

Reliability of analog resistive switching memory for neuromorphic computing

Cite as: Appl. Phys. Rev. **7**, 011301 (2020); doi: [10.1063/1.5124915](https://doi.org/10.1063/1.5124915)

Submitted: 18 August 2019 · Accepted: 11 October 2019 ·

Published Online: 2 January 2020



Meiran Zhao, Bin Gao,^{a)} Jianshi Tang, He Qian, and Huaqiang Wu^{a)}

AFFILIATIONS

Institute of Microelectronics, Tsinghua University, Beijing, China and Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing, China

Note: This paper is part of the special collection on Brain Inspired Electronics.

^{a)}Authors to whom correspondence should be addressed: gaobl@tsinghua.edu.cn and wuhq@tsinghua.edu.cn

ABSTRACT

As artificial intelligence calls for novel energy-efficient hardware, neuromorphic computing systems based on analog resistive switching memory (RSM) devices have drawn great attention recently. Different from the well-studied binary RSMs, the analog RSMs are featured by a continuous and controllable conductance-tuning ability and thus are capable of combining analog computing and data storage at the device level. Although significant research achievements on analog RSMs have been accomplished, there have been few works demonstrating large-scale neuromorphic systems. A major bottleneck lies in the reliability issues of the analog RSM, such as endurance and retention degradation and read/write noises and disturbances. Owing to the complexity of resistive switching mechanisms, studies on the origins of reliability degradation and the corresponding optimization methodology face many challenges. In this article, aiming on the high-performance neuromorphic computing applications, we provide a comprehensive review on the status of reliability studies of analog RSMs, the reliability requirements, and evaluation criteria and outlook for future reliability research directions in this field.

© 2020 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/1.5124915>

TABLE OF CONTENTS

| | |
|--------------------------------------------------|----|
| I. INTRODUCTION | 1 |
| II. RSM FOR NEUROMORPHIC COMPUTING | 2 |
| A. Analog RSM devices | 2 |
| B. RSM-based neuromorphic computing system | 3 |
| III. RELIABILITY CONCERNs | 4 |
| A. Basic reliability metrics | 5 |
| B. Functional reliability metrics | 6 |
| IV. REVIEW OF THE RELIABILITY STUDY | 9 |
| A. Basic reliability metrics study | 9 |
| 1. Retention | 9 |
| 2. Endurance | 10 |
| 3. Write/read disturb | 12 |
| 4. Noise | 13 |
| B. Functional reliability metrics study | 13 |
| 1. Nonlinearity | 13 |
| 2. Asymmetry | 14 |
| 3. Dynamic range | 14 |
| 4. Precision | 14 |
| 5. Variation | 14 |

| | |
|------------------------------|----|
| 6. Bit yield | 14 |
| V. SUMMARY AND OUTLOOK | 15 |

I. INTRODUCTION

By mimicking the mechanism of human brains, artificial intelligence (AI) has achieved remarkable success, with applications in image and natural-language processing,^{1,2} driving automation,³ big data analysis,⁴ and even vision-based robotic object handling.⁵ The rapid development of AI applications requires continuous hardware advancement; especially, a high-speed and energy-efficient hardware system is required. However, the traditional computing system with von Neumann architecture leads to the high energy consumption and latency due to a huge amount of data transfer between the separated memory unit and the logic unit.⁶ The speed gap between the two units further results in the considerable latency, which is called the “memory wall.”^{7,8} The neuromorphic computing system has been considered as a promising candidate for breaking this predicament.⁹

The neuromorphic computing systems refer to the systems that mimicked the biological brain from the structure and/or working mechanism.¹⁰ The emerging analog-type resistive switching memory

(RSM) based in-memory architecture is one of the key technologies to implement the neuromorphic computing system. Analog RSM refers to a kind of two-terminal nonvolatile memory device with multiple conductance levels. The stored information is determined by the conductance value.¹¹ In a neuromorphic computing system, the RSM devices act as synaptic weights to store information and process input signals.¹² Based on Ohm's law and Kirchhoff's law, the RSM with a crossbar array could naturally accomplish matrix–vector multiplication (MVM) within one step by collecting the accumulative output current.¹³ In this case, high parallelism could be realized to accelerate the computations without the latency and energy consumption between memory and computing units. Compared to the traditional memory devices, such as static random access memory (SRAM)¹⁴ and Flash,¹⁵ analog RSM has significant advantages. Although SRAM technology has a fast speed and matured manufactory technology following the CMOS scaling down, the limited area-inefficient and high standby power is undesired in large-scale SRAM arrays.¹⁶ In contrast, Flash is a kind of nonvolatile memory device with analog computing ability. Neuromorphic computing chips based on Flash technology have demonstrated excellent performance when compared to the conventional CMOS technology.^{17,18} However, compared to the Flash technology, analog RSMs show higher switching speed, lower programming voltage, and higher endurance ability.^{19,20} With these excellent characteristics and much higher area efficiency, the analog RSM array can stimulate great potential in future neuromorphic computing systems.

Recently, neuromorphic computing based on analog RSM has achieved significant progress from the synaptic devices to the array-level demonstrations. Jo *et al.*¹² first proposed implementing the synaptic functions by using analog RSM devices, which pioneered the development of RSM for neuromorphic computing. Preziosi *et al.*²¹ reported an array-level implementation of *in situ* training in a neuromorphic network based on $\text{Al}_2\text{O}_3/\text{TiO}_{2-x}$ analog RSM. Furthermore, several tasks, such as handwritten recognition,²² face classification,¹³ feature extraction,²³ and reinforcement learning,²⁴ have been demonstrated based on an analog RSM array. These achievements demonstrated the functional feasibility and performance advantage of analog RSM-based neuromorphic computing systems. However, to date, experimental reports have remained at small-array-level (hundreds to thousands of cells) primitive demonstrations.^{14,25} It is still very challenging to develop a large array or full chip to execute practical AI tasks. Furthermore, the computing accuracy of analog RSM-based systems is lower than that of a CPU.²⁶ The key challenge lies in the reliability issues of the analog RSM. With large write/read noises and disturbance,^{27,28} endurance and retention degradation,²⁹ and other reliability issues, it is difficult to achieve high performance with large arrays. This situation significantly blocks further research and applications of neuromorphic computing with analog RSMs. Therefore, it is important to provide a comprehensive analysis and summary of the reliability issues of analog RSM, which is the goal of this review.

Previously, several comprehensive reviews were published and discussed on the research progress of neuromorphic computing with emerging nonvolatile memory devices. They involved recent efforts in materials and mechanisms of synaptic devices^{30,31} and memory-based experimental demonstrations with novel algorithms and circuit architectures.^{16,32} It was also accompanied by an analysis of the desired device properties.^{33,34} However, in RSM-based neuromorphic systems

for the functional demonstration and practical application, reliability becomes a key challenge, which limits the performance and accuracy in a large-scale RSM array. Therefore, it is necessary to provide a comprehensive summary and discuss on the state of the art, challenges, and prospects of the reliability problems and their impacts on neuromorphic computing. With these considerations, this review offers the dedicated perspective of device reliability and its impact on neuromorphic system performance. In detail, we will review the reliability issues and possible solutions of analog RSM devices for neuromorphic computing. In Sec. II, we explain why and how to implement the neuromorphic computing with analog RSMs. Then, various reliability concerns of neuromorphic computing applications are discussed in Sec. III, including the basic and functional reliability metrics. In Sec. IV, state-of-the-art and representative works on the reliability of the analog RSM are reviewed. The physical mechanisms of reliability degradation and optimization methods for analog RSMs are also summarized. Finally, we provide the outlook and prospects of the unresolved reliability issues that urgently need to be addressed for analog RSMs for neuromorphic computing.

II. RSM FOR NEUROMORPHIC COMPUTING

A. Analog RSM devices

Compared to traditional memory types, RSMs have shown significant advantages in implementing neuromorphic computing systems. Hardware accelerators based on traditional memories such as SRAM show limitations for computing in terms of cell density (100–200 F² per bit cell). By contrast, analog RSM, as a synaptic device, demonstrates high storage density (4–16 F² per bit cell)²⁰ and fast parallel computing ability.³⁵ Recently, some three-terminal transistorlike synaptic devices have been proposed with better conductance tuning ability.^{36,37} However, in this article, we only focus on two-terminal resistorlike analog RSMs because they show better integration density and have been well studied on the reliability aspects.

The analog RSMs typically include filamentary-type resistive random access memory (RRAM) devices [Fig. 1(a)], non-filamentary-type RRAM devices [Fig. 1(b)], and phase change memory (PCM) devices [Fig. 1(c)]. Filamentary RRAMs can be further classified into cation type, anion type, and dual ionic type. The resistance value of the filamentary RRAM depends on the formation and rupture of conductive filaments (CFs),³⁸ as shown in Fig. 1(a). The CFs are composed of interstitial metal atoms (cation type),³⁹ oxygen vacancies (anion type),⁴⁰ or both (dual ionic type).⁴¹ Resistive switching is dominated by the migration of ions. The resistance value of the nonfilamentary RRAM is determined by the interfacial Schottky/tunneling barrier modulated by the electron trapping/detrapping or ion migration,⁴² as shown in Fig. 1(b). Different from the morphology or component change in RRAM, the resistance change in the PCM is determined by the thermally induced lattice phase change in a bulk region.⁴³ In PCM, the active layer is a chalcogenide-based material, which can maintain a crystalline or amorphous state for a long time, as shown in Fig. 1(c). The crystalline state shows a lower resistance value, whereas the amorphous state demonstrates semiconductor characteristics corresponding to a higher resistance state. The reversible switching is dependent on the Joule heating causing by the voltage/current pulses in the active region. Furthermore, some charge- or spin-based memory devices also show resistive switching behaviors, such as magnetic random access memory (MRAM) devices, domain wall devices, ferroelectric devices,

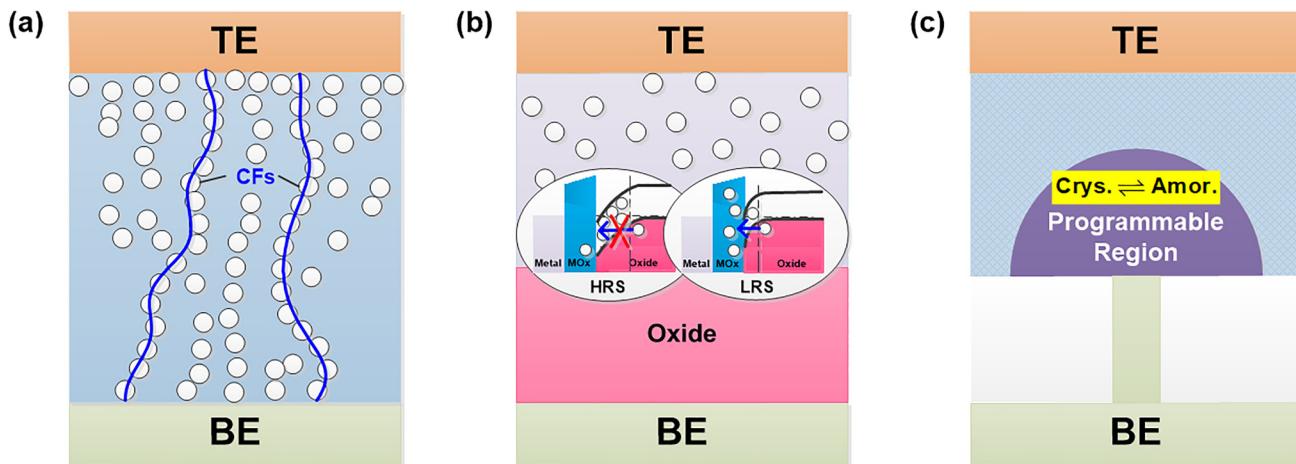


FIG. 1. Computing with the emerging analog-type RSM. (a) The structure and mechanism of filamentary RRAM. The rupture or connection of CFs represents the higher or lower resistance states, and multiple CFs contribute to the analog switching ability. (b) The structure and mechanism of nonfilamentary RRAM. The two insets illustrate the band diagrams of the interface in HRS (left) and LRS (right). (c) The structure and mechanism of PCM. The phase of the programmable region switches between the crystalline and amorphous states corresponding to the resistive switching between LRS and HRS, respectively.

and charge-trapping devices.^{44,45} However, these types of devices still need more studies to realize both analog-type resistive switching ability and long-term retention simultaneously.

To tune the conductance of analog RSM devices, an external voltage pulse is applied. If the device conductance increases with an applied pulse, we call this process “SET,” “weight increase,” or “potentiation.” Meanwhile, if a pulse causes a conductance decrease, we call this process “RESET,” “weight decrease,” or “depression.” Some of the RSMs are bipolar, which means that SET and RESET pulses should have different voltage polarities, and the others are unipolar, which means that SET and RESET are independent with voltage polarity. Most RSMs based on the ion-migration mechanism are bipolar. For analog RSMs, the lowest and highest resistance states are called LRS and HRS, respectively, and the other medium resistance states are all called MRS. Sometimes, when the device is switching between two MRSs, we call the pair a lower medium resistance state (L-MRS) and a higher medium resistance state (H-MRS).

B. RSM-based neuromorphic computing system

There are two approaches for implementing neuromorphic computing based on RSM with different information-encoding schemes. One is the deep neural network (DNN), which pursues high computation efficiency on data-intensive tasks. The other is the spiking neural network (SNN), which pursues excellent power efficiency by mimicking the biological neural network in the human brain with the neuron values encoded by spiking timing.⁴⁶ Analog RSM has been exploited for DNN chips, such as Umass’s 128×64 reconfigurable 1T1R memristor crossbars,⁴⁷ UCSB’s transistor-free 12×12 memristor crossbars,²¹ and Umich’s full integrated hardware system on a 54×108 reprogrammable memristor chip,⁴⁸ while the experimental demonstration of SNN based on the analog RSM chip remains to be studied. Because the DNNs are the mainstream neural networks and have demonstrated much better performance and accuracy than SNN,⁴⁹ this review will focus on the DNN with analog RSMs.

The processing of a neural network includes two phases: inference and training. Inference is a feedforward computing process by summing the weighted inputs from the prelayer neurons and generating output signals to the postlayer neurons, as shown in Fig. 2(a). The inference in multiple layers is carried out layer by layer sequentially. In an analog RSM array, the conductance (G_{ij}) of each RSM device acts as an analog weight (W_{ij}). As shown in Fig. 2(b), the current of each RSM (I_{ij}) is the product of G_{ij} and the biased voltage V_j based on Ohm’s law, whereas the total current I_i in one column is accumulated as the sum of the device current based on Kirchhoff’s current law.⁵⁰ For the hidden layer, an activation function unit is required to transform the output current $I_{h(m)}$ of the previous layer into the input voltage $V_{h(m)}$ of the subsequent layer. Therefore, the complex MVM can be naturally implemented by inputting the voltage from the word lines (WLs) and outputting the current through the bit lines (BLs), as shown in Fig. 2(b). In this case, significant amounts of energy and time are saved with the natural parallel operation without data transfer between memory and computing cells in the von Neumann architecture.^{12,51}

Training is critical for neuromorphic systems by the feedback algorithm, as shown in Fig. 2(a). This algorithm is used to update the weights in parallel according to the learning rules.⁵² There are two training methods for a neuromorphic system: *ex situ* and *in situ* training. For the *ex situ* method, training is executed in the software system first, and then the calculated weights are loaded to the analog RSM array.⁵³ The weights stored in the RSMs are not adjusted during the weight-loading process, regardless of the existence of variation or other undesired weight changes. It is important to maintain the conductance of the RSM unchanged after weight loading. Therefore, the requirements of retention, bit yield, and uniformity are very strict for analog RSMs under *ex situ* training.

