IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 18, 2019, accepted August 14, 2019, date of publication August 30, 2019, date of current version September 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2938729

Mixed-Signal Physically Unclonable Function With CMOS Capacitive Cells

KAMAL Y. KAMAL, (Student Member, IEEE), AND RADU MURESAN, (Member, IEEE)
School of Engineering, University of Guelph, Guelph, ON N1G 2W1, Canada

Corresponding author: Kamal Y. Kamal (kamalkamal@ieee.org)

ABSTRACT An electronic physically unclonable function usually includes an on-chip error-correcting code unit, which is vulnerable to security attacks and adds area, power, and data processing time overheads. This paper proposes a mixed-signal physically unclonable function circuit for authentication purposes, which we call the enhanced capacitive physically unclonable function. It divides the input challenge word over multiple computational groups to decrease processing time, increase security, and eliminate the need for error-correcting code units. Most of the challenge bits control capacitive networks grouped into several capacitive cells, while some are analogized through two digital-to-analog converters. One digital-to-analog converter controls the discharge loads of the capacitive cells; the other controls the reference voltage of comparator units. Each comparator controls a counter that digitizes the discharge time into a response chunk. Most of these counters operate at high frequencies for more precise time-to-digital conversion and are overflown to act as roulettes to promote unpredictability. One counter is not overflown to generate a reference response chunk to support error handling. The design allows for more intrinsic variations throughout the fabrication process, leading to unique response chunks. It applies an expanding challenge-response pair approach, generating a 128-bit response word for a 64-bit challenge word. The capacitive nature of the design supports various security features. Simulating the circuit using 45 nm complementary metal-oxide semiconductor technology resulted in an average power of 921.67 μW, a layout area of 22,470 μm^2, and an average data processing time of 118 μs.

INDEX TERMS Capacitive networks, chip authentication, EC-PUF, mixed-signal PUF, smart cards.

I. INTRODUCTION

Contactless smart cards, machine-readable documents, antitheft tags on books and drugs, and electronic keys for doors, which are known as proximity cards, are among the applications of radio frequency identification (RFID) systems which are vulnerable to physical cloning attacks [1]–[4]. Securing sensors, actuators, and other hardware devices used for supervisory control and data acquisition (SCADA) and the internet of things (IoT) requires a physical verification mechanism [5]–[9]. On-chip systems are susceptible to various breaches of security, such as counterfeiting the chip physically, modeling its function mathematically, eavesdropping on its data exchange, or bypassing its security [10], [11]. A physically unclonable function (PUF) can help encounter many security challenges in general and the physical clonability threat in particular. It is recommended to embed a PUF within smart cards, high-end RFID tags, and some other products [1], [3], [12], [13]. PUF systems are required to work under various environmental conditions; therefore, PUFs are usually tested under various temperatures and voltages. In general, such temporary variations can cause faulty responses, which decrease the reliability of PUFs [11]. Aging is another temporary source of errors; however, it is regarded as an irreversible variation that manifests after a long period of time [14]–[18]. A PUF circuit usually incorporates an error-correcting code (ECC) unit, which adds vulnerabilities to side-channel analysis attacks [19] and costs area, power, and data processing time [20]–[23]. Furthermore, the use of an ECC does not always guarantee that all faulty bits can be corrected, and therefore some PUFs contain a detection unit to flag nonstable response bits to omit them from the challenge-response pair (CRP) list during the enrollment phase [24]. Such issues reduce a PUF's reliability.

The capacitive physically unclonable function (C-PUF) [25] generates its response in such a way that the remote verifier software (VS) can handle the environmental-driven variations, eliminating the need for the ECC unit.

The associate editor coordinating the review of this manuscript and approving it for publication was Razi Iqbal.

The C-PUF [25] used square-shaped negative-channel metal–oxide–semiconductor field-effect transistor (n-MOSFET) capacitors to reduce the design area; however, such a shape does not lead to the most intrinsic variations throughout the chip manufacturing process. The C-PUF also relied on the internal frequency of the microcontroller, which is usually not high enough to indicate some minute differences among the electronic chips.

In this paper, we propose the enhanced capacitive physically unclonable function (EC-PUF). We outline MOSFET-related practices at technology, layout, and schematic levels that can be incorporated within our proposed PUF to enhance its intrinsic randomness throughout the chip fabrication process. The rest of this paper is organized as follows: Section II provides a background on the function, types, and security prospects of the electronic PUFs. The section also outlines several formulas for MOSFET variations. Section III presents the proposed design methodologies with aspects of variation and security enhancements. Section IV presents the experimental setup. Section V presents the simulation results and the discussion, and Section VI concludes the paper.

II. BACKGROUND

This section introduces general PUF functionality, common PUF designs, and security threats related to PUFs. It also discusses intrinsic variation sources related to MOSFETs affecting PUF fabrication.

A. HOW DO ELECTRONIC PUFS WORK?

Electronic PUFs refer here to MOSFET-based PUFs manufactured on silicon chips. Uncontrollable sources of randomness throughout chip fabrication processes can create distinct physical properties for each chip [26]. Throughout the fabrication process, a MOSFET suffers doping, oxide thickness (t_{ox}), and other geometrical intrinsic variations. The general concept of electronic PUFs is based on measuring the unique output *response* to an input *challenge* applied through a physical function. Although input and output signals are digital, the interactions with the physical devices are analog. Therefore, the working environments, such as temperature and voltage, usually influence the response. In electronic PUFs, the response methodology is based on one or more properties of the devices within the physical function, such as transistor threshold voltage, current, or delay.

A number of challenge words are applied to the PUF chip. The response word is measured against each applied challenge word to create a CRP list. Each CRP list is linked to the identification (ID) number of the smart card/RFID tag. The CRP list should be archived off-chip in a secure database, whereas the ID does not necessarily need to be secret and can be stored on-chip in non-volatile memory (NVM) unit. The step of recording the CRPs before distributing the product to the end-user is denoted the *enrollment phase*.

Later, when the end-user submits the smart card/RFID tag to be read by a card reader or scanned wirelessly, the ID is retrieved and sent to the verifier side, where a *verification phase* starts. The terms *verify*, *validate*, and *authenticate* are used interchangeably throughout this paper.

According to the claimed ID, the VS accesses (directly or through a third party) the related archived CRP list to randomly select a challenge word to send it to the PUF. The PUF then generates its response to be sent to the VS, which compares it to the archived one, to approve/disapprove the smart card/RFID tag authenticity. This authentication approach is also applicable to other electronic systems, such as SCADA and IoT systems.

B. TYPES OF ELECTRONIC PUFS

There are various classifications for electronic PUFs. According to their implementation approach, they are categorized here into digital and mixed-signal PUFs.

1) DIGITAL PUFS

Digital PUFs refer here to any PUF that does not use a digital-to-analog converter (DAC) unit or an analog-to-digital converter (ADC) unit. Digital PUFs can be classified into non-memory and memory-based PUFs. The first usually exploits the differences in propagation delays of the challenge bits, which race toward a combination component such as an exclusive-OR gate or multiplexer. The combination component represents the arbiter, and its output represents the response bit. By contrast, a memory-based PUF uses a memory component as an arbiter. A digital PUF can usually be implemented as a field-programmable gate array (FPGA) [27]–[34].

2) MIXED-SIGNAL PUFS

Mixed-signal PUFs are mostly based on some other physical properties rather than the propagation delay, such as the variations in threshold voltage, resistance, and capacitance. A mixed-signal PUF employs a DAC unit, incorporates analog measurement techniques, and uses an ADC stage to digitize the response. Examples of such PUFs are the integrated circuit identification (ICID) [35]–[37], the silicon nanokey PUF [24], [38], and the C-PUF [25].

PUFs are also classified into *strong* and *weak*based on their immunity from the modeling attacks. In today's computational measures, a strong PUF should be non-memory-based with at least 2^{64}CRPs [39].

C. SECURITY PROSPECTS

Data accessing to PUFs can be controlled by an application program interface (API) and hash or encrypting units to restrict any unauthorized applying of the challenges to the PUF. The API may also prevent repeating a response to add unreliability to side-channel modeling attacks [40]; therefore, such PUFs are called *controlled* PUFs [41]. Communications to the remote verifier are then protected along the communication path from end to end [42].

Security attacks are generally classified into invasive, semi-invasive, and non-invasive attacks.

1) INVASIVE ATTACKS

Invasive attacks require removing the chip's packaging by etching, drilling, or laser cutting, then using a micro-probing workstation to probe the chip [43]–[48]. An active post-process spray coating containing inhomogeneous particles can be included as introduced in [49], [50] to; 1) embed a unique signature for smart cards by measuring some electrical properties at certain spots within the inhomogeneous coating material, 2) protect the chip from the invasive attacks, and 3) harden the basic optical non-invasive attacks. The active coating included particles of various permeabilities, shapes, and sizes. In a later work, metal sensors were layered beneath the passivation layer to form an active-coating capacitive PUF [51]. In [52], the metal sensors were reshaped into pairs of metal comb capacitors to increase the surface area exposed to the coating. This model was elucidated further in [53]. A newer active-coating capacitive PUF stacked more than one layer of metal comb capacitors [54].

2) SEMI-INVASIVE ATTACKS

Semi-invasive attacks attempt to inject a temporary error into the PUF, either by a faulty instruction or by operating the chip in certain conditions. They would include testing the targeted chip with various input data, supply voltages, temperatures, and frequencies. However, such attacks do not permanently alter the properties of the chip [11], [55] [56]–[58].

3) NON-INVASIVE ATTACKS

Non-invasive attacks include optical, brute-force, man-in-the-middle (MITM), and side-channel attacks.

- *Optical Attacks:* These attacks aim to view the internal structure of the chip. Ray-resistant encapsulant shields can be used to resist such attacks.
- *Brute-Force Attacks:* An attacker tries multiple challenges to generate a specific response. This attack is more relevant to weak PUFs with a single CRP or a few CRPs. By contrast, it is not an imminent threat to strong PUFs, where each CRP is only used once or a few times.
- *MITM Attacks:*These attacks from eavesdropping, manipulating, to relaying are more relevant to uncontrolled PUFs, where communications to the verifier side are not encrypted. [59], [60]. Encrypted communications greatly decrease the viability of such threats.
- *Side-Channel Attacks:* These include optical emission power analysis (OEPA) [61], electromagnetic (EM) attacks [19], and differential power analysis (DPA) [62], [63]. In these attacks, an attacker analyzes a large number of power traces of the chip to discern its inward function. Decoupling capacitors can oppose such attacks [64]–[68].

To summarize the security prospects, applying a post-processing coating can protect internal data communication between the PUF and the microcontroller from invasive attacks. Encrypting the communication between the smart card and the verifier side resists MITM attacks.

D. INTRINSIC VARIATIONS OF A MOSFET

This section discusses some of the major intrinsic variations that occur throughout the fabrication process of a MOSFET-based chip. These random variations create distinct parameters for each MOSFET. A parametric variation is usually a source of disturbance for most applications, while for PUFs, it can be an excellent source of entropy to generate device-specific response bits [3]. This section discusses the intrinsic variations' effects on the threshold voltage (V_T). The initial threshold voltage value of a MOSFET at zero-bias (V_{T0}) depends on the flatband voltage, the bulk potential, the oxide capacitance per unit area (C_{ox}), and the depletion layer charges. The variation in the initial threshold voltage (ΔV_{T0}) plays a vital role in forming distinctive properties for a PUF chip. The polarities of V_T and V_{T0} are positive for an n-MOSFET and negative for a positive-channel metal–oxide–semiconductor field-effect transistor (p-MOSFET). In both MOSFET types, there is a roll-off voltage that contributes to $|\Delta V_{T0}|$, which reduces $|V_{T0}|$. For generic (non-PUF) applications, chip designers try to reduce the intrinsic variations by taking into consideration some factors, which are briefed here at the technology, layout and mask, and schematic levels.

1) TECHNOLOGY LEVEL

The introduction of strained silicon since 90 nm complementary metal-oxide-semiconductor (CMOS) technology and below enables the electrons to move faster, which ultimately makes MOSFETs switch faster [69]. It results in better MOSFET performance and lowers energy consumption; however, it causes more deterministic, or systematic variability in a MOSFET's parameters, such as C_{ox}, MOSFET channel width (W), and the charge carrier mobility (μ), which can be specified as mobility of electrons/holes for n/p channel as (μ_n/μ_p), respectively.

Statistical variability is another type of variation, it attributed to random variations in parameters like V_{T0}, μ, C_{ox}, L, or W. However, the variation in V_{T0} is the most prominent parameter and is primarily ascribed to random dopant fluctuation (RDF) within the channel and the gate; besides other factors, such as the random fluctuation of surface roughness, oxide charge, and the lithography driven line edge roughness (LER) [70]. The fluctuation in V_{T0} mainly depends on RDF, as the dopant concentration can vary up to ±10% [71].

The variation of t_{ox} gained more importance as the technology shirks it down. Using high-k insulators with larger physical t_{ox} abates its influence on variation. Unlike t_{ox}, the effect of the oxide charge on V_T value does not represent a significant factor in modern fabrication technologies, especially when a MOSFET works at strong inversion mode [72].

Let $K' = \mu C_{ox}(W/L)$, then an RDF of 10% with 5% variation in t_{ox} can lead to variation of 100 mV in V_{T0}, 15% in K', and 5% in the source-bulk (C_{SB}) and the drain-bulk (C_{DB}) junction capacitances [73].

Silicon-on-insulator (SOI) technology mitigates RDF, but an ultra-thin body SOI MOSFET raises more uniformity concerns. The high-k/ metal gate technology reduces RDF but creates metal gate granularity (MGG), which becomes among the major sources of ΔV_{T0} in scaled bulk-MOSFETs [74]. In general, a MOSFET fabrication technology tends to decrease $|\Delta V_{T0}|$ by; 1) strict control over doping fluctuation of the source and the drain; 2) strict control over the critical geometrical dimensions, such as the gate length and width of the MOSFET; 3) scaling up the gate insulator thickness; and 4) scaling up the junction depth [75].

Choosing a MOSFET manufacturing technology with less control of characteristics can better serve chip authenticity applications.

2) LAYOUT AND MASKING LEVELS

Low lithography resolution is a major source of systematic local process variations. Such imprecision is treated with resolution enhancement techniques, such as phase shift masking by adding phase information to the mask, off-axis illumination to optimize the angles of light illuminating the mask, source polarization to control of the polarization of the illumination, source mask optimization, and optical proximity correction (OPC) to optimize the mask pattern shape [70], [76]–[80]. Besides the lithography-related geometry variations, additional variations are added due to spacing, distances to the shallow trench isolation, and position variances of contacts. A regularized design and dummy features can greatly reduce the impact of systematic variability [81].

