A Highly Reliable SRAM PUF with a Capacitive Preselection Mechanism and pre-ECC BER of 7.4E-10

Avi Miller*, Yizhak Shifman*, Yoav Weizman, Osnat Keren and Joseph Shor
Dept. of Electrical Engineering, Bar-Ilan University
Ramat Gan, Israel
y.shifman@gmail.com

Abstract— An SRAM PUF with an internal error reduction mechanism is presented. A capacitive preselection test identifies potentially unstable cells with insufficient mismatch. The test can be accomplished in one VDD / temperature corner. An implementation in TSMC 65nm technology disqualified all the unstable cells (19.7%) in 14 800-bit arrays. The test has no impact on the randomness of the PUF and negligible impact on area. A Highly competitive pre-ECC BER of 7.4E-10 and an energy consumption of 16fJ/bit were achieved.

Keywords—PUFs; Hardware security; style; preselection;

I. INTRODUCTION

The growing demand for secure cryptographic keys generation and robust, reliable identification has boosted the interest in Physically Unclonable Functions (PUFs). PUF is a circuit that exploits the physical variations between the manufactured devices to produce a digital fingerprint of a semiconductor chip. When a PUF cell has an insufficient amount of internal variation, the resulting code may be dependent on noise and ambient effects, such as Voltage and Temperature (VT), which can cause the cell to be unstable between readings. Several different types of mechanisms exist to form a PUF [1]. However, the most accepted PUF in the industry for encryption and key storage (i.e. the weak PUF) is the SRAM or metastability PUF [2], since its mechanism is predictable, well understood, and scales well. Some of the recently reported mechanisms, such as leakage and oxide breakdown can yield a very compact cell [3][4]. However, they can be highly process dependent and may not scale well between technologies, across process corners and over the process lifetime.

Typically, up to 30% of the native SRAM PUF bit-cells may be unstable [2], but a bit-error rate (BER) as low as 10E−15 for a 128bit PUF response is often required by an end-user device. A few error reduction mechanisms have previously been proposed:

1) Direct preselection: the PUF is sampled multiple times across multiple operation conditions and unstable cells are disqualified and excluded [2].

2) Indirect preselection: A test is run on the PUF cells, such that a cell that passes the test is qualified [1,3,4,5].

3) Hardening: a burn-in procedure strengthens the direction of the response of the cells, making them less unstable [2].

4) Error Correction Codes (ECC): A code is run on groups of bits, resulting in smaller groups with smaller error probability. One of the simplest ECC is Temporal Majority Voting (TMV), in which each bit cell is sampled 15-20 times [2] and the response of the bit is decided as the majority of the samples.

Usually, an ECC such as BCH [1] is used as a final error reduction stage which guaranties the required error probability. However, BCH alone can be very inefficient for high native BER (~10%) such that up to 30-80 physical bits may be required to produce one stable bit [6]. Thus, other approaches are typically implemented on the native PUF bits prior to BCH. For example, [2] applied direct preselection, hardening and TMV before applying BCH. BCH starts to become area efficient when the pre-BCH BER is in the range of ~ 0.1%.

Some disadvantages should be pointed out in prior art implementations of the outlined error reduction mechanisms. Direct preselection by testing in multiple VDD and temperature corners has been demonstrated in [2], but this entails additional test costs. Some prior art SRAM indirect preselection mechanisms utilized precision differential analog voltages to identify unstable bits [5,6]. These voltages are usually not available to customers in the field and could not be used during enrollment unless an accurate voltage regulator is integrated into the PUF array, which adds cost, complexity, and power. Additionally, indirect preselection may disqualify some of the stable cells. Hardening requires heating the PUF to a high temperature, and thus increases the test costs. TMV includes a high redundancy itself (15-20X as in [2]) which can increase energy, timing and/or area costs, and BCH may require high redundancy, as well as additional power and time in each PUF reading.

