Reconfigurable and Dynamically Transformable In-Cache-MPUF System With True Randomness Based on the SOT-MRAM

Zhengyi Hou, Zhaohao Wang, *Senior Member, IEEE*, Chao Wang, Min Wang, You Wang, *Member, IEEE*, Xueyan Wang, *Member, IEEE*, Cenlin Duan, and Jianlei Yang, *Senior Member, IEEE*

***Abstract*— In this paper, we present a reconfigurable Physically Unclonable Functions (PUF) based on the Spin-Orbit-Torque Magnetic Random-Access Memory (SOT-MRAM), which exploits thermal noise as the true dynamic entropy source. Therefore, the MRAM cells could be configured to random final states with stochastic switching mechanism. The proposed PUF is constructed and reconfigured by combining the small-capacity true random number generator (TRNG) and high-reliability secure hash algorithm (SHA-512), realizing the dynamic transformation between SOT-MRAM based last level cache and PUF (In-Cache-MPUF). Thanks to the full reconfigurability and the high endurance of SOT-MRAM, the proposed In-Cache-MPUF can achieve 10^{14} maximum PUF bits per cell, which has greatly motivated the implementations compared with the traditional weak PUFs utilizing the static entropy source of process variations. The Monte-Carlo simulation results using 40 nm technology and a compact MTJ model show that the proposed PUF has desirable randomness as the digitized bit streams passing all the NIST tests, achieving 50.0428% uniqueness as well as 49.9236% uniformity. It also shows comparable reliability to the state-of-the-art works: a maximum bit error rate of 0.14% and 0.12% at 100 °C and 0.9 V, respectively. In addition, the system level performance is tested and validated by gem5.**

***Index Terms*— Physical unclonable function, SOT-MRAM, reconfigurable, last level cache, dynamic entropy, dynamic transformation.**

I. Introduction

THE era of big data with the information explosion has taken a great challenge to the security of systems in electrical equipment. Numerous researchers are motivated and devoted to inventing new devices with high security and portability [1]. In this circumstance, physical unclonable function (PUF) is widely used to address security issues as the efficient and reliable hardware primitive. The main mechanism of PUF is to take advantage of the random physical variations inherent to any manufacturing process, thus, mapping a set of challenges to a set of responses, which constitute challenge-response pairs (CRPs) [2]. Due to the unpredictability of the subtle physical variations among different devices, the CRPs of the PUFs are highly unpredictable, unique and unclonable [3].

In the past decades, various types of PUFs were proposed. For example, Gassend *et al.* introduced an arbiter PUF (APUF) in 2002 [4], which exploits the time delay difference caused by manufacturing variability between two signal propagation paths as the entropy source. APUF is simple and capable of generating an exponential number of CRPs. Moreover, some variants of APUFs were presented, such as the XOR-APUF [5]–[7] and the feed forward APUF [8]–[10], to increase the complexity for resisting modelling attacks based on machine learning algorithms. Another time delay-based PUF, named ring oscillator PUF (ROPUF), was presented in [6] and further improved in [11], [12]. Besides the aforementioned time delay-based PUFs, a number of PUFs using various entropy sources were proposed, including mismatch-based silicon PUFs, such as latch PUF [13], flip-flop PUF [14], butterfly PUF [15], the leakage current based PUF [16], and analog PUF [17]–[19].

Nowadays, some of PUFs are widely used in commercial products. However, as shown in Fig. 1(a), the implementation of dedicated circuits for PUFs inevitably increases the area and energy consumption [20]. Consequently, the memory-based PUFs (MemPUFs) are widely studied, which utilize the intrinsic characteristics of memory to realize security functions, thus, alleviating the area overhead problem. Holcomb *et al.* introduced an SRAM-based PUF in 2007 [21] and then plenty of SRAM PUFs were proposed in [22]–[24]. In [20], Li *et al.* presented a system-on-chip (SoC) with the capability to transform SRAM in the cache to a PUF. This transformation mechanism between cache and PUF has greatly saved the silicon area and power consumption. Nevertheless, SRAM-based PUFs need to be powered all the time and require an additional non-volatile memory (NVM) to save the secret keys. To solve this issue, various PUFs were designed with NVMs

Manuscript received September 30, 2021; revised January 19, 2022; accepted April 7, 2022. Date of publication April 27, 2022; date of current version June 29, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62171013, Grant 62072019, and Grant 62004011; in part by the National Key Research and Development Program of China under Grant 2021YFB3601303, Grant 2021YFB3601304, and Grant 2021YFB3601300; and in part by the State Key Laboratory of Computer Architecture under Grant CARCH201917. This article was recommended by Associate Editor K. Chen. *(Corresponding author: Zhaohao Wang.)*

Zhengyi Hou, Zhaohao Wang, Min Wang, Xueyan Wang, and Cenlin Duan are with the MIIT Key Laboratory of Spintronics, School of Integrated Circuit Science and Engineering, Fert Beijing Institute, Beihang University, Beijing 100191, China (e-mail: zhengyi_hou@buaa.edu.cn; zhaohao.wang@buaa.edu.cn).

Chao Wang is with the School of Electronics and Information Engineering, Fert Beijing Research Institute, Beihang University, Beijing 100191, China.

You Wang is with the School of Integrated Circuit Science and Engineering, Hefei Innovation Research Institute, Beihang University, Hefei 230013, China.

Jianlei Yang is with the School of Computer Science and Engineering, Fert Beijing Research Institute, Beihang University, Beijing 100191, China.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSI.2022.3168133.

Digital Object Identifier 10.1109/TCSI.2022.3168133

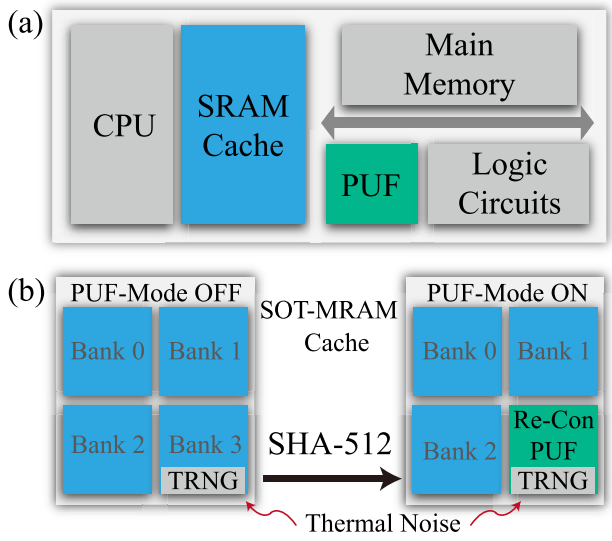

Fig. 1. (a) The normal structure of a system with a dedicated PUF module, (b) the proposed mechanism of In-Cache-MPUF based on SOT-MRAM.

(NV-PUFs), such as phase-change memory (PCM) [25], resistive random-access memory (RRAM) [26]–[28], spin-transfer torque memory (STT-MRAM) [29]–[31], and Spin-Orbit Torque memory (SOT-MRAM) devices [32], [33]. Among these NVM techniques, SOT-MRAM shows a better trade-off among power efficiency, operation speed, endurance, and scalability. Meanwhile, due to its sub-nanosecond switching speed, SOT-MRAM is a promising candidate for the next generation of cache by replacing SRAM [34]–[37]. The NV-PUFs feature much larger intrinsic stochasticity, and excel in terms of the area efficiency and native reliability. However, most of the NV-PUFs are limited by their finite CRP space, since the number of CRPs is linear with that of NVM cells, which is a common feature of MemPUFs.

On the other hand, all the above-mentioned PUFs utilize physical variations derived from semiconductor manufacturing as their entropy sources, which are well known as static entropy sources (SES). Several drawbacks of the SES-based PUF implementations are pointed out in [27]. Most importantly, the PUF with fixed entropy source lacks reconfigurability. This is unsafe when the PUFs suffer from owner changes or repeated uses in different environments. Hence, Ref. [26] proposed an RRAM PUF based on the 'post-processing randomness' to realize the reconfigurability. However, the randomness of the presented design has not been rigorously verified. To further solve the issues, Ref. [27] proposed a novel reconfigurable RRAM PUF based on the true random dynamic entropy source (DES) of jitter noise, which applies a ring oscillator true random number generator (RO-TRNG) to configure the RRAM cells by corresponding resistance states. This mechanism has increased the area of the whole system due to the dedicated TRNG. Furthermore, the RRAM cells chosen to be PUF would be occupied before reconfigured. As a result, once the PUF cells are reset to memory function, it would be impossible to establish the same PUF as the previous version, because of the real-time variation of the employed RO TRNG. Therefore, this approach is actually unable to realize the dynamic transformation between memory and PUF, and may dramatically degrade the performance of RRAM as an NVM.

To improve area inefficiency and reduce the impact of MemPUFs on memory performance, we propose a SOT-MRAM-based PUF (SOT-MPUF) utilizing the true DES of thermal noise. Moreover, a technique called the In-Cache-MPUF is presented, which employs part of the SOT-MRAM cells as TRNGs to realize the dynamic transformation between the SOT-MRAM and a reconfigurable PUF in the last-level cache (LLC). Thanks to the easily-controlled stochastic switching characteristic of SOT-MRAM, the TRNG could be realized inside the NVM for area saving without dedicated circuits proposed in [27]. Meanwhile, as shown in Fig. 1(b), the secure hash algorithm (SHA-512) [38], which converts a message of arbitrary length into a fixed 512 bits of message digest in ways of unidirectional mapping, is applied to realize the dynamic transformation. Meanwhile, this proposal can decrease the impact on memory performance of occupying the MRAM cells as PUF, by shrinking the large PUF space to limited TRNG space. Concretely, the proposed In-Cache-MPUF has three main features.

