Reliable Memory PUF Design for Low-Power Applications

Mohammad Saber Golanbari, Saman Kiamehr, Rajendra Bishnoi, and Mehdi B. Tahoori
Department of Computer Science, Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
Emails: {golanbari, kiamehr, rajendra.bishnoi, mehdi.tahoori}@kit.edu

***Abstract*—This paper presents a reliable memory-based Physical Unclonable Function (PUF) design for operating at low supply voltages, which is typically demanded in emerging Internet of Things (IoT) applications with stringent energy constraints. PUF is a promising approach for generating unique and secure IDs based on the intrinsic uncontrollable manufacturing process variation. A common approach is to use the power-up values of SRAM memory arrays as the PUF response. However, reliability of the PUF response is a major concern for such designs, in particular, at low supply voltage values where the impact of noisy operating environment becomes significant. As a result, a noisy PUF response due to the non-ideal reliability at low supply voltages, has to be transformed into a stable high-entropy key by a key extractor circuitry, which imposes significant area and power overhead. The proposed PUF design in this paper has the advantage of being highly reliable at low supply voltages allowing aggressive supply voltage reduction for lower power consumption and better energy efficiency with lower area and overhead of key extractor. In this paper, we first evaluate the reliability of the SRAM-based PUFs over a wide range of supply voltages from the super-threshold voltage regime down to the Near-Threshold Voltage (NTV) regime. Based on this analysis, we propose a new memory-based PUF design which provides higher reliability ($2.6\times$ improvement) while consuming less power ($\sim 2\times$) compared to the traditional SRAM PUF designs in the NTV region.**

I. Introduction

The number of *Internet of Things* (IoT) devices has been growing tremendously over the last decade. As per *World Economic Forum* prediction, the number of connected devices will reach up to 50 billion by 2020. High security and energy efficiency are the major requirements for such devices. In secure systems, one of the essential requirements is the generation and storage of the secret keys, which are typically used in cryptographic algorithms, for device identification, or preventing counterfeit devices. Non-volatile memory such as *Electrically Erasable Programmable Read Only Memories* can be used to store the secret keys, but this approach is very expensive and vulnerable to physical and software attacks [1].

A promising alternative solution for generation and storage of the secret keys is to use a *Physical Unclonable Function* (PUF), which derives secret keys from unique physical characteristics of the system, such as manufacturing process variation, instead of actually storing it [2]. When the secret key is required, the PUF is invoked by a challenge (i.e. input), and the PUF response is collected as the secret key. The quality of a PUF is mainly measured based on *uniqueness* and *reliability* metrics. The uniqueness metric describes the unpredictability of the responses of different PUF instances, whereas the reliability metric explains the stability of the PUF response for the same chip in the presence of environmental variations such as temperature, power supply, and noise [3].

Operating the circuits at voltages close to the threshold voltage of the transistors, commonly known as Near-Threshold Computing (NTC), can effectively improve the energy efficiency by up to 10 times [4]. Therefore, this is a promising solution for energy-constrained IoT platforms [5]. However, the reliability of the circuits is typically deteriorated in this voltage regime [4]. PUFs are used for identification and for key generation as a part of cryptography modules, therefore, they are frequently used throughout the system operation, particularly in IoT platforms. Despite being a small component of a secure System-on-Chip (SoC), the reliability of the PUF can significantly affect the power consumption and energy efficiency of the entire SoC by increasing the complexity of the key extractor and by limiting the voltage scaling capability of the SoC.

A PUF response is typically impacted by noise as well as runtime and environment variations because the PUF response is collected under different operating and environmental conditions. Therefore, a key extractor, such as the one proposed in [6], is required to extract a stable and high entropy key from a PUF response. The area and power consumption of a key extractor highly depend on the quality of the original PUF response, however, they typically dominate the area and power of the entire PUF circuitry. A PUF design with higher reliability needs a significantly smaller key extractor, hence, has lower area and power footprint. Since the amount of noise in the PUF response is increased when the supply voltage is reduced, the complexity and overheads of the required key extractor increase dramatically, which may not be tolerated in low-power and energy efficient designs and in IoT applications. On the other hand, due to the limited number of power domains in a typical ultra-low power SoC, the voltage downscaling capability, and hence energy saving of the entire SoC is limited by the component that has the lowest reliability at low voltages. In other words, the supply voltage of the entire SoC can be reduced down to the level that no component becomes unreliable. A PUF design with insufficient reliability at low voltages would enforce limited voltage scaling to the SoC, leading to significant energy inefficiency for the entire system. Therefore, a PUF design with higher reliability is crucial for low-power and energy- constrained application domains.

Nowadays, SRAM-based PUFs are prevalent designs used for the key generation in the commercial products, as they offer a mature and viable security component [7], [8]. The difference between the strength of the transistors in an SRAM cell due to process variation together with the feedback loops in the

SRAM cell could strongly bias the power-up value of the cell towards either "zero" or "one". Therefore, the impact of the process variation on the SRAM cells is leveraged to generate a unique identifier by powering up. Various techniques and PUF designs have been proposed to improve the reliability of the PUF response at nominal supply voltage [9], [10], [11]. However, the reliability of the PUF response at low supply voltage is mostly overlooked. The reliability of the SRAM-based PUF designs is also very sensitive to the variations and noise (both internal and environmental), especially in the NTC domain. As the amount of voltage scaling, and consequently the energy efficiency, is directly impacted by the circuit ability to operate reliably at such low supply voltages, a low-power PUF design that can reliably operate in the NTC domain without imposing large "key generator" overhead is highly beneficial in improving the overall circuit energy efficiency, resiliency, and security.

