On the Relation Between Reliability and Entropy in Physical Unclonable Functions

Vasilii Kulagin, Sergio Vinagrero Gutierrez, Tobias Kilian, Daniel Tille, Ulf Schlichtmann, Giorgio Di Natale, Ioana Vatajelu

On the Relation Between Reliability and Entropy in Physical Unclonable Functions

Vasilii Kulagin[1], Sergio Vinagrero Gutierrez[1], Tobias Kilian[2,3], Daniel Tille[2], Ulf Schlichtmann[3], Giorgio Di Natale[1], Elena-Ioana Vatajelu[1]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP*, TIMA, 38000 Grenoble, France
[2]Infineon Technologies AG, Munich, Germany
[3]Technical University of Munich, Germany

***Abstract*—Physical Unclonable Functions (PUFs) are integral for generating unique signatures, secret keys, and device identification, leveraging inherent manufacturing process variability. Mathematically defined as functions linking inputs (challenges) to outputs (responses), PUFs exhibit random properties. Key properties for high-quality PUFs include intra-device entropy (random distribution of responses within the same circuit), inter-device entropy (random distribution across different circuits for identical challenges), and reliability (response consistency for identical challenges and the same circuit). Inter-device entropy and reliability may be influenced by design discrepancies, systematic variability, noise, and aging. This paper addresses the correlation between entropy and reliability, providing evidence from an extensive set of circuits featuring diverse Ring Oscillators supplied by Infineon.**

***Index Terms*—Physical Unclonable Functions, Ring Oscillator, Reliability, Entropy**

I. Introduction

PUFs, or Physical Unclonable Functions, are cost-effective and tamper-proof mechanisms used for generating unique signatures, secret keys, and device identification. Instead of relying on non-volatile memories, PUFs leverage the inherent variability introduced during the manufacturing process [1]. This variability, encompassing both systematic and random variations, serves as the foundation for the unpredictability that makes each device distinct. Consequently, PUFs cannot be cloned because duplicating identical physical behaviour, even with full understanding of the circuit's mask, remains unachievable. In mathematical terms, a PUF [2] is a function characterized by random properties, which links an input (referred to as a challenge) to an output (recognized as a response). When various devices are given an identical challenge, they generate responses following a random distribution. Nevertheless, when the exact challenge is provided to a particular device, it consistently delivers the same output, thereby guaranteeing the reliability of the PUF response.

PUF architectures differ based on the physical characteristics they utilize [3], [4]. One extensively researched type of PUF relies on comparing nominally identical physical quantities, such as frequency, delay, or resistance value. Challenges enable the selection of a pair of elements for comparison, while responses are produced through this comparison. This paper specifically concentrates on Ring Oscillator PUFs (RO-PUFs), yet the findings and insights presented can be extrapolated to any other PUF architecture.

In the RO-PUF architecture, the outputs of two chosen Ring Oscillators (ROs) are connected to a counter, which tallies the number of cycles within a predetermined time period. The generated output is determined by comparing the counted cycle numbers. If the difference between the cycle counts is greater than zero, the response is '1'; otherwise, it is '0'. To generate multiple bits, successive challenges are applied, resulting in a series of interactions known as Challenge-Response Pairs (CRPs).

In order to achieve high-quality PUFs, the following key properties need to be ensured:

1) Random distribution of responses within the same circuit for different challenges (referred to as intra-device entropy, assessed by Uniformity Per Device - see Section II).
2) Random distribution of responses across different circuits for identical challenges (known as inter-device entropy, evaluated through Uniformity Per Challenge - see Section II).
3) The consistency of responses, meaning that for identical challenges and the same circuit, the response remains constant (evaluated by Reliability - see Section II).

