SCIENCE CHINA
Information Sciences

CrossMark
click for updates

• RESEARCH PAPER •

January 2024, Vol. 67, Iss. 1, 112404:1–112404:13
https://doi.org/10.1007/s11432-022-3722-8

A strong physical unclonable function with machine learning immunity for Internet of Things application

Pengpeng REN[1,2], Yongkang XUE[1,2], Linglin JING[2], Lining ZHANG[3], Runsheng WANG[4] & Zhigang JI[1*]

[1] *National Key Laboratory of Science and Technology on Micro/Nano Fabrication, Shanghai Jiao Tong University, Shanghai* 200240*, China;*
[2] *Department of Micro/Nano Electronics, Shanghai Jiao Tong University, Shanghai* 200240*, China;*
[3] *School of Electronic and Computer Engineering, Peking University, Shenzhen* 518055*, China;*
[4] *School of Integrated Circuits, Peking University, Beijing* 100871*, China*

Received 11 October 2022/Revised 28 December 2022/Accepted 14 March 2023/Published online 18 December 2023

Abstract The physical unclonable functions (PUFs) are novel cryptographic primitives in modern hardware security systems. Compared with traditional alternatives based on digital keys and non-volatile memory (NVM), the PUF system shows great unclonability, high efficiency, and physical attack resilience. However, the conventional PUF design suffers from weak machine learning immunity, high storage overhead, and unreliability, making it difficult to implement in the Internet of Things (IoT) and edge computing applications. This paper presents a new PUF design that could solve the proposed obstacles. By utilizing the emission probability of traps commonly found in nano-scaled transistors, a model-based PUF system with strong machine learning resistance could be achieved. This PUF design, called Prob-PUF, needs fewer challenge-response pairs (CRPs) space and reveals superior resistance to modeling attacks due to the mixture of stable/random bits in its output response. Moreover, the Prob-PUF system could reach a high level of uniqueness and robustness, making it a potential candidate for future cryptographically secured protocols within the IoT.

Keywords physical unclonable function (PUF), electron traps, authentication, encryption, cryptography, security

1 Introduction

With the explosive development of information technology, it will be a common phenomenon for everyday objects to be linked together. Billions of Internet of Things (IoT) devices could provide real-time computation and decision while maintaining connectivity and data communication with the "cloud" [1]. In order to support the massive amount of data interaction between different kinds of IoT devices, secure cryptographic protocols that are reliable and not vulnerable to being attacked should be built up to protect privacy information [2–4]. While the security protocols in traditional electronic entities commonly rely on digital keys which are stored in non-volatile memory (NVM), it is hard to implement the previous algorithm-based security systems to those "edge computing" devices [5]. Considering the limited energy budget and storage space of IoT applications, a lightweight and cost-efficient substitute is badly needed.

The physical unclonable functions (PUFs) [6–8], which have been discussed widely in recent years, could be an appropriate alternative approach to the existing security solutions. The PUF utilizes the random variation derived from the manufacturing process to generate unique "fingerprints" for each different hardware primitive. Due to their excellent unclonability, the PUF devices are supposed to have great advantages in the field of authentication or secret key generation to serve modern cryptographic algorithms [9,10] as shown in Figure 1. Only simple digital processing circuits are needed for the response output of PUF as the secret information is generated from the physical characteristics of the entity. The PUF hardware consumes less power and area overhead than NVM solutions to complete the same authentication process, making it more appropriate to deploy on mobile and embedded devices.

* Corresponding author (email: zhigangji@sjtu.edu.cn)

© Science China Press 2023

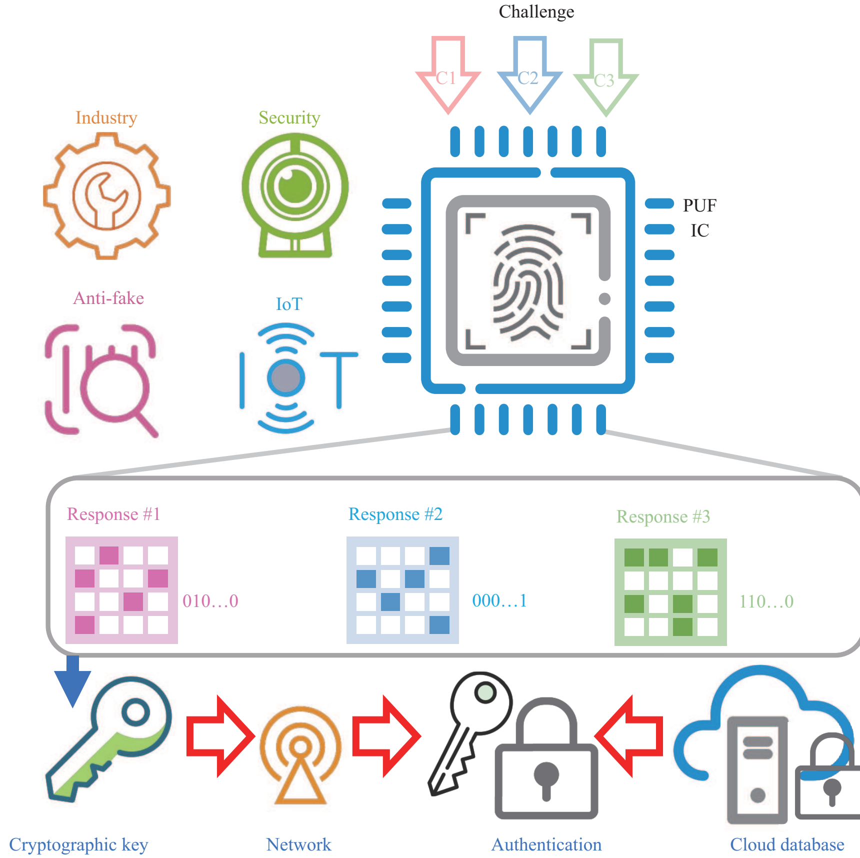

Figure 1 (Color online) PUF as an important component in modern cryptographic systems for various applications.

Though the earliest PUF designs were non-silicon PUFs, silicon PUFs are more popular at present compared with their bulky predecessors. There are two main types of silicon-based PUFs. The first type is time-delay based silicon PUFs like arbiter PUF (APUF) and ring-oscillator PUF (ROPUF). The second type is intrinsic silicon PUFs like static random-access memory PUF (SRAM PUF) [11]. However, both types of silicon-based PUFs suffer from some serious limitations:

Limited challenge-response pairs (CRPs). Some silicon-based PUFs like SRAM PUF could only support a small number of CRPs with a limited area budget. This kind of PUF is categorized as weak PUF. In contrast, those PUFs that can create enough CRPs are called strong PUFs [12]. According to the limited CRP numbers of weak PUFs, their functions are mainly restricted in the field of hardware fingerprints or key generation.