By contrast, a neuromorphic computing system with *in situ* training is capable of updating weights on chip and thus has better immunity to retention degradation, state-stuck issues, and variations than *ex situ* training.⁵⁴ Before *in situ* training, the weights stored in the

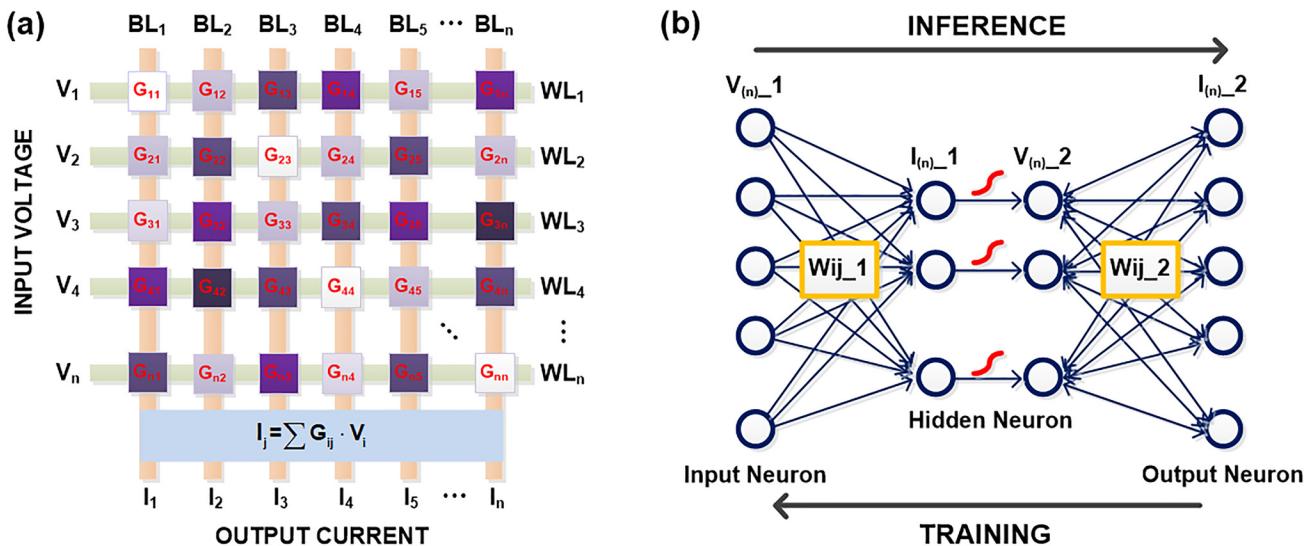


FIG. 2. (a) Schematic diagram of a two-layer neural network. Each neuron computes a weighted sum of its inputs and applies a nonlinear activation function. (b) The schematic of an analog RSM crossbar array implementation of the most critical part of the perceptron, the weighted sum, where the conductance of analog RSM acts as the synaptic weight in a neuromorphic network.

analog RSMs can be either random or started from the values that are calculated and loaded based on the preliminary *ex situ* training method.⁵⁵ The goal of *in situ* training is to maximize the inference accuracy by tuning the device conductance in the analog RSM arrays. During one training iteration, before conductance tuning, a preinference process is required to obtain the errors between the expected results and the calculated results. Then, the desired weight update value can be calculated. There are several ways to tune the conductance based on the calculated errors. The most general way is to use the conventional backpropagation (BP) learning rule. Using this method, the desired weight update values should be calculated exactly and mapped to analog RSM devices. A verification programming scheme should be introduced to ensure that the conductance of each analog RSM is tuned to the projected value.¹³ A simplified method is to use the sign-based BP (SBP) learning rule.⁵⁶ In this case, the exact weight update values are not necessary, and only the signs of weight update values are needed. If the sign is positive, a SET pulse is applied to the corresponding analog RSM device. By contrast, if the sign is negative, a RESET pulse is applied. The SBP method can largely reduce the overhead of periphery circuits and verification time and can be done in a parallel way rather than via one-by-one verification.⁵⁷ In addition, the training accuracy is strongly dependent on the characteristics of analog resistive switching behaviors, such as nonlinearity and asymmetry, which will be discussed in Sec. III. Recently, some novel learning rules for the trade-off between BP and SBP were also proposed.^{57,58} Furthermore, it should be noticed that a complete *in situ* training process contains many iterations, and thus, the conductance of each analog RSM is tuned thousands to millions of times according to the learning rule and learning task.^{51,59} Therefore, endurance becomes one of the most important concerns for analog RSMs under *in situ* training.

From the above discussion, it can be found that reliability requirements are largely dependent on the scenarios (*ex situ* training

or *in situ* training) and applications (for storage or for computing).⁶⁰ In Sec. III, we will summarize and define the main reliability concerns.

III. RELIABILITY CONCERN

Notably, there has been no universal evaluation criterion for the reliability of analog RSM devices. However, computing accuracy could be considered as a critical parameter to quantify the reliability metrics. The accuracy loss of analog RSM-based neuromorphic computing can be attributed to two main reasons: one is the nonideal effects of analog resistive switching, which causes the hardware training accuracy lower than the software training accuracy;⁵¹ the other is the conductance drift or fluctuation of the analog RSM devices, which causes the accuracy after training to degrade with time.⁶⁰ Based on this analysis, the device reliability metrics can be classified into basic reliability metrics and functional reliability metrics, as shown in Fig. 3. Basic reliability metrics are valid for both memory and computing applications, including retention, endurance, write/read disturbance, and random noise, whereas functional reliability metrics have attracted widespread attention mainly for computing applications, including nonlinearity, asymmetry, dynamic range, precision, variation, and bit yield. For analog RSMs, some basic reliability issues can degrade the functional reliability metrics and further decrease the accuracy, for example, endurance degradation usually triggers a decrease in the dynamic range and increase in nonlinearity.⁶¹ However, both the basic and functional reliability metrics are related to the accuracy loss.^{62,63} Therefore, the physical mechanisms and impact of reliability issues for analog RSM devices are very complex and require comprehensive studies. More importantly, the definition of the basic reliability metrics of the analog memory devices for neuromorphic computing application is different from that of the conventional memory application. It is highly desired to clarify their differences and provide a clear criterion for the study of reliability

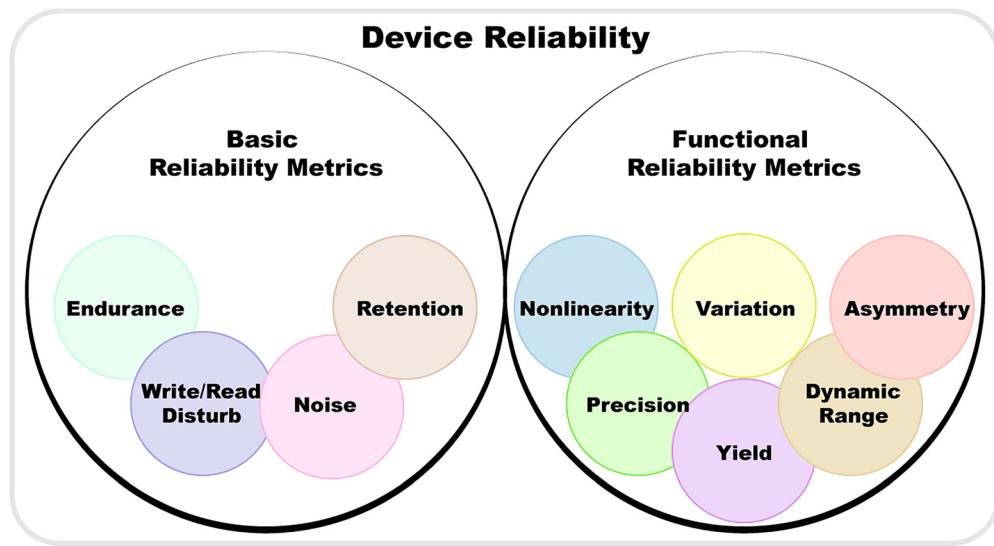


FIG. 3. Reliability metrics of the neuromorphic device. Device reliability metrics are classified into basic and functional reliability metrics, which degrade the accuracy during and after the training process.

physics in the future. In the following part of this section, we will discuss the basic and functional reliability metrics.

A. Basic reliability metrics

Basic reliability metrics refer to the common and essential reliability characteristics for both memory and neuromorphic computing

applications. Retention, endurance, write/read disturb, and noise are the critical reliability metrics of RSM devices. Their similarities and differences are illustrated in Fig. 4. The requirement of memory application focuses on the distinguishable resistance states, regardless of the change in the resistance value during the programming and data retention process.⁶⁴ Therefore, the reliability evaluation of memory applications mainly cares about the resistance windows between

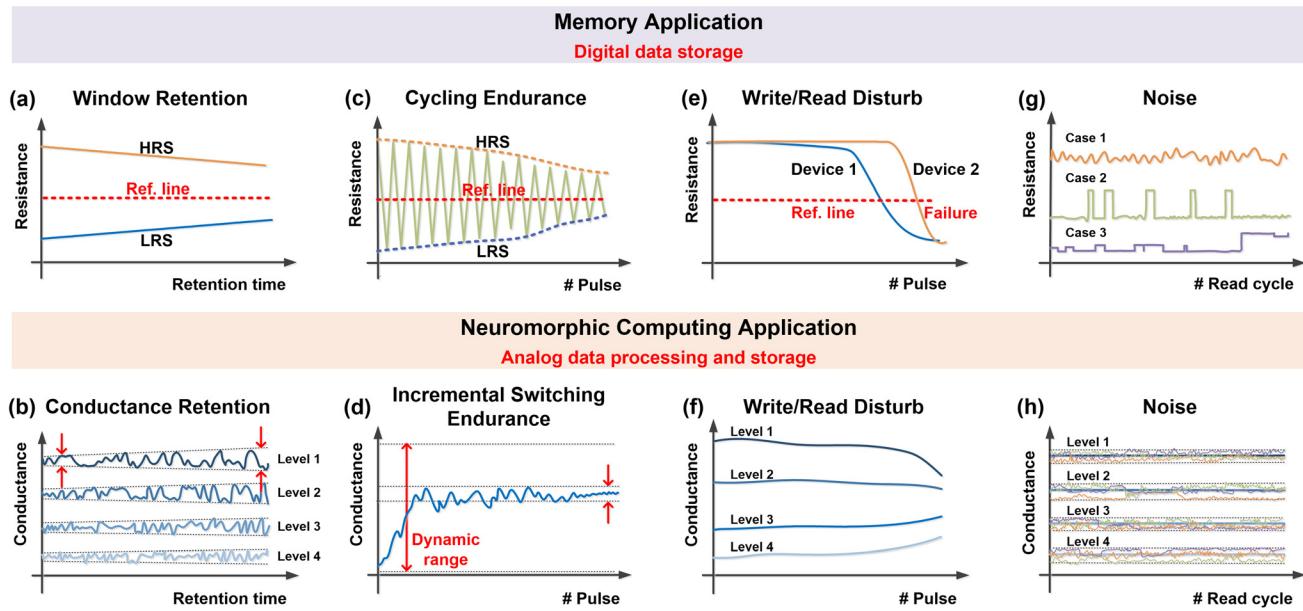


FIG. 4. Schematic diagram of different basic reliability metrics of memory application for digital data storage and neuromorphic computing application for analog data processing and storage. (a) Window retention of digital memory and (b) conductance retention of analog data in the computing process. (c) Cycling endurance of digital memory and (d) incremental switching endurance of analog data in the computing process. (e) Write/read disturb in memory and (f) computing. (g) Noises in memory and (h) computing.

different states. For example, the window retention means that the resistance window should maintain for more than 10 years at 85 °C according to the industry standard.³⁴ As long as the resistance value does not cross the reference line, the small range variation of the resistance value is permitted, as shown in Fig. 4(a). Cycling endurance also allows narrowing the switching window with the increased switching cycles unless the window disappears, as shown in Fig. 4(c). The largest endurance cycle is the most important evaluation criterion in the full window switching (between HRS and LRS).¹⁹ Write/read disturb refers to the unexpected resistance change during write or read process caused by long time accumulation of small voltage. Read disturb usually occurs on the selected cell in the array, caused by continuous read on the cell.⁶⁵ Write disturb occurs on the unselected cells in the array, caused by electric and thermal cross talk or sneak path effect. The occurrence of write disturb depends on the pulse voltage, array structure, location in the array,⁶⁶ and program scheme.²⁷ The criterion of disturbance is similar to retention, as shown in Fig. 4(e). Noise is an intrinsic characteristic of electron devices with a variety of forms, such as 1/f noise⁶⁷ and random telegraph noise (RTN),⁶⁸ as shown in Fig. 4(g). The noise limitation also lies in that the resistance with noises should not cross the reference line between HRS and LRS.

Compared to the reliability concerns of memory application, the concerns of neuromorphic computing application focus on the accurate conductance values, which are stricter from the device-level view. We hope that the conductance retention and conductance tuning process of the analog RSM remains stable, which requires the development of new mechanisms to further optimize the device, while from the system-level view, the requirements can be relaxed according to different applications, e.g., small degree of conductance fluctuation can be tolerated during inference,⁶⁹ which requires device-system co-optimization in the study. Specifically, the conductance retention of the analog RSMs pays attention to the individual conductance change of each analog level, as shown in Fig. 4(b). For endurance evaluation, neuromorphic computing focuses on the incremental switching, which means that the conductance of the analog RSM only changes a small value to mimic the weight update process.⁶⁰ The incremental switching endurance stands for the conductance tuning within varying levels and ranges as different numbers of pulses are applied, as shown in Fig. 4(d). The conductance evolution during different training algorithms may be quite different, and so the evaluation and measurement methods should also be adjusted according to different algorithms. The degradation of functional reliability metrics should also be considered during endurance tests.⁶¹ Write/read disturb shows similar behavior with the conductance retention by replacing the retention time with the pulse number as shown in Fig. 4(f). Because the operation schemes for inference and training are slightly different from the read and write schemes of memory application,^{13,65} particularly featured by the different parallelism, write/read disturb of the analog RSM for neuromorphic computing also needs further study and is largely dependent on the network structure and learning rule. Noise effects have more impacts on the neuromorphic computing application. The degree of read current fluctuation should be limited within a safety range to ensure a high degree of confidence on the neuromorphic computing [Fig. 4(h)].⁷⁰

To investigate the impact of conductance retention degradation on the performance of neuromorphic computing, a neural network simulator was developed to classify patterns from the MNIST

database, as shown in Fig. 5(a). A retention model was extracted to capture the conductance distribution evolution with the retention time.⁷¹ Apparent degradation ($\sim 4.6\%$) of the recognition rate was found after 10^4 s of baking at 175 °C.⁷² Based on the Arrhenius equation, this retention time was equivalent to 5.45 years at 85 °C [Fig. 5(b)]. When considering the impact of endurance degradation on the performance of neuromorphic computing, the impact of endurance on the functional metrics should be studied first. It was found that the functional reliability metrics, such as the dynamic range, nonlinearity, and asymmetry, degraded gradually with increasing numbers of increment switching cycling [Fig. 5(c)].⁶¹ The deformed incremental switching curve of conductance vs pulse number corresponded to the weight mapping model. Therefore, the changing mapping model resulted in the complexity of the update process and was attributed to the decrease in learning accuracy. Figure 5(d) shows that a significant accuracy loss occurred after 10^7 cycles when considering both nonlinearity and dynamic range degradation.⁶¹ Figure 5(e) shows the impact of read noise on accuracy loss with online (*in situ*) and offline (*ex situ*) training. It was found that accuracy loss became serious when the read noise was above 15%, and the abrupt loss occurred in online training.⁷⁰

B. Functional reliability metrics

Functional reliability metrics refer to the functional properties of analog RSMs for neuromorphic computing and have a direct influence on the training accuracy. Nonlinearity,⁷³ asymmetry,⁷⁴ dynamic range,⁷⁵ precision,³⁷ variation,⁷⁶ and yield⁷⁷ are important functional reliability metrics. The characteristics and impact of these functional metrics on training accuracy are presented in Fig. 6. The degradation of these functional metrics during the reliability test should be taken into careful consideration. Although many simulation works have been done on this topic and already provided many valuable design guidelines, experimental demonstrations on the impacts of these functional reliability metrics are still required to give more conclusive results in the future.