Ensuring intra-device entropy is relatively straightforward by creating exact duplicates of the same electrical components. Conversely, inter-device entropy may be influenced by design discrepancies or systematic variability, while reliability might be impacted by noise and aging. Although many studies have addressed the aforementioned issues independently, only a few have explored the relationship between entropy and reliability [5]. This paper aims, for the first time, to formally establish this correlation, both theoretically and experimentally, presenting results obtained from an extensive set of circuits embedding a wide array of ROs provided by *Infineon*.

The rest of this paper is organised as follows: Section II describes the theoretical background of PUFs and our research hypotheses; Sections III describes the experimental setup; the obtained results are presented in Section IV; finally Section V concludes the paper.

*Institut National Polytechnique Grenoble Alpes

II. PUF Metrics and Research Hypothesis

A. Metrics

In order to measure the quality of PUFs, there exists a set of established standard metrics detailed in [6]. These statistical indicators evaluate the system's ability to generate unique responses, as well as the entropy and the stability of each response. We briefly describe the most used canonical metrics and their security implications, by considering: D the set of devices, C the set of possible challenges, R the set of responses of a device. Moreover, the operator $\#X$ denotes the number of elements in the X set and lowercase letters are used to identify a single element of a set: d represents a single device; c represents a single challenge. The operator HD refers to the Hamming distances between two vectors and HD_F represents the fractional (normalised) Hamming distance.

Uniqueness determines the ability of the PUF to distinguish the devices in the system. It is calculated as weighted sum of Hamming distance between the responses of each possible pair of devices. The number of device pairs is computed as $nPairs = \#D(\#D-1)/2$.

$$Uniqueness = \frac{1}{nPairs} \sum_{i=1}^{\#D-1} \sum_{j=i+1}^{\#D} HD_F(R_i, R_j) \quad (1)$$

Uniformity Per Device (UPD) assesses the intra-device entropy by measuring the statistical distribution of all responses within a single device. A non-uniform distribution implies that an attacker could potentially predict a response from a set of known responses associated with the same device. To better highlight the average amount of information, we have defined UPD by applying the Shannon Entropy operator:

$$\mathcal{H}(p)) = -p \cdot log_2(p) - (1-p) \cdot log_2(1-p) \quad (2)$$

which provides the best score (i.e., 1) when the distribution is perfectly uniform (50% of '0'and 50% of '1').

$$UPD(d) = \mathcal{H}\left(\frac{1}{\#C} \sum_{r \in R_d} r\right) \quad (3)$$

Uniformity Per Challenge (UPC) evaluates the inter-device entropy by measuring the distributions of responses across devices for the same given challenge, as described in Equation 4. If a bias is present (also known as *bitliasing*, i.e., the number of zeros and ones are not equally distributed), a majority of devices produces the same response for a given challenge. This is undesirable as challenges that present bitaliasing can be used by an attacker to guess the responses of devices from the knowledge of the CRPs of other devices. As for UPD, the use of the Shannon entropy operator provides the best score (i.e., 1) when the distribution is perfectly uniform (50% of '0'and 50% of '1').

$$UPC(c) = \mathcal{H}\left(\frac{1}{\#D} \sum_{r \in R_c} r\right) \quad (4)$$

Reliability measures the ability of the PUF to provide the same response when the same challenge is applied multiple times to the same device. If subsequent evaluations of the PUF provide different responses, then the system will incorrectly identify the devices or generate wrong signatures/keys. Reliability is affected by variations in environmental conditions [7] and by aging [8], [9]. It is shown in [10] that the number of unreliable responses can be as high as 11% depending on the conditions. Per each challenge, the reliability can be calculated as the maximum number of identical responses over time (when assuming T measurements):

$$Reliability(c) = \max\left[\left(\sum_{t=1}^{T} R_c^t\right), \left(1 - \sum_{t=1}^{T} R_c^t\right)\right] \quad (5)$$

while the reliability of a device is defined as the average reliability of its challenges:

