<table>
<thead>
<tr>
<th>Title</th>
<th>Highly reliable spin-transfer torque magnetic RAM based physical unclonable function with multi-response-bits per cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Zhang, Le; Fong, Xuanyao; Chang, Chip-Hong; Kong, Zhi Hui; Roy, Kaushik</td>
</tr>
<tr>
<td>Date</td>
<td>2015</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/25495">http://hdl.handle.net/10220/25495</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: [Article DOI: <a href="http://dx.doi.org/10.1109/TIFS.2015.2421481">http://dx.doi.org/10.1109/TIFS.2015.2421481</a>].</td>
</tr>
</tbody>
</table>
Abstract—Memory-based Physical Unclonable Function (MemPUF) has gained tremendous popularity in the recent years to securely preserve secret information in computing systems. Most MemPUFs in the literature have unreliable bit generation and/or are incapable of generating more than one response-bit per cell. Hence, we propose a novel MemPUF exploiting the unique characteristics of Spin-Transfer Torque Magnetic RAM (STT-MRAM) that can overcome these issues. Bit generation in our STT-MRAM based MemPUF is stabilized using a novel automatic write-back technique. Also, the alterability of the Magnetic Tunneling Junction (MTJ) state is exploited to expand the response-bit capacity per cell. Our analysis demonstrated the advantage of our scheme in reliability enhancement (Bit-Error Rate from $\sim 10^{-1}$ to $\sim 10^{-6}$ in the worst-case under varying conditions) and response-bit capacity per cell improvement (from 1 bit to 1.48 bits). In comparison with the conventional MemPUFs, our approach is also better in terms of the average chip area and energy for producing a response-bit.

Index Terms—Automatic write-back, bit alteration, Spin-Transfer Torque Magnetic RAM (STT-MRAM) MemPUF, Physical Unclonable Function.

I. INTRODUCTION

Security has become one of the most pressing concerns in memory design after a growing number of attacks to obtain sensitive information (such as cryptographic keys) stored in conventional storage devices have been reported [1], [2]. Memory-based Physical Unclonable Function (MemPUF), has emerged as an inexpensive and yet highly secure way of generating and storing the cryptographic key in memory for operations such as authentication of devices, protection and communication of confidential information [3]–[7]. It harvests the intrinsic manufacturing process variations of memory or memory-like array to produce a set of chip-unique output bits (responses) from a set of addresses (challenges) that are prohibitively hard to simulate, emulate or predict. The exact mechanism of the challenge-response generation cannot be copied to another physical device due to the imperfections and uncertainties in the fabrication technology, hence the term Physical Unclonable Function (PUF). As the number of possible Challenge-Response Pairs (CRPs) is linearly related to the size of the memory, MemPUFs are usually classified as “weak PUF” or “obfuscating PUF”. This is in contrast to “strong PUF” that possesses exponential number of possible CRPs with respect to the number of physical devices [8]. The advantages of MemPUFs over other types of PUFs are their simple structure and the efficiency of implementation in memory sub-systems.

Although many MemPUFs have been proposed, most of them suffer from two key problems: 1) the reliability of response-bit generation is poor, and 2) only one bit can be generated from each cell. The most convenient way to solve the aforementioned issues is to use architectural techniques such as Error Correction Code (ECC) [9], [10] and interface module [11], [12]. ECC stabilizes the noisy response-bits generated from the MemPUF, while the interface module such as control-logic obfuscates the stabilized bits to equivalently increase the response-bit capacity for each MemPUF cell. However, if each cell can originally produce only one response-bit and reliability of the raw response-bits are poor, the overheads of these architectural techniques may be very costly. Circuit-level techniques can help to relax the burden of architecture-level methods if they are properly applied. Bhargava et al [13], [14] used the accelerated aging effects to skew the bi-stable PUF cells so as to produce reliable bits. However, aging acceleration requires dedicated circuit components or additional testing procedure, and also degrades the performance of the MemPUF cells if the MemPUF is used as normal memory [15]. Cortez et al [16] proposed to adapt the ramping time during power-up process of SRAM-PUF to temperature so as to improve the reliability of bit generation, but this degrades the speed of the bit generation process and thus makes the MemPUF inapplicable in certain conditions.

Another approach to improving MemPUF design is to expand the response-bit capacity by physical means, which may be achievable in certain MemPUFs. For example, the “re-configurable PUF” [7], [17], [18], which may be implemented using phase change material, is capable of producing refeshable response-bits when the PUF devices are reconfigured by external effects. Zheng et al in [19] makes use of SRAM write-failure effects to produce random bits, and tunes the supply voltage to increase the entropy contained in each cell. These MemPUFs may be compromised if adversaries know the digital signals that regulate the external effects (e.g., electrical pulse [7], [17] and supply voltage [19]) for bit expansion, i.e., the adversaries may exploit the obtained information to predict
the results of response-bit expansion.

To this end, a MemPUF with multi-response-bits per cell capacity and inexpensive effective reliability enhancement technique is therefore required. In this paper, we propose such a MemPUF by exploiting the physical variability of the emerging Spin-Transfer Torque Magnetic RAM (STT-MRAM). STT-MRAM has emerged as the most promising candidate for universal memory technology, due to its non-volatility, compatibility with CMOS process, near unlimited endurance and high integration density [20], [21]. Recently, some PUF designs based on other emerging non-volatile memory technologies have been proposed [22]–[24], but none have explicitly suggested any efficient design techniques to enhance the reliability and response-bit capacity for each memory cell. In our proposed PUF, multiple response-bits can be extracted by using resistance mismatches between two STT-MRAM cells that store the same state in the Magnetic Tunnel Junctions (MTJs). In addition, we also propose circuit-level design techniques that take advantage of the unique features of the MTJ to enhance the reliability of our PUF. Compared to the conventional methods such as those proposed in [7], [13], [14], [16]–[18], our design takes advantage of the intrinsic device characteristics which are natively exploitable via normal memory operations, and avoids the use of techniques such as accelerated aging, ramping time adaption, and external effects that alter device properties for regular function. It has eliminated the need for both a pseudo random number generator and a protected non-volatile memory for the generation, storage and retrieval of secret keys in the algorithmic description of cryptographic primitives. The main contributions of our work can be summarized as follows.