The dynamic range means the conductance ratio of HRS and LRS, also called the on/off ratio, which is from $2 \times$ to $100 \times$ with different materials and structures of analog RSM devices [Fig. 6(a)].⁷⁸ In the case of determining precision, the dynamic range is directly related to the number of conductance or weight levels during the training process. Chen and Yu⁶⁰ demonstrated the impact of dynamic range degradation on recognition accuracy, as shown in Fig. 6(b). The concepts of nonlinearity and asymmetry are derived from the relationship between the conductance change and the weight-update pulse number or voltage polarity. Nonlinearity corresponds to the degree of curvature of the weight update curve of analog RSMs, performing the incremental conductance change with the increasing weight update pulse number [Fig. 6(c)]. Several papers presented the nonideal linearity in different analog RSM devices.^{12,79–81} Inconsistent conductance changes make it difficult to tune the conductance to the target with identical pulses, resulting in poor convergence rates during the training process. Therefore, the nonlinearity directly causes more training accuracy loss. Chen *et al.*⁸² presented the accuracy degraded in a higher long-term depression (LTD) and long-term potentiation (LTP) nonlinearity region of three realistic synaptic devices in the sparse coding (SC) algorithm on-chip [Fig. 6(d)]. Asymmetry is a metric of the symmetry of the curves of conductance change vs pulse number during weight increase and decrease processes, as shown in Fig. 6(e). Tang

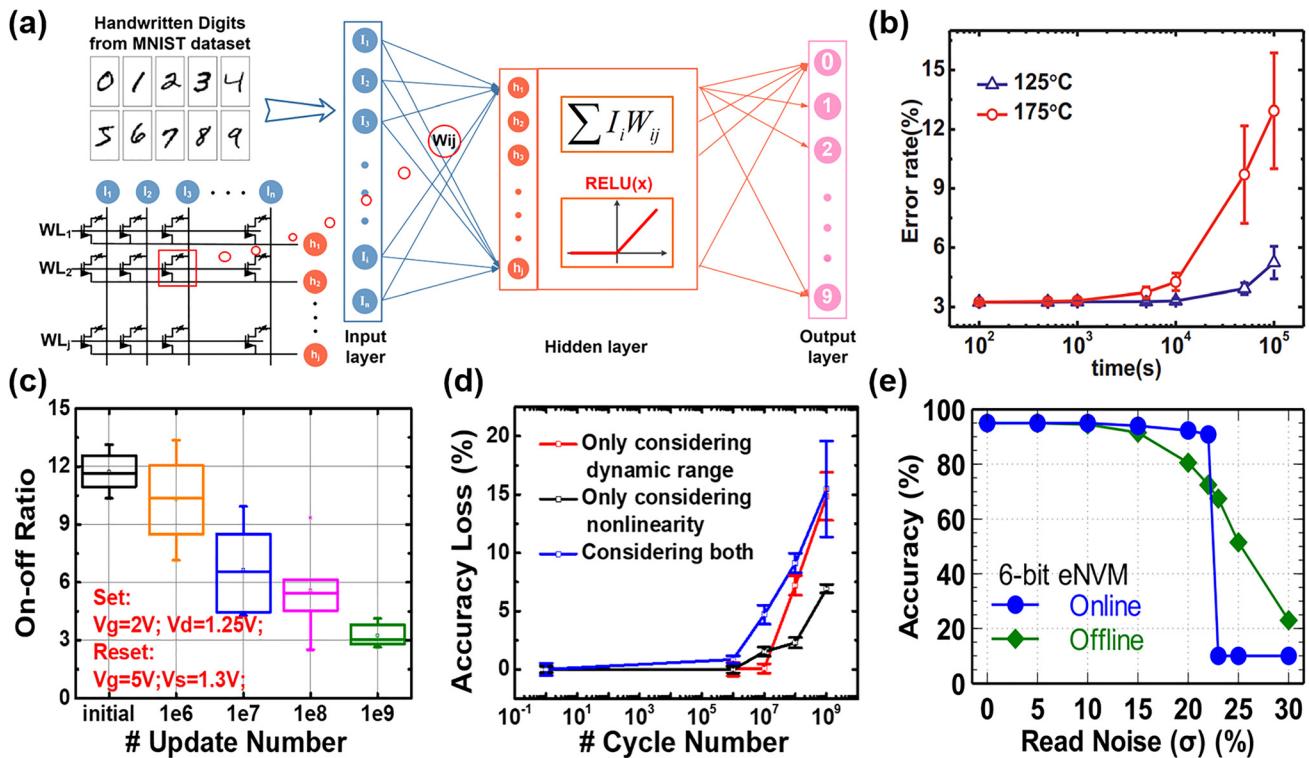


FIG. 5. Impact of basic reliability metrics on accuracy loss during MNIST image learning and recognition. (a) Schematic of the multilayer standard perceptron for recognizing images from the MNIST database, where the weights are implemented with the analog RSM array. (b) Impact of retention degradation on the recognition error rate. Reprinted with permission from Huang *et al.*, in IEEE International Electron Devices Meeting (IEDM) (2018), p. 40.4.1. Copyright 2018 IEEE. (c) On/off ratio decreased with the increasing weight update number due to endurance degradation. (d) The accuracy loss as a function of endurance cycle number. Reprinted with permission from Zhao *et al.*, in IEEE International Electron Devices Meeting (IEDM) (2018), p. 20.2.1. Copyright 2018 IEEE. (e) The impact of noise on accuracy loss in online and offline training. Reprinted with permission from Chen *et al.*, in IEEE International Electron Devices Meeting (IEDM) (2017), p. 6.1.1. Copyright 2017 IEEE.

*et al.*³⁷ implied that the accuracy mainly depended on the asymmetry, as shown in Fig. 6(f). Moreover, asymmetry influences the training accuracy, together with nonlinearity.⁸³ Li *et al.*⁸⁴ discussed the impact of nonlinearity and asymmetry on the training accuracy. It is found that the bidirectional symmetry incremental conductance change could keep good accuracy even with relatively large nonlinearity.⁸⁵ Precision indicates how many weight bits are provided by one device in the full dynamic range, determined by the analog switching ability [Fig. 6(g)].⁷⁶ For *ex situ* training, it has been proven that low bit precision can also implement reasonable inference accuracy.⁸⁶ However, it is necessary for a weight with high bit precision to realize an incremental weight update process during *in situ* training.⁸² Liu *et al.*⁸⁷ demonstrated the impact of weight precision on the accuracy for classifying the MNIST handwritten digits based on a perceptron neural network in the 1 kb 1T1R array. It suggested that the accuracy degraded gradually when the weight precision was lower than 4-bit for *ex situ* training [Fig. 6(h)]. Variations of analog RSMs for neuromorphic computing not only refer to the parameter (e.g., operation voltage, nonlinearity, and dynamic range) difference from one device to another and from one full switching cycle to another but also refer to the pulse-to-pulse variation during the one weight increase or decrease process [Fig. 6(i)]. For an inference-only system (*ex situ* training), device to device and cycle to cycle variations have a significant impact on accuracy loss.

Even with the verification programming scheme, these variations can lead to the deviation of the programmed conductance from the target value on each cell. The accumulated conductance programming deviation may cause computing error during inference.^{76,88} *In situ* training systems show better tolerance against the device to device and cycle to cycle variations, thanks to the self-adaptive ability.⁶⁹ But pulse-to-pulse variation can cause unregulated changes in the conductance change after applying one pulse. Due to the large number of weight update operation during training, pulse to pulse variation definitely results in the increased cost of training iteration and serious accuracy loss.^{82,89} Chen *et al.*⁷⁰ explored the impact of variation on accuracy loss, as shown in Fig. 6(j). With different nonlinearity during *in situ* training, high accuracy was only realized with small cycle-to-cycle variation and nonlinearity, and approximately 2% variation mitigated the accuracy degradation derived from the high nonlinearity. The bit yield refers to the percentage of RSM devices with analog switching behavior in the network. The low bit yield is caused by state-stuck or abrupt switching in some RSM devices, as shown in Fig. 6(k). Romero *et al.*⁷⁷ investigated the impact of the state-stuck effect on training accuracy. It was found that *in situ* training provided high tolerance to low bit yields to maintain reasonable accuracy, but accuracy loss remained inevitable [Fig. 6(l)]. As a supplement, there is a huge gap of accuracy loss shown in a small stuck ratio between *in situ* training and *ex situ* training, and

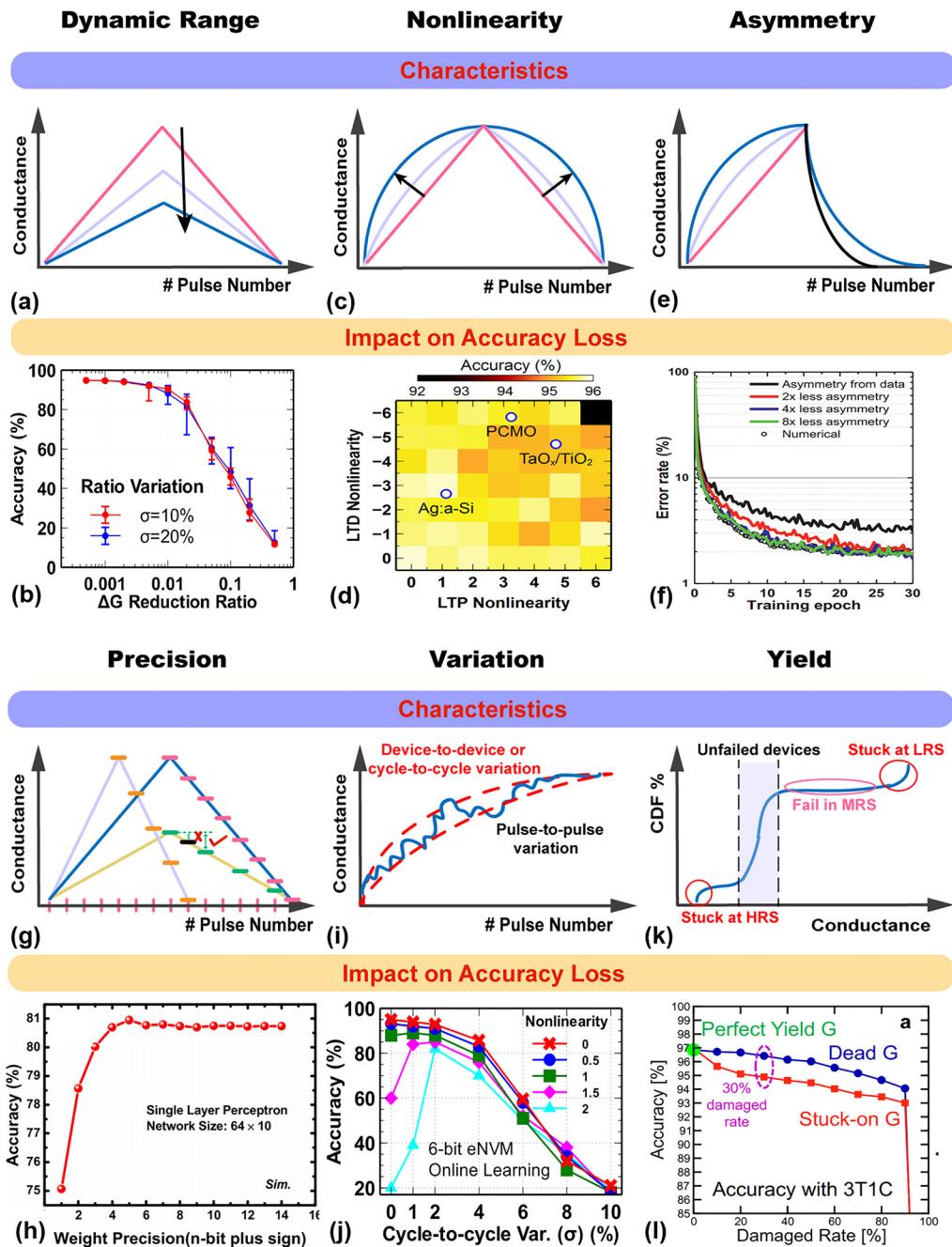


FIG. 6. The characteristics and definition of the functional reliability metrics of analog RSMs and the impact of these metrics on neuromorphic computing. (a), (c), (e), (g), (i), and (k) Schematic diagram of the dynamic range, nonlinearity, asymmetry, precision, variation, and yield. (b) Simulated accuracy as a function of dynamic range. The ΔG reduction ratio refers to the proportion of the dynamic range reduction. The error bar represents the impact of dynamic range variation on accuracy. Reprinted with permission from Chen *et al.*, in IEEE International Reliability Physics Symposium (IRPS) (2018), p. 5C.4. Copyright 2018 IEEE. (d) Test accuracy as a function of different nonlinearity magnitudes of three realistic RRAM devices. Reprinted with permission from Chen *et al.*, in IEEE/ACM International Conference on Computer-Aided Design (ICCAD) (2015), p. 3-A.3.194. Copyright 2015 IEEE. (f) MNIST simulation accuracy as a function of switching asymmetry. Reprinted with permission from Tang *et al.*, in IEEE International Electron Devices Meeting (IEDM) (2018), p. 13.1.1. Copyright 2018 IEEE. (h) Simulated accuracy as a function of precision bits. Reprinted with permission from Liu *et al.*, in European Solid-State Device Research Conference (2017), p. A3L-F18. Copyright 2017 IEEE. (j) Simulated accuracy loss due to pulse-to-pulse variation with several different nonlinearities during *in situ* training. Reprinted with permission from Chen *et al.*, IEEE International Electron Devices Meeting (IEDM) (2017), p. 6.1.1. Copyright 2017 IEEE. (l) Test accuracy as a function of stuck-on G rate and dead G rate after 20 epochs of training. Reprinted with permission from Romero *et al.*, Faraday Discuss. **213**(0), 371 (2019). Copyright 2019 RSC publishing.

the former performed better stuck at fault tolerance and a less accuracy loss than the latter due to its self-adaptation ability. Li *et al.*⁵⁴ reported that the presence of some nonresponsive devices leads to decreased accuracy in neuromorphic computing. It was found that multiple hidden layers in the neural network weakened the impact of stuck devices and further obtained higher accuracy than single-layer networks. This was because the hidden neurons can correspondingly adjust the connections to maintain the accuracy unchanged once a correlative device failed.

IV. REVIEW OF THE RELIABILITY STUDY

In this section, we focus on the state-of-the-art and representative works about the reliability of analog RSMs for neuromorphic computing. The studies of the basic and functional reliability metrics are reviewed, involving the measurement and characterization methods, landmark results, physical mechanism, and optimization methods. In this section, we also review some typical works on the reliability degradation mechanisms and optimization methods of binary RSMs because the mechanisms of binary RSMs have certain correlation with analog RSMs and can inspire in-depth studies of analog switching reliability in the future.