$$Reliability(d) = \frac{1}{\#C} \sum_{c \in C_d} Reliability(c) \quad (6)$$

B. Research Hypotheses

In this paper we focus on PUFs whose responses are generated by comparing nominally-identical physical quantities, such as frequency (e.g., RO-PUF), delay (e.g., Arbiter-PUF), or resistance value (e.g., MTJ-PUF). For example, RO-PUFs generate responses by comparing the frequencies of a pair of ROs. If the frequency of the first oscillator is greater than the frequency of the second oscillator, then the generated response is '1'. Otherwise, the response is '0'. Studies have shown that responses generated by pairs of elements with very similar physical properties (e.g., RO pairs with nearly identical oscillation frequencies) are more susceptible to noise and therefore less reliable [11], [12]. This susceptibility is due to the fact that any source of noise, such as variations in temperature, voltage, or aging, can potentially cause the relative magnitudes of the physical properties to change, leading to a different response output. Conversely, when the two compared elements exhibit significant differences, it may suggest the presence of bitaliasing, i.e., all devices tend to generate the same response, as demonstrated in simulation-based studies [5].

In this chapter we present a mathematical formulation of the relation between distance among physical quantities, and both Uniformity Per Challenge (UPC) and Reliability. Moreover, we formulate the direct relation between UPC and Reliability. Let us consider a response created by comparing the measurements of two elements *e1* and *e2* (e.g., two frequencies in RO-based PUF), as shown in Figure 1.a. The response *r* is calculate as:

$$r = \begin{cases} \text{'1'}, & \text{if } e1 - e2 > 0 \\ \text{'0'}, & \text{if } e1 - e2 < 0 \end{cases}$$

Let us suppose that the two elements have measurement values distributed (among all devices, d_1 to d_n) as normal distributions: $E1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $E2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, as

shown in Figure 1.b. Therefore, their difference (Δ, Figure 1.c) is distributed according to Equation 7.

$$\Delta = E1 - E2 \sim \mathcal{N}\left(\mu_2 - \mu_1,\ \sqrt{\sigma_1^2 + \sigma_2^2}\right) \tag{7}$$

Starting from the distribution Δ, the probability p('0') of obtaining a '0'as response (i.e., $e1 < e2$) corresponds to the cumulative distribution function $\Phi(0)$, i.e., the area of the probability density function of Δ from $-\infty$ to 0 (respectively, $1 - \Phi(0)$ being the probability p('1') of the response '1').

Our research hypotheses are:

Research hypothesis 1: in the case only random variability is present, μ_1 and μ_2 should have the same value (i.e., $\delta = \mu_1 - \mu_2 = 0$) thus leading to an equal probability of responses at '0'and '1'. Nevertheless, because of systematic process variability and design choices, the two values might differ (i.e., $\delta \neq 0$), in which case the probability of one of the responses becomes predominant. (Figure 1.c shows an example with positive δ, which leads to the probability $p('1')$ larger than $p('0')$. This larger $p('1')$ leads to a decrease of inter-device entropy, degrading the UPC. Figure 1.e shows the effect of increasing δ on UPC, which has been estimated by re-writing Equation 4 as $UPC = \mathcal{H}(p('1'))$.

Research hypothesis 2: it has been demonstrated that responses generated by elements with similar measurements (i.e., e1 - e2 $\approx$ 0) are less reliable. More precisely, as shown in [8], if $e1 - e2 < t$ the response is unstable in time; *t* depends on technology, design and expected operation conditions. Reliability can be estimated by re-writing Equation 5 as shown in Figure 1.d, which is determined by the area of the distribution Δ around 0 (i.e., between $-t$ and t). As for UPC, reliability depends on δ. Indeed, for $\delta \neq 0$, the window of unreliability slides towards the extremes of Δ distribution, therefore the are of concern becomes smaller, increasing thus the reliability of the PUF.

Corollary: the direct consequence of the two research hypotheses is that there is a correlation between inter-device entropy (UPC) and reliability. Indeed, since UPC decreases and reliability increases with increasing δ, we postulate that large reliability implies low inter-device entropy and vice-versa.