1) We analyzed process variations of STT-MRAM devices, and designed a novel MemPUF using STT-MRAM.
2) We proposed an Automatic Write-Back (AWB) technique to enhance the reliability of bit regeneration.
3) We leveraged the spin states of MTJ devices to expand the response-bit capacity of each MemPUF cell.
4) We performed a comprehensive evaluation of the proposed MemPUF with commonly used figure-of-merits for PUFs and compared our proposed design with traditional MemPUFs.

The rest of the paper is organized as follows. The fundamentals of STT-MRAM are reviewed in Section II. In Section III, the design of the proposed STT-MRAM is presented and the AWB scheme and bit alteration phenomenon are described. Next, the quality of our proposed STT-MRAM based MemPUF is evaluated and analyzed in Section IV. Finally, we conclude the paper in Section V.

II. STT-MRAM

A. Preliminaries

An STT-MRAM cell consists of an MTJ and an access transistor as shown in Fig. 1(a). The MTJ is the basic storage element in STT-MRAM. An MTJ consists of two ferromagnetic layers sandwiching a thin tunneling oxide barrier [see Fig. 1(b)]. One ferromagnetic layer is magnetically pinned (which we call the pinned layer, PL), while the other is not (which we call the free layer, FL). The MTJ state is stored as the magnetization of the FL with respect to that of the PL, and may be sensed as the MTJ resistance, \( R_{MTJ} \). When the magnetization of the FL is the same as that of the PL, the MTJ is in the parallel state (or P) and \( R_{MTJ} = R_P \) (low resistance). When the magnetization of the FL and the PL are opposite (known as anti-parallel, or AP state), \( R_{MTJ} = R_AP \) (high resistance). Hence, different logic bits can be obtained by sensing \( R_{MTJ} \): “0”: \( R_{MTJ} = R_P \); “1”: \( R_{MTJ} = R_AP \). The distinguishability between \( R_P \) and \( R_AP \) is defined as the Tunnel Magneto-Resistance Ratio (TMR), and is given by

\[
TMR = \frac{R_AP - R_P}{R_P}
\]  

(1)

The MTJ state may be programmed by passing a current pulse through it to change the FL magnetization. The current flows from FL to PL to program the MTJ from AP to P. The current direction is reversed to program the MTJ from P to AP. To successfully program the MTJ within a given delay, the current amplitude needs to be larger than a critical current (denoted as \( I_C \)). The write operation is accomplished by setting the voltages of the Bit-Line (BL) and the Source-Line (SL) [see Fig. 1(c)].

B. Variations

Process variations significantly affect the characteristics of STT-MRAM. The geometry of the STT-MRAM structure, for example, MTJ oxide thickness, \( t_{MgO} \), and cross-section area, \( Area \), NMOS channel length, \( L \), and width, \( W \), may vary due to imperfections in the fabrication process [26], [27]. In addition, point defect sites such as Oxygen Vacancies (OVs) may also be formed during deposition if the process condition
The RSD of $eOV$ is equal to 0.24% which causes the RSD of TMR to be 3% and matches the measurement results.

The resistance of each MTJ in STT-MRAM varies from 0.1 to 0.6 when the MTJ oxide layer is free of OV and $eOV$ is assumed to be 0 when the OV reaches its maximum value. $eOV$ was integrated into our MTJ device model and calibrated against the experimental data in Fig. 3 when the RSD of $eOV$ is assumed to be 1. The RSD of $eOV$ is 0.24%.

III. Extracting the Response-bits

The resistance of each MTJ in STT-MRAM varies from each other due to process variations. Hence, the measured current during the sensing of $R_{MTJ}$ also varies. Since $R_{MTJ}$ variation is an intrinsic property of STT-MRAM, it can be exploited for generation of random chip-specific response-bits for security applications.

A. Design of the circuit and system

The conventional STT-MRAM may be modified to design MemPUF. A pair of STT-MRAM cells forms a Secret Bit (SB) cell, which is replicated to construct an array (Fig. 5). Consider when both MTJs in an SB cell selected by the memory address (by setting $RSE$ and $WL$) are identically set to AP (or to P) and a read operation is performed on the cell (i.e., $RDEN = 1$). Due to mismatches between the STT-MRAM cell resistances in the SB cell, the outputs of the sense amplifier, which have been pre-discharged to $VDD/2$, will be driven to different potentials. Note that the differential nature of the sensing scheme cancels out the systematic variations of the STT-MRAM cells in the SB cell. The voltage at node $BIT$ is then latched as a full-swing voltage signal, which represents the response-bit extracted from the SB cell. To ensure that the output bits are dominated by the mismatches of the MTJs in the SB cell, the other transistors ($M1$, $M2$, $M5$, and $M6$, and those used in the sense amplifier) are up-sized to mitigate their impact on the output bit generated.

B. Automatic Write-Back

Maintaining the reliability of response-bit generation under varying conditions is a major challenge for MemPUF. The reliability of producing a response-bit is defined as the probability that bit $b_t$ generated from a selected SB cell at time point $t$, is reproduced as $b_{t+\delta t}$ at time $t + \delta t$ ($\delta t > 0$) (see Fig. 6). If the mismatches between memory cells in MemPUFs are too...
small, the reliability of bits generated from the cells may be degraded by noise effects from the memory devices themselves and/or cross-talk from external components. For example, the thermal fluctuation in the FL magnetization (measured as angle $\theta$ relative to the PL magnetization) [30] may cause the actual $R_{MTJ}$ to vary over time due to the thermal field [see Fig. 7(a)] [27], [30]. Hence, the reliability of our proposed STT-MRAM based MemPUF may be degraded if the thermal fluctuation effect affects the bit generation process more than the other intrinsic MTJ variations. Assuming that $\theta$ is distributed as in [30], Fig. 7(b) plots the complementary reliabilities (represented by Bit-Error Rates or BERs) at different temperatures as estimated by our simulation. It can be seen that the BERs increase with increasing ambient temperature and reaches 20\% in the worst case.