Among the basic layout approaches that are commonly applied to reduce such variations are; 1) matching the orientation of the transistors' layouts to decrease mobility variations, 2) laying out transistors as close as possible and using common centroid layout to minimize gradients throughout the fabrication process, 3) using dummy gates at the sides of transistor layouts, 4) not laying out contacts on top of active gates, and 5) not laying out metal lines across active gates [82].

Within an electronic PUF circuit, wherever the intrinsic variations are desired, the layout and masking processes can follow special approaches to oppose or at least avoid any variation-limiting effects, whereas the typical anti-variation approaches can still be implemented elsewhere.

3) DESIGN LEVEL

An intrinsic-based variation of a MOSFET's parameter among similarly designed MOSFETs within a chip is usually referred to as *intra-die variation* or *mismatch*.These variations can be due to random local effects (such as the non-flatness of the polysilicon gate due to granularity) and to t_{ox} gradients over the chip. The general trend of MOSFET technology evolution is toward smaller horizontal dimensions, smaller physical t_{ox}, smaller junction depth, heavier substrate doping, and (although not always) lower power supply voltage.

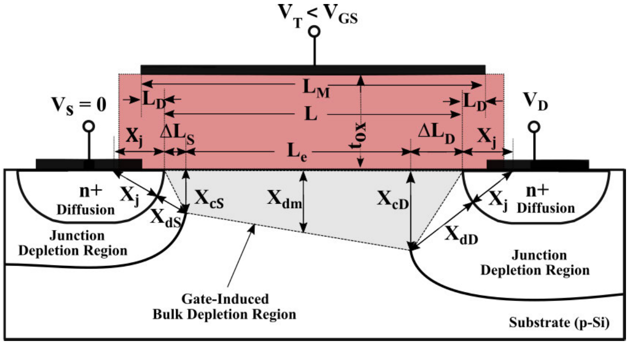


FIGURE 1. Regions of planar n-MOSFET.

In Fig. 1, a planar bulk MOSFET has as a mask length (L_M). This length is reduced due to depletion of the source and drain junctions by a length (L_{dep}) on each side, leaving a channel length (L), which is determined as [83]

$$L = L_M - 2L_{dep}. \tag{1}$$

It is important to emphasize that L, L_M, and L_{dep} exist even without applying any external voltage at the MOSFET. After applying a voltage on the drain/source, lateral extents of the p-n depletion regions of the bulk-source and the bulk-drain are produced, they are denoted here as (ΔL_S) and (ΔL_D), respectively. When a voltage is applied at the gate, there will be a shared gate-source/gate-drain influence on the depletion charges of the channel at the source/drain sides, respectively. The remaining distance between the source and drain depletion is considered the effective channel length (L_e), where the diffusion charges are influenced by the gate voltage only [84]. Then the L_e is calculated as

$$L_e = L\,(\Delta L_S + \Delta L_D). \tag{2}$$

The effective length exists only when the channel is formed by applying external voltages at the MOSFET, while the bulk-source and the bulk-drain depletions exist even without applying any external voltage to any of the MOSFET's terminals. A typical channel region contains only about $L_e^{1.5}$ dopant atoms; therefore, as a MOSFET fabrication process is scaled down, mismatch due to dopant fluctuation increases [85]. Furthermore, since V_T is inversely proportional to the square root of the device area [86], the current would, therefore, have a similar dependency on the area. For example, in analog designs, to improve the current matching by a factor of 2, the device area is quadrupled. One of the approaches to achieve that is doubling both W and L [73]. For PUF applications, minimizing the area of the MOSFET device can increase the current mismatching. In small-geometry devices, there are more dominant factors that contribute to ΔV_{T0}, such as non-uniform vertical and lateral doping concentrations, short channel, narrow width, and drain-induced barrier lowering.

Considering the short-channel effect (SCE) varies V_{T0} by a value denoted as (ΔV_{T0}^{SCE}), then:

$$V_T = V_{T0} \pm \Delta V_{T0}^{SCE} \tag{3}$$

where ΔV_{T0}^{SCE} here is considered positive, despite the channel type. The variation's sign in (3) is negative for n-channel and positive for p-channel. In other words, the term

$\pm\Delta V_{T0}^{SCE}$ always carries an opposite sign to V_{T0}, as the depletion regions around the junctions (the wells) reduce V_T, where $\left|\Delta V_{T0}^{SCE}\right| \propto (X_j/L)$ [87].

On the other hand, in narrow-channel MOSFETs, when W is on the same magnitude order as the maximum depletion region thickness X_{dm}, another source of variation that contributes to V_T must be considered. In addition to the oxide thickness above the channel (t_{ox}), there is a thick field oxide (FOX), which covers the region around the channel to prevent the surface leakage currents between adjacent MOSFETs. The overlapped area between the gate electrode and FOX develops a low depletion region, which raises V_T. This phenomenon is known as the narrow width effect (NWE) [88]. The narrow-channel-based variation of V_{T0} is denoted as (ΔV_{T0}^{NWE}), which always carries a positive sign. Then it affects V_T as [87]

$$V_T = V_{T0} \pm \Delta V_{T0}^{NWE} \tag{4}$$

where the variation sign in (4) is positive for n-channel and negative for p-channel. In other words, the term $\pm\Delta V_{T0}^{NWE}$ always has a similar sign to V_{T0}, and if we consider the absolute values, then we can say that the narrow channel causes extra depletion charge that ultimately increases $|V_T|$. Combining SCE and NWE, then V_T is determined as

$$V_T = V_{T0} \pm \Delta V_{T0}^{SCE} \pm \Delta V_{T0}^{NWE} \tag{5}$$

If we consider the case of a short-narrow n-channel, then the variations $(-\Delta V_{T0}^{SCE})$ and $(+\Delta V_{T0}^{NWE})$ tend to cancel each other out, and a similar argument (but with opposite signs) is valid for a short-narrow p-channel [87].

At the design level, a PUF designer should oppose the general anti-variation approaches which most analog designers follow, such as:

- To avoid the intra-chip mismatch among MOSFETs, which is attributed to statistical randomness, if the current should match, a high gate-to-source voltage (V_{GS}) is usually recommended to decrease the influence of V_T [89]. Another approach to decrease the current mismatch by a factor of 2 is by squaring the MOSFET's area [73]. On the other hand, if the voltage should match, it is recommended to keep low V_{GS}. That can be done by increasing the channel width-to-length (W/L) ratio [70], [89].
- In some cases, a designer can choose between using n-MOSFET or p-MOSFET. Usually, n-MOSFETs have more intrinsic variations than p-MOSFETs. The random discrete dopants, the LER, the polysilicon granularity of the gate electrode, and surface potential pinning at the poly-Si grain boundaries play important roles in the statistical variation of n-MOSFETs, while play negligible roles in p-MOSFET [90].
- For resistors, an N-well diffusion resistor is less susceptible to the intrinsic variations due to its lower doping and larger volume. Furthermore, it is made of monocrystalline materials, which reduces the impact of defects and grain borders [91]. Alternatively, a MOSFET-base resistor is more susceptible to the intrinsic variations.

An effective approach to improve the uniqueness of MOSFET-based PUF design is to include short-channel MOSFETs and/or narrow-channel MOSFETs. In either way, V_T would be more different among the equally scaled MOSFETs, which can eventually lead to more distinctive PUF chips. Since such differences are attributed to random intrinsic variations, it is quite probable to have unique properties for each PUF chip. Reversing or at least avoiding any variation-limiting approach allow for more intrinsic variations, which can create unique PUF chips.

III. METHODOLOGIES OF THE EC-PUF DESIGN

This section introduces the proposed EC-PUF system and schematics. It presents the EC-PUF design, protocol, error handling, CRP expansion, and other security aspects. The section also explains the applied variation and stability enhancements.

The EC-PUF is a mixed-signal PUF that is based on challenging groups of networked capacitors. Each group is referred to as a capacitive cell (CC). In the simulated design, the EC-PUF has eight CCs, as shown in Fig. 2. A time-to-digital converter (TDC) digitizes the discharge time of each CC into a 16-bit response chunk (RC). Concatenating all RCs yields a total response of 128 bit.

The uniqueness of the response is attributed to the uniqueness of the intrinsic variations throughout the fabrication process of the EC-PUF chip, especially of the analog units within the chip.

A. EC-PUF SCHEMATICS

The proposed EC-PUF comprises eight challenge-controlled CCs, eight comparators, eight 16-bit counters, one oscillator, one Schmitt trigger, and eight frequency dividers. It can accommodate an optional substitution box (S-box), as shown in Fig. 2. The EC-PUF also includes two DACs, that are DAC1 and DAC2; each of eight challenge bits referred to as M and N, respectively. The challenge-controlled reference voltage is controlled by DAC1, while the challenge-controlled discharge load is controlled by DAC2.

Each CC consists of six sections of several parallelly networked n-MOSFET capacitors. Each section is to be challenged by a challenge-bit in the field (ch63-ch16). Fig. 3 illustrates a simplified CC model with only one section that has only one n-MOSFET capacitor. The CCs were constructed with n-MOSFETs to reduce the silicon area since an n-MOSFET has a greater charge per unit area than a p-MOSFET. The primary n-MOSFET capacitor (NM0) is charged each time an authentication phase is initiated; this enables a counter to generate a non-zero RC, even when every challenge (ch) bit within the capacitive challenge bits group (L') is set to a high voltage (typically 1 V), which represents a logic 1. The charging operation of the auxiliary n-MOSFET

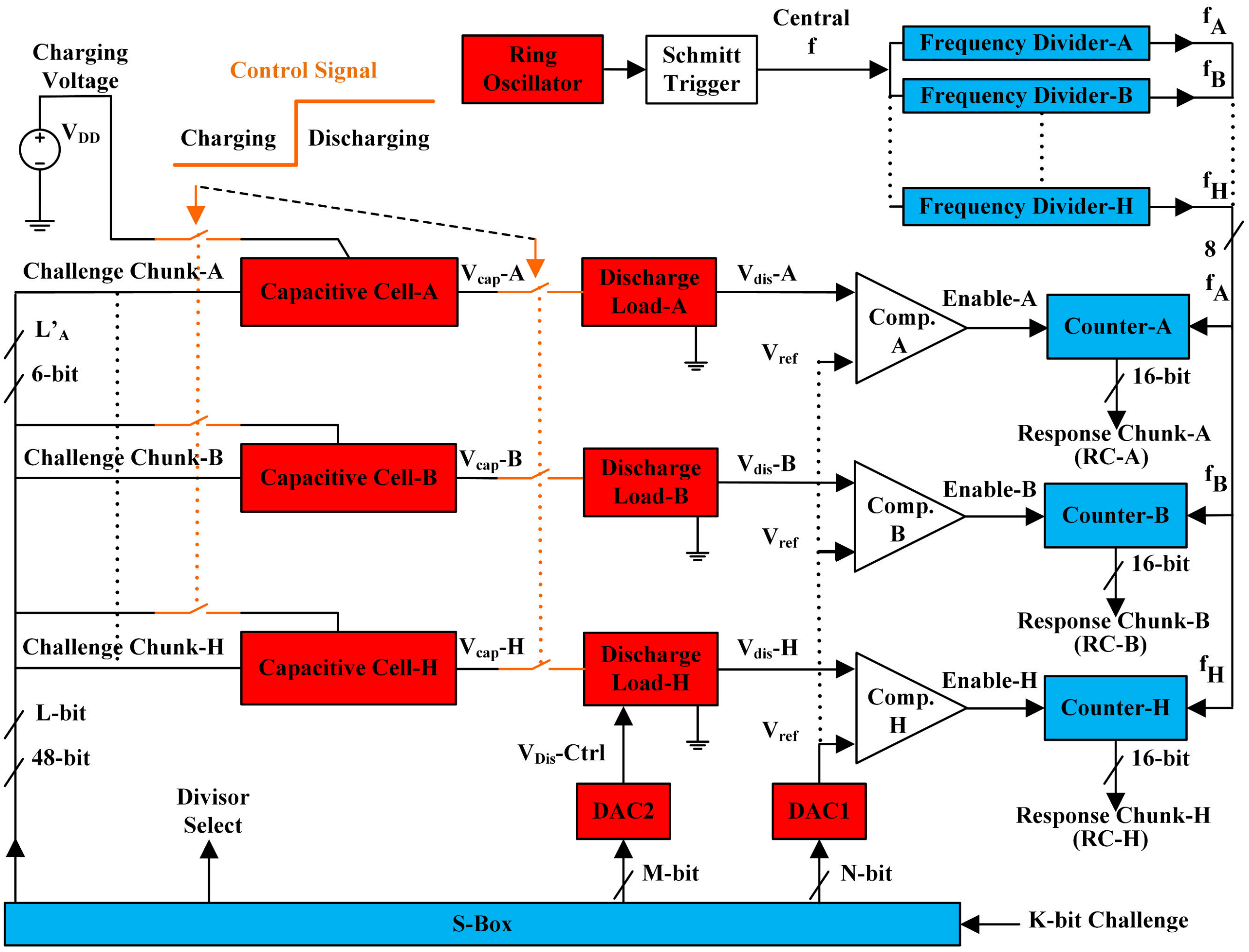


FIGURE 2. Enhanced capacitive physically unclonable function block diagram. DAC = digital-to-analog converter.

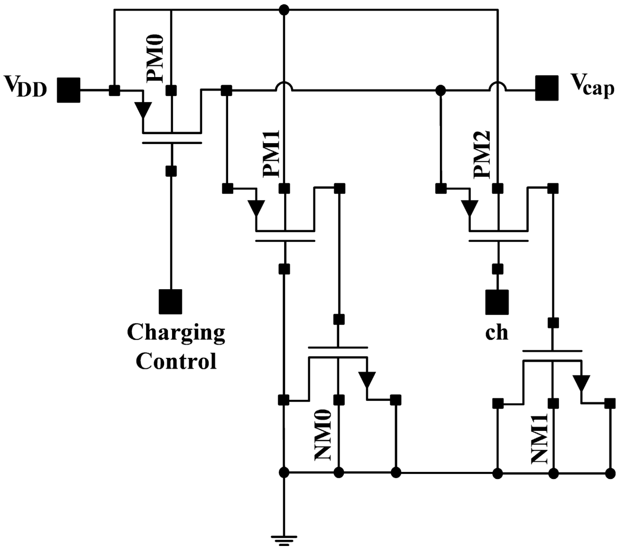


FIGURE 3. Capacitive cell module.

capacitor (NM1) is controlled by the ch bit. When ch is of a high voltage, and the charging control signal is applied at the gate of the main switch p-MOSFET (PM0), this makes the switch p-MOSFET (PM2) passes the charging current toward NM1.

To store larger charges in each capacitive section, NM0 and NM1 were extended parallelly with other n-MOSFETs. However, sizes and numbers of the n-MOSFETs within each section are random, to maintain unpredictability of the discharge times of each capacitive section enabled by a ch bit within an L' group of bits. The drains of all n-MOSFETs capacitors of all capacitive sections within a CC are joined into one terminal whose voltage is the capacitive cell's voltage (V_{cap}), as shown in Fig. 3. The electric path of each V_{cap} terminal toward the discharge load is controlled by a pass (or transmission) gate as a discharging control switch. Each of the eight pass gates consists of two parallelly connected MOSFETs; one p-MOSFET to better pass the high voltage around the drain voltage (V_{DD}) value, and one n-MOSFET to better pass the low voltage around the source voltage (V_{SS}) value.

