This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LED.2021.3130606, IEEE Electron Device Letters

A Probability-based Strong Physical Unclonable Function with Strong Machine Learning Immunity

Zezhong Tu, Yongkang Xue, Pengpeng Ren, Feng Hao, Runsheng Wang, *Member, IEEE*, Meng Li, *Member, IEEE,* Jianfu Zhang, Zhigang Ji, *Member, IEEE* and Ru Huang, *Fellow, IEEE*

Abstract — A novel strong physical unclonable function (PUF), called Probability-based PUF (Prob-PUF), is proposed using the stochastic process of trap emission in nano-scaled transistors. For the first time, the information of trap emission probability is used in the PUF design. This new approach offers ideal immunity to machine learning (ML) attacks. Since Prob-PUF only stores a mathematical model, it naturally avoids the dilemma between the requirement of a large number of challenge-response pairs (CRPs) and the limited storage space, making it a potential solution for future secure storage.

***Index Terms* — Physical unclonable functions (PUFs), electron defects, Charge trapping, security, authentication**

I. INTRODUCTION

Information security is one of the foundational requirements for the future society thriving on digital connectivity [1]. Compared with a traditional system that stores the algorithm-generated challenge-response pairs in non-volatile memories, physical unclonable functions (PUFs) provide low-cost and lightweight solutions. Although many PUF designs have been proposed, most of them only provide a limited number of challenge-response pairs (CRPs). Therefore, they are considered weak PUFs [2]. Examples include ring-oscillator PUF [3], SRAM PUF [4], and current-mirror PUF [5]. Since it is infeasible to exhaust all the CRPs, the extracted PUF features are mainly used as hardware fingerprints for anti-counterfeiting.

On the contrary, a strong PUF with a large number of CRPs can unlock more versatile applications like secure communications for the internet of things (IoTs). Therefore, the realization of strong PUFs has attracted attention from both academia and industry in recent years. Several structures have been proposed. For example, Jeloka et al. [6] modified a traditional SRAM PUF by adding an order of rows, achieving expanded CRP space. Xi et al. [7] proposed an SCA-PUF, which exploits the nonlinearity behaviors of MOSFETs in the subthreshold region. Their design achieved a CRP space of 2^{65}.

However, one major challenge for those strong PUFs is their vulnerability to machine learning (ML) attacks [2]. Moreover, key management is another challenge for strong PUFs, which has not received much attention in the past. Considering that these PUFs are to be implemented in a client-server environment, the increase in CRP space can make it more challenging to manage using conventional table-based solutions. This limits system scalability.

In this work, a strong PUF, named Probability-based PUF (Prob-PUF), is proposed to tackle the challenges mentioned above. The new design exploits the stochastic electron emission process, widely observed in commercial nano-scaled transistors. Unlike the state-of-art PUFs that either discard the uncertain outputs with error correction [8] or convert them into deterministic values using algorithms [9], our design makes use of the information from emission probability and thus achieves the strong immunity to ML attacks with an ideal prediction error of around 50%. In addition, the proposed Prob-PUF stores a mathematical model rather than a large CRP table and thus provides a secure and storage-efficient solution. Fig. 1 shows the feature of the proposed PUF in comparison with prior art.

PUF NAME	Ring-oscillator PUF [3]	Modified SRAM PUF [6]	Subthresold current PUF [7]	Prob-PUF
CRP space	Small	Large	Large	Large
ML Resistance	Weak	Weak	Weak	Strong
ECC Module	Need	Need	Need	Without
Storage Type	CRP Table	CRP Table	CRP Table	Model

Fig. 1. Comparison between major PUFs and the proposed design.