We propose an Automatic Write-Back (AWB) scheme that exploits the non-volatility of the MTJ to improve the reliability of raw response-bits generated from our proposed MemPUF. The basic idea of AWB is that when a response-bit is initially produced from an SB cell, the generated bit is automatically written back into the SB cell upon completion of the read operation to make the bit reproducible thereafter [see Fig. 8(a)]. Incorporated with the AWB scheme, STT-PUF operates in two phases:

- **Enrollment phase:** MTJs in a selected SB cell are initialized ($INIT = 1$) into the same state (either AP or P, depending on the value of Data_init). The mismatch of MTJ resistances in the SB cell is then sensed to produce a random bit. When the read operation finishes, the generated bit (Data_wb) is automatically (triggered by the completion signal of read operation) written back to the SB cell by setting the two MTJs into complementary states.

- **Regeneration phase:** Response-bits are directly produced from the SB cells whose MTJs have been set to the complementary states ($INIT = 0$ in this phase).

Fig. 8(b) and 8(c) show the timing diagrams of a write-back-after-read process in an SB cell with both MTJs initialized to “0” or P. Once the bit is read (= 1 in this case), the MTJ in the left branch is written back with “1” (i.e., switched to AP).

The security of AWB is guaranteed by its “spontaneous” execution upon the completion of read operation in the enrollment phase. Therefore, malicious write is prohibited during the write-back process. Note that our proposed AWB scheme may also be used in MemPUFs based on other emerging non-volatile memory (eNVM) technologies. However, most eNVMs have very limited endurance that may limit the total...
cycles of write-back operations. Table I lists the possible eNVMs and the relevant properties that are crucial to the MemPUF qualities. The write endurance of MTJs is much higher than other eNVM (e.g., Phase Change Memory and FLASH). Hence, the AWB scheme does not wear out STT-MRAM based MemPUF as fast as MemPUFs based on the other nonvolatile memory technologies. Furthermore, STT-MRAM has several advantages over other types of eNVMs: 1) the density of STT-MRAM is high, which makes it infeasible for tampering activities such as probing the internal nodes or wires; 2) since the current required for successful MTJ switching is not as large as that used in MRAMs [20], the electro-magnetic emanation may be too weak to be exploited by side-channel attacks. The other eNVMs listed in Table I may leak measurable information (e.g., electro-magnetic emanation, heat generation, and optical radiation) which can be exploited to guess the bit that is written back.

Compared to the reliability improvement techniques proposed in the literature [13, 14, 16], 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 prolonging the time delay which may degrade performance. Secondly, our method does not rely on aging effects on the memory devices for reliability enhancement. Aging effects such as negative-bias temperature instability and hot-carrier injection may change the properties of the CMOS devices permanently. Thus, the deliberately aged memory-based PUF is unsuitable for use as regular memory as the stability of read/write operations will be vastly deteriorated [15]. On the other hand, the STT-MRAM can be reused as a regular memory even with the AWB scheme incorporated. Though reading/writing MTJ devices excessively may also result in aging effect such as time-dependent dielectric breakdown, previous studies concluded that an MJT device can normally endure \( \sim 10^{15} \) cycles before breaking down [35]. This period of time is sufficient for the STT-MRAM to be used for both reliable PUF and regular memory. When switching from PUF to memory mode of operation, the MTJ states of a MemPUF cell will be erased (i.e., written by 0) before any read/write operations are performed. In this way, the secret bit generated in the PUF mode cannot be reproduced or predicted.

### C. Bit Alteration

While most MemPUFs are capable of producing only one bit from each cell, our proposed STT-MRAM based MemPUF is able to generate \( \geq 1 \) bit from each cell. This is achievable because the mismatches within the SB cell depend on the stored state of the MTJs. Consider two MTJs in an SB cell, represented by \( MTJ_A \) and \( MTJ_B \), and with parameters, \( t_{MgO,A} \) and \( eOV_A \), and \( t_{MgO,B} \) and \( eOV_B \), respectively. Assuming \( t_{MgO,A} = t_{MgO,B} \) but \( eOV_A > eOV_B \), then \( RA_A > RA_B \) and \( TMR_A < TMR_B \), where \( RA \) represents the resistance-area product of an MTJ. Since \( R_{AP} = (TMR+1)R_P \), the relationship of \( RA_{AP,A} \) and \( RA_{AP,B} \) is determined by \( \Delta RA_P = RA_{AP,A} - RA_{AP,B} \) and \( \Delta TMR = TMR_A - TMR_B \). If the signs of \( \Delta RA_P \) and \( \Delta TMR \) are opposite and the impact of \( \Delta TMR \) is more significant than \( \Delta RA_P \), \( RA_{AP,A} \) will be smaller than \( RA_{AP,B} \) (as shown in the case when \( t_{MgO,A} = t_1 \) in Fig. 9). At this time, \( RA_A > RA_B \) but \( RA_{AP,A} < RA_{AP,B} \), and the bit produced from the SB cell will be altered if the MTJ is initialized into different states during enrollment. We call this phenomenon bit alteration. If \( \Delta TMR \) is less significant than \( \Delta RA_P \) (\( t_{MgO,A} = t_2 \) in Fig. 9) or the signs of \( \Delta RA \) and \( \Delta TMR \) are the same (\( t_{MgO,A} = t_3 \) in Fig. 9), the relationships of \( RA_{AP,A} \) and \( RA_P \) between \( MTJ_A \) and \( MTJ_B \) are consistent and there is no bit alteration.