Vulnerable to modeling attacks. A PUF entity could be defined as a black-box challenge-response functional system. Moreover, the total construction of a specific PUF hardware could be concluded as a unique function:

$$r = f(c), \tag{1}$$

where c represents the input challenge, r reveals the returned response, and $f(\cdot)$ describes the challenge and response relationship.

When performing the tasks of IoT applications, the CRP interface of a PUF entity could be publicly accessible. This means that external attackers could gather the used CRP information. An experienced hacker could construct the unique function of the specific PUF based on the current information using mathematical methods, especially machine learning (ML) attacks. Once intruders crack the system function of the PUF, they could predict the rest of the CRPs. Nearly all kinds of delay-based PUFs are at risk of being cracked by powerful modeling attacks. Many variants sourced from APUFs and ROPUFs have been presented to improve their ML resilience. However, those solutions consume large area/power overhead or need unreasonable conditions [11].

Excessive storage expenditure. The key management of the PUF system is another intractable problem that is seldom discussed. The main discussion about PUF is only limited to the PUF circuit

design at present. Supposing that PUF entities are to be implemented in a server-client environment, the authentication process needs the server-side to secretly store all the CRPs in the form of a look-up table [13]. It is burdensome for the server-side to support a large amount of storage when there is frequent communication between thousands of PUF entities and the server. Eliminating the dilemma between the CRP and storage space is still a problem worth studying.

Almost no PUF exists that can solve all of the proposed problems at once. Therefore, realizing a storage efficient and robust strong PUF with no vulnerability to ML attacks is still an advancing topic that attracts attention from the academic and industry communities.

In this work, a new strong PUF, named probability-based PUF (Prob-PUF), is proposed to overcome all the deficiencies mentioned above. Utilizing the emission events of defects widely observed in commercial nano-scaled transistors, the proposed PUF design chooses the emission probability of the traps as the evaluation index to build a model-based PUF. Without any error correction circuit, the Prob-PUF could output a response mixed with deterministic and random bits, improving its resistance to modeling attacks such as ML. The proposed PUF shows strong immunity when facing ML attacks with an ideal prediction error of around 50%. The server side of the PUF system would just store the mathematical model to achieve the economical key management solution while maintaining a large CRP set.

The rest part of the paper is organized as follows. In Section 2, we explain the basic principle of the trap emission and the authentication process of the proposed PUF based on the emission probability. In Section 3, the specific structure of the proposed PUF circuit is demonstrated to reveal the total CRP space of the PUF system. In Section 4, a series of simulations are performed to evaluate the performance of the proposed PUF, including Hamming-distance, variation, and ML resilience. Finally, Section 5 concludes the whole paper.

2 Principle and authentication of Prob-PUF

2.1 Basic principle of defect emission

The oxide traps widely distributed in nano-scaled transistors have been discussed diffusely in recent years [14,15]. These traps are supposed to be the origin of a series of transistor aging mechanisms [16–20]. Given a settled gate voltage, the oxide traps would randomly capture and emit the channel carriers, causing the drain current variation in the time domain, as shown in Figure 2(a). Though the time required for an individual trap to emit its captured carrier, named the time-to-emit, varies due to its stochastic nature, the average time-to-emit remains the same. This average time-to-emit is called emission time constant τ_e, which is the most important parameter to describe the behavior of the trap.

Since both the activation energy and the location information of a specific trap are randomly distributed in transistors, the emission time constant of each trap is different. Moreover, the emission time constant of the trap remains stable after aging, which gives it the potential to become the “fingerprint” of the transistor [21]. On the other hand, τ_e of a single trap could be sensitive to the gate voltage applied to the transistor, as concluded in previous studies [22]. Figure 2(b) demonstrates the gate voltage dependence of the emission time constant for one typical trap generated with the randomly-chosen energy and spatial location following the standard simulation procedure. In this way, we could modify the emission speed of traps in a selected transistor to achieve the fast authentication process of a robust PUF system.

For a single trap, the probability of emitting its trapped carrier after elapsed time t is only related to the τ_e of the trap, as revealed in

$$P_e(t) = \left[1 - \exp\left(\frac{-t}{\tau_e}\right)\right]. \tag{2}$$

Therefore, given a preset time window t_w, the emission probability of different traps in this time window varies. This rule could be used as the entropy source of the PUF system. Unlike the previous trap-based design, which used the existence of traps in one transistor as the evaluation criteria [23], the Prob-PUF could generate larger CRP space and achieve strong ML immunity due to the randomness of the emission event.

2.2 Authentication process

As the PUF system works in the client-server environment, the same challenge is sent to both the server-side and client-side. While the client-side generates the physical response (real response) based on the

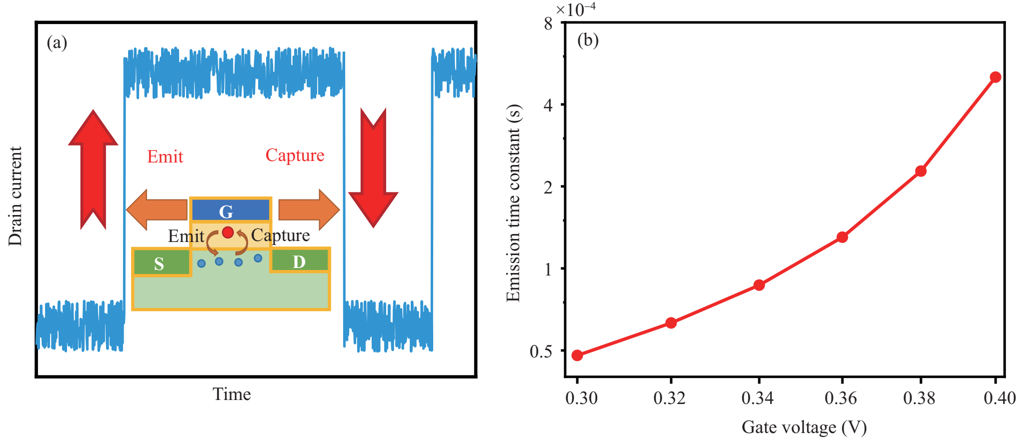

Figure 2 (Color online) (a) Capture and emission event of the traps in transistors; (b) variation of the emission time constant for one typical trap depending on the different gate voltage.

real emission situation, the server-side will calculate the corresponding theoretical emission probability according to the restored probability model based on (2) and generate the theoretical response. Finally, the real response will be transferred to the server and compared with the calculated result for completing the authentication.

