Design, Evaluation, and Optimization of Physical Unclonable Functions Based on Transient Effect Ring Oscillators

Abdelkarim Cherkaoui, Lilian Bossuet, *Senior Member, IEEE*, and Cédric Marchand

***Abstract*— This paper proposes a theoretical study and a full overview of the design, evaluation, and optimization of a PUF based on transient element ring oscillators (TERO-PUF). We show how, by following some simple design rules and strategies, designers can build and optimize a TERO-PUF with the state-of-the-art PUF characteristics in a standard CMOS technology. To this end, we analyzed the uniqueness, steadiness, and randomness of responses generated from 30 test chips in a CMOS 350-nm process in nominal and corner voltage and temperature conditions. Response generation schemes are proposed and discussed to optimize the PUF performances and reduce its area without noticeable loss in its output quality. In particular, we show that the large area of the basic blocks in the TERO-PUF is balanced by the high level of entropy extracted in each basic block. Guidelines are provided to balance reliability and randomness of the responses and the design area.**

***Index Terms*— Information security, cryptography, digital signatures, authentication.**

I. Introduction

A NOVEL approach for the identification and authentication of electronic devices emerged and has received quite some attention in the last few years. The new paradigm aimed at physically identifying hardware systems, instead of associating them with an explicitely programmed digital identity. The concept of physical unclonable functions (PUFs) was first introduced by Ravikanth in [1]. PUFs can extract unique secret keys from the physical characteristics of the device using a challenge and response procedure based on a physical interaction which is extremely hard or impossible to reproduce. Entropy is derived from a physical random variable such as the mismatch between transistor attributes (length, width, oxide thickness, etc.) caused by manufacturing process variability (MPV). The basic principle is that MPV is neither controllable (it is not predictable) nor reproducible, but can be measured. Ideally, when a PUF is challenged, its response is unique (each device has a unique, non reproducible response based on its unique physical characteristics), random (it is uniformely distributed and it cannot be predicted), steady (each device always gives the same response to a given challenge) and in some cases tamper resistant (probing the PUF changes its physical behavior and hence the obtained response).

Many silicon based PUF architectures exist, but two main approaches are used to extract entropy from MPV in digital devices: methods based on the measurement or comparison of timing and methods that exploit the resolution of a metastable state. SRAM-PUFs [2] and butterfly PUFs [3] rely on the settling state of cross-coupled elements: at the initialization of a SRAM, most cells' outputs are biased toward '1' or '0' depending on MPV. The arbiter PUF [4] relies on the race between two events (electrical transitions) in two identical delay lines. The ring oscillator based PUF [5] (RO-PUF) leverages the frequency mismatch between several identically designed ring oscillators (ROs). ROs can easily be implemented in both ASICs and FPGAs, and they have been widely used to measure and model MPV [6]. Numerous studies have shown that when correctly implemented, the uniqueness, steadiness and randomness of the RO-PUF are adequate, which is why it is currently considered to be one of the best PUF candidates [7], [8]. However, there are two main constraints in the use of ROs in a security primitive: the ROs must be independent, and their frequencies must be hidden.

Recent studies have shown that, in practice, ROs do not meet these requirements. When identical ROs are implemented in the same device, dependencies in their switching times may occur [9]. On rare occasions, two ROs may naturally lock on the same frequency. This state of locking can be caused intentionally either by manipulating the power supply voltage [9] or by harmonic electro-magnetic injection [10]. In the case of the RO-PUF, this latter contactless attack may render a large portion of the PUF responses predictable. On the other hand, RO frequencies and their location on the chip can be retrieved using electro-magnetic analysis [11]. Information leaked via the electro-magnetic channel can therefore help to mathematically clone the PUF.

To circumvent these issues, [12] proposes to use transient effect ring oscillators (TEROs) as an alternative to classical ROs. TEROs are supposed to be more robust against

Manuscript received June 19, 2015; revised October 20, 2015; accepted January 15, 2016. Date of publication February 3, 2016; date of current version March 23, 2016. This work was carried out in the framework of the SALWARE project ANR-13-JS03-0003 supported by the French Agence Nationale de la Recherche and by the French Fondation de Recherche pour l'Aéronautique et l'Espace. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mauro Barni.

A. Cherkaoui is with the Techniques de l'Informatique et de la Microélectronique pour l'Architecture des Systèmes Intégrés Laboratory, Université Grenoble Alpes, Grenoble 38031, France, and also with the Centre National de la Recherche Scientifique, Techniques de l'Informatique et de la Microélectronique pour l'Architecture des Systèmes Intégrés Laboratory, Grenoble 38031, France (e-mail: abdelkarim.cherkaoui@imag.fr).

L. Bossuet and C. Marchand are with the Laboratoire Hubert Curien, Centre National de la Recherche Scientifique, University of Lyon, Saint-Etienne 42000, France (e-mail: lilian.bossuet@univ-st-etienne.fr; cedric.marchand@univ-st-etienne.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2016.2524666

locking phenomena because they present only brief transient oscillations (and therefore also have a lower EM emanation). On the other hand, a TERO based PUF (TERO-PUF) extracts MPV by measuring the difference in the number of transient oscillations in two identical TEROs, making it possible to extract more bits per response than a simple comparison. Based on the work presented in [12], a similar approach can be used for RO-PUFs.

Bossuet *et al.* [12] present the TERO-PUF and focus on its proof of concept, TERO-PUFs were evaluated in FPGAs only in nominal voltage and temperature conditions. In this paper, we provide a full overview of the design, characterization and optimization of a TERO-PUF in a standard CMOS technology in nominal and corner voltage and temperature conditions. In [12], the randomness of responses generated using several bits per challenge were not evaluated. In this paper, we show that this response generation method needs to be set up correctly to avoid bit correlations in the responses and a major response deviation in corner voltage and temperature conditions. The contributions of this paper are: 1) a theoretical study of TERO-PUF behavior and its relevant parameters; 2) full characterization of 30 test chips in nominal and corner voltage and temperature conditions using different response generation schemes; and 3) guidelines for the optimization of the PUF's area and performances.

The paper is organized as follows: Section II presents the core architecture of the TERO-PUF, its general working principle, and compares it with the RO-PUF. The two following sections propose a theoretical study of the PUF. Section III focuses on the temporal behavior of TEROs. Section IV presents the entropy extraction mechanisms and their main parameters in the TERO-PUF. Section V describes the design process of TERO-PUFs in a CMOS 350 nm technology. Section VI provides experimental measurements and the characterization of 30 test chips in nominal and corner voltage and temperature conditions, and compares them with results for RO-PUF. Section VII discusses the results and provides guidelines for optimizing the design. Section VIII concludes the paper.

II. Background

This section introduces the TERO-PUF and highlights its main features compared to the RO-PUF.

A. Transient Effect Ring Oscillators

ROs are digital oscillators consisting of a chain of inverter elements connected to form a loop. Traditionally, they are composed of an odd number of inverters (which may include an initialization stage) to ensure steady oscillatory behavior. TEROs correspond to a very specific configuration of ROs in which the number of inverters is even. Contrary to classical ROs, TEROs require two initialization stages and have two functioning modes: a transient oscillatory state followed by a stable steady state. Fig. 1 shows the generic architecture of ROs and TEROs. The oscillatory behavior in both ROs and TEROs is due to the propagation of electrical transitions across the ring structure. When a stage i has the same

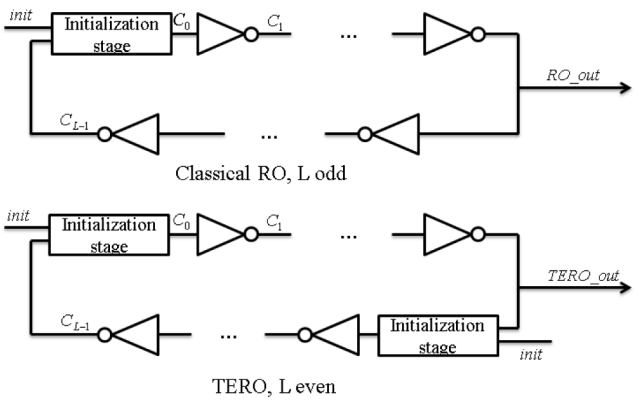

Fig. 1. Generic architecture of ROs and TEROs.

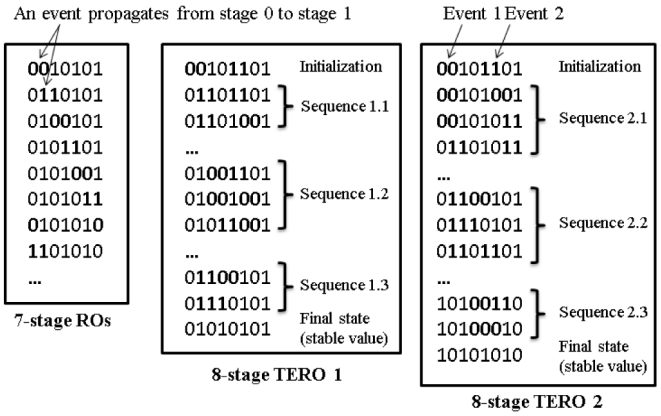

Fig. 2. Typical output sequences of ROs and TEROs.

input and output value, it switches its output after a time D corresponding to the propagation delay of the gate. We refer to this process as an "event" happening at the output of the stage i.

In a RO, there is always one stage which has conflictual output/input values due to the odd number of inverters. This stage switches its output after a propagation delay D, which transfers the output/input conflict to the following stage: an event propagates from stage i to stage $i+1$.

Fig. 2 shows RO and TERO typical output sequences. For 7-stage ROs, an output state is represented by a vector of 7 bits corresponding (from left to right) to the nodes C_0 to C_6 with respect to the index in Fig. 1. Whatever the propagation delays in the ring stages, all 7-stage ROs have the same sequence of states although the time of each state may vary. The oscillatory behavior of the RO corresponds to the constant propagation of one event across the ring structure.