Bit alteration may be exploited to increase the amount of extractable information from each SB cell, i.e., \( \geq 1 \) bit can be generated from one SB cell. Ideally 2 bits can be achieved since there are two initial states for an SB cell, i.e., AP and P. For convenience, the initial state is represented by \( s \) which belongs to the set \( \{0, 1\} \). Due to the imperfection of randomness originated from each cell, normally less than 2

### Table I

| Properties of NVM candidates for MemPUF with AWB scheme |
|-----------------|-----------------|-----------------|-----------------|
|                | STT-MRAM         | MRAM            | PCM             |
| Neighbor-disturbance | \( > 10^{15} \)  | Magnetostatic interaction \[31\] | Thermal disturbance \[32\] |
| Tamper evidence (density of the array) | No | High | High |
| Side-channel | No | Electro-magnetic emanation \[31\] | Heat profile \[32\] |
| Data of endurance were from \[20\]. | 10^10 | Cell-to-cell interference \[33\] | Optical fault injection \[34\] |

Fig. 9. \( RA - t_{MgO} \) (top) and \( TMR - t_{MgO} \) (bottom) characteristics of MTJ. Bit alteration occurs when \( t_{MgO} \) and \( eOV \) meet certain condition. Labels \( A \) and \( B \) identify the two MTJs of an SB cell.

1 Other process parameters (e.g., the MTJ contact area and parameters of the access transistor) do not affect both MTJ \( TMR \) and resistance, and are therefore not taken into account in this analysis.
response-bits can be generated from a cell.

It is possible to re-enroll the STT-MRAM based MemPUF so that its response-bit map can be refreshed (see Fig. 10): in a new enrollment phase, before the response bits are generated, the configuration states $s$ of the SB cells in an $N$-bit array are iteratively set. The configuration state $s_{i}$ ($1 \leq i \leq N$) of the $i$-th cell is generated using the function $Update(s_{i-1}, \tau)$, where $s_{i-1}$ is the configuration state of the $(i-1)$-th cell, and $\tau$ is a binary bit from a one-time pad produced using an on-chip random source. The function $Update(\ldots)$ may be implemented by an exclusive OR operator, i.e., $s_{i} = s_{i-1} \oplus \tau$. After the configuration states for all the $N$ cells are set, the response generation starts. Owing to the bit alteration effects, after re-setting the configuration state of the array, the response bits generated from the PUF are refreshed. This property is called “physical reconfigurability” [18]. It can be exploited for enlarging the key space of the PUF, protecting PUF-based persistent storage, recycling the PUF and enhancing the security of PUF against possible leakage [17], [18], [36], [37].

Note that physical refreshing of response-bits is different from other methods which logically re-map the input-output of a PUF. For example, control-logic interfaces, which can be implemented by a hash function or block cipher, have been used to diffuse the raw bits produced from an internal PUF [36]. Compared to the physical means, logical counterpart consumes larger chip area because the control-logic interfaces are hardware-expensive. The response-bit refreshability of our STT-MRAM MemPUF is originated from the intrinsic properties of the MTJ. No additional hardware entropy sources are needed to realize it. Furthermore, the response bit refreshability does not rely on any quantization of electrical signals (e.g., supply voltage or programming pulses, etc.) to increase the extractable entropy from each cell, as used in the contemporary designs [7], [19]. Leakage of these signals may help an adversary to guess the output response-bit from the MemPUF cell during/after the re-enrollment. For example, in the PCM-based MemPUF, learning the configuration state that regulates the programming pulses for response refreshment allows adversaries to predict the output bit according to the characteristics of the pulses. However, in the STT-MRAM MemPUF, the adversary has no idea about the response-bit generated from the $i$-th SB cell even if $s_{i}$ is known.

![Fig. 10. By resetting the state of each MemPUF cell (assuming the MemPUF array consists of $N$ cells) with randomly generated $s_{i}$ ($1 \leq i \leq N$), bit produced from the cell may be altered.](image)

### IV. QUALITY EVALUATION

In this section, we evaluate our proposed STT-MRAM based MemPUF in terms of reliability, uniqueness and randomness by simulations. Due to the inaccessibility to STT-MRAM fabrication, our analysis is based on the MTJ device model that has been validated against experimental data (Section II-B) and the 45 nm bulk CMOS Predictive Technology Model used for transistors in the circuitry [38]. The area and energy consumption of our design were compared with conventional MemPUFs. The parameters used in our simulation are listed in Table II. The simulations were performed in HSPICE and the results were processed and analyzed in MATLAB.

#### A. Reliability

Based on the definition introduced in Section III-B, three cases need to be considered for evaluating the MemPUF reliability:

1) $b_{t}$ and $b_{t+\delta t}$ are both generated in the enrollment phase. STT-MRAM MemPUF operating in this case can be regarded as a MemPUF without AWB scheme.
2) $b_{t}$ is generated in the enrollment phase while $b_{t+\delta t}$ is generated in the subsequent regeneration phase.
3) $b_{t}$ and $b_{t+\delta t}$ are both generated in the regeneration phase.

In case #1, as introduced earlier in Section III-B, the reliability can be significantly degraded by the noise effects under varying environmental conditions [see Fig. 7(b)]. In case #2 and case #3, the reliability is higher than that in case #1 because response-bits are reproduced from two MTJs (in an SB cell) which are in the complementary states after write-back. However, failure effects, i.e., read/write failures, may affect the reliability of bit regeneration. Write failure occurs when the current injected into the MTJ is not able to switch the MTJ state [39], thus resulting in unsuccessful write-back and the two MTJs in the SB cell will remain in the same state. As a consequence, the reliability becomes as low as that in case #1. Read-disturb failures affect the SB cell reliability.

#### TABLE II

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Nominal values</th>
<th>RSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact area</td>
<td>45 nm × 116 nm</td>
<td>5%</td>
</tr>
<tr>
<td>Oxide layer thickness</td>
<td>1.15 nm</td>
<td>2%</td>
</tr>
<tr>
<td>Effective oxygen vacancy</td>
<td>0.03</td>
<td>2.4%</td>
</tr>
<tr>
<td>MOSFET channel length</td>
<td>45 nm</td>
<td>10%</td>
</tr>
<tr>
<td>MOSFET threshold voltage</td>
<td>0.466 V</td>
<td>10%</td>
</tr>
<tr>
<td>Supply voltage</td>
<td>1 V</td>
<td></td>
</tr>
<tr>
<td>Write cycle</td>
<td>40 ns</td>
<td></td>
</tr>
<tr>
<td>Read cycle</td>
<td>20 ns</td>
<td></td>
</tr>
</tbody>
</table>
Based on the failure models in Fig. 11(a) and 11(b), BERs may become as low as that in case #2 and case #3, respectively. During a read operation in the regeneration phase, read-current that is unexpectedly larger than the critical current may switch the state of MTJ. Hence, the two MTJs in the SB cell are back to the identical state. Again, the reliability becomes as low as that in case #1. Sense-decision failure occurs when \( R_P \) is larger than \( R_{AP} \) of the other MTJ in the SB cell, and therefore, response-bits reproduced from this kind of cells will be different from that generated previously at \( t \).

