All-Spin PUF: An Area Efficient and reliable PUF Design with Signature Improvement for Spin-Transfer Torque Magnetic Cell-Based All-Spin Circuits

Kangwei Xu, Dongrong Zhang, Qiang Ren, Yuanqing Cheng, Patrick Girard

To cite this version:

Kangwei Xu, Dongrong Zhang, Qiang Ren, Yuanqing Cheng, Patrick Girard. All-Spin PUF: An Area Efficient and reliable PUF Design with Signature Improvement for Spin-Transfer Torque Magnetic Cell-Based All-Spin Circuits. ACM Journal on Emerging Technologies in Computing Systems, 2022, 18 (4), pp.1-20/71. 10.1145/3517811. lirmm-03768916

HAL Id: lirmm-03768916
https://hal-lirmm.ccsd.cnrs.fr/lirmm-03768916

Submitted on 5 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
All-Spin PUF: An Area Efficient and reliable PUF Design with Signature Improvement for Spin-Transfer Torque Magnetic Cell-Based All-Spin Circuits

KANGWEI XU, DONGRONG ZHANG, QIANG REN, and YUANQING CHENG, Beihang University, China
PATRICK GIRARD, the Laboratory of Computer Science, Robotics and Microelectronics of Montpellier, CNRS/University of Montpellier, France

Recently, spin-transfer torque magnetic cell (STT-mCell) has emerged as a promising spintronic device to be used in Computing-In-Memory (CIM) system. However, it is challenging to guarantee the hardware security of STT-mCell based all spin circuits. In this work, we propose a novel Physical Unclonable Function (PUF) design for STT-mCell based all-spin circuit (All-Spin PUF) exploiting the unique manufacturing process variation (PV) on STT-mCell write latency. A methodology is used to select appropriate logic gates in the all-spin chip to generate a unique identification key. A linear feedback shift register (LFSR) initiates All-Spin PUF and simultaneously generates a 64-bit signature at each clock cycle. Signature generation is stabilized using an automatic write-back technique. In addition, a masking scheme is applied for signature improvement. The uniqueness of the improved signature is 49.61%. With ±20% supply voltage, and 5°C-105°C temperature variations, All-Spin PUF shows a strong resiliency. In comparison with the state-of-the-art PUFs, our approach can reduce hardware overhead effectively. Finally, the robustness of All-Spin PUF against emerging modeling attack is verified as well.

CCS Concepts: • Security and privacy → Hardware security implementation; • Hardware → Spintronics and magnetic technologies.

Additional Key Words and Phrases: Spin-Transfer Torque magnetic Cell (STT-mCell); hardware security; physical unclonable function (PUF); automatic write-back; signature improvement.

1 INTRODUCTION

Due to the performance gap between processor and main memory in big data and neural network computing applications, memory access becomes the performance bottleneck of computing systems, which is called "Memory Wall". Computing-In-Memory (CIM) is a promising technique to solve this problem [1]. On the other hand, with the increasing integration density, power consumption rockets up and results in severe thermal problem, which is known as the "Power Wall". In order to solve "Power Wall" problem, several emerging semiconductor devices are proposed to achieve better power efficiency [2][3]. For example, the spintronic technology which exploits...
the spin polarization of electrons for information processing has been extensively studied in recent years [4]. STT-mCell is a kind of spintronic device supporting both data storage and computing. However, due to the non-volatility of STT-mCell, hardware security has become a big concern for STT-mCell based all-spin circuits. Physical Unclonable Function (PUF) is an effective technique to enforce the data privacy and helps in use cases such as prevention of semiconductor counterfeiting. PUF designs are based on some electrical properties, such as path delay and resistance value [7]. Many CMOS-based PUFs were proposed in the last decade to enhance the circuit security, such as Arbiter-PUF [8], ring-oscillator PUF [9], etc. However, the conventional CMOS PUFs have significant power consumption, and cannot be directly applied to all-spin circuits due to different logic switching mechanisms [6].

In this article, a novel area and energy-efficient PUF for all-spin circuits is proposed based on STT-mCell’s write delay variations. In addition, we also propose circuit-level design techniques to enhance the reliability of our PUF design. Compared to the state-of-the-art work, the proposed All-Spin PUF design can bring high energy efficiency enabled by all-spin device, and reduce the PUF area overhead effectively by using existing (Design-for-Testability) DFT structure. The main contributions of this work can be summarized as follows.

- We propose a PUF design for STT-mCell based all-spin circuits. With a K-stage linear feedback shift register (LFSR), the All-Spin PUF can generate signatures of an arbitrary length smaller than $2^K - 1$ for a given challenge.
- To enhance the reliability of signature generation, we proposed an Automatic Write-Back (AWB) technique. Moreover, a counter-based signature improvement technique is adopted in All-Spin PUF to enhance the signature’s uniformity and uniqueness.
- The throughout of signature generation in All-Spin PUF is high (64 bits per clock cycle), i.e., multiple response-bits can be extracted with a single cycle. Moreover, the proposed PUF design methodology can be extended to other non-volatile CIM devices, such as magnetoelectric spin-orbit (MESO) device [10], Composite-Input Magnetoelectric-based Logic Technology (CoMET) device [11] and so on.

The rest of this paper is organized as follows. Section 2 introduce the basics of STT-mCell and related works. Section 3.1 illustrates the motivation of PUF design for STT-mCell based circuits. Section 3 describes the working procedure and implementations of All-Spin PUF with write back scheme and the counter-based signature improvement technique. Comprehensive evaluations of the proposed PUF design are given in Section 4. Finally, we conclude the paper in Section 5.

2 PRELIMINARIES AND RELATED WORK

This section firstly introduces a four-terminal magnetoelectronic device named STT-mCell (referred to as “mCell” thereafter) [5] and then reviews the related work.