II. PROBABILITY-BASED PUF

A. Basic Principle

In recent years, the detrapping process for individual traps in nano-scaled transistors has been studied extensively [10]. Fig. 2(a) demonstrates the typical test procedure. After applying a high gate voltage to ensure the charging of one electron trap (Charging), the voltage is lowered to sense the detrapping event (Sensing). If detrapping occurs, an abrupt increase of the drain current can be detected. By repeating this charging-and-sensing procedure on one nMOSFET with 28nm technology node, the time-to-emit of the same trap varies as shown in Fig. 2(b) due to its stochastic nature. The probability of emitting the trapped charge after elapsed time t can be described in Eqn (1), wherein, τ_e is the emission time constant of the trap.

$$P_e(t) = [1 - \exp(\frac{-t}{\tau_e})] \quad (1)$$

Since both the energy and spatial locations of the trap are randomly distributed, τ_e measured in each transistor is different. Moreover, τ_e of the trap in each transistor does not change with time [11]. Therefore, it has been considered as the fingerprint of a transistor [12]. Recently, the trap-based PUF design has been proposed. For example, Chen et al. constructed a weak PUF by judging whether there is a pre-existing trap in the transistor [13]. In this work, we show that a strong Prob-PUF

This work is financially supported by the National Key Research and Development Program of China under grant 2019YFB2205005.

Zezhong Tu, Yongkang Xue, Pengpeng Ren and Zhigang Ji are with National Key Laboratory of Science and Technology on Micro/Nano Fabrication, Shanghai Jiaotong University (email: zhigangji@sjtu.edu.cn). Zezhong Tu and Yongkang Xue are also with Department of Micro/Nano Electronics, Shanghai Jiao Tong University. The first two authors contributed equally to this work.

Meng Li is with the Department of Electrical and Computer Engineering, University of Texas at Austin.

Feng Hao is with the Department of Computer Science, University of Warwick, UK.

Runsheng Wang and Ru Huang are with Peking University. Beijing, China.

Jianfu Zhang is with the Department of Electronics and Electrical Engineering, Liverpool John Moores University, Liverpool L3 3AF, UK.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LED.2021.3130606, IEEE Electron Device Letters

with high ML immunity can be achieved by further utilizing the information of the emission probability.

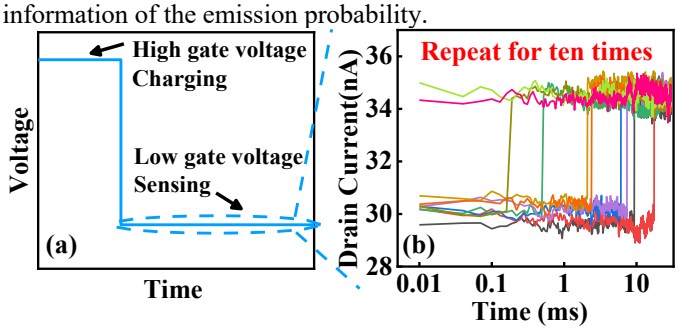

Fig. 2. (a) The Charging-Sensing Procedure to detect trap emission.(b) Typical measurement result with 10 repeating times. (on a 270nmx27 nm device with 0.5V high gate voltage and 0.05V low gate voltage)

B. Challenge-response-pairs and authentication

The *challenge* of Prob-PUF specifies the position of the transistor that is to be stimulated. Assuming that a group of transistors are selected in the PUF circuit, the charging-and-sensing voltage pattern is repeatedly applied onto each of these selected transistors 10 times. For each transistor, if no emission events are detected within the pre-set time window (t_w) for all these 10 times, a '0' bit is created. Similarly, if the emission can be detected all these 10 times, a '1' bit is created. When there are both '0' and '1' in these 10 repeats, a random bit (either '0' or '1') is generated. What is worth noting is that such repeated measurements are not for minimizing the unstable output bits like temporal majority voting methods [14, 15]. Instead, this procedure is to determine whether the output bit is stable or not. Both the stable and random bits will be used in the *response*, as illustrated on the left side of Fig. 3.