Failure effects in STT-MRAM can be attributed to process variations as well as thermal fluctuation effects [40], [41]. Based on the failure models in Fig. 11(a) and 11(b), BERs due to these effects can be estimated from the probabilities of relevant failure effects that make \( B_{t+\delta t} \neq b_t \). The Response Surface Modeling technique was used to numerically compute the probabilities given the distribution of the relevant variables (e.g. process parameters, initial angles, etc.) [27], [40]. In our STT-MRAM PUF, the read/write delays are fixed to meet the performance requirements (see Table II). The access transistors can be sized properly to achieve the minimum BER [27].

Fig. 12 shows the BERs for case #2 and case #3 with varying access transistor width in an SB cell. Increasing transistor width reduces write failures because the driving capability of the access transistor is enhanced. Also, the sense-decision failure reduces write failures because the equivalent resistance of the memory cell becomes dominated by the MTJ resistances. On the other hand, read disturbance failure is increased and may become significant when the transistor size is larger than certain value [39]. An optimum width of 1060 nm is found to have the lowest error rate considering both failure effects.

Failure effects in STT-MRAM can be attributed to process variations as well as thermal fluctuation effects [40], [41]. Based on the failure models in Fig. 11(a) and 11(b), BERs due to these effects can be estimated from the probabilities of relevant failure effects that make \( B_{t+\delta t} \neq b_t \). The Response Surface Modeling technique was used to numerically compute the probabilities given the distribution of the relevant variables (e.g. process parameters, initial angles, etc.) [27], [40]. In our STT-MRAM PUF, the read/write delays are fixed to meet the performance requirements (see Table II). The access transistors can be sized properly to achieve the minimum BER [27].

Fig. 12 shows the BERs for case #2 and case #3 with varying access transistor width in an SB cell. Increasing transistor width reduces write failures because the driving capability of the access transistor is enhanced. Also, the sense-decision failure reduces write failures because the equivalent resistance of the memory cell becomes dominated by the MTJ resistances. On the other hand, read disturbance failure is increased and may become significant when the transistor size is larger than certain value [39]. An optimum width of 1060 nm is found to have the lowest error rate considering both failure effects.

Failure effects in STT-MRAM can be attributed to process variations as well as thermal fluctuation effects [40], [41]. Based on the failure models in Fig. 11(a) and 11(b), BERs due to these effects can be estimated from the probabilities of relevant failure effects that make \( B_{t+\delta t} \neq b_t \). The Response Surface Modeling technique was used to numerically compute the probabilities given the distribution of the relevant variables (e.g. process parameters, initial angles, etc.) [27], [40]. In our STT-MRAM PUF, the read/write delays are fixed to meet the performance requirements (see Table II). The access transistors can be sized properly to achieve the minimum BER [27].

Fig. 12 shows the BERs for case #2 and case #3 with varying access transistor width in an SB cell. Increasing transistor width reduces write failures because the driving capability of the access transistor is enhanced. Also, the sense-decision failure reduces write failures because the equivalent resistance of the memory cell becomes dominated by the MTJ resistances. On the other hand, read disturbance failure is increased and may become significant when the transistor size is larger than certain value [39]. An optimum width of 1060 nm is found to have the lowest error rate considering both failure effects.

Failure effects in STT-MRAM can be attributed to process variations as well as thermal fluctuation effects [40], [41]. Based on the failure models in Fig. 11(a) and 11(b), BERs due to these effects can be estimated from the probabilities of relevant failure effects that make \( B_{t+\delta t} \neq b_t \). The Response Surface Modeling technique was used to numerically compute the probabilities given the distribution of the relevant variables (e.g. process parameters, initial angles, etc.) [27], [40]. In our STT-MRAM PUF, the read/write delays are fixed to meet the performance requirements (see Table II). The access transistors can be sized properly to achieve the minimum BER [27].

Fig. 12 shows the BERs for case #2 and case #3 with varying access transistor width in an SB cell. Increasing transistor width reduces write failures because the driving capability of the access transistor is enhanced. Also, the sense-decision failure reduces write failures because the equivalent resistance of the memory cell becomes dominated by the MTJ resistances. On the other hand, read disturbance failure is increased and may become significant when the transistor size is larger than certain value [39]. An optimum width of 1060 nm is found to have the lowest error rate considering both failure effects.
is low because write-failure is the dominant failure. When the supply voltage ramps up, the BERs drop and eventually converge to the BER for case #3 because read failures become dominating at high supply voltage. Fig. 13(b) shows BERs under different temperatures (from 200 K to 400 K) with supply voltage fixed at 1 V. BERs for all three cases increase with temperature because increasing temperature leads to more severe thermal fluctuation effects, which in turn leads to increased read-disturb failures. Furthermore, the increase in temperature results in a lower $T_{MR}$ [42] and increases sense-decision failures. Though the reliability of our STT-MRAM MemPUF with AWB scheme may be degraded by supply voltage fluctuations and temperature variations, the worst-case BER for all three cases under varying conditions is within the range of $10^{-6} \sim 10^{-5}$, which is far better than that of conventional MemPUFs or the original STT-MRAM MemPUF without AWB. As we will discuss later, the high reliability of raw response-bits significantly relaxes the burden of ECC by reducing the redundancies required for error correction.