Are also represented in Fig. 2 the sequences of output signals of two 8-stage TEROs, where each stage has a different timing characteristics, from node C_0 to node C_7 according to the index in Fig. 1. The goal of the initialization stages in TEROs is to trigger two events that can propagate simultaneously across the ring structure (by forcing the initial state in Fig. 2). The transient oscillatory regime in TEROs is due to the propagation of those two events across the structure.

Contrary to ROs, the output sequence in TEROs can vary from one TERO to another depending on specific timing

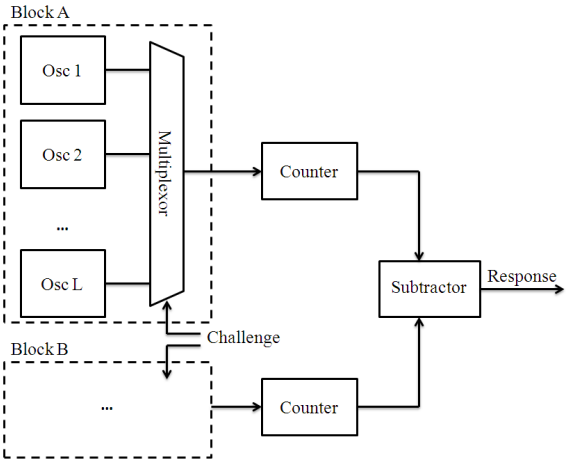

Fig. 3. Generic core architecture of RO-PUFs and TERO-PUFs.

TABLE I

PRACTICAL COMPARISON BETWEEN THE RO-PUF PRESENTED IN [5] AND THE TERO-PUF PRESENTED IN [12]

	RO-PUF [5]	TERO-PUF [12]
Entropy souce	MPV	MPV
Entropy extractor	RO, steady oscillating state	TERO, transient oscillating state followed by a stable steady state
Extraction vector	Oscillation frequency	Duration of the transient oscillations
Bit generation	Frequency comparison	Difference of the number of transient oscillations
Number of response bits per challenge	1 bit in [5], more than one bit possible in theory	Several
Leakages	RO frequencies (EM channel)	Currently unknown, possibly the duration of transient oscillations

characteristics of each ring stage. For instance, in TERO 1, the propagation delay of stage C_1 is shorter than the propagation delay of stage C_5, whereas in TERO 2, the propagation delay of stage C_1 is longer than the propagation delay of stage C_5. Starting from the same initial state "00101101" (Sequences 1.1 and 2.1), the following state in TERO 1 is "01101101" (Event 1 propagates first) whereas in TERO 2 it is "00101001" (Event 2 propagates first). More generally, the transition from one state to another depends on the propagation delays of the stages in which the two events occur, but also on the relative duration of the previous state (or in other words, the relative position of the two events). When two events happen at adjacent stages (Sequences 1.3, 2.2 and 2.3), the input level of one stage may vary before it has completely switched its output (for example to '1'), the output level starts switching in the opposite level (to '0') before the appropriate voltage level ('1') is detected by the next stage (Sequences 1.3 and 2.3). We refer to this situation as a collision between the two events. When events ultimately collide in a TERO, one of its two final steady states (either "01010101" or "10101010" in the 8-stage TERO) is reached.

In practice, the RO produces a digital clock with a duty cycle which approaches 50%. Its frequency depends on the number of inverters as well as on the propagation delay of each ring stage. The TERO produces a signal with transient oscillations and a variable duty cycle. The oscillation frequency depends on the propagation delays of the ring stages as well as on their number. The duty cycle corresponds to the "distance" between the two events as they propagate across the ring (*i.e.* the time that elapses between the two events seen at the output of one ring stage). The number of transient oscillations depends on the propagation delay of each stage but also on the capacitive charge and discharge parameters of its output, as will be shown in Section III.

B. RO-PUFs and TERO-PUFs

Based on [5] and [12], Fig. 3 proposes a generic architecture for oscillator based PUFs.

As originally proposed in [5], the RO-PUF extracts MPV in a digital circuit by comparing the oscillation frequencies of two identically implemented ROs. In Fig. 3, the oscillators are ROs whereas the subtractor is a comparator (which can be seen as a 1-bit subtractor). The number of oscillations counted during a fixed time window is compared to provide a one-bit response to the challenge, which consists in selecting two ROs.

In [12], ROs are replaced by TEROs and the comparator is replaced by an n-bit subtractor. The core architecture of a TERO-PUF is composed of two blocks of L identical TEROs. When selected and initialized, each TERO presents transient oscillations at its output (contrary to ROs, which present permanent oscillations). The number of transient oscillations varies between identically implemented TEROs due to MPV.

In a TERO-PUF, each challenge, denoted $C(i, j)$ (with $0 \leq i, j \leq L-1$), consists of selecting TERO i from Block A, and TERO j from Block B using multiplexors. The number of available challenges is L^2. An n-bit response is obtained from each challenge by subtracting the number of TEROs' oscillations. The goal of this subtraction is to reduce the influence of environmental global variations since they tend to affect each logic cell in the circuit equally. The L^2 challenges can be divided into k sets of m challenges in order to obtain responses of $m \times n$ bits.

When used in security primitives such as PUFs and TRNGs (true random number generators), TEROs have many advantages over classical ROs due to their transient oscillations. Locking phenomena are theoretically less likely due to the very brief time during which the TERO oscillates. In an RO, the number of oscillations in a fixed time window depends directly on the ring frequency. By acquiring the ring frequencies of a RO-PUF using the EM channel, an adversary can mathematically clone the PUF. In Section III and Section V, we show that the number of transient oscillations in a TERO-PUF does not only depend on its frequency, but also on the signal slopes. Measuring the ring frequencies using the EM channel and obtaining all the required parameters to build a model is obviously more challenging and cannot be achieved only using contactless scanning.

Table I is a practical comparison between the RO-PUF, as proposed in [5], and the TERO-PUF. In addition, TEROs are implemented using inverters and NAND or NOR gates. The TERO-PUF is based on digital elements and adapts very well in FPGA and ASIC design flows and process technologies.

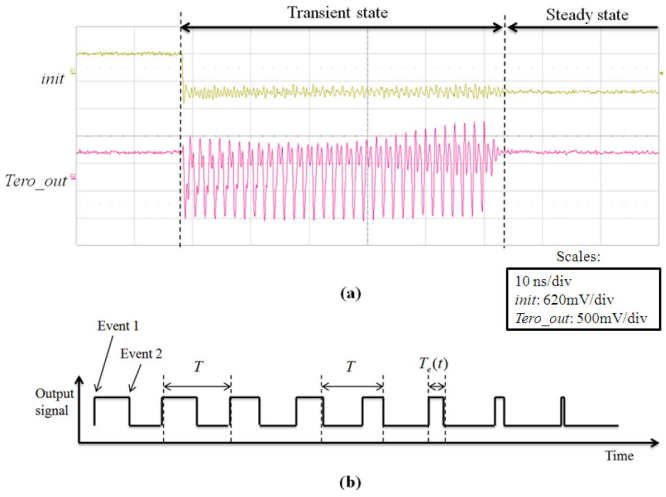

Fig. 4. (a) Output chronogram of a 6-stage TERO in a Xilinx Spartan 6 FPGA. (b) Digital representation of the output signal of a TERO.

The proposed architecture allows a large number of challenge and response pairs (CRPs): L^2 in Fig. 3. Responses can be built using one or more bits from the substractor value. The way the responses are generated significantly affects their statistical quality and the overall size of the design.

III. Temporal Behavior of TEROs and Associated Physical Phenomena

This section describes the temporal behavior of TEROs and analyses it using a mixed analog/digital model for the inverter gate.

A. Experimental Observations

We implemented a few TEROs in two technologies: an Altera Cyclone III FPGA (65 nm process) and a standard CMOS 350 nm process. Part (a) in Fig. 4 is an output chronogram of a 6-stage TERO in a Xilinx Spartan 6 FPGA. During the transient oscillation mode, the output signal presents a constant oscillation period but a variable duty cycle. A closer look at the duty cycle shows that it increases (or decreases) steadily until it reaches 100% (or 0%). The voltage level (peak to peak) of the analog output signal decreases in the final oscillations. When we implemented several 14-stage TEROs in a 350 nm CMOS chip, the mean number of oscillations in each TERO varied between 0 and 263 due to MPV. When we implemented the same TEROs in Xilinx Spartan 6 FPGAs, the number of oscillations varied from 0 to around 1,000. Some of the TEROs presented permanent oscillations in the FPGA implementation, but this phenomenon was not observed in ASICs.

Part (b) in Fig. 4 is a digital representation of the output signal of a TERO. The oscillation period T corresponds to the time required for one event to perform one lap around the TERO (crossing all the TERO stages once), and, according to our observations, remains constant. The time elapsed between Event 1 and Event 2 (seen at the output of the TERO), denoted $T_e(t)$, shrinks over time, which ultimately stops the oscillations. A very similar phenomenon was reported by Winstanley *et al.* who studied the propagation of events in asynchronous micropipelines [13]. It was called *drafting effect* because it caused events to propagate as bunches in asynchronous micropipelines featuring handshake communication protocols. The drafting effect can be explained by taking into account the signal slopes in the propagation delay model of the inverter gate (mixed analog/digital model).

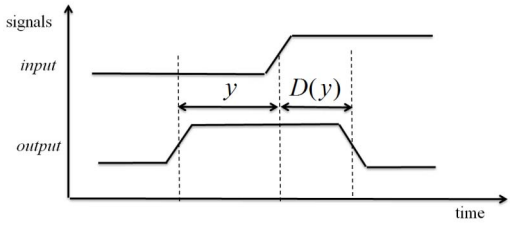

Fig. 5. Output chronogram of an inverter gate.

B. Temporal Model of an Inverter

In digital simulators, the propagation delay of a logic gate is usually modeled by a constant value. Because these simulators are mainly intended to simulate synchronous systems, they usually do not satisfactorily model fast oscillating analog signals (which may mostly consist of signal "slopes"). In practice, when a logic gate switches its output at a very high speed, its apparent propagation delay may be shorter than usual because the output voltage level does not fully reach GND or VCC between two events occuring at the input (instead, it converges assymptotically to VCC or GND). The less steep the signal slope, and the shorter the time between the events at the input of the logic gate, the shorter its propagation delay is. In practice, the signal slopes at the output of a ring stage are directly related to the parasitic output capacitance of the stage concerned. In [13], Winstanley *et. al.* studied the rapid propagation of events in asynchronous micropipelines composed of C-elements, and proposed a model for this analog effect based on the exponential charge and discharge model of a capacitance (it can therefore apply to any logic gate).

