Leveraging Circuit Reliability Effects for Designing Robust and Secure Physical Unclonable Functions

M. Kim, G. Park, P. Chiu, and C. H. Kim
Department of ECE, University of Minnesota, Minneapolis, MN 55455, USA, email: chriskim@umn.edu

***Abstract*—**Reliability mechanisms are undesirable from a product lifetime point of view, but their unique characteristics enable novel applications such as one-time-programmable memory, secure chip odometers, and physical unclonable functions (PUFs). In this invited paper, we will discuss how reliability mechanisms can be leveraged for the aforementioned applications, and then introduce a novel SRAM PUF design where the power up state of the SRAM cell is programmed into a local metal fuse using the electromigration phenomenon for improved stability.

I. Introduction

Product failures due to reliability mechanisms such as bias temperature instability (BTI), hot carrier injection (HCI), time dependent dielectric breakdown (TDDB), and electromigration (EM) can result in billions of dollars of losses for the semiconductor industry. Quality assurance teams around the world invest significant resources to understand and mitigate reliability issues. While reliability mechanisms are generally considered undesirable, the ability to selectively "burn" devices using a high voltage or high current makes them useful for applications such as one-time-programmable (OTP) memories, chip odometers, and physical unclonable functions (PUFs).

OTPs comprise metal fuses or gate dielectric antifuses which can be intentionally ruptured by applying an excessively high voltage or high current to program information [1]. OTPs enable post-silicon memory repair schemes where the location of failure bits must be stored within the memory macro itself. OTP takes advantage of the non-volatile, irreversible, and abrupt nature of TDDB and EM mechanisms. HCI is also irreversible, but unlike TDDB and EM which show an abrupt and clear response, HCI induced shifts are gradual and less pronounced. This makes HCI less attractive for OTPs but the preferred mechanism for tracking the chip's "age" as shown in [2]. Here, a secure chip odometer based on HCI degradation was proposed to measure the wear and tear of the chip. This information can be used to determine the remaining useful life of a chip and thereby preventing illegally recycled parts from entering the supply chain.

Reliability mechanisms exhibiting inherent randomness can be used as the entropy source for PUFs. PUFs exploit the manufacturing variability of a circuit to generate a unique and unpredictable response. Generating secret keys using a PUF offers a higher level of security compared to storing a software generated key in a non-volatile memory [6-7]. While majority of PUFs harness randomness from time-zero variability, recent works have shown that the random nature of TDDB time-to-breakdown can be exploited for PUF designs [3-5]. For instance, in [3], two identical transistors were subject to the same stress voltage, with the output response determined by the location of the first TDDB event; i.e. if device #1 breaks first, then the output is '0', and vice versa. A notable feature of this approach is that the stress voltage seen by the unbroken device is immediately lowered upon the first breakdown event ensuring that only one of the two devices exhibits a TDDB failure. This is achieved by the voltage divider action between the biasing device and the device with the first TDDB event. Another advantage of using TDDB over HCI for PUF applications is that the impact on device current is more pronounced which helps generate a stable and reliable response.

II. Case Study: Hardening PUF Responses Using Electromigration Mechanism

A critical requirement for PUFs is that the generated key should remain constant across a wide range of temperature, voltage, and aging conditions. However, PUFs rely on the intrinsic mismatch between two physically identical devices and therefore the output response can flip in the presence of random temporal noise. Post-processing algorithms such as error correction and de-biasing can be applied to correct unstable responses [8]. However, these techniques require significant hardware resources and necessitate helper data which is vulnerable to various security attacks. In this section, we present a PUF design where metal fuses are employed to harden the response of a SRAM PUF which obviates the need for high overhead error correction techniques.

A. Design Concept and Array Architecture

SRAM PUFs have gained popularity due to their simple design and ease of generating the response bits. An SRAM PUF utilizes the power up state of an SRAM cell as the source of entropy. A new SRAM PUF with internal metal fuses for storing the response bits is shown in Fig. 2. Having separate SRAM and fuse arrays can be risky from a security viewpoint as the data transfer between the two arrays can be intercepted by various physical attack methods. To address this security concern, we chose to embed the fuse inside each individual SRAM cell. A metal fuse was chosen over a gate dielectric antifuse due to its favorable scaling properties [1]. The array is composed of 4 rows and 32 columns with a total of 128 PUF cells. A program current higher than 25mA is required to reliably destroy the metal fuse in 65nm technology [9-10]. The header PMOS was sized to meet this requirement. Four PUF cells share a single PMOS header. Each row is controlled by the PWS signal and WL signal. The IN_HIGH signal is a dummy signal that is held high at all times. This signal is used

to keep the cell layout perfectly symmetric which is important for obtaining an unbiased PUF response.

