan, G. Yee, H. Zhang, A. Yap, J. Ouyang, T. Sasaki, A. Al-Shamma and e. a. Chinyu Chen, "A 130.7-mm^2 2-Layer 32-Gb ReRAM Memory Device in 24-nm Technology," *IEEE Journal of Solid-State Circuits,* vol. 49, no. 1, p. 140, 2014.

[77] S. Yu, "Neuro-inspired Computing with Emerging Nonvolatile Memory," *Proceedings of the IEEE,* vol. 106, no. 2, pp. 260-285, 2018.

[78] P. Jain, U. Arslan, M. Sekhar, B. C. Lin, L. Wei, T. Sahu, J. Alzate-vinasco, A. Vangapaty, M. Meterelliyoz, N. Strutt, A. B. Chen, P. Hentges, P. A. Quintero, C. Connor, O. Golonzka, K. Fischer and F. Hamzaoglu, "A 3.6Mb 10.1Mb/mm^2 Embedded Non-Volatile ReRAM Macro in 22nm FinFET Technology with Adaptive Forming/Set/Reset Schemes Yielding Down to 0.5V with Sensing Time of 5ns at 0.7V," in *IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA, 2019.

[79] A. Kawahara, R. Azuma, Y. Ikeda, K. Kawai, Y. Katoh, Y. Hayakawa, K. Tsuji, S. Yoneda, A. Himeno, K. Shimakawa, T. Takagi, T. Mikawa and K. Aono, "An 8 Mb Multi-Layered Cross-Point ReRAM Macro With 443 MB/s Write Throughput," *IEEE Journal of Solid-State Circuits,* vol. 48, no. 1, p. 178, 2013.

[80] N. Yamada, E. Ohno, K. Nishiuchi and N. Akahira, "Rapid-phase transitions of GeTe-Sb2Te3 pseudobinary amorphous thin films for an optical disk memory.," *J. Appl. Phys.,* Vols. 69, 2849, 1991.

[81] N. F. Mott and E. A. Davis, Electronic processes in non-crystalline materials, Clarendon Press: Oxford, 1979.

[82] D. Ielmini, D. Sharma, S. Lavizzari and A. L. Lacaita, in *Proceedings of the International Reliability Physics Symposium*, Phoenix, AZ, 2008, IRPS.

[83] D. Ielmini and Y. Zhang, "Analytical model for subthreshold conduction and threshold switching in chalcogenide-based memory devices.," *J. Appl. Phys.,* vol. 52., no. 102, p. 054517, 2007.

[84] M. Boniardi, A. Redaelli, C. Cupeta, F. Pellizzer, L. Crespi, G. D'Arrigo, A. L. Lacaita and G. Servalli, "Optimization metrics for phase change memory (PCM) cell architectures," *2014 IEEE International Electron Devices Meet. (IEDM),* 2014.

[85] S. Lavizzari, D. Ielmini and A. L. Lacaita, "Transient simulation of delay and switching effects in phase-change memories," *IEEE Transactions on Electron Devices,* vol. 57, no. 12, pp. 3257-3264, 2010.

[86] D. Ielmini, "Threshold switching mechanism by high-field energy gain in the hopping transport of chalcogenide glasses," *Physical Review B,* vol. 78, no. 3, p. 035308, 2008.

[87] M. Boniardi, D. Ielmini, S. Lavizzari, A. L. Lacaita, A. Redaelli and A. Pirovano, "Statistics of resistance drift due to structural relaxation in phase-change memory arrays," *IEEE Transactions on Electron Devices,* vol. 57, no. 10, pp. 2690-2696, 2010.

[88] M. Rizzi, N. Ciocchini, A. Montefiori, M. Ferro, P. Fantini, A. L. Lacaita and D. Ielmini, "Cell-to-cell and cycle-to-cycle retention statistics in phase-change memory arrays," *IEEE Transactions on Electron Devices,* vol. 62, no. 7, pp. 2205-2211, 2015.

[89] S. Lai and T. O. Lowrey, "A 180 nm nonvolatile memory cell element technology for stand alone and embedded applications.," in *International Electron Device Meeting IEDM*, 2001.

[90] S. W. Fong, C. M. Neumann and H.-S. P. Wong, "Phase-Change Memory—Towards a Storage-Class Memory," *IEEE TRANSACTIONS ON ELECTRON DEVICES,* vol. 64, no. 11, pp. 4374-4358, 2017.

[91] Y. Choi, I. Song, M.-H. Park, H. Chung, S. Chang, B. Cho, J. Kim, Y. Oh, D. Kwon, J. Sunwoo, J. Shin, Y. Rho, C. Lee, M. G. Kang, J. Lee, Y. Kwon, S. Kim, J. Kim, Y.-J. Lee, Q. Wang, S. Cha, S. Ahn, H. Horii, J. Lee, K. Kim, H. Joo, K. Lee, Lee, Yoo and Jeong, "A 20nm 1.8V 8Gb PRAM with 40MB/s program bandwidth," in *IEEE International Solid-State Circuits Conference*, 2012.

[92] D.-C. Kau, S. Tang, I. V. Karpov, R. Dodge, B. Klehn, J. A. Kalb, J. Strand, A. Diaz, N. Leung, J. Wu, S. Lee, T. Langtry, K. Chang, C. Papagianni, J. Lee, J. Hirst, S. Erra, E. Flores, N. Righos, H. Castro and G. Spadini, "A stackable cross point phase change memory," in *IEDM*, 2009.

[93] C. Chappert, A. Fert and F. N. V. Dau, "The emergence of spin electronics in data storage," *Nature Materials,* vol. 6, no. 11, p. 813– 823, 2007.

[94] J. Slaughter, N. Rizzo, F. Mancoff, R. Whig, K. Smith, S. Aggarwal and S. Tehrani, "Toggle and spin-torque MRAM: status and outlook," *Magnetics Japan,* vol. 5, no. 4, p. 171–177, 2010.

[95] J. C. Slonczewski, "Current-driven excitation of magnetic multilayers," *Journal of Magnetism and Magnetic Materials,* vol. 159, no. 1-2, pp. L1– L7,, 1996.

[96] L. Berger, "Emission of spin waves by a magnetic multilayer traversed by a current,”," *Physical Review B,* vol. 54, no. 13, p. 9353–9358, 1996.

[97] M. Tsoi, A. Jansen, J. Bass, W.-C. Chiang, M. Seck, V. Tsoi and P. Wyder, "Excitation of a magnetic multilayer by an electric current," *Physical Review Letters,* vol. 80, no. 19, p. 4281, 1998.

[98] E. Myers, D. Ralph, J. Katine, R. Louie and R. Buhrman, "Current-induced switching of domains in magnetic multilayer devices.»," *Science,* vol. 285.5429 , p. 867–870, 1999.

[99] J. Katine, F. Albert, R. Buhrman, E. Myers and D. Ralph, "Current-driven magnetization reversal and spin-wave excitations in Co/Cu/Co pillars.»," *Physical Review Letters,* vol. 84, no. 14, p. 3149, 2000.

[100] D. Apalkov, B. Dieny and J. Slaughter, "Magnetoresistive random-access memory," *Proceedings of the IEEE,* vol. 104, no. 10, p. 1796–1830, 2016.

[101] S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura and H. Ohno, "A perpendicular-anisotropy CoFeB–MgO magnetic tunnel junction," *Nature Materials,* vol. 9, no. 9, p. 721–724, 2010.

[102] H. Sato, S. Ikeda and H. Ohno, "Magnetic tunnel junctions with perpendicular easy axis at junction diameter of less than 20 nm.," *Japanese Journal of Applied Physics,* vol. 56, no. 8, p. 0802A6, 2017.

[103] A. Driskill-Smith, D. Apalkov, V. Nikitin, X. Tang, S. Watts, D. Lottis, K. Moon, A. Khvalkovskiy, R. Kawakami and X. Luo, "Latest advances and roadmap for in-plane and perpendicular STT-RAM.," in *3rd IEEE International Memory Workshop (IMW).*, 2011.

[104] D. C. Ralph and M. D. Stiles, "Spin transfer torques.," *Journal of Magnetism and Magnetic Materials,* vol. 320.7, p. 1190–1216, 2008.

[105] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki and K. Ando, "Giant room-temperature magnetore- sistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions," *Nature materials,* vol. 3, no. 12, p. 868–871, 2004.

[106] J. J. Nowak, R. P. Robertazzi, J. Z. Sun, G. Hu, J.-H. Park, J. Lee, A. J. Annunziata, G. P. Lauer, R. Kothandaraman, E. J. O'Sullivan, P. L. Trouilloud, Y. Kim and D. C. Worledge, "Dependence of voltage and size on write error rates in spin-transfer torque magnetic random-access memory," *IEEE Magnetics Letters,* vol. 7, pp. 1-4, 2016.

[107] R. Carboni, S. Ambrogio, W. Chen, M. Siddik, J. Harms, A. Lyle, W. Kula, G. Sandhu and D. Ielmini, "Modeling of Breakdown-Limited Endurance in Spin-Transfer Torque Magnetic Memory Under Pulsed Cycling Regime," *IEEE Transactions on Electron Devices,* vol. 65, no. 6, pp. 2470-2478, 2018.

[108] A. F. Vincent, N. Locatelli, J. O. Klein, W. S. Zhao, S. G.-. Retailleau and D. Querlioz, "Analytical macrospin modeling of the stochastic switching time of spin-transfer torque devices," *IEEE Transactions Electron Devices,* vol. 62, no. 1, p. 164–170, 2015.

[109] I. M. Miron, K. Garello, G. Gaudin, P.-J. Zermatten, M. V. Costache, S. Auffret, S. Bandiera, B. Rodmacq, A. Schuhl and P. Gambardella, "Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection," *Nature ,* vol. 476, p. 189–193, 2011.

[110] Y. Lv and J.-P. Wang, "A single magnetic-tunnel-junction stochastic computing unit," in *2017 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2017.

[111] H. Chen, S. Zhang, N. Xu, M. Song, X. Li, R. Li, Y. Zeng, J. Hong and L. You, "Binary and Ternary True Random Number Generators Based on Spin Orbit Torque," in *IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2018.

[112] L. Wei, J. G. Alzate, U. Arslan, J. Brockman, N. Das, K. Fischer, T. Ghani, O. Golonzka, P. Hentges, R. Jahan, P. Jain, B. Lin, M. Meterelliyoz, J. O'Donnell, C. Puls, P. Quintero, T. Sahu, M. M. Sekhar, A. Vangapaty, C. Wiegand and F. Hamzaoglu, "A 7Mb STT-MRAM in 22FFL FinFET Technology with 4ns Read Sensing Time at 0.9V Using Write-Verify-Write Scheme and Offset-Cancellation Sensing Technique," in *IEEE International Solid-State Circuits Conference*, 2019.

[113] T.Mikolajick, C.Dehm, W.Hartner, I.Kasko, M.J.Kastner, N.Nagel, M.Moert and C.Mazure, "FeRAM technology for high density applications," *Microelectronics Reliability,* vol. 41, no. 7, pp. 947-950 , 2001.

[114] D. Takashima, Y. Takeuchi, T. Miyakawa, Y. Itoh, R. Ogiwara, M. Kamoshida, K. Hoya, S. M. Doumae, T. Ozaki, H. Kanaya, K. Yamakawa, I. Kunishima and Y. Oowaki, "A 76-mm2 8-Mb chain ferroelectric memory," *IEEE Journal of Solid-State Circuits,* vol. 36, no. 11, pp. 1713 - 1720, 2001.

[115] S. Sakai, M. Takahashi, K. Takeuchi, Q.-H. Li, T. Horiuchi, S. Wang, K.-Y. Yun, M. Takamiya and T. Sakurai, "Highly Scalable Fe(Ferroelectric)-NAND Cell with MFIS(Metal-Ferroelectric-Insulator-Semiconductor) Structure for Sub-10nm Tera-Bit Capacity NAND Flash Memories," in *2008 Joint Non-Volatile Semiconductor Memory Workshop and International Conference on Memory Technology and Design*, Opio, France, 2008.