1. The SOT-MPUF is operated by configuring the MRAM cells to either '0' (low resistance state, LRS) or '1' (high resistance state, HRS) according to the random bits generated by the SOT-MRAM-based TRNG, and could be easily operated between the memory and PUF modes just by adjusting the driving currents.
2. The SHA-512 is utilized for mapping the shorter TRNG random bits to the longer PUF cells, thus, realizing the dynamic transformation between SOT-MRAM-based LLC and PUF.
3. The proposed MPUF is reconfigurable by refreshing the TRNG bits.

Due to the stochastic switching characteristic caused by the true random thermal noise, the proposed SOT-MPUF shows desirable randomness with the digitized bit streams passing all the NIST (the U.S. National Institute of Standards and Technology) tests [39], [40], achieving 50.0428% uniqueness as well as 49.9236% uniformity. The widely-exploited autocorrelation function (ACF) test validates the great randomness [41]. It also shows reliability comparable to the state-of-the-art work: a maximum bit error rate of 0.14% and 0.12% at 100 °C and 0.9 V, respectively. The In-Cache-MPUF after mapping by the SHA-512 also shows promising performance with negligible area and energy consumption. In addition, the influence of this novel technique on the whole system is tested using gem5 [42] and SPEC CPU 2017. The simulation results indicate that, compared with the normal SOT-MPUF, the proposed In-Cache-MPUF could improve the system performance by more than 30%.

The rest of this paper is organized as follows. Section II is a preliminary about the mechanisms utilized in this work, including both the deterministic and stochastic switching mechanisms of the SOT device, and the normal SOT-MPUF. The proposed In-Cache-MPUF is introduced in detail in Section III. The experimental results are presented, discussed,

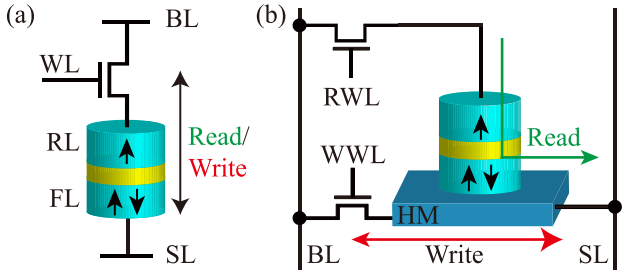

Fig. 2. Standard bit-cells of (a) STT-MRAM, and (b) SOT-MRAM.

and compared with the state-of-the-art implementations in Section IV. Finally, Section V concludes this work.

II. Preliminaries

In this section, we introduce the mechanisms of the deterministic and stochastic switching of SOT device, thus, constructing the proposed normal SOT-MPUF.

A. The Deterministic and Stochastic Switching Mechanisms of the SOT Device

Two types of basic MRAM cells are shown in Fig. 2, referring to the STT-MRAM and SOT-MRAM, respectively. The typical STT-MRAM cell (see Fig. 2(a)) consists of 1 NMOS transistor and 1 magnetic tunnel junction (MTJ). The core structure of the MTJ is composed of an ultra-thin oxide layer sandwiched between a free layer (FL) and a reference layer (RL). The relative orientation between the magnetization vectors of FL and RL (i.e. parallel or antiparallel, P or AP) represents the resistance state (i.e. LRS or HRS) and hence the stored data (i.e. 0 or 1). As a promising candidate for the next-generation memory technology, STT-MRAM has the advantage of non-volatility, high integration density, and high read speed. However, an incubation delay during STT switching limits its switching speed. In addition, read and write paths are coupled in an STT-MTJ, leading to difficulty in addressing the read disturb. In order to overcome these two drawbacks, the emerging SOT-MTJ has been widely investigated. A heavy metal (HM) layer carrying the write current is contacted to the FL of the MTJ, thus, the read and write paths of SOT-MTJ are separated, as shown in Fig. 2(b). Besides, the SOT induced current could achieve sub-nanosecond writing in a perpendicular-anisotropy MTJ.

In the conventional SOT scheme, an additional magnetic field is needed to achieve the deterministic writing. Recently, plenty of field-free SOT writing schemes were proposed. Among them, the toggle spin torques (TST) scheme, a SOT-assisted STT writing mechanism, has attracted much attention due to its ultra-low consumption and ultra-fast switching, which has been theoretically and experimentally validated in the recent studies [43]–[47]. As shown in Fig. 3(a)(b), two current pulses (I_{STT} and I_{SOT}) are sequentially applied to the device in the TST scheme. According to the direction of the current I_{STT} flowing through the MTJ device, the operations of write 1 (Fig. 3(a)) and write 0 (Fig. 3(b)) are unambiguously completed. The magnetization dynamics of the FL can be described by the modified Landau–Lifshitz–Gilbert (LLG) equation:

$$\begin{aligned}\frac{\partial \vec{m}}{\partial t} = &-\gamma\,\mu_0 \vec{m} \times \vec{H}_{eff} + \alpha \vec{m} \times \frac{\partial \vec{m}}{\partial t} - J_{STT}\xi P \vec{m} \\ &\times (\vec{m} \times \vec{m}_r) - J_{SOT}\xi\lambda_{DL}\vec{m} \times (\vec{m} \times \vec{\sigma}_{SOT}) \\ &- J_{SOT}\xi\lambda_{FL}\vec{m} \times \vec{\sigma}_{SOT} \end{aligned} \quad (1)$$

where the last three terms on the right side are STT, damping-like SOT, and filed-like SOT, respectively. $\vec{m}$ is the unit vector along the magnetization orientation of the FL, and $\vec{m}_r$ is the STT polarization vector. J_{STT} and J_{SOT} are the STT and SOT write current densities, respectively. $\vec{\sigma}_{SOT}$ is the SOT-induced spin polarization, and the other parameters could be found in [48].

TABLE I
Parameters of the Applied MTJ Device

Parameter	Mean (μ)	$3\sigma/\mu$
MTJ area	$40 \times 40 \times \pi/4\ nm^2$	/
Heavy metal dimension	$60 \times 40 \times 3\ nm^3$	/
MTJ free layer thickness	0.7 nm	0.03
MTJ oxide layer thickness	0.85 nm	0.03
Perpendicular magnetic anisotropy	1×10^5 A/m	/
Gilbert damping constant	0.05	/
Saturation magnetization	1.1×10^6 A/m	/
Spin Hall angle	0.3	/
TMR	120%	0.03

Specifically, hybrid CMOS/TST-MRAM simulations are conducted by applying 40 nm CMOS process technology and a SOT-MTJ compact model proposed in [48]. The MTJ parameters used for the simulations are shown in Table I. As illustrated in Fig. 3(c)(d), we have performed 1000 Monte-Carlo simulations to verify the deterministic writing of the TST mechanism within a 64×16 bits MRAM array. The typical 10 curves are displayed in each sub-figure for the clarity. The current I_{SOT} ($\sim$100 μA) is first applied to the HM layer for driving the magnetization of the MTJ FL to the in-plane direction ($m_z \approx 0$). Then, the I_{STT} current ($\sim$20 μA) is applied longitudinally, and the polarity of the written data is determined by the direction of the STT current. The results show that the writing delay of SOT-MTJ is close to 1 ns in agreement with the experimental measurement [45], and there is no write error in 1000 Monte-Carlo simulations (not shown completely in Fig. 3(c)). Therefore, the TST mechanism effectively decreases the writing delay without increasing any manufacturing complexity, also guarantees a high writing reliability. Fig. 3(e) illustrates the read operation of the SOT-MTJ, which requires a small current I_{read} ($\sim$15 μA) flowing through the device vertically.

In addition, the SOT-MTJ shows a very stable stochastic switching characteristic, which has a wide range of application prospects in hardware encryption. As shown in Fig. 3(f), a single current I_{SOT} is applied to the HM layer, and the spin accumulation induces the SOT effect and drives $\vec{m}$ to the in-plane direction. In this case, the magnetization dynamics

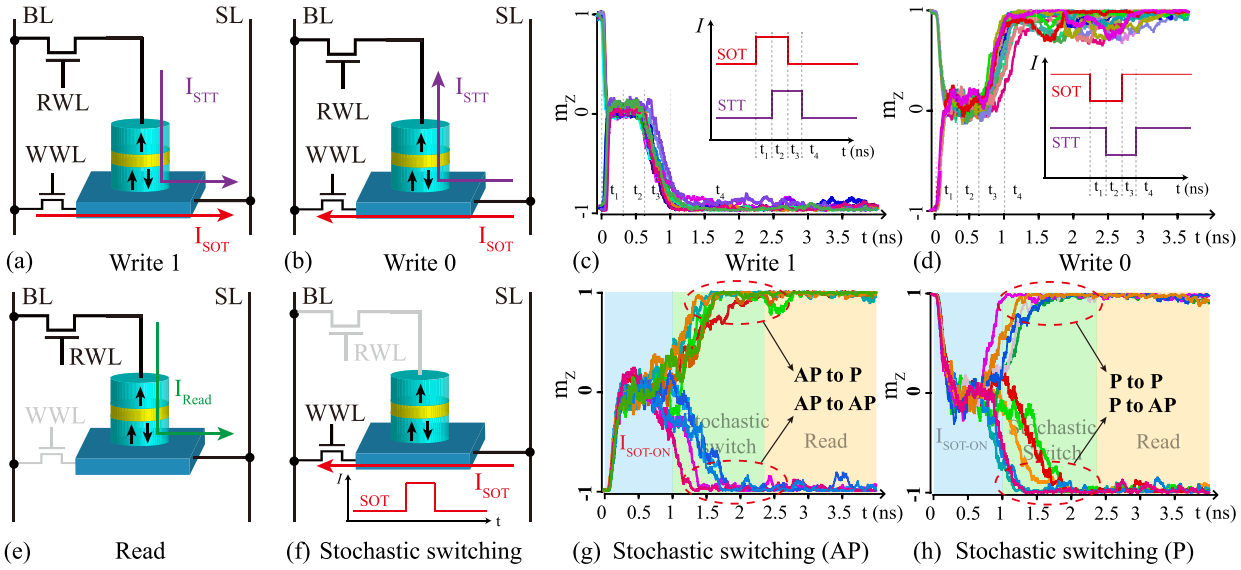

Fig. 3. Four basic operations of SOT-MRAM and the corresponding time evolution of z-component magnetization (m_z) under various initial states. (a)(b) The operations of writing 1 and writing 0, respectively; (c)(d) the Monte-Carlo simulation results corresponding to these two deterministic writing operations; (e) the operation of reading, (f) the operation of stochastic switching, and (g)(h) the Monte-Carlo simulation results of stochastic switching with the initial states as AP and P, respectively.

