0.3 pJ/Bit Machine Learning Resistant Strong PUF Using Subthreshold Voltage Divider Array

Abilash Venkatesh, Aishwarya Bahudhanam Venkatasubramaniyan, Xiaodan Xi, and Arindam Sanyal

***Abstract*—This brief presents a subthreshold voltage divider based strong physical unclonable function (PUF). The PUF derives its uniqueness from random mismatch in threshold voltage in an inverter with gate and drain shorted and biased in subthreshold region. The nonlinear current-voltage relationship in subthreshold region also makes the proposed PUF resistant to machine learning (ML) based attacks. Prediction accuracy of PUF response with logistic regression, support vector machine (SVM) and multi-layer perceptron (MLP) is close to 51%. A prototype PUF fabricated in 65nm consumes only 0.3pJ/bit, and achieves the best combination of energy efficiency and resistance to ML attacks. The measured inter and intra hamming distance (HD) for the PUF are 0.5026 and 0.0466 respectively.**

***Index Terms*—Physical unclonable function (PUF), hardware security, strong PUF, machine learning.**

I. Introduction

Si PHYSICAL unclonable function (PUF) are lightweight hardware primitives that leverage random variations in CMOS integrated circuits to generate a unique key that can be used for authentication protocols or chip identification. Compared to standard cryptographic algorithms like SHA or AES, Si PUFs provide on-chip security at a small fraction of power, thus making Si PUFs attractive candidate for secure Internet-of-Things (IoT) applications. Based on number of challenge-response pairs (CRPs), PUFs can be grouped into strong PUFs or weak PUFs. Strong PUFs have a large number of CRPs which grows exponentially with area while weak PUFs typically have small number of CRPs which grows linearly with area.

The first strong Si-PUF is an arbiter PUF [1] which uses variation in delay between two nominally identical paths to generate a 1-bit response. Variants on arbiter PUF include using XOR-ing of arbiter PUF outputs [2] and using feed-forward paths [3] to inject nonlinearity. Another widely used PUF is SRAM-PUF which can be used as both weak PUF [4] or as strong PUF [5]. The current-mirror based weak PUF [6] achieves high reproducibility without post-processing but at the cost of large area. To reduce area, [7] performs amplification using series of NAND gates while the weak PUF in [8] cascades 2-transistor subthreshold-region amplifiers to achieve rail-to-rail swing.

Recent works [9], [10] have shown that most existing PUF models can be broken and PUF response predicted with high accuracy (90 $\sim$ 99%) through the use of advanced machine learning (ML) models such as logistic regression or support-vector machine (SVM). The introduction of ML attacks pose a serious threat to security provided by PUFs. There has been a few recent works that have been shown to be robust against ML attacks. A recent strong PUF uses subthreshold current array [11] which has a strong nonlinearity arising out of MOSFET subthreshold operation leading to ML prediction accuracy of 60%. The work in [12] uses strong nonlinearity of convergence time in a bistable ring arising out of variations in threshold voltages to limit prediction accuracy of ML attacks to 50%.

In this brief, we propose a subthreshold voltage-divider array based strong PUF which achieves simultaneous low energy consumption and high resistance to ML attacks. We had previously presented simulation results on a voltage-divider array PUF in [13]. In this brief, we present more details on nonlinearity in the proposed PUF as well as demonstrate robustness of the proposed PUF against ML attacks. In addition, we present measurement results on a 65nm prototype which indicates that prediction accuracy of different ML algorithms, such as SVM, logistic regression and multi-layer perceptron, is close to 51% for the proposed PUF. The prototype PUF has an energy consumption of 0.3pJ/bit. The rest of this brief is organized as follows: Section II presents architecture of the proposed PUF and measurement results and comparison with state-of-the-art are presented in Section III. The conclusion is brought up in Section IV.

II. Proposed Architecture

The proposed PUF architecture is shown in Fig. 1(a). A unit PUF cell comprises an inverter with gate and drain shorted and an NMOS tail current source biased in subthreshold. A 1-b PUF output is obtained by comparing the drain voltage of two such unit PUF cells. To form a strong PUF, we use two arrays of N "nominally identical" unit PUF cells. The challenge inputs, C_1 through C_N, determine which of the N unit PUF cells in both arrays are connected to the differential inputs

Manuscript received August 16, 2019; accepted September 19, 2019. Date of publication September 23, 2019; date of current version August 4, 2020. This brief was recommended by Associate Editor W. N. N. Hung. *(Corresponding author: Abilash Venkatesh.)*

A. Venkatesh and A. Sanyal are with the Department of Electrical Engineering, University at Buffalo, Buffalo, NY 14260 USA (e-mail: abilashv@buffalo.edu).