[116] T. S. Boescke, J. Mueller, D. Brauhaus, U. Schroeder and U. Boettger, "Ferroelectricity in hafnium oxide thin films.," *Appl. Phys. Lett. ,* vol. 99, p. 102903, 2011.

[117] M. Takahashi and S. Sakai, "Self-Aligned-Gate Metal/Ferroelectric/Insulator/Semiconductor Field-Effect Transistors with Long Memory Retention," *Japanese Journal of Applied Physics,* vol. 44, no. 25, p. L 800–L 802, 2005.

[118] K. Florent, M. Pesic, A. Subirats, K. Banerjee, S. Lavizzari, A. Arreghini, L. D. Piazza, G. Potoms, F. Sebaai, S. R. C. McMitchell, M. Popovici, G. Groeseneken and J. V. Houdt, "Vertical Ferroelectric HfO2 FET based on 3-D NAND Architecture: Towards Dense Low-Power Memory," in *IEDM Tech. Dig. 43*, San Francisco, CA, USA, 2018.

[119] M. Trentzsch, S. Flachowsky, R. Richter, J. Paul, B. Reimer, D. Utess, S. Jansen, H. Mulaosmanovic, S. Müller, S. Slesazeck, J. Ocker, M. Noack, J. Müller, P. Polakowski, J. Schreiter, S. Beyer, T. Mikolajick and B. Rice, "A 28nm HKMG super low power embedded NVM technology based on ferroelectric FETs," in *IEDM Tech. Dig. 294*, San Francisco, CA, USA, 2016.

[120] H. Mulaosmanovic, J. Ocker, S. Müller, U. Schroeder, J. Müller, P. Polakowski, S. Flachowsky, R. v. Bentum, T. Mikolajick and S. Slesazeck, "Switching Kinetics in Nanoscale Hafnium Oxide Based Ferroelectric Field-Effect Transistors," *ACS Appl. Mater. Interfaces,* vol. 9, no. 4, p. 3792–3798, 2017.

[121] D. Schellekens, B. Preneel and I. Verbauwhede, "FPGA vendor agnostic true random number generator.," *International Conference on Field Programmable Logic and Applications, FPL'06. ,* pp. 1-6, 2006.

[122] B. Sunar, W. J. Martin and D. R. Stinson, "A provably secure true random number generator with built-in tolerance to active attacks.," *IEEE Transactions on Computers,* vol. 56, no. 1, p. 109–119, 2007.

[123] D. E. Holcomb, W. P. Burleson and K. Fu, "Power-Up SRAM State as an Identifying Fingerprint and Source of True Random Numbers," *IEEE Transactions on Computers,* vol. 58, no. 9, pp. 1198 - 1210, 2009.

[124] D. Ielmini, F. Nardi and C. Cagli, " Resistance-dependent amplitude of random telegraph-signal noise in resistive switching memories.," *Applied Physics Letters,* vol. 96, no. 5, p. 053503, 2010.

[125] E. Simoen, B. Kaczer, M. Toledano-Luque and C. Claeys, "Random telegraph noise: From a device physicist's dream to a designer's nightmare.," *ECS Transactions,* vol. 39, no. 1, pp. 3-15, 2011.

[126] Y. Yoshimoto, Y. Katoh, S. Ogasahara, Z. Wei and K. Kouno, "A ReRAM-based physically unclonable function with bit error rate < 0.5% after 10 years at 125 °C for 40nm embedded application.," in *2016 IEEE Symposium on VLSI Technology*, 2016.

[127] NIST, "A statistical test suite for random and pseudorandom number generators for cryptographic applications," *Special publication 800-22,* 2010.

[128] S. H. Jo, T. Chang, K.-H. Kim, S. Gaba and W. Lu, " Experimental, modeling and simulation studies of nanoscale resistance switching devices," *9th IEEE Conference on Nanotechnology IEEE-NANO 2009,* p. 493–495, 2009.

[129] S. H. Jo, K.-H. Kim and W. Lu, "Programmable resistance switching in nanoscale two-terminal devices," *Nano letters,* vol. 9, no. 1, p. 496–500, 2008.

[130] C. Cagli, F. Nardi and D. Ielmini, "Modeling of set/reset oper- ations in nio-based resistive-switching memory devices.," *IEEE Transactions on electron devices,* vol. 56, no. 8, p. 1712–1720, 2009.

[131] S. Gaba, P. Sheridan, J. Zhou, S. Choi and W. Lu, "Stochastic memristive devices for computing and neuromorphic applications," *Nanoscale,* vol. 5, no. 13, p. 5872–5878, 2013.

[132] H. Jiang, D. Belkin, S. E. Savel'ev, S. Lin, Z. Wang, Y. Li, S. Joshi, R. Midya, C. Li, M. Rao, M. Barnell, Q. Wu, J. J. Yang and Q. Xia, "A novel true random number generator based on a stochastic diffusive memristor.," *Nature communications,* vol. 8, no. 1, p. 882, 2017.

[133] T. Ohno, T. Hasegawa, T. Tsuruoka, K. Terabe, J. K. Gimzewski and M. Aono, "Short-term plasticity and long-term potentiation mimicked in single inorganic synapses," *Nature materials,* vol. 10, no. 8, p. 591, 2011.

[134] Z. Wang, S. Joshi, S. E. Savel'ev, H. Jiang, R. Midya, P. Lin, M. Hu, N. Ge, J. P. Strachan, Z. Li, Q. Wu, M. Barnell, G.-L. Li, H. L. Xin, R. S. Williams, Q. Xia and J. J. Yang, "Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing," *Nature materials,* vol. 16, no. 1, p. 101, 2017.

[135] A. Bricalli, E. Ambrosi, M. Laudato, M. Maestro, R. Rodriguez and D. Ielmini, " Resistive switching device technology based on silicon oxide for improved on–off ratio—part ii: Select devices.," *IEEE Transactions on Electron Devices,* vol. 65, no. 1, pp. 122-128, 2018.

[136] R. Midya, Z. Wang, J. Zhang, S. E. Savel'ev, C. Li, M. Rao, M. H. Jang, S. Joshi, H. Jiang, P. Lin, K. Norris, N. Ge, Q. Wu, M. Barnell, Z. Li, H. L. Xin, R. S. Williams, Q. Xia and J. J. Yang, "Anatomy of Ag/hafnia-based selectors with 10^10 nonlinearity.," *Advanced Materials,* vol. 29, no. 12, p. 1604457, 2017.

[137] S. Ambrogio, S. Balatti, S. Choi and D. Ielmini., "Impact of the mechanical stress on switching characteristics of electrochemical resistive memory.," *Advanced Materials,* vol. 26, no. 23, p. 3885–3892, 2014.

[138] W. Wang, M. Wang, E. Ambrosi, A. Bricalli, M. Laudato, Z. Sun, X. Chen and D. Ielmini, "Surface diffusion-limited lifetime of silver and copper nanofilaments in resistive switching devices," *Nature communications,* vol. 10, no. 1, p. 81, 2019.

[139] R. Carboni, W. Chen, M. Siddik, J. Harms, A. Lyle, W. Kula, G. Sandhu and D. Ielmini, "Random number generation by differential read of stochastic switching in spin-transfer torque memory," *IEEE Electron Device Letters,* vol. 39, no. 7, pp. 951-954, 2018.

[140] W. H. Rippard, R. Heindl, M. Pufall, S. Russek and A. Kos, "Thermal relaxation rates of magnetic nanoparticles in the presence of magnetic fields and spin-transfer effects," *Phys. Rev. B,* vol. 84, p. 064439, 2011.

[141] A. Mizrahi, N. Locatelli, R. Lebrun, V. Cros, A. Fukushima, H. Kubota, S. Yuasa, D. Querlioz and J. Grollier, "Controlling the phase locking of stochastic magnetic bits for ultra-low power computation," *Sci. Rep.,* vol. 6, no. 2016 , p. 30535.

[142] J. Hayakawa, S. Ikeda, Y. Lee, R. Sasaki, T. Meguro, F. Matsukura, H. Takahashi and H. Ohno, "Current-induced magnetization switching in MgO barrier based magnetic tunnel junctions with CoFeB/Ru/CoFeB synthetic ferrimagnetic free layer," *Jpn. J. Appl. Phys.,* vol. 45, p. L1057, 2006.

[143] W. H. Choi, Y. Lv, J. Kim, A. Deshpande, G. Kang, J.-P. Wang and C. H. Kim, "A magnetic tunnel junction based true random number generator with conditional perturb and real-time output probability tracking," *IEEE International Electron Devices Meeting (IEDM) ,* vol. 12–5, 2014.

[144] H. Zhao, A. Lyle, Y. Zhang, P. K. Amiri, G. Rowlands, Z. Zeng, J. Katine, H. Jiang, K. Galatsis, K. L. Wang, I. N. Krivorotov and J.-P. Wang, "Low writing energy and

sub nanosecond spin torque transfer switching of in-plane magnetic tunnel junction for spin torque transfer random access memory," *Journal of Applied Physics,* vol. 109, p. 07C720, 2011.

[145] S. Yuasa, A. Fukushima, K. Yakushiji, T. Nozaki, M. Konoto, H. Maehara, H. Kubota, T. Taniguchi, H. Arai, H. Imamura, K. Ando, Y. Shiota, F. Bonell, Y. Suzuki, N. Shimomura, E. Kitagawa, J. Ito, S. Fujita, K. Abe, K. Nomura and H. Noguchi, "Future prospects of MRAM technologies," in *IEEE International Electron Devices Meeting*, Washington, DC, USA, 2013.

[146] A. Fantini, L. Goux, R. Degraeve, D. Wouters, N. Raghavan, G. Kar, A. Belmonte, Y.-Y. Chen, B. Govoreanu and M. Jurczak, "Intrinsic switching variability in HfO2 RRAM.," *5th IEEE International Memory Workshop (IMW),* pp. 30-33, 2013.

[147] R. Carboni, S. Ambrogio, W. Chen, M. Siddik, J. Harms, A. Lyle, W. Kula, G. Sandhu and D. Ielmini, "Understanding cycling endurance in perpendicular spin-transfer torque (p-STT) magnetic memory," in *IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2016.

[148] R. Heindl, W. H. Rippard, S. E. Russek, M. R. Pufall and A. B. Kos., "Validity of the thermal activation model for spin-transfer torque switching in magnetic tunnel junctions.," *Journal of Applied Physics,* vol. 109, no. 7, p. 073910, 2011.

[149] Z. Li and S. Zhang, "Thermally assisted magnetization reversal in the presence of a spin-transfer torque.," *Physical Review B,* vol. 69, no. 13, p. 134416, 2004.

[150] E. Diehl, Ten Laws for Security., Springer, 2016.

[151] J. Mathew, R. S. Chakraborty, D. P. Sahoo, Y. Yang and D. K. Pradhan, "A novel memristor-based hardware security primitive.," *ACM Transactions on Embedded Computing Systems (TECS),* vol. 14, no. 3, p. 60, 2015.

[152] P. Kocher, J. Jaffe and B. Jun, "Differential power analysis," *Annual International Cryptology Conference,* p. 388–397, 1999.

[153] M.-D. Yu, R. Sowell, A. Singh, D. M'Raïhi and S. Devadas, "Performance metrics and empirical results of a puf cryptographic key generation ASIC.," *IEEE International Symposium on Hardware-Oriented Security and Trust (HOST),* 2012 .

[154] L. Zhang, Z. H. Kong, C.-H. Chang, A. Cabrini and G. Torelli, "Exploiting process variations and programming sensitivity of phase change memory for reconfigurable physical unclonable functions.," *IEEE Transactions on Information Forensics and Security,* vol. 9, no. 6, p. 921–932, 2014.

[155] L. Gao, P.-Y. Chen, R. Liu and S. Yu, "Physical unclon- able function exploiting sneak paths in resistive cross-point array.," *IEEE Transactions on Electron Devices,* vol. 63, no. 8, p. 3109–3115, 2016.