can be expressed by removing the STT term in (1). After the turn-off of I_{SOT}, the magnetization will be unstable under the disturbing of thermal noise and be stochastically switched to one final state, either up or down. The thermal noise related to the system temperature T is described by a Langevin random field $\vec{H}_{thermal}$ as (2) and added to the $\vec{H}_{eff}$ [49].

$$\vec{H}_{thermal} = \vec{\sigma}_i(t)\sqrt{\frac{2k_B T\alpha}{\gamma\,{\mu_0}^2 M_s V\,\Delta_t}}(i = x, y, z) \tag{2}$$

where $\vec{\sigma}(t)$ is a unit coefficient vector which obeys the Gaussian distribution with zero mean and unit standard deviation. Each of the x, y, z components has its own uncorrelated $\vec{\sigma}_i(t)$. k_B is the Boltzmann constant, V is the volume of the FL, and Δ_t is the simulation time-step.

It is worth noting that, when the I_{SOT} current is larger than the critical current (I_c, <100 μA with the applied device parameters), the switching probability of the FL for SOT-MTJ is stably close to 50% and resilient with the variation of temperature or pulse duration, which has been experimentally validated [33], [50]–[52]. However, for STT-MTJ, maintaining a 50% switching probability requires the extremely accurate control of the current and temperature [53].

For the stochastic switching characteristics of SOT-MRAM, we performed numerous Monte-Carlo simulations. Fig. 3(g) and (h) are the random switching curves of z-component m_z when the initial states are the '-1' and '1' states, respectively (similarly, we randomly selected 10 of the 1000 Monte-Carlo simulation curves). The simulation results show that the switching probability is always 50% in both cases consistent with the experimental data [33], [50], [51]. Moreover, as illustrated in Fig. 4, we tested the probability of random switching at different temperatures. With the temperature ranging from 230 K to 390 K, the switching probability still keeps close to 50%. We also test the influence of transient

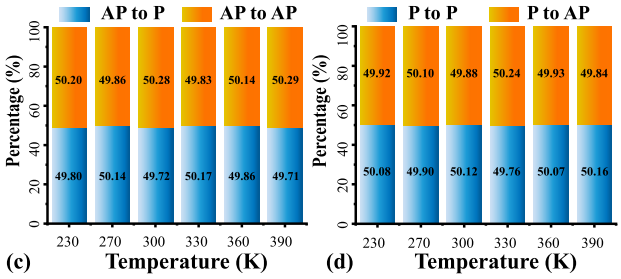

Fig. 4. Percentages of 0/1 after stochastic switching at different temperatures, with (a) AP and (b) P as the initial states.

noise ranging from 1-5 GHz on the switching probability, which exhibits maximum 0.39% deviation from 50% at 2 GHz. In brief, the stable switching probability makes SOT-MRAM an ideal candidate to generate true random numbers or establish a PUF.

B. The Normal SOT-MRAM Based Reconfigurable PUF

Most of the existing PUF designs use SES as the entropy source, which causes a series of problems [27]. In this article, we apply the true DES of thermal noise to achieve the completely stochastic switching of SOT-MPUF. As illustrated in Fig. 5(a), the proposed SOT-MPUF design maintains the normal structure of SOT-MRAM without other peripheral circuits but contains two operation modes: memory mode and PUF mode. For the memory mode, the deterministic writing operation adopts the TST method, which can significantly decrease the switching delay of MTJ devices [45], [47]. As for the read operation, the schematic of the applied two-stage read circuit is illustrated in Fig. 5(b). The first stage is a reference generator as shown in the right part. Two MTJs with opposite states (R_{AP} and R_P) are connected in parallel as a reference

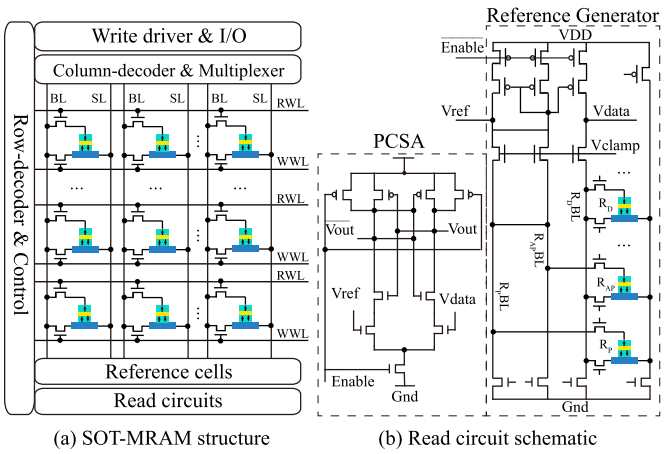

Fig. 5. (a) The normal SOT-MRAM structure, and (b) the utilized read circuits with a PCSA.

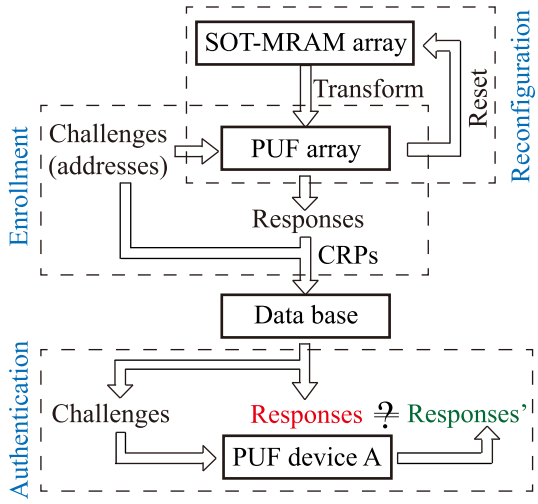

Fig. 6. The flowchart of SOT-MPUF execution.

cell for comparison with the data cell (R_D). A clamping voltage (V_{clamp}) is applied on both the data branch and the reference branch to guarantee the same initial voltage in the sensing paths. Meanwhile, by connecting the gates of the three loading transistors to the node of the reference voltage (V_{ref}), the current difference between the data cell and the reference cells is converted into a voltage difference between V_{ref} and V_{data}. Since the sensing voltage difference is usually too small and difficult to be read, the pre-charge sense amplifier (PCSA) is used here to amplify the difference, as shown in the left part of Fig. 5(b).

On the other hand, when the SOT-MRAM array is required to work in the PUF mode, the transformation from memory to PUF is implemented in two steps.

1. Initializing. Set all the RWLs and BLs to 0, while WWLs and SLs to 1. The SOT current larger than I_c flows through the HM layer of all the MTJ cells to switch the magnetization of FL to in-plane orientation, where the MTJs are in unstable states.

2. Stochastic switching. Turn off the SOT current by setting the WWLs and SLs to 0. Without the SOT current, all the MTJs will be randomly switched to a stable final state under the disturbance of the thermal noise. Finally, the SOT-MRAM array is completely transformed to a PUF array which can be utilized for security applications.

In addition, a reconfigured new PUF can be easily generated by repeating the above-mentioned two steps when all the CRPs have been used or the owner changed. Compared with the RRAM-based reconfigurable PUF [27], no external TRNG is needed to configure the PUF in this proposal due to the stochastic switching mechanism of SOT device. Thanks to the full reconfigurability of the SOT-MPUF and the time-variant property of the dynamic entropy of thermal noise, the maximum PUF bit per cell can reach 10^{14} (i.e., the endurance of the SOT-MRAM devices) [27], [54].

Fig. 6 shows the specific execution process for device authentication of the SOT-MPUF. After the SOT-MRAM array is transformed to a PUF array through the aforementioned method, the new SOT-MPUF must go through the enrollment process firstly. Possible challenges are applied to the PUF in turn, generating corresponding responses. All generated CRPs are saved in the database and the enrollment is completed. While performing the device authentication, a group of CRPs in the database are selected and the challenges are applied to the device A to be verified in turn. If the responses generated by device A match the responses registered in the PUF database, the device authentication succeeds, and vice versa.

III. In-Cache-MPUF Design and the Dynamic Transformation Mechanism

In Section II-B, we introduced the normal SOT-MPUF which utilized the thermal noise as its DES. SOT-MPUF can significantly increase the CRP space of memory-based weak PUF and improve the security due to its reconfigurability. Compared with the PUF design in [27], SOT-MPUF effectively saves the area serving for TRNG. Benefiting from the stochastic switching characteristics of SOT devices, each SOT-MTJ unit can work as a TRNG, and further constitute this reconfigurable PUF. However, this design faces the same two main shortcomings as [27]. Firstly, this design cannot achieve the dynamic transformation between PUF and memory. After the transformation from PUF to memory, it is impossible to rebuild the same PUF as the previous version because of the real-time variation of the thermal noise. Secondly, SOT-MPUF is still an independent PUF module in a computing system, which requires additional area consumption. In order to overcome both issues, we propose a method to implement reconfigurable PUF in the SOT-MRAM-based LLC, which can realize the dynamic transformation between PUF and SOT-MRAM.