2.1 Introduction to STT-mCell

The STT-mCell is a spintronic device with electrical insulation between the separated read and write paths. As shown in Fig. 1 (a), the four-terminals of the device form separated write path ($w^+, w^-$) and read path ($R, R'$). A magnetic tunnel junction (MTJ) is the basic storage element in a STT-mCell. It is composed of a tunnel barrier sandwiched between a pinned magnetic layer (PL) and a coupled free layer (FL) as shown in Fig. 1 (b). The MTJ resistance $R_{MTJ}$ is determined by the magnetization of the FL. When it is the same as that of the PL, the MTJ is in the parallel state and $R_{MTJ} = R_{Low}$ (low resistance). Otherwise, the MTJ is in the anti-parallel state and $R_{MTJ} = R_{High}$ (high resistance). The magnetization of the FL can be controlled by the domain wall motion underneath, which can be adjusted by injecting a spin current as shown in Fig. 1 (b). In an MTJ, the resistance
The ratio between $R_{\text{Low}}$ and $R_{\text{High}}$ is defined as Tunneling Magnetoresistance Ratio (TMR), and is given by

$$TMR = \frac{R_{\text{High}} - R_{\text{Low}}}{R_{\text{Low}}}$$  \hspace{1cm} (1)

The low and high resistances of a tunnel junction in a STT-mCell are given by (2) and (3) respectively [12].

$$R_{\text{Low}} = \frac{RA}{\text{Width} \times L_{\text{MTJ}}}$$  \hspace{1cm} (2)

$$R_{\text{High}} = \frac{RA}{\text{Width} \times L_{\text{MTJ}}} \times (1 + TMR)$$  \hspace{1cm} (3)

where $\text{Width}$ and $L_{\text{MTJ}}$ are device width and MTJ length respectively, as shown in Fig. 1.

Therefore, unlike conventional random access memory (RAM) chip technologies, data of STT-mCell are not stored as electric charge but instead stored by magnetic polarization of storage elements, which can be changed to represent either a ‘1’ (i.e., PL is anti-parallel to FL) or ‘0’ (i.e., PL is parallel to FL).

STT-mCells can also be used to build various logic gates with pull-up and pull-down networks like CMOS devices. As shown in Fig. 2(a) and (b), the STT-mCell based inverter/buffer can be constructed by two mCells, called pull-up mCell (UmC) and pull-down mCell (DmC). Taking the buffer cell as an example, the input is applied...
to the UmC (w−) port, and the DmC (w−) connects to ground, so that UmC and DmC can form a complementary pair [13]. The output current is

$$I = \frac{V(\frac{1}{R_{PU}} - \frac{1}{R_{PD}})}{k}, \quad k = 1 + R_{outpath}(\frac{1}{R_{PU}} + \frac{1}{R_{PD}})$$

where V is the magnitude of the power supply. $R_{PU}$ and $R_{PD}$ are the resistances of the UmC and DmC in the gate.

An mCell has low resistance $R_{Low}$ when the current is right-ward (w− to w+), and has high resistance $R_{High}$ when the current is left-ward (w+ to w−). In the STT-mCell based circuit, since logic value is only based on the direction of the current flow, the positive current represents logic ‘1’ and the negative current represents logic ‘0’.

In the buffer circuit shown in Fig. 2(b), when the applied input is $I_{in} = +25 \mu A$ (denote as logic ‘1’) in UmC of a buffer, $R_{PU} = R_{Low}$, $R_{PD} = R_{High}$, so the output current turns into a positive current ($I_{out} = +I_{in}$) according to (4), which means the output is logic ‘1’. The inverter cell works in a similar manner.

The detailed micromagnetic simulation of state switching of a buffer is shown in Figure 3 [12]. For a buffer with small write delay, if an applied challenge value ‘1’ is completely written into a buffer within the defined time window, the UmC is $R_{Low}$ when the current is right-ward (w− to w+), and the DmC is $R_{High}$ when the current is left-ward (w+ to w−). The domain wall moves in the opposite direction of the current, so the domain wall of UmC moves from w+ to w− (from right to left), and the domain wall of DmC moves from w− to w+ (from left to right).

In addition to the buffer/inverter shown in Fig. 2, other logic gates can be built by STT-mCells as well [12]. For instance, a 2-input NAND gate and its functional simulation waveform are illustrated in Fig. 4(a) and (b), respectively. Based on the NAND gate, we can build other logic gates accordingly.

The isolated read and write paths make STT-mCell based all-spin circuit ultra low power and more reliable [12]. Similar to conventional MRAM, the STT-mCell is feasible for power gating by enabling/disabling clock signals. The clock signal is called $pClock$ in this paper. As shown in Fig. 2(b), it illustrates the function of $pClock$ and the read/write operations of a buffer. When the $pClock$ is activated, the circuit can perform logic operations. When the $pClock$ is disabled, the currents can still flow into STT-mCells, but the STT-mCell based logic gate does not output current due to the disabled $pClock$. That is, the output of an STT-mCell based logic gate is
All-Spin PUF: An Area Efficient and reliable PUF Design with Signature Improvement for Spin-Transfer Torque Magnetic Cell-Based All-Spin Circuits • 5

Fig. 3. (a) The schematic diagram of a STT-mCell based buffer. (b) The low resistance ($R_{\text{Low}}$) and high resistance ($R_{\text{High}}$) of a working buffer. (c) Micromagnetic simulation of domain wall motion and state switching. The color indicates magnetic polarization [12].

Fig. 4. (a) The schematic of a NAND gate based on STT-mCell. (b) Transient waveforms of operating the NAND gate.