[156] U. Rührmair, J. Sölter, F. Sehnke, X. Xu, A. Mahmoud, V. Stoyanova, G. Dror, J. Schmidhuber, W. Burleson and S. Devadas., "PUF modeling attacks on simulated and silicon data.," *IEEE Transactions on Information Forensics and Security,* vol. 8, no. 11, p. 1976–1891, 2013.

[157] A. Vijayakumar and S. Kundu, "A novel modeling attack resistant puf design based on non-linear voltage transfer characteristics," *In Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition,* p. 653–658, 2015.

[158] A. Chen, "Utilizing the variability of resistive random access memory to implement reconfigurable physical unclonable functions.," *IEEE Electron Device Letters,* vol. 36, no. 2, p. 138–140, 2015.

[159] J. Zhou, K.-H. Kim and W. Lu, "Crossbar RRAM arrays: Selector device requirements during read operation.," *IEEE Transactions on Electron Devices,* vol. 61, no. 5, p. 1369–1376, 2014.

[160] E. Linn, R. Rosezin, C. Kügeler and R. Waser, "Complementary resistive switches for passive nanocrossbar memories.," *Nature Materials,* vol. 9, no. 5, pp. (403-6., 2010.

[161] M. A. Zidan, H. A. H. Fahmy, M. M. Hussain and K. N. Salama, "Memristor-based memory: The sneak paths problem and solutions," *Microelectronics Journal,* vol. 44, p. 176–183, 2013.

[162] E. Piccinini, R. Brunetti and M. Rudan, "Self-heating phase-change memory-array demonstrator for true random number generation," *IEEE Transactions on Electron Devices,* vol. 64, no. 5, pp. 2185-2192, 2017.

[163] E. Piccinini, M. Rudan and R. Brunetti, "Implementing physical unclonable functions using PCM arrays.," *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD),* pp. 269-272, 2017.

[164] A. Alaghi, W. Qian and J. P. Hayes, "The Promise and Challenge of Stochastic Computing," *IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems,* vol. 37, no. 8, pp. 1515-1531, 2018.

[165] B. R. Gaines, "Stochastic computing," *Proc. AFIPS Spring Joint Comput. Conf.,,* p. 149–156 , 1967.

[166] B. R. Gaines, "Stochastic computing systems," *Advances in Information Systems Science,* Vols. 2, Ed. Boston, MA, USA: Springer-Verlag, p. 37–172, 1969.

[167] W. J. Poppelbaum, C. Afuso and J. W. Esch, "Stochastic computing elements and systems," in *Proc. AFIPS Fall Joint Comput. Conf.*, Anaheim, CA, USA, 1967.

[168] A. W. Burks, H. H. Goldstine and J. V. Neumann, Preliminary Discussion of the Logical Design of an Electronic Computer Instrument, Princeton, NJ, USA: Inst. Adv. Study, 1946.

[169] W. Feller, An introduction to probability theory and its applications, John Wiley & Sons, 2008.

[170] J. M. Rabaey, A. P. Chandrakasan and B. Nikolic, Digital integrated circuits., Englewood Cliffs: Prentice hall, 2002.

[171] S. Gaba, P. Knag, Z. Zhang and W. Lu, "Memristive devices for stochastic computing," *2014 IEEE International Symposium on Circuits and Systems (ISCAS),* p. 2592–2595, 2014.

[172] Y. Shim, S. Chen, A. Sengupta and K. Roy, "Stochastic spin-orbit torque devices as elements for bayesian inference," *Scientific reports,* vol. 7, no. 1, p. 14101, 2017.

[173] F. V. Jensen, An introduction to Bayesian networks, London, UK: UCL press, vol. 210, 1996.

[174] D. Heckerman, D. Geiger and D. M. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical data.," *Mach. learning,* vol. 20, p. 197–243, 1995.

[175] T. D. Nielsen and F. V. Jensen, Bayesian networks and decision graphs, Springer Science & Business Media, 2009.

[176] M. Lin, I. Lebedev and J. Wawrzynek, "High-throughput bayesian computing machine with reconfigurable hardware.," *Proceedings of the 18th annual ACM/SIGDA International Symposium on Field Programmable Gate Arrays,* p. 73–82, 2010.

[177] V. K. Mansinghka, E. M. Jonas and J. B. Tenenbaum, "Stochastic digital circuits for probabilistic inference.," *Massachussets Inst. Technol. Tech. Rep. MITCSAIL-TR,* p. 2069, 2008.

[178] P. Mroszczyk and P. Dudek, "The accuracy and scalability of continuous-time bayesian inference in analogue CMOS circuits.," *IEEE International Symposium on Circuits and Systems (ISCAS),* p. 1576–1579, 2014.

[179] A. F. Murray, "Pulse arithmetic in VLSI neural networks," *IEEE MICRO,* vol. 9, p. 64–74, 1989.

[180] C. Thakur, S. Afshar, R. Wang, T. Hamilton, J. Tapson and A. van Schaik, "Bayesian estimation and inference using stochastic electronics.," *Frontiers Neuroscience,* vol. 10, 2016.

[181] C. Mead, "Neuromorphic electronic systems," *Proceedings of the IEEE,* vol. 78, no. 10, pp. 1629-1636, 1990.

[182] A. Hodgkin and A. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *The Journal of Physiology. ,* vol. 117, no. 4, p. 500–44, 1952.

[183] W. Gerstner, W. M. Kistler, R. Naud and L. Paninski, Neuronal Dynamics, Cambridge Univ. Press, 2014.

[184] G. Indiveri, E. Chicca and R. Douglas, "A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity.," *IEEE Trans. Neural Networks,* vol. 17, p. 211–221, 2006.

[185] G. Indiveri, B. Linares-Barranco, T. Hamilton, A. Schaik, R. Etienne-Cummings, T. Delbruck, S.-C. Liu, P. Dudek, P. Häfliger, S. Renaud, J. Schemmel, G. Cauwenberghs, J. Arthur, Hynna, Folowosele, Saighi, Serrano-Gotarredona, J. Wijekoon, Wang and Boahen, "Neuromorphic silicon neuron circuits," *Frontiers in neuroscience,* vol. 5, 2011.

[186] B. B. Averbeck, P. E. Latham and A. Pouget, "Neural correlations, population coding and computation.," *Nature Rev. Neurosci.,* vol. 7, p. 358–366, 2006.

[187] T. Tuma, A. Pantazi, M. L. Gallo, A. Sebastian and E. Eleftheriou, "Stochastic phase-change neurons," *Nature Nanotechnology,* vol. 11, pp. 693-700, 2016.

[188] S. R. Ovshinsky, "Analog neurons and neurosynaptic networks.". Patent US patent 6,999,953 B2, 2006.

[189] F. Xiong, A. D. Liao, D. Estrada and E. Pop, "Low-power switching of phase- change materials with carbon nanotube electrodes.," *Science,* vol. 332, p. 568–570, 2011.

[190] M. Suri, O. Bichler, D. Querlioz, G. Palma, E. Vianello, D. Vuillaume, C. Gamrat and B. DeSalvo, "CBRAM Devices as Binary Synapses for Low-Power Stochastic Neuromorphic Systems: Auditory (Cochlea) and Visual (Retina) Cognitive Processing Applications," in *2012 International Electron Devices Meeting*, San Francisco, CA, USA, 2012.

[191] H. B. Barlow, "Summation and inhibition in the frog's retina.," *J. Physiol.,* vol. 119, p. 69–88, 1953.

[192] A. Pasupathy and C. E. Connor, "Population coding of shape in area V4.," *Nat. Neurosci.,* vol. 5, p. 1332–1338, 2002.

[193] A. P. Georgopoulos, J. F. Kalaska, R. Caminiti and J. T. Massey, "On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex.," *J. Neurosci.,* vol. 2, p. 1527–1537, 1982.

[194] A. Mizrahi, T. Hirtzlin, A. Fukushima, H. Kubota, S. Yuasa, J. Grollier and D. Querlioz, "Neural-like computing with populations of superparamagnetic basis functions," *Nature Communications,* vol. 9, p. 1533, 2018.

[195] D. Heeger, Poisson model of spike generation, vol. 5, Handout, University of Standford 5, 1-13.

[196] A. Pouget, P. Dayan and R. Zemel, "Information processing with population codes," *Nat. Rev. Neurosci.,* vol. 1, p. 125–132, 2000.

[197] L. Chua, V. Sbitnev and H. Kim, "Neurons are poised near the edge of chaos.," *Int. J. Bifurc. Chaos,* vol. 22, p. 1250098, 2012.

[198] N. Bertschinger and T. Natschläger, "Real-time computation at the edge of chaos in recurrent neural networks.," *Neural Comput.,* vol. 16, p. 1413–1436, 2004.

[199] L. Chen and K. Aihara, "Chaotic simulated annealing by a neural network model with transient chaos.," *Neural Netw.,* vol. 8, p. 915–930, 1995.

[200] J. H. Shin, Y. Jeong, M. A. Zidan, Q. Wang and W. D. Lu, "Hardware Acceleration of Simulated Annealing of Spin Glass by RRAM Crossbar Array," in *IEEE International Electron Device Meeting*, 2018.

[201] F. Cai, S. Kumar, T. V. Vaerenbergh, R. Liu, C. Li, S. Yu, Q. Xia, J. J. Yang, R. Beausoleil, W. Lu and J. P. Strachan, "Harnessing Intrinsic Noise in Memristor Hopfield Neural Networks for Combinatorial Optimization," [Online]. Available: https://arxiv.org/pdf/1903.11194.pdf.

[202] S. Kumar, J. P. Strachan and R. S. Williams, "Chaotic dynamics in nanoscale NbO2 Mott memristors for analogue computing.," *Nature,* vol. 548, no. 7667, p. 318, 2017.

[203] J. Hopfield and D. Tank, "Computing with neural circuits: a model," *Science,* vol. 233, no. 4764, pp. 625-633, 1986.

[204] R. Kruse, C. Borgelt, C. Braune, S. Mostaghim and M. Steinbrecher, in *Computational Intelligence: A Methodological Introduction.*, Springer, 2016, pp. Ch. 8, 131–157.

[205] D. H. Ackley, G. E. Hinton and T. J. Sejnowski, "A learning algorithm for Boltzmann machines.," *Cognitive science,* vol. 9, no. 1, pp. 147-169, 1985.