The authentication phase starts when a low control signal of 0 V is applied at the gate of PM0. This control signal is generated from a timer unit within a microcontroller embedded within the smart card's chip. The width of the low-level pulse controls the charging time of the eight CCs and resets the eight counters. For our design, the minimum time required to fully charge the CCs in the proposed design is about 1 ns. It is not critical if the microcontroller is not capable of generating such a narrow pulse, as a wider charging time will not change the value the maximum charge (Q) of a CC.

The discharge phase starts when a high-level control signal of 1 V is applied by the microcontroller unit. This high-level signal shuts down PM0 of the eight CCs to stop the charging phase, enables the eight 16-bit counters, and simultaneously turns on the eight pass gates to connect each CC to its discharge load. Each discharge load is merely an n-MOSFET with a grounded source terminal.

When the control signal turns the eight pass gates on, this applies V_{cap} of each of the eight CCs at its related discharge load n-MOSFETs. In this case, the discharging voltage across the drain-to-source (V_{dis}) equals V_{cap} at the drain of the discharge load n-MOSFET.

The gates of all eight n-MOSFETs are driven by one discharge controlling voltage ($V_{Dis-Ctrl}$), as shown in Fig. 2. $V_{Dis-Ctrl}$ is the analog output voltage (Analog) of DAC2, as shown in Fig. 4. The eight challenge bits (ch15-ch8) control 256 levels of $V_{Dis-Ctrl}$, which drives the gates of the eight discharge-load n-MOSFETs. Alternatively, the analog voltage of DAC1 represents the reference voltage V_{ref} for all comparators, as shown in Fig. 2. In other words, the challenge bits (ch15-ch0) control 256 levels of V_{ref} at the eight comparators.

The voltage across each discharge load n-MOSFET that is V_{dis} drives a non-inverting input of a comparator. Each comparator gives an output voltage equals to V_{DD} when the applied V_{dis} at its non-inverting input is greater or equal to V_{ref}. Each comparator's output represents an enabling signal (E) that controls a 16-bit counter. The schematic in Fig. 4 illustrates DAC1/DAC2, where each is an 8-bit DAC and constructed as a MOSFET-based R-2R ladder. Each DAC was designed to output low analog voltages, especially DAC2, which drives all eight discharge loads. Each discharge load is implemented here as an n-MOSFET that functions within the subthreshold margin to maintain a discharging time that is large enough to be measured by digitizing it through a 16-bit counter that is clocked by a high frequency. The final digital value that appears at a counter is an RC, which does not necessarily represent the discharge time, as a large discharge time and a high-frequency clock collaborate toward overflowing the related counter. The actual digital representation for a discharge time of a CC is referred to here as the non-rouletted response chunk (NRC). For a 16-bit counter, the decimal representation of RC is the remainder after dividing NRC by 2^{16}. Such modulo (MOD) operation is represented here as $RC = MOD_{2^{16}}$ (NRC). Among the eight CCs, there is one CC with small charge storage capacity, whose discharge time is digitized by a relatively lower frequency to prevent its related 16-bit counter from overflowing. We call this CC the reference capacitive cell and its non-rouletted response the reference response chunk (RRC).

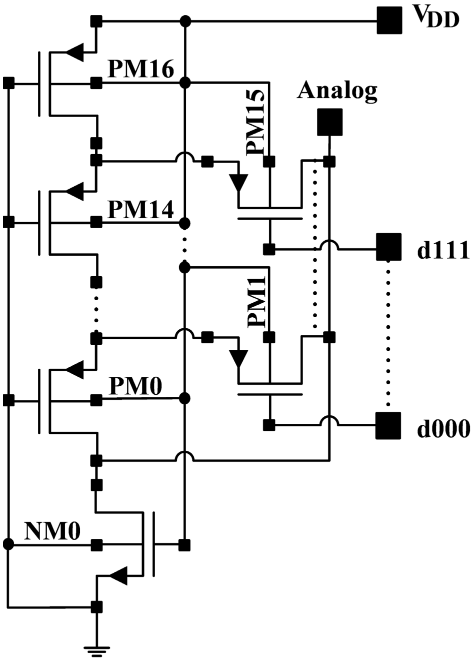


FIGURE 4. MOSFET-based R-2R ladder.

The discharge load n-MOSFETs were sized similarly in this design, although they can be scaled differently. Nevertheless, since these n-MOSFETs are scaled at minimum feature size, the intrinsic mismatch variations of the fabrication process would still contribute to the individuality of each RC locally within the same chip and globally among the similar EC-PUF chips. The eight comparators were also scaled identically, but they were designed, laid out, and to be masked according to conventional anti-variation approaches, yet some variation is also expected to contribute to the individuality of the response.

A supply voltage of analog and mixed-signal chips is usually controlled by a bandgap reference (BGR) regulating circuit, which can sustain the voltage within a variation window of a few millivolts. The proposed design did not include a voltage regulator, as it is assumed to be pre-existing among the other system-on-chip units; however, we used only a BGR circuit for the comparators. This enables the simulation to emphasize the effect of the voltage variation on the CCs and the DAC units, which ultimately shifts the response value up or down.

B. BASIC AUTHENTICATION PROTOCOL

The response of most PUFs is also used as an ID and/or cryptographic key [51], [92]. Combining the authentication with the identification/cryptography eliminates error-tolerability and hence an on-chip ECC unit becomes essential. Instead, the basic protocol of the EC-PUF requires to retrieve the ID information from an on-chip NVM, as most of today's smart cards and RFID tags, but embedding an on-chip PUF makes the ID data security less crucial. A basic verification phase protocol of the EC-PUF can be summarized in three main steps:

- The smart card/tag reader sends the ID to the verifier side. According to that ID, the VS randomly picks a 64-bit challenge and sends it to the EC-PUF. These challenges were measured and archived during the enrollment phase.
- The PUF receives a 64-bit input challenge and generates a 128-bit output response. The PUF sends the response to the verifier side without correcting it on-chip.
- The VS correlates the received RCs to the archived ones. To analyze the environmental-driven shifts within the NRCs' values, the VS may simply analyze the RRC's shift to evaluate the shifting trend of all NRCs. It can also apply an extrapolation approach to better achieve its verification task. It may also include an artificial intelligence structure to achieve its verification analysis, which can also help to handle the aging-driven shifts within the NRCs.

Ultimately, if the authenticator approves the authenticity of the EC-PUF chip, it allows the smart card to proceed to the next requested step.

C. THE EC-PUF MODEL

In this section, we model the EC-PUF design and discuss the major influential parameters. The considered parameters are listed in Table 1. Fig. 5 shows the basic parameters which control an RC. The analog units of the EC-PUF are drawn in red. The upper red block represents the factors that Q of a CC depends on, such as; 1) the CC related challenge bits (L'), where each bit decides which capacitive section to enable to be charged and which to leave disconnected; 2) the total size of the charged capacitors of all enabled sections (S) within each CC; 3) the structure organization (O) of each charged capacitive cell, although our basic capacitive cell consists only of capacitive devices networked parallelly, a cell can include other components such as inductors and resistors, and can be networked in various ways; 4) the intrinsic variation of each component within a CC, denoted here as (v); 5) the supplied voltage (V); and 6) the temperature (T). The total charge can then be depicted as Q (L', S, O, v, V, T), as in Fig. 5.

TABLE 1. Abbreviations and values.

Symbol	Details	Value
C	Number of bits of each counter	16 bits
f	Clock pulse frequency for a counter	varies
E	Comparator output	0/1 V
K	Total challenge bits	64 bits
L	Total challenge bits for C	48 bits
L'	Challenge bits for each capacitive cell (CC)	6 bits
M	Challenge bits for discharge control	8 bits
N	Challenge bits for reference voltage	8 bits
O	Organization of within CC	varies
P	# of CCs, discharge loads, comparators, and counters	8 units
Q	The maximum charge of a CC	varies
R	Total response bits of 8 counters	128 bits
R'	# of bits of each counter	16 bits
S	Size of the charged capacitors in a CC	varies
T	Temperature	−55°C to 125°C
V	Supplied voltage (V_{DD})	1 V to 1.2 V
v	Intrinsic variations of a CC	varies
v'	Intrinsic variations of DAC1&2	varies
v''	Intrinsic variations of discharge load	varies

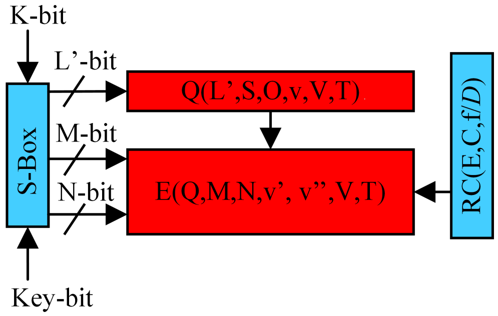


FIGURE 5. Data dependency of a response chuck.

The lower red block in Fig. 5 refers to the parameters which affect the period of the enable signal (E), where a high voltage level of 1 V represents an active E signal. A related counter is active by E and counts the applied clock pulses for the period of active E. Since the model in Fig. 5 refers to a single CC, the lower red block includes a single discharge load n-MOSFET, the output analog voltage of DAC1(v'), the output analog voltage of DAC1(v'), the supply voltage V, and the temperature T. Ultimately, E can be represented as a function E (Q, M, N, v', v'', V, T). The intrinsic variation of the comparator is another parameter that can affect E. However, the generic comparators such as the eight used in our EC-PUF are less affected by the intrinsic variations. Furthermore, the BGR that regulates the voltage for all comparators minimizes the effect of the voltage variation on the comparator.

The blue blocks in Fig. 5 refer to the digital portions of the EC-PUF model. Unlike the analog portions, the intrinsic variations usually do not alter the digital data processed with these units. The right blue block represents the parameters that impact an RC. It includes the period of counting activation E, the number of the bits of the counter (C), the clock pulse frequency applied at the counter after dividing the central frequency f by a divisor D. The RC of a CC is then represented as R (E, C, f/ D).

The left blue block in Fig. 5 illustrates the possibility of adding an S-box to add more *confusion* against model attacks. Furthermore, it is also possible to insert an authentication key at the S-box to add another level of security. The key in Fig. 5 can be provided from the authenticator side either as a permanent security code within an on-chip NVM (which represents a security threat) or through a secure communication channel as a short-term or even as a one-time password [93], [94]. Embedding an S-box is optional, and its structure can go from a few basic gates to plenty of various combinatorial logic blocks. It can be constructed in such a way that a chunk of the challenge bits (L') at one CC can influence (L') of other CCs.

In standard advanced encryption standard (AES), the S-box only applies the *confusion* concept of cryptography [95]. In our proposed EC-PUF design, an optional S-box can be added to serve both *confusion* and *diffusion* concepts. The diffusion within the EC-PUF can be applied when the design wires some of the challenge bits (L') of some or all CC as input to the S-box. For the EC-PUF, since the complexity of the S-box can vary widely depending on the application requirements, and since adding an S-box here is totally optional, the S-box is not included within the area and the power measurements.

D. EC-PUF FREQUENCY ASPECTS

Many microcontrollers operate at low internal frequencies. For example, the LPC8N04 microcontroller has a maximum internal frequency of 8 MHz [96]; then its timer can control the EC-PUF with a minimum period of 125 ns. In other words, although the EC-PUF's capacitors require only 1 ns to be fully charged, the charging time will be 125 ns. This is not a problem for the EC-PUF; instead, this makes it compatible with most of today's microcontrollers despite their internal frequencies. As after the charging control signal rises, the microcontroller can count a number of its internal clock pulses as waiting time and then read the 128-bit response, without the need for handshaking of control signals nor synchronization. The waiting time depends on the environmental conditions and corners of the chip. For the proposed chip, a waiting time of 750 μs is sufficient.

The clock pulse frequency here represents the sampling rate of the time-to-digital conversion. The higher a counter's clock pulse rate, the more precise the time-to-digital conversion would be. This ultimately enhances the ability to detect the distinct properties of each individual EC-PUF chip. A high frequency can overflow the seven non-reference counters, making them act like roulette wheels. This behavior promotes the *confusing* factor, which helps resist the modeling attacks. The proposed EC-PUF design includes a high-frequency ring oscillator instead of relying on the inner frequency of the microcontroller, which is generally low.

On the other hand, there are limitations on the frequency rate, as high frequency costs extra dynamic power. Besides, there is the limitation of propagation delay throughout each synchronous counter. This delay primarily depends on MOSFET technology, design, manufacturing, and environmental variations. Furthermore, the reference counter, which is affiliated with the smallest CC to generate the (non-rouletted) reference response, will also be overflown if a greater frequency is applied to it.

The maximum frequency that can run the 16-bit reference counter without rouletting was measured under these conditions; 1) the six input challenge bits which control the reference CC were all set to 0 V to achieve the maximum charge; 2) the eight input challenge bits which control the discharge loads were all set to 1 V to achieve the largest resistive load; 3) the eight input challenge bits which control the reference voltage at all comparators were all set to 1 V to achieve the largest time for the enabling signal E; 4) the simulation temperature was set at a minimum of -55^{o} C; and 5) the supply voltage (V_{DD}) was set at a minimum value of 1 V. Under those conditions, the discharge time of the reference CC was 29.38703 μs. Then, for the reference counter to count 65,535 without rouletting, its clock should not exceed 2.23 GHz. The initial frequency f of the oscillator at those conditions was 3.95 GHz. This frequency is central for all counters.

Our EC-PUF design includes a frequency division unit that contains eight digital frequency dividers. The central frequency is divided by 1, 2, 3, 4, 5, 6, 7, or 8 divisors. Some divisors can be selected by the challenge, either directly (as implemented here) or through the S-box. For example, one of the challenge bits in this design opts one of the divisors to be 3 or 2. The intrinsic and environmental variations within the oscillator and the Schmitt trigger add randomness and uniqueness for the EC-PUF chip; even though the Schmitt trigger per se does not alter the frequency, its loading effect on the oscillator does.

A digital frequency divider preserves a precise frequency division, which both the intrinsic and the environmental variations cannot alter. This helps the authorized VS to handle the environmental-driven effects that are supposed to shift the eight RCs.

In the proposed EC-PUF, the central frequency (f) is digitally divided by three, before running counter-A which is related to the smallest CC (CC-A). At low frequency ($f_A = f/3$), counter-A digitizes the small discharge time of CC-A (t_A), this ensures that counter-A will not be overflown, that is to consider RC-A the reference response chunk (RRC) for the EC-PUF.