On the server side, given the pre-stored τ_e for each selected transistor, the emission probability (P_e) at the end of the pre-set time window t_w can be calculated with Eqn (1). Since we repeat 10 times for each transistor in the client circuit, in principle, any trap with emission probability lower than 10% or higher than 90% can trigger an all- '0' or '1' bit stream, which in turn creates '0' or '1' as one output bit of the PUF. Since such bit is deterministic, we define it as the *stable bit*. Such 10% and 90% is the low P_e threshold (P_{eL}) and high P_e threshold (P_{eH}). Considering its stochastic nature, in practice, we can select P_{eL} and P_{eH} of 10 times larger (e.g. 1% and 99%). If the calculated P_e is within the range between P_{eL} and P_{eH}, the output bit of the PUF is random, and such output bit is defined as the *random bit*.

We only compare the *stable bits* between the PUF circuit output and the calculated value when running the authentication process. The PUF output bits that the server considers as *random bits* will not be used in the authentication, as illustrated in the middle of Fig. 3. Such benefits are two-fold: 1) Given each *challenge* (C), since the apparent *response* (R) contains both random and stable bits, even for C-R pairs that are collected by the attacker, they are difficult to be used for training in ML. 2) there is no need to save CRP tables on the server. Instead, only τ_e of each transistor needs to be saved. This can be highly efficient for key management.

C. Debiasing of the response bit

Considering that the response of the proposed PUF includes both the stable and random bits, the following two actions have been taken to ensure the balance between '0's and '1's.

- For the stable bits: We can measure multiple transistors used in the PUF and determine the distribution of the trap emission time. Then a time window can be selected to ensure the stable bits with 50% probability for being '0' or '1'. Since the emission time variability is specific to the fabrication process, the stable bit generation with 50% probability of '0' and '1' can be ensured regardless of the type of transistors selected.
- For the random bits: For any bit detected unstable in our proposed PUF, a true random number generator (TRNG) is triggered. This unstable bit is then replaced with the output of the random number generator. This ensures that the unstable bits also exhibit 50% probability of '0's and '1's.

D. Circuit structure for the Prob-PUF

The overall structure of the Prob-PUF design consists of several circuit blocks. The Charging and Biasing block creates the charging-and-sensing voltage to the specified transistors in the transistor matrices, while the Sense and Digitization block amplifies and processes the received signal. The voltage pattern will repeat 10 times. The output bit is detected as a stable or random bit by judging whether there is a current change in the time window. To increase reliability, the current sensing scheme [16] with Beta-Multiplier [17] can be used to suppress voltage variation. The soft dark-bit masking method [18] or designing specific rules in selecting transistors for random or stable bits can help suppress the temperature sensitivity. The transistors can also be stressed first to suppress the impact of