**B. Uniqueness And Randomness**

Uniqueness and randomness are crucial to the security of MemPUF [43]. The uniqueness measures how the generated response-bits using different cells vary from each other. This is estimated by computing the fractional Hamming distance (denoted by $H$) of $m$ $l$-bit strings $B = b_1||b_2||\ldots||b_l$ generated from randomly selected SB cells [44], i.e.,

\[
H = \frac{2l}{m \times (m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \text{HD}(B_i, B_j) \quad (2)
\]

where the function $\text{HD}$ computes the Hamming distance between two binary strings. Fig. 14(a) and Fig. 14(b) show the results calculated from 64-bit strings generated from SB cells initialized into P and AP states, respectively. The mean values of $H$ in both cases are close to 50%, which indicates a high uniqueness of the produced bits.

Entropy was used as a metric for evaluating the randomness of MemPUF, which is defined as follows,

\[
E = - \sum_{X \in \mathcal{X}} \log(\Pr(X))\Pr(X) \quad (3)
\]

where $X$ is a random event from the set $\mathcal{X}$. In our case, $X$ refers to an $l$-bit string generated using $l$ randomly selected SB cells. Since estimating the entropy of such a large output space using Eq. (3) is non-trivial, the Context-Tree Weighting (CTW) algorithm [43], [45] was adopted to compute the compression ratio of the generated bit strings, which is used as an alternative indicator of entropy. Fig. 15 shows the estimated entropies when the initial states are P and AP, with mean values equal to 0.99 and 0.98, respectively. These results were achievable due to the symmetric 2T2R structure of SB cells which cancels out the systematic variations of process parameters in the response-bit generation process. Therefore, the variations, due entirely to the mismatches between devices within an SB cell, offer excellent randomness for the generated bits.

Bit alteration also impacts the uniqueness and randomness of the generated bit strings when different initial states are used. Fig. 16(a) shows the bit-maps of a $32 \times 32$ SB cell array when the MTJs in each cell are initialized to P and AP states, respectively. Fig. 16(b) shows the bits obtained by XOR-ing the two bit-maps in Fig. 16(a), and 117 of the 1024 cells are alterable. The proportion of the alterable cells is fairly low due to self-correlation effects. Although the variations of $t_{MgO}$ and $\epsilon OV$ may result in bit alteration phenomenon, other relevant parameters (e.g., Area, $V_{th}$, etc.) create strong correlations between the cell resistances in the two spin states. For example, if the resistance of $MTJ_A$ is larger than that of $MTJ_B$ in P state mainly due to the difference in their junction areas, the resistance of $MTJ_A$ in AP state will be larger than that of $MTJ_B$ in AP state because $Area$ impacts MTJ resistances in P and AP states in the same way.

To measure the randomness of an SB cell considering bit alteration, we define equivalent entropy ($\bar{E}$) based on Eq. (3) but assuming that the random event $X$ here refers to the possible bit patterns that can be generated from one SB cell.
TABLE III
SECRET BITS GENERATED FROM ONE SB CELL WITH DIFFERENT INITIAL STATES.

<table>
<thead>
<tr>
<th>Bit pattern $(b_s=0, b_s=1)$</th>
<th>(0, 0)</th>
<th>(0, 1)</th>
<th>(1, 0)</th>
<th>(1, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>44.3%</td>
<td>4.9%</td>
<td>5.3%</td>
<td>45.5%</td>
</tr>
</tbody>
</table>

The probability that at least one bit is complemented in a cell among the selected cells is alterable, the bit string post-processed by the hash function will be completely diffused. The probability that at least one bit is complemented in an $l$-bit string given different initial states of the $l$ SB cells can be calculated as

$$
\Pr(\text{num}(b_s=0 \oplus b_s=1 = 1) \geq 1) = \sum_{i=1}^{l} \binom{l}{i} \Pr(b_s=0 \oplus b_s=1 = 1)^i (1 - \Pr(b_s=0 \oplus b_s=1 = 1))^{l-i}
$$

(6)

We varied the length of the response-bit string and obtained the results shown in Fig. 17(a). When the string-length is more than 512 bits, the probability that at least one in 512 cells is alterable is very close to 1. Further increase in string-length will converge the probability to 1. Using the post-processing hash function, the four bit patterns will have more balanced percentages [see Fig. 17(b)]. Hence, $E$ contained in each SB cell is improved to

$$
\hat{E} = - \left[ \log_2(0.252) \times 0.252 + \log_2(0.249) \times 0.249 + \log_2(0.248) \times 0.248 + \log_2(0.251) \times 0.251 \right] \approx 1.99997
$$

which is $\approx 2$.

C. Comparison

We compare our proposed STT-MRAM based MemPUF to conventional MemPUFs in terms of Hamming distance, BER (at different supply voltages and ambient temperatures), and equivalent entropy. The results are shown in Table IV. The proposed STT-MRAM based MemPUF without AWB exhibits randomness, uniqueness, and reliability that are comparable to conventional PUFs. The reliability of the STT-MRAM based MemPUF after adopting the AWB scheme is significantly improved ($\sim 10^4 \times$). Additionally, 1.48 bits can be produced from each cell of the proposed MemPUF owing to the bit alteration phenomenon, and this value can be improved to $\approx 2$ using a hash function to post-process the generated bits.

We also compare the instances in terms of energy and area consumption. Considering the entropy-loss due to the use of ECC\(^4\), we use the metric called “cell/bit” to measure the number of cells needed for generating a bit. Using an ECC which is specified by $[n, k, d]$ [49], cell/bit is equal to $n/k$. Also, due to the bit alteration effect, the number of raw bits (before being stabilized by ECC and obfuscated by hash function) that can be produced from each cell is equal to the equivalent entropy, $\hat{E}$, of the cell.

We used BCH-[255,k,t] as the ECC in our comparison, and the parameter $k$ was chosen to keep the worst-case BERs of the PUFs below $10^{-6}$ after the stabilization by ECC. Then area/bit was calculated by

$$
\text{area/bit} = \frac{\text{area/cell} \times n}{\hat{E}k}
$$

and energy/bit was calculated by

$$
\text{energy/bit} = \frac{\text{energy/cell} \times n}{\hat{E}k}
$$

\(^3\)For convenience of displaying the calculated values, complementary probability, i.e., $1 - \Pr(\text{num}(b_s=0 \oplus b_s=1 = 1) \geq 1)$ is plotted instead.