As mentioned above, due to the high power efficiency and fast switching speed, SOT-MRAM has been considered as the emerging candidate for the next-generation cache. Plenty of SOT-MRAM-based cache designs were presented in the literatures [35]–[37]. In general, the caches usually consist of various banks, which are fully-functional memory units and operate independently. The capacity of the LLC is usually 256 KB - 32 MB, the requirement of a PUF array

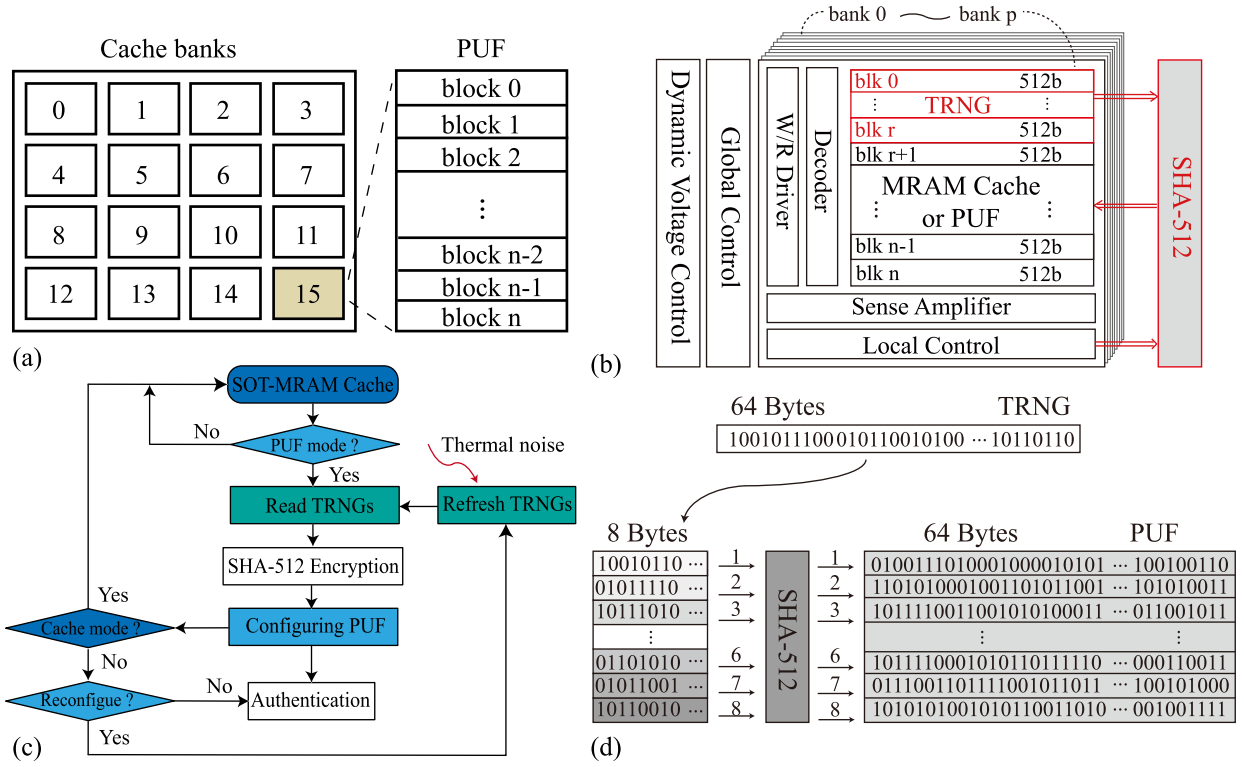

Fig. 7. The In-Cache-MPUF mechanism. (a) Use a complete bank in the cache as PUF, (b) the structure of the proposed In-Cache-MPUF utilize SHA-512 algorithm, (c) the specific execution flow of the proposed In-Cache-MPUF, and (d) an example of extending 64 bytes of TRNs to 8 × 64 bytes of TRNs through the SHA-512.

is much smaller. Hence, part of the cache can be selected to work as a PUF, while the other partitions can still work as the cache.

As shown in Fig. 7(a), the latest one of 16 banks in a cache is selected to work as the PUF. The scheme can be realized by only changing the control pattern without additional peripheral circuits, and the transformation is the same as SOT-MPUF mentioned in Section II-B. If the PUF mode is activated, all the MTJ cells in bank 15 will be selected and initialized by throwing I_{SOT} current. Then, the SOT-MTJ array with stochastic states will be used as a PUF for security operations. Meanwhile, the channel connecting CPU and bank 15 in the SOT-MRAM cache will be locked, for guaranteeing the exclusive use of bank 15 as the PUF. This method could save the area of a dedicated PUF module by implementing the PUF in the ubiquitous cache. However, it is still unable to realize the dynamic transformation between PUF and cache. PUF partitions will continuously occupy the memory capacity, which may degrade the cache performance.

Hence, a more flexible and area-efficient method is presented to perform the In-Cache-MPUF. As shown in Fig. 7(b), we use r rows of memory cells in the bank p of LLC as TRNGs, and utilize SHA-512 to encrypt and expand the generated true random number sequence (TRNs), thus, turning short random number sequences (SRNs) into long random number sequences (LRNs). Then, LRNs are assigned to the rest of the bank p, forming the PUF. In addition, since TRNGs are also composed of SOT-MRAM cells, the TRNs could be preserved for a long time. Therefore, in addition to TRNG, other parts of bank p can be dynamically switched between PUF and memory modes. Specifically, the whole working process is illustrated in Fig. 7(c) and a detailed description is as follows:

1. Generation of TRNs. When a cache bank works in PUF mode for the first time, part of the memory lines in the PUF bank are selected as TRNGs. The size of TRNGs should be smaller than the PUF bank, and determined by the application scenarios. Then, SOT currents are applied to the MTJ cells for performing the stochastic switching. Consequently, the random resistance values corresponding to TRNG units are read as the generated TRNs. Finally, the states of the TRNG units remain unchanged until requested by the system.

2. Extension of random numbers. The generated TRNs is divided into n segments (the bit-length of TRNs segments can be defined by customers and must be shorter than 512 bits), which are successively passed to the hash function SHA-512, and then are encrypted to output for n segments of 512-bits encrypted data. The encryption step has realized the extension of TRNs to PUF array. In an example shown in Fig. 7(d), a 64 bytes TRNs is divided into 8 segments, and then be extended to 8 numbers of 64 bytes random bit streams by the encryption of SHA-512.

3. Transformation between cache and PUF. The encrypted data are written to the least of cells in the PUF bank in turn, until all PUF cells are assigned. It is necessary to ensure that the number of TRNs segments meets the requirement that

the encrypted data can be assigned to the whole PUF bank. Hence, the transformation from the cache to PUF is realized. As for transformation back from PUF to cache, the data of all PUF cells are cleared, while the TRNGs remain unchanged. Because of the utilization of SHA-512, the saved TRNs can be extended to the same PUF array by the same strategy.

4. Reconfiguration of PUF. When the PUF is no longer secure or needs to be reconfigured, the TRNGs are refreshed through the operation of stochastic switching to generate new TRNs. Consequently, a reconfigured PUF could be formed by repeating the steps 2 and 3 with the new TRNs.

In summary, the above design could be realized with the normal structure of SOT-MRAM-based cache without additional dedicated circuits, which eliminates the area overhead of the dedicated PUF module by implementing PUF in cache. Meanwhile, the dynamic transformation between PUF and cache is realized through the use of SHA-512 algorithm and TRNG.

IV. Simulation Set-Up and Performance Analysis

To evaluate the proposed SOT-MPUF, we perform the simulation in two aspects. On the one hand, we implement a 64 × 16 bits SOT-MPUF array based on the 40 nm CMOS process design kit (PDK) and a SOT-MTJ compact model [48] with thermal noise considered. The performance metrics of the SOT-MPUF, including randomness, uniqueness and reliability, are researched in detail with more than 10000 Monte-Carlo simulations and 10M bit streams. The well-known NIST SP800-22 and NIST SP800-90B as well as ACF tests are employed to verify the performance of SOT-MPUF. On the other hand, we perform system-level simulation verification and performance analysis of the proposed In-Cache-MPUF using gem5 simulator and SPEC CPU 2017 benchmark.

A. Basic Performance of SOT-MRAM Based PUF

1) Randomness: As described in Section II, the SOT-MTJ has a stable switching probability of 50% under the action of thermal noise, which is irrelevant to the environment temperature and indicates the good randomness for SOT-MPUF. In this section, the uniformity is tested with 1000 number of 256-bits PUF responses, which is a primary performance metric representing the distribution of logic '0' and '1' in the PUF responses. The uniformity is defined as follows:

$$Uniformity = \frac{1}{n}\sum_{j=1}^{n} r_{i,j} \times 100\% \tag{3}$$

where $r_{i,j}$ is the j-th binary bit of an n-bits response vector i. As illustrated in Fig. 8, the distribution of uniformity is well-fitted by a Gaussian curve with the mean value $\mu = 50.10\%$ and the standard deviation $\sigma = 0.0258$, respectively. The testing results which are very close to the ideal value of 50% indicate that the SOT-MPUF has good randomness.

Meanwhile, the widely employed NIST test is exploited to evaluate the randomness of the bit stream generated by SOT-MPUF. Specifically, 1 MB data collected from the aforesaid Monte-Carlo simulations is fed into the NIST SP800-22 test suite [39] as 55 separate bit streams. As summarized in Table II, results of listed 13 NIST test items are passed with high proportion (where 52/53 is required) and P-Value > 0.01, which well verifies the true randomness of the SOT-MPUF with dynamic entropy from thermal noise. With respect to the NIST SP800-90B test [40], as shown in Table III, the min-entropy of the dynamic random thermal noise is tested as 0.993735, the proposed SOT-MPUF passed all the test items.