Alternatively, the central frequency (f) directly runs counter-B, in other words, ($f_B = f/1$). Counter-B is related to the largest CC (CC-B), which generates the largest discharge time (t_B). Aside from the environmental factors, such as temperature and voltage, the value of t_B mainly depends on the challenge bits L'_B, M and N. A large t_B along with a high f_B may overflow (roulette) counter-B, then the actual number of counts is the non-rouletted count (NRC-B); while the final count which counter-B will stop at, and a microcontroller will read and send to the VS is the response chunk (RC-B). Each counter within the proposed EC-PUF is of 16 bits, then:

$$\text{RC-B} = \text{MOD}_{2^{16}}\,(\text{NRC-B}) \tag{6}$$

In equation (6), both count values NRC-B and RC-B are in decimal. The modulo equation in (6) is applicable to all counters within the proposed EC-PUF. It is also obvious from (6) that if a counter is not overflown, then its RC will equal its NRC.

- At the enrollment phase, for each challenge word, it is possible to archive the non-rouletted response ratio of counter-B to counter-A, that is (NRC-B/RRC), which is denoted here as (RR-B).

- At a verification phase, unlike attackers, the VS can access RC-B, NRC-B, and RRC which were archived during the enrollment phase and can relate them to RC-B and RRC which are received at a verification phase. Similarly, the VS can overview all RCs to figure out the common environmental-driven shifting trend in their values. This enables the VS to reverse equation (6) to determine all NRCs. This is the main concept which enables the proposed EC-PUF to eliminate on-chip ECC unit or any other form of on-chip helper data.

If the enrollment phase of the EC-PUF is done under a central frequency f, then RR-B can be determined as:

$$\text{RR-B} = \frac{t_B}{t_A} \times \frac{f_B}{f_A} = \frac{t_B}{t_A} \times \frac{f}{\frac{f}{3}} = 3\frac{t_B}{t_A} \tag{7}$$

Since the central frequency (f) is digitally divided by constant divisors to generate a constant ratio of (f) for each counter, such as ($f_A = f/3$), then equation (7) indicates that having a different (f) at a verification phase does not alter RR-B, as long as there is a degree of constancy in the rate t_B/t_A under all environmental conditions. Centralizing the controlling voltage of all discharge loads by one DAC (DAC-2) and the reference voltage at all comparators by one DAC (DAC-1) helps to maintain a degree of stability for the discharge time ratios and the non-rouletted response ratios (RRs). Archiving the RRs as off-chip helper data is a key concept to handle the environmental-based effects on the NRCs.

E. ERROR HANDLING

During an authentication phase, a PUF usually generates a response with some inevitable rate of errors. Some errors are attributed to temporal variations in the environment, such as the variation in temperature, voltage, or frequency. To handle such errors, an on-chip ECC is usually embedded, but since it is insecure, eliminating it is one of the objectives of the proposed EC-PUF.

The EC-PUF exploits the correlations among the individual RCs to handle the environmental-driven shifts of the RCs' values and ultimately to authenticate the EC-PUF chip. There are two main features which support the VS to handle the environmental-driven shifts in the generated RCs:

- Within the analog portion, there is a central M-bit set to control all discharge loads and a central N-bit set to control the reference voltage V_{ref} of all comparators.
- Within the digital portion, there is a central f for all counters, although it is divided by various divisors since they are known to the VS.

Among the 128-bit response, the total effect of the environmental effects is conveyed to the remote VS by the 16-bit RRC. The only specialty about the RRC is that its CC is relatively small, and its counter's frequency should be relatively low to avoid rouletting its 16-bit counter. Alternatively, there are no limitations to the CC size nor the counter's frequency regarding the other seven 16-bit RCs. A VS is supposed to analyze each RC with a degree of tolerance, which eliminates the need for an on-chip ECC unit.

The VS can reconfigure its tolerance margins to tolerate the environmental effects on the EC-PUF chip. The tolerance margins control the false acceptance rate (FAR) and the false rejection rate (FRR) of the authentication phase. The proposed EC-PUF not only eliminates on-chip ECC but also gives the authorized VS flexibility to handle errors.

F. CRP EXPANSION

A group of similarly produced objects is referred to as a *class* [97]. A class of PUF chips can be challenged by one 64-bit challenge word only since each chip generates a unique response, which eventually creates a unique CRP for each individual chip. It is evident that only the number of response bits can limit the population of a class. The output response word of the proposed EC-PUF consists of 128 bits and hence a difference of one bit permits a population of about 3.4E+38 chips. The CRP expansion aims to balance constraints of area, power, population, and security.

G. THE EC-PUF VARIATION ASPECTS

In the EC-PUF schematic, we applied variation-aware aspects, based on the type and the scale of some components:

- The C-PUF in [25] has square-shaped n-MOSFET capacitors to save the area, whereas the EC-PUF has short-wide n-MOSFET capacitors with a minimum channel length to support attaining more intrinsic variation throughout the fabrication process.
- The two DAC ladders introduced in [25] were based on diffusion N-well resistors. Instead, the EC-PUF ladders were designed using MOSFETs in the subthreshold region as resistors. Such ladders can cause more random variations and save area. Each of these MOSFET-based resistors has a long, narrow channel. The non-square MOSFETs in the two DAC ladders achieve a high electrical resistivity with less physical size and more intrinsic variation.

H. APPLICABLE VARIATION ENHANCEMENTS

At layout and masking levels, causing fabrication variations requires the layout and masking engineer to act exactly in the opposite direction against any variation-limiting approaches. For example, PUF masks are better with PUF-aware OPC applied to improve their uniqueness [98], [99]. This approach can be applied wherever a variation is required in a portion of the circuit, whereas a regular OPC can normally be applied elsewhere.

Choosing a fabrication technology with high-rate intrinsic variations can help acquire distinct PUF chips. Besides, it is sometimes possible to influence certain fabrication stages to achieve more variations as a kind of PUF-aware fabrication.

I. MEASUREMENT ENHANCEMENT

Most generic microcontrollers adopt low-power approaches, including lowering the clock pulse of the internal oscillator. Our first C-PUF [25] relied on the internal frequency of the microcontroller, assuming a frequency of 100 MHz. This frequency represented the sampling rate of the TDC units. Alternatively, the EC-PUF includes a high-frequency oscillator for higher sampling rates, which enhances the precision of the time-to-digital conversion to better indicate tinier intrinsic variations in the form of RCs.

J. CRP ENHANCEMENT

Increasing the total challenge bits to 64 in the EC-PUF, instead of 21-bit in the C-PUF [25], enhances the security by two aspects; longer challenge word string and more possible CRP choices.

Furthermore, since it is safer to use each CRP only once throughout a PUF's lifetime, then extending the number of the possible CRPs allows more authentications. This can support an extended lifetime for a PUF chip as well as for its attached product, such as a smart card or RFID tag.

K. THE EC-PUF STABILITY ASPECTS

The stability-enhancing is discussed here in terms of thermal aspects and transient aspects.

1) THERMAL ASPECTS

The magnitudes of the drain currents vary with temperature due to variations of the threshold voltage and the mobilities of the charged carriers. It is well known that both the threshold voltage and the mobility are inversely proportional to the temperature. In linear and saturation modes, increasing the threshold voltage increases the drain current; decreasing the mobility decreases the current. At lower V_{GS}, the variation in V_T dominates; therefore, increasing the temperature increases the drain current. At higher V_{GS}, mobility dominates; therefore, increasing the temperature decreases the drain current. At some V_{GS}, both effects cancel each other out, and the drain current does not change with temperature. As a rule of thumb, when V_{GS} is much less than V_T, the temperature rises the drain current, whereas when ($V_{GS} = V_{DD}$), the temperature reduces the drain current.

Furthermore, when working in subthreshold mode, a subthreshold drain current of p-MOSFET is very much less affected by temperature variations than n-MOSFET. On the other hand, at saturation mode, the drain current of a p-MOSFET is slightly more affected than in n-MOSFET [100]. In the proposed EC-PUF, most MOSFETs within both R-2R ladders (DAC1 and DAC2) operate in subthreshold mode; therefore, the ladders' implementation was mostly based on p-MOSFETs, for more thermal stability compared to n-MOSFETs; however, to implement better pull-down circuit, the foot of each ladder is an n-MOSFET, as shown in Fig. 4.

2) TRANSIENT ASPECTS

The DAC ladders are to be initiated before the charge/discharge controlling pulse, to assure stability when the discharge begins. The DAC ladders do not have any switching activity through the entire challenging time of the EC-PUF circuit. Therefore, using slow resistive MOSFETs does not set back the DAC performance. The pulse width time of the charge/discharge control signal depends on the microcontroller's internal clock frequency. On the other hand, for the MOSFET capacitors, even the time of 2 ns was enough to have the capacitors fully charged; however, most microcontrollers already run at much slower pulse rates. This makes the EC-PUF maintain its stability with most generic microcontrollers.

L. SECURITY ASPECTS

Similar chips are referred to as a class. They can be challenged by one 64-bit challenge, as the unique response of each chip is what matters most. However, for security measures, each CRP is safer to be used once only, to thwart MITM attacks. At each authentication attempt, the authorized authenticator randomly chooses a challenge among the archived CRP list, so an attacker cannot predict the upcoming challenge. In addition, the EC-PUF can have an S-box, as shown in Fig. 2. and Fig. 5. This optional S-box serves both *confusion* and *diffusion* purposes at the same time. Furthermore, a key can either be inserted by the smart card reader's keypad, stored in on-chip NVM, or received from the verifier side. It is also possible to relate that key to the encrypting key of the communication with the verifier side. A brute-force modeling attack should overcome the environmental variations of voltage, temperature, and frequency, besides the large number of possible response combinations, which is raised by the roulette-effect.

The communications between the PUF and the microcontroller are internal, and the communications with the verifier side over the internet are secured as well. Exploring the internal structure of the targeted chip, whether by basic optical attacks or by x-rays, is a vital step in planning a suitable approach to attack the targeted chip. However, knowing the internal structure still does not ensure accurate modeling due to the different doping rates and other complex factors of each MOSFET within the EC-PUF design. Besides, in general, it is difficult to model the environmental effects on mixed-signal PUFs. It is worth mentioning that to the best of the authors' knowledge, no successful attack on any mixed-signal PUF was reported until the writing time.

For the EC-PUF, a DPA attack will have to analyze a short charging phase of about 1 ns; furthermore, all CCs withdraw their charging currents simultaneously. It is true that the total current is correlated to the challenge-activated capacitors, but their distribution among the CCs is not known to an outsider unless the attacker can probe the individual current of each CC. This hardens the DPA attack unless achieving direct contacts for the microprobes to each CC. The physical access to the internal details of the EC-PUF chip implies an invasive attack, which is not feasible against the suggested active coating packaging for the EC-PUF chip.

The best bet for an attacker would be applying EM probing, measuring the eight different frequencies and the counters' stop times to reveal the NRC of all counters. However, among the 128 flip-flops, only 16 work simultaneously as a 16-bit

TABLE 2. Test framework.

Test	Objective	Metric
Temperature variation	Measuring the thermal effect on the response	Discharge time and digital response
Voltage variation	Measuring the voltage effect on the response	Discharge time and digital response
Monte Carlo simulation–local	Measuring the effect of local intrinsic variations on the discharge time of the capacitive cells (CCs)	Discharge time's coefficient of variation
Monte Carlo simulation–global	Measuring the effect of global intrinsic variations on the discharge time of the CCs	Discharge time's coefficient of variation
Monte Carlo simulation–total	Measuring the collective effect of the local and the global intrinsic variations on the discharge time	Discharge time's coefficient of variation

counter, and some counters may meet at the same rising edge at different times. This involves determining the physical wiring of each response bit, which can be more difficult if a designer arbitrarily wires the 128 response bits to the microcontroller. The VS is to be programmed to sort out the proper position of the received response bit among the eight RCs. The PUF-to-microcontroller data bus is internal and to be protected by an active coating; therefore, if an attacker tries to indicate the stop time of each counter, this requires removing the active coating and possibly some higher metal layers, which will destroy the EC-PUF system.

IV. EXPERIMENTAL SETUP

This section presents the tests we performed on the proposed EC-PUF design. We also discuss the basic verification approaches for the EC-PUF. In our simulation, we focused on the discharge time and RC of the smallest and largest CCs. Table 2 lists the related objective and metrics of each test. We carried out two categories of variation tests; environmental and intrinsic.

A. ENVIRONMENTAL VARIATION TESTS

The EC-PUF chip is intended to be embedded in smart cards and other portable devices that would work under various environmental conditions. Therefore, the effects of the variations in temperature and voltage were tested. The RCs of the smallest and the largest CCs were compared as part of the ECC elimination approach. The environmental variation tests carried out are categorized as follows:

1) THERMAL VARIATION TESTS

The thermal variation effect on the discharge times and the RCs of the smallest and largest CCs were simulated in the range (−55° to 125°) C.

2) VOLTAGE VARIATION TESTS

The effect of the supply voltage variation on the discharge times and the RCs of the smallest and the largest CCs was simulated in the range (1-1.2) V.

3) THERMAL AND VOLTAGE VARIATION TESTS

The collective effect of both thermal and voltage variations was measured to study the worst error case. This was to attain the maximum tolerance rate, which the VS should adopt to maintain a suitable FAR and FRR.

B. INTRINSIC VARIATION TESTS

Monte Carlo simulations were conducted to test the effect of the intrinsic randomness of the manufacturing process on the uniqueness of EC-PUF chips. That is by measuring the variance of their discharge times, which ultimately reflect on the EC-PUF chips' responses. We adopted the coefficient of variation (CV) of the discharge time as a metric for the uniqueness of the EC-PUF chip. The first Monte Carlo simulation was to measure the effect of the *global, inter-die, or process* variation among the EC-PUF chips within a silicon wafer on the discharge times of the CCs. The second simulation of Monte Carlo also added the *local, intra-die, or mismatch* variation among devices within each EC-PUF chip. This was to assess the effect of the total variations on the discharge times.

In both simulations, the discharge times of the smallest and largest CCs were measured to evaluate the influence of a CC's size on the uniqueness of the discharge time.

The CV of a discharge time is a generic parameter of an RC uniqueness. The frequency of the clock pulse to each counter is not of concern here, as it is mostly high enough to indicate even the minute differences in discharge times, and in the end, the concatenation of the eight distinct 16-bit RCs results in a unique 128-bit response for each EC-PUF chip.

V. RESULTS AND DISCUSSION

This section discusses the area of the layout and the average consumed power. The section also discusses the effect of simulations of temperature, voltage, and intrinsic variation on the time of discharge and the responses generated.

It would be erroneously assumed that a lower supply voltage would store less charge in the capacitors; therefore, a shorter discharge time would be expected. But in fact, the influence of the supply voltage on the two DAC R-2R converters is more significant. Therefore, supplying a lower supply voltage leads to longer discharge times for all eight CCs.