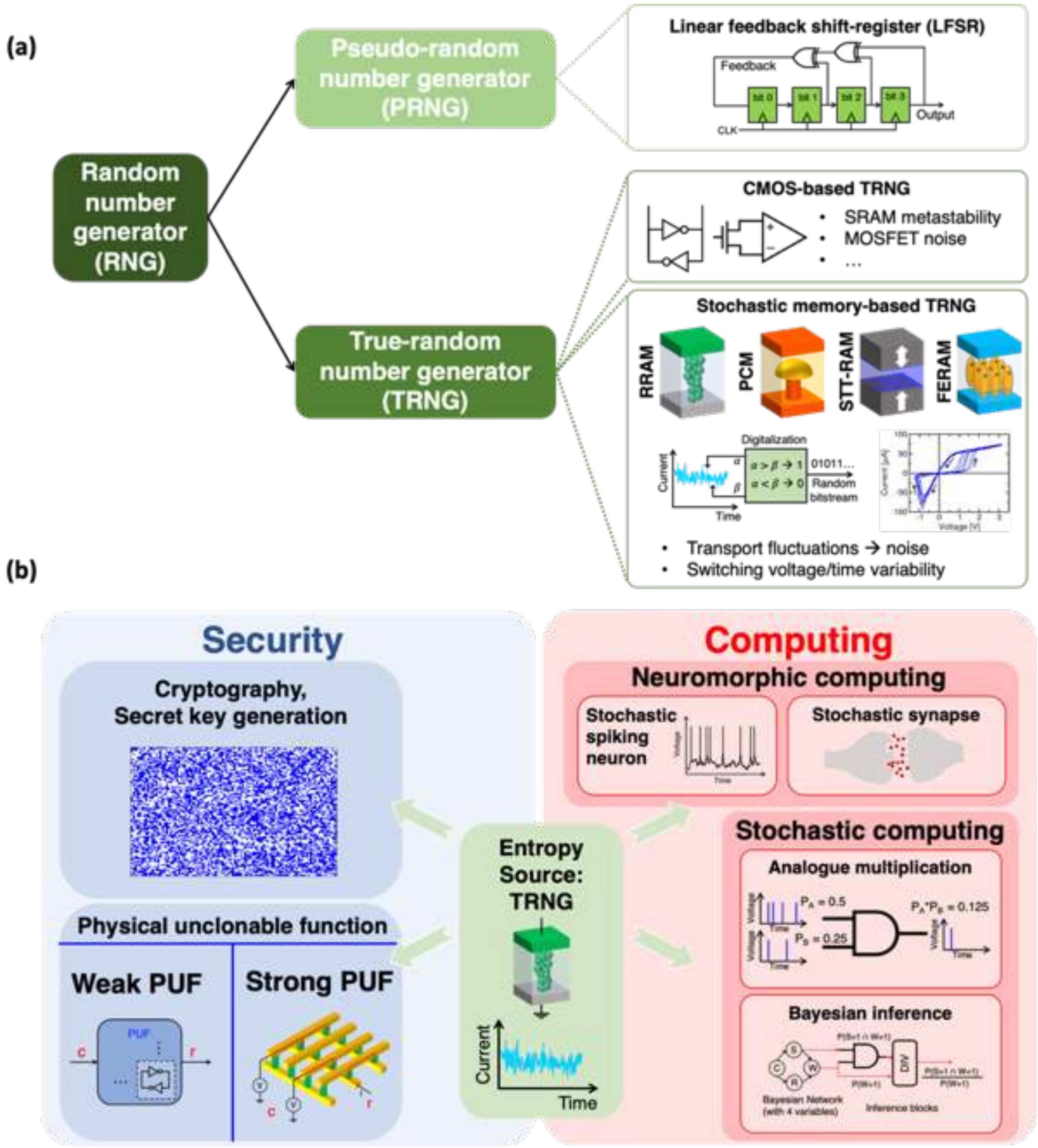


Figure 1. Current scenario of hardware primitives for security and computing. (a) The fundamental building block is the random number generator (RNG), which can be implemented either as a pseudo-RNG (PRNG), such as the linear-feedback shift register (LFSR) in the figure, or as a true-random number generator (TRNG), which harnesses the stochasticity of physical phenomena to generate a random bitstream. TRNG have been demonstrated using CMOS technology, while more recently solutions based on transport and switching fluctuations in stochastic memory devices (RRAM, PCM, STT-MRAM and FERAM) showed promising performances. (b) Stochastic memory-based TRNGs have natural implementation as on-chip entropy sources, enabling applications in both security and computing. Regarding security, the TRNG has a crucial role in data and hardware security by generating secret keys and providing the basis for the physical unclonable function (PUF) concept. A PUF introduces a challenge (c) response (r) relation, where r = f (c) and f (·) is given by the physical details defining the specific PUF instance. Depending on the available number of unique challenge response pairs (CRPs), a PUF is defined as weak, such as the

SRAM-based concept, or strong, such as the concept based on sneak paths in RRAM crosspoint. Also, TRNG-based entropy sources are pivotal for emerging computing paradigms relying on stochastic primitives, such as neuromorphic and stochastic computing. Both stochastic spiking neurons and stochastic synapses were demonstrated thanks to the stochastic behavior of memristive devices. Stochastic neurons were also applied to the simple analogue multiplication, consisting in an AND gate elaborating stochastic spiking signals. In this computing paradigm, spiking signals encode analogue number from 0 to 1 as the spiking probability.

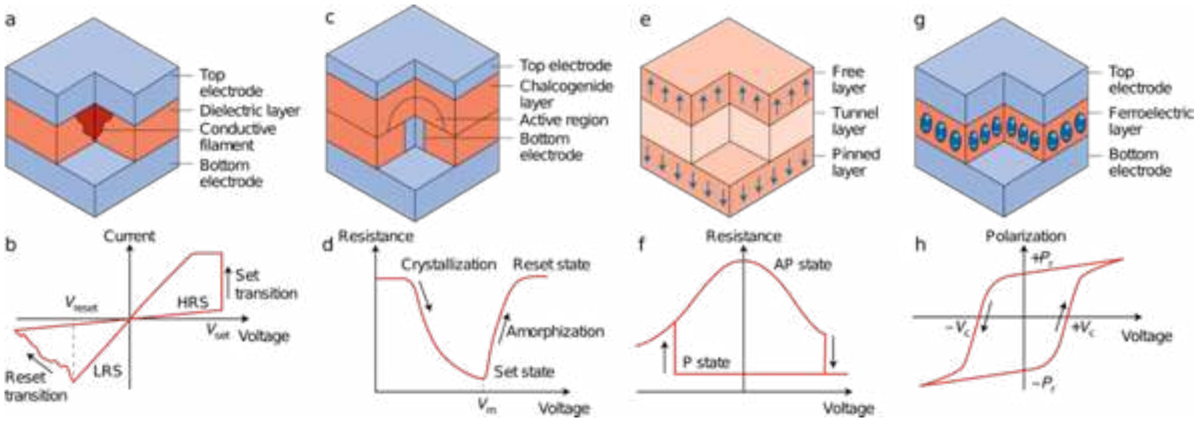


Figure 2. Resistive switching memory (RRAM) structure **(a)** and corresponding current-voltage (I-V) characteristic **(b)** for a bipolar switching RRAM. The set transition from the high resistance state (HRS) to the low resistance state (LRS) happens at a positive voltage V_{set}, corresponding to the formation of a conductive filament (CF) from the top electrode to the bottom electrode, thus shunting the dielectric layer. Reset transition, from LRS to HRS, happens at a negative voltage V_{reset}, marking the disconnection of the CF.
Phase-change memory (PCM) structure **(c)** and resistance change characteristic **(d)**. A positive voltage pulse is applied to the PCM device with the chalcogenide active material in its amorphous state, inducing the growth of the crystalline volume, hence a resistance reduction. A voltage higher than the melting-point voltage (V_m) induces the growth of the amorphous volume, hence the increase of resistance.
Magnetic tunnel junction (MTJ) metal-insulator-metal structure **(e)** and corresponding resistance-voltage (R-V) characteristics **(f)** of a spin-transfer torque magnetic memory (STT-MRAM). Positive pulses higher than a critical voltage cause set transition from AP to P state, *i.e.* from high to low resistance. The opposite reset transition, from P to AP, is attained by applying a negative voltage pulse.
Ferroelectric random access memory (FERAM) structure **(g)** and polarization-voltage hysteretic characteristics **(h)**. In this structure, the orientation of the electric dipoles in the ferroelectric layer causes a permanent polarization. The application of a voltage higher than the coercive voltage V_c leads to dipole reorientation.
Reproduced with permission. [13] Copyright 2018, Macmillan Publishers Limited.

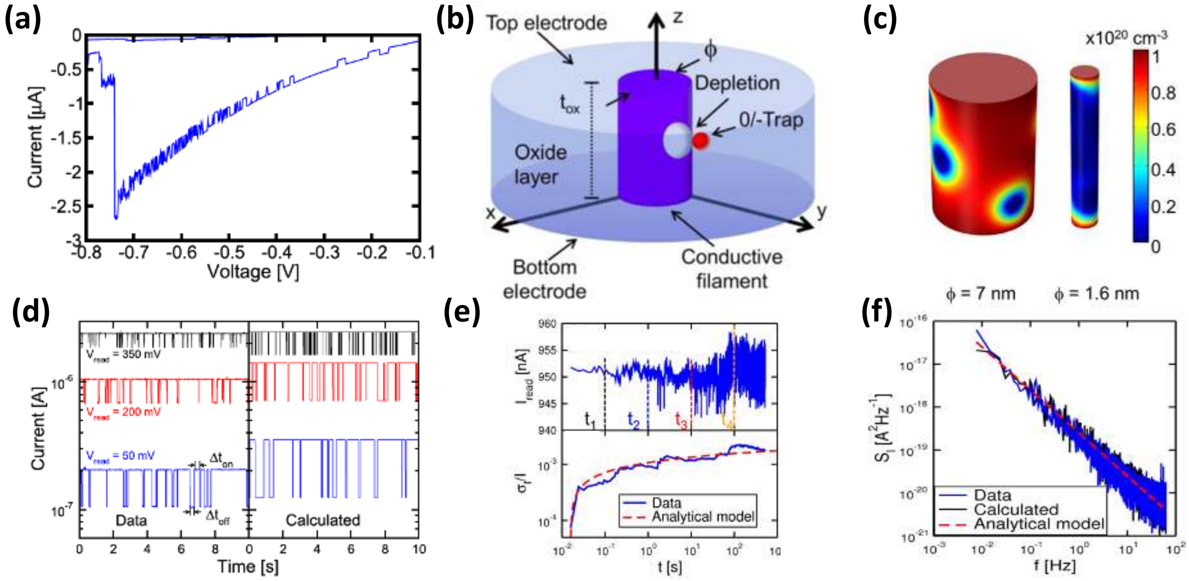


Figure 3. (a) Measured I-V characteristics of an RRAM in LRS for negative voltage showing RTN. Schematic representation of the conductive filament (CF) of a RRAM device, showing the carrier depletion induced by a negatively charged defect site following an electron trapping event (b). Electrostatic simulation of carrier density within a CF, demonstrating the increased impact of charge state fluctuations for reduced diameter filament (c). (d) Measured read current as a function of time for read voltage V_{read} = 50, 200 and 350 mV and corresponding simulations. The RTN switching times Δt_{ON} and Δt_{OFF} decrease with V_{read}. (e) Measured read current as a function of time for a device in LRS with R = 10 kΩ and corresponding relative standard deviation σI/I. (f) PSD of experimental and calculated noise, showing a 1/*f* behavior. Results from the analytical model of [71] are reported in (e) and (f). Reprinted with permission from [70] [71]. Copyright (2014-2015) IEEE.

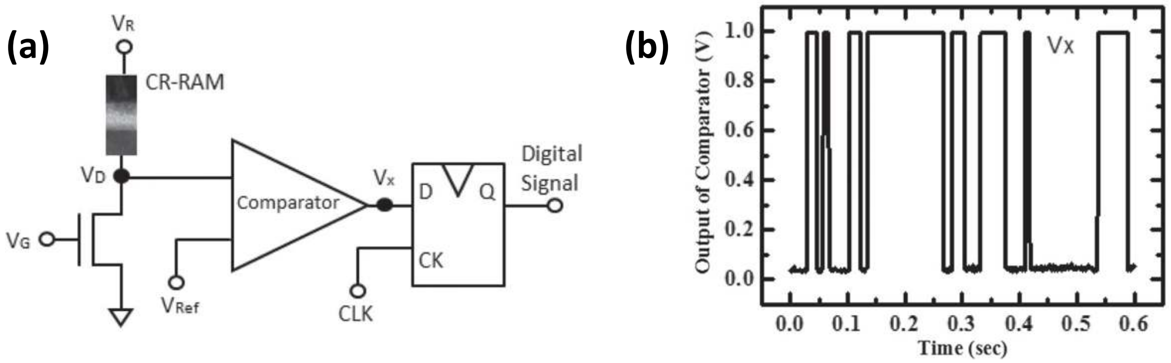


Figure 4. (a) Schematic representation of the TRNG block diagram, including CRRAM, comparator and clock control circuit. (b) Comparator output, showing binary random digital behavior. Reprinted with permission from [37]. Copyright (2012) IEEE.