On the client-side, the PUF circuit consists of a series of transistor matrices, each of which consists of a large number of transistors with the smallest size to maximize the impact of trap emission. The input challenge specifics the exact position of a set of transistors that are to be selected and stimulated in this authentication. After selection, a high voltage will be applied on the gate of the selected transistors for a while to ensure the capture event of oxide traps in these transistors (charging process). Then, the voltage applied on the gate would be lowered to sense the emission event of the traps (sensing process), as shown in Figure 3. If an emission event occurs in the preset time window t_w, an abrupt increase of the drain current could be detected in the subsequent circuit. The proposed charging-and-sensing voltage pattern would be repeatedly applied on these transistors several times (for example, ten times). For each transistor, if no emission event happens in t_w for all these ten times of sensing, the PUF circuit will output a stable 0 bit. In contrast, if there are emission events in all these ten times, a stable 1 bit will be generated. However, if both detrapping and non-detrapping events happen in these ten times, a random bit will be generated. Compared with the previous similar operation based on repeated measurements such as majority voting methods [24, 25], our procedure determines whether the specific bit is stable or not. Whether each bit is stable or not will be utilized for authentication and ML resistance. In order to keep the balance between the numbers 0 and 1, the random bit will be replaced with the output of a true random number (TRN) generator [26] before formal output. Finally, after combining all the single bits output, a complete response (real response) will be generated and sent to the server-side for further identification.

Given the same challenge, the server-side stored the emission time constant τ_e of all the transistors on the client-side. According to (2), the emission probability (P_e) of all the selected transistors could be calculated. To determine the output response, any transistor whose calculated emission probability higher than a predefined high P_e threshold ($P_{\rm eH}$) or lower than the low P_e threshold ($P_{\rm eL}$) would trigger a stable 1 or 0 bit. On the contrary, if the calculated probability is between the range of high threshold and low threshold, the output of this bit will be defined as a random bit. In fact, the high/low threshold is defined by the number of times the voltage pattern repeats. For example, if the client circuit repeats ten times of stimulation for each transistor, any traps whose emission probability is higher than 90% or lower than 10% will occur detrapping events or non-detrapping events in all the ten times of sense. This 90% and 10% value will be determined as high and low thresholds. Allowing for the possibility of multiple defects in one transistor, we only store the fastest τ_e on the server-side to ensure correctness. In the real application of the PUF system, the repeat times and probability threshold could change flexibly, achieving the trade-off between the reliability and authentication speed of the Prob-PUF system. Finally, the calculated theoretic response is generated after combining all the calculated bits. After discarding the random bits, the server compares only the stable bits between the theoretical response and the

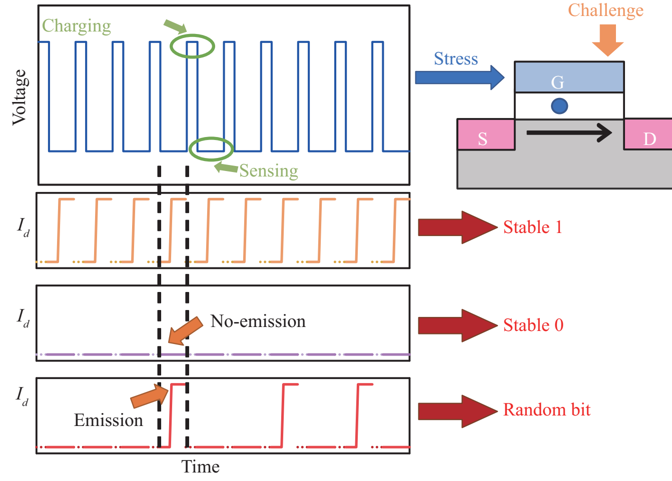

Figure 3 (Color online) Detailed process for the client-side to generate a one-bit real response in the Prob-PUF system.

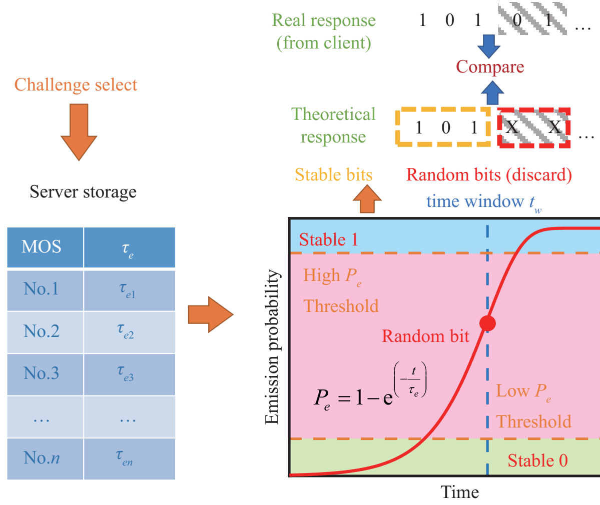

Figure 4 (Color online) Detailed procedure for the server-side to generate the theoretic response and compare it with the real response to complete the authentication.

real response transmitted from the client-side to complete the authentication. The whole procedure is illustrated in Figure 4.

The whole PUF system is constructed based on the emission probability model of the oxide traps. As a result, instead of saving an immense CRP table like the traditional paradigm [13], the proposed Prob-PUF only needs to save the τ_e of each transistor which could greatly reduce the storage expenditure. Moreover, as the response of the Prob-PUF circuit contains both stable and random bits, even if the attackers filch them, it is impossible to build an accurate model based on the obtained response, achieving strong ML resilience. The random bits in the response would poison the training process to hinder the normal construction of a clear separation boundary used by ML algorithms [27]. To further improve the reliability of the PUF system, a proper time window t_w would be selected to eliminate the biasing situation of the stable bits in the PUF response. After the fabrication, multiple transistors used in the PUF would be measured to determine the whole distribution of the trap emission time in the PUF circuit. The time window could be selected to ensure the stable bits with a 50% cumulative distribution function (CDF) probability of '0's and '1's.

3 Circuit structure of the proposed PUF

The overall architecture of the proposed PUF is shown in Figure 5. Supposing a PUF system with N bit input challenge and N bit response, the whole circuit hierarchy can be arranged with $N/8$ sub-

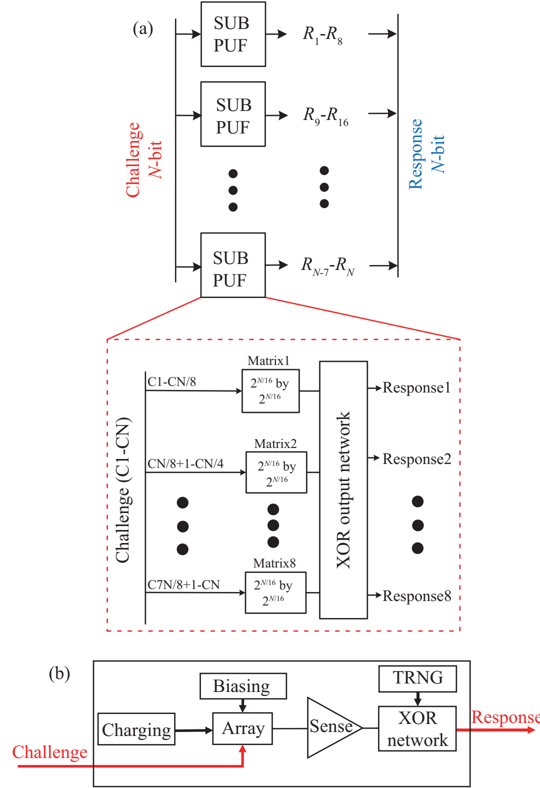

Figure 5 (Color online) Circuit structure of Prob-PUF. (a) Hierarchy structure of the whole PUF system and sub-PUF; (b) detailed circuit implementation of function blocks to generate a one-bit response.