Fig. 5 represents the chronogram of an inverter gate. $D(y)$ is the propagation delay of the inverter gate which is a function of y, the time that elapses between the input event and the last output commutation of the logic gate. Applied to the inverter gate, the temporal model presented in [13] translates into the following equation, which represents the propagation delay of the inverter gate as a function of the time elapsed after its last commutation:

$$D(y) = D_m - \alpha e^{-\frac{y}{\tau}} \quad (1)$$

D_m corresponds to the static propagation delay of the gate (when $y \to \infty$). The exponential term is due to the exponential model of capacitive charge and discharge. α is a magnitude factor (in units of time) whereas τ is a duration factor (also in units of time). The most remarkable property of this temporal model is the fact that the propagation delay of the inverter decreases with a decrease in the time elapsed since its last commutation. This has a major consequence for the timing of events propagating in TEROs.

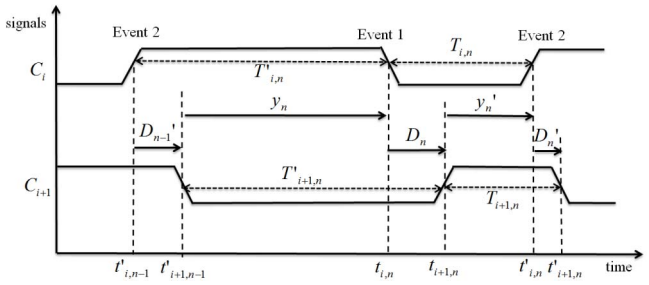

Fig. 6. Typical output chronogram of two adjacent TERO stages.

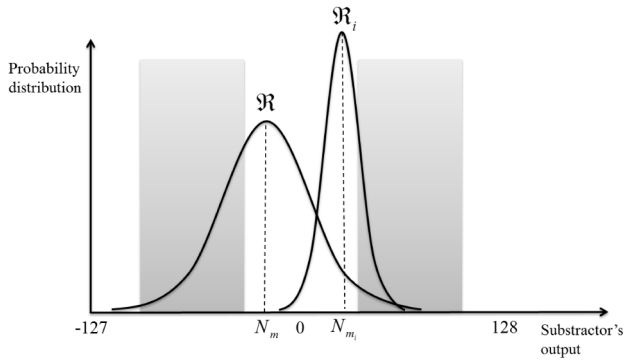

Fig. 7. Expected theoretical distribution of the 8-bit subtractor's output ($\mathcal{R}$ describes extra-chip variations whereas $\mathcal{R}_i$ describes intra-chip variations) and graphic representation of ones (gray) and zeros (white) for the bit 2^5 of its Gray code.

C. The Drafting Effect in TEROs

Fig. 6 represents a generic output chronogram of two adjacent stages in a TERO. An oscillation cycle denoted n corresponds to the period required by one event (such as that defined in section II-A and in part (b) in Fig. 4) to cross the same stage after performing one lap around the ring structure. D_n (repectively D'_n) is the propagation delay of stage $i+1$ when Event 1 (repectively Event 2) crosses this stage in cycle n. $T_{i,n}$ (receptively $T'_{i,n}$) is the time that elapsed between Event 2 and Event 1 (repectively between Event 1 and Event 2) seen in stage i in cycle n. Based on the previous temporal model of the inverter gate, the temporal behavior of a TERO can be explained as follows:

- Let us consider $i \in \{0, 1, \ldots, L-1\}$. If, initially, $T_{i,0} < T'_{i,0}$ then it can be shown that $(T_{i,n})_{n\in\mathbb{N}}$ is a decreasing series while $(T'_{i,n})_{n\in\mathbb{N}}$ is an increasing series. The recursive demonstration is based on the following remark: if $T_{i,n} < T'_{i,n}$, then $y_n > y'_n$ (see Fig. 6), which means that $D_n > D'_n$ according to Eq. (1) ($D(y)$ is an increasing function). Therefore $T_{i+1,n} = T_{i,n} + D'_n - D_n < T_{i,n}$. Note that due to the loop structure of the TERO, we can assume (by convention) that $T_{L,n} = T_{0,n+1}$ (stage L does not actually exist). Hence, the previous relationship leads to $T_{i,n+1} < T_{0,n+1}$ and $T_{L,n} < T_{i+1,n} < T_{i,n}$ with $T_{0,n+1} = T_{L,n}$, therefore $T_{i,n+1} < T_{i,n}$.
- In the same way, if $T_{i,0} > T'_{i,0}$, then $(T_{i,n})_{n\in\mathbb{N}}$ is an increasing series while $(T'_{i,n})_{n\in\mathbb{N}}$ is a decreasing series.
- If $T_{i,0} = T'_{i,0}$, then $(T_{i,n})_{n\in\mathbb{N}}$ and $(T'_{i,n})_{n\in\mathbb{N}}$ are constant series, which means that the TERO oscillates infinitely. However, although possible, this is very unlikely in practice.

The initial conditions depend largely on MPV in each ring stage. The speed of convergence of $(T_{i,n})_{n\in\mathbb{N}}$ also depends on the capacitive charge and discharge parameters in each cell output, *i.e.* the strength of the drafting effect (and they are also affected by MPV), which makes the measurement of the number of transient oscillations a very interesting vector to extract MPV.

IV. Entropy Extraction

As mentioned in Section II, in a TERO-PUF, challenges consist in selecting two TEROs whereas responses are built using the subtractor's output. In this section, we describe the different schemes that can be used to generate the actual PUF response from the subtractor's output, and how noise and MPV affect those responses.

A. Subtractor's Output Distribution and Properties

The objective of the entropy extraction in the TERO-PUF is to maximize the amount of MPV information in the generated responses, and to minimize the influence of noise and environmental variations on those responses. Let us consider a population of M chips, each containing two identical TEROs. A pair of TEROs in chip i has a differential number of oscillations N_i which varies with time due to noise: N_i is a draw of a random variable $\mathcal{R}_i$ with a mean value N_{m_i} and a variance σ_{noise_i}, related to the magnitude of random noise in this particular TERO configuration. N_{m_i} varies from one chip to another due to MPV: N_{m_i} is itself a draw of a random variable $\mathcal{R}$ with a mean value N_m, and a variance σ_{MPV}, related to MPV parameters of the selected technology. In practice, σ_{noise_i} may also depend on N_{m_i} because the longer the TEROs oscillate, the higher the variability of their output (random noise accumulates over time). This means that responses generated in different TEROs may have different noise stability. However, in general, the higher σ_{MPV}, the better the overall uniqueness of the PUF (its ability to generate unique responses); and the lower the magnitude of random noise in the selected technology, the better the overall steadiness of the PUF (its ability to generate responses that do not vary over time).

Based on the assumption of a Gaussian model for the distribution of propagation delays of inverters, the expected theoretical distribution of $\mathcal{R}$ and $\mathcal{R}_i$ is plotted in Fig. 7 supposing an 8-bit subtractor and Gaussian distributions of its output. In this figure, we also represented the ones (gray) and zeros (white) for the bit 2^5 (*i.e.* the 5^{th} bit starting from the LSB) of the subtractor output coded in Gray code. A look at this figure provides a quick evaluation of the sensitivity to noise and to MPV of responses generated using this particular bit. By estimating the sum of the gray areas delimited by R, we obtain the probability of bit flips between different chips due to MPV. In the same way, by estimating the sum of the gray areas delimited by R_i for a particular pair of TEROs, we obtain the probability of bit flips due to noise. It is clear

from Fig. 7 that data coding has an important influence on the uniqueness and steadiness of the responses.

In general, LSB subtractor bits are more sensitive to MPV than MSB bits, but this is also true for random noise variations [12]. However, sensitivity to random noise can be considerably reduced by using a noise filter, by acquiring several samples from the subtractor output over time, and by computing a mean value. Nonetheless, this does not improve the robustness of the responses to environmental variations, which cannot be filtered in the same way because they are not necessarily random and centered around a nominal value. Therefore, filtering noise is generally not helpful since it only enhances the steadiness in nominal voltage and temperatures conditions.

In most practical cases, $\sigma_{noise_i} < \sigma_{MPV}$. In recent technologies, the noise related standard deviation for propagation delays is of the order of one thousandth, whereas the MPV related standard deviation for propagation delays is of the order of one hundredth. This means that several subtractor bits may be affected by MPV merely because they are sensitive to noise, meaning the noise filtering process is not mandatory (which reduces the size of the design and increases its throughput).

B. Response Generation Schemes

The most simple and straightforward response generation scheme consists in selecting the sign bit (MSB) in the subtractor as a response for the corresponding challenge. This method makes it possible to maximize robustness to voltage and temperature: if TERO 1 oscillates longer than TERO 2 in certain temperature and voltage conditions, there is a high probability that is will continue to oscillate longer in other conditions (the sign bit is less likely to be affected by such variations). The main drawback here is that only a one bit response can be obtained per challenge, which maximizes the size of the design. The approach is then similar to that consisting in counting two RO outputs during a fixed time period and comparing them. However, in TEROs, the number of oscillations does not only depend on their oscillation frequency but also on the capacitive charge and discharge paramaters at the output of each cell (which may also be affected by MPV). In other words, we still benefit from a design that is more difficult to clone and model because obtaining the parameters (for instance, via electro-magnetic characterization) is harder.

Responses can be generated using two or more bits from the subtractor's output, as proposed in [12]. Using this approach may affect the performances of the PUF (especially in terms of randomness), as it will be shown in Section VI. However, the size of the design is significantly reduced since each challenge makes it possible to generate more response bits. The subtractor bits used to generate the responses must be carefully chosen with respect to the noise and MPV parameters of the selected CMOS technology and the parameters of the TERO, which determine the nominal number of oscillations in the implemented design.