A. Basic reliability metrics study

1. Retention

Retention is a metric to evaluate how long the device can maintain its conductance value. For binary RSM devices, the widely accepted standard for reasonable retention is more than 10 years at 85 °C. Obviously, it is unrealistic to test for such a long time, and so the typical method is to accelerate the resistance drift at high temperature. The Arrhenius equation is used to convert the projected retention time at desired temperature from the experimental results at the measured temperature.⁹⁰ For analog RSM devices, it pursues a long retention time while keeping the conductance of multiple resistance levels unchanged and aims to predict the conductance distribution at variable temperatures and times. The measurement method is similar to the binary retention test, except it uses fine sampling to explore the evolution of conductance changes of various levels.

In the early exploration of memory application, a lot of works have been devoted to retention behavior of binary RSMs. Wei *et al.*⁹¹ showed the stable window-retention properties of Ta₂O₅-based anion-type binary RRAM in an 8-kbit array for approximately 3000 h at 150 °C. Rizzi *et al.*^{92,93} studied the retention statistics in a 1-Gb binary PCM array with Ge₂Sb₂Te₅ chalcogenide for 10⁵ s at 160 °C. For increasing conductance levels, the mean value decreases and the relative spread increases within a population of 16k PCM cells for calculations repeated for 100 cycles.⁹³ With the development of analog RSMs, the goal of research on retention ability is not only the long retention lifetime but also the tight conductance distribution of each level with time. Zhao *et al.*⁷¹ illustrated a statistical research on analog filamentary RRAM retention. Further work also found that the conductance distribution of each analog levels showed normal distribution at 175 °C.⁹⁴ After baking for 12 000 s at 3 μA, the standard deviation was about 0.50 μA. Lin *et al.*⁶² also studied a statistical retention test in a 4-level 1 Mb 1T1R tungsten-oxide RRAM array for 2 × 10⁵ s at 150 °C. The extended conductance distribution of each level was also showed with baking time. Stanisavljevic *et al.*⁹⁵ provided statistical experimental characteristics of analog PCM with conductance drift and elevated

temperatures. Over a retention time of 10⁵ s, all the 4 levels were presented by using the eM-metric at 80 °C with little overlap between each state.

The physical mechanisms of retention failure are different, owing to different device structures, material stacks, and switching types. The stochastic diffusion of ion or oxygen vacancy (Vo) could result in retention degradation. Wang *et al.*³⁸ directly observed the rupture and connection of CF through Ag nanoparticle migration using *in situ* TEM. The paper provided a direct characterization on the microscopic origin of resistive switching and diffusion process in a cation based RSM device. To explain complex scenarios, several simulation models were proposed based the physical principle and probability analysis such as Monte Carlo simulation. Zhao *et al.*⁹⁶ developed a physical compact model to explain the retention degradation of cation-type analog RRAM, as shown in Fig. 7(a). The retention degradation relied on the diffusion of metal atoms toward the lower concentration region resulting in the rupture (HRS) and connection (LRS) of the percolation path. Chen *et al.*⁹⁷ investigated two possible physical mechanisms for retention failure of anion-type binary RRAM, as shown in Fig. 7(b). It was found that the diffusion and recombination of mobile O²⁻ and Vo determined the change in filament morphology and further caused retention degradation. In PCM, retention degradation usually occurs in the amorphous state, resulting in tail bits and threshold voltage changes.⁹⁸ Russo *et al.*⁹⁹ proposed that the retention failure of binary PCM was derived from the spontaneous and thermal-activated crystallization of the programmed amorphous chalcogenide, as shown in Figs. 7(c)-7(e). A physical model is developed to obtain the resistance evolution with time in single device and statistical retention failure in a whole array by extracting the crystallization parameters (both geometry and electrical properties). It provided a valuable basis of statistic prediction for PCM retention performance.

To mitigate the retention degradation, some feasible solutions were provided regarding the process technology, innovative materials and structures of RSM devices, and programming schemes. Chen *et al.*⁹⁷ demonstrated a fabrication process improvement by adding an additional annealing operation for full stack after the cell patterning in the process flow based on HfO₂/Hf binary RRAM. The HfO₂ intermixed with Hf under the thermal effect, which caused the mobile oxygen to combine with the Hf. In this case, an HfO₂ interface layer was formed to slow the oxygen movement¹⁰⁰ and further mitigated the retention degradation. In addition, higher forming energy, provided by sources such as large current, long pulse width, or high temperature, resulted in better retention due to stronger CF formation. Huang *et al.*¹⁰¹ developed a 1-kbit array based on HfO₂ anion-type binary RRAM for retention optimization. It was found that oxygen anneals after HfO₂ atomic layer deposition, yielding a significant improvement in retention and uniformity. After applying Al₂O₃ mixed with the HfO₂ layer, the tail bits in retention failure were suppressed, especially for HRS, which provided a method of improving the tail-bit retention.¹⁰² Moon *et al.*¹⁰³ adopted a Mo electrode to control the redox reaction at the interface to obtain good uniformity and retention characteristics of nonfilamentary analog RRAM based on Mo/Pr_{0.7}Ca_{0.3}MnO₃ (PCMO) by increasing the activation energy for oxygen migration. They further reported another material improvement by inserting an MnO_x buffer layer to realize bidirectional analog switching in nonfilamentary RRAM based on Al/Mo/PCMO with better retention and dynamic ranges.¹⁰⁴ Several previous works were

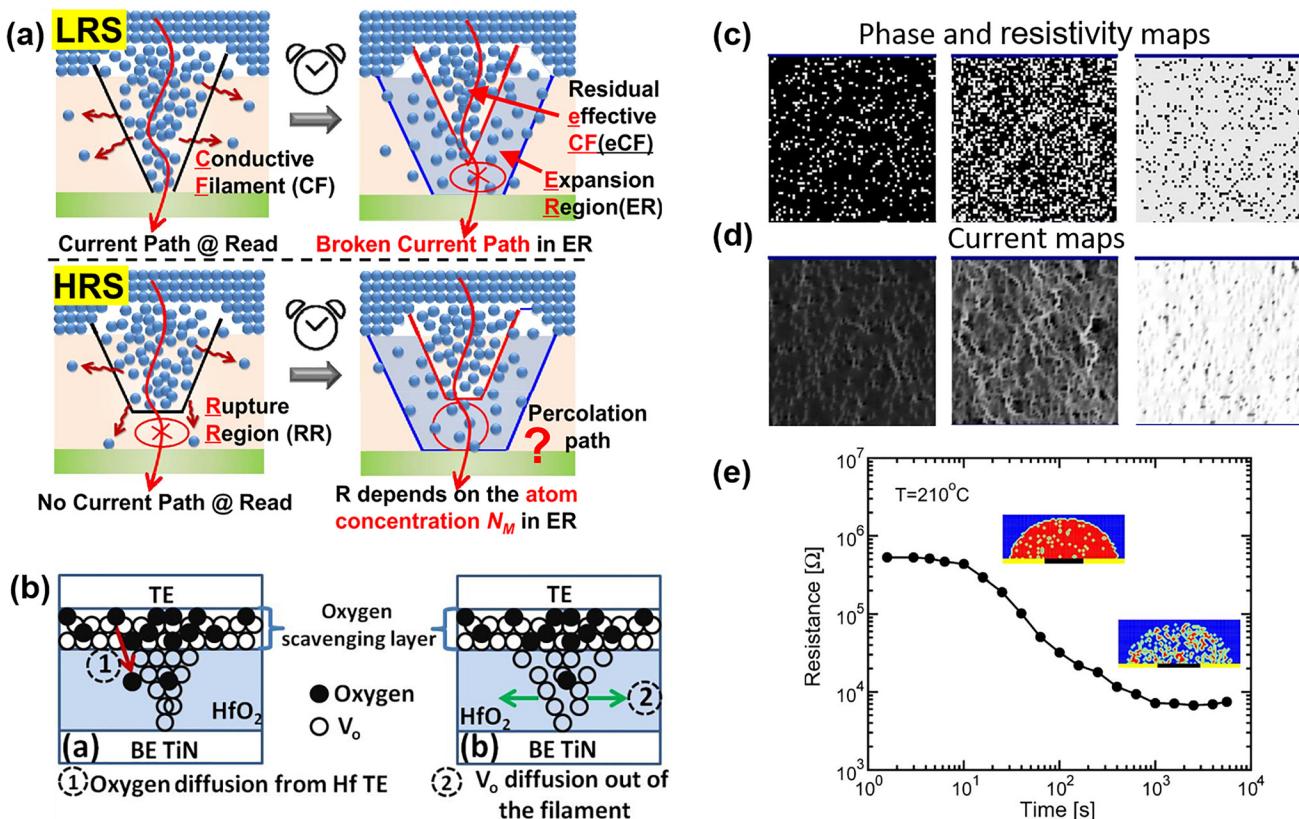


FIG. 7. Schematics of retention degradation. (a) Retention degradation process of cation-type RRAM. In LRS, the metal atoms in CF diffuse gradually toward the low atom concentration region with the increasing time, which may result in a broken current path in the expansion region. By contrast, diffusive atoms enter the rupture region (RR) corresponding to the increasing conductance (HRS). Reprinted with permission from Zhao *et al.*, IEEE Electron Device Lett. **40**, 647 (2019). Copyright 2019 IEEE. (b) Two possible retention degradation mechanisms in anion-type RRAM. Oxygen scavenged by the Hf cap layer diffuses back into HfO₂ and recombines with V_o in the filament; V_o diffusion out and dissolution of the filament. Reprinted with permission from Chen *et al.*, in IEEE International Electron Devices Meeting (IEDM) (2013), p. 10.1.1. Copyright 2013 IEEE. (c) and (d) Retention degradation process with the morphology change with the increasing time. (c) Simulation results of phase and resistive maps with different crystalline fractions. Light gray represents the crystalline elements, whereas the amorphous phase is in black. (d) Corresponding current maps with the morphology above. Lighter gray represents a higher current density. (e) Experimental results of resistance vs retention time at 210 °C. The inset shows the mixed-phase structure of short (left) and long (right) baking times. Reprinted with permission from Russo *et al.*, IEEE Trans. Electron Devices **53**, 3032 (2006), Copyright 2006 IEEE.

reported on improving the retention lifetime and performance of the multilevel PCM-based system through the trade-off between retention and write latency.^{105,106} This was because a longer write can achieve a better retention and high precision due to the tolerance of conductance drift, but it also resulted in a longer write latency, while a shorter write scheme caused a reduced retention time and a larger number of refresh operations. Zhang *et al.*¹⁰⁷ proposed region retention monitor (RRM) to balance the write latency and retention by automatically identifying hot access of each device and dynamically assigning the proper write schemes. Their further work demonstrated a lightweight scheme (called quick-and-dirty) to improve performance by 30.9% with a retention lifetime of 7.85 years on the geometric mean on a 2-bit PCM chip.¹⁰⁸

2. Endurance

Endurance is expressed as the maximum weight update number during the training process. In the endurance measurement, the conductance switches between several levels with alternating SET and

RESET programming pulses. The endurance behaviors of binary and analog RSM devices have been studied extensively. Compared to the endurance requirement of binary RSMs, analog RSMs such as synaptic devices attach importance not only to the sustainable endurance cycles between any two or more conductance levels to mimic the actual weight update process but also to the device performance degradation, such as some functional reliability metrics.

Various demonstrations on elevating the endurance ability in RRAM and PCM are reviewed; Lee *et al.*¹⁰⁹ reported an asymmetric and antiselial RRAM with Pt/Ta₂O_{5-x}/TaO_{2-x}/Pt bilayer structures that demonstrated excellent cycling endurance over 10¹² switching in binary switching mode. It also showed that the endurance lifetime increased with the resistance of the switching window and the decreasing oxygen partial-pressure conditions. Yeh *et al.*¹⁰⁹ achieved about 10⁹ programming endurance cycles and 10¹¹ read endurance cycles in the binary PCM device. However, incremental switching endurance of analog RSMs is required to satisfy the number of weight update in computing application. To gain an intuitive understanding, the

research on conductance evolution of various endurance switching pairs has been performed. Zhao *et al.*⁶¹ investigated the incremental endurance behaviors of analog RRAM for neuromorphic computing. Multiple conductance switching windows at different levels showed a huge gap in the weight update times. With efficient bidirectional verification, over 10^{11} incremental switching endurance cycles at low resistance levels were performed in a 1k analog RRAM array. In an analog PCM array, Athmanathan *et al.*¹¹⁰ demonstrated an endurance ability about 10^6 cycles of 3 bits/cell PCM with a variation of $\pm 1\sigma$ in a 64k array by combining drift-immune cell-state metrics and drift-tolerant coding and detection schemes.

To investigate the physical mechanism of endurance degradation, various approaches are applied including observation using high-resolution microscopy and analysis using compact models. Lee *et al.*¹⁹ observed the metal Ta clusters (white color) in the Ta_2O_{5-x} layer after

10^6 cycles by high-resolution TEM [Fig. 8(a)]. Chen *et al.*¹¹¹ reported three types of cycling endurance failure behaviors in anion-type binary RRAM [Figs. 8(b)–8(d)]. It should be noted that the oxygen reservoir is important to maintain good endurance behavior. Failure type I refers to an interfacial electron barrier induced by oxidation of the metal electrode at large power/current and high temperature. As the aforementioned switching mechanism, the interfacial barrier limits the transport of electrons and ions and further results in the endurance degradation. The reason for failure type II is that the electric field and accompanying heat lead to redundant Vos generation, enlarging the radius of filaments. In this case, the device typically fails at LRS. Failure type III refers to gradual changes of HRS. The excessive consumption of O^{2-} after frequent cycling causes the rupture of filaments due to the decreasing recombination rate. The TEM image illustrates that the atom clusters grow after 10^6 cycles. Switching between arbitrary two