B. Unit PUF Cell Design

The PUF cell consisting of an SRAM cell and a metal fuse was designed from scratch to ensure that the schematic and layout are perfectly symmetric. Otherwise, the PUF response will be skewed which makes the response more predictable for attackers. Fig. 3 shows the detailed schematic and layout of the hybrid PUF cell. The schematic contains a cross-coupled inverter pair with a power up PMOS which serves as the entropy source. The key novelty of the design is in the read out and programming circuits denoted as (2), (3), (4), and (6). The cell layout is shown in Fig. 3 (bottom). Extra care was taken to ensure the entire layout is perfectly symmetric. Post-layout simulation results shown in Fig. 4 confirm that the cell layout is perfectly symmetric.

C. Program and Read Operation

Fig. 5 shows the detailed program and read operations of the proposed hybrid PUF cell. During program operation, the shared PMOS header is turned on. For a power up state of Q=1 and QB=0, the tri-state inverter on the right-hand side of the SRAM is enabled which resets the DA signal. This activates the NMOS device connected to the fuse, conducting a large program current and breaking the metal fuse. For the opposite power up state of Q=0 and QB=1, the tri-state inverter on the left-hand side is enabled. Since IN_HIGH is held high throughout the program operation, DA is low which leaves the fuse intact. For read operation, both the read out NMOS passgate and the shared PMOS header are enabled. The SRAM cell is powered off and the bitline voltage is predischarged by the weak pull down keeper. If the fuse is blown (i.e. high impedance), the bitline signal remains low. Otherwise, the bitline signal is pulled up to VDD by the current through the pull up header PMOS.

D. Metal Fuse Design

The specific layout of the metal fuse structure may affect their program efficiency. Five different metal fuse shapes shown in Fig. 6 were implemented in the same 128 bit array for comparison purpose. The metal fuse is formed by M2 and M3 metal layers. The program current flows from M2 to M3 as illustrated in the figure. The lengths of the M2 and M3 wires were varied to understand the program time trends for different bottom and top metal lengths. For instance, as shown in Fig. 6, the fuse denoted as (a) has a 2μm M2 lead and a 2μm M3 lead with a 1μm extension region for M3. One of the fuse structures has an extended metal region to verify if a metal reservoir might assist in programming the fuse. Other fuse types have metal lengths ranging from 1μm to 3μm. The width of all metal wires is 0.1μm.

III. MEASUREMENT RESULTS

Fig. 7 shows the die photo and key features of the 128 bit hybrid SRAM + metal fuse PUF chip fabricated in a 1.0V 65nm GP process. First we measured the power up state of the 128 SRAM cells from 10 different chips. The array maps are shown in Fig. 8. The ratio between the number of '1's and the number of '0's is well balanced with no noticeable skew. Next, we sequentially program the metal fuses based on the SRAM power up state obtained in Fig. 8. For detailed analysis, we also measured the program time for the different metal fuse shapes using a program current of 30mA. Results for the 5 different fuse types at 25°C and 90°C are shown in Fig. 9. Our results indicate that the program time is a function of both the top and bottom metal lengths. The type (b) fuse, which has the longest M2 and M3 wires of the five metal fuses, showed the shortest program time. In contrast, type (c) which has the shortest metal wire had the longest program time. These results generally agree with previous studies showing that EM is more pronounced in longer wires [10-11]. Fig. 10 shows the intra-chip and inter-chip Hamming distance distributions of the conventional SRAM PUF and the proposed hybrid PUF. The inter-chip Hamming distance distributions of both PUFs are centered around the ideal value of 0.5. The intra-chip Hamming distance distribution of the conventional SRAM PUF is wide due to its susceptibility to temperature, voltage, and noise effects. In contrast, the hybrid PUF has an intra-chip Hamming distance of zero. Fig. 11 shows the 128 bit key map of two PUF chips before and after baking the chip at 150°C for 192 hours, showing no changes in the key map.

ACKNOWLEDGMENT

This work was supported in part by the Semiconductor Research Corporation (SRC) and the Texas Analog Center of Excellence (TxACE).