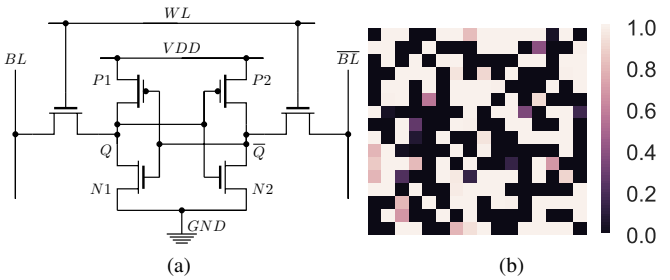
In this paper, we propose a novel memory-based PUF design that is suitable for low-power applications with low-power footprints and good reliability. We first develop a simulation flow to analyze the effectiveness of the SRAM PUFs over a wide range of supply voltages from the nominal (super-threshold) voltage all the way down to the near-threshold voltage region. In the proposed flow, we evaluate the impacts of different factors such as supply voltage, temperature, and sizing of SRAM on the characteristic of the PUFs. Moreover, we compare our proposed PUF design to other SRAM based PUF designs. Simulation results ascertain that the sensitivity of the power-up state of the PUF to noise is significantly smaller in the proposed PUF design, and hence, the reliability of the proposed PUF design is up to $2.6\times$ better than the conventional SRAM PUFs in the NTC domain.

The rest of the paper is organized as follows. Section II presents brief preliminaries regarding memory-based PUFs as well as PUF performance metrics. The analysis methodology using a conventional SRAM-based PUF are presented in Section III. In Section IV, the proposed PUF design is presented. Finally, Section V concludes the paper.

II. Preliminaries

A. SRAM PUF

The most common SRAM cell design is a 6-transistor (6T) design consisting of two back-to-back connected inverters and two access transistors, as shown in Figure 1a. When an SRAM cell is powered up, it gets a value of either "one" or "zero" depending on the threshold voltage of the transistors and noise. We can define P_0 as the probability of settling at "zero" after power-up for an SRAM cell, and $P_0 = 1 - P_1$ as the probability of settling at "zero". These values (P_0, P_1) explain the *skewness* of the SRAM cell. When the SRAM cell is not skewed (i.e. $P_1 \approx P_0 \approx 0.5$), the power-up value is solely depending on the noise. However, in the presence of process variation, some of the transistors might become stronger and the SRAM is skewed. Depending on the strength of the SRAM transistors, P_0 could be higher or lower than P_1. Figure 1b shows P_1 for a 256-bit SRAM array. Each small tile in the figure corresponds to P_1 of one of the cells in the SRAM array. As shown in this figure, most of the cells are skewed (dark or white), which means their power-up values are determined by process variation. Due to random process variation, this pattern



(a) (b)

Fig. 1: a) Schematic of a 6T SRAM, b) Probability of power-up value being "one" (P_1) for a 256-bit SRAM-based PUF

is unique among different chips and could be used as a device fingerprint. Therefore, we can assume the power-up value of an SRAM array as a *PUF response*. However, as the power-up values of the non-skewed cells are mostly dependent on the noise, there might be a slight variance between different PUF responses. In other words, the presence of the non-skewed cells in the SRAM-based PUF impacts the PUF reliability.

B. Related Work

It has been shown that SRAM PUFs are more resilient to temperature variation and are more area efficient compared to the other memory-based PUFs [12]. Here, we review the most related work on SRAM-based PUF designs as our proposed PUF design lies within the same area.

Several low-power PUF design techniques have been proposed in the literature to overcome the reliability degradation of memory-based PUFs. In [13], the authors proposed to use SRAM PUFs at ultra-low voltages but the detailed analysis of the reliability and uniqueness of PUF at low voltages in the presence of temperature variation was missing. Authors in [9] have proposed a PUF design which improves the reliability by a two stage identification method. Their design is very compact, however, a large amount of the power (more than 50% of total power) is consumed in the bias circuitry during the first stage, which resulted in an inefficient design. A number of techniques have been proposed in [10] for improving the reliability of the PUF designs by reducing the impact of noise, however, the analysis of the proposed designs at low supply voltage range is missing. In [11], the reliability of the proposed 8T-SRAM based PUF is studied at low voltage, however, the sneak leakage path in the design leads to energy inefficiency. Hence, a reliable low-power PUF design which can operate in NTC domain in order to attain high energy efficiency of NTC is still missing.