PUFs, each of which will generate an 8-bit response. As illustrated in Figure 5(a), each sub-PUF contains eight $2N/16$-by-$2N/16$ transistor matrices, and each matrix supports $N/8$-bit width address (challenge). As for each matrix to generate a single bit response, the subsequent functional circuit consists of several blocks. As demonstrated in Figure 5(b), the charging and biasing block creates the charging-and-sensing voltage pattern to the gate of the selected transistor in transistor matrices. If the emission event occurs in the time window t_w, a voltage step signal will appear on the direct output point of the transistor matrix. Finally, the sense block will amplify the voltage signal from the transistor matrix and compare the voltage value at the beginning and end of the time window to determine whether the output value is '0' or '1'. The same voltage pattern is cycled ten times to detect the specific bit as the stable or random bit. After the comparison, the result will be sent to an XOR-network for further processing.

As the output bits from different matrices are independent, the XOR-network [28] is used to combine the independent bits from the output of each transistor matrix. The XOR-network connects the independent output bits to generate the final response, avoiding the risk of possible information leaks caused by repeated input challenges. The detailed structure of the XOR-network is shown in Figure 6. Wherein, to hinder the random bit from disrupting the stable bits, the bit '0' is utilized to replace the random bit as the input of the XOR network using multiplexers (MUXs) to ensure the predictability of every bit in the output of the network. MUXs are also added to the output of the XOR-network to restore the unpredictability of random bits. If a certain bit is determined to be random in the previous procedure, then the corresponding mixed bit at that position will be replaced by a TRN. This operation ensures that every response generated from the PUF circuit will be mixed with stable and random bits.

4 Security and robustness performance

4.1 CRP space compared with storage overhead

The CRP space decides the degree of security of a PUF system. However, the CRP storage of the PUF system on the server-side would finally determine the application scenarios and the number of PUF circuits that could be implemented simultaneously. In terms of CRP storage, the conventional

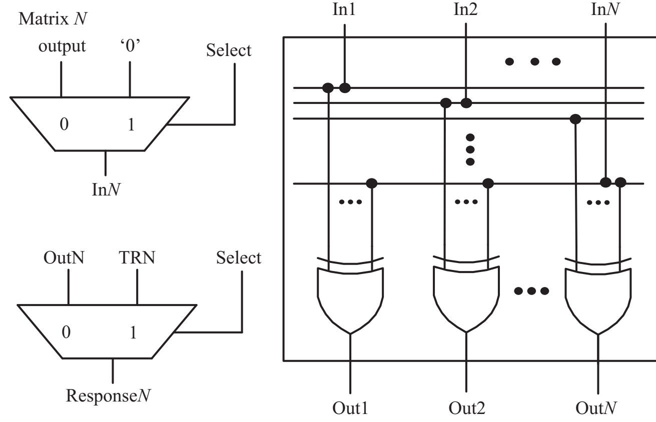

Figure 6 Structure of XOR-network before generating the final response. The 'Select' signal is generated from the charging-and-sensing blocks before.

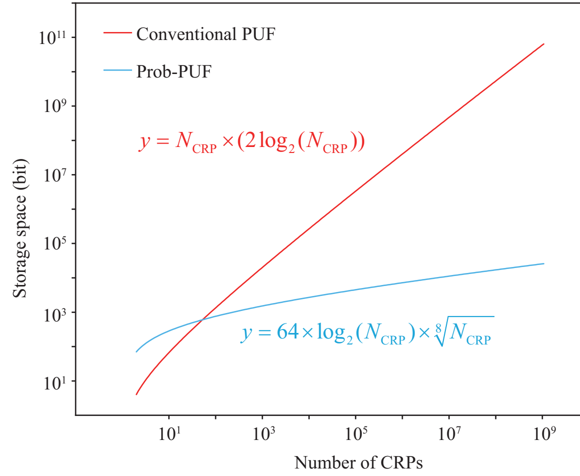

Figure 7 (Color online) Comparison of the storage expenditure for CRPs between the conventional PUF solution (table-based) and the Prob-PUF (model-based).

PUFs use the table-based solution [13]: CRPs are stored in the form of a table in the database on the authentication server. However, the proposed Prob-PUF uses the model-based solution [12] in which only model parameters are stored on the server. The model-based storage solution could effectively reduce the storage overhead of the server. In our particular case, the model parameters are the MOS-ID and the emission time constant for each transistor that can generate one bit for the PUF response.

To intuitively compare the required storage between these two methods, a proposed Pro-PUF system with N-bit challenges and N-bit responses is assumed for ease of simplicity. The CRP space of this PUF ($N_{\rm CRP}$) equals 2^N which grows exponentially with the challenge width, hinting that this PUF is strong.

For the conventional table-based method, each CRP includes one N-bit challenge and one N-bit response, which occupies $2N$ bits of space in total. For the CRP number of ($N_{\rm CRP}$), the total required storage is $N_{\rm CRP} \times (2N)$ bits, i.e., $N_{\rm CRP} \times (2 \times \log_2 (N_{\rm CRP}))$ bits. Meanwhile, using the proposed Prob-PUF hierarchy structure described in Section 3, there are $2^{(N/8)} \times N$ transistors in total. The MOS-ID and the emission time constant for each transistor should be stored, i.e., $2 \times \left(N \times 2^{(N/8)}\right)$ parameters. By adopting the standard 32-bit IEEE floating-point format, the required storage is $32 \times 2 \times \left(N \times 2^{(N/8)}\right)$ bits, i.e., $64 \times \log_2 (N_{\rm CRP}) \times (N_{\rm CRP})^{1/8}$ bits.

The comparison of storage overhead between conventional PUFs and our Pro-PUF as the increase of CRP amount is shown in Figure 7. It is clear that when the PUF system includes more than 100 CRPs, the Prob-PUF shows a significant reduction in storage.