REFERENCES

[1] S. Kulkarni, Z. Chen, B. Srinivasan, et al., "A High-Density Metal-Fuse Technology Featuring a 1.6 V Programmable Low-Voltage Bit Cell With Integrated 1 V Charge Pumps in 22 nm Tri-Gate CMOS", IEEE Journal of Solid-State Circuits, vol. 51, no. 4, April 2016

[2] N. Akkaya, B. Erbagci, K. Mai, "Secure Chip Odometers Using Intentional Controlled Aging", IEEE International Symposium on Hardware Oriented Security and Trust (HOST), 2018

[3] K. Chuang, E. Bury, R. Degnaeve, et al., "Physically Unclonable Function Using CMOS Breakdown Position", International Reliability Physics Symposium (IRPS), 2017

[4] K. Chuang, E. Bury, R. Degraeve, et al., "A Multi-bit/cell PUF Using Analog Breakdown Positions in CMOS", International Reliability Physics Symposium (IRPS), 2018

[5] M. Wu, T. Yang, L. Chen, et al., "A PUF Scheme Using Competing Oxide Rupture with Bit Error Rate Approaching Zero", International Solid-State Circuits Conference (ISSCC), 2018

[6] C. Herder, M. D. Yu, F. Koushanfar, et al., "Physical Unclonable Functions and Applications: A Tutorial," Proceedings of the IEEE, vol. 102, no. 8, Aug. 2014.

[7] "PUF-Physical Unclonable Functions-Protecting Next-Generation Smart Card ICs with SRAM-based PUFs." NXP Semiconductor N.V. Feb. 2013. Available: http://www.nxp.com/documents/other/75017366.pdf

[8] B. Karpinskyy, Y. Lee, Y. Choi, et al., "Physically Unclonable Function for Secure Key Generation with a Key Error Rate of 2E-38 in 45nm Smart-Card Chips." International Solid-State Circuits Conference (ISSCC), 2016

[9] T. Ueda, H. Takaoka, M. Hamada, et al., "A Novel Cu Electrical Fuse Structure and Blowing Scheme utilizing Crack-assisted Mode for 90-45nm-node and beyond", VLSI Technology Symposium, 2006.

[10] K. Wu, C. Tseng, C. Wong, et al., "Investigation of Electrical Programmable metal Fuse in 28nm and beyond CMOS Technology" IEEE International Interconnect Technology Conference (IITC), 2011

[11] I. Blech, "Electromigration in thin aluminum films on titanium nitride," J. Appl. Phys. 47, pp. 1203-1208, 1976.

Reliability mechanism	Key properties	PUF hardening	PUF entropy source
Bias temperature instability	Strong recovery, gradual shift	No	No
Hot carrier injection	Irreversible, gradual shift	Maybe	No
Time dependent dielectric breakdown	Irreversible, abrupt shift	Yes	Yes
Electromigration	Irreversible, abrupt shift	Yes	Yes

Fig. 1. Device reliability mechanisms, their key features, and application to PUF design.

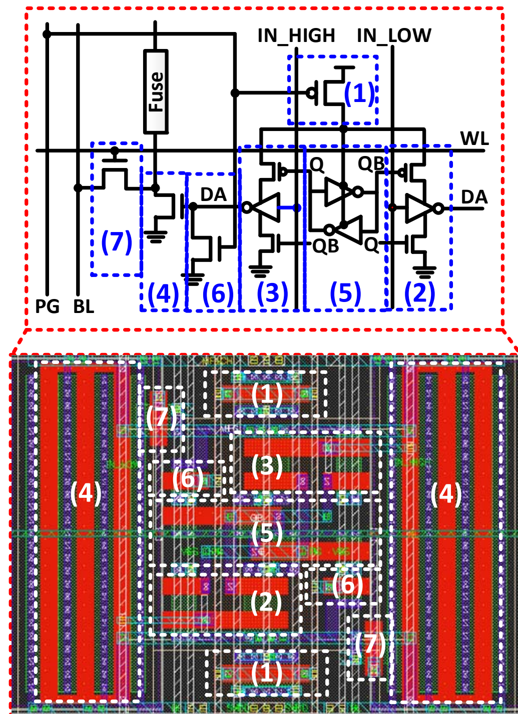


Fig. 3. Unit PUF cell schematic and layout. A perfectly symmetric layout ensures that the PUF response is random. Device definition: (1) SRAM power gating PMOS, (2-3) Tri-state buffers, (4) Split NMOS program device, (5) SRAM cell, (6) Split discharge NMOS, (7) Split read NMOS.