For the purpose of our evaluation, we generated responses using the binary code of the subtractor output, or its Gray code. The Gray code has two useful properties: the hamming distance between two successive symbols is always equal to one (which may have the advantage of reducing the responses' sensitivity to noise) and a distinct sign bit (contrary to binary codes in which each bit can become a sign extension). Having a distinct sign bit is a necessary condition to generate more than 1-bit responses as it will be shown in section VI-D. Note that the Gray coder can be easily implemented by adding bitwise (XOR) the subtractor output bits and does not add much hardware to the design.

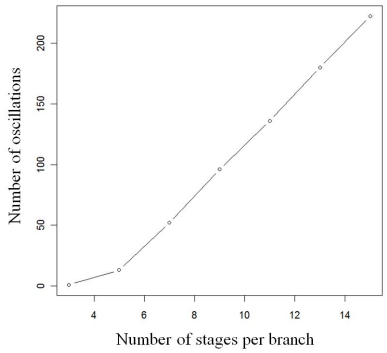

Fig. 8. Nominal number of oscillations of a TERO as a function of its number of stages (CMOS 350 nm, electrical simulations in *Cadence* environment).

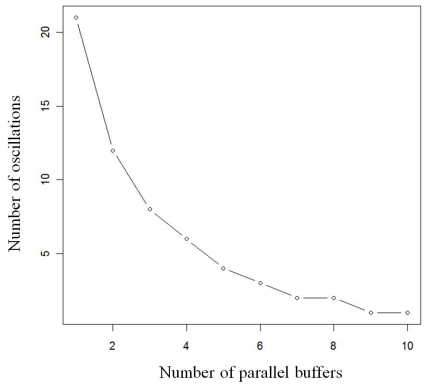

Fig. 9. Nominal number of oscillations of a 10-stage TERO as a function of its output capacitance represented by the number of output parallel buffers (CMOS 350 nm, electrical simulations in *Cadence* environment).

In Section VI, these response generation schemes (*i.e.* using the Gray or binary code, and one or more subtractor bits in the response generation) are analyzed and compared in terms of PUF performances (uniqueness, steadiness and randomness) as well as the area of the PUF.

V. Design of TERO-PUF in a 350 nm CMOS Technology and Raw Measurements

This section provides an overview of the design process of a TERO-PUF in a standard CMOS technology.

A. Development and Design of the Test Chips

The two most important factors to take into account when designing TERO-PUFs are the number of stages in each TERO and the capacitive charge and discharge parameters of each ring stage. The higher the number of stages, and the lower the output parasitic capacitance per ring stage, the higher the nominal number of oscillations N_m, as shown in Fig. 8 and Fig. 9. In fact, a higher output capacitance means

a stronger drafting effect, and hence a shorter time before the two events collide, based on the analysis in Section III-C. Other relevant factors are falling/rising propagation delays, and the dissymmetry between the two branches (to take full advantage of MPV, they should be identical). The more symmetrical they are, the higher the number of transient oscillations. N_m can be set up using electrical simulation (preferably post-layout with parasitic capacitance extraction).

The choice of N_m usually enables a trade-off between overall uniqueness and steadiness. In fact, choosing a too high N_m allows high variability (due to MPV) between identically implemented TEROs, but also increases the variability due to noise in each TERO (jitter variations are additive in time). The size of the counters and the subtractor need to be selected as a function of N_m.

The setup of our design was based on electrical simulations. We varied the output capacitance of the logic gates by modifying the size of transistors and we selected the number of ring stages in order to control the number of transient oscillations in electrical simulations (using *Cadence* environment). We used 8-bit counters at the output of the TEROs, and therefore selected the TERO parameters in such a way as to obtain 128 transient oscillations (for the convenience of data analysis). The corresponding configuration of the TEROs is 8 stages per branch and one output buffer per branch. The base width of the transistors is 0.6 μm for the NMOS and 0.9 μm for the PMOS.

The careful layout of the ring stages is critical to obtain symmetrical branches in the TEROs. This cannot be left to automated routing placement and routing tools but should be performed manually. Each TERO uses 36 transistors. Symmetry is obtained by designing one branch of a TERO, and then using it as a pattern to build all the other blocks.

We implemented the architecture depicted in Fig. 3. It consists of two blocks of 128 TEROs per block along with the selection and initialization circuitery and two 8-bit counters. For our experimentation and evaluation purposes, we did not implement the 9-bit subtractor, but rather acquired data directly from the outputs of the counters in order to test the different generation schemes proposed in Section IV. Each test chip also contains a control module which provides all the credentials for connectivity and rapid data transfer. We manufactured and packaged 30 test chips with 256 TEROs per chip giving a total of 7680 TEROs for our characterization.

B. Raw Measurements of the Outputs of the Counters

Prior to the statistical evaluation, we measured the counter outputs directly in order to plot the distribution of the number of transient oscillations in each TERO ($\mathcal{R}_i$), and the distribution of the mean number of oscillations from different TEROs ($\mathcal{R}$). The target is a population of 256 TEROs from one test chip. For each TERO, 960 response samples were acquired. We measured the mean number of oscillations for each TERO using the 960 response samples. We first plotted the probability distribution of the mean number of oscillations from this population of 256 TEROs: this distribution is due to MPV (noise is filtered). It has a maximum density of around 30 oscillations. The mean value is $N_m = 46$ and the standard deviation approaches $\sigma_{MPV} = 39$. N_m is relatively low compared to the nominal value in electrical simulations (128). In fact, 128 corresponds to the nominal number of oscillations when the two branches of the TERO structure are strictly identical (it actually corresponds to a maximum value). Fig. 10 shows the distribution of the mean differential number of oscillations for a population of 256 TEROs. The resulting distribution is relatively symmetrical and centered around zero, suggesting that the two blocks of 128 TEROs are statistically identical.

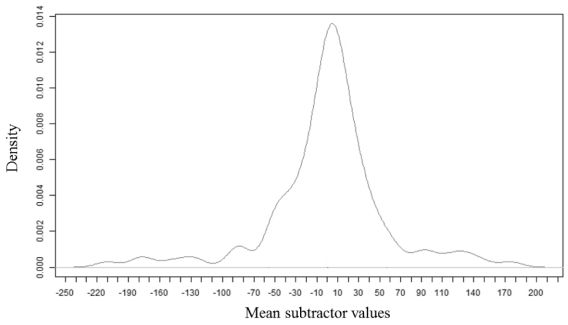

Fig. 10. Measured distribution of the mean subtractor output values from a population of 256 TEROs in one device.

We plotted the effective number of oscillations for three TEROs using a population of 960 samples for each TERO. The effective number of oscillations of TERO i is described by a distribution $\mathcal{R}_i$: here, variability is due to the intrinsic noise in each TERO. For the first TERO, $N_{m_1} = 156$ and $\sigma_{noise_1} = 34$. For the second TERO, $N_{m_2} = 86$ and $\sigma_{noise_2} = 5$. And for the last TERO, $N_{m_3} = 25$ and $\sigma_{noise_3} = 0$ (noise does not affect this TERO at all). As we can see, the lower the mean number of oscillation of a TERO N_{m_i}, the lower the standard deviation (noise related variability) of its effective number of oscillations σ_{noise_i}. This is because noise variations accumulate over time causing an increase in the standard deviation for TEROs which oscillate longer.

In practice, it is possible to roughly set the mean number of oscillations N_m for the whole population of TEROs by carefully selecting the number of ring stages and the sizes of the transistors in each cell. However, it is not possible to determine the mean number of oscillations of one TERO N_{m_i} in advance, since MPV is unpredictable, and it is therefore not possible to determine in advance if a particular TERO is more or less reliable than other TEROs.

VI. Evaluation of TERO-PUFs in a CMOS 350 nm Technology

This section describes the characterization and the results of the evaluation of the 30 test chips designed in a 350 nm CMOS technology.

A. The Evaluation Approach

The characterization of a PUF must at least consider three factors: its uniqueness, its steadiness (also called reliability or robustness, computed in nominal and corner temperatures

and voltage conditions) and finally its randomness. Let us consider a set of N devices denoted $(d_i)_{1\leq i\leq N}$. n-bit responses are extracted L times from the N different chips. We denote $r_{i_j}(p)$ (with $1 \leq i \leq N$ and $1 \leq j \leq L$) the j^{th} response of device d_i to the challenge c_p.

1) Evaluating Uniqueness: For a given challenge c_p, the responses from two devices d_i and d_j must differ with high probability ($r_{i_l}(p) \neq r_{j_m}(p)$ with $1 \leq l, m \leq L$). One common indicator of this characteristic is the inter-chip hamming distance, which is computed using the following equation:

$$EC = \frac{1}{N(N-1)L}\sum_{i=1}^{N}\sum_{k=1,k\neq i}^{N}\sum_{j=1}^{L}\frac{HD\left(r_{i_j}(p), \bar{r_k}(p)\right)}{n} \times 100\% \tag{2}$$

where $r_{i_j}(p)$ is the j-th response sample from the PUF i, $\bar{r_k}(p)$ is the mean value of the L response samples from the PUF k, n is the size of the response vectors, and HD is the hamming distance between the two vectors. The optimal value for this indicator is 50%. In the following, we refer to this indicator simply as uniqueness. In a practical application of the PUF, uniqueness will determine the size of the identifiers needed to distinguish a certain number of devices.

2) Evaluating Steadiness: For a given challenge c_p repeated several times, the responses of one device d_i should be always the same ($r_{i_j}(p)$ does not vary over time). A common indicator used for steadiness (also called reliability or robustness) is the intra-chip hamming distance which is computed, for a PUF i, at the temperature T and power supply voltage V, using the following equation:

$$IC_i(T,V) = \frac{1}{L}\sum_{j=1}^{L}\frac{HD(r_{i_j}(p), r_{ref}(p))}{n} \times 100\% \tag{3}$$

$r_{ref}(p)$ is a reference response (associated with the challenge c_p and the device d_i) obtained in nominal voltage V_n and temperature T_n conditions. The optimal value of this indicator is 0%. In the following, we refer to this indicator simply as steadiness. Low steadiness values mean that the PUF is reliable. In a practical application of the PUF, its steadiness determines the cost of error correction needed to obtain reliable identifiers and/or the voltage and temperature ranges for the proper fonctioning of the PUF. For our setup, $V_n = 3.3$ V and $T_n = 23$ °C. We evaluated steadiness between −20 °C and 70 °C, and at power supply voltages within the limit of the correct functioning of the device between 3 V and 3.6 V.