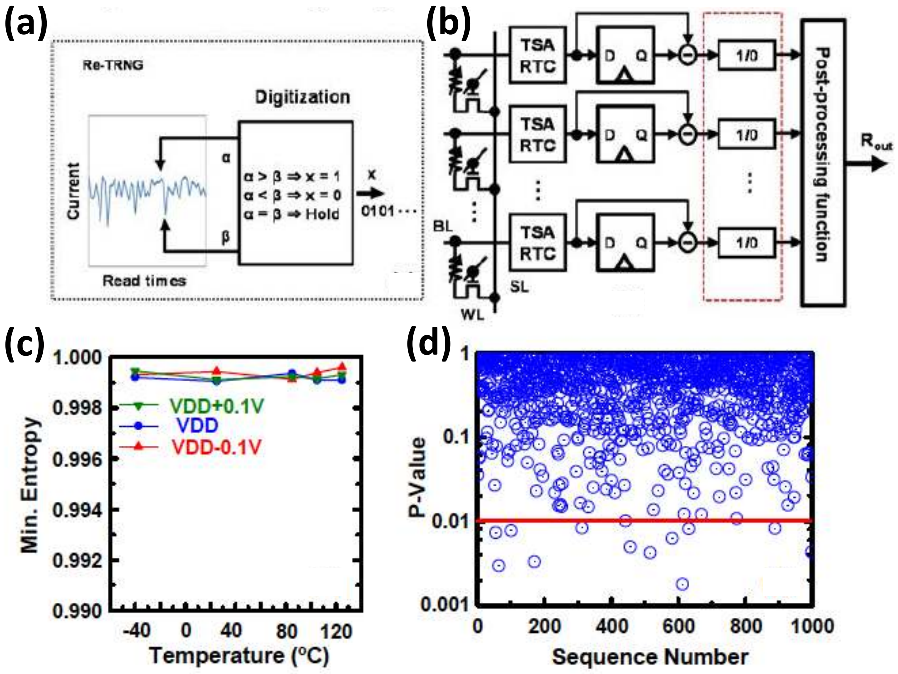


Figure 5. (a) Conceptual representation of the entropy harvesting algorithm for TRNG. (b) Block diagram of the parallel TRNG circuit, which allows for a 32 Mbps random bitstream. (c) Minimum entropies are higher than 0.999 over broad range of operative temperature and voltages. (d) P-value of 1000 sequences of 1 Mb bitstreams for the NIST frequency test. Reprinted with permission from [40]. Copyright (2016) IEEE.

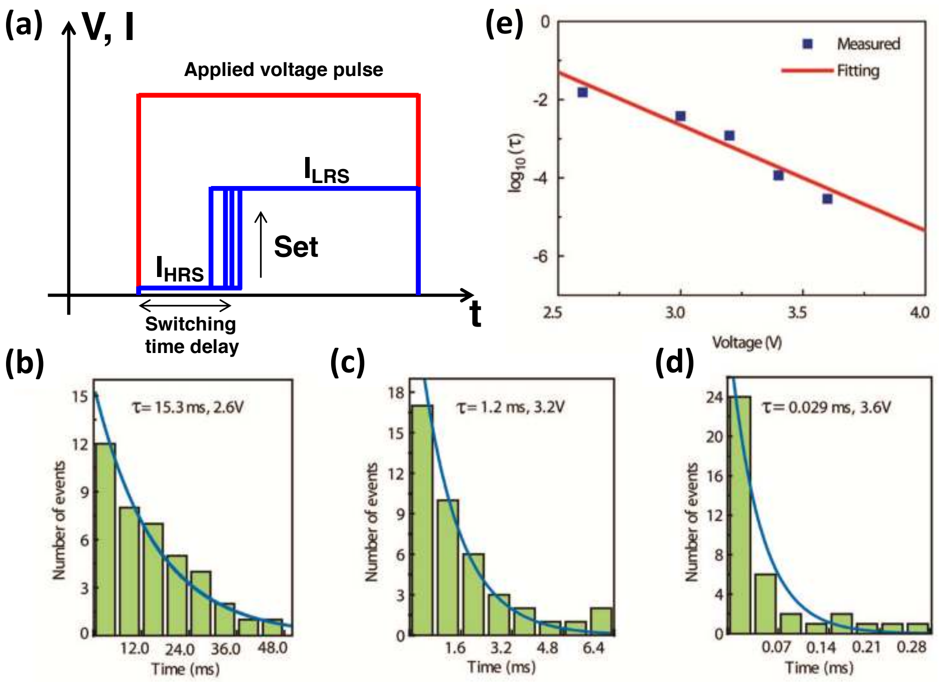


Figure 6. (a) Schematic representation of the stochastic switching events occurring in a RRAM device. The voltage pulse is applied to the device in the HRS and induces the set transition after a stochastic time delay Δt. Distributions of switching time delay for applied voltage of 2.6 V (b), 3.2 V (c), and 3.6 V (d), with their corresponding fitting with Poisson distribution. The only fitting parameter was τ = 15.3 ms, 1.2 ms and 0.029 ms for figure (b), (c) and (d), respectively. (e) shows the voltage dependence of τ, demonstrating a voltage-controlled average set transition delay Δt. Reprinted with permission from [129]. Copyright (2008) American Chemical Society.

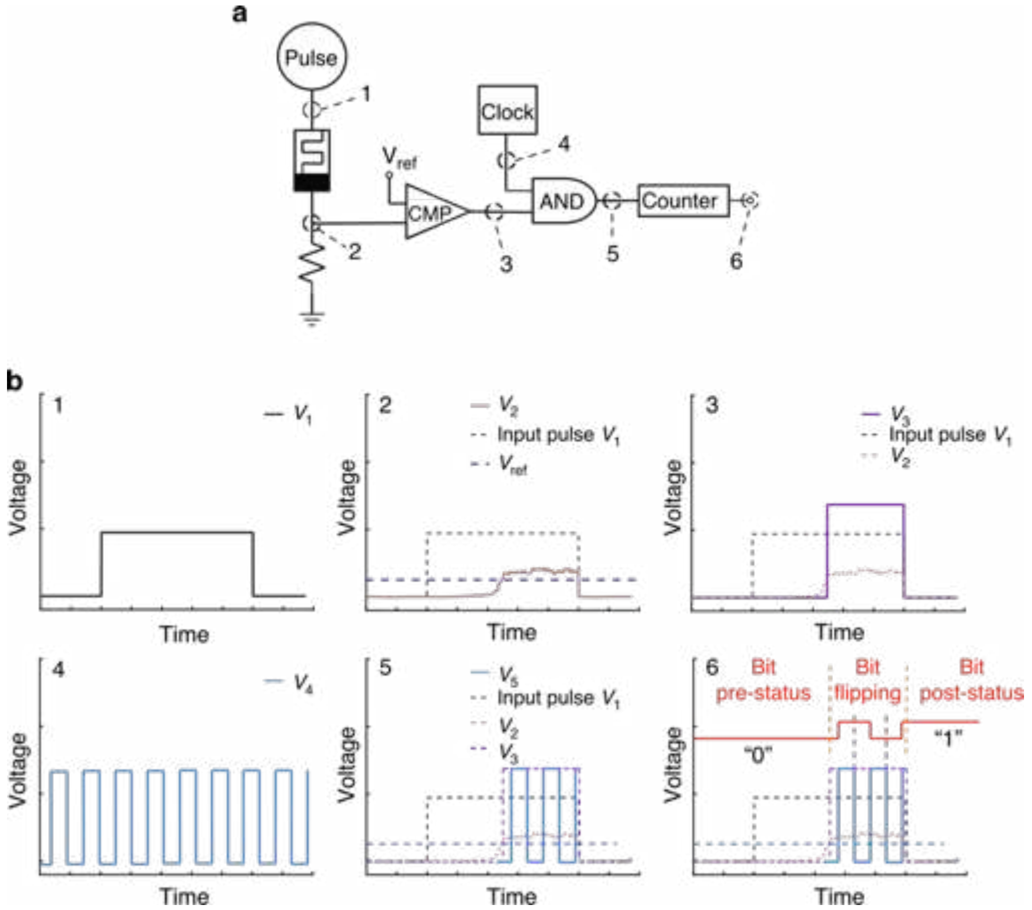


Figure 7. (a) Schematic representation of the TRNG circuit block diagram, comprising a memristive device, a comparator, an AND gate, and a counter. (b) Pulsed waveforms at each stage of the circuit, explaining the working principle of the TRNG. Reprinted from [132]. Creative Commons (2017).

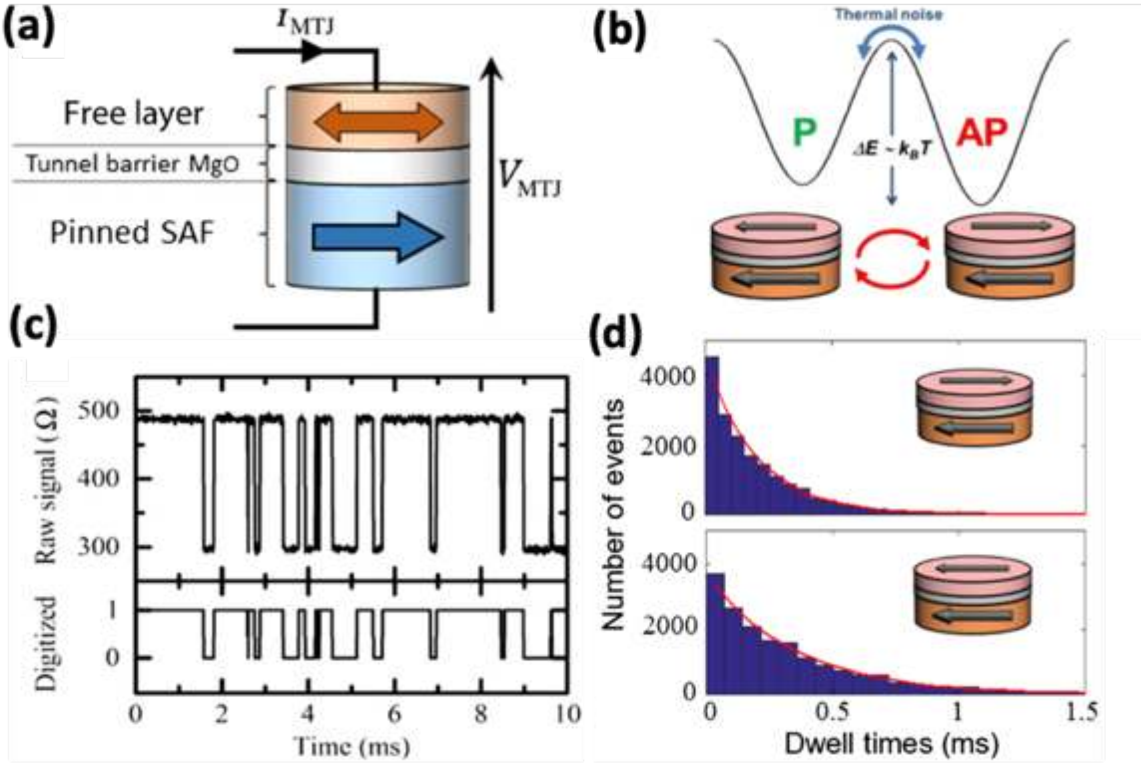


Figure 8. Device structure of the superparamagnetic tunnel junction (a) with a schematic representation of the two stable states (P and AP) separated by an energy barrier ΔE (b). Measured resistance as a function of time, evidencing random switching transitions (c) and corresponding digitized signal. Reprinted with permission from [21]. Copyright (2017) American Physical Society.