The longest discharge time was attained at a minimum temperature of −55° C. The discharge load controlling bits ch15−ch8 were set to a high voltage of 1V as logic 1s. Similarly, the comparator's controlling bits ch7 − ch0 were set to logic 1s; all the capacitors within that CC were enabled and charged by a minimum VDD of 1V. Under typical corners conditions for the MOSFETs, the average data processing time was 118 μs. To determine the maximum time which a microcontroller should wait before reading the RCs, the longest discharge time was determined using Monte Carlo simulation for the total intrinsic variations of 100 chip samples. The simulation showed that the largest CC required 715 μs to discharge, and this is the required time to generate its RC. It would, therefore, be safe to program a microcontroller timer to wait for about 750 μs before reading the 128-bit response.

The proposed design processes a 64-bit input challenge to generate a 128-bit output response. The design area is 22,470 μm2, and it consumes an average power of 921.67 μW. The analog portion consumed an average power of 131.5 μW, this includes the oscillator, Schmitt trigger, two DACs, and eight comparators. Charging all CCs consumed an average power of 67 nW, with a peak instantaneous current of 385.5 μA. The digital units, which include the frequency dividers, clock gating, and counters, consumed 792.4 μW.

The results and discussion are divided into environmental and intrinsic variations.

A. ENVIRONMENTAL VARIATION RESULTS

It is essential to study the effects of the environmental variations on the proposed EC-PUF chip. Under the environmental variations, the simulation results are divided into thermal and voltage variations.

1) THERMAL VARIATION RESULTS

To study the EC-PUF at its worst-case scenario of thermal variations, the basic EC-PUF design did not include any frequency stabilization. Fig. 6 shows that the central frequency (f) varies inversely to the temperature. However, it is assumed here that the supplied voltage is regulated to 1V at both the enrollment and the verification phases.

Fig. 7 shows the temperature variation effect on the discharge time and the decimal representation of the RRC. It is clear from Fig. 7 that the time-to-digital conversion produces an RC that is linear to the discharge time. Fig. 7 shows that the discharge time varies reversely to the temperature; however, the response value does not overflow counter-A. On the other hand, Fig. 8 shows that excess of t_B overflows counter-B; therefore, it acts as a digital roulette, and the value of RC-B varies even against the same applied L'challenge chunk at CC-B.

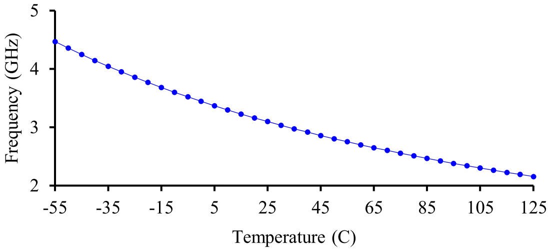


FIGURE 6. The central frequency (f) vs. the temperature.

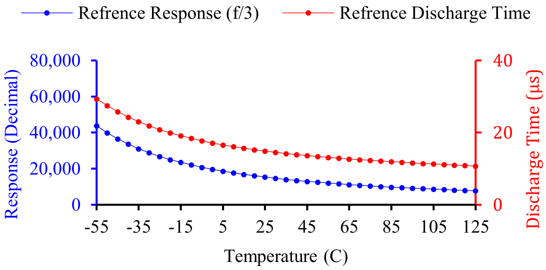


FIGURE 7. Discharge time and response of cell-A vs. the temperature.

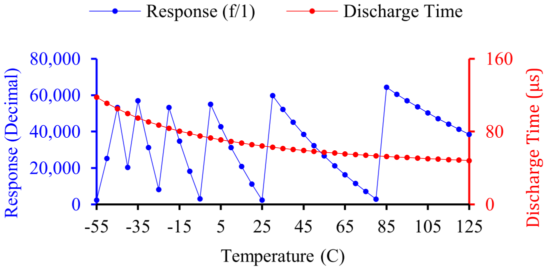


FIGURE 8. Discharge time and response of cell B vs. the temperature.

Fig. 9 shows the ratio of the discharge times' ratio (t_B/t_A) and Fig. 10 shows the ratio of the non-rouletted value of RC-B to RC-A, that is (NRC-B/RRC), which is denoted as (RR-B), both vary with the temperature. One of the reasons behind such inconsistency is that the generic comparator used in this design has an overdrive of 50 mV with a response time of 38 ns. This delay enables the related counter to count for an extended time, and with such a high frequency, this makes a significant difference in the counted value. Each different CC creates a different input impedance, which also varies with

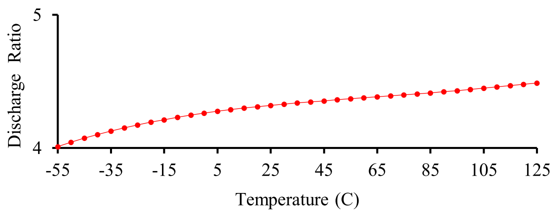


FIGURE 9. Discharge time ratio of cell-B to cell-A vs. the temperature.

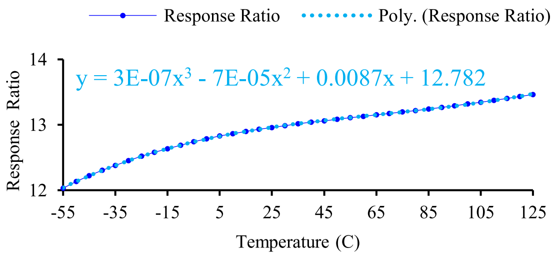


FIGURE 10. Response ratio of cell-B to cell-A vs. the temperature.

temperature, and impacts the transition slope from 1 V to 0 V of each comparator's output.

Each output represents an enable signal E for its related counter. With such high frequencies, even a slight change in the slope can impact the NRC of the counter. We studied the worst-case scenario with the largest CC, that is CC-B, which at certain conditions generates the largest NRC. Dividing this NRC by the RRC yields the largest RR, which is related to the smallest CC, that is labeled here as CC-A. We set all the bits of N challenge field to 1s to have the least V_{ref}, this is to achieve the maximum counted values related to t_A and t_B. We chose the challenging field L' (for both CCs) of all 0's to stored the maximum charge in CC-A and CC-B. We also chose an M challenging field of all ones; this makes all the discharge loads implemented by n-MOSFETs at their maximum impedance. When the discharge phase is initiated by the charging/discharging control signal, the impedance of each CC becomes in parallel to its related n-MOSFET discharge load. The total impedance of each CC with its parallel discharge n-MOSFET affects the input impedance of its related comparator. Having a larger n-MOSFET impedance gives the impedance of the CC more influence. Furthermore, since each CC consists of tens of parallel n-MOSFETs, this makes a CC have less impedance and more susceptibility to the thermal effects. This explains the variations in the ratio of discharge times (t_B/t_A) in Fig. 9 and RR-B in Fig. 10 when the temperature varies. However, the RR variation shrinks against other challenges that cause shorter discharge times and when the size difference between the CCs is less, as such conditions reduce the discharge time ratios, hence the RR.

Here we discuss two different verification handling approaches that a VS can apply to verify a CC like CC-B despite the thermal-driven shifts in its RC-B and NRC-B values, without employing an on-chip ECC unit.

a: RATIO-BASED VERIFICATION

We have simulated the enrollment phase at 25^o C and V_{DD} of 1 V for the challenge that causes the maximum discharge time. Counter-B generated a binary NRC-B which in decimal equals 197375. The authorized enroller knows that counter-B counted 197375 and shows a final count of 767. Alternatively, counter-A generated a 16-bit binary RRC, which equals 15232 in decimal, and since it is smaller than 65536, it did not roulette counter-A. In other words, counter-A did not apply a MOD function to the base 65536. Then RR-B is 12.9579.

During the enrollment phase, the seven RCs, NRCs, and their RRs besides the RRC are to be archived in a safe server where the enrollment data are only accessible by the authorized authenticator or a trusted third party, depending on the authentication protocol.

At a verification phase, under unknown temperature, the authenticator can assume that all the received RCs were shifted similarly, trend-wise but not quantity-wise, and all RCs are generated under similar, but not necessarily identical, environmental factors.

As an example of a verification phase at a different temperature, we have set the temperature to −10° C. We considered the worst-case scenario where the microcontroller does not convey the temperature to the verifier side. Counter-A gave 22090. Then the authenticator expects NRC-B to be around $\lceil 12.9579 \times 22090 \rceil$, that is 286240. After applying the MOD function to base 65536, this yields 24096. In the simulation, RC-B at counter-B was 18161. The difference between the expected and the generated values can be considered for that CRP as a tolerance margin for NRC-B.

The frequency was not stabilized; at 25^o C it was 3.09684 GHz, while at –10° C was 3.59,985 GHz; however, this should not affect the response rate, as the frequency is eliminated as in (7). A more extreme case was simulated under −55° C, counter-A had 43775, then the authenticator expects NRC-B of approximately $\lceil 12.9579 \times 43775 \rceil$, that is 567232, whereas counter-B had an RC-B of only 526675.

If the VS finds the shifting trends among the eight RCs consistent, such shifts can be attributed to natural causes rather than security threats. This comparison helps the VS to decide whether to accept or reject the entire 128-bit response. The consistency among the eight RCs also played a key role in eliminating the on-chip ECC unit.

Using the same RR of an NRC at different temperatures can save both enrollment and verification phases' time, cost, and computational resources. However, it requires the VS to adopt wider margins of error tolerance. This raises the FAR and lowers the FRR, which may still be acceptable depending on the authentication application, especially since analyzing the general shift trend among all NRCs plays a major role in the authentication process.

b: RATIO-BASED VERIFICATION GUIDED BY INDIVIDUAL TRENDS

As RRs such as RR-B vary with temperature, as shown in Fig. 10, a more precise approach requires archiving the individual trend of the RR and the individual trend of each RC at different temperatures. This implies:

- At the enrollment phase, each challenge is applied at several temperatures; this creates a set of different RRs for each CC. These data sets of all CCs are archived at the server. This enrollment phase is slow and requires controlling the thermal condition. It is possible to reduce the required number of measurements by choosing the most dominant temperatures for the targeted application. For example, for banking smart cards in Europe and North America, the trend of the RCs through temperatures (−25, 0, and 25)° C can make a good fit for smart card/RFID applications. The trend equation (y) of the RRC and NRC of each of the other seven chunks against the temperature (x) are driven and archived at the server. Fig. 11 shows the trends of chunk-A (RRC) and the non-rouletted value of chunk-B (NRC-B) against the temperature variation.

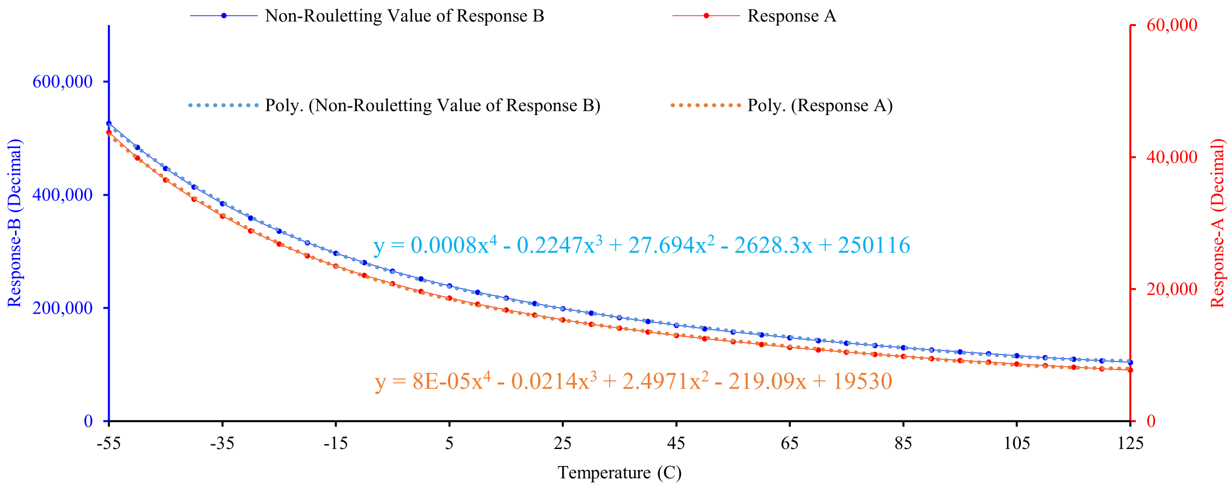


FIGURE 11. Discharge times and trends of the smallest and the largest CCs vs. the temperature.

The number of the measured temperature points was (N = 37). When relating the NRC-B data set to the RRC data set, this makes the degree of freedom (DF) equal to (N– 2 = 35). Pearson's correlation factor (r) was determined from the related data and is usually represented as r(35) = 0.9998. This correlation value represents a largely positive relationship between the two sets of responses. We also found that the statistical correlation coefficient, a.k.a. the probability value (P), has a value of 5.8998E-63. This extremely small P indicates a significant statistical correlation between the two response sets.

- At a verification phase, the VS can utilize the archived trends. Starting with the received value of RC-A, which is also the RRC, the VS substitutes the RRC value into y to find the temperature x. If RRC is out of that archived range of responses, that means the temperature at the EC-PUF is out of the archived range. In such cases, the VS can apply the extrapolation approach based on the closest archived RRC value to estimate the temperature x.
- Since all CCs are assumed to have the same temperature, the VS can substitute the calculated value of x into the trends of the other NRCs, like NRC-B. The VS applies the MOD function to the base 65536 to find RC-B, to compare this estimated value to the received one. Similarly, the VS can estimate all the other RCs.
- The VS can analyze the shift trends of all eight RCs to check their consistency, then ultimately decide whether to accept or reject the entire 128-bit response.
- The VS can also substitute the determined value of x into the trend equation, as shown in Fig. 10, to determine the RR at this temperature. If the value of x is out of the archived range, the VS can apply the extrapolation approach based on the closest archived x value to estimate the RR. This way, RR is determined at that specific temperature, which makes it more accurate. The VS can then apply the ratio-based verification approach explained in (a), but this time, the difference diminishes between the determined RC and the received one. This enables the VS to set finer margins of error tolerance and become more definite about the authenticity of the EC-PUF chip.

2) VOLTAGE VARIATION RESULTS

Voltage variation is another vital environmental variation factor that influences all the RCs. An increment in the supplied V_{DD} increases the oscillator's frequency, as shown in Fig. 12. The oscillator's frequency represents the central frequency for all frequency dividers. The enroller does not need to measure the frequency, as it is eliminated when determining the RR between the RR, that is RC-A, and each of the other seven RCs. Fig. 13 shows that the discharge time varies inversely to

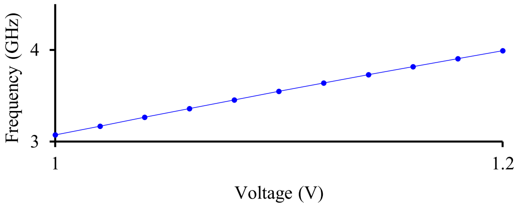


FIGURE 12. The central frequency (f) vs. the voltage.