3) Evaluating Randomness: For a given challenge c_p, the responses $r_{i_j}(p)$ (with $1 \leq j \leq L$) should be unpredictable and uniformly distributed. In a practical application of the PUF, randomness determines the vulnerability of the PUF to brute force and modeling attacks. Evaluating randomness can be challenging for PUFs because a large amount of data (and therefore a large number of test chips) is required. However, a few statistical tests from NIST SP 800-22 [14] can be adapted to a small amount of data, although these tests will have a low confidence level. The main objective of this evaluation is to rapidly eliminate responses which are not random (proving thoroughly that the sequences are random requires a very large set of data). On the other hand, we want to evaluate randomness, which is exclusively caused by MPV. Therefore, we compute a mean 128-bit response for each PUF using 960 samples. The 30 noise filtered 128-bit responses are then merged to obtain a 3840-bit sequence, which is tested using 6 statistical tests (based on the minimal input sequence length parameters recommended in [14]). We denote T1 the frequency test, which measures the bias of the sequence. T2 is the frequency within a block test, which we apply for 2-bit, 3-bit and 4-bit block lengths. T3 is the cumulative sums test. T4 is the runs tests (which evaluates the distribution of sequences composed of successive ones). T5 is the longest run of ones test. And finally, T6 is the approximate entropy test which evaluates the distribution of overlapping M-blocks in the sequences (we use this test with M = 2, M = 3 and M = 4). These tests are applied to series of 10 sequences, which corresponds to a 0.1 confidence level. Pass rates are given in the NIST software STS 2.1.1 available on the NIST government website.[1] T4 and T6 are particularly suitable for our study in which several bits of the subtractor value are used to build the responses. For example, in the binary coded output, a few bits can serve as a sign extension in some pair of TEROs but can have significant value in others: using them along with the sign bit to build the responses can lead to direct correlation in blocks of two successive bits (which can be easily detected using T4 and T6).

In addition to the statistical tests, we estimate the overall bias of the responses using the Shannon entropy formula applied to one-bit vectors. Here again, responses are filtered from noise in order to estimate bit flips which are due to MPV and not to random noise. We compute H using the following formula:

$$\bar{H} = \frac{1}{n \times N}\sum_{k=1}^{n}\sum_{i=1}^{N} -p_{k,i}log_2(p_{k,i}) - (1-p_{k,i})log_2(1-p_{k,i}) \tag{4}$$

where $p_{k,i}$ is the probability that $\bar{r_k}(p) = \frac{1}{L}\sum_{j=1}^{L} r_{k,j}(p)$ is equal to 1. The optimal value of this indicator is 1.

Note that most previous works on PUFs only compute the bias for this randomness evaluation (which is equivalent to the above Shannon entropy applied to 1-bit vectors). This approach is clearly not sufficient to detect correlations between the generated bits as will be shown in Section VI-D.

B. Rapid Evaluation of One Test Chip

For this rapid evaluation, each pair of TEROs in one chip is considered as a PUF which generates a 1-bit response using one of the subtractor output bits ($n = 1$). A total of 960 samples were generated for each of the 128 couples of TEROs ($L = 960$ and $N = 128$). We evaluated the responses in terms of uniqueness (inter-chip hamming distance EC using Eq. 2), steadiness (intra–chip hamming distance IC using Eq. 3) and entropy (H using Eq. 4) with $T = 23$ °C and a 3.3 V power supply. Since each pair of TEROs has its own

[1] http://csrc.nist.gov/groups/ST/toolkit/rng/index.html

TABLE II

UNIQUENESS (EXTRA-CHIP VARIATION EC), STEADINESS (INTRA-CHIP VARIATION IC) AND RANDOMNESS (ENTROPY H) OF EACH BIT OF THE SUBSTRACTOR OUTPUT FOR A SET OF 256 TEROS WITH 960 SAMPLE PER RESPONSE

Bit nb.	EC (%)	IC_{mean} (%)	σ_{IC} (%)	H
0	50.11	30.21	18.46	0.99
1	50.17	28.90	18.13	0.99
2	48.92	20.46	19.20	0.96
3	50.19	11.82	17.51	0.99
4	50.08	6.90	13.81	0.99
5	49.52	3.82	9.82	0.98
6	50.36	2.44	8.21	0.99
7	50.28	1.30	6.12	0.99
8	50.28	0.06	0.48	0.99

TABLE III

UNIQUENESS (EXTRA-CHIP VARIATION EC), STEADINESS (INTRA-CHIP VARIATION IC) AND RANDOMNESS (ENTROPY H) OF EACH BIT OF THE SUBTRACTOR OUTPUT CODED IN GRAY CODE FOR A SET OF 256 TEROS WITH 960 SAMPLES PER TERO

Bit nb.	EC (%)	IC_{mean} (%)	σ_{IC} (%)	H
0	49.37	27.71	18.45	0.97
1	50.00	19.74	18.42	0.99
2	50.31	15.15	17.85	0.99
3	50.12	8.36	14.91	0.99
4	48.01	3.30	8.72	0.96
5	40.38	1.37	5.95	0.85
6	24.83	1.14	5.73	0.59
7	6.33	1.24	6.11	0.20
8	50.28	0.06	0.48	0.99

steadiness value, we computed the mean value IC_{mean} and the standard deviation σ_{IC}. The results are presented in Table II and Table III. In Table II, the subtractor output is binary coded, whereas it is Gray coded in Table III.

As it can be seen in Table II, the responses are unique ($EC \simeq 50\%$) and unbiased in all the configurations tested. This is due to the combination of two factors: first, the binary code of a positive integer and its negative value have opposite digits (except for the LSB); and second, the distribution of the mean differential number of oscillations is roughly symmetrical and centered around zero, as shown in Fig. 10. Considering this probability distribution and the binary code, each bit in the subtractor output can potentially be used as a sign extension for a pair of TEROs. Therefore, whatever the bit used to build the response and for a given response value, the opposite value is equally probable. Contrary to the binary code, in the Gray code, the sign information is only contained in the MSB. As suggested in part (b) in Fig. 7, the distribution of zeros and ones is not centered around zero for Gray coded digits other than the MSB. For example, by using the bit 2^5 with the Gray code, the PUF will generate more ones than zeros if the distribution of differential oscillations is the one shown in part (b) in Fig. 7 (the white area under the distribution $\mathcal{R}$ is larger than the gray area).

As expected in theory, overall steadiness increases from the LSB to the MSB with better results with the Gray code.

TABLE IV

COLOR CODE FOR TABLE VI, TABLE VII AND TABLE VIII

Quality Metrics	Uniqueness	Steadiness	Randomness
Very good	$\mathbf{49\% \leq EC}$	$\mathbf{IC \leq 5\%}$	$\mathbf{0.99 \leq H}$
Average	$45\% \leq EC \leq 49\%$	$5\% \leq IC \leq 10\%$	$0.90 \leq H \leq 0.99$
Insufficient	$EC \leq 45\%$	$10\% \leq IC$	$H \leq 0.90$

TABLE V

COMPARISON OF TERO-PUF AND RO-PUF IMPLEMENTATIONS

PUF	TERO-PUF [12]	TERO-PUF (this work)	RO-PUF [8]	RO-PUF [15]
Implementation	Altera Cyclone II (90 nm)	CMOS 350 nm	Xilinx Spartan3E (90 nm)	CMOS 65 nm
Uniqueness	48%	49.7%	47.3%	49.5%
Steadiness	1.7%	0.6%	0.9%	2.8%
Steadiness (T.V. corners)	not reported	6.2%	15%	3.9%
Basic blocks	2 NANDs and 2 inverters	2 NANDs and 14 inverters	1 NAND and 4 inverters	1 NAND and 40 inverters

High steadiness values are mainly caused by TEROs which have a high mean number of oscillations, and which are therefore highly sensitive to noise. Nonetheless, for this configuration, the sign bit (MSB) is very reliable ($IC_{mean} = 0.06\%$).

C. Evaluation of Responses Generated Using One Subtractor Bit

This section details the evaluation of 128-bit responses generated using one subtractor bit in 30 test chips. Each 128-bit response is obtained using the set of challenges $\{C(i,i)\}_{0 \leq i \leq 127}$, *i.e.* by successively selecting TERO i from block A and TERO i from block B.

1) Uniqueness and Steadiness: We generated 960 samples of 128-bit response vectors from each of the 30 test chips at 23 °C and a 3.3 V power supply ($n = 128$, $N = 30$ and $L = 960$). Responses were generated by selecting the same subtractor output bit from 128 couples of TEROs in each chip. We first tested their uniqueness (inter-chip hamming distance) and steadiness (intra-chip hamming distance). For binary and Gray codes, results are given in the third and fourth column in Table VI. $EC(23\ °C, 3.3V)$ is the computed inter-chip hamming distance value (uniqueness), and $IC_m(23\ °C, 3.3V)$ is the mean intra-chip hamming distance value (steadiness) for the 30 test chips. To be able to quickly identify the relevant results in Table VI, we created a color code (white, light gray and dark gray) which is depicted in Table IV the darker the corresponding color for each indicator, the better the PUF output quality.

Results are similar to those shown in Table II and Table III. Responses built using the MSB are unique ($EC = 49.72\%$)

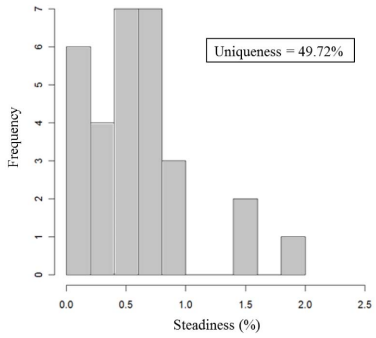

Fig. 11. Distribution of steadiness values for responses generated using the MSB subtractor bit at 23 °C and 3.3 V.

and steady ($IC_m = 0.61\%$). In this particular configuration (*i.e.* using one subtractor bit), the Gray code has no impact since the MSB (which is the same in both the binary code and the Gray code) is systematically preferred to build the response. A few chips have a steadiness value approaching 2% as it is shown in Fig. 11, but most have a steadiness value lower than 1%.