Moreover, the evaluation is further extended to the ACF test [41], which is another well-known mathematical function to inform how the responses are biased. 10M random bits are collected through the Monte Carlo simulation of the SOT-MRAM to perform the ACF test, and the results are

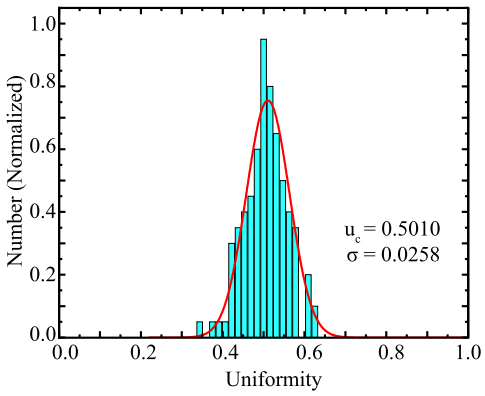

Fig. 8. The uniformity distribution of the SOT-MPUF.

TABLE II
NIST SP800-22 Test Results

Test	P-value	Proportion	Pass ?
Frequency	0.071177	53/55	YES
BlockFrequency	0.224821	52/55	YES
CumulativeSums	0.534146	55/55	YES
Runs	0.011883	53/55	YES
LongestRun	0.275709	53/55	YES
Rank	0.213309	53/55	YES
FFT	0.010988	52/55	YES
NonOverlappingTemplate	0.122325	55/55	YES
OverlappingTemplate	0.035174	55/55	YES
ApproximateEntropy	0.012199	54/55	YES
Serial	0.181557	53/55	YES
LinearComplexity	0.350485	54/55	YES

TABLE III
NIST SP800-90B Test Results

Num. of binary samples	1M bits
Bit-wide per symbol	7
H_original	0.993735
H_bitstring	0.635028
Min-entropy	0.993735
Chi square tests	Pass
Chi square independence	Pass (Score=0.7809, Dof=-2)
Chi square goodness of fit	Pass (Score=8.1643, Dof=9)

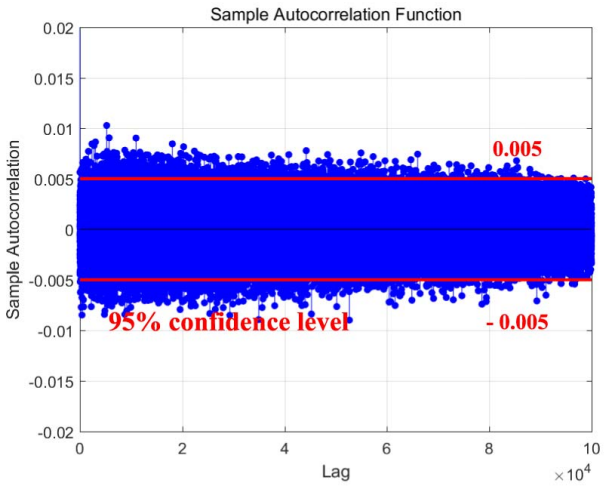

Fig. 9. ACF test results for 10M bit streams.

shown in Fig. 9. It is observed that within 95% confidence bound of a Gaussian distribution, the ACF value only fluctuates between -0.005 and 0.005, which is comparable to the state-of-the-art works [27], [30].

2) Uniqueness: Uniqueness demonstrates the difference of responses generated from the same challenges on different PUF instances, which is qualified by inter-Hamming distance HD_{Inter}. The average HD_{Inter} is defined as (4):

$$HD_{Inter} = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \frac{HD(r_i, r_j)}{n} \tag{4}$$

where r_i and r_j are n-bits responses of same challenges for different k PUF instances. The ideal value of HD_{inter} is 50%.

Additionally, the diffuseness is measured by calculating the mean Hamming distance (HD) of different responses generated by the same PUF for different challenges. The ideal value of diffuseness is 50%. The diffuseness is defined as (5):

$$Diffuseness = \frac{1}{d(d-1)} \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} \frac{HD(r_i, r_j)}{n} \times 100\% \tag{5}$$

where d is the number of randomly selected responses, r_i and r_j are two different n-bits responses in the d responses.

The simulation results of uniqueness and diffuseness are illustrated in Fig. 10. The solid lines in red are the fitting lines, which follow the normal distribution. The mean (μ_c) and variance (σ) of uniqueness and diffuseness are 0.5003, 0.0352, and 0.4975, 0.0286, respectively. According to (4) and (5), we can calculate the value of uniqueness and diffuseness, which are 50.0428% and 49.9236%, respectively. Both values are very close to their ideal values of 50%. In addition, to show full reconfigurability of the proposed SOT-MPUF, we compared 1000 responses generated from SOT-MPUFs before and after the reconfiguration of one MRAM device. The HD calculation results follow the normal distribution with a mean HD equal to 0.50282. The results indicate that the proposed SOT-MPUF has a great reconfigurability and that

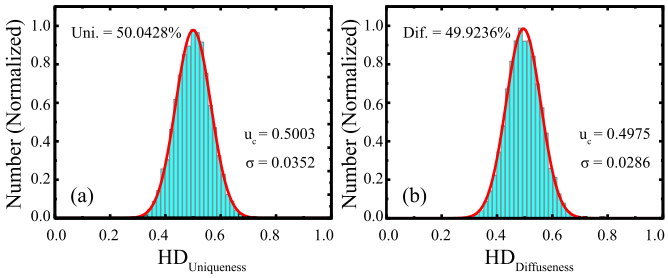

Fig. 10. Distribution of the uniqueness and diffuseness for the SOT-MPUF.

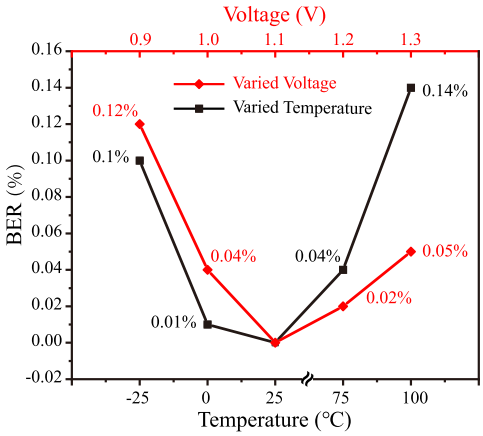

Fig. 11. BER of the SOT-MPUF with the temperature ranging from −25 °C to 100 °C (black), and the voltage ranging from 0.9 V to 1.3 V (red).

the responses generated before and after reconfiguration are almost independent.

3) Reliability: In this subsection, we investigate the reliability of the proposed PUF under various environmental temperature and supply voltage conditions, which is crucial in designing a PUF to perform a cryptographic algorithm. The PUF is required to generate constant responses under temperature and voltage variations. Under the nominal condition (25 °C temperature and 1.1 V supply voltage), no read error occurs through 1000 Monte-Carlo simulations of the repeated readout of 100 challenges and responses, indicating that the native unstable bits rate is less than 0.001%. Then the reliability or BER with environmental variations is defined as follows:

$$BER = 1 - Reliability = \frac{1}{k} \sum_{j=1}^{k} \frac{HD(r_{i,0}, r_{i,j})}{n} \times 100\% \tag{6}$$

where $r_{i,0}$ is the n-bits response of an arbitrary challenge under the nominal condition, then the same challenge is applied k times to the same PUF instance under different operation conditions to obtain k responses, $r_{i,j}$ for $j = 1, 2, \ldots, k$.

As shown in Fig. 11, BER is measured with the operating temperature varying from −25 °C to 100 °C at 1.1 V and the supply voltage varying from 0.9 V to 1.3 V at 25 °C. The worst-case BERs are measured as 0.14% at −25 °C and 0.12% at 0.9 V. As a result, the averaged BER per 10 °C and BER per 0.1 V are lower than 0.02% and 0.06%, respectively. The

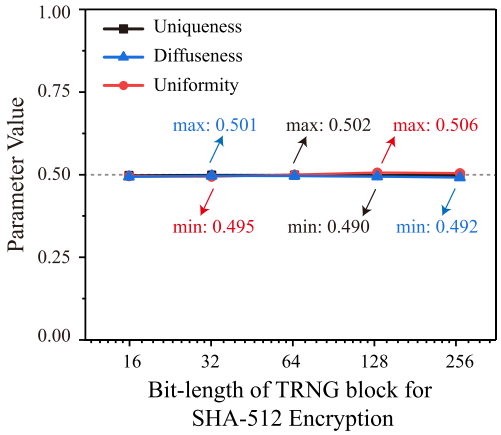

Fig. 12. Three performance metrics of the proposed In-Cache-MPUF with various bit lengths of TRNG blocks.

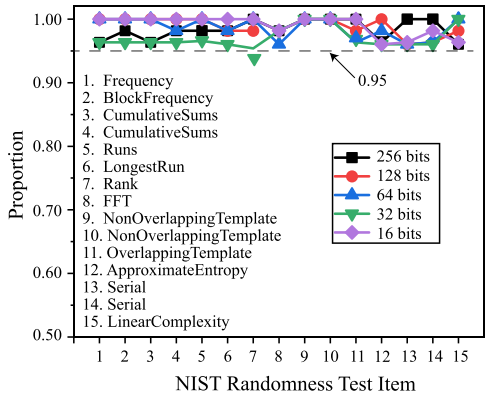

Fig. 13. NIST test results of the bit streams after encryption of different bit lengths of TRNG blocks.

reported ultra low unstable bits and BER are attributed to the unique TST mechanism of SOT-MRAM with the separated read and write paths, which can greatly reduce the impact of leakage current and read current on cell data, thereby minimizing the probability of false-flipping of MTJ states.