A. B. Venkatasubramaniyan is with Intel Corporation, Nonvolatile Memory Section Group, Folsom, CA 95630 USA.

X. Xi is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 USA.

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

Digital Object Identifier 10.1109/TCSII.2019.2943121

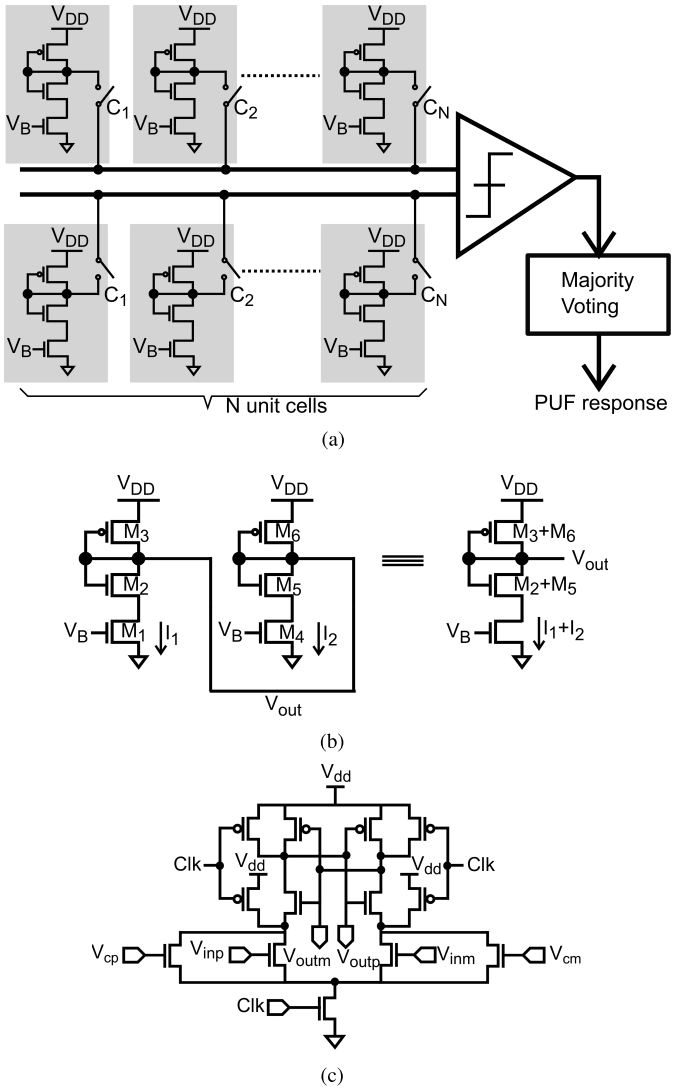

Fig. 1. (a) Proposed subthreshold voltage divider array PUF (b) nonlinearity in PUF array with 2 unit cells (c) comparator schematic.

of a comparator. For this design, we have 60 unit PUF cells in each array corresponding to 2^{60} possible CRPs. The PUF array used in this design has an intrinsic advantage over the current-array PUF of [11] in that the comparator input common-mode voltage does not vary significantly with the challenge pattern. Variation of comparator input common-mode voltage results in challenge pattern dependent offset and is problematic for comparator offset cancellation [11]. The proposed PUF does not need any common-mode feedback loop which reduces energy consumption significantly compared to [11].

Fig. 1(b) shows how nonlinearity is introduced in output voltage of the proposed PUF using 2 unit cells. The currents I_1 and I_2 through the two unit PUF cells can be written as

$$I_1 = I_s \exp\left(\frac{V_B - V_{th1}}{\eta V_T}\right);\ I_2 = I_s \exp\left(\frac{V_B - V_{th4}}{\eta V_T}\right) \quad (1)$$

where V_{th1} and V_{th4} denotes threshold voltages of M_1 and M_4 respectively, V_T is thermal voltage kT/q, V_B is biasing voltage for NMOS tail current source and it is assumed that V_{ds} of M_1 and M_4 is greater than 100mV. When the two unit PUF cells are connected together, the output voltage V_{out} can be expressed as