2) Randomness: The result of the evaluation of the randomness of PUF responses are listed in the right hand columns in Table VI. Whatever the subtractor bit used in the binary code, the responses have an almost null bias and pass the selected test battery. This is due to the symmetrical distribution of the mean differential number of oscillations, as previously explained with respect to uniqueness. With the Gray code, only a few configurations pass the test battery. Some have an important bias which is detected by both the indicator $\bar{H}$ and the frequency test (T1). The failure of the majority of other tests is inherent to this marked bias. Once again, the prefered configuration is the one using the MSB (bit 8). Hence, using the Gray code is not appropriate in this situation (only one bit is extracted).

3) Robustness to Voltage and Temperature Variations: We elaborated the PUF responses steadiness in a −20 °C to −70 °C temperature range for a randomly selected test chip. In Table VI, $IC_6(23\ °C, 3.3V)$ is the intra-chip hamming distance for the selected chip at ambient temperature, $IC_{max}(0\ °C$ to $50\ °C)$ is the maximum steadiness in the 0 °C to 50 °C temperature range, $IC_{max}(-20\ °C$ to $70\ °C)$ is the maximum steadiness in the −20 °C to 70 °C temperature range. The reference response is generated at 25 °C. As can be seen in Table VI, most of the tested configurations do not fair well with temperature variations. However, the configuration using the MSB gives maximum steadiness values which remain within good limits: 3.75% in the 0 °C to 50 °C range and 6.21% in the −20 °C to 70 °C temperature range. Fig. 12 shows the steadiness at different temperatures for this configuration.

We evaluated the steadiness of the PUF responses within the power supply voltage limits for a proper functioning of the circuit (between 3.0 V and 3.6 V). In Table VI, $IC_{max}(3.3V \pm 0.1V)$ is the maximum steadiness in the 3.2 V to 3.4 V voltage range, $IC_{max}(3.3V \pm 0.3V)$ is the maximum steadiness in the 3.0 V to 3.6 V voltage range. The reference response is generated at 3.3 V. Table VI shows that the TERO-PUF is not significantly affected by variations in voltage in the voltage range tested. The configuration using the MSB has very good maximum steadiness values: 0.96% in the 3.2 V to 3.4 V voltage range and 2.82% in the 3.0 V to 3.6 V range. Fig. 12 details the steadiness at each voltage for this configuration.

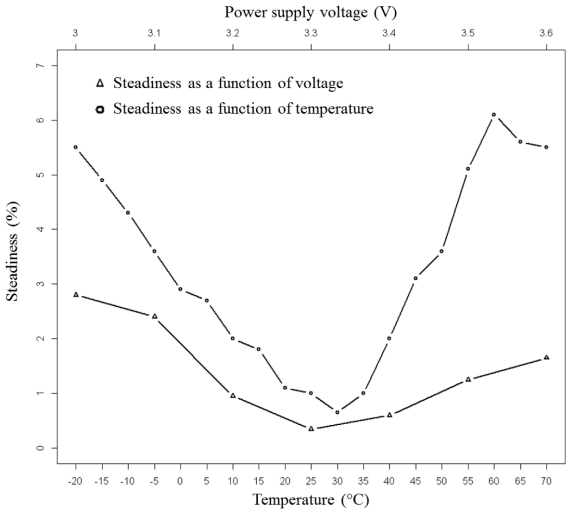

Fig. 12. Steadiness of responses generated using the MSB subtractor bit in a −20 °C to 70 °C temperature range (reference response at 25 °C) and in a 3.0 V to 3.6 V power supply voltage range (reference response at 3.3 V).

D. Evaluation of Responses Generated Using Two or More Subtractor Bits

For 2-bit (respectively 3-bit) response per challenge configurations, each 128-bit PUF response is obtained using the set of challenges $\{C(i,i)\}_{0\leq i\leq 63}$ (respectively $\{C(i,i)\}_{0\leq i\leq 42}$). When experimenting with responses built using several subtractor bits, we identified two main problems that were not reported in [12]: the uniqueness and the steadiness of the responses generated using this approach were very good in nominal conditions, but in corner conditions, their robustness and their randomness dropped drastically. In fact, in the binary code, each subtractor bit may (with a certain probability) serve as a sign extension for a number of TERO pairs, which results in important correlations when building responses using two bits ("00" and "11" sequences are more probable than "01" and "10"). However, here we show that this can be mitigated by using of the Gray code, which restrains the sign information in the MSB.

We evaluated responses generated using several combinations of two bits of the subtractor output. Those bits are carefully selected based on the characterization in Table VI (subtractor bits with low performances are not used for the construction of the 2-bit response). Table VII lists the results of the evaluation for a few configurations which are relevant for our analysis. As it can be seen in the table, the two main issues are the randomness and steadiness of the responses in corner voltage and temperature conditions. Using two bits in the binary code based on their individual performances can have disastrous repercussions on the responses' randomness.

TABLE VI

EVALUATION OF THE UNIQUENESS, THE STEADINESS AND RANDOMNESS OF 128-BIT RESPONSES GENERATED USING ONE SUBTRACTOR BIT IN 30 TEST CHIPS

Subtractor output	Configuration	Uniqueness	Steadiness				Randomness						
		$EC(23°C, 3.3V)$	$IC_m(23°C, 3.3V)$	$IC_6(23°C, 3.3V)$	$IC_{max}(0° \text{ to } 50°C)$	$IC_{max}(-20° \text{ to } 70°C)$	$IC_{max}(3.3V \pm 0.1V)$	$IC_{max}(3.3V \pm 0.3V)$	H	T1	T2	T3	T4	T5	T6
Binary	**Bit 0**	**49.06 %**	26.21 %	28.64 %	44.39 %	47.76 %	40.22 %	46.63 %	**0.99**	√	√	√	√	√	√
	Bit 1	**50.10 %**	22.07 %	27.22 %	49.43 %	51.15 %	40.96 %	47.50 %	**0.99**	√	√	√	√	√	√
	Bit 2	**49.68 %**	12.11 %	18.71 %	39.80 %	46.90 %	30.20 %	44.19 %	**0.99**	√	√	√	√	√	√
	Bit 3	**49.68 %**	5.98 %	11.51 %	33.30 %	40.29 %	22.12 %	32.15 %	**0.99**	√	√	√	√	√	√
	Bit 4	**50.07 %**	**3.06 %**	7.94 %	29.73 %	33.18 %	15.83 %	26.75 %	**0.99**	√	√	√	√	√	√
	Bit 5	**49.70 %**	**1.55 %**	**4.98 %**	18.25 %	25.95 %	10.43 %	15.98 %	**0.99**	√	√	√	√	√	√
	Bit 6	**49.68 %**	**0.78 %**	**1.91 %**	12.03 %	14.96 %	8.55 %	14.07 %	**0.99**	√	√	√	√	√	√
	Bit 7	**49.72 %**	**0.61 %**	**1.05 %**	7.34 %	11.49 %	5.42 %	7.03 %	**0.99**	√	√	√	√	√	√
	Bit 8	**49.72 %**	**0.61 %**	**1.05 %**	**3.75 %**	6.21 %	**0.96 %**	**2.82 %**	**0.99**	√	√	√	√	√	√
Gray	**Bit 0**	**50.05 %**	21.68 %	27.56 %	46.85 %	49.14 %	40.18 %	45.62 %	**0.99**	√	√	√	√	√	√
	Bit 1	**49.89 %**	12.40 %	19.47 %	42.76 %	43.72 %	31.95 %	39.85 %	**0.99**	√	√	√	√	√	√
	Bit 2	**50.25 %**	6.20 %	13.73 %	31.40 %	41.73 %	27.58 %	32.24 %	**0.99**	√	√	√	√	√	√
	Bit 3	**49.86 %**	**2.92 %**	6.70 %	26.00 %	31.12 %	17.90 %	23.80 %	**0.99**	√	×	√	√	√	√
	Bit 4	47.01 %	**1.52 %**	**3.80 %**	19.86 %	21.35 %	14.70 %	24.88 %	0.96	×	×	×	×	×	×
	Bit 5	39.75 %	**0.77 %**	**2.13 %**	13.78 %	20.53 %	7.52 %	13.10 %	0.84	×	×	×	×	×	×
	Bit 6	3.80 %	**0.17 %**	**0.94 %**	**4.74 %**	6.19 %	6.50 %	10.02 %	0.13	×	×	×	×	×	×
	Bit 7	0.00 %	**0.00 %**	**0.00 %**	**4.78 %**	6.02 %	**4.56 %**	5.24 %	0.00	×	×	×	×	×	×
	Bit 8	**49.72 %**	**0.61 %**	**1.05 %**	**3.75 %**	6.21 %	**0.96 %**	**2.82 %**	**0.99**	√	√	√	√	√	√

TABLE VII

EVALUATION OF THE UNIQUENESS, THE STEADINESS AND RANDOMNESS OF 128-BIT RESPONSES GENERATED USING TWO SUBTRACTOR BIT IN 30 TEST CHIPS