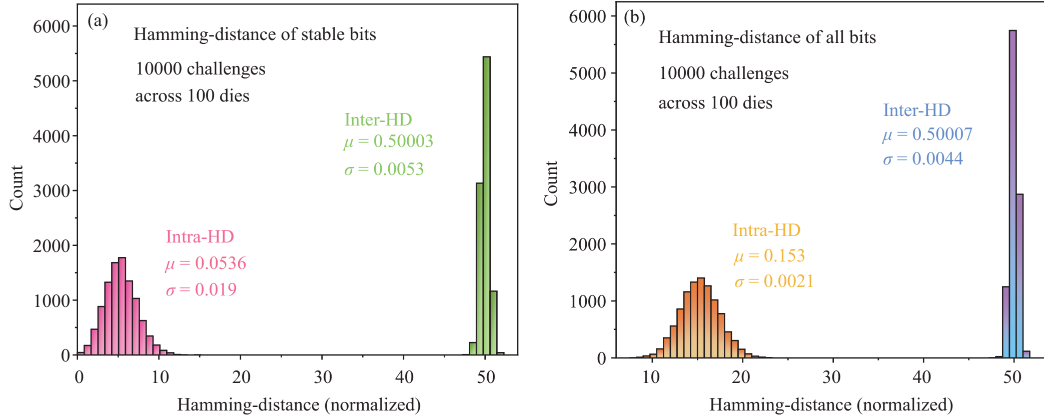

Figure 8 (Color online) Normalized inter- & intra-HD of Prob-PUF. (a) Only stable bits; (b) all bits.

4.2 Uniqueness and uniformity performance

Uniqueness and uniformity are two key properties to confirm the security of a PUF design [29]. Given a constant challenge, one PUF entity should output the same response set every time in the ideal case. On the contrary, each different PUF entity could ideally produce a unique set of responses to the same challenge. The uniformity of a PUF is formally quantified as the intra Hamming distance (HD), which demonstrates the stability of responses when a die is given repeated challenges. The theoretical normalized intra-HD value of a perfect PUF design is 0. Further, the normalized inter-HD is calculated to measure the uniqueness of a specific PUF design. The inter-HD is generally used to distinguish the difference between two identical PUF instances. The ideal normalized inter-HD is 0.5, which declares that 50% of the response bits generated from two different PUF instances to the same challenge are different. A proper normalized inter/intra-HD close to 0.5/0 illustrates the superior identifiability and reliability of the specific PUF.

To test the uniqueness and uniformity of the Prob-PUF, more than 10^6 response bits across 100 different dies are gathered. As shown in Figure 8(a), the measured normalized inter-HD of stable bits has an average of $\mu = 0.50003$ and a standard deviation of $\sigma = 0.0053$. In contrast, the normalized intra-HD of stable bits has an average of $\mu = 0.0536$ and a standard deviation of $\sigma = 0.019$. Compared with stable bits, the HD of complete response tends to 0.5 due to the mixture of random bits. The gathered normalized intra-HD of all bits has an average of $\mu = 0.153$ and a standard deviation of $\sigma = 0.0021$, which illustrates the perfect confusability of the embedded random bits. The normalized inter-HD of all bits has an average of $\mu = 0.50007$ and a standard deviation of $\sigma = 0.0044$. The detailed distribution is shown in Figure 8(b). It can be seen that both types of response bits have uniqueness close to the ideal value of 0.5.

The randomness of the proposed Prob-PUF system is evaluated by the NIST statistical test suite for random and pseudorandom number generators, which is performed on 200000 bits from the PUF system. As shown below, the outputs of the PUF system pass the tests.

4.3 Design for robustness

The output response of Prob-PUF is mainly determined by the emission time constants of transistors in the PUF circuit. In practice, the emission time constants could be affected by external influences, including voltage variation, temperature variation, and aging. Though it has been confirmed in the previous work that the emission time constant stays settled through aging [30]. The other two elements could change the emission time constant during the operation, and the influence of the supply voltage and ambient temperature should be evaluated.

The temperature sensitivity can be a common difficulty for PUF design, especially RO-based PUFs [31] and arbiter-based PUFs [32]. For the proposed Prob-PUF, the change in ambient temperature leads to the change in the emission time of the trap, as described in

$$\tau_e \propto \exp\left(\frac{E_a}{kT}\right). \tag{3}$$

Table 1 Randomness evaluation results of the Prob-PUF system

No.	Type of test	P-value	Status
1	Frequency test (monobit)	0.508050616	Passed
2	Frequency test within a block	0.368881744	Passed
3	Run test	0.29212373	Passed
4	Longest run of ones in a block	0.094089647	Passed
5	Binary matrix rank test	0.605071974	Passed
6	Discrete Fourier transform (spectral) test	0.538167126	Passed
7	Non-overtapping template matching test	0.071815941	Passed
8	Overlapping template matching test	0.601304996	Passed
9	Linear complexity test	0.708402088	Passed
10	Serial test	0.71623516	Passed
11	Approximate entropy test	0.744039427	Passed
12	Cummulative sums (forward) test	0.292174959	Passed
13	Cummulative sums (reverse) test	0.822250312	Passed

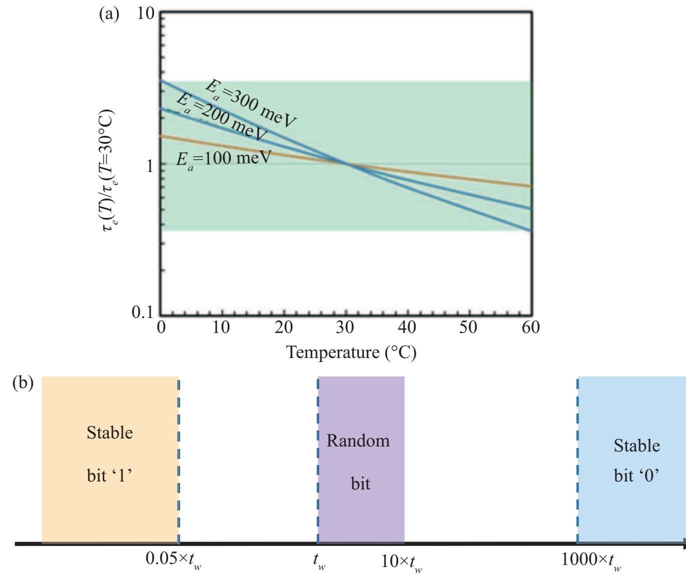

Figure 9 (Color online) (a) Impact of temperature on the emission time for different activation energy E_a. The emission time is normalized against the value at 30°C. (b) The designed rule in selecting suitable transistors for the random and stable bits. The target is to accommodate traps with E_a less than 300 mV under temperatures between 0°C and 60°C.

One possible solution to suppress this effect is selecting suitable transistors that exhibit weak temperature dependence for authentication during registration. This is similar to the soft dark-bit masking method already in use [33]. In order to select proper transistors for random and stable bits, specific selection rules are designed. One typical rule is given below, the target of which is to accommodate any transistors with a trap which activation energy (E_a) is smaller than 300 meV for the temperature ranging from 0°C to 60°C.

The emission time at 30°C is used as the standard reference stored in the server, denoted as τ_{e0}. As calculated by (3) and shown in Figure 9(a), when the temperature changes between 0°C and 60°C, the emission time deviates from τ_{e0}. For the extreme case with the activation energy of 300 meV, the emission time can vary from $4\tau_{e0}$ to $0.25\tau_{e0}$.