$$I_1 + I_2 = I_{s1} \exp\left(\frac{V_{DD} - V_{out} - |V_{thp}|}{\eta V_T}\right) \quad (2)$$

$$V_{out} = V_{DD} - |V_{thp}| - \ln\left(\frac{I_s}{I_{s1}}\right) - \ln\left[\exp\left(\frac{V_B - V_{th1}}{\eta V_T}\right) + \exp\left(\frac{V_B - V_{th4}}{\eta V_T}\right)\right] \quad (3)$$

where $|V_{thp}|$ is the threshold voltage of PMOS transistor. It can be seen that (3) is a transcendental equation and V_{out} is nonlinear in terms of threshold voltages of the NMOS tail current sources. Second-order effects such as drain induced barrier lowering further increase the coupling between threshold voltage and V_{out}. Since threshold voltage exhibits large intrinsic variation due to random dopant fluctuation, each PUF unit element has large entropy. As shown in (3), entropy of different PUF unit elements is not simply additive unlike that in arbiter PUF [10]. Hence, the proposed PUF is expected to be robust against modeling attacks including ML attacks as we will show later in Section III. To show that entropy of unit elements is not additive, we performed simulations on a 4-element model of the proposed PUF. We first applied one-hot challenges, ({0001}, {0010}, {0100}, and {1000}), to extract the entropy, Δ_i, for i-th unit element in the two arrays. We used noiseless, ideal comparator and performed 500 monte-carlo runs for each challenge. We then modeled the PUF using a linear equation $t = sgn(\vec{\Delta}^T\vec{\phi})$ where $\vec{\Delta}$ encodes entropy vector and $\vec{\phi}$ is the challenge vector. We then used the linear additive model to predict the PUF output for the other 11 challenge vectors. The prediction accuracy varied from 0.58 to 0.71 with mean of 0.66 and standard deviation of 0.067, thus showing that systematic one-hot challenges cannot be used to accurately model the proposed PUF, unlike PUFs with additive entropy. It should be noted here that the prediction accuracy corresponds to a scenario in which the attacker has access to comparator input nodes for extracting the entropy values using one-hot challenges. For non-invasive modeling attacks, only the sign of entropy values are available to the attacker, and the mean prediction accuracy for linear additive model drops to 0.537.

Fig. 1(c) shows schematic of the comparator used in this design. A strong-arm latch is used as comparator. The comparator has two auxiliary input transistors for offset calibration. The auxiliary transistors are controlled by the voltages V_{cp} and V_{cm} which are used to tune the comparator offset [11]. During offset calibration phase, the comparator inputs are shorted, the comparator is fired multiple times and the distribution of '1' in the comparator output is observed. If the comparator has an offset, its output will have unequal distribution of '0' and '1'. The voltages V_{cp} and V_{cm} are used to bias the auxiliary input transistors to ensure the comparator has approximately equal distribution of '0' and '1'. For this design, comparator offset calibration is done in the foreground at nominal conditions of 0.9V power supply and at room temperature.

Fig. 2 shows the distribution of PUF differential output voltage ΔV_{out} for two extreme cases: (a) when only 1 challenge

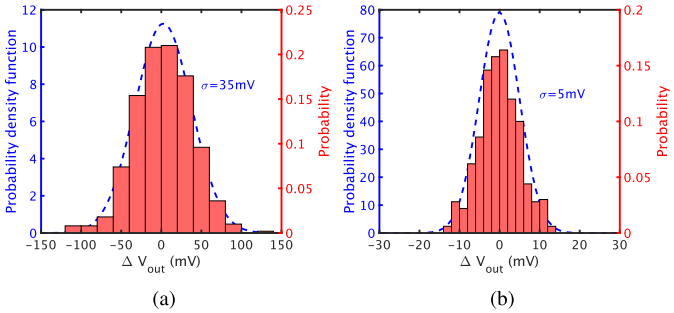

Fig. 2. Distribution of PUF differential output voltage for (a) when only 1 challenge input is '1' (b) when all challenge inputs are '1'.