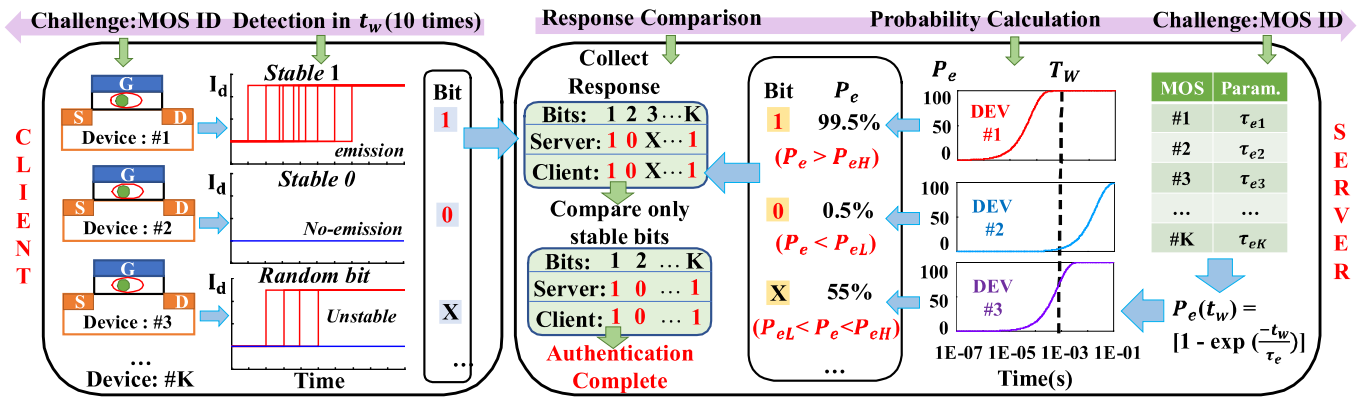

Fig.3. The authentication process of the proposed Prob-PUF. The server and client generate calculated and circuit response separately. After discarding the unstable bits in the response, the stable bits are left to be compared for authentication.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LED.2021.3130606, IEEE Electron Device Letters

aging [11]. To increase the security of the PUF, the XOR-mixed network [19] is used to combine the independent bits from the output of each transistor matrix. The bit '0' is used to replace the random bit before being fed into the XOR network using multiplexers (MUXs) to ensure the response at the output is always predictable, as shown in Fig. 4. At the outputs of the XOR network, MUXs are also added. If any bit is detected unstable, the mixed bit is replaced with a true random number to balance '0's and '1's.

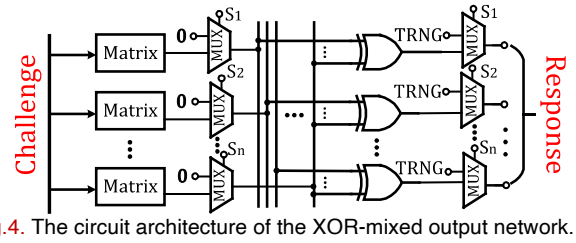

Fig.4. The circuit architecture of the XOR-mixed output network. The select signals $S_1 \ldots S_n$ are generated in the proposed test process.

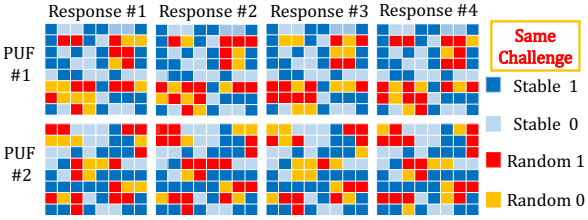

Fig.5.The variation of response bits in two different PUFs when the same Challenge was repeatedly applied.

E. CRP Space and Security Evaluation

An ideal strong PUF provides a change-response space that grows exponentially with the number of challenge bits [20]. For Prob-PUF, the transistors can be divided into K matrices, and each matrix consists of N transistors. Therefore, the whole CRP number can reach N^K, suggesting an exponential increase with challenge width. For CRP storage on the server side, Prob-PUF adopts the model-based solution [20] rather than typical table-based solution [21]. Therefore, for applications that require a large number of CRPs, the storage can be significantly reduced.

CRP space is only related to the number of unique challenges that the PUF can process [20]. Whether or not there are random bits in the response, the CRP space will not be affected. In the registration stage after fabrication, a portion of the CRPs can be randomly chosen for future authentication. The huge CRP space can effectively remove the potential threat that an attacker gets access to PUF hardware and tries to record all the CRPs, but this attack is infeasible due to the exponential number of CRPs.

To demonstrate the robustness and uniqueness of Prob-PUF, we constructed the circuit with the structure proposed in Section D. The simulation flow for stochastic process of traps is adopted [22], wherein, the energy and spatial distributions of the trap measured in transistors are used [11]. 64 transistors are randomly selected, and the authentication process is repeated for 4 times. This is to mimic the case that one *challenge is to be applied on the same PUF for multiple times*. The corresponding responses are shown in Fig. 5, where the stable and random bits are marked with different colors. As we can see, the stable bits keep the same, providing the deterministic authentication. When another batch of 64 transistors is selected to mimic the same challenge applying on different PUFs, Fig. 5 shows that the stable bits do not change with time when comparing with measurements from the same PUF, but they change with different PUFs. This supports the uniqueness of our design.