As calculated in (3), the emission probability of the trap, P_e, equals $1-\exp(t_w/\tau_e)$. Wherein, t_w is the time window for the Prob-PUF. According to the variation of time constant and probability, the specific rules for choosing different kinds of bits are given as follows.

• For random bits, we can select the trap with τ_{e0} ranging between t_w and $10t_w$. P_e is 0.632 for $\tau_{e0} = t_w$. When temperature variation is considered, P_e spreads from 0.22 to 0.98. Similarly, for $\tau_{e0} = 10t_w$, the corresponding P_e is 0.095. With temperature variation, P_e spreads between 0.025 and 0.33. Therefore, P_e of random bits can vary between 0.025 and 0.98 when the temperature varies between 0°C and 60°C.

• For stable bits '1', we can select the trap with τ_{e0} smaller than $0.05t_w$. In this case, P_e is always

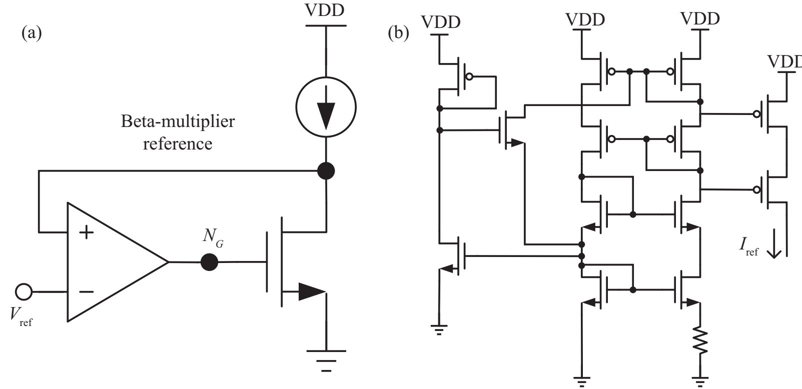

Figure 10 (a) The structure of the monitoring circuit. The impact of the single-trap-induced threshold shift. (b) The transistor-level BMR circuit derived from [35].

higher than 0.993 (when $\tau_e = 4\tau_{e0}$) when the temperature varies between 0°C and 60°C.

• For stable bits '0', we can select the trap with τ_{e0} larger than $1000t_w$. In this case, P_e is always smaller than 0.004 (when $\tau_e = 0.25\tau_{e0}$) when the temperature varies between 0°C and 60°C.

The above selection rule can be summarized in Figure 9(b). During the authentication process, the server-side could modify the high threshold/low threshold to 0.99/0.01 to differentiate between random and stable bits. As the criterion threshold becomes stricter, it is necessary to increase the repeat time of the sensing (For example, from 10 times to 100 times). The Prob-PUF system could sacrifice the speed for a higher tolerance of temperature variation.

Another main issue affecting the PUF response is the supply voltage variation. The emission time constant of traps in transistors could change dramatically with the gate voltage of the sensed transistor, as shown in Figure 2. However, the fluctuation of the voltage supply from the external input could disturb the normal operating point of the sensed transistor, and this will give rise to the error of response bits in the worst case.

In order to suppress the voltage variation on the sensed transistor, we adopted the widely-used constant current scheme [34] to monitor the emission event, as shown in Figure 10(a). Wherein, the sensing current level $I_{\rm ref}$ is set by the Beta-multiplier reference structure (BMR). Various designs have been proposed to ensure the BMR current source insensitive to $V_{\rm DD}$ variation [35, 36]. Our design modified the traditional BMR structure to a cascaded type to further improve its stability, as shown in Figure 10(b). The negative feedback configuration for the op-amp ensures the single-trap-induced-$V_{\rm th}$-shift is monitored as the abrupt change at the node of the gate of the transistor (Ng). When trap emission occurs during the sensing process, the $V_{\rm th}$ of the sensed transistor decreases abruptly. The transient gate voltage decreases as the drain current needs to be kept constant. Therefore, the emission event can be detected.

To test the robustness of the proposed structure against supply voltage variation, the proposed Prob-PUF circuit is set up in a Cadence environment with 65 nm commercial process design kits (PDK). The gate voltage at the quiescent point (Ng) is set to 0.35 V for a quick emission event by adjusting the drain current value. The emission time constant of the trap in a typical transistor is repeatedly measured and extracted while the supply voltage varies from 0 to 1.2 V. This emission time constant is around 100 μs under the proposed quiescent point and normal supply voltage (1.2 V). As illustrated in Figure 11(a), when $V_{\rm DD}$ is higher than 0.6 V, the maximum change in the emission time constant is about $+/-20\%$. When $V_{\rm DD}$ is below 0.6 V, the BMR structure itself becomes malfunctional. Furthermore, we checked the impact of the reference voltage $V_{\rm ref}$. We compared the emission time extracted under different $V_{\rm ref}$ conditions while maintaining $V_{\rm DD}$ at 1.2 V. As shown in Figure 11(b), the emission time varies at most 20% when $V_{\rm ref}$ changes between 0.3 and 0.7 V (the normal operating voltage of $V_{\rm ref}$ is 0.5 V). In general, the change in emission time constant caused by voltage fluctuation is much smaller than that caused by temperature variation. Like the solution used to suppress the temperature effect, specific design rules could be utilized to select suitable transistors for random and stable bits.

4.4 Machine-learning resilience

As a strong PUF, the challenge-response interface of Prob-PUF is public-facing. An attacker could gather all the used CRPs to build a mathematical model. As the traditional PUF circuit structure could be

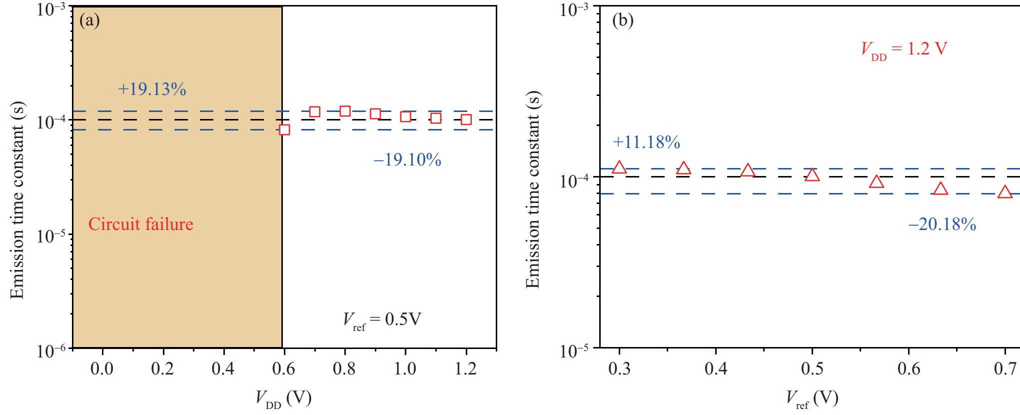

Figure 11 (Color online) PUF robustness measurement against supply voltage variation. (a) The emission time constant (τ_e) under different $V_{\rm DD}$. (b) The largest positive and negative alterations on the emission time constant. The emission time constant (τ_e) and variation of τ_e for different $V_{\rm ref}$.

generalized as an underlying function with several unknown parameters, hackers could approach this function using automatic update algorithms like machine-learning [12]. As a result, an adversary could predict all the remaining CRPs based on the observed small set of CRPs and crack the entire PUF circuit. The ability of a strong PUF to resist ML attacks is a key assessment criterion of its security.