As shown in Fig. 5(a), a PUF is a disordered physical system $S$ that, when interrogated by a challenge (input), generates a unique device response (output). This response shall depend on the applied challenge and on the specific disorder and device structure of the PUF. The exact mechanism of the challenge-response generation enabled/disabled by pClock. Since STT-mCell can be used for both storage and logic operations, it is a promising technology inherently suitable for Computing-In-Memory system [12].

2.2 Physical Unclonable Function

As shown in Fig. 5(a), a PUF is a disordered physical system $S$ that, when interrogated by a challenge (input), generates a unique device response (output). This response shall depend on the applied challenge and on the specific disorder and device structure of the PUF. The exact mechanism of the challenge-response generation enabled/disabled by pClock. Since STT-mCell can be used for both storage and logic operations, it is a promising technology inherently suitable for Computing-In-Memory system [12].
cannot be copied to another physical device due to the imperfections and uncertainties in the fabrication technology.

The PUF is used for chip authentication. As shown in Fig. 5(b), the Central Authority (CA) selects a threshold and a challenge from challenge-response-pair (CRP) list in the remote server, and sends the challenge to the authorized user. The user applies this challenge to the All-Spin PUF of the chip # 1, after obtaining the correct response, the user can use the chip.

2.3 Related Work

Considering the existing PUFs, many area and energy-efficient designs have been proposed as depicted as a radar chart, which directly reflect the performance through the size of area. As shown in Fig. 6, CRC-PUF has the best overall performance, followed by MemPUF, Buskeeper PUF and ARO-PUF. The overall performance of Bi-stable PUF and DWM-PUF are weaker than other PUFs.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Uniqueness</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cost</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Attack Resistance</td>
<td>Not Mentioned</td>
<td>Resistant to Side-Channel Attacks</td>
<td>Not Mentioned</td>
<td>Not Mentioned</td>
<td>Resistant to Reverse Engineering</td>
<td>Not Mentioned</td>
<td>Resistant to Modeling Attack</td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(1) ✓ Indicates that the proposed scheme can improve the specific metric of PUF;
(2) △ Indicates that the proposed scheme may improve the specific metric of PUF;
(3) × Indicates that the proposed scheme may deteriorate the specific metric of PUF.

Fig. 6. The reliability, uniqueness, cost and security comparisons of existing PUF schemes.

2.3.1 Reliability Improvement. Bhargava et al. used the accelerated aging effects to skew the bi-stable PUF cells so as to produce reliable bits [13][14]. However, aging acceleration required dedicated circuit components or additional testing procedure, and also degraded the performance of the PUF when it worked as the normal
All-Spin PUF: An Area Efficient and reliable PUF Design with Signature Improvement for Spin-Transfer Torque Magnetic Cell-Based All-Spin Circuits

memory [15]. Zhang et al. proposed bit generation technique in a STT-MRAM-based MemPUF [16], which was stabilized with a novel automatic write-back scheme. But the PUF required an independent write-back module that resulted in larger area overheads.

The most common way to solve the reliability issue is to use Error Correction Code (ECC) [17][18]. ECC stabilizes the noisy response-bits generated from the PUF. However, if the bit error rate (BER) of the raw response is high, the ECC overhead may be very costly.

2.3.2 Uniqueness Improvement. Zheng et al. took advantage of SRAM write-failure effect to produce random bits, and the supply voltage was used to increase the uniqueness of the PUF [19]. But it may be compromised if adversaries know the digital signals that can regulate the external effects (e.g., supply voltage [20] and electrical pulse [21]) for bit expansion. Another domain wall memory PUF (DWM-PUF) was proposed in [22]. This design provided additional knobs, e.g., shift pulse, number of access ports to expand the set of challenge-response pairs. The results showed excellent uniqueness. However, the domain wall can only be shifted forward and backward by injecting current from the left/right-shift contact. Read is performed by shifting the desired bit under the read head using spin polarized currents. For random access, the worst case latency may be very high.

2.3.3 Cost Improvement. Buskeeper PUF utilizes a buskeeper cell, which is smaller than a D flip-flop (DFF). However, it is a DFF-based PUF and requires additional addressing circuit [23]. The aging-resistant ring oscillator PUF (ARO-PUF) [24] offers a considerably smaller PUF footprint since it requires lighter ECC scheme. Although it shows less area overhead, the design is limited to custom layout design.

2.3.4 Resistance to Attack. In [25], a nonlinear voltage transfer function was instantiated to create a complex mapping between the challenge and response of each circuit module, which was impossible to attack. However, the nonlinear voltage transfer function varies with aging, and the error is accumulated at each stage, leading to low PUF reliability. A lightweight PUF construction, a Cyclic Redundancy Check (CRC) PUF was proposed in [26], where input challenges were de-synchronized from output responses to make a PUF model difficult to learn. However, its security evaluation is only for modeling attacks and does not take other attacks into consideration.

Most of existing spintronic PUFs suffer from non-negligible area overhead, and lack of considering the all-spin circuit design enabled by STT-mCells. To this end, an All-Spin PUF design with high energy and area efficiency is desirable.

3 WRITE DELAY BASED STT-MCELL PUF ARCHITECTURE

In this section, we first present the motivation of this work and then detail the proposed All-Spin PUF architecture.

3.1 Motivation of PUF Design for STT-mCell Based Circuits

The geometry of a STT-mCell structure, e.g., device width, MTJ length, TMR, and resistance area product (RA), may vary due to imperfect fabrication process [12][27][28]. During the fabrication of a STT-mCell, one of the greatest challenges is to obtain high TMR with low RA of the MTJ. Low RA can reduce the power consumption of the circuit, but decreasing RA also results in a drop in TMR due to metallic conduction through the junction. Therefore, there is a trade-off between RA and TMR [29]. Besides, improved MTJ properties (low RA) and low critical current density in the write-path further reduce the supply requirements. For an inverter, compared with VDD=0.9V in 32nm CMOS, the supply voltages in 32nm STT-mCell can be well under 50mV if the MTJs have a low enough RA [29], the STT-mCell based circuit can achieve significant power savings due to much lower supply voltage.