C. PUF Evaluation Metrics

Several metrics have been proposed to explain the quality of a PUF [3], [14]. Here, we base our analysis on the most important metrics which are *uniqueness*, *uniformity*, *bit-aliasing*, and *reliability*. The first three metrics explain the quality of the PUF in terms of having a non-skewed distribution of values over the responses of different PUFs, over the bits within a PUF response, and over a specific bit position of different PUF responses, respectively. The reliability metric describes the stability of the response of a specific PUF at different readouts and over different operating conditions. These metrics are calculated based on the *Fractional Hamming Distance* (FHD) of the PUF responses. The FHD explains the fraction

of the bits which differs from one bit array (A) to another (B):

$$FHD(A,B) = \frac{1}{N}\sum_{i=1}^{N}|A_i - B_i|. \quad (1)$$

Here, we can assume that A and B are N-bit PUF responses. Therefore, the aforementioned metrics of SRAM-based PUF at different supply voltages and different temperatures are obtained according to the following definitions [14]:

1) Uniqueness: The PUF response should be unique which means that the responses of two different PUFs (for two different devices) should be different. The FHD distance of the responses of different PUFs can be used to show the uniqueness of PUFs, and this value should be as large as possible. Therefore, for a set of PUFs, we obtain the FHD between every two PUF responses (A, B) and the average value of the FHDs should be ideally equal to 50%.

2) Uniformity: Uniformity explains how well the bits of a PUF response are distributed between "zero" and "one" values. For each PUF, it is calculated as the average of all the bit values in the responses of the PUF. A uniformity close to 50% is desired as we want the PUF responses not to be skewed towards "zero" or "one".

3) Bit-aliasing: It is defined as the distribution of the values at a specific bit-position over different PUF responses. A bit-aliasing close to 50% for a bit-position ensures that the bit-position has a fair distribution of "zero" and "one" over different PUF responses, thus not being skewed toward any value. The bit-aliasing of bit-position k is calculated as the average of all the k^{th} bit of all PUF responses.

4) Reliability: The response of a particular PUF has to ideally remain the same at different readouts and across various environmental conditions, such as temperature variation and power supply noise. Here, the reliability of a PUF is defined as the average FHD between the different responses of the same PUF subjected to different environmental and noise conditions. Please note that, a lower reliability value represents a more stable PUF, as the change in the response has to be as small as possible (close to zero). In this paper, the first PUF response obtained at room temperature ($T = 25°C$) is considered as the reference response for obtaining the FHD values.

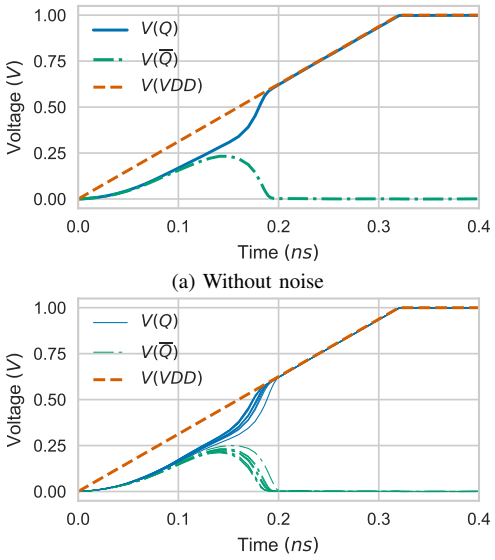
III. Memory-based PUF Analysis Methodology and Application to 6T-SRAM PUF

In this section, we first present the methodology to evaluate memory-based PUFs. Then, we apply the methodology to study 6T SRAM-based PUFs. For this purpose, we perform accurate SPICE simulations and the metrics presented in the previous section are calculated based on the simulation results.

A. Analysis setup

We assume that there are twelve SRAM-based PUFs each having 256 bits. The power-up values of the SRAM arrays are obtained 100 times, each time considering the noise using accurate SPICE simulation. The simulations are performed for a wide range of supply voltages from the NTV region to the nominal supply voltage and for different temperatures:

$$Vdd \in \{0.45, 0.5, 0.6, 0.8, 1.0\}\ V, \qquad T \in \{-25, 25, 75\}\ °C.$$



(a) Without noise

(b) With noise Monte-Carlo simulation (10 samples)

Fig. 2: Transient simulation of a 6T SRAM cell to obtain power-up values

The SPICE simulation are done with 32nm SAED technology library. Then, the performance of the PUF designs are compared at different supply voltages and temperatures by calculating the PUF metrics. Please note that, we intentionally consider a higher number for the readouts (100 times) as we are going to accurately compare the reliability of the PUF designs. In summary, $\sim$ 500 million SPICE simulations have to be executed for the 6T SRAM-based PUF as well as other PUF designs discussed in Section IV (a Schmitt-Trigger based PUF and our proposed PUF).

B. SRAM simulation flow

In our simulation flow, the impacts of various process variation sources such as L, W, V_{th}, and T_{ox} are considered as a lumped shift in the threshold voltage of the internal transistors for each SRAM cell separately [15]. Then, we simulate the power-up of the cells with HSPICE to extract the power-up values, i.e. PUF responses. For this purpose, the word-line node (WL) is connected to ground and the nodes Q and $\overline{Q}$ are initialized with a "zero" value. The supply voltage is ramped-up (i.e. the SRAM cell is turned on) to obtain the value of Q and $\overline{Q}$ after a given transient time, as shown in Figure 2a.