B. Performance of the Proposed In-Cache-MPUF

1) Area and Energy: A detailed study to explore the SOT-MRAM based cache design space using NVSim [42] is conducted in this section. The inner cell area and energy for the simulation are shown in Table IV, which are the results of circuit level simulation. As shown in Fig. 14, we compared the SOT-MRAM with SRAM and STT-MRAM for different capacities at the 40 nm technology node. When the capacity is greater than 512 KB, the area of SRAM is no longer advantageous compared to SOT-MRAM and STT-MRAM because of its larger number of transistors. In terms of write power consumption, SOT-MRAM exhibits extremely high advantages as the TST mechanism requires extremely low writing energy. Regardless of its capacity, its write power consumption is always lower than the other two types of memories. In summary, SOT-MRAM outperforms than SRAM and STT-MRAM in area and write energy, and also comparable to SRAM in

TABLE IV
CIRCUIT LEVEL PARAMETERS USED IN SIMULATIONS

Parameter	STT-MRAM	SOT-MRAM
$S(F^2)$	40.33	85
$P_{write}(fJ)$	573 for '1' to '0' 648 for '0' to '1'	54 for '1' to '0' 67.2 for '0' to '1'

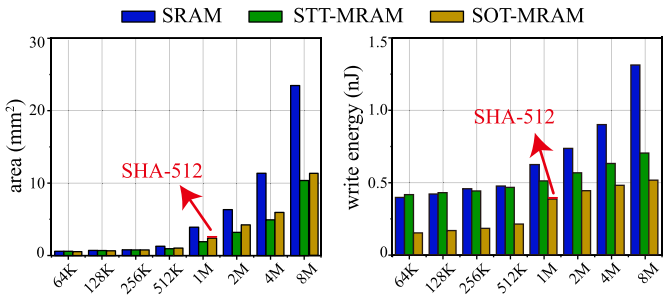

Fig. 14. Comparisons of performance among SRAM, STT-MRAM, and SOT-MRAM.

read and write speed, which proves the great potential of SOT-MRAM as an LLC.

Considering the utilization of SHA-512, we conduct an SHA-512 using 40 nm CMOS design kit and EDA tools, its area consumption is about 0.03 mm^2 and energy costs 3.38 pJ. While comparing with the 1 MB SOT-MRAM array of LLC, whose area is larger than 2.4 mm^2 and energy is about 0.4 nJ according to the NVSim results, the SHA-512 only occupied 1/80 and 1/118 of cache area and energy consumption as illustrated in Fig. 14, respectively. Meanwhile, the SHA-512 module could be reused in some cryptographic scenarios, which may decrease the waste of resources inside the system.

2) Basic Performance: According to our proposal, the bit stream generated by TRNG is divided into n segments of the m-bits data stream. To investigate the relationship between the length of bit stream (m) before encryption and the performance of PUF configured by the enhanced data, we measured the uniformity, uniqueness, and diffuseness under the various values of m ranging from 16 to 256. As illustrated in Fig. 12, the minimum and maximum of the three parameters are 0.495, 0.490, 0.492 and 0.506, 0.502, 0.501, respectively. The results indicate that there is no significant correlation between m and the performance of encrypted PUF. Furthermore, we also carried out the NIST test with the different conditions of various m values. The outcome presented in Fig. 13 indicates that the NIST test passes with the p-value > 0.01 and a high proportion larger than 95% for all the values of m. This means that the PUF established by the TRNG and SHA-512 function has great reliability and randomness.

3) System-Level Simulation Analysis: To further investigate the performance of the system with In-Cache-MPUF, a system-level simulation has been taken through gem5 and SPEC CPU 2017. We have tested the 1 MB LLC with 8 banks and 4 banks, corresponding to the different proportions of PUF as 1/8 and 1/4 of LLC. The two comparison groups are shown in Fig. 15(a) and Fig. 15(b), respectively. 20 different

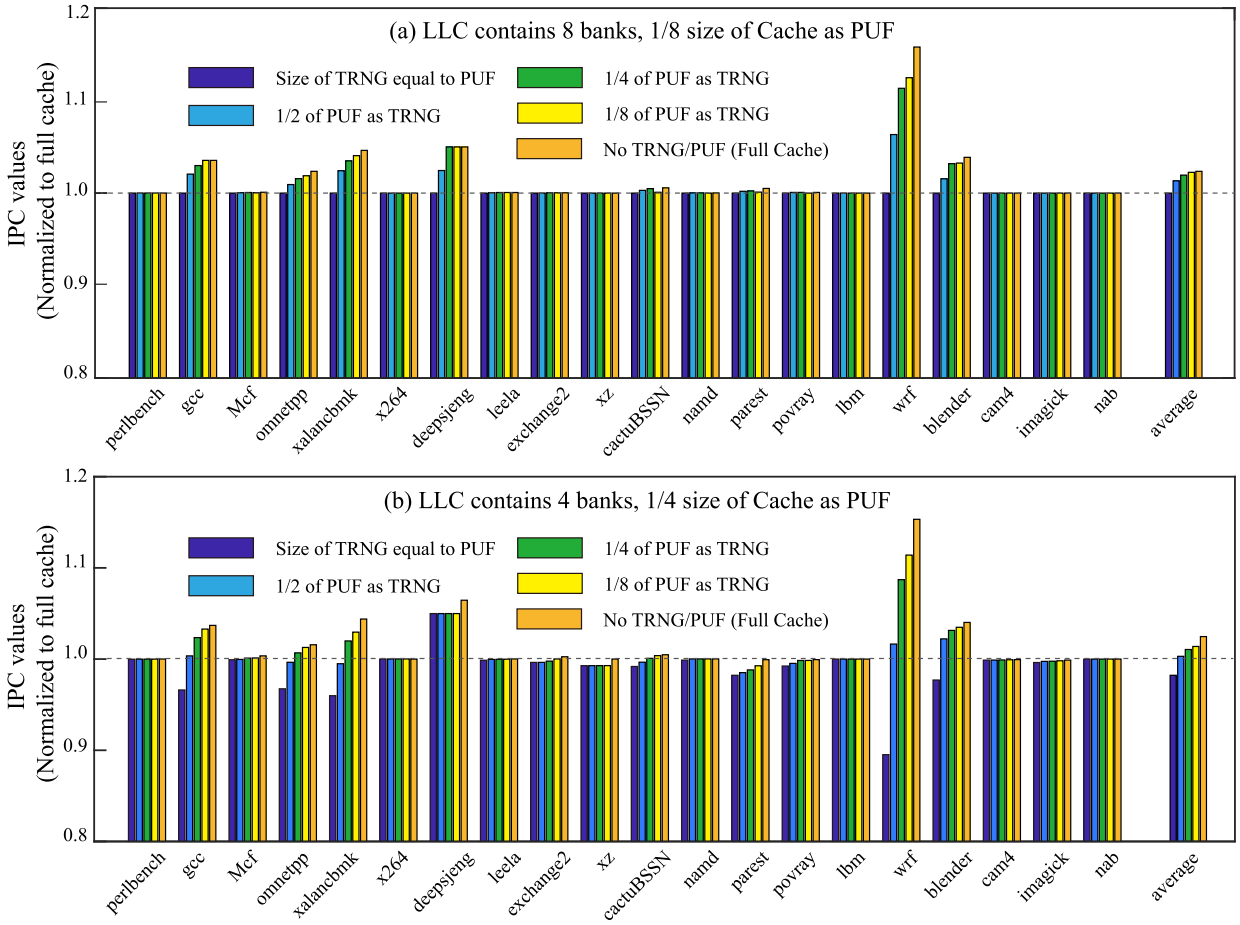

Fig. 15. The gem5 simulation results of 1 MB SOT-MRAM-based LLC with (a) 8 banks while 1/8 as PUF, and (b) 4 banks while 1/4 as PUF.

TABLE V
THE CMP ARCHITECTURE CONFIGURATION

Processor	4-core @ 3.3 GHz, out-of-order, ARM
L1 Cache	I-cache 32 KBytes 8-way set associate
	D-cache 32 KBytes 8-way set associate
L2 Cache	1 MBytes 8-way set associate, shared
	16 bits width, 8 cacheline entries write buffer
	MOESI cache coherence protocol
Main Memory	4 GBytes DDR3 DRAM
Benchmark	Spec CPU 2017

benchmarks of SPEC CPU 2017 are used in simulations. Meanwhile, five different capacity situations of TRNG are analyzed, which are equal to 100%, 1/2, 1/4, and 1/8 of PUF along with no PUF in LLC. As illustrated in Fig. 15, most of the histogram sets increase with the decrease of TRNG capacity, which means that the system performance is better when the TRNG occupies less space of LLC. Especially, in the test of 'wrf', the instructions per cycle (IPC), which reflects the system performance, in the case of no TRNG/PUF is 15% and 30% higher than the case where TRNG occupies a whole PUF bank (where the capacity of TRNG is equal to PUF) of 8 banks and 4 banks LLC, respectively. The results show that our proposed scheme using TRNG and SHA-512 can implement scalable PUF and effectively alleviate the negative impact of PUF occupying space in the LLC. Meanwhile, the same conclusion can be drawn from the average group of two comparative experiments.

From the discussion in the two subsections above, the smaller the TRNG space is, the less the impact on system IPC performance will be. In addition, the bit length of TRNs segments used for encryption has little effect on PUF performance. However, considering the operation complexity, more encryption operations would be taken in the case of longer TRNs segments. Meanwhile, when the length of a segment is 256 bits, the repetition probability of the encrypted bit stream is $1/2^{\mathbf{256}}$. However, the repetition probability increases exponentially to $1/2^{\mathbf{16}}$ when the length is 16. This greatly reduces the security of PUF. Hence, customers need to make a trade-off between the TRNs segments length and the security of PUF.