To successfully program the MTJ within a given delay, the current amplitude needs to be larger than a critical reference current (denoted as \(I_{ref}\)). As shown in Fig. 2 (b), a positive 25\(\mu\)A write current is injected to the buffer
at 0ns, and the buffer completely outputs 25μA at 2.5ns. High TMR with low RA of the MTJ can reduce the write delay. The smaller $L_{MTJ}$ is, the shorter the write delay is.

An etching step is required to fabricate and separate two MTJs that share the same free layer. The contact is difficult to fabricate because precisely etching the MgO barrier without damaging the magnetic free layer is a challenging task. In this paper, we consider the process variations of $L_{MTJ}$, RA and TMR simultaneously. When $L_{MTJ}$, RA and TMR change due to the process variations [12], the domain wall can have slightly different moving speeds, that can affect the write delay of a STT-mCell. Based on the observation of process variation impact on STT-mCell write delay, we can construct the all spin PUF.

3.2 The working procedure of All-Spin PUF Architecture

As shown in Fig. 7, all logic gates in the circuit are built by STT-mCells. Like conventional CMOS VLSI design, we assume the DFT structure is embedded in the all-spin circuits, including scan flip flops and scan chains, which are implemented with STT-mCells as well. We select the logic gates following the SFFs for delay characterization, and their inputs can be controlled by scan flip-flops (SFFs) in the scan chain.

3.2.1 Challenge Generation. A PUF can be stimulated with external inputs, called challenges, upon which it generates with corresponding outputs, called responses. The generated response $R_{C1}$ depends on its internal physical disorder and the input challenge. As shown in Fig. 1, the STT-mCell is designed symmetrically for storing logic '0' and '1'. In actual manufacturing, due to the process variations, it is impossible to fabricate complete symmetrical left and right path ($w^+$, $w^-$), and the domain wall moves from the center to the left or right with slightly different distances due to process variations.

As shown in Fig. 7, at the beginning of working procedure, we initialize all logic gates through the 'SI' terminals of SFFs. Then, a 16-bit external challenge vector is provided as the initial seed of a linear feedback shift register (LFSR). When the LFSR is enabled, it can generate a signature of an arbitrary length smaller than $2^{16}-1$ for a given challenge. For the simulation purpose, we selected 64 bits of the $2^{16}-1$ bits as the All-Spin PUF’s internal challenge vector ($C_0, \ldots, C_m, C_{m+1}, \ldots, C_{63}$). $C_0, \ldots, C_m, C_{m+1}, \ldots, C_{63}$ are propagated through the circuit to the inputs of scan flip-flops (SFFs) in parallel, and are simultaneously applied to the logic gates following the SFFs.

3.2.2 Response Extraction. In STT-mCell logic, the logic value is based on the current direction, the positive current represents logic ‘1’ and the negative current represents logic ‘0’. Read operations of proposed All-Spin PUF are performed by using the current directions of the selected buffers as mentioned above. As shown in Fig. 7, in the initialization phase, through the ‘SI’ terminals of the scan flip-flops (SFF), the inverse values of challenge
bits are scanned into the selected buffers behind the SFFs. This initialization makes a '0' → '1' or '1' → '0' transition during the response extraction.

Therefore, the challenge bits generated by LFSR are written into the selected buffers. The buffer with an initial value '0' is written to '1', and the buffer with an initial value '1' is written to '0'. The reference write time window (Threshold, ∆T) is determined by the output current distribution of all selected buffers during the read operation. To ensure the uniformity of response bits, an independent ∆T is set for each All-Spin PUF.

The gate selection principle is based on that the gate inputs can be controlled by scan flip-flops (SFFs) in the scan chain. That is, the challenges can simultaneously reach the inputs of logic gates following the SFFs when the unified clock edge arrives at SFFs. In Fig. 7, we denote the selected gates for delay characterization with $1, \cdots, m, m+1, \cdots, n$. The write delays of these gates are different due to process variations. In actual STT-mCell circuit design, a buffer/an inverter is generally placed behind the D flip-flop to improve driving capability. Therefore, we select these buffers as the logic gates to generate response bit in our PUF design.

As shown in Fig. 7, the first buffer (labeled as '0' in the figure) with an initial value '0' has been written completely. As a result, the signature bit of this buffer is logic value '1'. However, the buffer (labeled as 'm' in the figure) with an initial value '0' may not receive enough current to reach the switching threshold within the same time period. So the signature bit corresponding to buffer 'm' is still logic value '0'. Then, these selected buffers can generate different '0'/'1' bits simultaneously.

In a single-cycle query, the proposed PUF can be interrogated by a small number of fixed challenges, and it is similar to a weak PUF. In multiple-cycle queries, the number of CRPs is exponential with respect to the number of components used for building the proposed PUF, and it can be considered as a strong PUF [7].

3.3 The Structure of All-Spin PUF

The proposed All-Spin PUF is composed of an LFSR, and the scan chain existing in the DFT structure. By utilizing the existing components as much as possible, All-Spin PUF can reduce the area and power overheads effectively. The main components and their functionalities are described as follows.