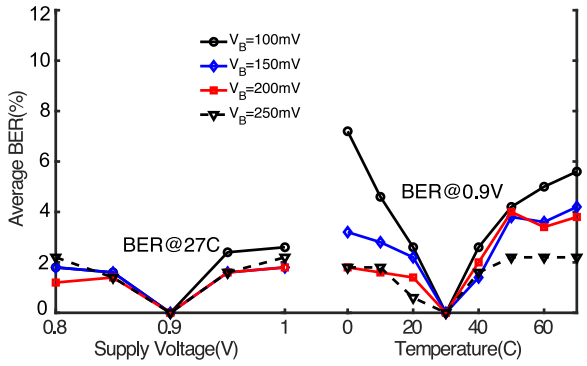

Fig. 3. Average simulated BER across supply voltage and temperature.

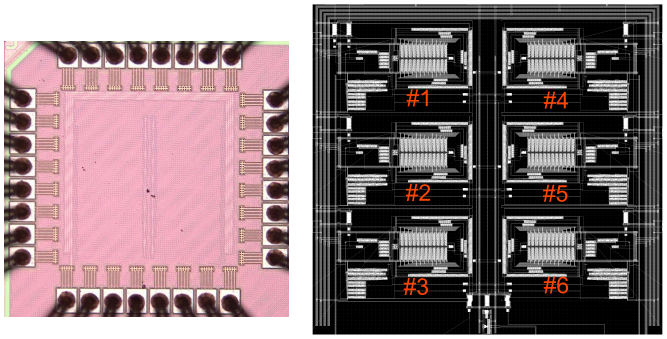

Fig. 4. Die photo and layout.

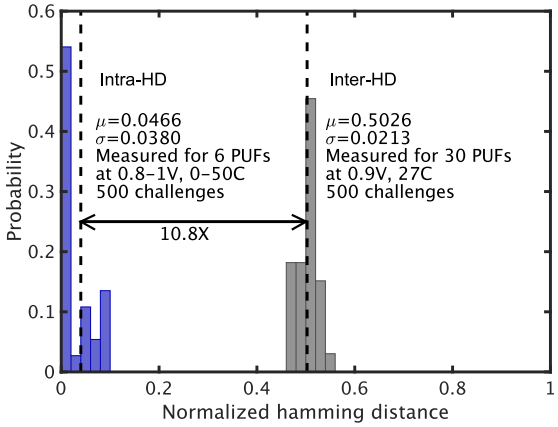

Fig. 5. Measured normalized intra and inter-HD.

input is '1' and (b) when all challenge inputs are '1'. The distributions are extracted from 500 monte-carlo runs. When only 1 challenge input is '1', ΔV_{out} has a large spread and a standard deviation, σ_{mis}, of 35mV. When all challenge inputs are '1', σ_{mis} reduces to 5mV.

The comparator is the dominant noise source and it has to be designed such that comparator noise is much smaller than the worst-case σ_{mis} of 5mV for the PUF to have high native stability. There is a trade-off between power, comparator noise and bit-error rate (BER) which can be optimized by tuning the biasing voltage, V_B. As V_B is reduced, the current through unit PUF cells reduces, which reduces total power, but the comparator input common-mode voltage increases, which increases comparator noise [14]. Assuming the comparator has a noise standard deviation of σ_n, the native stability of the PUF can be written as stability $= 1 - \text{erf}(\sigma_n/\sqrt{2}/\sigma_{mis})$ For this design, V_B is set to 200mV such that the comparator noise has a standard deviation of 350μV which corresponds to a native stability of 94.4%. In order to improve PUF native stability, we perform temporal majority voting of the comparator output. Application of majority voting of 7 reduces σ_n to 132μV which improves PUF native stability to 97.9%. We use a 3-bit counter which counts up everytime the comparator output is '1'. The MSB of the counter is used as 1-b PUF output. The counter is reset every 8th cycle of the comparator clock. Reduction in V_B also increases susceptibility to temperature and worst-case BER increases for small V_B as shown in Fig. 3. The BER values are calculated for 5 PUFs with 500 monte-carlo simulations for each voltage and temperature points and with an offset-less comparator. The worst-case BER is less than 5% across voltage and temperature corners for V_B of 200mV.

III. Measurement Results

A test chip is fabricated in 65nm CMOS process. The die micro-photograph and layout are shown in Fig. 4. Each test chip contains 6 PUFs. Each PUF has an area of 110μm×170μm with the core (PUF array+comparator) occupying 40μm×70μm. Each PUF consumes 3.8μW power from 0.9V supply with a throughput of 12.5M samples/s. Out of the 3.8μW power, the comparator consumes 1.2μW power while the PUF array consumes 2.6μW power.