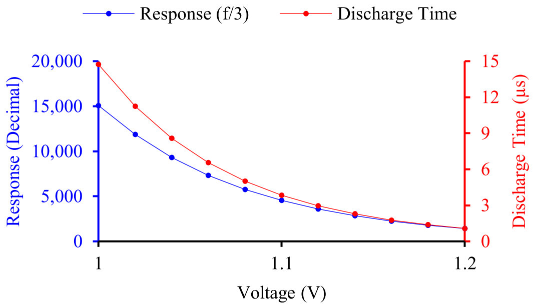


FIGURE 13. Discharge time and response of cell-A vs. the voltage.

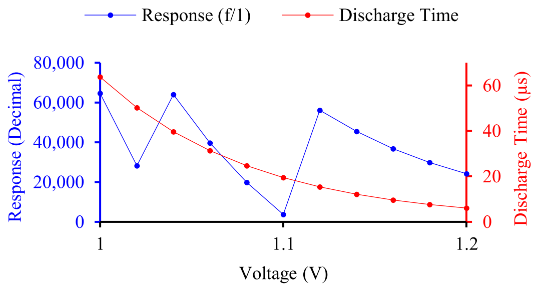


FIGURE 14. Discharge time and response of cell-B vs. the voltage.

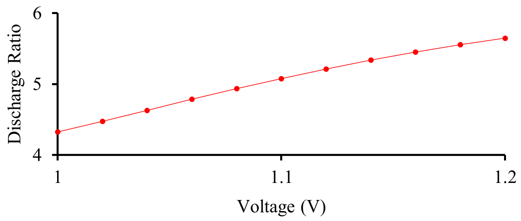


FIGURE 15. Discharge time ratio of cell-B to cell-A vs. the voltage.

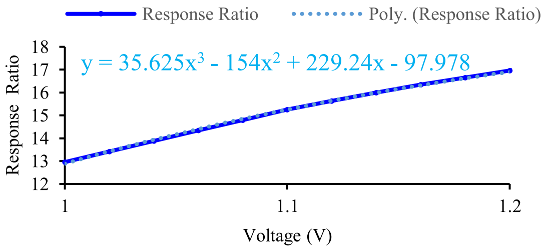


FIGURE 16. Response ratio of cell-B to cell-A vs. the voltage.

the voltage; however, the response value does not overflow counter-A. On the other hand, Fig. 14 shows that excess of t_B overflows counter-B; therefore, it acts as a digital roulette. Ideally, both the rate of discharge times (t_B/ t_A) in Fig. 15 and the rate of responses (NRC-B/ BRC) in Fig. 16 should be constant; however, both vary inversely to the voltage. The output slope for each comparator varies with the voltage, the CC size, and the structure of the networked capacitors within each CC. Increasing the supplied voltage significantly affects the eight discharge loads and the reference voltage at the comparator inputs, which are centrally controlled by DAC1 for all eight comparators. In comparison to the other analog components of the EC-PUF, the comparators are less affected by voltage variation, as their voltage is regulated by a central BGR circuit.

The VS can handle the voltage variation effects similar to the temperature variation. However, a large voltage variation is not such an inevitable factor, as precise voltage regulation is possible with circuits such as the BGR, which can regulate the supplied voltage with an error of just a few microvolts. Voltage regulation is mainly essential for the analog portion of the EC-PUF chip. On the other hand, the digital units such as the S-box, the eight AND gates, and the eight counters do not require such a tight voltage regulation.

We studied the worst-case considering the voltage supplied to the oscillator, the Schmitt trigger, the CCs, and the DAC ladders is not regulated. As a worst-case scenario for the voltage variation of 100 mV, we simulated a verification phase at a voltage of 1.1 V; cell-A generated 4525 counts. In this case, the VS presumes counter-B should have an NRC-B of $\lceil 12.9579 \times 4525 \rceil$, that is 58635 counts. Since it is less than 65536, the VS does not need to apply MOD function, or even if it does, it also gets 58635 counts. Through simulation, counter-B gave an NRC-B of 69035 counts. This difference of 10400 counts can be tolerated by the VS. We also studied a more extreme variant of 200 mV; at 1.2 V, counter-A had 1415 counts. Then the VS presumes that counter-B should have NRC-B of $\lceil 12.9579 \times 1426 \rceil$, that is 18478 counts. The simulation showed that counter-B gave an NRC-B of 24001 counts. This 5623 difference can be considered within the tolerance margin for NRC-B. It is obvious that increasing the voltage makes the discharge times shorter. This was the worst-case scenario where the voltage varies by 200 mV, and we studied the largest NRC. That was generated by the largest CC while being challenged with (L', M, N) which cause the largest discharge time.

Although the simulated EC-PUF design can function at the voltage range (1–1.2) V, we optimized it for 1 V and assumed the enrollment and the verification phases to be done in the range (1–1.1) V. This means the VS can set narrower tolerance margins for the NRCs than in the worst-case scenario studied above. The correlation between the response rates under various voltages is shown in Fig. 17. The number of the measured voltage points was (11). When relating NRC-B data set to RRC data set, this makes DF equal to (9), and r(9)= 0.9989. This correlation value represents a largely positive relationship between the two sets of responses. We also found that P has a value of 2.47E-13. This extremely small P indicates a significant statistical correlation between the two response sets.

The total shift in the value of any NRC depends on both temperature and voltage shifts. It is clear from Fig. 9 and Fig. 15 that the effects of temperature and voltage on the discharge time oppose each other. The resulted response is also influenced by the frequency if it is not stabilized. Also, Fig. 6 and Fig. 12 show that the temperature and the voltage have opposing effects on the central frequency (f). The influence of the frequency is eliminated anyway, as in (7). The inaccuracy of the generic comparators used in the design makes the discharge time ratio of cell-B to cell-A increase with both temperature and voltage variations, therefore the VS should set a margin of tolerance when analyzing the response at a verification phase.

Even when considering the worst-case scenario which is when neither the temperature nor the voltage is conveyed to the verifier side, and neither the voltage nor frequency is firmly stabilized, the proposed EC-PUF is still able to handle these variations, and the VS can authenticate the chip.

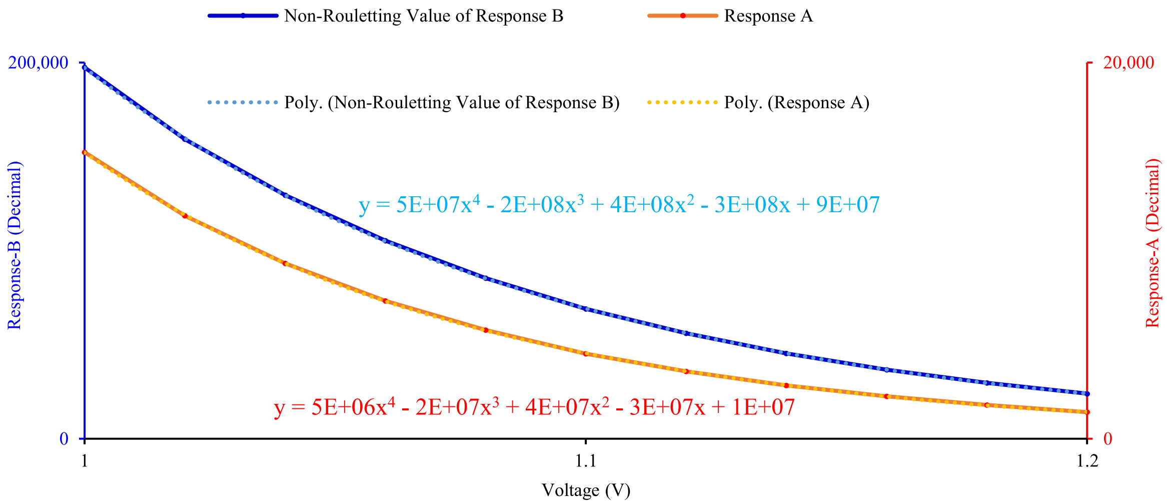


FIGURE 17. Discharge times and trends of the smallest and the largest CCs vs. the voltage.

B. INTRINSIC VARIATION RESULTS

The intrinsic fabrication variations were simulated to estimate the randomness contribution to the individuality of each EC-PUF chip if it is fabricated by a typical MOSFET technology. 100 hundred chips were generated and tested using Monte Carlo simulation guided by the generic process design kit gpdk045 for oxide-polysilicon gated planar MOSFET with BSIM4- model. Throughout the simulation, a supply voltage of 1V and a temperature of 25° C were considered. The intrinsic variation is usually divided into global and local variations. For the EC-PUF chip, Monte Carlo simulation was run for the global and total variations.

1) LOCAL INTRINSIC VARIATION

The effects of the local variations of the largest CC (cell-B) and the smallest CC (cell-A) were simulated. An input challenge was applied to both CCs. The resulted discharge time distributions of both CCs are shown in Fig. 18 and Fig. 19, respectively. The discharge time t_B in Fig. 18 is to be digitized by a counter runs at a frequency f/1; similarly, the discharge time t_A in Fig. 19 is to be sampled by a frequency f/3. For cell-B, Fig. 18 shows that the discharge time's mean of 100 chip samples against an arbitrary challenge was 68.1642 μs, and the standard deviation was 17.8477 μs. Fig. 19 illustrates the discharge time distribution of cell-A; its mean was 17.255 μs, and its standard deviation was 4.95769 μs. When we compared the standard deviation to the mean of each CC, we found that the largest CC had a CV of 26.18%, compared with 28.73% for the smallest CC.

2) GLOBAL INTRINSIC VARIATION

The effects of the global variations of the largest CC (cell-B) and the smallest CC (cell-A) were simulated. Fig. 20 shows that the discharge time's mean of 100 chip samples against

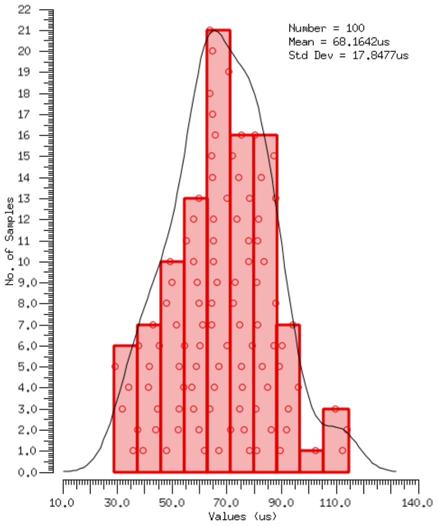


FIGURE 18. Simulated effect of the local randomness on the discharge time of the largest capacitive cell.

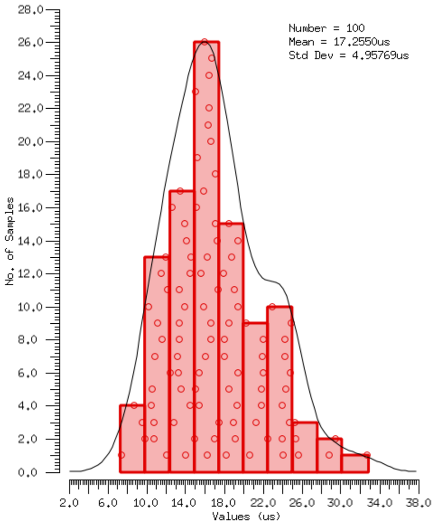


FIGURE 19. Simulated effect of the local randomness on the discharge time of the smallest capacitive cell.

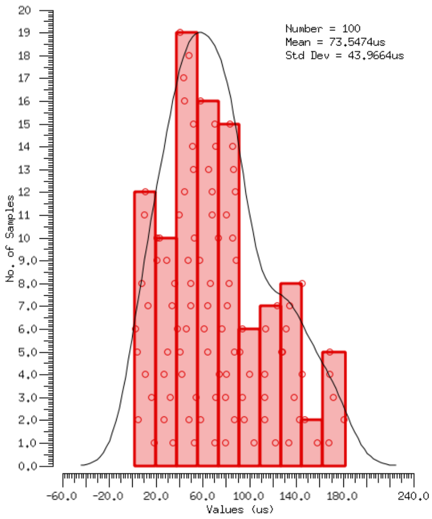


FIGURE 20. Simulated effect of the global randomness on the discharge time of the largest capacitive cell.

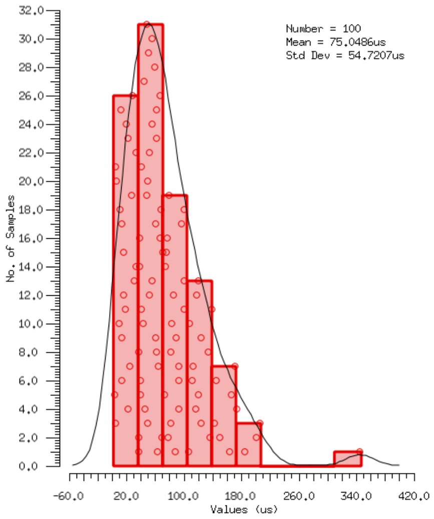


FIGURE 22. Simulated effect of the total randomness on the discharge time of the largest capacitive cell.

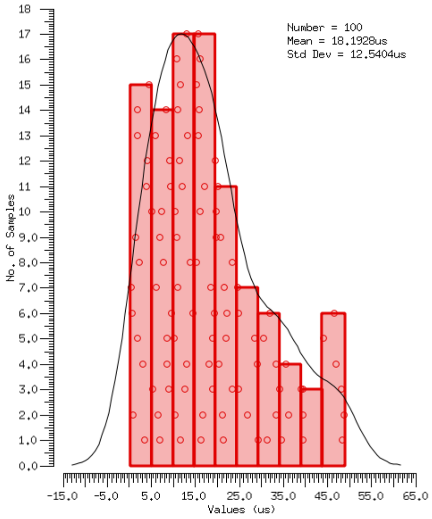


FIGURE 21. Simulated effect of the global randomness on the discharge time of the smallest capacitive cell.

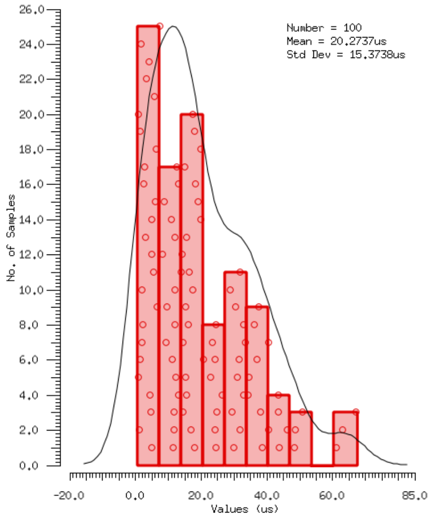


FIGURE 23. Simulated effect of the total randomness on the discharge time of the smallest capacitive cell.

an arbitrary challenge was 73.5474 μs and the standard deviation was 43.9664 μs. Fig. 21 illustrates the discharge time distribution of CC-A; its mean was 18.1928 μs, and its standard deviation was 12.5404 μs. When we compared the standard deviation to the mean of each CC, we found that the largest CC has a CV of 59.78%, compared with 68.93% for the smallest CC.