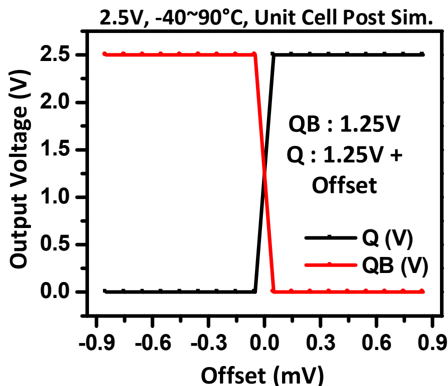


Fig. 4. Post layout simulation results showing the SRAM cell's power up state versus voltage offset between Q and QB nodes. The Q and QB nodes switch at the zero offset condition which confirms a perfectly symmetric PUF cell layout.

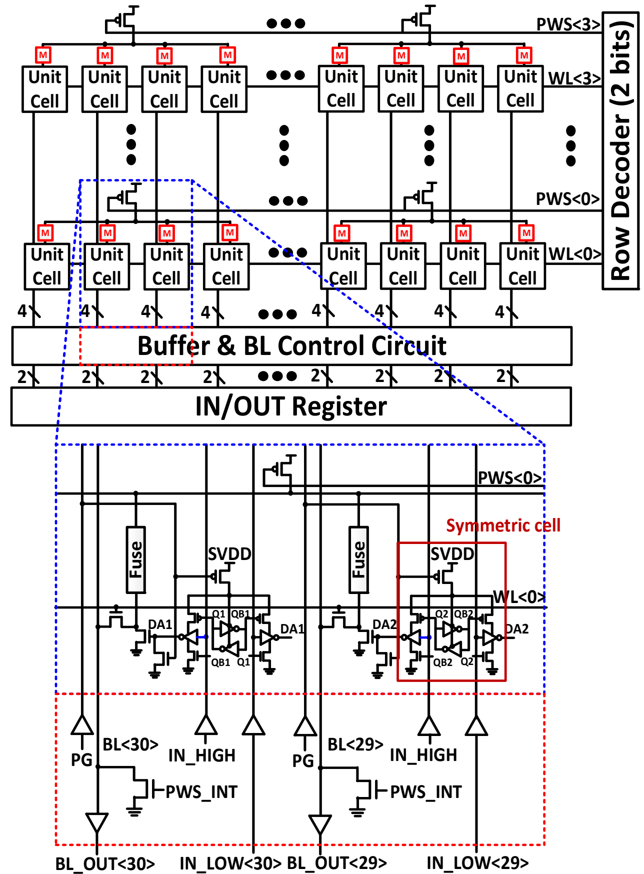


Fig. 2. Hybrid SRAM + metal fuse PUF cell and array architecture. Each row is controlled by the PWS (header PMOS power switch enable) and WL (word line enable) signals. Each column is controlled by the IN_LOW (input low enable) signal. IN_HIGH is a dummy signal which ensures that the SRAM cell layout is perfectly symmetric.