Fig. 5 shows the measured normalized intra and inter-HD of the PUF. Intra-HD is measured for 6 PUFs over a supply range of 0.8V-1V and temperature range of 0-50C for 500 challenges. The measured intra-HD is 0.0466 with a standard deviation of 0.038. Inter-HD is measured for 30 PUFs at 0.9V supply and 27°C for 500 challenges. The measured inter-HD is 0.5026 with a standard deviation of 0.0213. The ratio between inter-HD to intra-HD is 10.8 which indicates that the proposed PUF has a high uniqueness. Fig. 6 shows the measured average BER across supply voltage and temperature. The BER with variation in supply voltage and temperature is measured with respect to PUF output at 0.9V supply and 27°C temperature. Temperature variation affects BER more than supply voltage variation. The worst case BER is 10.9%.

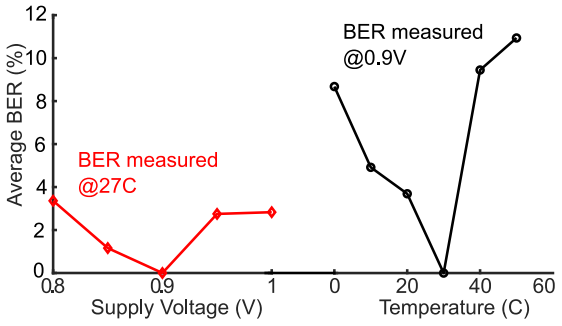

Fig. 6. Average BER across supply voltage and temperature.

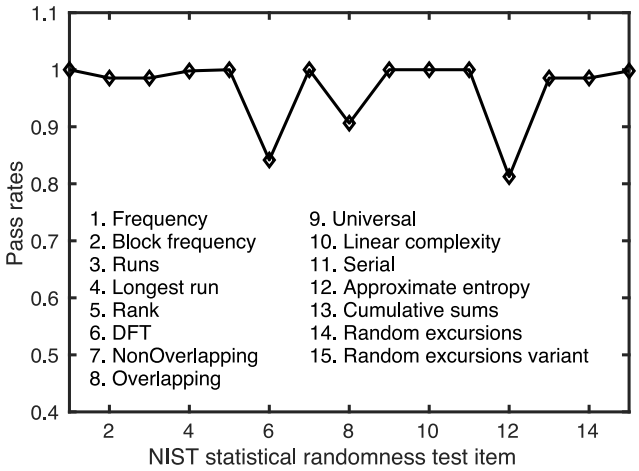

Fig. 7. NIST randomness test results.

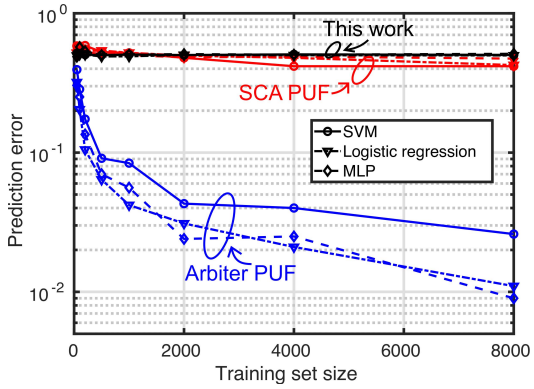

Fig. 8. ML Prediction error vs training set size.

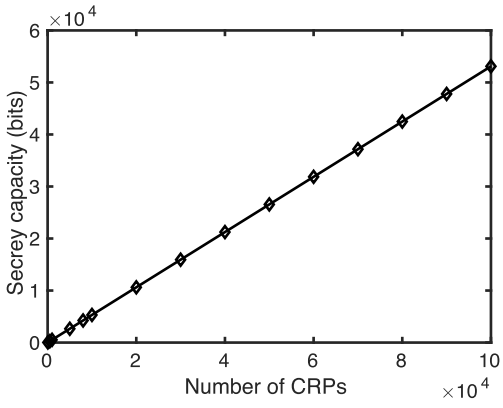

Fig. 9. Measured secrecy capacity versus CRPs.

In order to test the randomness of our PUF, we used NIST SP 800-22 randomness tests on 30 PUF devices from 5 different test chips. For each PUF, we recorded the response for 16 different times. Thus, the NIST tests are performed on 480 PUF response streams. Fig. 7 shows graphically the results of the NIST tests. For 12/15 NIST tests, the minimum pass rate was greater than 0.95. The pass rate for 3 tests, DFT, overlapping template and approximate entropy, was between 0.85-0.9. The NIST test results indicate good randomness of the proposed PUF.