3) TOTAL INTRINSIC VARIATION

The local variation was added to the global one to assess the total variation of the discharge time of 100 simulated EC-PUF chips against the same challenge. Fig. 22 and Fig. 23 show the discharge time variation for the largest and the smallest CC, respectively. The randomness simulation exhibited the mean discharge time of the largest CC as 75.0486 μs, and its standard deviation was 54.7207 μs. The smaller CC yielded 20.2737 μs and 15.3738 μs, respectively. The collective anticipated CV of both the mismatch and the total variations is then 72.91% for the largest CC and 75.83% for the smallest CC. The high variation in discharge times shows the validity of the proposed EC-PUF design, as applying a higher clock pulse frequency at a counter means a higher sample rate for the time-to-digital conversion, which can more precisely indicate the physical uniqueness of the analog units related to that counter. The random variation in a discharge time of a CC comply with the theoretical concept about the relation between the size of a MOSFET-based object versus the random variations in its physical structure, as the smallest CC showed more CV. The CV values also exhibit the far greater impact of the global intrinsic variations compared to the local variations. This can help draw a road map to guide the variation-aware layout and manufacturing processes to further enhance the uniqueness of the EC-PUF chips.

VI. CONCLUSION

In this article, various design, layout, and manufacturing aspects were discussed in order to acquire more intrinsic variations throughout a MOSFET-based manufacturing process to form distinctive properties for a MOSFET-based PUF chip.

The proposed EC-PUF design includes several power-saving techniques. It also reduces the effects of the environmental variations of the temperature and the supply voltage on the generated response chucks (RCs).

The proposed EC-PUF is based on the discharge time delay of networked MOSFET-based capacitors. A high-frequency oscillator drives eight frequency dividers, which run the counters. A higher frequency at a counter means a higher sampling rate, which makes a more precise time-to-digital conversion, represented by an RC. This helps make an EC-PUF chip generate distinct RCs among other EC-PUF chips, even when only meager intrinsic differences are obtained throughout the fabrication process.

The capacitive nature of the EC-PUF has pros and cons. The CCs are interactive to the metal comb arrays and to the random post-process sprayed coating. This can inherently increase the intrinsic randomness and the chip uniqueness, optically oppose the basic optical inspecting attacks, and electrically shield the chip against the invasive attacks. Other than the verification phase time, the CCs can be utilized for other purposes, such as decoupling and current flattening applications to counter the DPA attacks.

On the other hand, the capacitive nature of the EC-PUF causes environmental-based constrains, which imply that the VS should adopt a broad margin of tolerance at the verification phase.

Security-wise, a large CC and a high-frequency clock pulse can cause an overflow in a counter, which then acts as a roulette wheel. This helps the unpredictability of an RC against the modeling attacks. Furthermore, the confusion and diffusion concepts of cryptography can be enhanced further by including a custom-made S-box to raise the security level against the modeling attacks. The on-chip microcontroller encrypts the data using conventional security schemes.

Decrypting the chip-to-verifier communications does not directly lead the adversary to model the PUF, as the 128 bits of the counters are hard-wired in a secret sequence to the microcontroller; therefore, an invasive attack would still be needed to disclose the physical meaning of each bit.

Environmental-wise, parameters of voltage and temperature can be measured from on-chip sensors and conveyed to the verifier software (VS) at the remote verifier side. However, since in this mixed-signal PUF design, errors are not in the form of randomly flipped bits, but in the form of shifts in the digital representations of the discharge times of all the capacitive cells (CCs), and since the 128-bit response will always include at least one non-rouletted reference response chunk (RRC), then even without measuring those environmental parameters, the VS can analyze the collective shift caused by those parameters and verify the EC-PUF with a margin of tolerance.

Dividing the response-generating process of the 128-bit over eight clusters helps the remote VS to handle the environmental shifts, control the acceptance/rejection tolerance margins of the responses, detect the odd shifts which would be attributed to invasive attacks, and eliminate the need for an on-chip error-correcting code (ECC) unit.

Monte Carlo randomness simulated the fabrication variations of 100 EC-PUF chip samples. The global variations have shown much more impact on the discharge time of a CC than the local variations. Using a high frequency to digitize the discharge time improves the time-to-digital converting accuracy. The fabrication variations within the embedded ring oscillator will add more uniqueness to the generated response.

The design was simulated under temperatures in the range ($-55°$ to $125°$) C and supplied voltages in the range (1-1.2) V. The design was simulated using 45 nm CMOS technology with a chip area of 22,470 μm^2. The average response time was 118 μs, measured at 1 V, 25° C, and typical corners for both n and p-MOSFETs. The average power was 921.67 μW.

Based on the area, average power, and compatibility with a range of voltages, temperatures, and microcontrollers, the proposed design is a feasible solution for authentication and anti-counterfeiting
applications.

APPENDIX

The layout of the proposed EC-PUF chip is shown in Fig. 24. The dimensions are 149.899 μm × 149.899 μm.

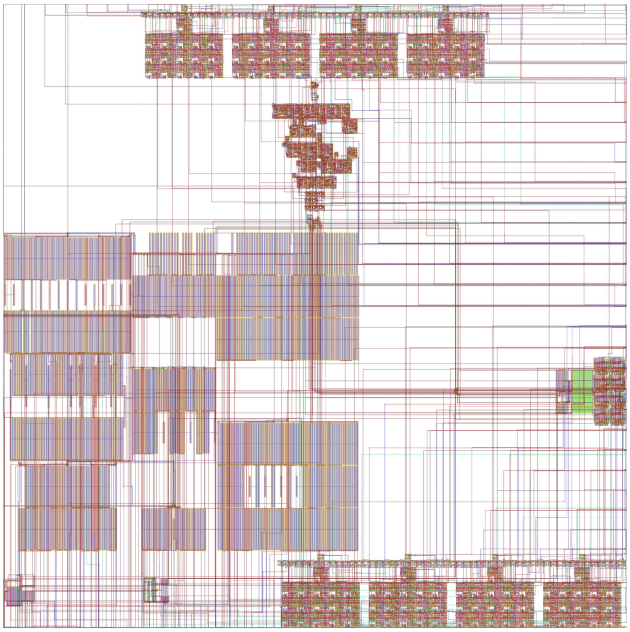


FIGURE 24. Layout of the EC-PUF chip.

REFERENCES

[1] L. Bolotnyy and G. Robins, "Physically unclonable function-based security and privacy in RFID systems," in *Proc. 5th Annu. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2007, pp. 211–220.

[2] P. H. Cole and D. C. Ranasinghe, *Networked RFID Systems and Lightweight Cryptography: Raising Barriers to Product Counterfeiting*. Berlin, Germany: Springer-Verlag, 2008.

[3] S. Devadas, E. Suh, S. Paral, R. Sowell, T. Ziola, and V. Khandelwal, "Design and implementation of PUF-based 'unclonable' RFID ICs for anti-counterfeiting and security applications," in *Proc. IEEE Int. Conf. RFID*, Apr. 2008, pp. 58–64.

[4] S. Ahson and M. Ilyas, Eds., *RFID Handbook: Applications, Technology, Security, and Privacy*. Boca Raton, FL, USA: CRC Press, 2008.

[5] F. Skopik and P. Smith, Eds., *Smart Grid Security: Innovative Solutions for a Modernized Grid*. Amsterdam, The Netherlands: Elsevier, 2015.

[6] K. Markantonakis and K. Mayes, Eds., *Secure Smart Embedded Devices, Platforms and Applications*. New York, NY, USA: Springer, 2014.

[7] C. Marchand, L. Bossuet, U. Mureddu, N. Bochard, A. Cherkaoui, and V. Fischer, "Implementation and characterization of a physical unclonable function for IoT: A case study with the TERO-PUF," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 37, no. 1, pp. 97–109, Jan. 2018.

[8] K. Yang, D. Yang, and M. Tehranipoor, "ReSC: An RFID-enabled solution for defending IoT supply chain," *ACM Trans. Design Autom. Electron. Syst.*, vol. 23, no. 3, p. 29, 2018.

[9] B. Halak, M. Zwolinski, and M. S. Mispan, "Overview of PUF-based hardware security solutions for the Internet of Things," in *Proc. IEEE 59th Int. Midwest Symp. Circuits Syst.*, Oct. 2016, pp. 1–4.

[10] H. Ferradi, R. Géraud, D. Naccache, and A. Tria, "When organized crime applies academic results: A forensic analysis of an in-card listening device," *J. Cryptograph. Eng.*, vol. 6, no. 1, pp. 49–59, Apr. 2016.

[11] A. Barenghi, L. Breveglieri, I. Koren, and D. Naccache, "Fault injection attacks on cryptographic devices: Theory, practice, and countermeasures," in *Proc. IEEE*, vol. 100, no. 11, pp. 3056–3076, Apr. 2012.

[12] R. Maes, *Physically Unclonable Functions: Constructions, Properties and Applications*. Berlin, Germany: Springer-Verlag, 2013.

[13] A.-R. Sadeghi and D. Naccache, Eds., *Towards Hardware-Intrinsic Security: Foundations and Practice*. Berlin, Germany: Springer-Verlag, 2010.

[14] A. Maiti, L. McDougall, and P. Schaumont, "The impact of aging on an fpga-based physical unclonable function," in *Proc. IEEE Int. Conf. Field Program. Logic Appl.*, Sep. 2011, pp. 151–156.

[15] D. Ganta and L. Nazhandali, "Study of IC aging on ring oscillator physical unclonable functions," in *Proc. IEEE 15th Int. Symp. Qual. Electron. Design*, Mar. 2014, pp. 461–466.

[16] I. Verbauwhede and R. Maes, "Physically unclonable functions: Manufacturing variability as an unclonable device identifier," in *Proc. 21st Great Lakes Symp. VLSI*, May 2011, pp. 455–460.

[17] T. Md Rahman, D. Forte, J. Fahrny, and M. Tehranipoor, "ARO-PUF: An aging-resistant ring oscillator PUF design," in *Proc. Design, Autom. Test Eur. Conf. Exhib.*, Mar. 2014, pp. 69–74.

[18] M. T. Rahman, F. Rahman, D. Forte, and M. Tehranipoor, "An aging-resistant RO-PUF for reliable key generation," *IEEE Trans. Emerg. Topics Comput.*, vol. 4, no. 3, pp. 335–348, Jul./Sep. 2016.

[19] L. Tebelmann, M. Pehl, and G. Sigl, "EM side-channel analysis of BCH-based error correction for PUF-based key generation," in *Proc. Workshop Attacks Solutions Hardw. Secur.*, 2017, pp. 43–52.

[20] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *Proc. 44th Annu. Design Autom. Conf.*, 2007, pp. 9–14.

[21] Y. Lao, B. Yuan, C. H. Kim, and K. K. Parhi, "Reliable PUF-based local authentication with self-correction," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 36, no. 2, pp. 201–213, Feb. 2017.

[22] P. Koeberl, J. Li, A. Rajan, and W. Wu, "Entropy loss in PUF-based key generation schemes: The repetition code pitfall," in *Proc. IEEE Int. Symp. Hardw.-Oriented Secur. Trust*, May 2014, pp. 44–49.

[23] M.-D. Yu, D. M'Raihi, R. Sowell, and S. Devadas, "Lightweight and secure PUF key storage using limits of machine learning," in *Proc. Int. Workshop Cryptograph. Hardw. Embedded Syst.*, 2011, pp. 358–373.

[24] D. Puntin, S. Stanzione, and G. Iannaccone, "CMOS unclonable system for secure authentication based on device variability," in *Proc. 34th Eur. Solid-State Circuits Conf.*, Sep. 2008, pp. 130–133.

[25] K. Kamal and R. Muresan, "Capacitive physically unclonable function," in *Proc. IEEE 30th Can. Conf. Elect. Comput. Eng.*, Apr. 2017, pp. 1–6.

[26] M. Dietrich and J. Haase, Eds., *Process Variations and Probabilistic Integrated Circuit Design*. New York, NY, USA: Springer, 2012.

[27] T. Xu and M. Potkonjak, "Digital bimodal functions and digital physical unclonable functions: Architecture and applications," in *Secure System Design and Trustable Computing*, C. Chang and M. Potkonjak, Eds. Cham, Switzerland: Springer, 2016, pp. 83–113.

[28] J. Kong, F. Koushanfar, P. K. Pendyala, A.-R. Sadeghi, and C. Wachsmann, "PUFatt: Embedded platform attestation based on novel processor-based PUFs," in *Proc. 51st Annu. Design Autom. Conf.*, 2014, pp. 1–6.

[29] J. H. Anderson, "A PUF design for secure FPGA-based embedded systems," in *Proc. Asia South Pacific Design Autom. Conf.*, 2010, pp. 1–6.

[30] T. Xu, J. B. Wendt, and M. Potkonjak, "Secure remote sensing and communication using digital PUFs," in *Proc. 10th ACM/IEEE Symp. Archit. Netw. Commun. Syst.*, Oct. 2014, pp. 173–184.

[31] J. X. Zheng and M. Potkonjak, "A digital PUF-based IP protection architecture for network embedded systems," in *Proc. 10th ACM/IEEE Symp. Archit. Netw. Commun. Syst.*, Oct. 2014, pp. 255–256.

[32] S. U. Hussain, S. Yellapantula, M. Majzoobi, and F. Koushanfar, "BIST-PUF: Online, hardware-based evaluation of physically unclonable circuit identifiers," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design*, Nov. 2014, pp. 162–169.

[33] D. P. Sahoo, S. Saha, D. Mukhopadhyay, R. S. Chakraborty, and H. Kapoor, "Composite PUF: A new design paradigm for physically unclonable functions on FPGA," in *Proc. IEEE Int. Symp. Hardw.-Oriented Secur. Trust*, 2014, pp. 50–55.

[34] M. Majzoobi and F. Koushanfar, "Time-bounded authentication of FPGAs," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 1123–1135, Sep. 2011.

[35] K. Lofstrom, W. R. R. Daasch, and D. Taylor, "IC identification circuit using device mismatch," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2000, pp. 372–373.

[36] K. Lofstrom, "System for providing an integrated circuit with a unique identification," U.S. Patent 6 161 213 A, Dec. 12, 2000.

[37] K. Lofstrom, "ICID—A robust, low cost integrated circuit identification method," SiidTech, Richmond, BC, Canada, Tech. Rep., Mar. 2007. [Online]. Available: http://www.siitech.com

[38] S. Stanzione and G. Iannaccone, "Silicon physical unclonable function resistant to a 10^{25}-trial brute force attack in 90 nm CMOS," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2009, pp. 116–117.