In Section IV we will provide experimental evidence of these hypotheses.

III. Circuit description

The circuit under investigation is a prototype of a large general-purpose SoC built for exploratory purposes in the 28nm CMOS technology design. The circuitry is designed to be used in an industrial environment under severe operation conditions. Besides large areas of digital logic, a mixed-signal part, onboard memories, monitoring, and learning structures, the SoC has an exhausting amount of on-chip Ring Oscillators (ROs).

Those ROs are grouped in six identical modules spatially distributed over the SoC (Figure 2.a). All six modules (A÷F) are identically designed. Each module contains 255 ROs of different path topologies, i.e, some paths are built from homogeneous gates like inverter gates, NOR gates, or NAND gates, while other ROs are path replicas of the dedicated paths in the design. All ROs oscillate in the range of 50 to 400MHz (when measured at nominal conditions) and they are intended to be used for timing and process monitoring purposes. Although they are not designed to act as PUF, since the six modules are identically designed, an RO-based PUF can be imagined where frequencies of theoretically identical ROs are compared to generate a PUF response. In other words, to emulate a PUF behavior based on the existing ROs, we compare the frequencies of the ROs in the same position in two of the six modules (Figure 2.b).

The ROs are measured using an on-chip counter, which is the same procedure used in RO-PUFs. The RO frequencies are measured during the manufacturing stages at the wafer level (front-end, FE) and on the packed SoC in the back end (BE). Each frequency read-out is conducted at two temperatures (cold, hot) and two voltages (minimum, nominal). The frequency measurement is accurate, and the voltage and temperature are controlled for perfect conditions. The frequency values of all ROs are logged and used for this experiment. This results per SoC in 8 measured sets of 6 x 255 (1530) RO frequencies along the manufacturing process.

IV. Experimental results

In this section we show the results that are validating our hypotheses presented in Section II. We have collected data from 400 chips, each containing 1530 ROs, as described in Section III. One RO will be referred to as $RO^d_{i,b}$ where i ranges from 1 to 255 (representing the number of ROs in one block), b represents the block (A,B,C,D,E,F) and d ranges from 1 to 400 (representing the number of chips). We have measured their frequencies in the 8 conditions (all combinations of back- and front-end, two temperatures, and two voltages). One measured frequency will be referred to as $\mathrm{f}^c_{i,b,d}$, where c represents the measurement conditions.