Fig. 6(a) evaluated the normalized inter-hamming distance. 1280k bits across 100 PUFs are used in the simulation. No matter whether the response contains random bits or not, the median of inter-hamming distance is close to 50% with a tight variation of 2%, suggesting the desired uniqueness of our PUF design. The intra-hamming distance is evaluated in Fig. 6(b) using 100 PUF instances. Without unstable bits, its mean value is close to 5%, and with unstable bits, the mean value reaches about 15%. This suggests the good reproducibility of each PUF.

The ability of a strong PUF to maintain unpredictability under ML attacks is an important measure of its security. Due to the stochastic nature of the trap emission, the random bits in the response could poison the training data to hinder the effective construction of the model [23]. The ML resistance of Prob-PUF is evaluated by using three widely-used methods: logistic regression (LR), support vector machines (SVM), and neural networks (NN) [24]. The arbiter PUF, 4-XOR PUF, and 4-XOR lightweight (LW) PUF were implemented by following previous work [25]. For a fair comparison, all the PUFs are provided with 64-bit challenges.

As shown in Fig. 6(c), the prediction error drops below 0.01 when training samples are more than 10^5. This agrees with the results in the literature [25], suggesting the effectiveness of our ML setup. Using the same setup, our PUF exhibits an ideal prediction error of 50% even when the training set approaches one million, suggesting its good immunity to the ML attack.

To strengthen security against physical attacks, the e-fuse based circuits can be designed [26,27]. After manufacturing and before selling, the e-fuse is used by a trustworthy party to access the PUF responses for test and diagnosis. When PUFs are ready for deployment, the fuse is burned so that nobody can directly access the PUF responses through the fuse anymore.

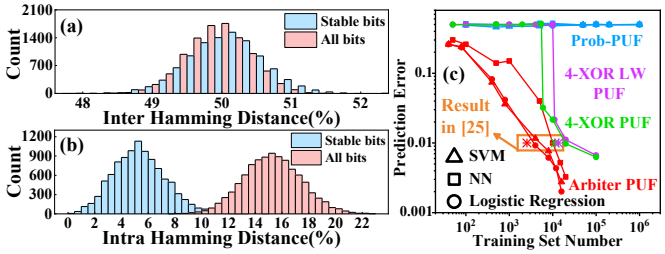

Fig.6. The evaluation of uniqueness and robustness of proposed Prob-PUF with (a) the inter-Hamming distance and (b) intra-Hamming distance for 128 bits output. (c) Machine learning result of proposed PUF compared with other classical PUF systems [25].

III. CONCLUSION

This letter proposed a novel strong PUF design, called Probability-based PUF (Prob-PUF). Prob-PUF uses the probability of stochastic charge emission events in nano-scaled FETs. The proposed design provides a large CRP space that increases exponentially with circuit size and shows excellent inter- and intra- hamming distance. The uniqueness and robustness of the proposed PUF are also demonstrated. The PUF response is mixed with stable and random bits and thus achieves strong ML-immune capability. In addition, instead of managing the CRP table, the new PUF architecture stores the mathematical models at the server side. It thus can be efficient for key management in future large scale PUF systems.

REFERENCES

[1] U. Rührmair, S. Devadas, and F. Koushanfar, "Security Based on Physical Unclonability and Disorder," in *Introduction to Hardware Security and Trust*, M. Tehranipoor and C. Wang, Eds., New York, NY, USA: Springer New York, 2012, pp. 65–102, doi: 10.1007/978-1-4419-8080-94.

[2] Y. Gao, S. F. Al-Sarawi, and D. Abbott, "Physical unclonable functions," *Nat. Electron.*, vol. 3, no. 2, pp. 81–91, Feb. 2020, doi: 10.1038/s41928-020-0372-5.

[3] G. E. Suh and S. Devadas, "Physical Unclonable Functions for Device Authentication and Secret Key Generation," in *Proc. Des. Autom. Conf. (DAC)*, Jun. 2007, pp. 9–14. doi: 10.1145/1278480.1278484.