In order to test the susceptibility of the proposed PUF to machine learning attacks, we used three different machine learning algorithms: support vector machine with nonlinear radial bias function (RBF) kernel, logistic regression, and multi-layer perceptron (MLP) with two layers and 30 hidden neurons. Fig. 8 shows the results of machine learning attacks on our PUF as well as on the switched current array (SCA) PUF [11] and arbiter PUF. For small training set sizes, the worst-case prediction error rate for the proposed PUF and SCA is close to 0.5, while the worst-case prediction error rate for arbiter PUF is close to 0.3. As the training set size is increased, the worst-case prediction error rate for our PUF does not change significantly and remains close to 0.5. The SCA PUF also exhibits similar performance and as training set size increases, the wost-case prediction error rate for SCA PUF is 0.4. On the other hand, the arbiter PUF has a worst-case prediction error rate of 0.3 at small training set sizes which reduces quickly to 0.009 as the training set size is increased to 8000. Thus, use of subthreshold nonlinearity makes the proposed PUF and SCA PUF robust against the three different machine learning attacks we tried, while the arbiter PUF can be easily modeled with training set sizes > 200.

An interesting aspect of PUF performance measurement is its usability as secure key generator. As shown in [15], PUF response can be considered as fuzzy secrets from which secure keys can be extracted. Since response of the same PUF to different challenges may not be completely independent, the question is how many secure bits can be extracted from an M-bit PUF response? The secrecy capacity, $S(X^M)$ is defined as the maximum number of secure bits that can be extracted from an M-bit PUF response [15]. As shown in [15], the upper bound on $S(X^M)$ is defined as

$$S(X^M) \leq \sum_{i=1}^{M} h(SR(i-1)) - M \cdot h(p_e) \tag{4}$$

where $SR(i)$ is the prediction accuracy of PUF modeling attack based on an i-bit PUF response, $h()$ is the entropy function defined as $h(p) = -p \cdot log_2(p) - (1-p) \cdot log_2(1-p)$ and p_e is the bit-error rate of the PUF. Fig. 9 shows the measured secrecy capacity versus number of CRPs for our PUF. The secrecy capacity of our PUF increases linearly with number of bits in the PUF response, thus showing that the proposed PUF can be used to generate secure keys. In contrast, as shown in [11], [15] the secrecy capacity of arbiter PUF does not increase linearly

TABLE I
COMPARISON WITH STATE-OF-THE-ART STRONG PUFS

	This Work	[11] VLSI'17	[2] ACM'07	[8] ISSCC'17	[1] VLSI'04	[5] VLSI'17	[12] TCAS2'18
Technology (nm)	**65**	130	90	40	180	28	65
Type of PUF	**Voltage array**	Current array	Ring oscillator		Arbiter	SRAM	Bistable ring
Possible CRPs	**1.15 x 10^{18}**	$\approx 3.7 \times 10^{19}$	523776	$\approx 5.5 \times 10^{28}$	$\approx 1.4 \times 10^{20}$	1.17×10^{11}	131071
ML prediction error (10^4 CRPs)	**49%**	40%	1%	–	1%	10.6%	50%
Worst-case BER	**10.9%**	9% (0.4%*)	0.48%	9%	4.8%	3.17%	–
Energy/bit (pJ/bit)	**0.3**	11	–	17.75	–	0.097	–
Voltage range (V)	**0.8-1**	1.08-1.32	1.08-1.2	0.7-1.2	1.75-1.85	0.5-0.9	1.6-1.8
Temperature range (°C)	**0-to-50**	−20-to-80	20-to-120	−25-to-125	20-to-70	0-to-80	20-to-50
Inter-HD	**0.5026**	0.499	0.4615	0.5007	0.4	0.481-0.495	–
Intra-HD	**0.0466**	0.058	0.0048	0.0101	0.0357	0.0317	–

* after discarding 42% CRPs

with number of bits in PUF response and the arbiter PUF has a maximum secrecy capacity of 600 for 5000-bit response. This result is expected as it has been shown that machine learning attacks can successfully model arbiter PUF with high accuracy. Thus, for long PUF responses ($M > 200$), the response bits for arbiter PUF can be easily predicted. For the proposed PUF, the prediction accuracy for long PUF responses remains low, thus resulting in a secrecy capacity which increases linearly with M.