It is crucial to consider the internal noise of the circuit elements (thermal noise, flicker noise, and shot noise) during the simulations [16], [17], otherwise all the responses of a PUF under the same environmental condition would be the same. For this purpose, considering a correct and accurate noise model during the simulation is very important, as it directly impacts the measured PUF responses. An accurate noise model would minimize the gap between the simulation results and the real measurement results for a single cell. The noise model should have a reasonable spectrum and should affect the circuit during the entire simulation. In some of the related work, the noise is considered as a fixed source during the entire simulation or at the beginning of the simulation which does not reflect the real properties of the noise. To achieve accurate noise properties, we employ the state-of-the-art noise model

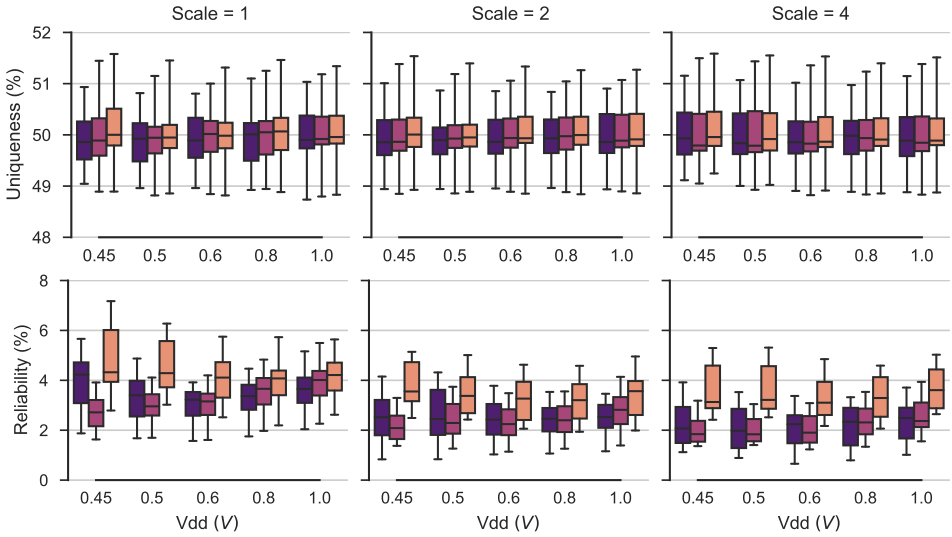


Fig. 3: Uniqueness of 6T SRAM PUF which is not sensitive to the temperature, the applied supply voltage, and the size of cell.

Fig. 4: Reliability of 6T SRAM PUF is significantly affected at low supply voltages.

implemented in the HSPICE circuit simulator. This way, a valid noise model is considered for all circuit elements. With a tuned noise configuration, the simulation results for the reliability of a 256-bit SRAM-based PUF conform with the experimental SRAM-based PUF results presented in [7], [8], [18], i.e. the reliability metric is $\sim 5\%$ at nominal voltage. The results of ten SPICE simulations considering noise for the same SRAM cell are presented in Figure 2b. Here, all the simulations led to the same response, which is $V(A) = 1$. Moreover, the impact of transistor sizes on the characteristics of the SRAM PUF is studied by considering three different SRAM sizes: scale=1 in which the transistor sizes are adopted from [19], scale=2 and scale=4 with two and four times larger transistors, respectively.

C. Simulation Results

1) Uniqueness, Uniformity, and Bit-Aliasing: Figure 3 shows the uniqueness of SRAM PUF for different supply voltages, temperatures, and sizes. As shown in the figure, the uniqueness of SRAM PUF has almost negligible sensitivity to the temperature, the applied supply voltage, and the size of SRAM cell and the average uniqueness is close to 50%. This is because of the inherent symmetry in the SRAM cell. For the same reason, uniformity and bit-aliasing also have similar results, which is omitted for brevity. From this experiment, we can conclude that the voltage scaling and temperature variations do not impact these metrics for the SRAM PUF.

2) Reliability: Figure 4 depicts the reliability of the SRAM PUF for different supply voltages, temperatures, and sizes. The impact of different factors can be summarized as follows:

- **Supply voltage**: Reducing the supply voltage increases the circuit sensitivity to process variation, however, the feedback strength of the back-to-back inverters of the SRAM cell also decreases. These two mechanisms have opposite impacts on the reliability. Hence, by decreasing supply voltage from the super-threshold down to the NTV region, the reliability fluctuates slightly. However, in the NTV region (when $Vdd \rightarrow 0.45V$) the reliability is worse due to the rapid decrease in the strength of the transistors.
- **Temperature**: The reliability at high and low temperatures are lower compared to the room temperature. Particularly, the reliability decreases significantly at high temperature. This is not only because of the impact of noise but also due to the fact that the reference PUF response is obtained at room temperature ($25°C$). The response of SRAM cells with small skewness might change at low and high temperatures, which results in slightly different PUF responses from time to time.
- **SRAM size**: The reliability values of the larger SRAM PUFs are slightly better than the smallest size SRAM PUF. This could be explained by the lower impact of noise on larger transistors. Additionally, the reliability of the larger SRAM PUFs are less sensitive to temperature variation and voltage scaling, especially at NTV.