\(^4\)Using an ECC will induce an entropy-loss of the originally generated bits, i.e., a number of bits which are equal to the ECC redundancy will be truncated.

---

\(^2\)More specifically, strict avalanche criterion requires that one bit flip in the input string to the hash function will cause each bit of the output string to flip with an average probability of 50%.
where “area/cell” is estimated based on the \( \lambda \)-based rule (\( \lambda \) is equal to half of the smallest feature size) for 2-finger SB cell [see Fig. 18(b)] layout which consumes less area than 1-finger layout (Fig. 18(a)) [50]. For STT-MRAM based MemPUF incorporated with the AWB scheme, “energy/cell” in the enrollment phase is calculated by

\[
\text{energy}_{\text{EB}} = \frac{1}{2} \sum_{s=\{0,1\}} \left[ \text{energy}_{(s)}(\text{INIT}) + \text{energy}_{(s)}(\text{READ}) + \text{energy}_{(s)}(\text{AWB}) \right]
\]

and the energy consumed in the regeneration phase is

\[
\text{energy}_{\text{FR}} = \text{energy}(\text{READ})
\]

For STT-MRAM based MemPUF without AWB scheme, only read operation has to be considered.

The calculated results are listed in Table V. Note that the energy consumed by enrollment is relatively high due to the high write current required by an MTJ for state switching, while regenerating bits, i.e., sensing \( R_{MTJ} \), consumes much lesser energy (\( \approx 2 \times \) less compared to the least of the referred MemPUF, i.e., Latch-based). Note also that the frequency of enrollment is related to the security level of the system.

Intuitively, the more frequent the enrollment is performed, the more difficult it is for adversaries to predict the response-bits. In practical designs, when the energy budget and hardware resources are limited, optimization between energy consumption and system-security should be taken into consideration.

SB cell implemented using STT-MRAM has a smaller footprint compared to the conventional implementations using CMOS technology and hence, our proposed STT-MRAM based MemPUF consumes less area (\( \approx 11 \times \) less than the least of the referred instances (i.e., SRAM-based)). Furthermore, using AWB scheme reduces the area/bit tremendously by almost 200\% (see Table V) due to the high reliability of the raw response-bits, which relaxes the ECC requirements to achieve an acceptably high reliability.

MemPUF designs based on other types of non-volatile memory technologies [22]–[24], which may not specifically tackle secure key storage problem, use different method to extract the entropies. For example, in [22], a Public PUF (PPUF) is proposed for secure time-bounded authentication. The principle is similar to the classical public-key cryptography except that it relies on the complexity discrepancy between using a hardware simulator and physical device to evaluate the same amount of response bits. It is difficult to make a fair comparison with these designs since their reliabilities are not reported in the references. Otherwise, the response randomness (measured in uniformity) and uniqueness of these PUFs and ours are comparable (the differences are all within one percent). As far as entropy, or equivalent entropy as defined and used in our paper, is concerned, our proposed STT-MRAM PUF has an advantage since the other instances produce only one bit from an individual cell.

### Table IV

<table>
<thead>
<tr>
<th>Instances</th>
<th>Fractional ( H )</th>
<th>BER (235 K)</th>
<th>BER (298 K)</th>
<th>BER (358 K)</th>
<th>Equivalent Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(0.9 V)</td>
<td>(1.0 V)</td>
<td>(1.1 V)</td>
<td></td>
</tr>
<tr>
<td>SRAM</td>
<td>49.7%</td>
<td>8%</td>
<td>6%</td>
<td>8%</td>
<td>0.99</td>
</tr>
<tr>
<td>FF</td>
<td>41.8%</td>
<td>17%</td>
<td>4%</td>
<td>21%</td>
<td>1</td>
</tr>
<tr>
<td>Latch</td>
<td>36.9%</td>
<td>28%</td>
<td>3%</td>
<td>18%</td>
<td>-</td>
</tr>
<tr>
<td>Buskeeper</td>
<td>49.1%</td>
<td>11%</td>
<td>4%</td>
<td>20%</td>
<td>0.99</td>
</tr>
<tr>
<td>STT-MRAM, original</td>
<td>49.9%/50.2%</td>
<td>5%</td>
<td>8%</td>
<td>21%</td>
<td>( \approx 1.48 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8%</td>
<td>8%</td>
<td>8%</td>
<td>( \approx 2 )</td>
</tr>
<tr>
<td>STT-MRAM, with AWB</td>
<td>49.9%/50.2%</td>
<td>( 4.1 \times 10^{-6} )</td>
<td>( 4.2 \times 10^{-6} )</td>
<td>( 4.6 \times 10^{-6} )</td>
<td>( \approx 1.48 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.6 \times 10^{-6}</td>
<td>4.2 \times 10^{-6}</td>
<td>4.0 \times 10^{-6}</td>
<td>( \approx 2 )</td>
</tr>
</tbody>
</table>

1. Data of the fractional Hamming distances and the BERs of SRAM, FF, Latch and Buskeeper were from [46], [47]. The data of entropies were from [47].
2. Worst-case BERs under varying conditions are considered in the comparison. It should be noted that the referred PUF instances [46], [47] were fabricated using 65-nm technology while our proposed STT-MRAM was evaluated using 45 nm technology, but we argue that the worst-case reliability of these MemPUF instances will not be better if more scaled CMOS technology is used.
TABLE V
COMPARISON OF THE PROPOSED STT-MRAM PUFs WITH THE CONVENTIONAL DESIGNS IN TERMS OF energy/bit AND area/bit