Table I compares this brief with state-of-the-art strong PUFs. Compared to existing work, our PUF achieves simultaneous high energy efficiency and strong resistance to ML attacks. While the proposed PUF has similar resistance to ML attacks as [11], energy efficiency of the proposed PUF is 36× better than [11]. The SRAM PUF of [5] has 3× better energy efficiency than the proposed PUF, but is 5× more susceptible to ML attacks than the proposed PUF.

IV. CONCLUSION

A subthreshold voltage-divider array based strong PUF is proposed in this brief. Voltage output of the proposed PUF has a strong nonlinear dependence on threshold voltage which results in robustness against ML based modeling attacks. A 65nm prototype consumes only 0.3pJ/bit and has prediction accuracy of 51% with three different ML algorithms. The BER can be improved by post-processing or CRP reduction.

REFERENCES

[1] J. W. Lee, D. Lim, B. Gassend, G. E. Suh, M. Van Dijk, and S. Devadas, "A technique to build a secret key in integrated circuits for identification and authentication applications," in *Proc. IEEE Symp. VLSI Circuits*, Honolulu, HI, USA, 2004, pp. 176–179.

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

[3] B. Gassend, D. Lim, D. Clarke, M. Van Dijk, and S. Devadas, "Identification and authentication of integrated circuits," *Concurrency Comput. Pract. Exp.*, vol. 16, no. 11, pp. 1077–1098, 2004.

[4] S. K. Mathew *et al.*, "A 0.19 pJ/b PVT-variation-tolerant hybrid physically unclonable function circuit for 100% stable secure key generation in 22nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers (ISSCC)*, 2014, pp. 278–279.

[5] 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*, Kyoto, Japan, 2017, pp. C270–C271.

[6] A. B. Alvarez, W. Zhao, and M. Alioto, "Static physically unclonable functions for secure chip identification with 1.9–5.8% native bit instability at 0.6–1 V and 15 fJ/bit in 65 nm," *IEEE J. Solid-State Circuits*, vol. 51, no. 3, pp. 763–775, Mar. 2016.

[7] B. Karpinskyy, Y. Lee, Y. Choi, Y. Kim, M. Noh, and S. Lee, "8.7 Physically unclonable function for secure key generation with a key error rate of 2E-38 in 45nm smart-card chips," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, San Francisco, CA, USA, 2016, pp. 158–160.

[8] K. Yang, Q. Dong, D. Blaauw, and D. Sylvester, "14.2 A physically unclonable function with BER< $10e^{-8}$ for robust chip authentication using oscillator collapse in 40nm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, San Francisco, CA, USA, 2015, pp. 1–3.

[9] S. S. Zalivaka, A. A. Ivaniuk, and C.-H. Chang, "Low-cost fortification of arbiter PUF against modeling attack," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Baltimore, MD, USA, 2017, pp. 1–4.

[10] U. Rührmair *et al.*, "PUF modeling attacks on simulated and silicon data," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 11, pp. 1876–1891, Nov. 2013.

[11] X. Xi, H. Zhuang, N. Sun, and M. Orshansky, "Strong subthreshold current array PUF with 2^{65} challenge-response pairs resilient to machine learning attacks in 130nm CMOS," in *Proc. IEEE Symp. VLSI Circuits*, Kyoto, Japan, 2017, pp. C268–C269.

[12] Y. Tanaka, S. Bian, M. Hiromoto, and T. Sato, "Coin flipping PUF: A novel PUF with improved resistance against machine learning attacks," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 65, no. 5, pp. 602–606, May 2018.

[13] A. B. Venkatasubramaniyan and A. Sanyal, "Physically unclonable function based on voltage divider arrays in subthreshold region," in *Proc. IEEE Int. Midwest Symp. Circuits Syst. (MWSCAS)*, 2018, pp. 845–848.

[14] L. Chen, A. Sanyal, J. Ma, X. Tang, and N. Sun, "Comparator common-mode variation effects analysis and its application in SAR ADCs," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Montreal, QC, Canada, 2016, pp. 2014–2017.

[15] G. Hospodar, R. Maes, and I. Verbauwhede, "Machine learning attacks on 65nm Arbiter PUFs: Accurate modeling poses strict bounds on usability," in *Proc. IEEE Int. Workshop Inf. Forensics Security (WIFS)*, Tenerife, Spain, 2012, pp. 37–42.