[4] J. Guajardo, S. S. Kumar, G.-J. Schrijen, and P. Tuyls, "FPGA Intrinsic PUFs and Their Use for IP Protection," in *Cryptographic Hardware and Embedded Systems - CHES 2007*, Berlin, Heidelberg, Germany: Springer Berlin Heidelberg, 2007, pp. 63–80, doi: 10.1007/978-3-540-74735-2_5.

[5] R. Kumar and W. Burleson, "On design of a highly secure PUF based on non-linear current mirrors," in *Proc. Int. Symp. Hardw.-Oriented Secur. Trust (HOST)*, May. 2014, pp. 38–43, doi: 10.1109/HST.2014.6855565.

[6] S. Jeloka, K. Yang, M. Orshansky, D. Sylvester, and D. Blaauw, "A sequence dependent challenge-response PUF using 28nm SRAM 6T bit cell," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2017, pp. C270–C271, doi: 10.23919/VLSIC.2017.8008504.

[7] H. Zhuang, X. Xi, N. Sun, and M. Orshansky, "A Strong Subthreshold Current Array PUF Resilient to Machine Learning Attacks," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 67, no. 1, pp. 135–144, Jan. 2020, doi: 10.1109/TCSI.2019.2945247.

[8] L. Santiago, V. C. Patil, C. B. Prado, T. A. O. Alves, L. A. J. Marzulo, F. M. G. Franca, and S. Kundu, "Realizing strong PUF from weak PUF via neural computing," in *Proc. IEEE Int. Symp. Defect Fault Toler. VLSI Nanotechnol. Syst. (DFT)*, Oct. 2017, pp. 1–6, doi: 10.1109/DFT.2017.8244433.

[9] W. Zheng, X. Pan, and X. Zhao, "A low power current mode PUF based on winner-take-all scheme," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May. 2019, pp. 2–6, doi: 10.1109/ISCAS.2019.8702712.

[10] T. Grasser, "Stochastic charge trapping in oxides: From random telegraph noise to bias temperature instabilities," *Microelectron. Reliab.*, vol. 52, no. 1, pp. 39–70, Jan. 2012, doi: 10.1016/j.microrel.2011.09.002.

[11] J. Brown, R. Gao, Z. Ji, J. Chen, J. Wu, J. Zhang, B. Zhou, Q. Shi, J. Crowford, and W. Zhang, "A low-power and high-speed True Random Number Generator using generated RTN," in *Proc. Symp. VLSI Technol. (VLSIT)*, Jun. 2018, pp. 95–96, doi: 10.1109/VLSIT.2018.8510671.

[12] T. Grasser, H. Reisinger, P. J. Wagner, and B. Kaczer, "Time-dependent defect spectroscopy for characterization of border traps in metal-oxide-semiconductor transistors," *Phys. Rev. B - Condens. Matter Mater. Phys.*, vol. 82, no. 24, pp. 1–10, Dec. 2010, doi: 10.1103/PhysRevB.82.245318.

[13] J. Chen, T. Tanamoto, H. Noguchi, and Y. Mitani, "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 *Proc. Symp. VLSI Technol. (VLSIT)*, Jun. 2015, pp. T40–T41, doi: 10.1109/VLSIT.2015.7223695.

[14] S. K. Mathew, S. K. Satpathy, M. A. Anders, H. K, S. K. Hsu, A. Agarwal, G. K. Chen, R. J. Parker, R. K. Krishnamurthy and V. De, "16.2 A 0.19pJ/b PVT-variation-tolerant hybrid physically unclonable function circuit for 100% stable secure key generation in 22nm CMOS," in *Proc. IEEE Int. Solid State Circuits Conf. (ISSCC)*, Feb. 2014, pp. 278–279, doi: 10.1109/ISSCC.2014.6757433.

[15] A. Vijayakumar, V. C. Patil, and S. Kundu, "On Improving Reliability of SRAM-Based Physically Unclonable Functions," *J. Low Power Electron. Appl.*, vol. 7, no. 1, 2, Jan. 2017, doi: 10.3390/jlpea7010002.