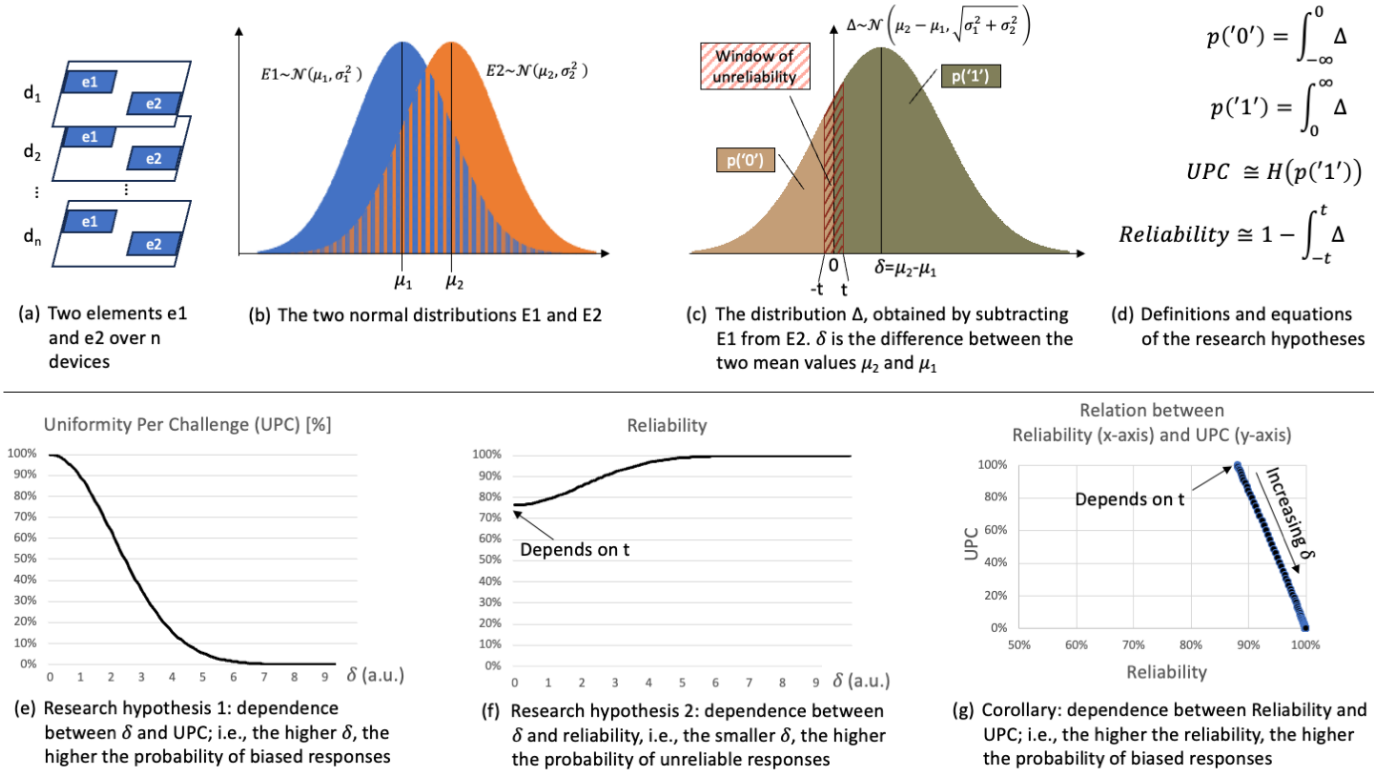
We conceived the PUF as proposed in Figure 2b: for each position i of an RO we have defined 9 possible challenges, i.e., all combinations of blocks $(b1\ b2)$={AD, AE, AF, BD, BE, BF, CD, CE, CF}, leading to overall 2295 CRPs. The set of CRPs enabled us to compute UPC and reliability as defined in Equations 4 and 5.

In order to demonstrate our first hypothesis (i.e., UPC decreases when δ increases), we analysed the full set of frequencies and the corresponding CRPs extracted under nominal conditions (i.e., measured at back-end, cold temperature, and nominal supply voltage). For each RO, we calculated the average frequency over all chips as follows:

$$\overline{f^{be,cold,nom}_{i,b}} = \frac{1}{400}\sum_{d=1}^{400} f^{be,cold,nom}_{i,b,d} \tag{8}$$

For each challenge $(i, b1\ b2)$ we have calculated the UPC and the corresponding δ as follows:

$$\delta_{i,b1\ b2} = \mid \overline{f^{be,cold,nom}_{i,b1}} - \overline{f^{be,cold,nom}_{i,b2}} \mid \tag{9}$$



(a) Two elements e1 and e2 over n devices

(b) The two normal distributions E1 and E2

(c) The distribution Δ, obtained by subtracting E1 from E2. δ is the difference between the two mean values μ_2 and μ_1

(d) Definitions and equations of the research hypotheses

(e) Research hypothesis 1: dependence between δ and UPC; i.e., the higher δ, the higher the probability of biased responses

(f) Research hypothesis 2: dependence between δ and reliability, i.e., the smaller δ, the higher the probability of unreliable responses

(g) Corollary: dependence between Reliability and UPC; i.e., the higher the reliability, the higher the probability of biased responses