This work introduces a novel, all digital, SRAM-based PUF circuit with an internal preselection test to be applied prior to BCH. The test circuitry is fully integrated, doesn’t require any external analog components and may be run at a single VDD and temperature, saving test costs. The PUF was manufactured in TSMC 65nm process and the preselection mechanism has disqualified all the unstable cells in 14 800-bit arrays across 14 dies. This allows an efficient implementation of BCH, such that the overall area efficiency of both these two error

*Equally-Credited Authors
reduction mechanisms is very high. The preselection test doesn't damage the random qualities of the PUF and has minor impact on its area utilization and its power consumption.

II. CIRCUIT AND TEST ARCHITECTURE

The PUF cell (Fig. 1) is a bistable circuit composed of two back-to-back inverters, similar in structure to [2,5]. During normal operation, the switches s1-s2 are non-conducting and the cell converges to a state dictated by the mismatch and the ambient conditions. The preselection test essentially measures the internal mismatch of the PUF cell. If it is not high enough, the cell is considered "potentially unstable" and disqualified. The test procedure is described in Fig. 2. During the test, a skew, or "tilt", is introduced to the PUF cell and biases it towards either a “0” or “1”, depending on the direction of the tilt. If the internal mismatch of the cell is higher than the introduced tilt, the cell will not change its state between the two opposite tilts and can be considered stable and thus qualified for use. However, if the opposite tilts yield opposite responses, the cell is disqualified.

The tilting is realized through an addition of capacitance to either of the internal latch nodes, H and Hb (Figs. 1 and 3). The node with the higher capacitance is tilted towards ‘1’. The capacitors are MOS capacitors, controlled by biasing the source and drain to VSS for inversion, a larger capacitance, or to VDD for depletion, a much smaller capacitance. The capacitors are binary-weighted, to enable control of the tilt magnitude. To minimize area, a single capacitor bank serves four PUF cells (Fig. 4), and the tilt is done serially, each time on one-fourth of the array. The use of gate capacitance in this manner is more area efficient than MOM (Metal-Oxide-Metal) or MIM (Metal-Insulator-Metal) capacitors. The binary control of capacitor array via the source/drain of the MOS capacitors enables the inverters (Fig. 3) to be shared among 4 capacitor banks, which avoids the need for more switches and further minimizes the area.

The tilt preselection is done once in the lifetime of the device. To minimize noise effects during tilting, the test can be repeated multiple times. Since this is a one-time test, it has no impact on in-field power consumption. The resulting mask may either be generated during IC testing or by the end-user during enrollment.

III. IMPLEMENTATION AND MEASUREMENT RESULTS

The PUF was designed as an array of 800 bit cells in TSMC 65nm (Fig. 5). The size of the PUF cell is 10µm² while the two capacitors banks are 5.37 µm² each, and the capacitors are shared between four PUF cells. The devices of the back-to-back inverters were sized to the minimum design rule (Fig. 1) and laid out in an isolated manner, without diffusion sharing, in
order to increase the mismatch between them. Each capacitance step is about 250aF (simulated). The average energy per bit is 16fJ (simulated). Fourteen dies were measured, a total of 11.2K PUF cells. To check the stability of the cells, each cell was sampled a total of 120K times across the operation range, 1.0-1.2V and -10 – 85°C (6 volt-temp conditions, 20K times each). 2206 (19.7%) of the cells didn’t converge to the same response for all the measurements and were considered unstable. The tilt test was run 100 times, and a cell that exhibited the same result for all the runs was qualified. The post-test BER is defined as in [5, 6],

\[ BER = \frac{1}{ \text{(# of qualified bits X # of runs)}} \]  

(1)

assuming, pessimistically, that the next run will be erroneous. Fig. 6(a) presents how the test can recognize unstable cells, by showing the measured impact of the tilt on the responses of 100 of the unstable cells. Each of the unstable cells can be “tilted” either to a “0” or “1” state depending on the magnitude of the tilt. Fig. 6(b) presents the measured tilt impact on one PUF array, stable and unstable cells alike. It is observed that ~40% of the cells have high mismatch and do not change their state even for the full range of the tilting. Fig. 7 presents the results of the test on all of the measured PUF cells. For a tilt of 4.75fF (19 LSBs), all the unstable cells were disqualified. The calculated post-tilt BER is 7.4E-10. As the efficiency of BCH improves for a low BER (~0.1%), a user may select the most efficient trade-off between the BER reduction of the tilt test and that of the BCH.