[16] M. Simicic, S. Morrison, B. Parvais, P. Weckx, B. Kaczer, K. Sawada, H. Ammo, S. Yamakawa, K. Nomoto, M. Ohno, D. Linten, D. Verkest, P. Wambacq, G. Groeseneken and G. Gielen, "A fully-integrated method for RTN parameter extraction," in *Proc. Symp. VLSI Technol. (VLSIT)*, Jun. 2017, pp. T132–T133, doi: 10.23919/VLSIT.2017.7998151.

[17] S. Liu and R. J. Baker, "Process and temperature performance of a CMOS beta-multiplier voltage reference," in *Proc. Midwest Symp. Circuits Syst.*, Aug. 1998, pp. 33–36, doi: 10.1109/MWSCAS.1998.759429.

[18] V. Suresh, R. Kumar, M. Anders, H. Kaul, V. De, and S. Mathew, "A 0.26% BER, 10^{28} Challenge-Response Machine-Learning Resistant Strong-PUF in 14nm CMOS Featuring Stability-Aware Adversarial Challenge Selection," in Proc. *IEEE. Symp. VLSI Circuits*, Jun. 2020, pp. 1–2, doi: 10.1109/VLSICircuits18222.2020.9162890.

[19] M. Majzoobi, F. Koushanfar, and M. Potkonjak, "Lightweight secure PUFs," in *Proc. IEEE ACM Int. Conf. Comput. Des. (ICCAD)*, Nov. 2008, pp. 670–673, doi: 10.1109/ICCAD.2008.4681648.

[20] C. Herder, M. Yu, F. Koushanfar, and S. Devadas, "Physical Unclonable Functions and Applications: A Tutorial," *Proc. IEEE*, vol. 102, no. 8, pp. 1126–1141, May. 2014, doi: 10.1109/JPROC. 2014.2320516.

[21] H. Awano and T. Sato, "Ising-PUF: A machine learning attack resistant PUF featuring lattice like arrangement of Arbiter-PUFs," in *Proc. Des., Autom. Test Europe Conf. Exhib. (DATE)*, March. 2018, pp. 1447–1452, doi: 10.23919/DATE.2018.8342239.

[22] V. V. A. Camargo, B. Kaczer, T. Grasser, and G. Wirth, "Circuit simulation of workload-dependent RTN and BTI based on trap kinetics," *Microelectron. Reliab.*, vol. 54, no. 11, pp. 2364–2370, Nov. 2014, doi: 10.1016/j.microrel.2014.06.003.

[23] S. Wang, Y. Chen, and K. S. Li, "Adversarial Attack against Modeling Attack on PUFs," in *Proc. Des. Autom. Conf. (DAC)*, Jun. 2019, pp. 1–6, doi: 10.1145/3316781.3317761.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017, doi: 10.1145/3065386.

[25] U. Ruhrmair, J. Solter, F. Sehnke, X. Xu, A. Mahmoud, V. Stoyanova, G. Dror, J. Schmidhuber, W. Burleson, and S. Devadas, "PUF Modeling Attacks on Simulated and Silicon Data," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 11, pp. 1876–1891, Nov. 2013, doi: 10.1109/TIFS.2013.2279798.

[26] J. Ye, Q. Guo, Y. Hu, and X. Li, "Deterministic and Probabilistic Diagnostic Challenge Generation for Arbiter Physical Unclonable Function," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 37, no. 12, pp. 3186–3197, Dec. 2018, doi: 10.1109/TCAD.2018.2801224.

[27] M. Rostami, M. Majzoobi, F. Koushanfar, D. S. Wallach, and S. Devadas, "Robust and Reverse-Engineering Resilient PUF Authentication and Key-Exchange by Substring Matching," *IEEE Trans. Emerg. Top. Comput.*, vol. 2, no. 1, pp. 37–49, March. 2014, doi: 10.1109/TETC.2014.2300635.