Subtractor output	Configuration		Uniqueness	Steadiness				Randomness						
	Bit 1	Bit 2	$EC(23°C, 3.3V)$	$IC_m(23°C, 3.3V)$	$IC_6(23°C, 3.3V)$	$IC_{max}(0° \text{ to } 50°C)$	$IC_{max}(-20° \text{ to } 70°C)$	$IC_{max}(3.3V \pm 0.1V)$	$IC_{max}(3.3V \pm 0.3V)$	H	T1	T2	T3	T4	T5	T6
Binary	8	6	**50.06 %**	**0.58 %**	**1.83 %**	8.21 %	9.04 %	**4.12 %**	6.26 %	**0.99**	√	×	√	×	×	×
	8	3	**49.88 %**	**3.35 %**	7.29 %	17.26 %	23.66 %	10.96 %	16.27 %	**0.99**	√	×	√	×	√	√
	8	2	**49.90 %**	6.41 %	10.53 %	23.44 %	24.22 %	13.87 %	22.07 %	**0.99**	√	√	√	√	√	√
	6	3	**49.92 %**	**3.41 %**	8.55 %	22.81 %	26.00 %	14.07 %	21.43 %	**0.99**	√	×	√	×	√	√
	6	2	**49.94 %**	6.48 %	11.78 %	26.39 %	29.50 %	17.26 %	27.24 %	**0.99**	√	√	√	√	√	√
	5	3	**49.73 %**	**3.83 %**	9.85 %	24.47 %	33.75 %	16.05 %	23.30 %	**0.99**	√	√	×	√	√	√
	4	3	**50.05 %**	**4.61 %**	11.89 %	30.46 %	36.76 %	17.47 %	27.29 %	**0.99**	√	√	√	√	√	√
Gray	8	5	44.84 %	**0.67 %**	**1.58 %**	9.99 %	14.68 %	**4.93 %**	7.61 %	0.97	×	√	×	×	×	×
	8	4	48.26 %	**1.04 %**	**2.80 %**	14.42 %	15.19 %	5.59 %	11.03 %	**0.99**	√	√	√	√	√	√
	8	3	**49.86 %**	**1.82 %**	**4.45 %**	14.09 %	16.90 %	8.04 %	13.27 %	**0.99**	√	√	√	√	√	√
	5	3	44.67 %	**1.98 %**	5.46 %	19.03 %	26.22 %	11.97 %	19.78 %	0.94	×	√	×	×	×	×
	4	2	48.21 %	**3.89 %**	10.04 %	28.35 %	30.97 %	17.85 %	26.51 %	**0.99**	√	×	√	√	√	×
	4	1	48.26 %	7.03 %	14.35 %	34.36 %	35.54 %	20.67 %	30.55 %	**0.99**	×	√	×	√	√	×
	3	2	**49.82 %**	**4.68 %**	11.69 %	29.63 %	34.91 %	18.22 %	28.75 %	**0.99**	√	√	√	√	√	√

TABLE VIII

EVALUATION OF THE UNIQUENESS, THE STEADINESS AND RANDOMNESS OF 128-BIT RESPONSES GENERATED USING THREE SUBTRACTOR BIT IN 30 TEST CHIPS

Subtractor output	Configuration			Uniqueness	Steadiness					Randomness						
	Bit 1	Bit 2	Bit 3	$EC(23°C, 3.3V)$	$IC_m(23°C, 3.3V)$	$IC_6(23°C, 3.3V)$	$IC_{max}(0° \textit{ to } 50°C)$	$IC_{max}(-20° \textit{ to } 70°C)$	$IC_{max}(3.3V \pm 0.1V)$	$IC_{max}(3.3V \pm 0.3V)$	H	T1	T2	T3	T4	T5	T6
Binary	8	6	5	**49.88 %**	**0.85 %**	**3.28 %**	12.90 %	15.54 %	7.71 %	9.44 %	**0.99**	√	×	√	×	×	×
	8	6	2	**49.96 %**	**4.44 %**	8.17 %	19.22 %	20.84 %	12.46 %	17.72 %	**0.99**	√	×	√	×	×	×
	8	5	3	**49.89 %**	**2.67 %**	6.27 %	17.74 %	24.61 %	12.34 %	16.07 %	**0.99**	√	×	√	×	×	×
	6	5	3	**49.88 %**	**2.71 %**	7.40 %	23.81 %	27.94 %	15.30 %	19.01 %	**0.99**	√	×	√	×	×	×
Gray	8	5	4	46.15 %	**0.99 %**	**2.30 %**	14.76 %	17.44 %	7.06 %	11.78 %	0.97	×	√	×	√	√	×
	8	5	3	47.05 %	**1.45 %**	**3.26 %**	15.10 %	18.36 %	9.00 %	12.73 %	0.98	×	√	×	√	√	√
	8	4	3	**49.51 %**	**1.73 %**	**4.66 %**	16.08 %	18.87 %	9.27 %	14.48 %	**0.99**	√	√	√	√	√	√
	5	4	3	46.30 %	**1.82 %**	**4.75 %**	20.20 %	24.78 %	11.93 %	19.01 %	0.95	×	×	×	×	×	×

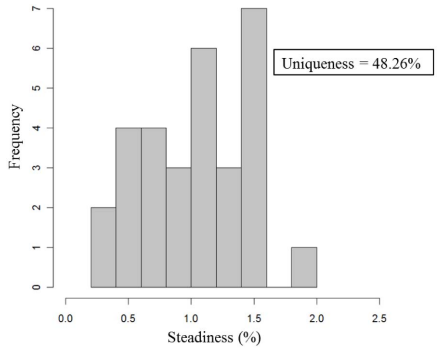

Fig. 13. Distribution of steadiness values for responses generated using bits 8 and 4 in the Gray code at 23 °C and 3.3 V.

For example using bits 8 and 6 (both of whose 1-bit configurations pass the individual statistical tests, as shown in Table VI), the responses are no longer random even though they are generaly unbiased. These statistical defects are easy to detect with the runs test (T4) and the approximate entropy test (T6). Conversely, when building responses using two bits in the Gray representation, the main issue is the overall bias, which can be detected individually in each 1-bit configuration in Table VI.

Based on these measurements, the 2-bit configuration (*i.e.* using two of the subtractor outputs) with the best performance trade-off is the one using the 8 and 4 in the Gray coded output. This configuration passes the randomness test batteries and displays a good uniqueness (48.26%) and steadiness (1.04%) in nominal temperature and voltage conditions. The distribution of steadiness values for this configurations is shown in Fig. 13. However, its main drawback is its low robustness to temperature variations (maximum intra-chip variation 15.19%), although it can be used within limited voltage and temperature ranges with adequate error correction. To guarantee a maximum steadiness lower than 10%, the temperature must be between −15 °C and 45 °C, and voltage must be between 3.1 V and 3.5 V.

Based on the above results, 3-bit configurations are carefully selected and evaluated using the same approach. These results are summarized in Table VIII. None of the configurations tested in the binary code passed the randomness tests. The robustness of the responses was also strongly reduced. However, one configuration using the Gray code (bits 8, 4 and 3) offered a satisfying trade-off. It passes all the selected randomness tests and has a uniqueness of 49.51% and a steadiness of 1.73% in nominal voltage and temperature conditions. The distribution of steadiness values for this configurations is shown in Fig. 14. For this configuration, maximum steadiness is 15.50% in the 0 °C to 50 °C temperature range. To guarantee maximum steadiness lower than 10%, the temperature must be between 15 °C and 40 °C, and voltage must be between 3.2 V and 3.4 V.

In conclusion, when generating responses using one subtractor bit, the MSB can be used systematically whereas the Gray code is not appropriate. However, when building responses using several subtractor bits, using the Gray code improves significantly the PUF randomness (with a very low

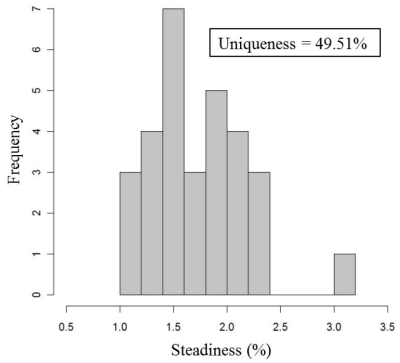

Fig. 14. Distribution of steadiness values for responses generated using bits 8, 4 and 3 in the Gray representation at 23 °C and 3.3 V.

design effort) because the Gray code makes it possible to reduce correlations within the generated blocks by isolating the sign information in the MSB (which can be therefore safely used along with other bits without direct correlations). By generating more bits per response, the area of the design is significantly reduced, but at the cost of a lower robustness to environmental variations, especially temperature.

E. Comparison With RO-PUFs

Both RO-PUFs and TERO-PUFs size may vary depending on the targeted PUF performances for a given technology (which has its own noise and MPV parameters), which makes comparisons between the published works difficult. Nonetheless, in this section, we compare the above results with evaluation results of two RO-PUF implentations found in the literature, although results must be interpreted cautiously. To make this comparison relevant, we only compared the TERO-PUF configuration with a one bit response per challenge with RO-PUFs which also generate a one bit response per challenge. This comparison is given in Table V.

First, we remark that for both PUFs, the ASIC implementations require larger oscillators than the FPGA implementations while having similar PUF performances in nominal T.V. (temperature and voltage) conditions. The ASIC implementation of the RO-PUF uses more than twice basic cells than the TERO-PUF's ASIC implementation, while having slightly lower PUF performances in nominal T.V. conditions. However, it is more robust to voltage and temperature variations. On the contrary, the RO-PUF FPGA implementation, which has a lower area, is very unreliable in T.V. corners.

For both the RO-PUF and the TERO-PUF, responses can easily be made more unique and more robust to environmental variations by increasing the number of stages in the oscillators, which makes them at the same time less reliable to intrinsic noise. Thus, the size of the PUF in a given technology is a result of a trade-off between its uniqueness and its reliability. Nonetheless, we remark from Table V that TERO-PUFs are more unique than RO-PUFs while using less stages in both the FPGA and ASIC implementations. This suggests that more entropy due to MPV is extracted using TEROs than ROs having the same size.

VII. Synthesis and Analysis

The design of TEROs is mainly based on electrical simulations which make it possible to roughly set the targeted nominal number of oscillations during the development step. The main relevant parameters are the number of stages per TERO and the capacitive charge and discharge parameters of their outputs. In practice, the mean number of oscillations of a TERO depends on MPV, whereas the effective number of oscillations varies depending on intrinsic noise fluctuations. TEROs which oscillate longer tend to be more affected by noise fluctuations but they are also more affected by MPV.