3) Power and Area: Table I summarizes the leakage power of SRAM PUF for different sizes and supply voltages. According to the results, the leakage power decreases significantly by reducing the supply voltage. Although, the larger SRAM PUFs are slightly better in terms of reliability, due to the excessive area and power overhead of these PUFs, the smaller size SRAM PUF is preferred. However, as shown in Section III-C2, smaller SRAMs need to be improved for better reliability.

IV. Reliable Memory-based PUF for NTV operation

A. Proposed SRAM PUF

According to the results presented in the previous section, reducing the supply voltage of 6T SRAM PUF to the NTV region deteriorates the reliability significantly. The impact of decreasing supply voltage on the characteristic of the SRAM PUF is two-fold:

- The sensitivity to variation increases which leads to a reliability improvement.
- The feedback of the back-to-back connected inverters becomes weaker. This leads to a higher sensitivity to noise, hence, the reliability worsens.

Based on this analysis, we propose a new reliable memory-based PUF design, which is presented in Figure 5. Our proposed PUF cell has 10 transistors and consists of two different parts; I) *pre-charge circuit* consisting of $P3$, $P4$, and $P5$ which charges the SRAM cell nodes in a pre-charge phase and is very sensitive to process variation, and II) *back-to-back*

TABLE I: Leakage power of SRAM PUF at different supply voltages. The results are normalized to the case with nominal $Vdd = 1.0V$ and $scale = 1.0$

	Vdd (V)				
	1.0	0.8	0.6	0.5	0.45
Scale = 1	1.000	0.326	0.116	0.078	0.068
Scale = 2	1.672	0.500	0.135	0.068	0.048
Scale = 4	3.144	0.943	0.256	0.129	0.090

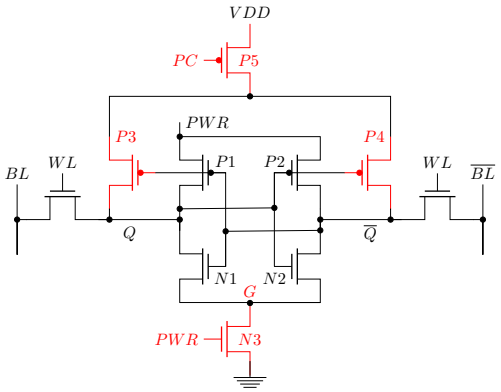


Fig. 5: Schematic of the proposed memory-based PUF

inverter structure which amplifies the voltage difference of the internal nodes, obtained in the pre-charge phase.

The pre-charge phase starts when the voltage of node PC is set to zero. The pre-charge time is relatively short. Here, we assumed this time as twice the delay of a minimum size inverter. After the completion of the pre-charge phase, the back-to-back inverter is activated by ramping the voltage of node PWR, which is done in the same way as discussed in Section III-B. This signaling scheme is shown in Figure 6a.

1) Pre-charge circuit: Using this circuit, nodes Q and $\overline{Q}$ of the cell are pre-charged to a value between $0V$ and Vdd. The final values of the nodes are strongly dependent on the strength

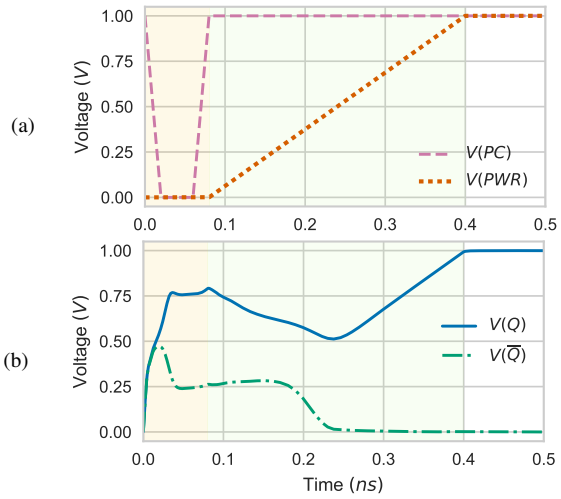


Fig. 6: (a) Pre-charge (PC) and Power-up (PWR) signals for the proposed memory-based PUF, (b) Transient behaviour of the proposed memory-based PUF

of PMOS transistors $P3$ and $P4$, and hence their process variations as shown in Figure 6b. Transistors $P3$ and $P4$ are minimum size transistors for achieving higher sensitivity to process variation according to Pelgrom's model [15]. Additionally, during the pre-charge phase, NMOS transistors $N1$ and $N2$ are disconnected from the ground by control transistor $N3$. At the beginning of the pre-charge phase, transistors $P3$ and $P4$ are turned on because of the low initial voltage at nodes Q and $\overline{Q}$, and they start charging these nodes. Increased voltage level at Q and $\overline{Q}$ nodes increases the gate-source voltage of transistors $N1$ and $N2$. Therefore, the transistors start charging node G slowly. However, as $N3$ is turned off, the voltage at node G also increases, which keeps $N1$ and $N2$ in weak inversion. This eventually increases the sensitivity of $N1$ and $N2$ to process variation leading to a clear signal separation by the end of the pre-charge phase, as shown in Figure 6b.