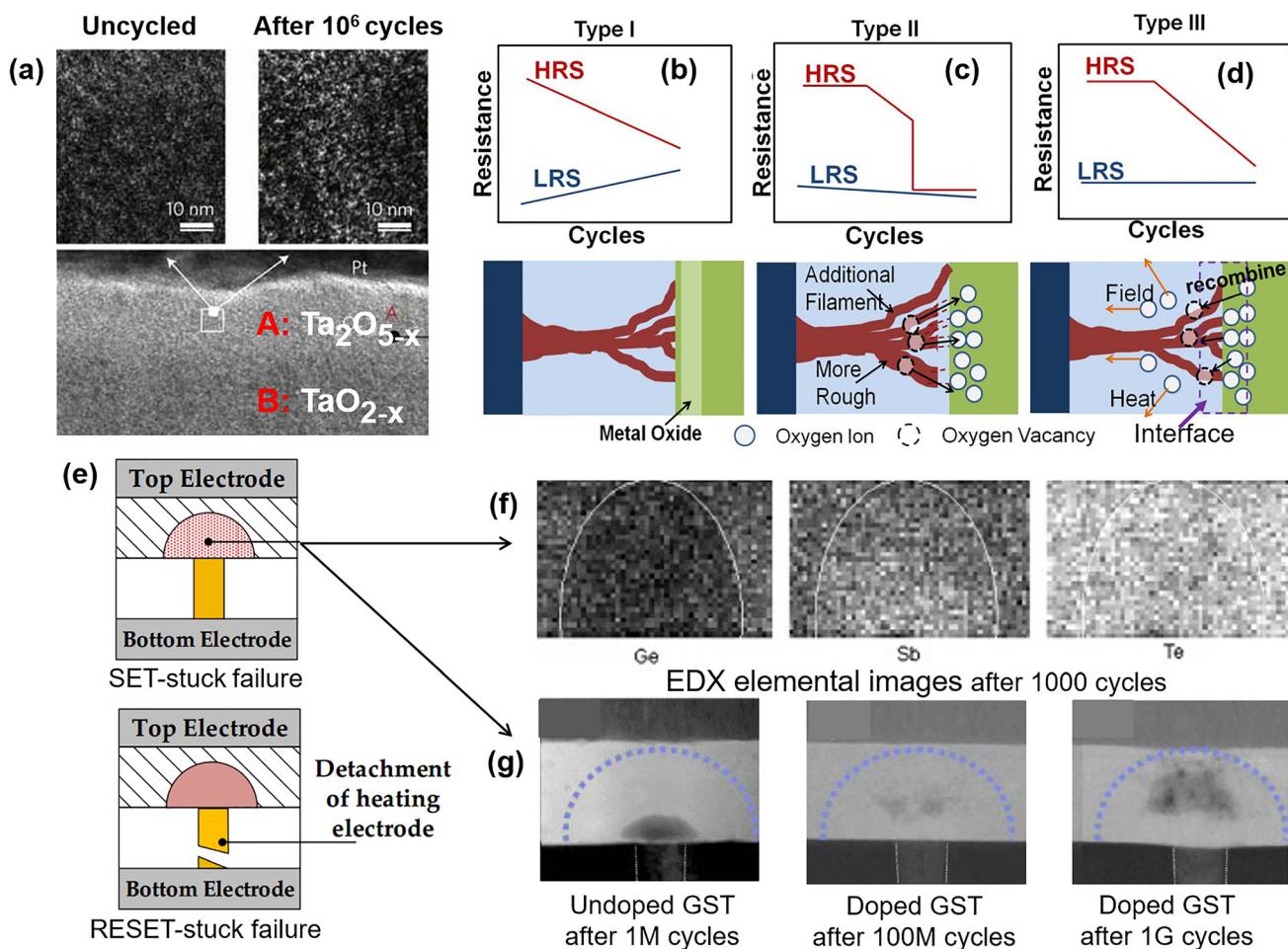


FIG. 8. Physical mechanism of endurance degradation of RSM. (a) TEM image of Ta metal clusters formed after cycles. Reprinted with permission from Lee *et al.*, Nat. Mater. **10**, 625 (2011). Copyright 2011 Macmillan Publishers. (b)–(d) Schematic of the endurance failure mechanism of RRAM. Three failure types of endurance degradation are illustrated. Reprinted with permission from Chen *et al.*, in IEEE International Electron Devices Meeting (IEDM) (2011), p. 12.3.1. Copyright 2011 IEEE. (e) The endurance failure mechanism of PCM. Schematic of SET-stuck failure and RESET-stuck failure. Reprinted with permission from Tavana *et al.*, in Proceedings of the International Symposium on Memory Systems (2017), p. 385. Copyright 2017 ACM. SET-stuck failure includes two main failure modes: (f) EDX (energy-dispersive X-ray spectroscopy) images after 1000 cycles, Ge depletion. Reprinted with permission from Raoux *et al.*, Microelectron. Eng. **85**, 2330 (2008). Copyright 2008 Elsevier. (g) TEM image of undoped and doped GST. Reprinted with permission from Chen *et al.*, in IEEE International Memory Workshop (2009), p. 1. Copyright 2009 IEEE.

conductance levels performed various endurance cycles, resulting from different physical origins. Zhao *et al.*¹¹² proposed the physical mechanism of endurance degradation of analog RRAM. The impact of switching windows with different resistance levels on the endurance lifetime was explained. The morphology of multiple weak CFs in the smaller switching windows was easily maintained after the endurance cycles. Figure 8(e) shows the endurance degradation of the PCM device. The failure was explained as two modes from SET- or RESET-stuck failure.¹¹³ SET-stuck failure is caused by (1) Ge depletion due to element separation [Fig. 8(f)]¹¹⁴ and (2) sustained void formation in the switching region near the bottom electrode. However, it has been reported that doped $\text{Ge}_2\text{Sb}_2\text{Te}_5$ (GST) could defer the appearance of clusters and improve the cycle endurance lifetime, as shown by the TEM images in Fig. 8(g).¹¹⁵ The reset-stuck failure may originate from the rupture and detachment of the heating electrode.

Through the analysis and conclusion of the physical mechanism, the endurance optimization methods are summarized from material/structure selection, programming schemes, and circuit design. Chen *et al.*¹¹⁶ proposed that the doping effect of Ti, Si, and Al in HfO_2 binary RRAM influences the cycling endurance lifetime. Single pulse endurance about 10^9 cycles was obtained because the dopants influenced the formation of the oxygen exchange layer. Grossi *et al.*¹¹⁷ proposed an end-to-end approach combining the programming scheme and system resilience techniques to overcome endurance and temporary bit error rates for deep learning application. They set the upper write number in each remapping period to reduce the write number of each RRAM device, basically decreasing the possibility of irreversible breakdown/dissolution of CFs. Yamaga *et al.*¹¹⁸ proposed the highly reliable approximate-RRAM to implement real-time image recognition with pixel-to-pixel data matching (P2P-DM) and interpixel error-correction code (ECC). In this case, compared to the Bose–Chaudhuri–Hocquenghem (BCH) ECC, the acceptable retention time and the endurance of the most significant bits (7th bit) suffered from relatively serious errors have improved by $5\times$ and $3.3\times$, respectively. Recent advances have been made to overcome limited endurance for Multi-level Cell (MLC) PCM. Pan *et al.*¹¹⁹

developed a write operation selection algorithm and task scheduling to improve endurance and energy efficient of MLC PCM-based systems.

3. Write/read disturb

Write disturb refers to a location-dependent write error of unselected devices after certain programming pulses, especially in large-scale crossbar arrays.²⁷ The $\text{Vdd}/2$ scheme is the typical solution for write disturb, where the unselected WLs and BLs are biased at the half voltage in the programming process [Fig. 9(a)].¹²⁰ If the half voltage can drive some half-selected devices to the wrong resistance values, a write disturb will occur. Moreover, the write disturb of PCM usually occurs in reset operation when these neighboring PCM cells are in the amorphous state because of the thermal cross talk, and the problem becomes more serious in sub-20-nm technology.¹²¹ The classical approach of avoiding the write disturb is to allocate a large intercell space¹²² and adopt a strong ECC.¹²³ In contrast, read disturb is a time-dependent current error of selected devices, affected by the programming and endurance.¹²⁴ It has been reported that read disturb issue exists inevitably in any circuit regimes. Similar to different reliability concerns of retention degradation, the largest write or read cycle number of the RSM device is the most important focus in the memory storage scenario, while in neuromorphic computing, the conductance fluctuation with increasing write/read pulses should be paid much attention because the accumulation of the effect of read disturb causes failure acceleration, resulting in the reduced learning accuracy.

The physical mechanism of write/read disturb of RSM devices is introduced from the perspective of formula derivation by experimental results and physical models with various materials and structures of devices. The mechanism of disturbance immunity of RRAM comes from the stability of ions in the switching region. Wang *et al.*¹²⁵ explained the reason for better disturbance immunity of $\text{Cu}_{x}\text{Si}_y\text{O}_z$ -based binary RRAM than Cu_xO . Higher activation energy suppressed the copper vacancy migration and reduced the probability of write and read disturbance. Using a physical model, Li *et al.*²⁷ provided the explanation of the physical mechanism of write disturbance in binary

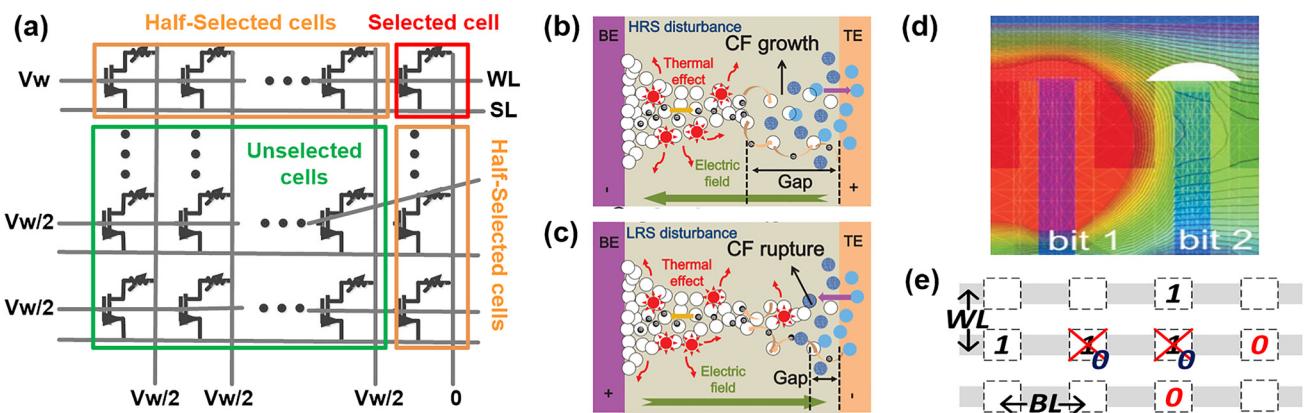


FIG. 9. Write disturbance of (a)–(c) RRAM and (d) and (e) PCM. (a) Schematic diagram of the RRAM crossbar array with the $1/2$ voltage scheme. (b) The physical mechanism of HRS disturbance of RRAM, incorrect resistive switching from HRS to LRS. (c) LRS disturbance, incorrect resistive switching from LRS to HRS. Reprinted with permission from Li *et al.*, in IEEE International Reliability Physics Symposium (IRPS) (2014), p. MY.3.1. Copyright 2014 IEEE. (d) Calculated temperature maps of write disturbance within two adjacent PCM cells at 45-nm technology node. The left cell was being programmed, and the right cell was disturbed by the increasing temperature, which was originally in the reset state (0). Reprinted with permission from Russo *et al.*, IEEE Trans. Electron Devices 55, 515 (2008). Copyright 2008 IEEE. (e) Vulnerable cells are colored in red. Reprinted with permission from Jiang *et al.*, in IEEE/IFIP International Conference on Dependable Systems and Networks (2014), p. 216. Copyright 2014 IEEE.

RRAM, as shown in Figs. 9(b) and 9(c). With the electric field and thermal effect, the broken CFs grew, accompanying the generated Vo moving along the electric field, shortening the gap, and inducing the disturbance of the resistance state. In LRS disturbances, Vo escaped from the CF region under the force of thermal and electric fields, resulting in the gap formation and resistance shifting. Figure 9(d) demonstrates the scenario of the write disturbance between two adjacent binary PCM cells under the thermal diffusion caused by the RESET programming pulse.¹²⁶ In particular, the situation only occurred when the neighboring cell was in the RESET state (storing “0”), and there was basically no disturbing influence for the cell in the SET state (storing “1”), as shown in Fig. 9(e). When the left cell was programmed to a high temperature, the heat spread horizontally to reach the right cell. The temperature of bit 2 caused crystallization but not melting, and thus, the cell in the RESET state turned into the crystallization state.

Aiming to control precisely the cell-location-dependent selected cells and avoid disturbance with unselected cells, Chen *et al.*²⁸ proposed an optimization method to control the cell-location-dependent selected cells precisely and avoid disturbance with unselected cells. It was found that inserting a thin AlO_x buffer layer under the resistance switching layer (HfO₂) can improve the tolerance to read disturbances in the binary RRAM array. Li *et al.*²⁷ proposed that the Vdd/3 scheme showed much better data preservation ability than the Vdd/2 scheme after a certain number of write pulses. This was because the lower programming voltage was applied to the unselected device in the crossbar RRAM array. Wang *et al.*¹²⁷ proposed a fine-grained write method to mitigate the write disturbance by utilizing the imbalance distribution in the binary PCM array. The imbalance referred to only a few cell groups that played a divisive role in the performance degradation based on the programming regime in memory. However, few works have been reported on the physical mechanism and optimization methods of the write/read disturb of analog RSMs, and this topic remains to be explored in depth.

4. Noise

Read noise is classified into three types: thermal noise,¹²⁸ $1/f^\alpha$ -like noise,^{67,129} and random telegraph noise (RTN).⁶⁸ As the name suggests, thermal noise derives from the carrier movement or ion migration due to the voltage induced thermal effect. $1/f^\alpha$ -like noise ($\alpha \sim 1$ for LRS and $\alpha \sim 2$ for HRS) refers to a kind of low-frequency current fluctuation. RTN is a dominant pattern of low-frequency noise (LFN) with the conductance oscillating between two states, originating from the filling or emptying of one or more traps. RTN determines the read-disturb immunity and bit precision in the analog RSM.¹³⁰ It has been proven that a certain weight standard deviation due to noise results in the accuracy loss in neuromorphic computing.⁷⁰

The physical mechanisms of the mentioned various noises were elucidated on RRAM and PCM. Huang *et al.*¹³¹ provided the phenomenon and physical explanation of RTN in binary RRAM and presented a triangular programming pulse scheme to suppress the tail bits. Similar works on the physical mechanism of noise were reported on estimating the diameter of CFs,¹³² establishing the electron tunneling mechanism for the filamentary conductive process based on LFN behavior,⁶⁷ detecting the Vo count and its properties (activation and deactivation) in the filament region,¹³¹ and so on. Based on the

thorough analysis, several optimization methods were developed to mitigate the impact of noise and conversely utilized the noise as the source for special applications. To reduce the $1/f$ noise, Kim *et al.*¹³³ utilized a metal nitride liner in a multilevel PCM device to provide another conductive path in the amorphous region, which was proven as the dominant source of large noise.¹³⁴ It was found that more than 4 times noise reduction can improve the multilevel performance and program-and-verify ability. Giannopoulos *et al.* demonstrated an 8-bit projected analog PCM. It showed remarkable immunity to $1/f$ noise with the introduction of the noninsulating projected segment in parallel to the phase-change segment. Noise could be utilized to carry out probabilistic inference through sampling in the neural network¹³⁵ and has solved several problems. Lin *et al.*¹³⁶ proposed a generative adversarial network (GAN) based on an analog RRAM array by utilizing the intrinsic noise as inputs to diversify the generated outputs. Cai *et al.*¹³⁷ demonstrated an optimization in speed and energy efficiency with the intrinsic analog noise as the computing resource for an RRAM-based Hopfield network.

B. Functional reliability metrics study

1. Nonlinearity

Nonlinearity, attracting unprecedented attention, can be tracked back to the analog RSM device acting as the electrical synapses in the neural network. It reflects the rate of conductance change with the number of voltage pulses, which degrades the training accuracy, as shown in Fig. 6(d). In addition, large nonlinearity leads to complex weight modulation and high energy and time costs in the training process. Therefore, it is necessary to improve the nonlinearity of analog RSM for higher accuracy.

Based on the device characterization and the understanding of the physical mechanism of resistive switching, several representative works have been dedicated to trying novel material and structures of RSM devices to improve nonlinearity. Wu *et al.*¹³⁸ introduced a methodology to improve the linearity of analog filamentary RRAM for both the SET and RESET processes by inserting a electrothermal modulation layer (ETML) over the switching layer (HfO₂). ETML was reported to not only control the distribution of the electric field to suppress the change in the electric field in the filament region for RESET linearity but also to control the thermal distribution to make the Vo distribution uniform for SET linearity. Chandrasekaran *et al.*¹³⁹ introduced Al dopants to improve the nonlinearity of the HfO₂-based analog RRAM. The uneven doping method resulted in oxygen-rich and oxygen-poor regions in the switching layer to confine the filament formation, which decreased the nonlinearity by 14% in potential and 31% in depression, respectively. Moon *et al.*⁶³ designed a 1T2R structure with the nonfilamentary analog RRAM as a synapse device to achieve linear conductance changes. With an additional serial-connected resistor for voltage division, the identical programming pulses can be converted to incremental pulses to improve the nonlinearity. Besides, some representative works on programming schemes have been reported to improve the nonlinearity.¹⁴⁰ Chen *et al.*⁸² proposed a smart programming scheme for linear weight update with a pair of positive and negative pulses to mitigate the overshoot effect of the previous pulse. Furthermore, the pulse duration was controlled to vary with the conductance levels to slow the weight update at the beginning of depression and potential, which also inevitably burdened the peripheral circuitry. Cai *et al.*⁴⁸

mitigated the device I-V nonlinearity by pulse modulation and custom analogue-to-digital converter (ADC) in a fully functional, hybrid memristor chip to reduce the multiplication error.