Fig. 1: Research hypotheses

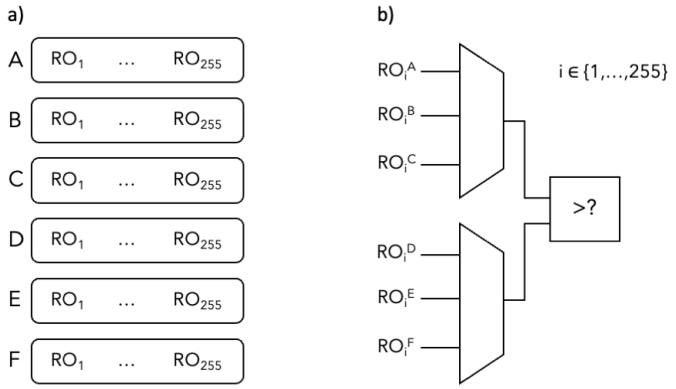


Fig. 2: Schematic representation of the circuit under study: a) the six modules containing 255 ROs; b) the prinicple of operation of the PUF emulator

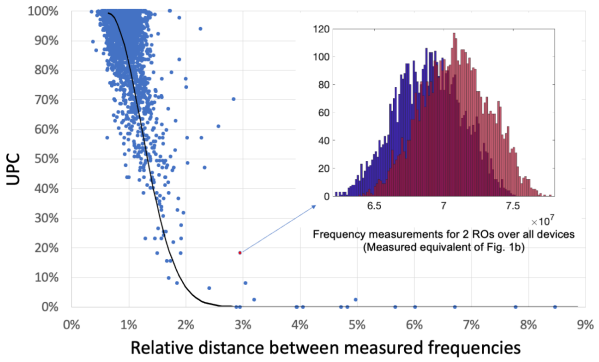


Fig. 3: Uniformity Per Challenge (UPC) in function of the relative distance between measured frequencies.

Moreover, for consistency in representation, because the range of frequencies in a block is wide, we have calculated the relative distance between frequencies, which is give by:

$$\left(\delta_{i,b1\ b2} / \overline{f_{i,b1}^{be,cold,nom}}\right) * 100[\%] \tag{10}$$

Figure 3 illustrates the correlation between the UPC and the relative distance between measured frequencies. Each individual blue data point represents the UPC for a distinct challenge. The black line represents the fitting of the experimental data to the theoretical model previously depicted in Figure 1e. Moreover, the inset provides a visual representation of the frequency distributions utilized in the computation of a single challenge, denoted by the red data point in the figure. The blue and red histograms correspond to the frequency distributions of the two Ring Oscillators (ROs) associated with the challenge, serving as the empirical equivalent of the conceptual representation in Figure 1b. The findings presented in this figure provide substantial evidence supporting the validity of our initial hypothesis.

To substantiate our second hypothesis, which posits that reliability increases concurrently with an increase in δ, we conducted an analysis of the comprehensive set of frequencies

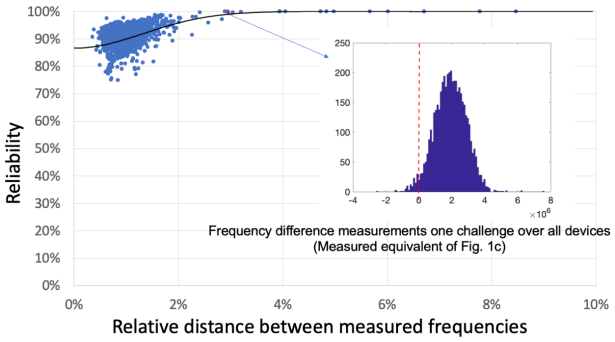


Fig. 4: Reliability in function of the relative distance between measured frequencies.

and their corresponding CRPs obtained under all experimental conditions. For each challenge, we computed the reliability according to the definition provided in Equation 5. Figure 4 exhibits the correlation between reliability and the relative distance between measured frequencies. Each blue data point signifies the reliability for a distinct challenge. The black line represents the fitting of the experimental data to the theoretical model previously illustrated in Figure 1f. The inset provides a visual representation of the distribution Δ of frequency differences for the same challenge as depicted Figure 3. The red line denotes the 0 value of Δ. Upon examination, it is evident that the 0 is located at the extreme left of the distribution. This observation substantiates both the bias, wherein the majority of responses exhibit a value of '1', and the high reliability of the responses, as the region around 0 is relatively small. Consequently, this empirical distribution serves as the practical equivalent of the conceptual representation depicted in Figure 1c. The findings presented in this figure provide substantial evidence supporting the validity of our second hypothesis.