[39] K.-H. Chuang, R. Degraeve, A. Fantini, G. Groeseneken, D. Linten, and I. Verbauwhede, "A cautionary note when looking for a truly reconfigurable resistive RAM PUF," in *Proc. Int. Assoc. Cryptologic Res. Trans. Cryptograph. Hardw. Embedded Syst.*, no. 1, 2018, pp. 98–117.

[40] Y. Gao, H. Ma, S. F. Al-Sarawi, D. Abbott, and D. C. Ranasinghe, "PUF-FSM: A controlled strong PUF," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 37, no. 5, pp. 1104–1108, May 2018.

[41] B. Gassend, M. van Dijk, D. Clarke, E. Torlak, S. Devadas, and P. Tuyls, "Controlled physical random functions and applications," *ACM Trans. Inf. Syst. Secur.*, vol. 10, no. 4, pp. 15:1–15:22, 2008.

[42] M. Montgomery, A. Ali, and K. Lu, "Secure network card: Implementation of a standard network stack in a smart card," in *Proc. FIP 18th World Comput. Congr. TC8/WG8.8 TC11/WG11.2 6th Int. Conf. Smart Card Res. Adv. Appl.*, 2004, pp. 193–208.

[43] R. J. Anderson and M. Kuhn, "Tamper resistance—A cautionary note," in *Proc. 2nd USENIX Workshop Electron. Commerce*, 1996, pp. 1–11.

[44] O. Koemmerling and M. G. Kuhn, "Design principles for tamper-resistant smartcard processors," in *Proc. USENIX Workshop Smartcard Technol.*, 1999, p. 13.

[45] D. Nedospasov, J. P. Seifert, C. Helfmeier, and C. Boit, "Invasive PUF analysis," in *Proc. 10th Workshop Fault Diagnosis Tolerance Cryptogr.*, 2013, pp. 30–38.

[46] I. Verbauwhede, Ed., *Secure Integrated Circuits and Systems*. Springer, 2010.

[47] M. Tehranipoor and C. Wang, Eds., *Introduction to Hardware Security and Trust*. Springer, 2012.

[48] C. Helfmeier, C. Boit, D. Nedospasov, and J. P. Seifert, "Cloning physically unclonable functions," in *Proc. IEEE Int. Symp. Hardw.-Oriented Secur. Trust*, Jun. 2013, pp. 1–6.

[49] R. Posch, "Protecting devices by active coating: A method to build a signature based microsafe," *J. Universal Comput. Sci.*, vol. 4, no. 7, pp. 652–668, 1998.

[50] P. Tuyls, G.-J. Schrijen, B. Škorić, J. van Geloven, N. Verhaegh, and R. Wolters, "Read-proof hardware from protective coatings," in *Proc. Int. Workshop Cryptograph. Hardw. Embedded Syst. (CHES)*, Berlin, Germany, 2006, pp. 369–383.

[51] B. Škorić, S. Maubach, T. Kevenaar, and P. Tuyls, ''Information-theoretic analysis of capacitive physical unclonable functions,'' *J. Appl. Phys.*, vol. 100, no. 2, pp. 24902–24911, May 2006.

[52] D. Roy, J. H. Klootwijk, N. A. M. Verhaegh, H. H. A. J. Roosen, and R. A. M. Wolters, ''Comb capacitor structures for measurement of post-processed layers,'' in *Proc. IEEE Conf. Microelectron. Test Struct.*, Mar. 2008, pp. 205–209.

[53] D. Roy, J. H. Klootwijk, N. A. M. Verhaegh, H. H. A. J. Roosen, and R. A. M. Wolters, ''Comb capacitor structures for on-chip physical uncloneable function,'' *IEEE Trans. Semicond. Manuf.*, vol. 22, no. 1, pp. 96–102, Feb. 2009.

[54] H. P. Duan and K. Peng, ''Physical unclonable function (PUF) chip and fabrication method thereof,'' U.S. Patent 2018 0 181 775 A1, Jun. 28, 2018.

[55] I. Verbauwhede, D. Karaklajic, and J.-M. Schmidt, ''The fault attack jungle—A classification model to guide you,'' in *Proc. Workshop Fault Diagnosis Tolerance Cryptogr.*, 2011, pp. 3–8.

[56] J. Delvaux and I. Verbauwhede, ''Fault injection modeling attacks on 65 nm arbiter and RO sum PUFs via environmental changes,'' *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 6, pp. 1701–1713, Jun. 2014.

[57] R. Korkikian, ''Side-channel and fault analysis in the presence of countermeasures: Tools, theory, and practice,'' Ph.D. dissertation, Math. Sci., Paris Center, Paris, France, 2016.

[58] R. J. Anderson and M. Kuhn, ''Low cost attacks on tamper resistant devices,'' in *Proc. 5th Int. Workshop Secur. Protocols*, 1997, pp. 125–136.

[59] B. Gassend, D. E. Clarke, M. Van Dijk, and S. Devadas, ''Controlled physical random functions,'' in *Security With Noisy Data: On Private Biometrics, Secure Key Storage Anti-Counterfeiting*, P. Tulys, B. Skoric, and T. Kevenaar, Eds. London, U.K.: Springer, 2007, pp. 235–253.

[60] J. Delvaux, ''Machine-learning attacks on PolyPUFs, OB-PUFs, RPUFs, LHS-PUFs, and PUF–FSMs,'' *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 8, pp. 2043–2058, Jan. 2019.

[61] S. Skorobogatov, ''Using optical emission analysis for estimating contribution to power analysis,'' in *Proc. Workshop Fault Diagnosis Tolerance Cryptogr.*, Sep. 2009, pp. 111–119.

[62] A. Mahmoud, U. Rührmair, and M. Majzoobi, ''Combined modeling and side channel attacks on strong PUFs,'' IACR Cryptol. ePrint, Lyon, France, Tech. Rep. 2013/632, 2013.

[63] U. Rührmair, X. Xu, J. Sölter, A. Mahmoud, M. Majzoobi, F. Koushanfar, and W. Burleson, ''Efficient power and timing side channels for physical unclonable functions,'' in *Proc. Int. Workshop Cryptograph. Hardw. Embedded Syst.*, 2014, pp. 476–492.

[64] M. Mayhew and R. Muresan, ''Implementation of a decoupling based power analysis attack countermeasure,'' *IET Circuits, Devices Syst.*, vol. 10, no. 6, pp. 528–535, 2016.

[65] R. Muresan and M. Mayhew, ''On-chip decoupling architecture with variable nMOS gate capacitance for security protection,'' in *Proc. IEEE 56th Int. Midwest Symp. Midwest Symp. Circuits Syst.*, Aug. 2013, pp. 1342–1345.

[66] M. Mayhew and R. Muresan, ''Modeling the effect of NMOS gate capacitance in an on-chip decoupling capacitor PAA countermeasure,'' in *Proc. IEEE 57th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2014, pp. 121–124.

[67] M. Mayhew, ''Design of an on-chip power analysis attack countermeasure incorporating a randomized switch box,'' Ph.D. dissertation, Dept. Eng., Univ. Guelph, Toronto, ON, Canada, 2016.

[68] M. Mayhew and R. Muresan, ''An overview of hardware-level statistical power analysis attack countermeasures,'' *J. Cryptogr. Eng.*, vol. 7, no. 3, pp. 213–244, Sep. 2017.

[69] D. L. Pulfrey, *Understanding Modern Transistors and Diodes*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[70] M. Fulde, *Variation Aware Analog and Mixed-Signal Circuit Design in Emerging Multi-Gate CMOS Technologies*. Springer, 2010. doi: 10.1007/978-90-481-3280-5.

[71] T. Yannis, *Mixed Analog-Digital VLSI Devices and Technology*. London, U.K.: Oxford Univ. Press, 2002.

[72] T. Ytterdal, Y. Cheng, and T. Fjeldly, *Device Modeling for Analog and RF CMOS Circuit Design*. Hoboken, NJ, USA: Wiley, 2003.

[73] T. C. Carusone, D. A. Johns, and K. W. Martin, *Analog Integrated Circuit Design*, 2nd ed. Hoboken, NJ, USA: Wiley, 2012.

[74] O. Kononchuk and B.-Y. Nguyen, *Silicon-on-Insulator (SOI) Technology: Manufacture and Applications*. Waltham, MA, USA: Woodhead Publishing, 2014.

[75] E. Salman and E. G. Friedman, *High Performance Integrated Circuit Design*. New York, NY, USA: McGraw-Hill, 2012.

[76] A. Balasinski, *Design for Manufacturability: From 1D to 4D for 90–22 nm Technology Nodes*. New York, NY, USA: Springer, 2014.

[77] H. Kaeslin, *Digital Integrated Circuit Design: From VLSI Architectures to CMOS Fabrication*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[78] C. C. Hu, *Modern Semiconductor Devices for Integrated Circuits*. London, U.K.: Pearson, 2010.

[79] C. Chiang and J. Kawa, *Design for Manufacturability and Yield for Nano-Scale CMOS*. Dordrecht, The Netherlands: Springer, 2007.

[80] R. J. Anderson, M. Bond, J. Clulow, and S. Skorobogatov, ''Cryptographic processors—A survey,'' *Proc. IEEE*, vol. 94, no. 2, pp. 357–369, Jan. 2006.

[81] A. Asenov, ''Statistical nano CMOS variability and its impact on SRAM,'' in *Extreme Statistics in Nanoscale Memory Design* (Integrated Circuits and Systems), A. Singhee and R. A. Rutenbar, Eds. New York, NY, USA: Springer, 2010, pp. 17–49.

[82] C. Bohm and M. Hofer, *Physical Unclonable Functions in Theory and Practice*. New York, NY, USA: Springer, 2013.

[83] P. K. Ko, T. Y. Chan, A. T. Wu, and C. Hu, ''The effects of weak gate-to-drain (source) overlap on MOSFET characteristics,'' in *IEDM Tech. Dig.*, vol. 32, Dec. 1986, pp. 292–295.

[84] Y. Hori, H. Kang, T. Katashita, A. Satoh, S. Kawamura, and K. Kobara, ''Evaluation of physical unclonable functions for 28-nm process field-programmable gate arrays,'' *J. Inf. Process.*, vol. 22, no. 2, pp. 344–356, 2014.

[85] A. Marshall, *Mismatch and Noise in Modern IC Processes*. San Rafael, CA, USA: Morgan & Claypool, 2009.

[86] P. R. Kinget, ''Device mismatch and tradeoffs in the design of analog circuits,'' *IEEE J. Solid-State Circuits*, vol. 40, no. 6, pp. 1212–1224, Jun. 2005.

[87] S.-M. Kang, Y. Leblebici, and C. Kim, *CMOS Digital Integrated Circuits: Analysis and Design*, 4th ed. New York, NY, USA: McGraw-Hill, 2017.

[88] D. K. Bhattacharya and R. Sharma, *Solid State Electronic Devices*, 2nd ed. London, U.K.: Oxford Univ. Press, 2013.

[89] A. Hastings, *The Art of Analog Layout*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.

[90] A. Asenov, A. Cathignol, B. Cheng, K. P. McKenna, A. R. Brown, A. L. Shluger, D. Chanemougame, K. Rochereau, and G. Ghibaudo, ''Origin of the asymmetry in the magnitude of the statistical variability of n- and p-channel poly-Si gate bulk MOSFETs,'' *IEEE Electron Device Lett.*, vol. 29, no. 8, pp. 913–915, Aug. 2008.

[91] F. Maloberti, *Analog Design for CMOS VLSI Systems*. Norwell, MA, USA: Kluwer, 2003.

[92] U. Chatterjee, V. Govindan, R. Sadhukhan, D. Mukhopadhyay, and R. S. Chakraborty, ''PUF+IBE: Blending physically unclonable functions with identity based encryption for authentication and key exchange in IoTs,'' IACR, Lyon, France, Tech. Rep. 2017/422, 2017, p. 15.

[93] B. K. K. Reddy and B. I. Reddy, ''A comparative analysis of various multifactor authentication mechanisms,'' *Int. J. Sci. Res. Comput. Sci., Eng. Inf. Technol.*, vol. 3, no. 5, pp. 1098–1103, 2018.

[94] S. M. Bellovin, ''Frank Miller: Inventor of the one-time pad,'' *Cryptologia*, vol. 35, no. 3, pp. 203–222, 2011.

[95] C. Paar and J. Pelzl, *Understanding Cryptography: A Textbook for Students and Practitioners*. Berlin, Germany: Springer-Verlag, 2010.

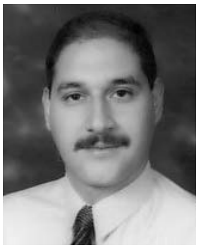
[96] *LPC8N04: 32-Bit ARM Cortex-M0+ Microcontroller; 32 kB Flash and 8 kB. SRAM; NFC/RFID ISO 14443 Type A Interface*, NXP Semicond., Eindhoven, The Netherlands, 2018.

[97] M. A. Usmani, S. Keshavarz, E. Matthews, L. Shannon, R. Tessier, and D. E. Holcomb, ''Efficient PUF-based key generation in FPGAs using per-device configuration,'' *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 27, no. 2, pp. 364–375, Nov. 2019.

[98] D. Forte and A. Srivastava, ''On improving the uniqueness of silicon-based physically unclonable functions via optical proximity correction,'' in *Proc. 49th Annu. Design Automat. Conf.*, 2012, pp. 96–105.

[99] D. Forte and A. Srivastava, ''Improving the quality of delay-based PUFs via optical proximity correction,'' *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 32, no. 12, pp. 1879–1891, Dec. 2013.

[100] M.-B. Lin, *Introduction to VLSI Systems: A Logic, Circuit, and System Perspective*. Boca Raton, FL, USA: CRC Press, 2012.



KAMAL Y. KAMAL received the B.Sc. degree in electrical engineering from Al-Mustansiriya University, Baghdad, Iraq, in 1999, and the M.Sc. degree in electronics from Quaid-i-Azam University, Islamabad, Pakistan, in 2003. He is currently pursuing the Ph.D. degree in engineering systems and computing at the University of Guelph, Canada. His current research interests include mixed-signal VLSI design, cryptography, system-on-chip design, security and intelligent embedded systems design, on-chip side-channel attack countermeasures, and physically uncleanable function modules. He is a member of the IEEE Circuits and System Society, the IEEE Solid-State Circuits Society, and ACM organizations.

RADU MURESAN received the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Canada. He is currently an Associate Professor with the Engineering Department, University of Guelph, Canada. His main research interests include VLSI design, system-on-chip design, and security and intelligent embedded systems design. In the area of security, he studies the integration of highly secure cryptographic components into intelligent embedded systems. Specifically, the design and integration of on-chip countermeasure circuits against side-channel attacks, physically uncleanable function modules, and chaotic ciphers. He is an Ontario Professional Engineer and a member of the IEEE Circuits and System Society, the IEEE Solid-State Circuits Society, and ACM organizations.