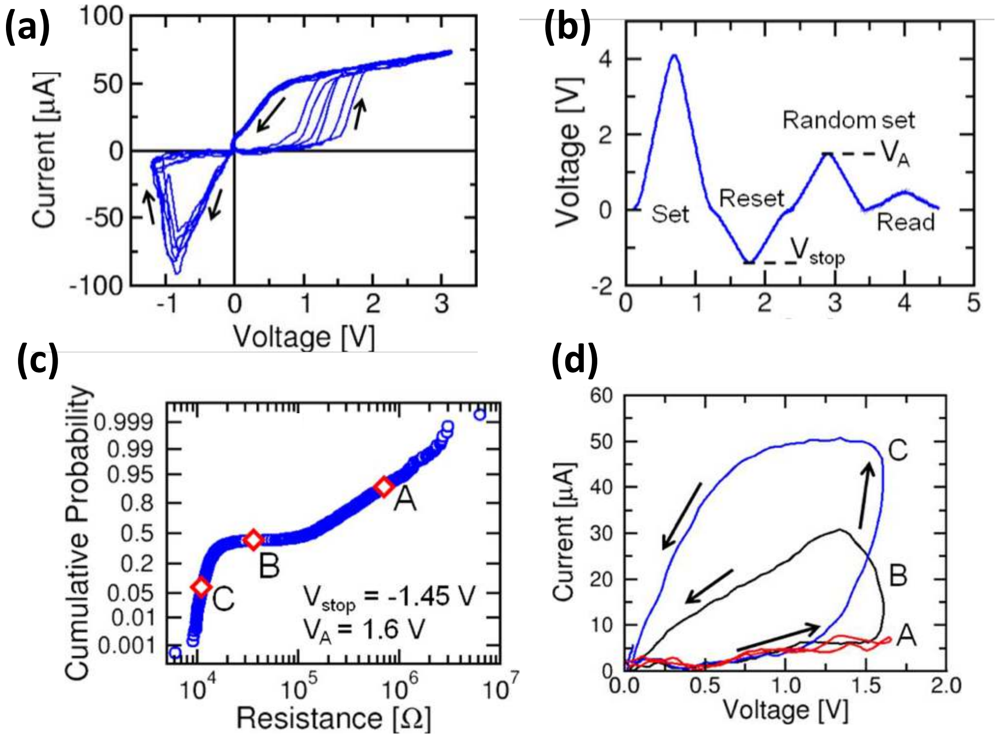


Figure 9. (a) Measured current-voltage (I-V) characteristics for six consecutive cycles on the same 1T1R structure, evidencing stochastic switching. (b) Sequence of applied pulses for TRNG, with (c) the corresponding cumulative distribution of read resistance. Random set process is highlighted in the three I-V curves (d), corresponding to states A, B and C in (c). Reprinted with permission from [41]. Copyright (2015) IEEE.

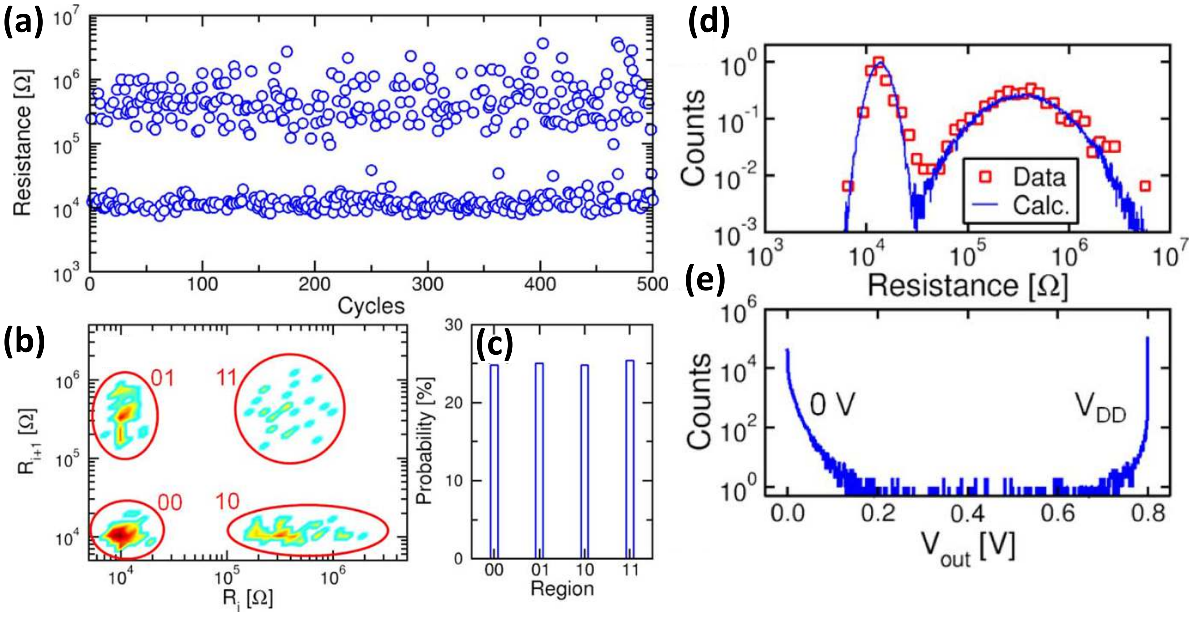


Figure 10. Measured resistance for 500 random set cycles with V_{stop} = – 1.45 V and V_A = – 1.6 V (a), correlation of R in cycle i+1 as a function of R in cycle i (b) and (c) population of the four regions in (b), demonstrating absence of memory-effect. In (d), the measured and calculated distributions of the RRAM resistance are shown. (e) Simulated distribution of the digital comparator stage output voltage V_{out}. Reprinted with permission from [41]. Copyright (2015) IEEE.

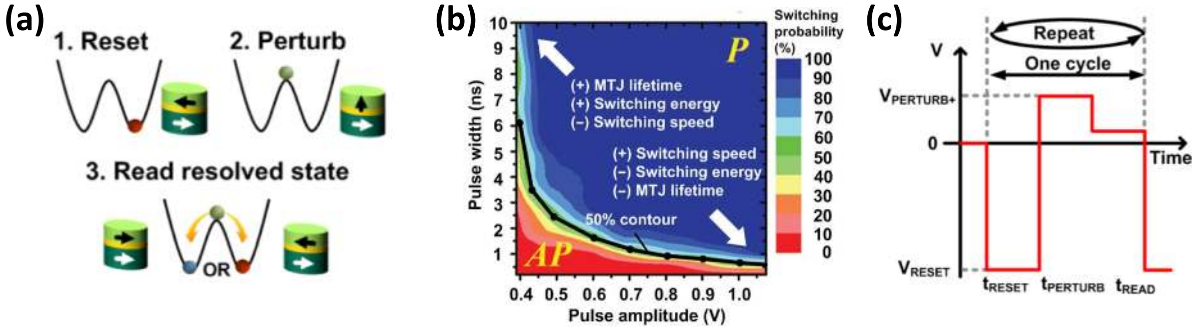


Figure 11. Schematic illustration of the reset/perturb/read scheme for TRNG (a). Spin-transfer torque-induced switching probability for AP→P transition, with blue indicating a high switching and red a low switching probability (b). Applied voltage pulses during a single TRNG cycle (c). Reprinted with permission from [143]. Copyright (2014) IEEE.

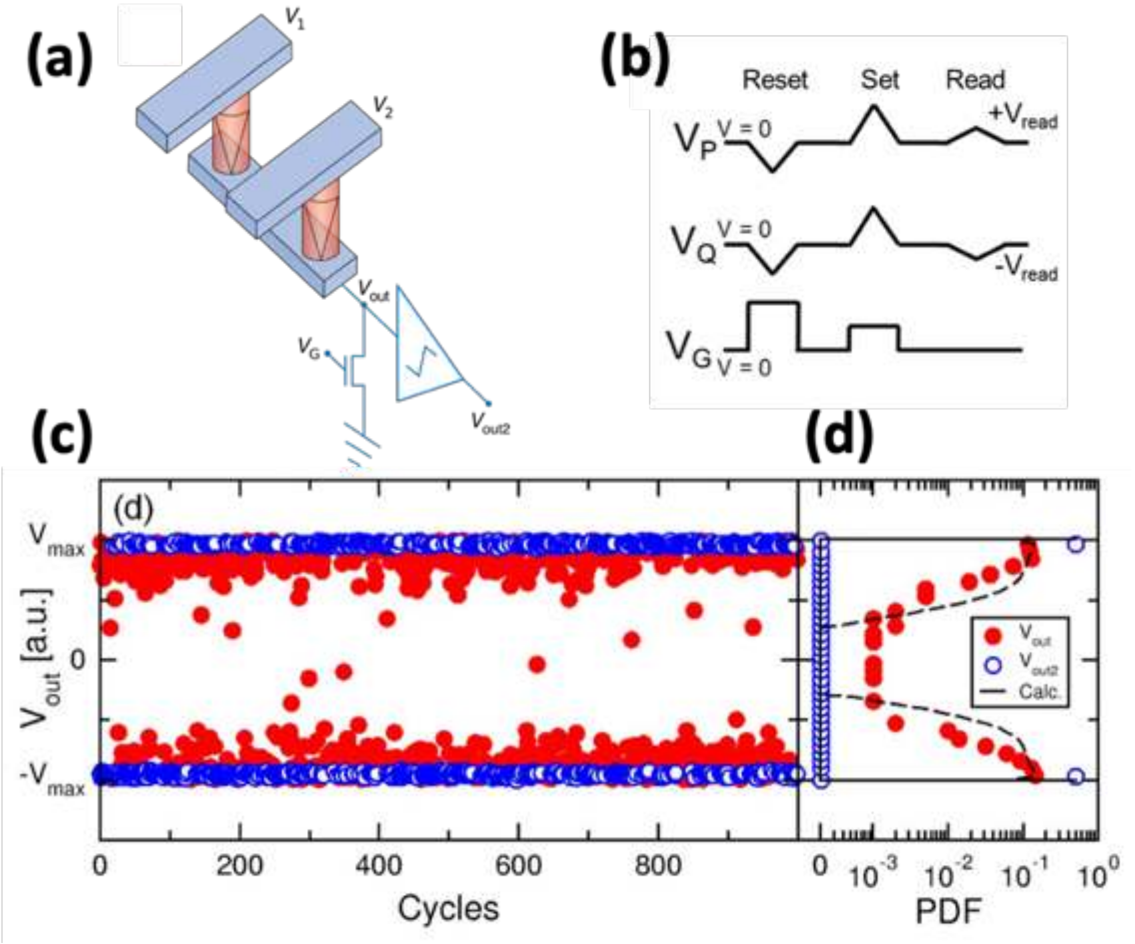


Figure 12. (a) Parallel set differential scheme and (b) sequence of applied signals. From the LRS, the cells are first independently reset, the subjected to parallel set, and finally read with voltage-divider configuration. (c) Measured V_{out} and V_{out2} during 1000 consecutive RNG cycles, and (d) corresponding PDF. Reprinted with permission from [42]. Copyright (2016) IEEE. Reproduced with permission. [13] Copyright 2018, Macmillan Publishers Limited.

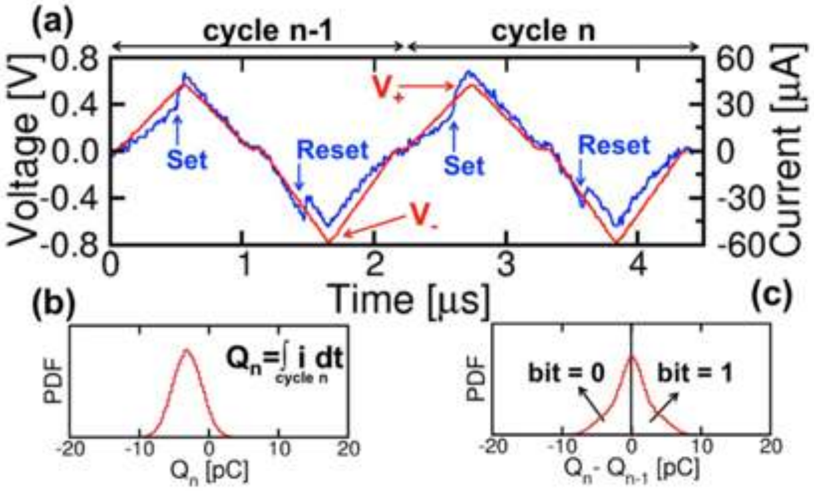


Figure 13. (a) Measured triangular voltage pulses and current response for 2 consecutive cycles n–1 and n, (b) PDF of the integrated current Q_n and (c) PDF of differential charge $\Delta Q_n = Q_n - Q_{n-1}$. The pulse sequence includes positive and negative triangular pulses for stochastic set and reset transitions, respectively, as evidenced by the abrupt steps in the current response. The random bit is assigned from the value of ΔQ_n in (c). Reprinted with permission from [139]. Copyright (2018) IEEE.