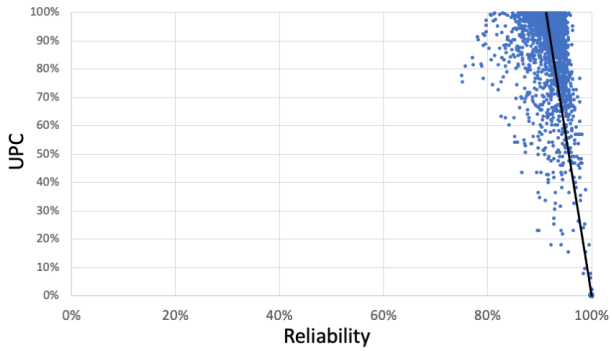


Fig. 5: Relationship between uniformity per challenge (UPC) and reliability.

Corollary: The direct implication of the two research hypotheses is the establishment of a correlation between inter-device entropy (measured by UPC) and reliability. As UPC decreases and reliability increases concurrently with the augmentation of δ, we demonstrate that high reliability corresponds to low inter-device entropy, and vice versa. Figure 5 illustrates this dependency between UPC and reliability. Each blue data point represents the UPC and reliability of a single challenge. The black line represents the fitting of the experimental data to the theoretical model previously depicted in Figure 1g. This correlation between UPC and reliability further supports the validity of our research hypotheses and their implications for inter-device entropy and reliability.

V. Conclusion

We introduced a theoretical approach to determine the reliability and entropy of RO-PUF, based on the average differences between the frequencies of two ROs. To support our theoretical hypotheses, we conducted extensive measurements on thousands of ROs from a set of industrial SoC prototypes. Our findings revealed an important insight: RO pairs with an average frequency difference close to zero tend to be less reliable, while pairs with a larger average difference have insufficient entropy. We also demonstrated the interdependence of these two effects, i.e., that there is a strong inverse correlation between reliability and inter-device entropy.

These findings contribute to the comprehension of PUF dynamics, a critical factor in fortifying PUF-based security protocols across a multitude of applications. While existing solutions focus on filtering out frequency pairs with minimal frequency differences, we suggest that future solutions should also consider excluding responses generated by RO pairs with frequencies that are significantly distinct from one another.

References

[1] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *Proceedings of the 44th Annual Design Automation Conference*, ser. DAC '07, San Diego, California: Association for Computing Machinery, 2007, pp. 9–14, ISBN: 9781595936271. DOI: 10.1145/1278480.1278484. [Online]. Available: https://doi.org/10.1145/1278480.1278484.

[2] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld, "Physical one-way functions," *Science*, vol. 297, no. 5589, pp. 2026–2030, 2002.

[3] B. Gassend, D. Clarke, M. Van Dijk, and S. Devadas, "Delay-based circuit authentication and applications," in *Proceedings of the 2003 ACM symposium on Applied computing*, 2003, pp. 294–301.

[4] Z. Cherif, J.-L. Danger, F. Lozac'h, Y. Mathieu, and L. Bossuet, "Evaluation of delay pufs on cmos 65 nm technology: Asic vs fpga," in *Proceedings of the 2nd International Workshop on Hardware and Architectural Support for Security and Privacy*, 2013, pp. 1–8.

[5] S. Vinagrero Gutierrez, G. Di Natale, and E.-I. Vatajelu, "On-line method to limit unreliability and bit-aliasing in ro-puf," in *2023 IEEE 29th International Symposium on On-Line Testing and Robust System Design (IOLTS)*, IEEE, 2023, pp. 1–6.