C. Performance Comparison

Table VI shows the performance comparison between this work and the state-of-the-art PUF implementations based on SRAM, RRAM, and STT-MRAM. Similar to the RRAM PUF, our proposal uses DES as the entropy source of PUF. Compared with the PUFs using SES, DES enables reconfigurable PUFs, which exponentially increases the maximum bit number

TABLE VI
PERFORMANCE COMPARISON WITH THE STATE-OF-THE-ART PUF DESIGNS

Comparison	TIFS'15 [29]	JSSC'18 [20]	TIFS'20 [30]	TCAS-1'20 [27]	**This work**
Tech node	45nm	65nm	65nm	130nm	40nm
Scheme	STT-MRAM	SRAM	STT-MRAM	RRAM + TRNG	**SOT-MRAM**
Entropy source	SES MTJ resistance	SES Transistor V_{th}	SES MTJ resistance	DES Jitter noise	**DES Thermal noise**
Cell area (F^2)	212	/	172	108	**85**
Uniqueness (%)	50.2	50.17	/	50.00	50.04
Max bit num per cell	2	1	1	$\sim 10^7$	$\sim \mathbf{10^{14}}$
Temperature (°C)	-75 $\sim$ 125	0 $\sim$ 100	-25 $\sim$ 75	-50 $\sim$ 150	-25 $\sim$ 100
BER per 10 °C	0.05%	0.5%	0.01%	0.03%	0.02%
Supply voltage (V)	0.9 $\sim$ 1.1	0.5 $\sim$ 1	0.8 $\sim$ 1.2	1.44 $\sim$ 2.16	0.9 $\sim$ 1.3
BER per 0.1 V	0.05%	0.6%	0.04%	0.01%	0.06%
ACF @95% confidence	/	/	0.0306	0.006	**0.005**
Entropy	0.99	/	0.9977	0.9999	0.9937
Reconfigurability	No	No	No	Yes	**Yes**
In Cache	No	Yes	No	No	**Yes**
Dynamic transformation	No	Yes	No	No	**Yes**
Randomness	low	low	low	high	**high**
Tape out	No	Yes	No	Yes	No

per cell. Due to the high endurance of SOT-MRAM, the maximum bit number per cell of SOT-MPUF is 10^7 times that of RRAM PUF. Moreover, the area of the dedicated PUF and TRNG circuits are saved by implementing PUF in the cache like Ref. [20]. In addition, the 50.04% uniqueness of this work is better than STT and SRAM, only second to RRAM. Furthermore, the switching probability of the SOT-MRAM cell shows higher reliability and better controllability. In terms of reliability, the BER of 0.02% per 10 °C and 0.06% per 0.1 V is comparable to the STT-MRAM and SRAM-based PUFs. More importantly, TRNG and SHA-512 enable simultaneous reconfigurability and dynamic switching between PUF and cache.

V. CONCLUSION

In this paper, we proposed a SOT-MRAM based reconfigurable PUF, and further realized dynamic transformation between the SOT-MPUF and the SOT-MRAM based LLC by utilizing the in-cache TRNG and SHA-512. As a result, the area overhead caused by the dedicated security circuit is eliminated by implementing PUF inside the cache. The maximum bit number per cell can achieve 10^{14} due to the full reconfigurability under the thermal noise as its DES, which greatly improves the CRP space compared with the traditional SES based PUFs. Through the circuit simulation conducted with the 40 nm technology and a compact SOT-MTJ model, the SOT-MPUF with true randomness shows good uniqueness and reliability. In addition, the In-Cache-MPUF using SHA-512 to encrypt random numbers of different bit lengths also exhibits promising performances. By using gem5 and spec CPU 2017 to build a system-level simulation, the results show that our proposed method with TRNG effectively improves the IPC performance by 30% compared to the method of using an entire cache bank as PUF. Indeed, further improvement in the area efficiency, energy efficiency and performance could be achieved with a more complicated design, as our prospect may be the further work. Due to the full compatibility with the promising SOT-MRAM cache, the method proposed in this paper has good application prospects in the emerging fields, such as embedded SoC and portable wearable devices.

REFERENCES

[1] Y. Gao, S. F. Al-Sarawi, and D. Abbott, "Physical unclonable functions," *Nature Electron.*, vol. 3, no. 2, pp. 81–91, 2020.

[2] Y. Wang, C. Wang, C. Gu, Y. Cui, M. O'Neill, and W. Liu, "A dynamically configurable PUF and dynamic matching authentication protocol," *IEEE Trans. Emerg. Topics Comput.*, early access, Apr. 12, 2021, doi: 10.1109/TETC.2021.3072421.

[3] C. Herder, M.-D. Yu, F. Koushanfar, and S. Devadas, "Physical unclonable functions and applications: A tutorial," *Proc. IEEE*, vol. 102, no. 8, pp. 1126–1141, Aug. 2014.

[4] B. Gassend, D. Clarke, M. van Dijk, and S. Devadas, "Controlled physical random functions," in *Proc. 18th Annu. Comput. Secur. Appl. Conf.*, Dec. 2002, pp. 149–160.

[5] D. Lim, J. W. Lee, B. Gassend, G. E. Suh, M. van Dijk, and S. Devadas, "Extracting secret keys from integrated circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 13, no. 10, pp. 1200–1205, Oct. 2005.

[6] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *Proc. 44th ACM/IEEE Design Autom. Conf.*, Jun. 2007, pp. 9–14.

[7] P. Santikellur and R. S. Chakraborty, "A computationally efficient tensor regression network-based modeling attack on XOR arbiter PUF and its variants," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 40, no. 6, pp. 1197–1206, Jun. 2021.

[8] Y. Lao and K. K. Parhi, "Statistical analysis of MUX-based physical unclonable functions," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 33, no. 5, pp. 649–662, May 2014.

[9] J. Delvaux and I. Verbauwhede, "Fault injection modeling attacks on 65 nm arbiter and RO sum PUFs via environmental changes," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 6, pp. 1701–1713, Jun. 2014.

[10] S. V. S. Avvaru, Z. Zeng, and K. K. Parhi, "Homogeneous and heterogeneous feed-forward XOR physical unclonable functions," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2485–2498, 2020.

[11] C. Q. Liu, Y. Cao, and C. H. Chang, "ACRO-PUF: A low-power, reliable and aging-resilient current starved inverter-based ring oscillator physical unclonable function," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 64, no. 12, pp. 3138–3149, Dec. 2017.

[12] H. Mandry, A. Herkle, S. Muelich, J. Becker, R. F. H. Fischer, and M. Ortmanns, "Normalization and multi-valued symbol extraction from RO-PUFs for enhanced uniform probability distributions," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 12, pp. 3372–3376, Dec. 2020.

[13] Z. Huang, C. Zhao, Q. Wang, and Z. Wang, "Implementation and analysis of improved RO PUFs with latch structure," in *Proc. Int. Conf. Netw. Netw. Appl. (NaNA)*, Oct. 2018, pp. 318–323.

[14] R. P. Challa, S. A. Islam, and S. Katkoori, "An SR flip-flop based physical unclonable functions for hardware security," in *Proc. IEEE 62nd Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2019, pp. 574–577.

[15] S. S. Kumar, J. Guajardo, R. Maes, G.-J. Schrijen, and P. Tuyls, "Extended abstract: The butterfly PUF protecting IP on every FPGA," in *Proc. IEEE Int. Workshop Hardw.-Oriented Secur. Trust*, Jun. 2008, pp. 67–70.

[16] J. Lee, D. Lee, Y. Lee, and Y. Lee, "A 354F^2 leakage-based physically unclonable function with lossless stabilization through remapping for low-cost IoT security," *IEEE J. Solid-State Circuits*, vol. 56, no. 2, pp. 648–657, Feb. 2021.

[17] Y. Cao, W. Zheng, X. Zhao, and C.-H. Chang, "An energy-efficient current-starved inverter based strong physical unclonable function with enhanced temperature stability," *IEEE Access*, vol. 7, pp. 105287–105297, 2019.

[18] A. Alvarez, W. Zhao, and M. Alioto, "14.3 15fJ/b static physically unclonable functions for secure chip identification with <2% native bit instability and 140× inter/intra PUF Hamming distance separation in 65nm," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2015, pp. 1–3.

[19] X. Zhao *et al.*, "A 124 fJ/bit cascode current mirror array based PUF with 1.50% native unstable bit ratio," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 9, pp. 3494–3503, Sep. 2019.

[20] J. Li, T. Yang, M. Yang, P. R. Kinget, and M. Seok, "An area-efficient microprocessor-based SoC with an instruction-cache transformable to an ambient temperature sensor and a physically unclonable function," *IEEE J. Solid-State Circuits*, vol. 53, no. 3, pp. 728–737, Mar. 2018.

[21] D. E. Holcomb, W. P. Burleson, and K. Fu, "Initial SRAM state as a fingerprint and source of true random numbers for RFID tags," in *Proc. Conf. RFID Secur.*, vol. 7. Jan. 2007, pp. 1–12.

[22] S. K. Mathew *et al.*, "16.2 A 0.19pJ/b PVT-variation-tolerant hybrid physically unclonable function circuit for 100% stable secure key generation in 22nm CMOS," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2014, pp. 278–279.

[23] J. Li, T. Yang, and M. Seok, "A technique to transform 6T-SRAM arrays into robust analog PUF with minimal overhead," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2017, pp. 1–4.

[24] K. Liu, Y. Min, X. Yang, H. Sun, and H. Shinohara, "A 373-F2 0.21%-native-BER EE SRAM physically unclonable function with 2-D power-gated bit cells and V_{SS} bias-based dark-bit detection," *IEEE J. Solid-State Circuits*, vol. 55, no. 6, pp. 1719–1732, Jun. 2020.

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

[26] Y. Pang *et al.*, "25.2 A reconfigurable RRAM physically unclonable function utilizing post-process randomness source with $< 6 \times 10^{-6}$ native bit error rate," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2019, pp. 402–404.

[27] Q. Zhao, W. Zheng, X. Zhao, Y. Cao, F. Zhang, and M.-K. Law, "A 108 F^2/bit fully reconfigurable RRAM PUF based on truly random dynamic entropy of jitter noise," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 11, pp. 3866–3879, Nov. 2020.