TABLE IX
SELECTED CONFIGURATIONS FOR THE TERO-PUF IN A CMOS 350 nm TECHNOLOGY

PUF configuration	Number of response bits per challenge	Subtractor output	Used bits	Max. response size using $2n$ TEROs	Nb. of transistors per response bit
C1	1	Binary or Gray	8 (MSB)	n^2	$72/n$
C2	2	Gray	8 and 4	$2n^2$	$36/n$
C3	3	Gray	8, 4 and 3	$3n^2$	$24/n$

For this PUF, the development step consists in selecting the output subtractor bits used to build the PUF responses depending on the targeted performances (steadiness, uniqueness and randomness) and the size of the design. In the case of our study (CMOS 350 nm), we identified three potential configurations, which, depending on the designer's requirements in terms of temperature and voltage ranges, could be used to obtain relatively reliable responses. Those configurations are summerized in Table IX.

With two blocks of n TEROs, the number of available challenge and response pairs (CRPs) is n^2. If each challenge yields a one-bit response (C1), then the maximum response size (using the whole set of challenges) is n^2 bits. The number of transistors required for entropy extraction (whose size is fixed) does not exceed a few hundreds. The size of the selection circuitery (multiplexors) depends on the number of TEROs, but is negligible compared to the size of the PUF. Since each TERO in our design uses 36 transistors, then the number of transistors per response bit can be estimated with $(36 \times (2n))/n^2 = 72/n$. This ratio is lower for C2 and C3 since each CRP yields more response bits. The three selected configurations have the following characteristics:

- C1 has 49.72% uniqueness, 0.61% steadiness in nominal voltage and temperature conditions, and it generates sequences that passed the selected randomness test battery. Using 256 TEROs ($n = 128$), it can generate a response of up to 16, 384 bits. It requires 0.56 transistor per response bit.
- C2 has 48.26% uniqueness, 1.04% steadiness in nominal voltage and temperature conditions, and it generates sequences that passed the selected randomness test battery. Using 256 TEROs, it can generate a response

TABLE X
TEMPERATURE AND VOLTAGE RANGES FOR A STEADINESS OF LESS THAN 10%

Configuration	Temperature range for a ref. at 23 ^{o}C	Voltage range for a ref. at 3.3 V
C1	-20 ^{o}C to 70 ^{o}C	3.0 V to 3.6 V
C2	-15 ^{o}C to 45 ^{o}C	3.1 V to 3.5 V
C3	15 ^{o}C to 40 ^{o}C	3.2 V to 3.4 V

of up to 32,768 bits. It requires 0.28 transistor per response bit.

- C3 has 49.51% uniqueness, 1.73% steadiness in nominal voltage and temperature conditions, and it generates sequences that passed the selected randomness test battery. Using 256 TEROs, it can generate a response of up to 49,152 bits. It requires 0.19 transistor per response bit.

In nominal voltage and temperature conditions, the three configurations have similar performances in terms of uniqueness, steadiness and randomness. They are also the same size. The main difference between the three configurations are the maximum response size, as shown in Table IX (and therefore the number of transistors required per response bit), but also the temperature and voltage ranges required to obtain reliable responses. The temperature and voltage ranges that guarantee a steadiness lower than 10% for each configuration are listed in Table X. As can be seen, C2 and C3 require less area to achieve the same maximum response size than C1, but are less robust to temperature and voltage variations. Finally, note that C1 actually has a maximum steadiness below 7% in all the tested temperature and voltage ranges.

VIII. CONCLUSIONS

This paper proposed a theoretical study and a full overview of the design, evaluation and optimization of a PUF based on transient element ring oscillators (TERO-PUF). The TERO-PUF generates unique responses from each chip by extracting MPV in oscillating structures called TEROs. We proposed a theoretical study of the PUF which showed how the drafting effect influences the number of transient oscillations in TEROs. Electrical simulations confirmed that the number of transients oscillations in TEROs depend on the signal slopes (thus, the drafting effect parameters) and on the number of ring stages. Based on those remarks, guidelines were provided to optimize the design using electrical simulations.

We analyzed the PUF using conventional PUF evaluation criteria (uniqueness, steadiness and randomness). The analysis was based on a statistical study of the PUF responses which involved data from 30 test chips in a 350 nm CMOS process. The steadiness of the PUF responses was analyzed in nominal and corner temperature (-20 °C to -70 °C) and power supply voltage ($V_{cc} \pm 10\% V_{cc}$) conditions. One of the main contributions of this paper is the analysis of different response generation schemes (using a Gray code or a binary code and using one or several subtractor bits to construct the responses) making it possible to propose several PUF configurations (each with its advantages and drawbacks) for the CMOS 350 nm technology. Based on this analysis, three TERO-PUF configurations, denoted C1, C2 and C3, were selected and presented in Section VII.

Configurations C2 and C3 (which provide a more than 1-bit response per challenge) require less hardware but fail to produce reliable responses in T.V. corners. Although they present good PUF performances in nominal T.V. conditions, they can only be used in limited ranges of voltage and temperature which makes them not practical except for some very specific use cases (*e.g.* a shipment tracker in a refrigerated compartment). C1 presents the best trade-off between uniqueness (49.72%), steadiness in nominal conditions (0.61%) and T.V. corners (6.21%).

The design approach and the optimizations presented in this paper can be used to implement TERO-PUFs in different technologies while taking into account specific requirements in terms of area and performance.

ACKNOWLEDGMENTS

The authors would like to thank Matthias Monnier for his help in performing the measurements. They would also like to show their gratitude to Alain Aubert and Yoann Fanthou for their assistance in designing the test chips.

REFERENCES

[1] P. S. Ravikanth, "Physical one-way functions," Ph.D. dissertation. Massachusetts Inst. Technol., Cambridge, MA, USA, 2001.
[2] D. E. Holcomb, W. P. Burleson, and K. Fu, "Power-up SRAM state as an identifying fingerprint and source of true random numbers," *IEEE Trans. Comput.*, vol. 58, no. 9, pp. 1198–1210, Sep. 2009.
[3] S. S. Kumar, J. Guajardo, R. Maes, G.-J. Schrijen, and P. Tuyls. "Extended abstract: The butterfly PUF protecting IP on every FPGA," in *Proc. IEEE Int. Workshop*, Jun. 2008, pp. 67–70.
[4] D. Lim, J. W. Lee, B. Gassend, G. E. Suh, M. van Dijk, and S. Devadas, "Extracting secret keys from integrated circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 13, no. 10, pp. 1200–1205, Oct. 2005.
[5] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *Proc. 44th ACM/IEEE Design Autom. Conf. (DAC)*, Jun. 2007, pp. 9–14.
[6] H. Onodera, "Variability: Modeling and its impact on design," *IEICE Trans. Electron.*, vol. E89-C, no. 3, Mar. 2006, pp. 342–348.
[7] S. Devadas *et al.*, "'Unclonable' RFID ICs for anti-counterfeiting security applications," in *Proc. IEEE Int. Conf. RFID*, Apr. 2008, pp. 58–64.
[8] A. Maiti, J. Casarona, L. McHale, and P. Schaumont, "A large scale characterization of RO-PUF," in *Proc. IEEE Int. Symp. Hardw.-Oriented Secur. Trust (HOST)*, Jun. 2010, pp. 94–99.
[9] N. Bochard, F. Bernard, V. Fischer, and B. Valtchanov, "True-randomness and pseudo-randomness in ring oscillator-based true random number generators," *Int. J. Reconfigurable Comput.*, vol. 2010, Dec. 2010, Art. ID 879281.
[10] P. Bayon *et al.*, "Contactless electromagnetic active attack on ring oscillator based true random number generator," in *Proc. Constructive Side-Channel Secure Design (COSADE)*, Darmstadt, Germany, 2010, pp. 151–166.
[11] P. Bayon, L. Bossuet, A. Aubert, and V. Fischer, "Electromagnetic analysis on ring oscillator-based true random number generators," in *Proc. IEEE ISCAS*, May 2013, pp. 1954–1957.
[12] L. Bossuet, X. T. Ngo, Z. Cherif, and V. Fischer, "A PUF based on a transient effect ring oscillator and insensitive to locking phenomenon," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 1, pp. 30–36, Mar. 2014.
[13] A. Winstanley and M. Greenstreet, "Temporal properties of self-timed rings," in *Proc. 11th IFIP WG Adv. Res. Working Conf. Correct Hardw. Design Verification Methods (CHARME)*, 2001, pp. 140–154.
[14] A. Rukhin *et al.* (2001). *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications, NIST Special Publication 800-22*. [Online]. Available: http://www.nist.gov/manuscript-publication-search.cfm?pub_id=906762
[15] R. Maes, V. Rozic, I. Verbauwhede, P. Koeberl, E. van der Sluis, and V. van der Leest, "Experimental evaluation of physically unclonable functions in 65 nm CMOS," in *Proc. ESSCIRC (ESSCIRC)*, Sep. 2012, pp. 486–489.

Abdelkarim Cherkaoui received the M.S. degree in microelectronics from the University Joseph Fourier of Grenoble, in 2010, and the Ph.D. degree in applied cryptography from the Hubert Curien Laboratory at Saint-Etienne, in 2014. He is currently a Postdoctoral Researcher with the Techniques de l'Informatique et de la Microélectronique pour l'Architecture des Systèmes Intégrés Laboratory and Grenoble INP. His research activities cover two main topics, secured embedded systems and asynchronous design oriented toward low power.

Lilian Bossuet (SM'15) received the M.S. degree in electrical engineering from INSA, Rennes, France, in 2001, and the Ph.D. degree in electrical engineering and computer sciences from the University of South Britanny, Lorient, France, in 2004. From 2005 to 2010, he was an Associate Professor, and the Head of the Embedded System Department with the Bordeaux Institute of Technologies. Since 2010, he has been an Associate Professor with the University of Lyon/Saint-Etienne, and a member of the Hubert Curien Laboratory. He holds the special Centre National de la Recherche Scientifique Chair of Applied Cryptography and Embedded System Security. He has published over 110 refereed publications in these areas. His main research activities focus on embedded systems hardware security, IP protection, crypto-processor design, and reconfigurable architecture.

Cédric Marchand received the M.S. degree in electrical engineering from the Ecole Nationale Supérieure des Mines de Saint-Etienne, France, in 2013. He is currently pursuing the Ph.D. degree with the University of Lyon/Saint-Etienne, France. He is a member of the Hubert Curien Laboratory. He is working on salutary hardware for IP protection. His work is funded by the French ANR project SALWARE.