In order to evaluate the machine-learning resilience of proposed Prob-PUF, several typical ML methods (logistic regression (LR), support vector machines (SVM) and neural networks (NN)) which have been widely used in [37, 38] are implemented and utilized to attack Prob-PUF. For comparison, the same attacks are run on APUF [6], 4-XOR PUF [5], and 4-XOR lightweight (LW) PUF [28]. All the PUFs included in the test are given the same challenge length of 64 bits and response length of 1 bit. The CRPs of Prob-PUF used for training and testing are divided from the circuit simulation, while the CRP datasets of three referenced PUF used for training are generated from the simulated program that was open-sourced and available in [39]. The number of CRP used for training in evaluation ranges from 50 to 10^6. The ratio of the total number of samples that are used for training and inference respectively was set to 4 by 1, which is commonly used in the machine-learning setup.

LR is one of the simplest and most effective machine-learning algorithms in PUF modeling attacks. The greatest advantage of LR in PUF attacks is that this algorithm could also specify nonlinear decision boundaries apart from the conventional linear decision boundary. In the test, the previous successful procedure was followed, in which the LR model was set up in a Python environment using the Sckikit-learn library [40]. The Newton-conjugate gradient is used as the loss function. To attack 4-XOR PUF and LW PUF, we used the RProp gradient descent algorithm [41] as the optimization method for its improvement in convergence speed and stability of the LR [39]. This paradigm is also applied for the Prob-PUF attack due to its similar XOR output network.

SVM is a widely applied ML algorithm for classification which could be fairly effective when facing a training set with small numbers. It could use nonlinear kernel functions to improve its ability to distinguish samples with high dimensional features. To attack the APUF, the SVM model was set up with the same configuration as [37], which is based on the linear kernel and uses the hinge function as the loss function. According to [38], the learning ability of an SVM is higher when employing a nonlinear kernel function. Therefore, to attack Prob-PUF, the radial basis function (RBF) [42] is applied as the nonlinear kernel function.

NNs are specific machine-learning structures that consist of interconnected computing nodes. A typical NN implementation called the multi-layer perceptron (MLP) [43] model was implemented with a $64 \times 800 \times 128 \times 2$ structure. The input layer of 64 neurons corresponds to the 64 bits of challenge, while the output layer of 2 neurons indicates that the output value is 0/1. Eight hundred and 128 neurons are used in the first and second hidden layers to store sufficient internal parameters. The rectified linear unit (ReLu) [44] function is used as the activation function for each hidden layer to bring in nonlinear transformation.

During the evaluation, we found that the SVM and NN will not converge during training for the 4-XOR PUF and 4-XOR LW PUF, which is similar to [40]. Therefore, we applied all three methods to

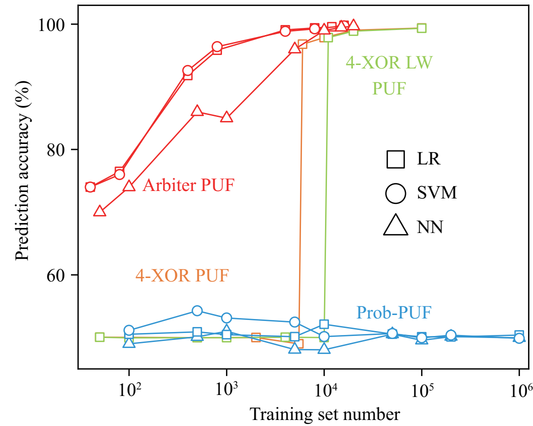

Figure 12 (Color online) Machine-learning attacks result of proposed Prob-PUF compared with other typical PUF systems.

the APUF and the LR method to the 4-XOR PUF & 4-XOR Lightweight PUF. As shown in Figure 12, the prediction accuracy rises above 99% when training samples are more than 10^5. This is similar to the results in [39, 40]. When the same ML methods are adopted for Prob-PUF, the prediction accuracy stays at only 50% even when the training sample numbers reach 10^6. This result confirms that the Prob-PUF exhibits good immunity to the ML attack.

5 Conclusion

In this paper, a strong PUF that derives from the emission probability of transistor traps is presented. This novel PUF design includes both underlying physical mechanisms and top layer system implementation. The advantages of the proposed Prob-PUF include large CRP space, small storage overhead, resilience to machine-learning attacks, and lower power expenditure which makes it a potential candidate for securing future large-scale systems in IoT applications. In order to evaluate its performance and robustness, the PUF circuit was constructed with a 65 nm commercial PDK. The proposed PUF shows great uniformity and uniqueness through various tests. Moreover, this PUF construct solves two major problems (machine-learning resistance and storage management) that have plagued the PUF structure for a long time and could have value in triggering further PUF design ideas in the field of hardware security.

Acknowledgements This work was partly supported by National Key R&D Program of China (Grant No. 2019YFB2205005).

References

1 Wang Z Q, Du Y, Wei K J, et al. Vision, application scenarios, and key technology trends for 6G mobile communications. Sci China Inf Sci, 2022, 65: 151301
2 Shi W S, Cao J, Zhang Q, et al. Edge computing: vision and challenges. IEEE Int Things J, 2016, 3: 637–646
3 Cai D H, Fan P Z, Zou Q Y, et al. Active device detection and performance analysis of massive non-orthogonal transmissions in cellular Internet of Things. Sci China Inf Sci, 2022, 65: 182301
4 Xu J D, Yuen C, Huang C W, et al. Reconfiguring wireless environments via intelligent surfaces for 6G: reflection, modulation, and security. Sci China Inf Sci, 2023, 66: 130304
5 Suh G E, Devadas S. Physical unclonable functions for device authentication and secret key generation. In: Proceedings of the 44th ACM/IEEE Design Automation Conference, 2007. 9–14
6 Zhang S, Zhang J, Li S H, et al. Reconfigurable physical unclonable cryptographic primitives based on current-induced nanomagnets switching. Sci China Inf Sci, 2022, 65: 122405
7 Gassend B, Clarke D, Van Dijk M, et al. Controlled physical random functions. In: Proceedings of the 18th Annual Computer Security Applications Conference, 2002. 149–160
8 Jeloka S, Yang K, Orshansky M, et al. A sequence dependent challenge-response PUF using 28nm SRAM 6T bit cell. In: Proceedings of Symposium on VLSI Circuits, 2017. 270–271
9 Liu C Q, Cao Y, Chang C H. ACRO-PUF: a low-power, reliable and aging-resilient current starved inverter-based ring oscillator physical unclonable function. IEEE Trans Circ Syst I, 2017, 64: 3138–3149
10 Potkonjak M, Goudar V. Public physical unclonable functions. Proc IEEE, 2014, 102: 1142–1156
11 Gao Y S, Al-Sarawi S F, Abbott D. Physical unclonable functions. Nat Electron, 2020, 3: 81–91
12 Herder C, Yu M D, Koushanfar F, et al. Physical unclonable functions and applications: a tutorial. Proc IEEE, 2014, 102: 1126–1141
13 Awano H, Sato T. Ising-PUF: a machine learning attack resistant PUF featuring lattice like arrangement of arbiter-PUFs. In: Proceedings of the Design, Automation & Test in Europe Conference & Exhibition (DATE), 2018. 1447–1452
14 Grasser T. Stochastic charge trapping in oxides: from random telegraph noise to bias temperature instabilities. MicroElectron Reliab, 2012, 52: 39–70