[28] B. Lin *et al.*, "A highly reliable RRAM physically unclonable function utilizing post-process randomness source," *IEEE J. Solid-State Circuits*, vol. 56, no. 5, pp. 1641–1650, May 2021.

[29] L. Zhang, X. Fong, C.-H. Chang, Z. H. Kong, and K. Roy, "Highly reliable spin-transfer torque magnetic ram-based physical unclonable function with multi-response-bits per cell," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 8, pp. 1630–1642, Aug. 2015.

[30] S. Lim, B. Song, and S.-O. Jung, "Highly independent MTJ-based PUF system using diode-connected transistor and two-step postprocessing for improved response stability," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2798–2807, 2020.

[31] B. Song, S. Lim, S. H. Kang, and S.-O. Jung, "Environmental-variation-tolerant magnetic tunnel junction-based physical unclonable function cell with auto write-back technique," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 2843–2853, 2021.

[32] G. Finocchio *et al.*, "Spin–orbit torque based physical unclonable function," *J. Appl. Phys.*, vol. 128, no. 3, 2020, Art. no. 033904.

[33] J. Zhang *et al.*, "Spin–orbit torque-based reconfigurable physically unclonable functions," *Appl. Phys. Lett.*, vol. 116, no. 19, 2020, Art. no. 192406.

[34] B. Dieny *et al.*, "Opportunities and challenges for spintronics in the microelectronics industry," *Nature Electron.*, vol. 3, no. 8, pp. 446–459, 2020.

[35] L. Chang, Z. Wang, Y. Gao, W. Kang, Y. Zhang, and W. Zhao, "Evaluation of spin-Hall-assisted STT-MRAM for cache replacement," in *Proc. IEEE/ACM Int. Symp. Nanosc. Architectures (NANOARCH)*, Jul. 2016, pp. 73–78.

[36] C. Wang, Z. Wang, S. Peng, Y. Zhang, and W. Zhao, "Advanced spin orbit torque magnetic random access memory with field-free switching schemes (invited)," in *Proc. IEEE 15th Int. Conf. Solid-State Integr. Circuit Technol. (ICSICT)*, Nov. 2020, pp. 1–4.

[37] B. Wu *et al.*, "Field-free 3T2SOT MRAM for non-volatile cache memories," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 12, pp. 4660–4669, Dec. 2020.

[38] M. Sumagita and I. Riadi, "Analysis of secure hash algorithm (SHA) 512 for encryption process on web based application," *Int. J. Cyber-Secur. Digit. Forensics*, vol. 7, pp. 373–381, Oct. 2018.

[39] NIST. *NIST SP800-22: Documentation and Software—Random Bit Generation*. Accessed: Jul. 9, 2014. [Online]. Available: https://csrc.nist.gov/Projects/Random-Bit-Generation/Documentation-and-Software

[40] NIST. *NIST SP800-90B: Entropy Sources Used for Random Bit Generation*. Accessed: Jan. 2018. [Online]. Available: https://csrc.nist.gov/publications/detail/sp/800-90b/final

[41] X. Zhao, C. Xie, Q. Zhao, and X. Pan, "A dual-entropy-superposed PUF with in-cell entropy sign-based stabilization," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 69, no. 1, pp. 284–286, Jan. 2022.

[42] N. Binkert *et al.*, "The gem5 simulator," *ACM SIGARCH Comput. Archit. News*, vol. 39, no. 2, pp. 1–7, 2011.

[43] E. Grimaldi *et al.*, "Single-shot dynamics of spin–orbit torque and spin transfer torque switching in three-terminal magnetic tunnel junctions," *Nature Nanotechnol.*, vol. 15, no. 2, pp. 111–117, Feb. 2020.

[44] N. Sato, F. Xue, R. M. White, C. Bi, and S. X. Wang, "Two-terminal spin–orbit torque magnetoresistive random access memory," *Nature Electron.*, vol. 1, no. 9, pp. 508–511, Sep. 2018.

[45] W. Cai *et al.*, "Sub-ns field-free switching in perpendicular magnetic tunnel junctions by the interplay of spin transfer and orbit torques," *IEEE Electron Device Lett.*, vol. 42, no. 5, pp. 704–707, May 2021.

[46] Z. Wang *et al.*, "Proposal of toggle spin torques magnetic RAM for ultrafast computing," *IEEE Electron Device Lett.*, vol. 40, no. 5, pp. 726–729, May 2019.

[47] M. Wang *et al.*, "Field-free switching of a perpendicular magnetic tunnel junction through the interplay of spin-orbit and spin-transfer torques," *Nature Electron.*, vol. 1, no. 11, pp. 582–588, Nov. 2018.

[48] Z. Wang, W. Zhao, E. Deng, J.-O. Klein, and C. Chappert, "Perpendicular-anisotropy magnetic tunnel junction switched by spin-Hall-assisted spin-transfer torque," *J. Phys. D, Appl. Phys.*, vol. 48, no. 6, Jan. 2015, Art. no. 065001.

[49] R. H. Koch, J. A. Katine, and J. Z. Sun, "Time-resolved reversal of spin-transfer switching in a nanomagnet," *Phys. Rev. Lett.*, vol. 92, Feb. 2004, Art. no. 088302.

[50] P. Debashis and Z. Chen, "Experimental demonstration of a spin logic device with deterministic and stochastic mode of operation," *Sci. Rep.*, vol. 8, no. 1, p. 11405, Jul. 2018.

[51] P. Debashis, V. Ostwal, R. Faria, S. Datta, J. Appenzeller, and Z. Chen, "Hardware implementation of Bayesian network building blocks with stochastic spintronic devices," *Sci. Rep.*, vol. 10, no. 1, Sep. 2020, Art. no. 16002.

[52] G. Finocchio *et al.*, "Spin–orbit torque based physical unclonable function," *J. Appl. Phys.*, vol. 128, no. 3, 2020, Art. no. 033904.

[53] Y. Kim, X. Fong, and K. Roy, "Spin-orbit-torque-based spin-dice: A true random-number generator," *IEEE Magn. Lett.*, vol. 6, 2015, Art. no. 3001004.

[54] Y. C. Wu *et al.*, "Voltage-gate-assisted spin-orbit-torque magnetic random-access memory for high-density and low-power embedded applications," *Phys. Rev. A, Gen. Phys.*, vol. 15, no. 6, Jun. 2021, Art. no. 064015.

Zhengyi Hou received the B.S. degree from the Changchun University of Science and Technology in 2018. He is currently pursuing the Ph.D. degree with the School of Integrated Circuit Science and Engineering, Beihang University, Beijing, China. His current research interests include emerging-device-based physical unclonable function design and peripheral circuit design for MRAM.

Zhaohao Wang (Senior Member, IEEE) received the B.S. degree in microelectronics from Tianjin University, China, in 2009, the M.S. degree in microelectronics from Beihang University, China, in 2012, and the Ph.D. degree in physics from University Paris-Saclay, France, in 2015. He is currently an Associate Professor at the School of Integrated Circuit Science and Engineering, Beihang University. His current research interests include the modeling of non-volatile nano-devices and the design of emerging non-volatile memories and logic circuits.

Chao Wang received the B.S. degree from Beihang University, Beijing, China, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include the modeling of non-volatile nano-devices, circuit level and architecture level design and optimization of STT-MRAM, SOT-MRAM, and design of in-memory computing architecture.

Min Wang received the M.S. degree in system engineering and the B.S. degree in applied mathematics from Beihang University, Beijing, China, in 2018 and 2021, respectively, where she is currently pursuing the Ph.D. degree in microelectronics with the School of Integrated Circuit Science and Engineering. Her current research interests include field-free spin-orbit torque writing and interfacial physical effects and controlling.

You Wang (Member, IEEE) received the Ph.D. degree in electrical engineering from the Institut Mines-Telecom, Telecom Paristech, France, in 2017. He is currently working as an Associate Research Fellow at the Hefei Innovation Research Institute, Beihang University, China. He has authored/coauthored more than 50 scientific papers. His research interests include spintronic devices modeling, circuit reliability-aware design, and novel designs for low power computing methods and security applications.

Xueyan Wang (Member, IEEE) received the B.S. degree in computer science and technology from Shandong University, Jinan, China, in 2013, and the Ph.D. degree in computer science and technology from Tsinghua University, Beijing, China, in 2018. From 2015 to 2016, she was a Visiting Scholar at the University of Maryland, College Park, MD, USA. She is currently an Assistant Professor with the School of Integrated Circuit Science and Engineering, Beihang University, Beijing. Her current research interests include processing-in-memory architectures, AI chip, and hardware security.

Cenlin Duan received the B.S. degree in electronic science and technology from the University of Electronic Science and Technology of China, Chengdu, China, in 2015, and the M.S. degree in software engineering from Xidian University, Xi'an, China, in 2018. She is currently pursuing the Ph.D. degree with the School of Integrated Circuit Science and Engineering, Beihang University, Beijing, China. Her current research interests include processing-in-memory architectures and energy-efficient deep-learning-network training processor design.

Jianlei Yang (Senior Member, IEEE) received the B.S. degree in microelectronics from Xidian University, Xi'an, China, in 2009, and the Ph.D. degree in computer science and technology from Tsinghua University, Beijing, China, in 2014.

From 2014 to 2016, he was a Post-Doctoral Researcher with the Department of ECE, University of Pittsburgh, PA, USA. He is currently an Associate Professor with the School of Computer Science and Engineering, Beihang University, Beijing. His current research interests include computer architectures and neuromorphic computing systems.

Dr. Yang was a recipient of the first/second place on ACM TAU Power Grid Simulation Contest in 2011/2012. He was a recipient of the IEEE ICCD Best Paper Award in 2013, the ACM GLSVLSI Best Paper Nomination in 2015, the IEEE ICESS Best Paper Award in 2017, and the ACM SIGKDD Best Student Paper Award in 2020.