<table>
<thead>
<tr>
<th>Instances</th>
<th>energy/bit (pJ)</th>
<th>area/bit (μm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRAM</td>
<td>9.78 × 10⁻³</td>
<td>4.81</td>
</tr>
<tr>
<td>FF</td>
<td>4.33 × 10⁻²</td>
<td>35.47</td>
</tr>
<tr>
<td>Latch</td>
<td>1.48 × 10⁻²</td>
<td>20.16</td>
</tr>
<tr>
<td>Buskeeper</td>
<td>5.10 × 10⁻³</td>
<td>14.43</td>
</tr>
<tr>
<td>Proposed, original</td>
<td>0.63 × 10⁻³</td>
<td>1.28</td>
</tr>
<tr>
<td>Proposed, with AWB</td>
<td>14.31(E), 0.69 × 10⁻³(R)</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Energy/bit of the referenced PUF designs were estimated in HSPICE with 45 nm bulk PTM. Area/bit were estimated with the cell layout based on the 22-nm-based design rules.

response-bits produced from the STT-MRAM can be improved from ∼ 10⁻¹ to ∼ 10⁻⁶ in the worst case under varying operating conditions. Using the bit alteration phenomenon in MTJ devices, 1.48 bits can be generated from each SB cell and the equivalent bit capacity can be further improved to ≈ 2 by employing a hash function. The high reliability of our proposed STT-MRAM based MemPUF greatly lowers the burden of ECC. Furthermore, response-bit map of the STT-MRAM MemPUF can be dynamically refreshed when it is required so that predicting the response-bits generated from the MemPUF becomes more difficult. Our comparative studies demonstrated the superiority of our schemes in area and energy consumption compared to the conventional MemPUFs (≥ 11× smaller in area/bit, and ≥ 2× smaller in energy/bit).

ACKNOWLEDGMENT
This project is supported by the Singapore Ministry of Education Tier II Research Grant No. MOE2013-T2-2-017. X. Fong and K. Roy’s work were supported in part by National Science Foundation, Semiconductor Research Corporation, and by the Center for Spintronics Materials, Interfaces, and Architecture (C-SPIN), a StarNet center funded by MARCO and DARPA.

REFERENCES


Le Zhang (S’11) received the B.Eng. degree from Harbin Institute of Technology, China in 2010. He is now working towards Ph.D at School of Electrical and Electronic Engineering in Nanyang Technological University. He was a visiting scholar to Purdue University in 2013. His current research interests are low-power and process variation tolerant circuits and systems, non-volatile memory, hardware security and physical unclonable function.

Xuanyao Fong (M’06) received the Ph.D. and B.Sc. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 2004 and 2006, respectively.

From January to August 2007, he was an Intern Engineer with Advanced Micro Devices, Inc., Boston Design Center, Boxboro, MA. He was a Research Assistant to Professor Kaushik Roy in the Nanoelectronics Research Laboratory, Purdue University. Currently, he is a Postdoctoral Research Assistant to Professor Kaushik Roy in the Nanoelectronics Research Laboratory, Purdue University. His research interests include device/circuit/architecture co-design for Si and non-Si nanoelectronics; design of VLSI logic and memory systems using spintronic devices, circuits, and architectures; and non-Boolean and analog computing paradigms using emerging technologies.

Dr. Fong received the “AMD Design Excellence Award” at Purdue in 2008, and the best paper award at the 2006 International Symposium on low power electronics and design.

Chip-Hong Chang (S’92-M’98-SM’03) received the B.Eng. (Hons.) degree from the National University of Singapore in 1989, and the M.Eng. and Ph.D. degrees from Nanyang Technological University (NTU) in 1993 and 1998, respectively.

He served as a Technical Consultant in industry prior to joining the School of Electrical and Electronic Engineering (EEE) of NTU in 1999, where he is currently an Associate Professor. He holds joint appointments with the university as Assistant Chair of Alumni of the School of EEE since June 2008, Deputy Director of the Center for High Performance Embedded Systems from 2000 to 2011, and the Program Director of the Center for Integrated Circuits and Systems from 2003 to 2009. He has coedited one book, published four book chapters and around 200 research papers in refereed international journals and conferences. His current research interests include hardware security and trust, low power arithmetic circuits and digital filter design.

Dr. Chang has served as Associate Editor of IEEE Access since 2013, IEEE Transactions on Circuits and Systems-I from 2010-2013, IEEE Transactions on Very Large Scale Integration (VLSI) Systems since 2011 and Integration, the VLSI Journal since 2013, and the Editorial Advisory Board Member of Open Electrical and Electronic Engineering Journal and Journal of Electrical and Computer Engineering. He also guest edited several journal special issues and served in many international conference advisory and technical program committees. He is a Fellow of the IET.

Zhi Hui Kong received the B.Eng. (Hons) degree from University of Technology, Malaysia in 2000 and Ph.D. from Nanyang Technological University (NTU), Singapore in 2006. From 2000 to 2002, Dr. Kong worked as a Research Engineer, first at Centre for Signal Processing then at Institute for Infocomm Research. She became a Project Officer in NTU in 2005 and Research Fellow in 2006. Dr. Kong worked as a Teaching Fellow with the School of EEE at NTU in 2007 and she is now an Assistant Professor with the Division of Circuits and Systems. She has published extensively in refereed international journals and conferences. Her international edition book is adopted as an undergraduate textbook. Her research interest is on energy efficient and variability-induced circuit design.
Kaushik Roy (F’02) received B.Tech. degree in electronics and electrical communications engineering from the Indian Institute of Technology, Kharagpur, India, and Ph.D. degree from the electrical and computer engineering department of the University of Illinois at Urbana-Champaign in 1990. He was with the Semiconductor Process and Design Center of Texas Instruments, Dallas, where he worked on FPGA architecture development and low-power circuit design. He joined the electrical and computer engineering faculty at Purdue University, West Lafayette, IN, in 1993, where he is currently Edward G. Tiedemann Jr. Distinguished Professor. His research interests include spintronics, device-circuit co-design for nano-scale silicon and non-Silicon technologies, low-power electronics for portable computing and wireless communications, and new computing models enabled by emerging technologies. Dr. Roy has published more than 600 papers in refereed journals and conferences, holds 15 patents, graduated 65 PhD students, and is co-author of two books on Low Power CMOS VLSI Design (John Wiley & McGraw Hill).