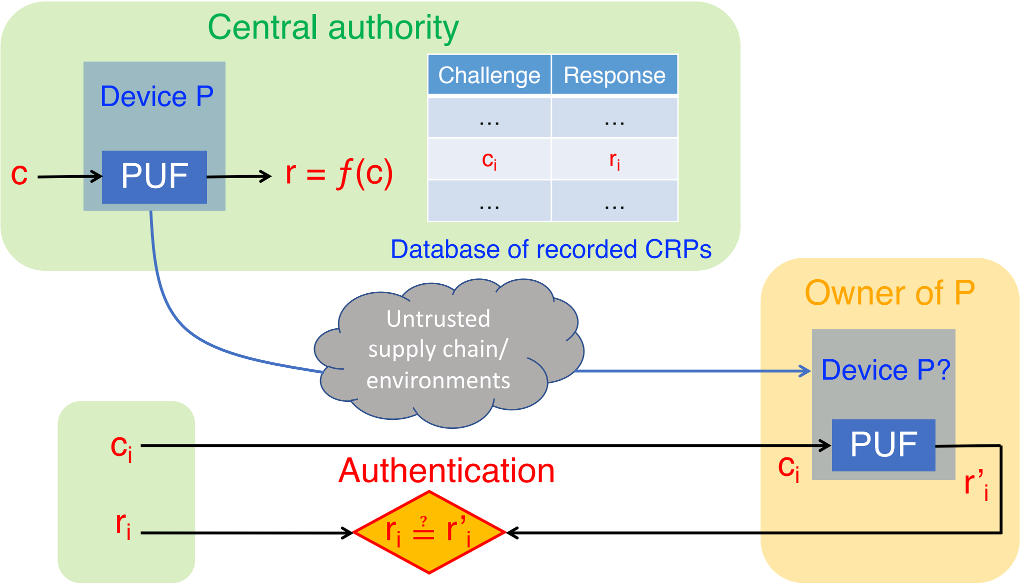


Figure 14. Physical Unclonable Function (PUF) principle of operation, involving a central authority collecting the challenge-response pairs (CRPs) of a particular PUF instance, which can be later used to authenticate that specific hardware. Adapted from [15].

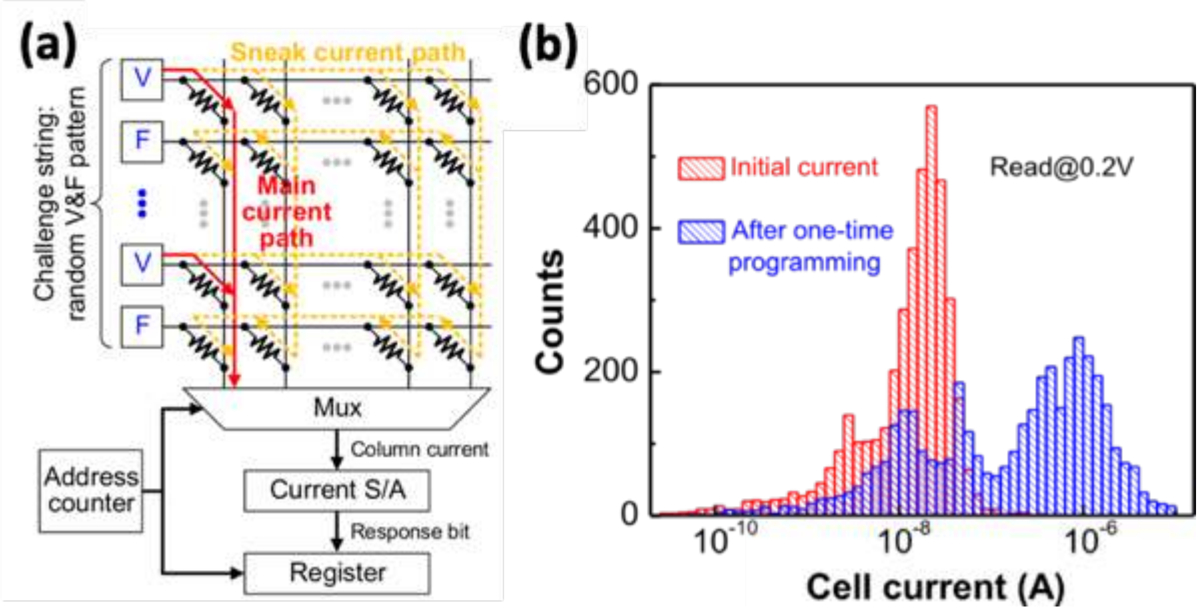
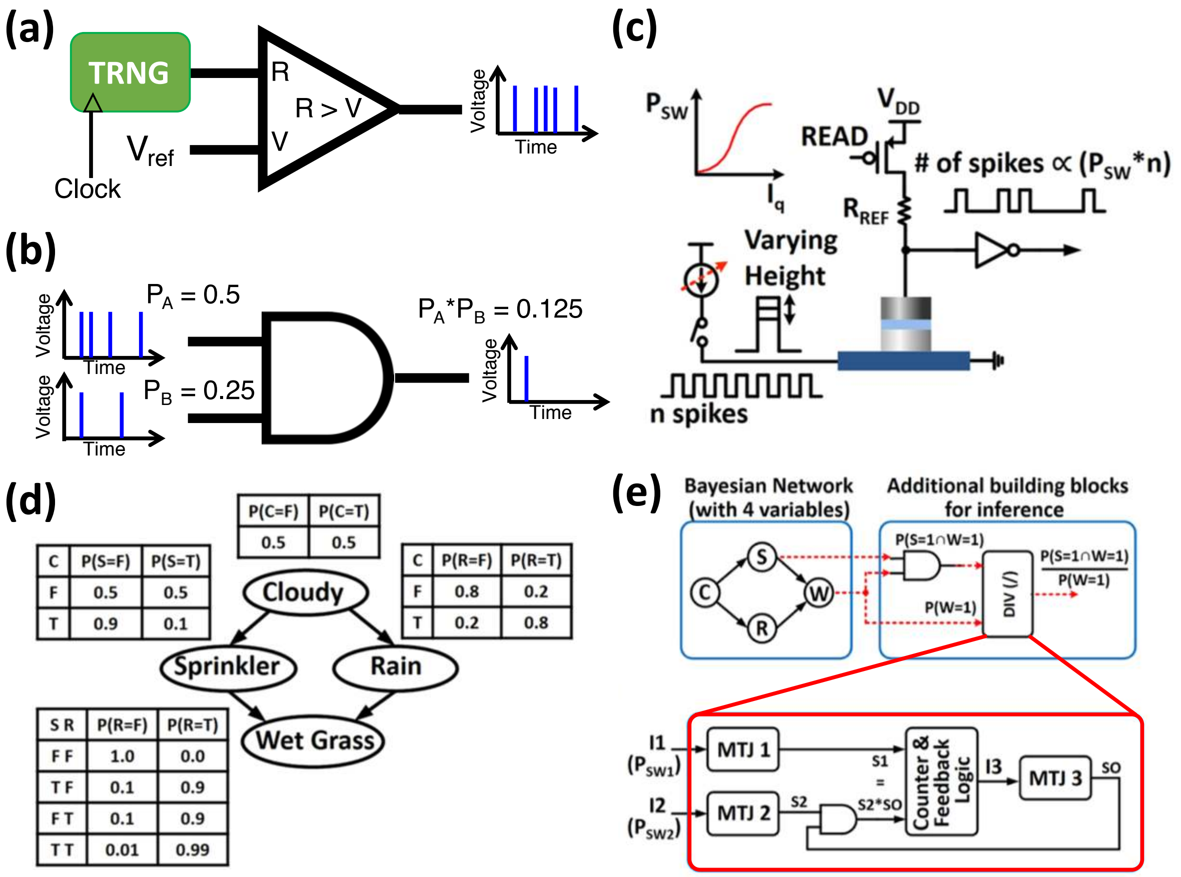


Figure 15. (a) Schematic illustration of the resistive crosspoint array used for implementing a strong PUF leveraging the exponential number of available sneak paths. (b) Distributions of cell current before and after the one-time programming, showing quite large analogue distribution. Reprinted with permission from [155]. Copyright (2016) IEEE.



P(C=F)	P(C=T)
0.5	0.5

C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1

C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

S R	P(R=F)	P(R=T)
F F	1.0	0.0
T F	0.1	0.9
F T	0.1	0.9
T T	0.01	0.99

Figure 16. Stochastic number generator (SNG) based on the comparison between a clocked multibit RNG output and a reference voltage V_{ref} at the input of an analogue comparator. The output shows the stochastic representation of the analogue values V_{ref} (a). Analogue multiplication can be easily implemented with an AND gate by adopting stochastic number representations (b). (c) Stochastic spike generator based on spin-orbit torque (SOT) magnetic tunnel junction and additional CMOS peripherals. (d) Bayesian network with 4 variables, with the corresponding Conditional Probability Table (CPT) to describe the conditional independence between variables. (e) Implementation of the inference operation through multiplication and division between Poissonian spiking signals. The particular demonstrate the internal architecture of the division block. Reprinted from [172]. Creative Commons 2017.

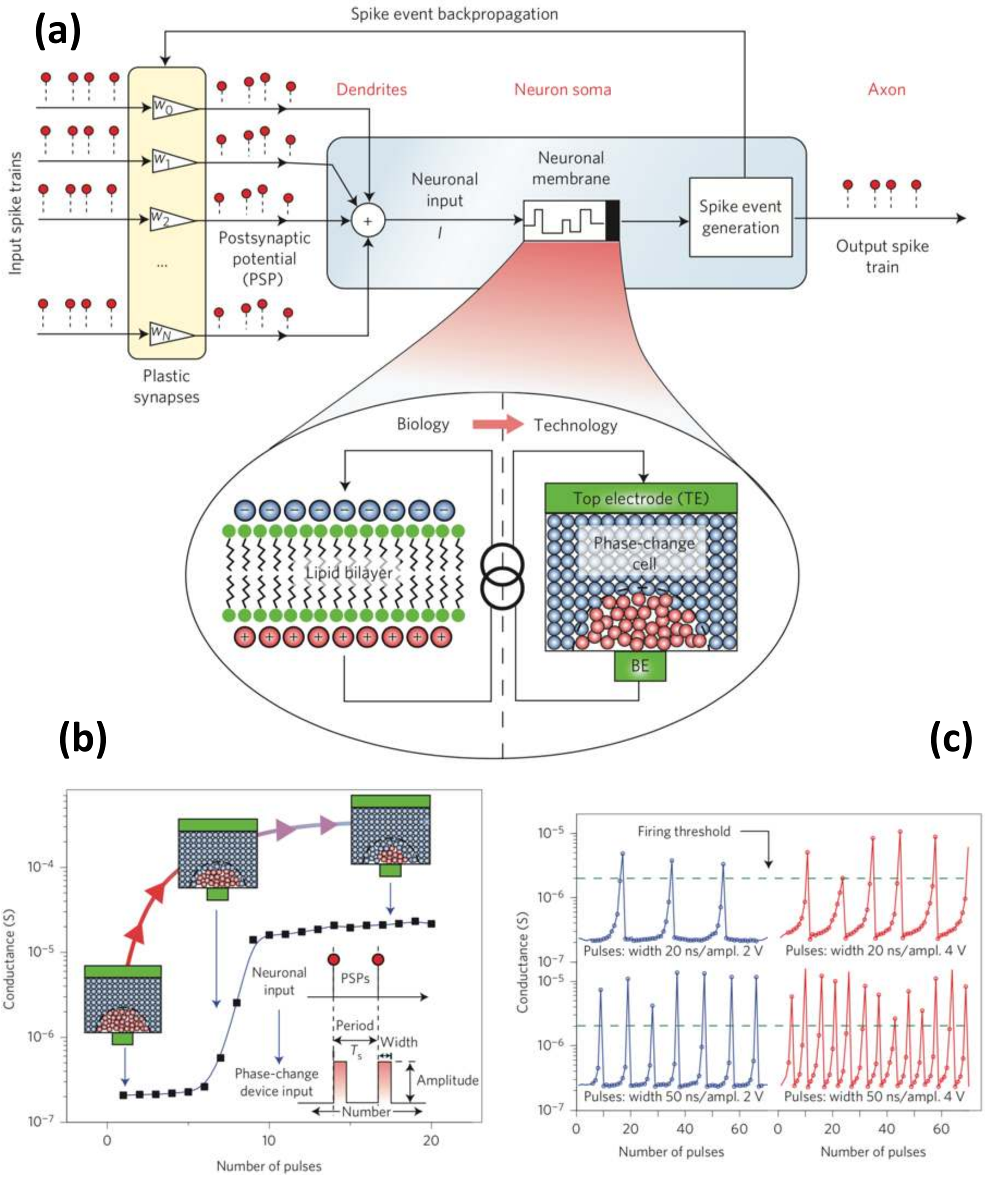


Figure 17. (a) Artificial neuron schematic consisting in the input (dendrites), the neuron soma constituted by the neuronal membrane and the spike generator, and the axon to deliver the output spike train. The dendrites are connected to plastic synapses linking the neuron to other neurons in the network. The neuronal membrane is composed by a lipid bilayer in biological systems, while it can be implemented with a phase-change memory in nanoelectronics systems. The associated neuronal membrane potential is described by means of the cumulative crystallization dynamics induced by multiple pulses (b), which allows for emulating the biological integrate and fire mechanism (c). Reproduced with permission from [187]. Copyright 2016, Macmillan Publishers Limited.