2. Asymmetry

Unlike the characterization of nonlinearity, asymmetry is used to indicate the degree of difference in the conductance change of a certain conductance level between the potential and depression stages, as shown in Fig. 6(e). Similar to the optimization method for nonlinearity, Li *et al.*⁵⁴ developed a two-pulse conductance programming scheme to achieve the linear and symmetric tunable analog behavior. Based on linear and symmetric weight updates, only 2.4% lower accuracy than the ideal value was achieved in *in situ* training. Lee *et al.*¹⁴¹ utilized a fixed resistor connected to the analog RRAM device in series to implement compensational voltage division. In this case, the asymmetric conductance changes can be improved by controlling the induced oxide to form smoothly at the interface under identical pulse bias. Consequently, the optimized asymmetry attributed to the significant promotion of recognition accuracy from 30% to 96%. Asymmetry usually appears with nonlinearity, together causing the accuracy loss. Ambrogio *et al.*⁵¹ demonstrated software-equivalent DNN accuracy using the analog memory unit of 2PCM+3T1C ("3 transistors, 1 capacitor") devices. Compared to PCM, the lower asymmetry can be obtained to implement high training efficiency. Haensch¹⁴² broke the connection between asymmetry and nonlinearity. It was believed that the bidirectional devices as synapses did not rely on the linearity, but the symmetric response, i.e., the mirror images of the incremental conductance changes, is indeed required in the SET and RESET process.

3. Dynamic range

The dynamic range is the on/off ratio between the highest and lowest conductance values. A larger dynamic range can result in high precision and the weight mapping ability,³⁴ further providing higher accuracy. To increase the dynamic range, some innovative device structures are developed. Moon *et al.*⁶³ demonstrated the excellent dynamic range for more than 100 based on a 1T2R analog RRAM device. To achieve a large dynamic range, a parallel connection of an RRAM and a transistor forced the transistor to operate in the steep subthreshold region of the MOSFET. Then, a small voltage change in the RRAM induced a large shift in the drain current by controlling the gate voltage bias. Choi *et al.*¹⁴³ developed a transistor-free SiGe epiRAM to control the formation of metal filaments in a customized channel. They claimed that the confined CFs dramatically enhanced uniformity and reliability with a large dynamic range, resulting in a high online learning accuracy of 95.1%. Ambrogio *et al.*⁵¹ introduced a 2T2R+3T1C unit cell with the increased dynamic range by applying different scale factors on read current for two pairs of conductance. The designed unit cell can also improve the update symmetry, contributing to good training accuracy.

4. Precision

Precision refers to the achievable maximum of weight bits in the full switching window. High precision is required for both *ex situ* and *in situ* training in neuromorphic computing. In order to explain the formation of high precision, Gao *et al.*¹⁴⁴ investigated the physical

mechanism of abrupt and analog switching using kinetic Monte Carlo simulation. It was suggested that achieving high precision of analog RSMs should avoid the formation of strong CFs. Several optimization methods were explored and developed to obtain high precision in RRAM and PCM with considering their physical principles. Stathopoulos *et al.*¹⁴⁵ demonstrated a bilayer 2-terminal metal-insulator-metal (MIM) structure of the analog nonfilamentary RRAM device with up to 6.5 bits capacity based on $\text{Al}_x\text{O}_y/\text{TiO}_2$ stack. The key technology of high precision lies in introducing a thin interfacial barrier layer between the active layer and one electrode benefiting to the device stability¹⁴⁶ and increasing the number of conductance levels, while the conductance variations and noise of the RSM devices seriously limit the high precision implemented. Giannopoulos *et al.*¹⁴⁷ reported the projected analog PCM with 8-bit precision. A simple temperature compensation method for the PCM device was developed to correct and the temperature variation and noise. In this case, the 100% classification accuracy was observed in a single-layer neural network using a crossbar with 30 projected PCM devices.

5. Variation

Different from noise and fluctuation, variation here emphasizes the spatial variation from device to device and the temporal variation from pulse to pulse.⁸² Broadly speaking, variation should include the variation in functional reliability metrics, such as linearity, symmetry, dynamic range, and precision. From the perspective of physical mechanism, variation is an intrinsic instability due to stochastic ion migration. Chen *et al.*⁸² showed that the tolerable accuracy loss can be performed in *in situ* learning with ~30% device variation. The impact of pulse-to-pulse variation on learning accuracy should be taken seriously, and approximately 22% temporal variation corresponds to less than 90% learning accuracy with the best linearity, based on the simulation results (Fig. 10). To reduce the influence of variation, Prezioso *et al.*²¹ varied the titanium dioxide compositions and layer thickness to select the optimal parameter range in RRAM stacks to realize low device variability by experimental search. Besides, Montano and Cheng¹⁴⁸ utilized the resistance ratio to encode information using two series RRAM cells connected with a transistor. Gao *et al.*¹⁴⁹ proposed a three-dimensional vertical structure with several parallel RRAM devices on the same nanopillar to suppress the intrinsic variation. In this case, the recognition accuracy was improved from 65% to 90%, based on the simulation results. Alibart *et al.*⁷⁶ designed a simple feedback algorithm to reduce the variation by adopting the resistance state within 1% relative accuracy of the dynamic range.

6. Bit yield

The bit yield represents the proportion of cells with normal analog resistive switching ability to the total number of devices in one array. It results from the fabrication and integration process and undesired device reliability degradation. Tran *et al.*¹⁵⁰ reported an ultrahigh yield (almost 100%) on a 6-in. wafer based on binary HfO_x -RRAM devices with Si-diode selectors. However, an acceptable yield of large-scale RSM arrays is not easy to obtain owing to the immature fabrication technology and undesired reliability degradation. Since the optimization from the device level is difficult to substantially increase the bit yield, some works have proposed creative write/read strategies and algorithms. Shih *et al.*¹⁵¹ improved the yield of 128-kb binary HfO_2 -

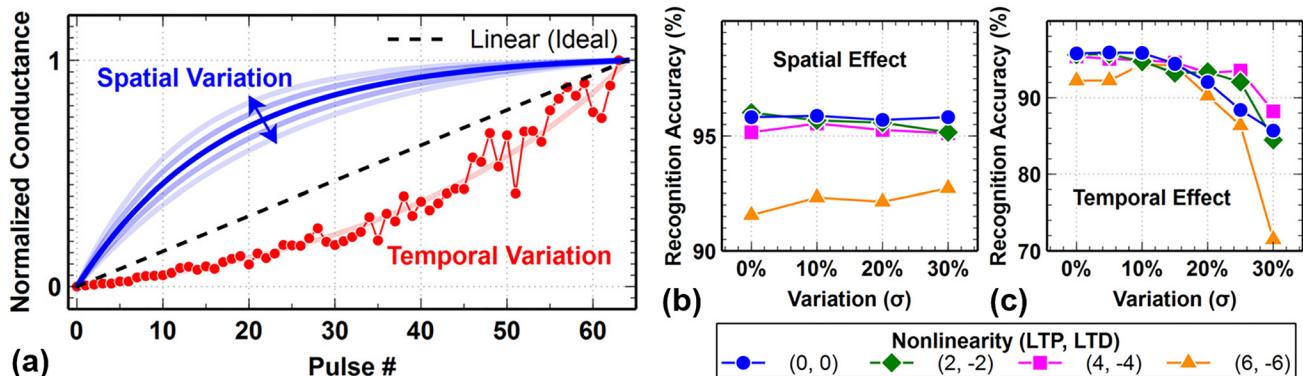


FIG. 10. Device variation and its impact on accuracy. (a) Illustration of spatial variation and temporal variation in the weight update process. Different devices show slight non-linear differences due to spatial variation. Temporal variation refers to the fluctuation of conductance with the incremental pulses of one device. (b) and (c) Recognition accuracy as a function of the standard deviation of device variation. The curves of different colors and shapes represent the nonlinearity baseline from (0, 0) to (6, -6) of long-term potentiation (LTP) and long-term depression (LTD). Compared to the spatial effect, the impact of the temporal effect on recognition accuracy is more critical. Reprinted with permission from Chen *et al.*, in IEEE/ACM International Conference on Computer-Aided Design (ICCAD) (2015), p. 3-A.3.194. Copyright 2015 IEEE.

RRAM circuits from 38.01% to 93.96% by addressing the overforming problem based on the training sequence. Xia *et al.*¹⁵² presented a fault-tolerance framework to reduce the number of stuck-at-fault RRAM devices. In detail, a mapping algorithm was proposed and attributed to the improved recognition accuracy of MNIST. Furthermore, hardware-level schemes with algorithm-level methods were explored to optimize the fault tolerance.¹⁵³ Xue *et al.*¹⁵⁴ developed self-adaptive write/read modes, improving the read bit yield of a 0.13- μm 8-Mb Cu_xSi_yO binary RRAM macro from 98% to 100% at 125 °C. Although several optimizations of the yield were proposed, improvements in the yield of large-scale analog RSM arrays for neuromorphic computing are still required. In addition, few publications involved the physical mechanism of the yield of an analog RSM array, which should be valued and further explored.

V. SUMMARY AND OUTLOOK

In the past few years, neuromorphic computing based on the emerging analog RSM has made notable progress. However, the research on the reliability of analog RSMs for neuromorphic computing still faces serious challenges in three aspects: (1) the reliability concerns and characterization methods of analog RSM devices are quite different, and well-accepted evaluation criteria are still lacking; (2) because of the complexity of the physical mechanism of analog resistive switching, mechanism studies for device reliability are difficult; (3) cross-layer codesign from the device to the system/algorithim is critical for neuromorphic computing; thus, a single-device-level study is not sufficient. For these reasons, this topic requires much effort for the reliability study of analog RSM-based neuromorphic computing.

In particular, we suggest several research directions that should be given much attention in the future. First, atom-level *in situ* characterization for the switching mechanisms is important for the reliability study. Direct observations of the dynamics of ion migration in the active region can offer critical evidence to understand the degradation mechanisms. Second, a complete reliability evaluation is required based on statistical measurements under different temperature conditions. The investigation should focus on the tail bits of a crossbar array to capture the stochastic behaviors of reliability degradation. Different

reliability metrics should be studied simultaneously to determine how each metric influences the others. Developing new techniques that can quickly finish the statistical measurement is required for this purpose. Third, the reliability evaluation should be performed with close correlation with a specific algorithm and system. At the initial stage, the mainstream AI algorithms could be considered, such as the convolutional neural network and recurrent neural network. Then, other algorithms like SNN and GAN can be considered. The reliability of 3D arrays is also important for the neuromorphic computing study. Finally, physical modeling and compact modeling of reliability degradation is also a key direction in the future study. Physical modeling is powerful for providing guidelines for reliability optimization. The compact model of reliability degradation must be added in future system-level simulators for performance benchmark and circuit design.

In this review, we have summarized the significant research studies on the reliability topic of analog RSM devices for neuromorphic computing. The landmark works involve the cross-layer reliability analysis, physical mechanism of device reliability, and optimization methods from the device characteristics to the algorithm and system. A set of evaluation methods of device reliability has been proposed to provide a guideline for further reliability research. Neuromorphic computing has enabled complex tasks at less cost than that of von Neumann architecture. As the problems of the large-scale integration of emerging analog RSM devices are solved, massive commercialization of neuromorphic computing chips will be realized. With excellent computing ability, there is no doubt that the neuromorphic computing chips will be widely used in various applications of the medical field, aerospace, and some areas related to human life. We expect that significant breakthroughs in reliable and energy-efficient neuromorphic computing chips based on analog RSM will be achieved in the near future.

ACKNOWLEDGMENTS

This work was supported in part by the National Key R&D Program of China (No. 2017YFB0405604), NSFC (Nos. 61851404,

61874169, 61674089, 61674092, and 61674087), National Major Research Program (No. 2017ZX02315001-005), Beijing Municipal Science and Technology Project (Nos. Z181100003218001 and Z191100007519008), and Beijing Innovation Center for Future Chips (ICFC).