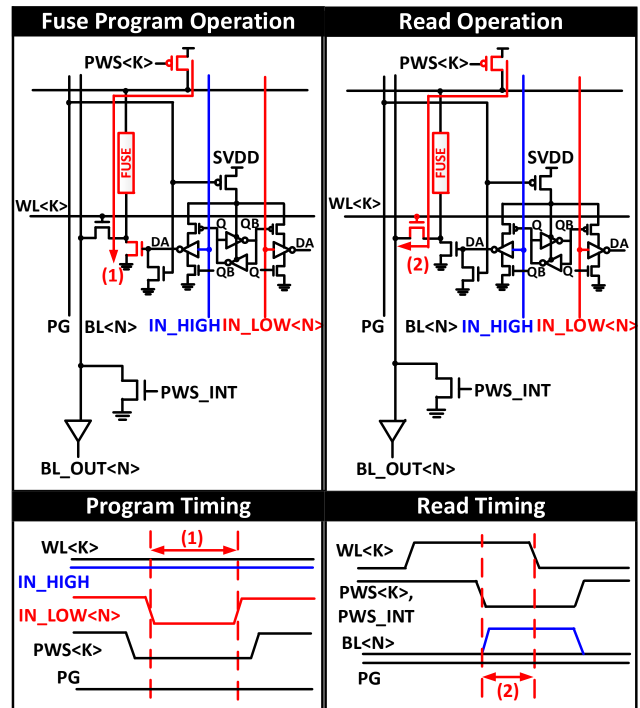
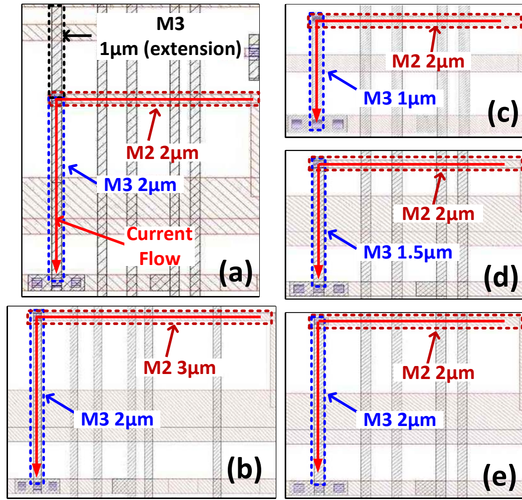
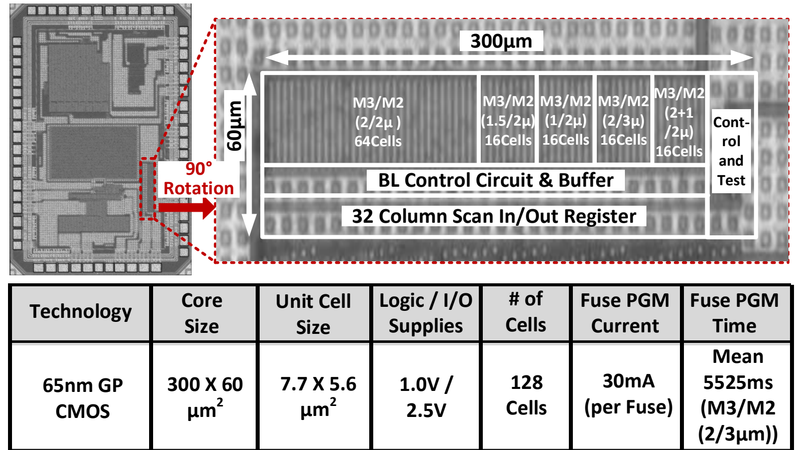


Fig. 5. Program and read operation of proposed hybrid PUF cell. (1) Metal fuse program time. (2) BL (Bit line) charging time.





Technology	Core Size	Unit Cell Size	Logic / I/O Supplies	# of Cells	Fuse PGM Current	Fuse PGM Time
65nm GP CMOS	300 X 60 μm^2	7.7 X 5.6 μm^2	1.0V / 2.5V	128 Cells	30mA (per Fuse)	Mean 5525ms (M3/M2 (2/3µm))

Fig. 7. 65nm chip microphotograph and feature summary table

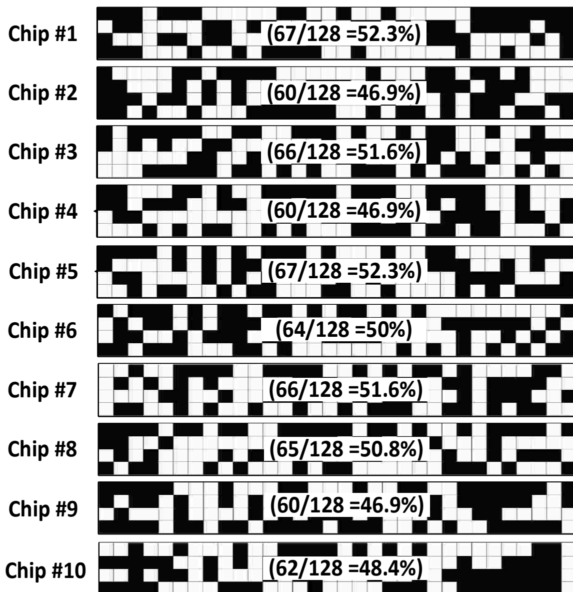


Fig. 6. Metal fuse layouts with different M3 and M2 lengths implemented in the same test chip. (a) 2+1(extension)µm and 2µm, (b) 2µm and 3µm, (c) 1µm and 2µm, (d) 1.5µm and 2µm, (e) 2µm and 2µm. The widths of M2 and M3 are 0.1µm. The program current flows from M2 to M3 in all fuse types.