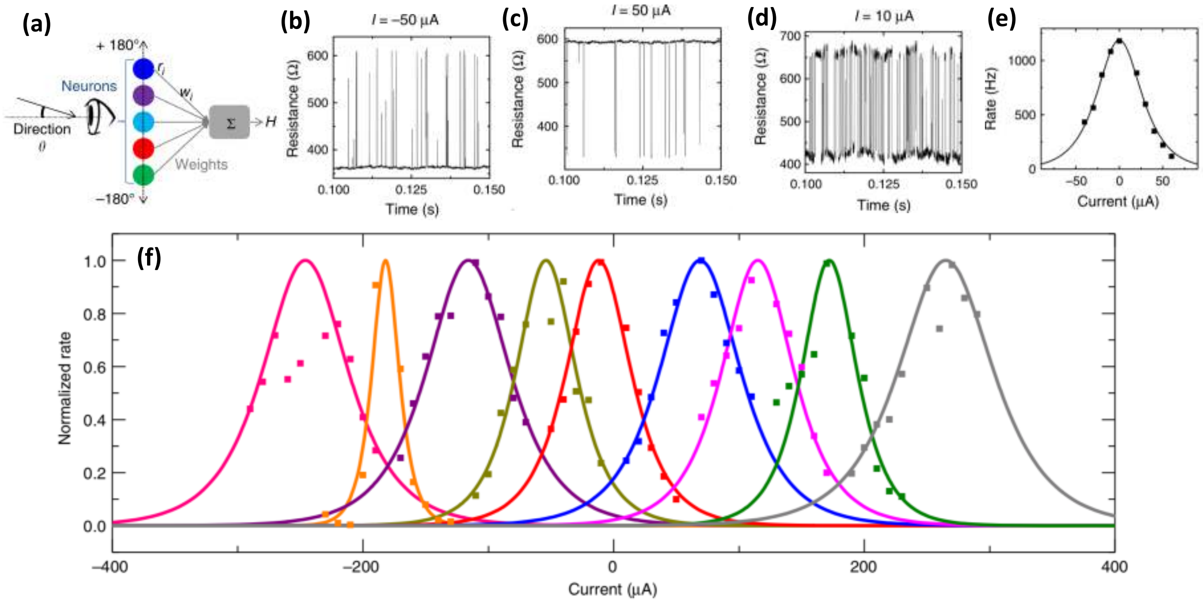


Figure 18. Schematized illustration of the information reconstruction process by means of population coding happening in the visual cortex (a). (b-d) Input current-dependent stochastic switching in superparamagnetic tunnel junction, with relatively low similar spiking frequencies for the cases –50 µA and 50 µA and relatively high spiking frequency at 10 µA input current. This rate versus current dependence is summarized in (e) for a single tunnel junction, representing a neuronal tuning curve. Measured data and calculations for nine different tuning curves associated to different junctions and constituting a basis set for population coding (f). The rate values are normalized by the corresponding natural rate r_0. Reproduced from [194] under Creative Commons license.

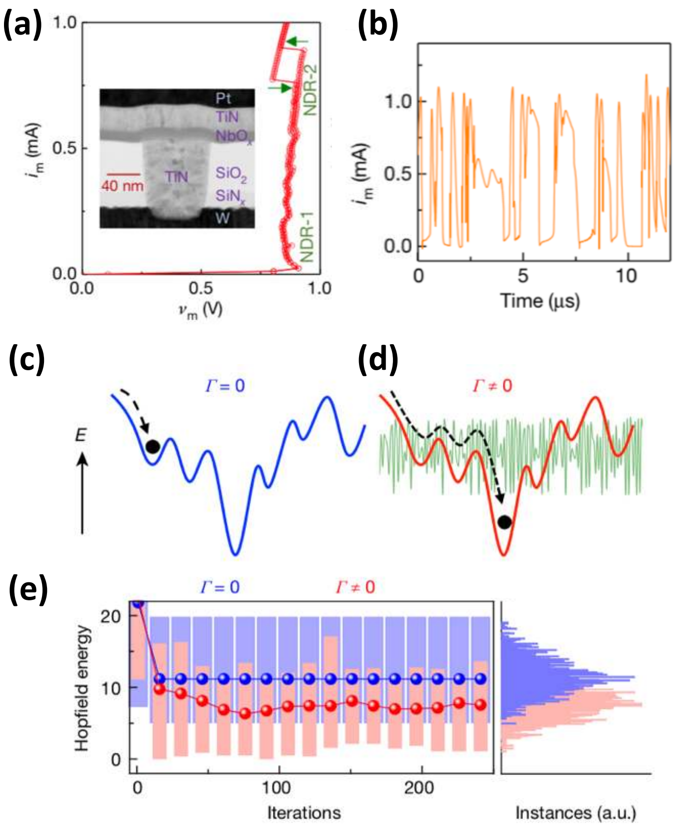


Figure 19. (a) Experimental I-V characteristics for the NbO2 Mott memristor, showing two negative differential resistance regions (NDR-1 and NDR-2). (b) Simulated chaotic oscillations of the device current under harmonic temperature perturbation. (c-d) Energy landscape associated to a Hopfield network. In the absence of zero oscillations, the final state gets stuck in a local minima (c), while in the presence of stochastic oscillations the global minima can be reached (d). (e) Evolution of the Hopfield energy as a function of the iteration number, with (red) and without (blue) chaotic fluctuations. 1000 random iterations were performed for a six-city travel salesman problem. Reproduced with permission from [202] Copyright 2017, Macmillan Publishers Limited.

Table 1. Stochastic memory devices comparison.

Technology	Resistive switching memory (RRAM)	Phase-change memory (PCM)	Spin-transfer torque magnetic RAM (STT-MRAM)	Ferroelectric memory (FERAM)
Cell size	4~12 F^2	4~30 F^2	6~50 F^2	15~34 F^2
Multibit	2	2	1	1
Voltage	< 3 V	< 3 V	< 1.5 V	< 4 V
Read time	< 10 ns	< 10 ns	< 10 ns	< 10 ns
Write time	< 10 ns	~ 50 ns	1-10 ns	~ 30 ns
Write current	10 – 100 µA	80 – 200 µA	> 50 µA	< 100 µA
Retention	10 years	10 years	10 years	10 years
Endurance	> 10^6 - 10^{12}	> 10^9	> 10^{15}	~ 10^{10}
ON/OFF ratio	10^{1-4}	10^{1-4}	TMR = 100%-200%	10^{2-3}
Write Energy	0.1 - 1 pJ/bit	10 pJ/bit	~ 0.1 pJ/bit	~ 0.1 pJ/bit
Main Advantages	- Simplicity/Low cost - High density	- Maturity	- High performance - Novel mechanisms (SOT) for improved performance	- Low power - Ferroelectricity in doped HfO_x
Key Challenges	- Reliability - Cycling endurance	- Switching power	- MTJ patterning/etch - BEOL thermal budget	- Reliability

Table 2. TRNG concepts based on stochastic memory devices.

Technology [reference]	CRRAM [37]	RRAM [40]	Diffusive memristor ($Ag{:}SiO_2$) [132]	SPMTJ [21]	RRAM [41]	p-MTJ [38]	RRAM [42]	p-MTJ [139]
Entropy source	Sampling of LRS RTN	$1/f^{\beta}$ noise in LRS	Stochastic switching time	Stochastic switching time	Stochastic switching voltage	Stochastic switching voltage	2-cell differential scheme (parallel set)	1-cell differential reading scheme
Additional circuitry	Comparator and FF	Sense amplifier + RTC	Comp., AND gate, counter	Precharge -sense amplifier	CMOS inverter	Comparator	Comparator	Integrating capacitor, ADC
Area	45 μm^2	0.256 μm^2	NA	4000 μm^2	NA	NA	NA	NA
Bit-rate	50 b/s – 1 kb/s	32 Mb/s	6 kb/s – 300 kb/s	0.5 kb/s – 6 kb/s	0.2 Mb/s	0.6 Mb/s	0.16 kb/s	0.5 Mb/s
Power/ Energy	NA	0.04 nJ/bit	NA	3 pJ/bit	NA	NA	NA	NA
Benchmark	NIST: 5 tests passed	NIST: All tests passed	NIST: All tests passed	NIST: All tests passed	Output bitstream distribution	NIST: All tests passed	NIST: 11 tests passed	NIST: 12 tests passed
Probability tracking	Yes	No	Yes	No	Yes	Yes	No	No
Post-processing	No	Post-proc. function	No	Cascaded XOR	No	Cascaded XOR	Von Neumann algorithm	No

Roberto Carboni received the B.S. degree and the M.S. degree in electrical engineering from Politecnico di Milano, Milan, Italy in 2013 and 2016, respectively, where he is currently pursuing the Ph.D. in electrical engineering. His main research interests are the characterization and modeling of resistive switching memory (RRAM) and spin-transfer torque magnetic memory (STT-MRAM) for memory and computing applications.

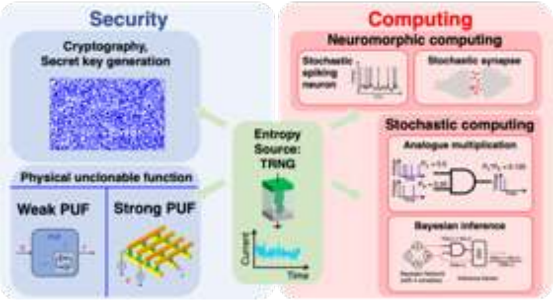
Daniele Ielmini is a Full Professor at the Dipartimento di Elettronica, Informazione, e Bioingegneria, Politecnico di Milano. He conducts research on emerging nanoelectronic devices, such as phase change memory (PCM) and resistive switching memory (RRAM), and their application in computing. He received the Intel Outstanding Researcher Award in 2013, the ERC Consolidator Grant in 2014, and the IEEE EDS Rappaport Award in 2015.

Keyword True-random number generator (TRNG)

Roberto Carboni and Daniele Ielmini* Corresponding Author*

Stochastic memory devices for security and computing

True-random number generation is the basis for a broad range of security and computing applications, such as secret key-based data cryptography, hardware authentication, and neuromorphic and stochastic computing. To deploy these innovations to IoT and mobile devices, new materials and devices with controllable entropy sources must be developed. Here, the current status and outlook of stochastic devices based on advanced materials and memory concepts (RRAM, PCM and STTRAM) are reviewed.


Security
Cryptography,
Secret key generation
Physical unclonable function
Weak PUF
Strong PUF
Entropy Source: TRNG
Current
Time
Computing
Neuromorphic computing
Stochastic spiking neuron
Stochastic synapse
Stochastic computing
Analogue multiplication
Bayesian inference