- **Linear Feedback Shift Register (LFSR):** An external challenge vector as an initial seed is provided to the LFSR, and it generates a 64-bit internal challenge vector ($C_i$) in each clock cycle. Note that the challenge length is determined by the number of selected gates (i.e., buffers as mentioned above), and any LFSR capable of producing a 64-bit pseudo-random number can be used for this purpose. Note that the LFSR need to be initialized before every CRP generation. Because each CRP can only be used once, and the CRP-list on a server shrinks over time [7].

- **The scan chain:** Note that DFT techniques are widely used in contemporary ICs [32]. All-Spin PUF reuses the scan chain to implement its functionality. The outputs (internal challenge $C_i$) of the LFSR act as the inputs to the scan chain, and are written to the SFFs in parallel.

The more gates we choose, the higher randomness of the authentication signature is. If there are $n$ logic gates to be chosen, $2^n - 1$ different response combinations can be generated.

In the following, we will introduce an automatic write-back (AWB) scheme to improve the reliability of response bit generation, and a counter-based signature improvement technique to improve the uniqueness of the All-Spin PUF further.

3.4 The Automatic Write-Back Scheme

Maintaining the reliability of response bit generation under varying working conditions is a major challenge for the spintronic PUFs [16]. The thermal fluctuation in the output current may cause the actual write delay vary over time due to the thermal noise. The reliability of producing a response-bit is defined as the probability that bit
$b_t$, generated from a selected logic gate at time point $t$, is reproduced as $b_{t+\delta t}$ at time $t + \delta t$ ($\delta t > 0$). Therefore, we proposed the write-back path to guarantee the accuracy of the response bit output at time $t$.

![Image](image_url)

**Fig. 8.** (a) The schematic of the AWB scheme. (b) The transient timing diagram of latching a bit in the scan flip-flop for a buffer with large write delay. (c) The transient timing diagram of latching a bit in the scan flip-flop for a buffer with small write delay.

To precisely read the response bit $b_t$ at time $t$, we propose an automatic write-back (AWB) scheme to enhance the reliability of the response bit generation. As shown in Fig. 7, in the write-back phase, the switch ‘S’ and the terminal ‘TE’ are set to 0, which means that the switch ‘S’ is open, and the response is written back to the ‘SI’ terminal to prevent interference between every two stages of the SFFs. Note that when the circuit performs normal logic operations, TE=1, and the D terminal is selected [7].

The working principle of the AWB scheme is described as follows. As shown in Fig. 8, when the write time of the selected logic gates reaches the delay threshold $T_1$, all SFFs are triggered by the rising edge of the clock signal, the generated bits are automatically written back into the scan flip-flops (SFFs).

Due to the process variation of buffer ‘A’, its write delay between write ($A_{in}$) and read ($A_o$) is unique. Fig. 8(b) shows the timing diagrams of the write-back process in the scan chain. Firstly, a current (line ‘D’) is injected to the SFF, after the first rising edge of the clock, the output current (‘challenge’, line ‘Q’) reaches $A_{in}$ of the buffer ‘A’ at time $T_0$. With the write-back scheme, TE is set to 0. Then, the SFF is triggered by the second clock rising edge, the output response current ($A_o$) is written to the SFF with a fixed current value at time $T_1$. Finally, the response can be shifted out by the scan chain for post-processing.

To describe AWB scheme more clearly, the buffer with large process variation (large write delay due to a rough process) in Fig. 8(b) was denoted as Buffer ‘A’, and the buffer with small process variation (small write delay) in Fig. 8(c) was denoted as Buffer ‘B’.

As shown in Fig. 8(b), after the first rising edge of the clock, the buffer ‘A’ with an initial value ‘0’ is written with challenge ‘1’ (Signal $A_{in}$). As the the buffer ‘A’ with large process variation has a large write delay, after the second clock rising edge, the challenge value ‘1’ has not been written completely. Therefore, the output response (Signal $A_o$) is still the logic value ‘0’.

As shown in Fig. 8(c), after the first rising edge of the clock, the buffer ‘B’ with an initial value ‘0’ is written with challenge ‘1’ (Signal $B_{in}$). As the buffer ‘B’ with small process variation has a small write delay, after the second clock rising edge, the challenge value ‘1’ has been written completely. Therefore, the output response (Signal $B_o$) is the logic value ‘1’.

Compared to the reliability improvement techniques proposed in the literature [13] [14] [39], our method is superior in several aspects. Firstly, unlike the approach in [16], generation of the response bits in our scheme does not necessarily require setting the two MTJs into complementary states to represent write-back value, which may
add extra hardware overhead and the complexity of write-back operation. Secondly, our method does not rely on aging effects on the memory devices for reliability enhancement. Aging effects such as hot-carrier injection and negative-bias temperature instability may change the inherent properties of the devices permanently, and the stability of read/write operations can be significantly deteriorated [15].

3.5 Counter-Based Signature Improvement

In the actual circuit, logic gates may be interfered by the surrounding environment conditions or manufacturing process variations (PV), and a small portion of the ‘0’/‘1’ signature may be biased [40]. Meanwhile, with the modern semiconductor supply chain, the foundry may control the process parameter to bias the output signature, and ship out-of-spec/defective devices [7]. To improve the quality and security of signature, a counter-based signature improvement technique is proposed for All-Spin PUF as shown in Fig. 7.

The counter-based signature improvement module consists of a counter, a multiplexer and an XOR gate. A signature masking scheme is used to enhance the uniformity and uniqueness of All-Spin PUF. The control and processing unit initiates the mask generation circuit to generate the mask for the raw signature. An external 16-bit seed is applied to the LFSR and a 64-bit signature is used for mask generation. The multiplexer is used to select bits ‘1’ in this new signature. The total occurrences of ‘1’s in the signature are counted, and is converted to binary code as the m-bit mask, which is stored in STT-mCell. When All-Spin PUF is activated, the control unit loads the m-bit mask value from memory into the Shift Register as shown in Fig. 7.