REFERENCES

- ¹H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa, *Mobile Networks Appl.* **23**(2), 368 (2018).
- ²T. Young, D. Hazarika, S. Poria, and E. Cambria, *IEEE Comput. Intell. Mag.* **13**(3), 55 (2018).
- ³R. Hadsell, P. Sermanet, J. Ben, A. Erkan, M. Scoffier, K. Kavukcuoglu, U. Muller, and Y. Lecun, *J. Field Rob.* **26**(2), 120 (2009).
- ⁴M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muhamagic, *J. Big Data* **2**(1), 1 (2015).
- ⁵S. Sundaram, P. Kellnhofer, Y. Li, J.-Y. Zhu, A. Torralba, and W. Matusik, *Nature* **569**(7758), 698 (2019).
- ⁶M. A. Zidan, J. P. Strachan, and W. D. Lu, *Nat. Electron.* **1**(1), 22 (2018).
- ⁷W. A. Wulf and S. A. McKee, *ACM Sigarch Comput. Archit. News* **23**(1), 20 (1995).
- ⁸M. V. Wilkes, *ACM Sigarch Comput. Archit. News* **23**(4), 4 (1995).
- ⁹H. Tanaka, M. Akai-Kasaya, A. TermehYousefi, L. Hong, L. Fu, H. Tamukoh, D. Tanaka, T. Asai, and T. Ogawa, *Nat. Commun.* **9**(1), 2693 (2018).
- ¹⁰T. E. Potok, C. Schuman, S. Young, R. Patton, F. Spedalieri, J. Liu, K. Yao, G. Rose, and G. Chakrma, *Journal on Emerging Technologies in Computing Systems* **14**(2), 1 (2018).
- ¹¹D. Ielmini and H. S. P. Wong, *Nat. Electron.* **1**(6), 333 (2018).
- ¹²S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, *Nano Lett.* **10**(4), 1297 (2010).
- ¹³P. Yao, H. Wu, B. Gao, S. B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H. P. Wong, and H. Qian, *Nat. Commun.* **8**, 15199 (2017).
- ¹⁴Y. Chen, T. Krishna, J. Emer, and V. Sze, in IEEE International Solid-State Circuits Conference (ISSCC) (2016), p. 262.
- ¹⁵M. R. Mahmoodi and D. Strukov, in ACM/ESDA/IEEE Design Automation Conference (DAC) (2018), p. 20.2.1.
- ¹⁶G. W. Burr, R. M. Shelby, A. Sebastian, S. Kim, S. Kim, S. Sidler, K. Virwani, M. Ishii, P. Narayanan, A. Fumarola, L. L. Sanches, I. Boybat, M. L. Gallo, K. Moon, J. Woo, H. Hwang, and Y. Leblebici, *Adv. Phys. X* **2**(1), 89 (2017).
- ¹⁷X. Guo, F. M. Bayat, M. Bavandpour, M. Klachko, M. R. Mahmoodi, M. Prezioso, K. K. Likharev, and D. B. Strukov, in IEEE International Electron Devices Meeting (IEDM) (2017), p. 6.5.1.
- ¹⁸F. Merrikh-Bayat, X. Guo, M. Klachko, M. Prezioso, K. K. Likharev, and D. B. Strukov, *IEEE Trans. Neural Networks Learn. Syst.* **29**(10), 4782 (2018).
- ¹⁹M.-J. Lee, C. B. Lee, D. Lee, S. R. Lee, M. Chang, J. H. Hur, Y.-B. Kim, C.-J. Kim, D. H. Seo, and S. Seo, *Nat. Mater.* **10**(8), 625 (2011).
- ²⁰J. J. Yang, D. B. Strukov, and D. R. Stewart, *Nat. Nanotechnol.* **8**, 13 (2013).
- ²¹M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, and D. B. Strukov, *Nature* **521**(7550), 61 (2015).
- ²²M. Prezioso, I. Kataeva, F. Merrikh-Bayat, B. Hoskins, G. Adam, T. Sota, K. Likharev, and D. Strukov, in IEEE International Electron Devices Meeting (2015), p. 17.4.1.
- ²³B. Desalvo, E. Vianello, D. Garbin, O. Bichler, and L. Perniola, in IEEE International Memory Workshop (IMW) (2015), p. 1.
- ²⁴M. Cheng, L. Xia, Z. Zhu, Y. Cai, Y. Xie, Y. Wang, and H. Yang, in Proceedings of the 54th Annual Design Automation Conference (2017), p. 1.
- ²⁵M. Hu, J. P. Strachan, Z. Li, E. M. Grafals, N. Davila, C. Graves, S. Lam, N. Ge, J. J. Yang, and R. S. Williams, in ACM/EDAC/IEEE Design Automation Conference (DAC) (2016), p. 1.
- ²⁶C. Chang, J. Liu, Y. Shen, T. Chou, P. Chen, I. Wang, C. Su, M. Wu, B. Hudec, C. Chang, C. Tsai, T. Chang, H. P. Wong, and T. Hou, in IEEE International Electron Devices Meeting (IEDM) (2017), p. 11.6.1.
- ²⁷H. Li, H. Chen, Z. Chen, B. Chen, R. Liu, G. Qiu, P. Huang, F. Zhang, J. Zizhen, B. Gao, L. Liu, X. Liu, S. Yu, H. P. Wong, and J. Kang, in IEEE International Reliability Physics Symposium (2014), p. MY.3.1.
- ²⁸Y. S. Chen, H. Y. Lee, P. S. Chen, P. Y. Gu, C. W. Chen, W. P. Lin, W. H. Liu, Y. Y. Hsu, S. S. Sheu, P. C. Chiang, W. S. Chen, F. T. Chen, C. H. Lien, and M. Tsai, in IEEE International Electron Devices Meeting (2009), p. 1.
- ²⁹S. Deora, G. Bersuker, M. G. Sung, D. C. Gilmer, P. D. Kirsch, H. Li, H. Chong, and S. Gausepohl, in IEEE International Reliability Physics Symposium (IRPS) (2013), p. MY.2.1.
- ³⁰Z. Wang, L. Wang, M. Nagai, L. Xie, M. Yi, and W. Huang, *Adv. Electron. Mater.* **3**(7), 1600510 (2017).
- ³¹J. Lee and W. D. Lu, *Adv. Mater.* **30**(1), 1702770 (2018).
- ³²D. S. Jeong, K. M. Kim, S. Kim, B. J. Choi, and C. S. Hwang, *Adv. Electron. Mater.* **2**(9), 1600090 (2016).
- ³³Q. Xia and J. J. Yang, *Nat. Mater.* **18**(4), 309 (2019).
- ³⁴S. Yu, *Proc. IEEE* **106**(2), 260 (2018).
- ³⁵G. Zhu, W. Chen, D. Wang, H. Xie, Z. Zhao, P. Gao, J. Schutt-Aine, and W. Yin, *IEEE Trans. Electron Devices* **66**(4), 1747 (2019).
- ³⁶M. Jerry, P. Chen, J. Zhang, P. Sharma, K. Ni, S. Yu, and S. Datta, in IEEE International Electron Devices Meeting (IEDM) (2017), p. 6.2.1.
- ³⁷J. Tang, D. Bishop, S. Kim, M. Copel, T. Gokmen, T. Todorov, S. Shin, K. Lee, P. Solomon, K. Chan, W. Haensch, and J. Rozen, in IEEE International Electron Devices Meeting (2018), p. 13.1.1.
- ³⁸Z. 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, *Nat. Mater.* **16**(1), 101 (2017).
- ³⁹J. R. Jameson, P. Blanchard, C. Cheng, J. Dinh, A. Gallo, V. Gopalakrishnan, C. Gopalan, B. Guichet, S. Hsu, D. Kamalanathan, D. Kim, F. Koushan, M. Kwan, K. Law, D. Lewis, Y. Ma, V. McCaffrey, S. Park, S. Putthenthermadam, E. Runnion, J. Sanchez, J. Shields, K. Tsai, A. Tysdal, D. Wang, R. Williams, M. N. Kozicki, J. Wang, V. Gopinath, S. Hollmer, and M. V. Buskirk, in IEEE International Electron Devices Meeting (IEDM) (2013), p. 30.1.1.
- ⁴⁰S.-G. Koh, K. Kurihara, A. Belmonte, M. I. Popovici, G. L. Donadio, L. Goux, and G. S. Kar, *IEEE Electron Device Lett.* **39**(4), 480 (2018).
- ⁴¹A. Wedig, M. Luebben, D.-Y. Cho, M. Moors, K. Skaja, V. Rana, T. Hasegawa, K. K. Adeppalli, B. Yildiz, and R. Waser, *Nat. Nanotechnol.* **11**(1), 67 (2016).
- ⁴²S. Asanuma, H. Akoh, H. Yamada, and A. Sawa, *Phys. Rev. B* **80**(23), 235113 (2009).
- ⁴³M. Boniardi, A. Redaelli, C. Cupeta, F. Pellizzer, L. Crespi, G. D. Arrigo, A. L. Lacaita, and G. Servalli, in IEEE International Electron Devices Meeting (2014), p. 29.1.1.
- ⁴⁴S. Oh, T. Kim, M. Kwak, J. Song, J. Woo, S. Jeon, I. K. Yoo, and H. Hwang, *IEEE Electron Device Lett.* **38**(6), 732 (2017).
- ⁴⁵A. D. Kent and D. C. Worledge, *Nat. Nanotechnol.* **10**, 187 (2015).
- ⁴⁶R. C. Froemke and Y. Dan, *Nature* **416**(6879), 433 (2002).
- ⁴⁷C. Li, M. Hu, Y. Li, H. Jiang, N. Ge, E. Montgomery, J. Zhang, W. Song, N. Davila, C. E. Graves, Z. Li, J. P. Strachan, P. Lin, Z. Wang, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, and Q. Xia, *Nat. Electron.* **1**(1), 52 (2018).
- ⁴⁸F. Cai, J. M. Correll, S. H. Lee, Y. Lim, V. Bothra, Z. Zhang, M. P. Flynn, and W. D. Lu, *Nat. Electron.* **2**(7), 290 (2019).
- ⁴⁹Y. LeCun, in IEEE International Solid-State Circuits Conference (2019), p. 12.
- ⁵⁰P. Narayanan, A. Fumarola, L. L. Sanches, K. Hosokawa, S. C. Lewis, R. M. Shelby, and G. W. Burr, *IBM J. Res. Dev.* **61**(4/5), 11:1 (2017).
- ⁵¹S. Ambrogio, P. Narayanan, H. Tsai, R. M. Shelby, I. Boybat, C. di Nolfo, S. Sidler, M. Giordano, M. Bodini, N. C. P. Farinha, B. Killeen, C. Cheng, Y. Jaoudi, and G. W. Burr, *Nature* **558**(7708), 60 (2018).
- ⁵²F. Alibart, E. Zamanidoost, and D. B. Strukov, *Nat. Commun.* **4**, 2072 (2013).
- ⁵³P. A. Merolla, J. V. Arthur, A. I. Rodriguez, A. S. Cassidy, S. Jun, A. Filipp, B. L. Jackson, I. Nabil, G. Chen, and N. Yutaka, *Science* **345**(6197), 668 (2014).
- ⁵⁴C. Li, D. Belkin, Y. Li, P. Yan, M. Hu, N. Ge, H. Jiang, E. Montgomery, P. Lin, Z. Wang, W. Song, J. P. Strachan, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, and Q. Xia, *Nat. Commun.* **9**(1), 2385 (2018).
- ⁵⁵C. Liu, M. Hu, J. P. Strachan, and H. Li, in ACM/EDAC/IEEE Design Automation Conference (2017), p. 1.
- ⁵⁶Q. Zhang, H. Wu, P. Yao, W. Zhang, B. Gao, N. Deng, and H. Qian, *Neural Networks* **108**, 217 (2018).
- ⁵⁷Y. Liao, N. Deng, H. Wu, B. Gao, Q. Zhang, and H. Qian, *Front. Neurosci.* **12**, 167 (2018).

- ⁵⁸D. Negrov, I. Karandashev, V. Shakirov, Y. Matveyev, W. Dunin-Barkowski, and A. Zenkevich, *Neurocomputing* **237**, 193 (2017).
- ⁵⁹S. Yu, Z. Li, P. Chen, H. Wu, B. Gao, D. Wang, W. Wu, and H. Qian, in IEEE International Electron Devices Meeting (2016), p. 16.2.1.
- ⁶⁰P. Chen and S. Yu, in IEEE International Reliability Physics Symposium (IRPS) (2018), p. 5C.4.
- ⁶¹M. Zhao, H. Wu, B. Gao, X. Sun, Y. Liu, P. Yao, Y. Xi, X. Li, Q. Zhang, K. Wang, S. Yu, and H. Qian, in IEEE International Electron Devices Meeting (2018), p. 20.2.1.
- ⁶²Y. Lin, C. Wang, M. Lee, D. Lee, Y. Lin, F. Lee, H. Lung, K. Wang, T. Tseng, and C. Lu, *IEEE Trans. Electron Devices* **66**(3), 1289 (2019).
- ⁶³K. Moon, M. Kwak, J. Park, D. Lee, and H. Hwang, *IEEE Electron Device Lett.* **38**(8), 1023 (2017).
- ⁶⁴F. M. Puglisi, C. Wenger, and P. Pavan, *IEEE Electron Device Lett.* **36**(10), 1030 (2015).
- ⁶⁵S. Yu and P.-Y. Chen, *IEEE Solid-State Circuits Mag.* **8**(2), 43 (2016).
- ⁶⁶H. Chen, G. Bin, H. Li, L. Rui, H. Peng, C. Zhe, C. Bing, Z. Feifei, Z. Liang, J. Zizhen, L. Lifeng, L. Xiaoyan, K. Jinfeng, S. Yu, Y. Nishi, and H. P. Wong, in Symposium on VLSI Technology (2014), p. 1.
- ⁶⁷S. Yu, R. Jeyasingh, W. Yi, and H. P. Wong, in International Electron Devices Meeting (2011), p. 12.1.1.
- ⁶⁸D. Vekslar, G. Bersuker, L. Vandelli, A. Padovani, L. Larcher, A. Muraviev, B. Chakrabarti, E. Vogel, D. C. Gilmer, and P. D. Kirsch, in IEEE International Reliability Physics Symposium (2013), p. MY.10.1.
- ⁶⁹D. Garbin, O. Bichler, E. Vianello, Q. Rafhay, C. Gamrat, L. Perniola, G. Ghibaudo, and B. DeSalvo, in IEEE International Electron Devices Meeting (IEDM) (2014), p. 28.4.1.
- ⁷⁰P. Chen, X. Peng, and S. Yu, in IEEE International Electron Devices Meeting (IEDM) (2017), p. 6.1.1.
- ⁷¹M. Zhao, H. Wu, B. Gao, Q. Zhang, W. Wu, S. Wang, Y. Xi, D. Wu, N. Deng, S. Yu, H. Y. Chen, and H. Qian, in IEEE International Electron Devices Meeting (2017), p. 39.4.1.
- ⁷²P. Huang, Y. C. Xiang, Y. D. Zhao, C. Liu, B. Gao, H. Q. Wu, H. Qian, X. Y. Liu, and J. F. Kang, in IEEE International Electron Devices Meeting (2018), p. 40.4.1.
- ⁷³S. Sidler, I. Boybat, R. M. Shelby, P. Narayanan, J. Jang, A. Fumarola, K. Moon, Y. Leblebici, H. Hwang, and G. W. Burr, in European Solid-State Device Research Conference (2016), p. 440.
- ⁷⁴J. Park, M. Kwak, K. Moon, J. Woo, D. Lee, and H. Hwang, *IEEE Electron Device Lett.* **37**(12), 1559 (2016).
- ⁷⁵J. Welser, J. W. Pitera, and C. Goldberg, in IEEE International Electron Devices Meeting (2018), p. 1.3.1.
- ⁷⁶F. Alibart, L. Gao, B. D. Hoskins, and D. B. Strukov, *Nanotechnology* **23**(7), 075201 (2012).
- ⁷⁷L. P. Romero, S. Ambrogio, M. Giordano, G. Cristiano, M. Bodini, P. Narayanan, H. Tsai, R. M. Shelby, and G. W. Burr, *Faraday Discuss.* **213**, 371 (2019).
- ⁷⁸S. Ambrogio, S. Balatti, V. McCaffrey, D. C. Wang, and D. Ielmini, *IEEE Trans. Electron Devices* **62**(11), 3805 (2015).
- ⁷⁹L. Gao, I. T. Wang, P. Y. Chen, S. Vrudhula, J. S. Seo, Y. Cao, T. H. Hou, and S. Yu, *Nanotechnology* **26**(45), 455204 (2015).
- ⁸⁰S. Park, A. Sheri, J. Kim, J. Noh, J. Jang, M. Jeon, B. Lee, B. R. Lee, B. H. Lee, and H. Hwang, in IEEE International Electron Devices Meeting (2013), p. 25.6.1.
- ⁸¹J. Woo, K. Moon, J. Song, S. Lee, M. Kwak, J. Park, and H. Hwang, *IEEE Electron Device Lett.* **37**(8), 994 (2016).
- ⁸²P. Chen, B. Lin, I. Wang, T. Hou, J. Ye, S. Vrudhula, J. Seo, Y. Cao, and S. Yu, in IEEE/ACM International Conference on Computer-Aided Design (ICCAD) (2015), p. 194.
- ⁸³C. C. Chang, P. C. Chen, T. Chou, I. T. Wang, B. Hudec, C. C. Chang, C. M. Tsai, T. S. Chang, and T. H. Hou, *IEEE J. Emerging Sel. Top. Circuits Syst.* **8**(1), 116 (2018).
- ⁸⁴Y. Li, S. Kim, X. Sun, P. Solomon, T. Gokmen, H. Tsai, S. Koswatta, Z. Ren, R. Mo, C. C. Yeh, W. Haensch, and E. Leobandung, in IEEE Symposium on VLSI Technology (2018), p. 25.
- ⁸⁵W. Haensch, in 2018 IEEE International Electron Devices Meeting (IEDM) (2018), p. 3.2.1.
- ⁸⁶K. Ni, J. A. Smith, B. Grisafe, T. Rakshit, B. Obradovic, J. A. Kittl, M. Rodder, and S. Datta, in 2018 IEEE International Electron Devices Meeting (IEDM) (2018), p. 13.2.1.
- ⁸⁷R. Liu, H. Lee, and S. Yu, in European Solid-State Device Research Conference (ESSDERC) (2017), p. 18.
- ⁸⁸H. Wu, P. Yao, B. Gao, W. Wu, Q. Zhang, W. Zhang, N. Deng, D. Wu, H. P. Wong, S. Yu, and H. Qian, in IEEE International Electron Devices Meeting (IEDM) (2017), p. 11.5.1.
- ⁸⁹G. W. Burr, R. M. Shelby, S. Sidler, C. D. Nolfo, J. Jang, I. Boybat, R. S. Shenoy, P. Narayanan, K. Virwani, and E. U. Giacometti, *IEEE Trans. Electron Devices* **62**(11), 3498 (2015).
- ⁹⁰A. Pirovano, A. Redaelli, F. Pelizzier, F. Ottogalli, M. Tosi, D. Ielmini, A. L. Lacaita, and R. Bez, *IEEE Trans. Device Mater. Reliab.* **4**(3), 422 (2004).
- ⁹¹Z. Wei, Y. Kanzawa, K. Arita, Y. Katoh, and M. Oshima, in IEEE International Electron Devices Meeting (2008), p. 1.
- ⁹²M. Rizzi, N. Cioccioni, A. Montefiori, M. Ferro, P. Fantini, A. L. Lacaita, and D. Ielmini, in IEEE International Electron Devices Meeting (2013), p. 21.7.1.
- ⁹³M. Rizzi, N. Cioccioni, A. Montefiori, M. Ferro, and D. Ielmini, *IEEE Trans. Electron Devices* **62**(7), 2205 (2015).
- ⁹⁴H. Wu, M. Zhao, Y. Liu, P. Yao, Y. Xi, X. Li, W. Wu, Q. Zhang, J. Tang, B. Gao, and H. Qian, in IEEE International Reliability Physics Symposium (2019), p. 1.
- ⁹⁵M. Stanisavljevic, A. Athmanathan, N. Papandreou, H. Pozidis, and E. Eleftheriou, in IEEE International Reliability Physics Symposium (2015), p. 5B.6.1.
- ⁹⁶Y. Zhao, P. Huang, Z. Zhou, C. Liu, S. Qin, L. Liu, X. Liu, H. P. Wong, and J. Kang, *IEEE Electron Device Lett.* **40**(4), 647 (2019).
- ⁹⁷Y. Y. Chen, M. Komura, R. Degraeve, B. Govoreanu, L. Goux, A. Fantini, N. Raghavan, S. Clima, L. Zhang, A. Belmonte, A. Redolfi, G. S. Kar, G. Groeseneken, D. J. Wouters, and M. Jurczak, in IEEE International Electron Devices Meeting (IEDM) (2013), p. 10.1.1.
- ⁹⁸Y. H. Shih, J. Y. Wu, B. Rajendran, M. H. Lee, R. Cheek, M. Lamorey, M. Breitwisch, Y. Zhu, E. K. Lai, C. F. Chen, E. Stinzianni, A. Schrott, E. Joseph, R. Dasaka, S. Raoux, H. L. Lung, and C. Lam, in IEEE International Electron Devices Meeting (2008), p. 1.
- ⁹⁹U. Russo, D. Ielmini, A. Redaelli, and A. L. Lacaita, *IEEE Trans. Electron Devices* **53**(12), 3032 (2006).
- ¹⁰⁰S. Clima, Y. Y. Chen, R. Degraeve, M. Mees, K. Sankaran, B. Govoreanu, M. Jurczak, S. D. Gendt, and G. Pourtois, *Appl. Phys. Lett.* **100**(13), 133102 (2012).
- ¹⁰¹X. Huang, H. Wu, D. C. Sekar, S. N. Nguyen, K. Wang, and H. Qian, in IEEE International Memory Workshop (IMW) (2015), p. 1.
- ¹⁰²X. Huang, H. Wu, B. Gao, D. C. Sekar, L. Dai, M. Kellam, G. Bronner, N. Deng, and H. Qian, *Nanotechnology* **27**(39), 395201 (2016).
- ¹⁰³K. Moon, E. Cha, J. Park, S. Gi, M. Chu, K. Baek, B. Lee, S. Oh, and H. Hwang, in IEEE International Electron Devices Meeting (2015), p. 17.6.1.
- ¹⁰⁴K. Moon, A. Fumarola, S. Sidler, J. Jang, P. Narayanan, R. M. Shelby, G. W. Burr, and H. Hwang, *IEEE J. Electron Devices Soc.* **6**, 146 (2018).
- ¹⁰⁵D. Kang, S. Baek, J. Choi, D. Lee, S. H. Noh, and O. Mutlu, in Symposium on Mass Storage Systems and Technologies (MSST) (2015), p. 1.
- ¹⁰⁶M. Zhang, L. Zhang, L. Jiang, F. T. Chong, and Z. Liu, in IEEE International Conference on Computer Design (ICCD) (2017), p. 585.
- ¹⁰⁷M. Zhang, L. Zhang, L. Jiang, Z. Liu, and F. T. Chong, in IEEE International Symposium on High Performance Computer Architecture (HPCA) (2017), p. 385.
- ¹⁰⁸M. Zhang, L. Zhang, L. Jiang, F. T. Chong, and Z. Liu, *IEEE Trans. Comput.* **68**(9), 1365 (2019).
- ¹⁰⁹C. W. Yeh, W. C. Chien, R. L. Bruce, H. Y. Cheng, I. T. Kuo, C. H. Yang, A. Ray, H. Miyazoe, W. Kim, F. Carta, E. K. Lai, M. BrightSky, and H. L. Lung, in IEEE Symposium on VLSI Technology (2018), p. 205.
- ¹¹⁰A. Athmanathan, M. Stanisavljevic, N. Papandreou, H. Pozidis, and E. Eleftheriou, *IEEE J. Emerging Sel. Top. Circuits Syst.* **6**(1), 87 (2016).
- ¹¹¹B. Chen, Y. Lu, B. Gao, Y. H. Fu, F. F. Zhang, P. Huang, Y. S. Chen, L. F. Liu, X. Y. Liu, J. F. Kang, Y. Y. Wang, Z. Fang, H. Y. Yu, X. Li, X. P. Wang, N. Singh, G. Q. Lo, and D. L. Kwong, in International Electron Devices Meeting (2011), p. 12.3.1.