The uniqueness and randomness of the PUF were measured for the qualified cells. The Auto Correlation Function (ACF), the inter-chip, the intra-chip Hamming Distances (HD) and the percentage of ‘1’s (Hamming Weight) for consecutive 20-bit blocks of qualified cells are presented in Fig. 8. The average

| TABLE I. Performance Summary and Comparison with Prior-Art |
|-----------------|-----------|-----------|-----------|-----------|-----------|-----------|
| process (nm)    | SRAM PUF  | 65        | 22        | 65         | 65         | 65         | 14         |
| Post-TMV energy per bit (fJ) | 16    | 190       | 73        | --         | 163*       | 15*        | 60         |
| Area per bit (F²) | 3001   | 9628      | 11627     | 561        | 12000      | 6000       | 9388       |
| Native Unstable Cells | 19.7%   | 30%       | 27%       | 27%        | 4.51%      | 4.56%      | --         |
| pre-ECC BER after error reduction | <7.4e-10 | 0.97%     | 2.7e-10   | <1e-7      | --         | --         | --         |
| Preselection Method | Digital | Direct    | preselection | Requires precision analog voltages | --         | --         | --         |
| % of ‘1’s in the qualified cells | 51.4% | --        | 49.60%     | 45.0%      | 50.46%     | 50.7%      | --         |
| Inter chip HD   | 50.27%   | 49%       | 49.3%      | 50.17%     | 50.2%      | 50.1%      | 49%        |

*TMV or another error reduction mechanism was not implemented.
HD is 50.27% and the average percentage of 1’s is 51.4%, both are very close to the ideal 50%. The intra-chip HD is 0, indicating that all the qualified PUF cells are stable.

![Graph](image)

**Fig. 7.** Measured percentage of qualified unstable cells and disqualified stable cells Vs. capacitance tilt size on 11.2K PUF cells (14 chips).

**IV. CONCLUSION**

Table I presents a comparison to prior work. This PUF has a very competitive energy consumption of 16fJ even though it is in an older process node than [2] and [9], as it doesn’t require TMV. It has an extremely low pre-ECC BER, together with a competitive area. Low BER’s are also demonstrated in [5, 6], but require precision external analog voltages. If these analog voltages are provided during IC testing, then the mask needs to be transferred to the customer using non-volatile memory, which adds cost and reduces security. The analog voltages could also be provided by an on chip regulator, but this adds power, area and complexity. The present capacitive tilting preselection is fully integrated and can be accomplished in a single VDD-temperature corner, making it highly practical for testing by the end-user during enrollment. This PUF does not rely on process dependent parameters, such as leakage or oxide breakdown, which may be less suitable for high volume manufacturing. It is based on SRAM metastability, which is an established mechanism and can scale well between technologies and over the process lifetime. The techniques shown can be equally applied to a 6T-SRAM cell for further area savings. Yet, we chose to base our implementation on the cell of [2, 5, 9], where each cell contains its own power-gate and the latching nodes are initialized to VDD. This cell is bigger than the 6T SRAM, but enables a highly controlled wakeup, which enhances the cell stability.

**ACKNOWLEDGMENT**

This research was sponsored by the Israel Innovation Authority through the Kamin program.

**REFERENCES**


![Graph](image)

**Fig. 8.** (a) Measured ACF (Autocorrelation Function). (b) HD (Hamming Distance) and HW (Hamming Weight) of 4710 qualified cells (14 chips) shows no spatial correlations, strong uniqueness, uniformity and reliability.

**Table I**

<table>
<thead>
<tr>
<th>Work</th>
<th>Energy Consumption</th>
<th>Process Node</th>
<th>Area</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>[2]</td>
<td>0.19pJ/b</td>
<td>22nm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[9]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This PUF</td>
<td>16fJ</td>
<td>Older</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>