15 Grasser T, Reisinger H, Goes W, et al. Switching oxide traps as the missing link between negative bias temperature instability and random telegraph noise. In: Proceedings of IEEE International Electron Devices Meeting (IEDM), 2009. 1–4
16 Ji Z G, Chen H B, Li X Y. Design for reliability with the advanced integrated circuit (IC) technology: challenges and opportunities. Sci China Inf Sci, 2019, 62: 226401
17 Ren P, Gao R, Ji Z, et al. Understanding charge traps for optimizing Si-passivated Ge nMOSFETs. In: Proceedings of IEEE Symposium on VLSI Technology, 2016. 1–2
18 Ji Z G, Zhang X, Franco J, et al. An investigation on border traps in III-V MOSFETs with an $In_{0.53}Ga_{0.47}As$ channel. IEEE Trans Electron Dev, 2015, 62: 3633–3639
19 Brown J, Gao R, Ji Z G, et al. A low-power and high-speed true random number generator using generated RTN. In: Proceedings of IEEE Symposium on VLSI Technology, 2018. 95–96
20 Zhan X P, Shen C D, Ji Z G, et al. A dual-point technique for the entire I_D-V_G characterization into subthreshold region under random telegraph noise condition. IEEE Electron Dev Lett, 2019, 40: 674–677
21 Grasser T, Reisinger H, Wagner P J, et al. Time-dependent defect spectroscopy for characterization of border traps in metal-oxide-semiconductor transistors. Phys Rev B, 2010, 82: 245318
22 Nagumo T, Takeuchi K, Hase T, et al. Statistical characterization of trap position, energy, amplitude and time constants by RTN measurement of multiple individual traps. In: Proceedings of International Electron Devices Meeting, 2010
23 Chen J, Tanamoto T, Noguchi H, et al. Further investigations on traps stabilities in random telegraph signal noise and the application to a novel concept physical unclonable function (PUF) with robust reliabilities. In: Proceedings of Symposium on VLSI Technology, 2015. 40–41
24 Mathew S K, Satpathy S K, Anders M A, et al. 16.2 A 0.19 pJ/b PVT-variation-tolerant hybrid physically unclonable function circuit for 100 stable secure key generation in 22nm CMOS. In: Proceedings of the IEEE International Solid-State Circuits Conference, 2014. 278–279
25 Vijayakumar A, Patil V, Kundu S. On improving reliability of SRAM-based physically unclonable functions. J Low Power Electron Appl, 2017, 7: 2
26 Brown J, Zhang J F, Zhou B, et al. Random-telegraph-noise-enabled true random number generator for hardware security. Sci Rep, 2020, 10: 17210
27 Wang S J, Chen Y S, Li K S M. Adversarial attack against modeling attack on PUFs. In: Proceedings of the 56th ACM/IEEE Design Automation Conference (DAC), 2019. 1–6
28 Majzoobi M, Koushanfar F, Potkonjak M. Lightweight secure PUFs. In: Proceedings of the IEEE/ACM International Conference on Computer-Aided Design, 2008. 670–673
29 Maiti A, Gunreddy V, Schaumont P. A systematic method to evaluate and compare the performance of physical unclonable functions. 2013. https://eprint.iacr.org/2011/657.pdf
30 Liu C, Lee K T, Lee H, et al. New observations on the random telegraph noise induced Vth variation in nano-scale MOSFETs. In: Proceedings of IEEE International Reliability Physics Symposium, 2014
31 Rahman M T, Forte D, Fahrny J, et al. ARO-PUF: an aging-resistant ring oscillator PUF design. In: Proceedings of the Design, Automation & Test in Europe Conference & Exhibition (DATE), 2014. 1–6
32 Zhou C, Parhi K K, Kim C H. Secure and reliable XOR arbiter PUF design: an experimental study based on 1 trillion challenge response pair measurements. In: Proceedings of the 54th Annual Design Automation Conference, 2017. 1–6
33 Suresh V, Kumar R, Anders M, et al. A 0.26 BER, 1028 challenge-response machine-learning resistant strong-PUF in 14nm CMOS featuring stability-aware adversarial challenge selection. In: Proceedings of IEEE Symposium on VLSI Circuits, 2020. 1–2
34 Simicic M, Morrison S, Parvais B, et al. A fully-integrated method for RTN parameter extraction. In: Proceedings of Symposium on VLSI Technology, 2017. 132–133
35 Liu S, Baker R J. Process and temperature performance of a CMOS beta-multiplier voltage reference. In: Proceedings of Midwest Symposium on Circuits and Systems, 1998. 33–36
36 Vittoz E, Fellrath J. CMOS analog integrated circuits based on weak inversion operations. IEEE J Solid-State Circ, 1977, 12: 224–231
37 Hospodar G, Maes R, Verbauwhede I. Machine learning attacks on 65nm arbiter PUFs: accurate modeling poses strict bounds on usability. In: Proceedings of IEEE International Workshop on Information Forensics and Security (WIFS), 2012. 37–42
38 Zhuang H Y, Xi X D, Sun N, et al. A strong subthreshold current array PUF resilient to machine learning attacks. IEEE Trans Circ Syst I, 2019, 67: 135–144
39 Rührmair U, Sehnke F, Sölter J, et al. Modeling attacks on physical unclonable functions. In: Proceedings of the 17th ACM Conference on Computer and Communications Security, 2010. 237–249
40 Ruhrmair U, Solter J, Sehnke F, et al. PUF modeling attacks on simulated and silicon data. IEEE Trans Inform Forensic Secur, 2013, 8: 1876–1891
41 Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: Proceedings of the IEEE International Conference on Neural Networks, 1993. 586–591
42 Chang Y W, Hsieh C J, Chang K W, et al. Training and testing low-degree polynomial data mappings via linear SVM. J Mach Learn Res, 2010, 11: 1471–1490
43 Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw, 1989, 2: 359–366
44 Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning, 2010