[6] A. Maiti, J. Casarona, L. McHale, and P. Schaumont, "A large scale characterization of ro-puf," in *2010 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*, IEEE, 2010, pp. 94–99.

[7] R. Maes, "An accurate probabilistic reliability model for silicon pufs," in *International Conference on Cryptographic Hardware and Embedded Systems*, Springer, 2013, pp. 73–89.

[8] M. Barbareschi et al, "A ring oscillator-based identification mechanism immune to aging and external working conditions," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. PP, pp. 1–23, Aug. 2017. DOI: 10.1109/TCSI.2017.2727546.

[9] N. Karimi, J.-L. Danger, and S. Guilley, "Impact of aging on the reliability of delay pufs," *Journal of Electronic Testing*, vol. 34, no. 5, pp. 571–586, 2018.

[10] S. Katzenbeisser, Ü. Kocabaş, V. Rožić, A.-R. Sadeghi, I. Verbauwhede, and C. Wachsmann, "Pufs: Myth, fact or busted? a security evaluation of physically unclonable functions (pufs) cast in silicon," in *International Workshop on Cryptographic Hardware and Embedded Systems*, Springer, 2012, pp. 283–301.

[11] A. Schaub, J.-L. Danger, S. Guilley, and O. Rioul, "An improved analysis of reliability and entropy for delay pufs," in *2018 21st Euromicro Conference on Digital System Design (DSD)*, 2018, pp. 553–560. DOI: 10.1109/DSD.2018.00096.

[12] H. Martin et al, "On the reliability of the ring oscillator physically unclonable functions," in *2019 IEEE 4th International Verification and Security Workshop (IVSW)*, IEEE, 2019, pp. 25–30.

Vasilii Kulagin is currently pursuing a Master Electronics, Electrical Energy, Automation at Grenoble-Alpes University (UGA), France.

Sergio Vinagrero Gutierrez is currently a PhD Candidate at UGA/TIMA Laboratory. He received the M.S. degree in Electronic Engineering from UGA in 2021. His research interests are on Hardware Security.

Tobias Kilian received the B.Sc. and M.Sc. degrees in electrical engineering and information technology from the Technical University of Munich (TUM), Munich, Germany. He is currently pursuing the Ph.D. degree as part of a collaborative project between Infineon Technologies A.G. and the TUM. His research focus lies on architectures and monitoring structures for next-generation microcontrollers. He is currently working as an SoC Architect at Infineon Technologies AG.

Daniel Tille received his Diploma Degree in Computer Science from University of Halle, Germany, and his PhD Degree in Engineering from University of Bremen, Germany, in 2006 and 2011, respectively. Since 2012, he has been with Infineon Technologies in Munich, Germany, in different DFT-related roles. Currently, he is Director of the Verification and Validation department in the Automotive Smart Power business unit. His research interests cover all relevant DFT methods for automotive applications, especially ATPG, LBIST and Test Point Insertion

Ulf Schlichtmann received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering and information technology from the Technical University of Munich (TUM), Germany, in 1990 and 1995, respectively. He is Professor and the Head of the Chair of Electronic Design Automation, TUM. His current research interests include computer-aided design of electronic circuits and systems, with an emphasis on designing reliable and robust systems.

Giorgio Di Natale received the PhD in Computer Engineering in 2003. He works as Director of Research with CNRS. His research interests include hardware security and trust, secure circuits design and test, reliability evaluation and fault tolerance.

Elena-Ioana Vatajelu is researcher with CNRS on the design, test and reliability of Integrated Circuits. She obtained her PhD from UPC Spain in 2011. Her expertise is on the reliability and the robustness assessment, design-for- reliability, test strategies and security primitives for CMOS and beyond CMOS RAMs in traditional and non-Von Neumann computing paradigms.