- ¹¹²M. Zhao, H. Wu, B. Gao, Y. Liu, P. Yao, Y. Xi, W. Wu, X. Li, Q. Zhang, N. Deng, and H. Qian, in Electron Devices Technology and Manufacturing Conference (2019), p. 267.
- ¹¹³M. K. Tavana, A. K. Ziabari, M. Arjomand, M. Kandemir, C. Das, and D. Kaeli, in Proceedings of the International Symposium on Memory Systems (2017), p. 385.
- ¹¹⁴S. Raoux, R. M. Shelby, J. Jordan-Sweet, B. Munoz, M. Salina, Y.-C. Chen, Y.-H. Shih, E.-K. Lai, and M.-H. Lee, *Microelectron. Eng.* **85**(12), 2330 (2008).
- ¹¹⁵C. Chen, A. Schrott, M. H. Lee, S. Raoux, Y. H. Shih, M. Breitwisch, F. H. Baumann, E. K. Lai, T. M. Shaw, P. Flaitz, R. Cheek, E. A. Joseph, S. H. Chen, B. Rajendran, H. L. Lung, and C. Lam, in IEEE International Memory Workshop (2009), p. 1.
- ¹¹⁶Y. Y. Chen, R. Roelofs, A. Redolfi, R. Degraeve, D. Crotti, A. Fantini, S. Clima, B. Govoreanu, M. Komura, L. Goux, L. Zhang, A. Belmonte, Q. Xie, J. Maes, G. Pourtois, and M. Jurczak, in IEEE Symposium on VLSI Technology (2014), p. 1.
- ¹¹⁷A. Grossi, E. Vianello, M. M. Sabry, M. Barlas, L. Grenouillet, J. Coignus, E. Beigne, T. Wu, B. Q. Le, M. K. Wootters, C. Zambelli, E. Nowak, and S. Mitra, *IEEE Trans. Electron Devices* **66**(3), 1281 (2019).
- ¹¹⁸Y. Yamaga, Y. Deguchi, S. Fukuyama, and K. Takeuchi, in IEEE Symposium on VLSI Technology (2018), p. 109.
- ¹¹⁹C. Pan, M. Xie, J. Hu, Y. Chen, and C. Yang, in Proceedings of the 2014 International Conference on Hardware/Software Codesign and System Synthesis (2014), p. 33.1.
- ¹²⁰C. Yi-Chou, C. F. Chen, C. T. Chen, J. Y. Yu, S. Wu, S. L. Lung, R. Liu, and L. Chih-Yuan, in IEEE International Electron Devices Meeting (2003), p. 37.4.1.
- ¹²¹K. Byeungchul, S. Yoonjong, S. Ahn, K. Younseon, J. Hoon, A. Dongho, S. Nam, J. Gitae, and C. Chilhee, in IEEE International Conference on ASIC (2011), p. 279.
- ¹²²S. J. Ahn, S. Yoonjong, J. Hoon, K. Byeungchul, K. Youn-Seon, A. Dong-Ho, Y. Kwon, S. W. Nam, J. Gitae, K. Hokeyu, and C. Chilhee, in International Electron Devices Meeting (2011), p. 12.6.1.
- ¹²³L. Jiang, Y. Zhang, and J. Yang, in IEEE/IFIP International Conference on Dependable Systems and Networks (2014), p. 216.
- ¹²⁴M. Kato, N. Miyamoto, H. Kume, A. Satoh, T. Adachi, M. Ushiyama, and K. Kimura, in IEEE International Electron Devices Meeting (IEDM) (1994), p. 45.
- ¹²⁵M. Wang, W. J. Luo, Y. L. Wang, L. M. Yang, W. Zhu, P. Zhou, J. H. Yang, X. G. Gong, Y. Y. Lin, R. Huang, S. Song, Q. T. Zhou, H. M. Wu, J. G. Wu, and M. H. Chi, in Symposium on VLSI Technology (2010), p. 89.
- ¹²⁶U. Russo, D. Ielmini, A. Redaelli, and A. L. Lacaita, *IEEE Trans. Electron Devices* **55**(2), 515 (2008).
- ¹²⁷R. Wang, L. Jiang, Y. Zhang, L. Wang, and J. Yang, in ACM/EDAC/IEEE Design Automation Conference (2015), p. 1.
- ¹²⁸M. S. Gupta, *Proc. IEEE* **70**(8), 788 (1982).
- ¹²⁹Z. Fang, H. Y. Yu, W. J. Fan, G. Ghiaudo, J. Buckley, B. DeSalvo, X. Li, X. P. Wang, G. Q. Lo, and D. L. Kwong, *IEEE Trans. Electron Devices* **60**(3), 1272 (2013).
- ¹³⁰N. Raghavan, R. Degraeve, A. Fantini, L. Goux, S. Strangio, B. Govoreanu, D. J. Wouters, G. Groeseneken, and M. Jurczak, in IEEE International Reliability Physics Symposium (2013), p. 5E.3.1.
- ¹³¹P. Huang, D. B. Zhu, C. Liu, Z. Zhou, Z. Dong, H. Jiang, W. S. Shen, L. F. Liu, X. Y. Liu, and J. F. Kang, in IEEE International Electron Devices Meeting (IEDM) (2017), p. 21.4.1.
- ¹³²D. Ielmini, F. Nardi, and C. Cagli, *Appl. Phys. Lett.* **96**(5), 053503 (2010).
- ¹³³S. Kim, N. Sosa, M. BrightSky, D. Mori, W. Kim, Y. Zhu, K. Suu, and C. Lam, *IEEE Trans. Electron Devices* **63**(10), 3922 (2016).
- ¹³⁴G. F. Close, U. Frey, M. Breitwisch, H. L. Lung, C. Lam, C. Hagleitner, and E. Eleftheriou, in International Electron Devices Meeting (2010), p. 29.5.1.
- ¹³⁵W. Maass, *Proc. IEEE* **102**(5), 860 (2014).
- ¹³⁶Y. Lin, H. Wu, B. Gao, P. Yao, W. Wu, Q. Zhang, X. Zhang, X. Li, F. Li, J. Lu, G. Li, S. Yu, and H. Qian, in IEEE International Electron Devices Meeting (2018), p. 3.4.1.
- ¹³⁷F. 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, e-print [arXiv:1903.11194](https://arxiv.org/abs/1903.11194) (2019).
- ¹³⁸W. Wu, H. Wu, B. Gao, P. Yao, X. Zhang, X. Peng, S. Yu, and H. Qian, in IEEE Symposium on VLSI Technology (2018), p. 103.
- ¹³⁹S. Chandrasekaran, F. M. Simanjuntak, R. Saminathan, P. D. Panda, and T.-Y. Tseng, *Nanotechnology* **30**(44), 445205 (2019).
- ¹⁴⁰I. T. Wang, C.-C. Chang, L.-W. Chiu, T. Chou, and T.-H. Hou, *Nanotechnology* **27**(36), 365204 (2016).
- ¹⁴¹C. Lee, S. Koo, J. Oh, and D. Lee, *IEEE J. Electron Devices Soc.* **6**, 403 (2018).
- ¹⁴²W. Haensch, in IEEE International Electron Devices Meeting (IEDM) (2018), p. 1.
- ¹⁴³S. Choi, S. H. Tan, Z. Li, Y. Kim, C. Choi, P.-Y. Chen, H. Yeon, S. Yu, and J. Kim, *Nat. Mater.* **17**(4), 335 (2018).
- ¹⁴⁴B. Gao, H. Wu, W. Wu, X. Wang, P. Yao, Y. Xi, W. Zhang, N. Deng, P. Huang, X. Liu, J. Kang, H. Chen, S. Yu, and H. Qian, in IEEE International Electron Devices Meeting (2017), p. 4.4.1.
- ¹⁴⁵S. Stathopoulos, A. Khiat, M. Trapatseli, S. Cortese, A. Serb, I. Valov, and T. Prodromakis, *Sci. Rep.* **7**(1), 17532 (2017).
- ¹⁴⁶D.-Y. Cho, M. Luebben, S. Wiefels, K.-S. Lee, and I. Valov, *ACS Appl. Mater. Interfaces* **9**(22), 19287 (2017).
- ¹⁴⁷I. Giannopoulos, A. Sebastian, M. L. Gallo, V. P. Jonnalagadda, M. Sousa, M. N. Boon, and E. Eleftheriou, in IEEE International Electron Devices Meeting (2018), p. 27.7.1.
- ¹⁴⁸M. A. Lastrasmontaño and K. T. Cheng, *Nat. Electron.* **1**(8), 466 (2018).
- ¹⁴⁹B. Gao, Y. Bi, H.-Y. Chen, R. Liu, P. Huang, B. Chen, L. Liu, X. Liu, S. Yu, H. S. P. Wong, and J. Kang, *ACS Nano* **8**(7), 6998 (2014).
- ¹⁵⁰X. A. Tran, H. Y. Yu, Y. C. Yeo, L. Wu, W. J. Liu, Z. R. Wang, Z. Fang, K. L. Pey, X. W. Sun, A. Y. Du, B. Y. Nguyen, and M. F. Li, *IEEE Electron Device Lett.* **32**(3), 396 (2011).
- ¹⁵¹S. Hsiu-Chuan, C. Ching-Yi, W. Cheng-Wen, C. Lin, and S. Sheu, in VLSI Test Symposium (2011), p. 146.
- ¹⁵²L. Xia, W. Huangfu, T. Tang, X. Yin, K. Chakrabarty, Y. Xie, Y. Wang, and H. Yang, *IEEE J. Emerging Sel. Top. Circuits Syst.* **8**(1), 102 (2018).
- ¹⁵³L. Xia, M. Liu, X. Ning, K. Chakrabarty, and Y. Wang, in Design Automation Conference (2017), p. 26.1.1.
- ¹⁵⁴X. Y. Xue, W. X. Jian, J. G. Yang, F. J. Xiao, G. Chen, X. L. Xu, Y. F. Xie, Y. Y. Lin, R. Huang, Q. T. Zhou, and J. G. Wu, in Symposium on VLSI Circuits (2012), p. 42.