The XOR gate is used to XOR the raw signature with the mask. As shown in Fig. 9, controlled by the rising edge of the clock signal, the n-bit raw signature shifted in the scan chain is XORed with the m-bit mask shifted out by the shift register. That is, the m-bit mask is XORed with the m-bit of n-bit signature every cycle until all n-bit are XORed to generate the final signature.

As shown in Fig. 9, with the masking scheme, the uniqueness of All-Spin PUF’s signature can be improved. Assume that there are different k bits between two n-bit raw signatures, and different j bits between two m-bit masks. Then, there are n-k same bits between two raw signatures.

With the masking scheme, only $\frac{n-k}{m} \times 100\%$ different mask bits can affect the raw signature bits. Among these affected bits, $\frac{n-k}{n} \times 100\%$ would be flipped compared to the raw signatures, which increases the uniqueness of the generated signature. However, there would be $\frac{k}{n} \times 100\%$ bits flipped to the same values, which reduces the uniqueness (because they were actually different in raw signatures). So the overall uniqueness enhancement can
be represented as:

$$\Delta u = \frac{j}{m} \cdot \frac{n-k}{n} - \frac{j}{m} \cdot \frac{k}{n} = \frac{j}{m} \left(1 - \frac{2k}{n}\right)$$  \hspace{1cm} (5)

As shown in (5), the whole uniqueness is improved for cases with $\frac{k}{n} < 0.5$, and the lower the initial uniqueness $\frac{k}{n}$ is, the more improvement $\Delta u$ can be achieved. Also, it should be noted that an instable bit within m-bit mask can cause $\frac{n}{m}/n = \frac{1}{m}$ final signature bit error rate (BER) due to the repeated mask application scheme. Therefore, it is necessary to store the mask value in non-volatile STT-mCell, to improve its resiliency to attacks. As shown in Fig. 7, the control unit is used to control the activation of All-Spin PUF, enable counter-based signature improvement module, store and load mask/threshold values.

In summary, the workflow of All-Spin PUF with signature improvement is shown in Fig. 10. The main steps in the workflow is described as follows.

- **Challenge Generation**: Initialize all selected buffers following SFFs. Then, a 16-bit external seed is applied to a 16-stage LFSR, and a 64-bit long input challenge (seed) $C_i$ is obtained.

- **The Mask Generation**: Controlled by the first rising edge of SFFs’ clock, $C_i$ simultaneously reaches the inputs of the logic gates following the SFFs. 64 bits signature (String A) shifted out to the signature improvement module. The multiplexer is used to select bit ‘1’s in this signature. The total occurrences of bit ‘1’s in ‘String A’ are accumulated in the counter, and is converted to a binary code as an m-bit mask. Then, the m-bit mask value is stored in STT-mCell memory. When an XOR operation is performed, the mask is shifted out by the shift register.

- **The Raw Signature Generation**: The raw signature generation process is basically the same as the signature generation process used for the mask as mentioned above. First, a new 16-bit seed (different from the previous one) is applied to the LFSR. Then, a new 64-bit sequence of ‘0’/’1’s can be generated as the raw signature (‘String B’).

- **The Improved Signature Generation**: The n-bit raw signature is shifted out bit by bit through the scan chain and XORed with the m-bit mask in sequence. Finally, the n-bit masked signature can be served as the final masked signature.

4 EXPERIMENTAL RESULTS

In this section, we evaluated the three most important metrics of All-Spin PUF, i.e., uniqueness, uniformity and reliability. The STT-mCell model that has been validated against experimental data was used in our simulations [46]. The area and energy consumption of our design were compared with the state-of-the-art works.
Table 1. Parameters Used in the Monte Carlo Simulation

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Nominal values</th>
<th>Process Variation (RSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of MTJ</td>
<td>$L_{MTJ} = 12\text{nm}$</td>
<td>10%</td>
</tr>
<tr>
<td>Distance of two MTJs</td>
<td>$L_{EXT} = 8\text{nm}$</td>
<td>10%</td>
</tr>
<tr>
<td>RA of the mCell</td>
<td>$R_A = 0.1\Omega \mu m^2$</td>
<td>5%</td>
</tr>
<tr>
<td>Width of the mCell</td>
<td>$W = 10\text{nm}$</td>
<td>2.4%</td>
</tr>
<tr>
<td>Supply Voltage</td>
<td>50mV</td>
<td>-</td>
</tr>
<tr>
<td>Temperature</td>
<td>25°C</td>
<td>-</td>
</tr>
</tbody>
</table>

4.1 Experimental Setup

The STT-mCell based logic gates were built based on the Verilog-A model of STT-mCell [46], and simulated in Cadence Spectre [41]. The simulation results were calibrated with the prototype in [46] to guarantee the accuracy of the logic gate modeling. The power consumption of each logic gate was calculated in Virtuoso. With these logic gates, the All-Spin PUF circuit was built and simulated with Cadence Spectre. Parameters used for Monte Carlo simulation are given in Table 1. During simulations, the same challenge set was applied to the LFSR of all 1000 All-Spin PUF samples at 25°C with 50mV supply voltage to produce the final 64-bit signature. Finally, the statistical data were generated and analyzed in Matlab. ITC99 [42], Gaisler [43] and ISCAS [44] benchmarks were used for evaluations.

4.2 Uniformity Analysis

The uniformity denotes a good random distribution of logic values ‘0’ and ‘1’ in each PUF signature. The average uniformity of the raw signatures is 49.03% (ideally 50%) as shown in Fig. 11.

Fig. 11. Uniformity of the proposed PUF ($\mu = 49.03\%$).