2) Back-to-back inverter structure: Once activated by increasing $V(PWR)$, the back-to-back inverter structure amplifies the voltage difference between Q and $\overline{Q}$ nodes obtained during the pre-charge phase. The transistors of the back-to-back inverter structure are larger (4x) than those of the pre-charge circuit for two reasons:

- By using larger size transistors, the strength of the back-to-back inverter structure increases, and therefore, the sensitivity to the voltage difference between Q and $\overline{Q}$ is amplified, and the sensitivity to the noise is also reduced leading to a higher reliability of PUF.
- Larger transistors are less impacted by process variation (according to Pelgrom's model [15]), and hence, the back-to-back inverter structure cannot suppress the voltage difference obtained from the pre-charge phase.

Please note that, as we only read from the proposed memory-based PUFs, the size of the rest of the transistors should be set to maximize the Read Noise Margin of the cell. A PUF array based on the proposed design can also be placed next to the conventional SRAM array to be read with the same memory controller by accessing the corresponding PUF address. For this, the existing memory controller design can be used without any modification. However, before reading from the PUF address, it is required to initialize the PUF cells by activating the PC and PWR signals, as explained in Figure 6. The design of the initializer circuit is simple as it generates and drives only two signals. Therefore, the area overhead of such a circuit is negligible compared to the PUF array.

B. Simulation results and discussion

In this section, PUF metrics such as reliability, uniqueness, uniformity, and bit-aliasing, as well as power consumption of the proposed memory-based PUF are studied and compared to other designs. We use an optimized 6T SRAM-based PUF design as the *baseline* for comparison. Since our proposed cell has 10 transistors, its characteristics are also compared to an optimized SRAM cell with 10 transistors, for a fair comparison. Therefore, we compare the characteristics of the proposed SRAM with the Schmitt Trigger based sub-threshold SRAM (10TST PUF) proposed in [20] which has better characteristics compared to the baseline. The PUF metrics and the power consumption of these designs are obtained using the proposed simulation flow in Section III at different supply

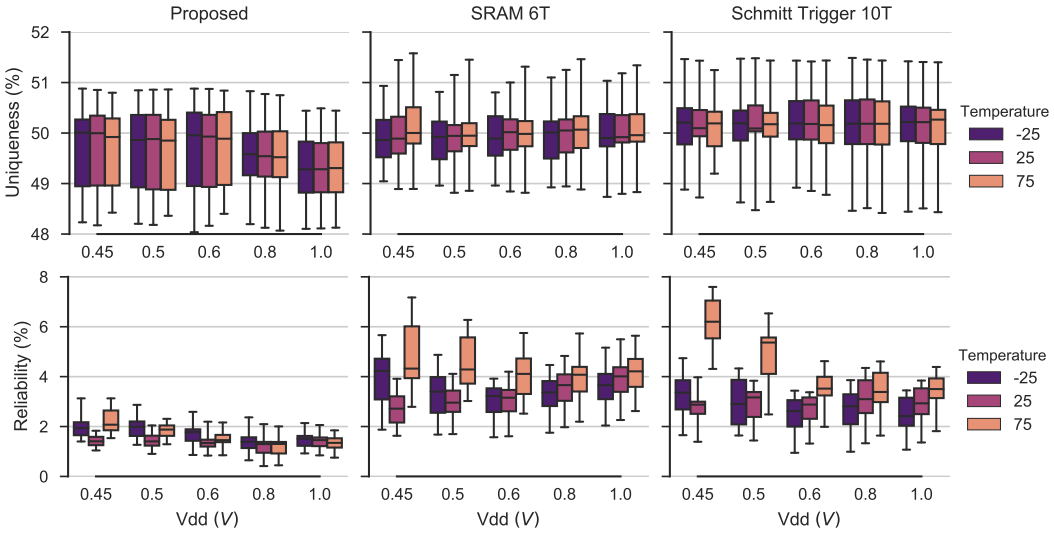


Fig. 7: Uniqueness of the proposed PUF compared to 6T SRAM PUF and Schmitt Trigger-based PUF [20]

Fig. 8: Reliability of the proposed PUF is significantly better than the reliability of 6T SRAM PUF and Schmitt Trigger-based PUF [20]

voltages, from the nominal supply voltage down to the NTV, and temperatures.

1) Uniqueness, uniformity, bit-aliasing: As can be seen in Figure 7, the average uniqueness of the proposed memory-based PUF as well as that of 10TST PUF is close to 50% and has negligible sensitivity to supply voltage and temperature similar to the uniqueness of the conventional 6T SRAM PUF. Table II compares the average values of the uniqueness, uniformity, and bit-aliasing metrics for the PUF designs. The results of uniformity and bit-aliasing are similar to the uniqueness results and the corresponding figures are eliminated for brevity.