Fig. 8. 128 bit PUF output measured from 10 chips. The ratio between the number of '1's and number of '0's is roughly 50%.

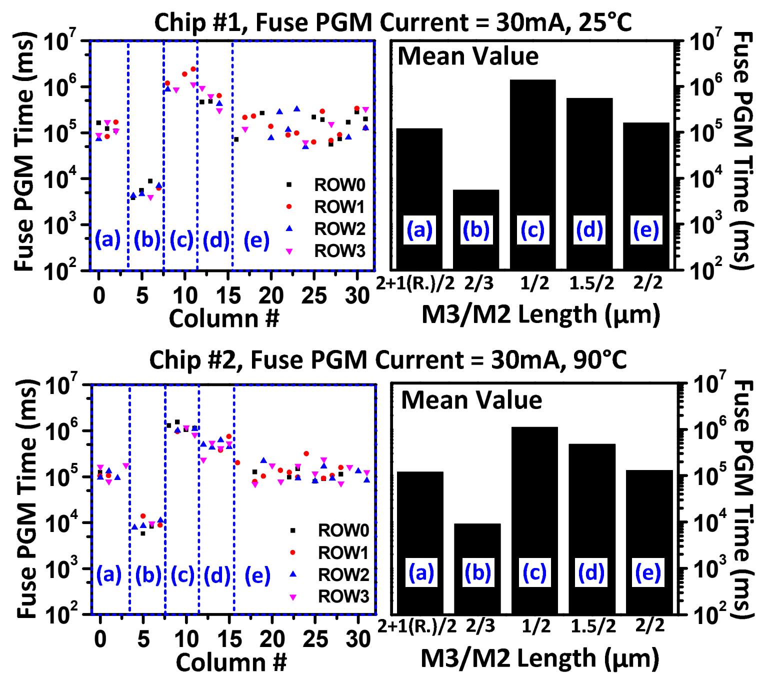


Fig. 9. Program time of different fuse structures measured at 25°C and 90°C for a program current 30mA. The program time trends are similar at two temperatures.

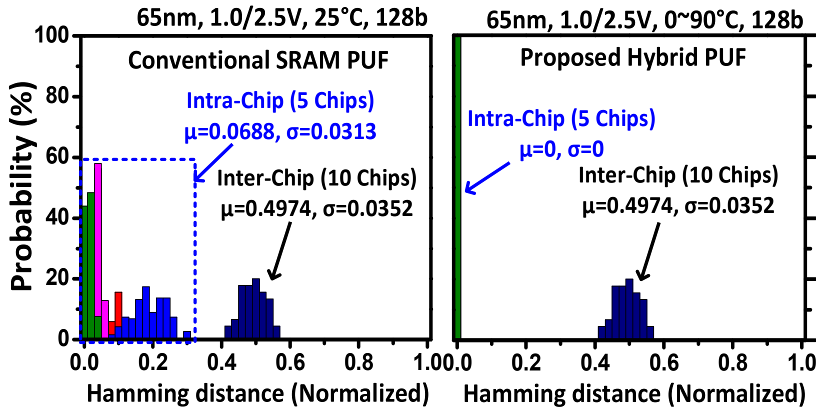


Fig. 10. Intra-chip and inter-chip Hamming distance comparison between conventional SRAM PUF and proposed hybrid SRAM + metal fuse PUF. Intra-chip Hamming distance on left figure was measured from 5 chips, and is based on 20 power up trials.

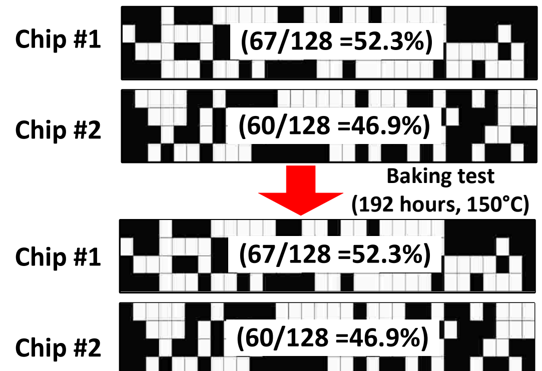


Fig. 11. PUF reliability measured after 192 hours of baking at 150°C. No errors occurred after the baking test indicating a 100% stable PUF response.