The bit-aliasing indicates the balanced random distribution of ‘0’s and ‘1’s for the same bit position in a PUF population. We count the number of occurrences of bit ‘1’ for the same bit position. The bit-aliasing of the raw signatures and the masked signatures are shown in Fig. 12 (a) and (b), respectively. We can observe that the masking scheme reduces the bit-aliasing of the All-Spin PUF signature effectively.

4.3 Uniqueness Analysis

Uniqueness is often quantified as the average Hamming distance between the signatures to the same challenge obtained from all All-Spin PUF samples measured in the same environmental condition. Let $R_u$ and $R_v$ be the $n$-bit
Fig. 12. Bit-aliasing comparisons. (a) Bit-aliasing of the raw signature, $\mu = 42.35\%$. (b) Bit-aliasing of the masked signatures, $\mu = 48.47\%$.

Fig. 13. The distribution of the Hamming Distance obtained from All-Spin PUF using $N = 64$ logic gates for 1000 chips. (a) HD of the raw signatures, $\mu = 37.73\%$. (b) HD of the improved signatures, $\mu = 49.61\%$.

As shown in Fig. 13, the uniqueness of the improved signature approaches to 49.61% (ideally 50%), which indicates a high uniqueness of the produced signatures.

4.4 Reliability Analysis

The reliability measures the stability of PUF responses to the same challenge under temperature and supply voltage variations. It was evaluated by comparing the response bits generated at different operating corners with those at the nominal corner (25°C, 50mV). The proposed All-Spin PUFs were measured with the supply voltage, in the range of 50mV±20% with a step of 5mV. The environmental variation models for STT based Cells have been thoroughly discussed in some studies [12] [47] [48]. To facilitate thermal analysis of STT-mCell at the circuit level, we added the temperature variable to the mCell Verilog-A model as [47] and varied it from 5°C to 105°C.

Let $R_i$ be an $n$-bit response to an input challenge $C_i$ produced by a PUF chip $i$ under the nominal operating condition. The same set of challenges were then applied $k$ times to obtain the response $R_{i,j}$ for $j = 1, 2, \ldots, k$, where
Fig. 14. Bit error rates estimated for 1000 All-Spin PUFs operating under supply voltage and temperature variations. The reliability $S$ of chip $i$ can be computed by

$$S = 1 - BER = 1 - \frac{1}{k} \sum_{j=1}^{k} \frac{HD(R_i, R_{i,j})}{n} \times 100\%$$

Fig. 14 illustrates the bit error rate (BER) with environmental variations. The average worst BERs with voltage and temperature variations are 4.92% and 4.77%, respectively. The BER ≤ 10% is acceptable for PUF designs since error-correcting codes can be used with a low cost [51]. The results demonstrate that All-Spin PUF is robust.

4.5 Security Considering Emerging Modeling Attack

Adversarial machine learning-based modeling attack is an emerging threat to the security of the PUF design. Modeling attack is generally carried out in three steps: 1) the adversary collects lots of CRPs of a specific PUF; 2) the adversary uses large CRPs to train the predictive model; and 3) the adversary applies new challenges and uses the predictive model for the specific PUF to obtain the complete PUF CRPs, to break its security. The effectiveness of modeling attack is based on the principle that similar challenges tend to generate similar responses, because signal propagation delay can be well represented by an additive linear delay model with a limited number of unknown parameters [56]. The premise of modeling attacks is that the collected CRP sets have certain regularity.

However, the mCell-based circuit consists of nonlinear magnetic devices. There are complex dependencies between the related parameters, such as magnetic field variations inside the device. Meanwhile, the threshold current for different All-Spin PUFs may be different, which also makes it difficult to perform modeling attacks.

To prove the security of All-Spin PUF under modeling attack, 16 one-hot challenges are applied. Fig. 15 (a) depicts the bitmap of 16 16-bit one-hot external challenge seeds input to LFSR. After an external seed input to the 16-level LFSR, it outputs a 64-bit string as an internal challenge applied to the input terminals of the scan chain. After passing through a scan chain and a mask module, these different responses will be generated (Fig. 15 (b)). The responses are different because LFSR can convert similar external challenge seeds into different internal challenges. The average hamming distance of the 16 responses is 38.63%, indicating that All-Spin PUF is capable of yielding significantly diverged signatures for similar challenges. Therefore, it is difficult for adversaries to succeed in modeling attacks.

4.6 Optimal Setting of Mask and Signature

Table 2 summarizes the results obtained for the different signature and mask sizes. The minimal size, for which the conditions of unpredictable All-Spin PUF are met, is $length_{signature}=32$, $length_{mask}=4$ (marked in blue in Table 2). However, this size would be sensitive to additional noise that may affect its reliability. A better option...
is $\text{length}_\text{signature} = 64$ and $\text{length}_\text{mask} = 6$ (marked in green in Table 2). This option guarantees the reliability of All-Spin PUF with sufficient margins for additional noise and relatively low BER. Moreover, this option has better signature uniqueness.