2) Reliability: The reliability of the proposed memory-based PUF, as well as the SRAM-based PUFs, are presented in Figure 8. As can be seen, the reliability of the 10TST PUF is in the range of the conventional 6T SRAM PUF, however, our proposed memory-based PUF can improve the reliability by up to 2.6× compared to the other PUF designs. This is due to the fact that the power-up state of the cell is determined by the pre-charge phase, and hence, the noise impact on the back-to-back inverter is negligible. As shown in the figure, the reliability of the proposed memory-based PUF is reduced at low and high temperatures compared to the room temperature. Moreover, the sensitivity to temperature variation increases at lower supply voltages. The reason for the lower reliability at these temperatures is that the reference PUF response is measured at room temperature ($T = 25°C$).

Table II compares the reliability metric for the proposed PUF design and other SRAM-based designs. The reliability results presented in this table are the worst-case values for all operating voltages and temperatures.

3) Power and Area: A PUF design for ultra-low power applications should not have high current demand because the power sources that are typically used in ultra-low power applications, such as energy harvesters, are unable to provide high peak power.

The leakage power of proposed memory-based PUF and other SRAM-based PUFs are reported in Table III for different supply voltages. The reported numbers are normalized to the power of conventional 6T SRAM PUF at nominal supply voltage and with minimum size transistors. As reported, the

TABLE II: Comparison of the PUF metrics for different PUF designs. The presented reliability is the worst case over all temperatures. For uniqueness, uniformity, and bit-aliasing the average values over all operating conditions are reported.

	Vdd (V)	6T SRAM	10T Schmitt Trigger [20]	Proposed	Improvement over 6T SRAM
Reliability (%) (worst-case)	1.0	5.64%	4.39%	2.14%	62% (2.6×)
	0.8	5.73%	4.61%	2.37%	59% (2.4×)
	0.6	5.75%	4.62%	2.59%	55% (2.2×)
	0.5	6.27%	6.53%	2.87%	54% (2.2×)
	0.45	7.17%	7.60%	3.13%	56% (2.3×)
Uniqueness (%)	all	50.0%	50.1%	49.6%	–
Uniformity (%)	all	50.9%	50.5%	49.7%	–
Bit-aliasing (%)	all	50.9%	50.5%	49.7%	–

power consumption of the proposed PUF is almost half of traditional SRAM PUF at low voltages. As the PUF module could be used regularly during the IoT device operation for cryptography or identification purposes, a lower power PUF design is beneficial. This makes the proposed PUF design a suitable candidate for reliable low-power PUF for the NTV operation. The proposed 10T memory-based PUF occupies 49% more area compared to the conventional 6T SRAM PUF, while the 10TST PUF is 31% larger than the 6T SRAM PUF.

At the system level, the area overhead of a PUF array is typically negligible as the key extractor circuit, which is mandatory for all possible weak PUF designs, is typically

TABLE III: Leakage power of 10TST PUF [20] and proposed PUF desighn at different supply voltages. The results are normalized to the leakage power of 6T SRAM PUF with nominal $Vdd = 1.0V$ and Scale=1.

	Normalized leakage power vs. Vdd (V)				
	$1.0V$	$0.8V$	$0.6V$	$0.5V$	$0.45V$
6T SRAM (Scale = 1)	1.000	0.326	0.116	0.078	0.068
Schmitt Trigger SRAM [20]	0.494	0.147	0.049	0.032	0.026
Proposed	0.682	0.207	0.068	0.041	0.033
Improvement over 6T SRAM	32%	37%	41%	47%	51%

much larger than the PUF array. In fact, the power and area of a PUF circuit (including the key extractor) is totally dependent on the reliability of the PUF response at low supply voltage: the higher the reliability, the lower the overall area and power consumption. To have an estimate of the overheads due to key extractor, we followed the simple code approach presented in [21] and extracted the suitable codes which can attain a desired error probability of 10^{-6}, according to the bit error probability values extracted from our simulations. For the 10TST PUF design, a $[255, 37, 91]$-BCH code, which is a 255-bit code with 37 information bits and the ability to correct 45=$\lceil \frac{91-1}{2} \rceil$ errors, is required. However for the proposed PUF design, a much smaller $[127, 29, 43]$-BCH is sufficient. We synthesized the encoder and decoder for these BCH codes using synopsys design compiler with SAED 32nm cell library. The total cell area and leakage power of the $[255, 37, 91]$-BCH encoder/decoder are approximately $2.4\times$ larger than the area and power of the $[127, 29, 43]$-BCH encoder/decoder. Note that the BCH decoder for 10TST PUF needs to correct more errors compared to the decoder required for the proposed PUF, and therefore, it imposes more area and power overhead. Additionally, the area of the $[127, 29, 43]$-BCH code decoder circuit is more than $80\times$ larger than the entire PUF array area. Therefore despite consuming more power compared to the 10TST PUF, the proposed PUF design has significantly lower area and power at the system level (when the key extractor is also considered), thanks to its better reliability.

4) Reliability under aging: Transistors are impacted by a variety of aging mechanism during their operating lifetime. Bias Temperature Instability (BTI) and Hot Carrier Injection (HCI) are among the most important aging mechanism which reduce the driving capability of a transistor by increasing the threshold voltage of the transistor. BTI impact is highly dependent on the overdrive voltage of the transistor whereas HCI is dependent on the current density. Both effects are highly dependent on temperature. In an ultra-low voltage system, the impact of both BTI and HCI is negligible, as the overdrive voltage is much smaller and the current density is orders of magnitude smaller compared to those of the nominal supply voltage (please refer to [22]). The temperature is mostly determined by the environment as the power dissipation of circuits is reduced by orders of magnitude. Therefore, unlike the PUFs operating at nominal voltage, the overall impact of aging on the proposed PUF operating at NTC is negligible over the circuit operating lifetime.

V. Conclusion

The power-up pattern of a SRAM memory array can be used as a unique identifier of the chips for device authentication and identification. The emerging IoT application domains require security and authentication needs on the one hand, and low energy constraints on the other hand. Therefore, it is important to have PUFs operating reliably at low voltages. In this paper, we performed detailed simulation-based analysis of SRAM PUFs in low voltage ranges. Based on our observations, we have proposed a new memory-based PUF design which operates very reliably for a wide voltage range. Such reliable operation at low-supply voltages allows aggressive voltage scaling for the entire circuit, which leads to large energy efficiency when the circuit is operated at NTC domain. Our simulation results show that our proposed PUF design is $2.6\times$ more reliable than conventional SRAM PUF while consuming almost half the leakage power, making it a promising candidate for energy-constrained IoT applications.

References

[1] U. Rührmair *et al.*, "Modeling attacks on physical unclonable functions," in *ACM conf. Computer and Communications Security (CCS)*, 2010, pp. 237–249.

[2] R. Pappu *et al.*, "Physical one-way functions," *Science*, vol. 297, no. 5589, pp. 2026–2030, 2002.

[3] A. Maiti, V. Gunreddy, and P. Schaumont, "A systematic method to evaluate and compare the performance of physical unclonable functions," in *Embedded systems design with FPGAs*. Springer, 2013, pp. 245–267.

[4] R. Dreslinski *et al.*, "Near-Threshold Computing: Reclaiming Moore's Law Through Energy Efficient Integrated Circuits," *Proceedings of the IEEE*, vol. 98, no. 2, pp. 253–266, Feb 2010.

[5] H. Jayakumar *et al.*, "Powering the Internet of Things," in *ISLPED*, 2014, pp. 375–380.

[6] Y. Dodis, L. Reyzin, and A. Smith, *Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 523–540.

[7] R. Maes *et al.*, "Experimental evaluation of Physically Unclonable Functions in 65 nm CMOS," in *ESSCIRC*, 2012, pp. 486–489.

[8] G.-J. Schrijen and V. van der Leest, "Comparative analysis of SRAM memories used as PUF primitives," in *DATE*, 2012, pp. 1319–1324.

[9] C. Böhm *et al.*, "A reliable low-area low-power puf-based key generator," in *TRUEDEVICE*, 2016, pp. 1319–1324.

[10] V. C. Patil *et al.*, "Improving reliability of weak pufs via circuit techniques to enhance mismatch," in *HOST*, 2017, pp. 146–150.

[11] J.-W. Jang and S. Ghosh, "Design and analysis of novel sram pufs with embedded latch for robustness," in *ISQED*, 2015, pp. 298–302.

[12] S. Katzenbeisser *et al.*, "PUFs: Myth, fact or busted? A security evaluation of physically unclonable functions (PUFs) cast in silicon," *CHES*, pp. 283–301, 2012.

[13] M. Kassem *et al.*, "A sub-threshold sram based puf," in *Int. Conf. Energy Aware Computing*, 2010, pp. 1–4.

[14] M. Claes, V. van der Leest, and A. Braeken, "Comparison of SRAM and FF PUF in 65nm technology," in *NordSec*, 2011, pp. 47–64.

[15] M. J. Pelgrom *et al.*, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, 1989.

[16] J. Chang, A. Abidi, and C. Viswanathan, "Flicker noise in CMOS transistors from subthreshold to strong inversion at various temperatures," *IEEE Trans. Electron Devices*, vol. 41, no. 11, pp. 1965–1971, 1994.

[17] D. P. Triantis, A. N. Birbas, and D. Kondis, "Thermal noise modeling for short-channel MOSFETs," *IEEE Trans. Electron Devices*, vol. 43, no. 11, pp. 1950–1955, 1996.

[18] V. van der Leest *et al.*, "Efficient implementation of true random number generator based on sram pufs," in *Cryptography and Security*. Springer, 2012, pp. 300–318.

[19] Y. Morita *et al.*, "An area-conscious low-voltage-oriented 8T-SRAM design under DVS environment," in *VLSI Circuits*, 2007, pp. 256–257.

[20] J. P. Kulkarni, K. Kim, and K. Roy, "A 160 mV, fully differential, robust schmitt trigger based sub-threshold SRAM," in *ISLPED*, 2007, pp. 171–176.

[21] C. Bösch *et al.*, "Efficient helper data key extractor on fpgas," *Cryptographic Hardware and Embedded Systems–CHES 2008*, pp. 181–197, 2008.

[22] W. Wang *et al.*, "Compact modeling and simulation of circuit reliability for 65-nm CMOS technology," *IEEE Transactions on Device and Materials Reliability*, vol. 7, no. 4, pp. 509–517, 2007.