![Fig. 15. The binary bitmaps of one-hot challenges and responses. The '0's and '1's are represented by light blue pixels and dark blue pixels, respectively.](image)

### Table 2. PUF reliability and uniqueness with different sizes of raw signatures and the masks.

<table>
<thead>
<tr>
<th>$L_{\text{sig}}$</th>
<th>$L_{\text{mask}}$</th>
<th>No. of masking cycles</th>
<th>BER</th>
<th>$f_{\text{HD}}$ [%]</th>
<th>$f_{\text{HW}}$ [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\mu$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>4</td>
<td>4.56</td>
<td>35.77</td>
<td>10.6</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>8</td>
<td>9.19</td>
<td>38.95</td>
<td>9.1</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>6</td>
<td>11.42</td>
<td>41.78</td>
<td>9.4</td>
</tr>
<tr>
<td>32</td>
<td>4</td>
<td>8</td>
<td>12.72</td>
<td>48.02</td>
<td>5.4</td>
</tr>
<tr>
<td>32</td>
<td>5</td>
<td>7</td>
<td>10.03</td>
<td>45.55</td>
<td>8.6</td>
</tr>
<tr>
<td>48</td>
<td>5</td>
<td>10</td>
<td>7.81</td>
<td>49.41</td>
<td>3.1</td>
</tr>
<tr>
<td>64</td>
<td>4</td>
<td>16</td>
<td>6.85</td>
<td>49.58</td>
<td>1.6</td>
</tr>
<tr>
<td>64</td>
<td>6</td>
<td>11</td>
<td>5.14</td>
<td>49.61</td>
<td>1.5</td>
</tr>
</tbody>
</table>

### 4.7 Area Overhead

As mentioned above, the two parts: LFSR and scan chains are reused parts in the existing DFT structure without incurring extra hardware overhead. Therefore, the area overhead mainly comes from the buffers and the signature improvement module (including a counter, a multiplexer, an XOR gate and a STT-mCell memory).

![image]

**Table 3. The Area Overhead of All-Spin PUF.**

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>leon2</th>
<th>leon3mp</th>
<th>b22</th>
<th>s38584</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area Overhead (%)</td>
<td>0.0024</td>
<td>0.0031</td>
<td>0.09</td>
<td>0.26</td>
</tr>
</tbody>
</table>
Referring to the methods by [46] [49] to calculate the spin circuit area [50], we designed the layouts of various STT-mCell based logic gates and derived their areas to create the standard cell library file. The All-Spin PUF circuit was synthesized in Design Compiler with the generated cell library in 45nm technology node.

In general, to ensure the security of PUF, the response bit is more than 32 bits. However, as the number of PUF response bits increases, the area overheads of PUF array will also increase. For instance, without considering other security enhancements, PUFs with 64-bit signatures generally have higher security than PUFs with 32-bit signatures, but 64-bit PUFs generally have higher area overhead. If the total area is used to measure the area overhead, it may not be fair and accurate. Therefore, the metric for measurement was taken as area per bit instead of only total area. The area per bit \( \frac{\text{area}}{\text{bit}} \) can be calculated by

\[
\frac{\text{area}}{\text{bit}} = \frac{\text{total area} - \text{reused area}}{n} \tag{8}
\]

where \( \text{total area} \) is the area of All-Spin PUF, the \( \text{reused area} \) includes the area of LFSR and scan chain, \( n \) is the number of response bits. We also evaluated the area overheads for some benchmark circuits (i.e., Gaisler, ITC99 and ISCAS) as shown in Table 3. We can observe that the area overhead is relatively lower, especially for large-scale circuits.

### Table 4. Comparison of the Proposed All-Spin PUF with Other State-of-the-art PUF Designs.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology Node</td>
<td>45nm</td>
<td>65nm</td>
<td>65nm</td>
<td>65nm</td>
<td>65nm</td>
<td>45nm</td>
<td>45nm</td>
<td>180nm</td>
<td>65nm</td>
</tr>
<tr>
<td>Energy/bit(pJ)</td>
<td>0.0625</td>
<td>0.93</td>
<td>0.548</td>
<td>1.1</td>
<td>1.5\times10^3</td>
<td>0.69\times10^3</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Uniqueness(%)</td>
<td>49.61</td>
<td>50.6</td>
<td>50.01</td>
<td>33.21</td>
<td>50.14</td>
<td>49.9</td>
<td>49.8</td>
<td>50.0</td>
<td>49.1</td>
</tr>
<tr>
<td>Temp. range(°C)</td>
<td>5-105°C</td>
<td>0-80°C</td>
<td>0-80°C</td>
<td>25-85°C</td>
<td>25-85°C</td>
<td>-40-85°C</td>
<td>0-90°C</td>
<td>-25-85°C</td>
<td>-40-85°C</td>
</tr>
<tr>
<td>BER per 10°C (%)</td>
<td>1.07</td>
<td>0.68</td>
<td>0.44</td>
<td>6.67</td>
<td>0.47</td>
<td>2.16</td>
<td>0.9</td>
<td>1.4</td>
<td>1.14</td>
</tr>
<tr>
<td>BER per 10% volt. (%)</td>
<td>2.10</td>
<td>1.82</td>
<td>0.13</td>
<td>&gt;16.67</td>
<td>1.3</td>
<td>N/A</td>
<td>3.65</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Area/bit(μm²)</td>
<td>1.14\times10^2</td>
<td>29.86</td>
<td>7.42</td>
<td>N/A</td>
<td>N/A</td>
<td>0.43</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 4 compares the figures of merit among several state-of-the-art PUF designs with the proposed All-Spin PUF, where All-Spin PUF shows satisfying security performance. Compared with other PUFs in terms of design effort, the proposed All-Spin PUF reuses the scan chain to replace the PUF array for response generation. By using the existing all-spin circuit, the signature is reliable with high throughout (64 bits per clock cycle), and All-Spin PUF has significant area benefits over other designs.

5 CONCLUSION

In this work, we propose a novel delay PUF (All-Spin PUF) design for the emerging STT-mCell based all-spin circuits. By incorporating the automatic write-back scheme, the generation of raw response-bits produced from
the All-Spin PUF can be stabilize. The uniqueness of the signature is improved by signature masking. Simulation results show that All-Spin PUF is reliable considering environment and process variations. Additionally, the All-Spin PUF reusing the existing DFT structure has higher energy and area efficiency compared to the state-of-the-art PUF designs.

REFERENCES


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," 2015 IEEE International Solid-State Circuits Conference -


