Ultra-Low Power Physical Unclonable Function with Nonlinear Fixed-Resistance Crossbar Circuits

M. R. Mahmoodi*, H. Nili*, Z. Fahimi, S. Larimian, H. Kim, and D. Strukov*
UC Santa Barbara, Santa Barbara, CA 93106-9560, U.S.A. email: {mrmahmoodi, hnili, strukov}@ece.ucsb.edu

Abstract - The proliferation of networked mobile devices and smart gadgets in the IoT landscape has accelerated the demand for lightweight, low power and operationally compatible cryptographic solutions. As a result, emergent hardware-intrinsic security architecture must demonstrate manufacturability, power efficiency and platform compatibility, in addition to robust security performance. Here we present a novel design of physical unclonable function, called VRPUF, and prototype it using unformed 4K-ReRAM passive crossbar circuits fabricated with a CMOS-compatible process, suitable for the back-end-of-line (BEOL) integration. The architecture utilizes intrinsic process variations in crossbar circuits, manifested as variations in device I-V nonlinearities and the leakage currents, and allows for extremely large (~10^{25}) number of challenge-response pairs (CRPs). The VRPUF design does not require forming/ programming crosspoint devices, which simplifies peripheral circuits, leading to ~ 4× better density compared to the architectures which rely on switching the states of ReRAM devices. Moreover, uniform I-Vs of the virgin-state devices, coupled with lower conductance and stronger static nonlinearity, allow for ~100× improvement in power consumption and more robust security metrics. The intrinsic reliability of VRPUF primitives is also ~4× better compared to CMOS architectures, which can be further improved via selective de-enrolling of worst-case CRPs prior to deployment.

I. Introduction

In the era of IoT, the need for robust, lightweight, and power-efficient localized security platforms is more urgent than ever. Such need is further amplified due to the rapid advances in big-data extraction and analysis, e.g. machine learning algorithms, and public access to virtually unlimited remote computing resources. Hardware-intrinsic security primitives such as physical unclonable functions (PUFs) are promising candidates for such security approaches [1]. However, ensuring robust, resilient, and secure operation in such primitives has proven challenging [2-4]. Moreover, to prove viable, these primitives must be low-cost, conducive to integration into existing platforms with minimal overhead, and conform to the power-constraints of mobile platforms. PUF primitives based on passive analog ReRAM crossbar circuits have been shown to be capable of satisfying the stringent requirements for hardware-intrinsic security primitives [1,5]. The rich space of intrinsic variations in these devices, along with their low-power selector-free operation provides unique opportunities for robust security solutions.

Herein, we utilize large passive ReRAM crossbar circuits, fabricated via a CMOS-compatible process to construct robust ultra-low power PUF primitives with small footprint, quick manufacturing and deployment turnaround (Fig. 1). The VRPUF architecture utilizes intrinsic process-induced variations in as-fabricated (unformed) crossbars circuits to extract cryptographic keys, eliminating the need for forming and programming operations and their corresponding peripheral circuitry. The virgin-state operation reduces the power consumption due to the lower conductance state of individual devices, further improving the power efficiency. The intrinsic variations in inter-device I-V nonlinearities, the main source of entropy in VRPUF response, is also significantly (~10×) more pronounced in as-fabricated devices as compared to formed devices, further improving the VRPUF response unpredictability and attack resilience.

II. VRPUF Design

VRPUF has been prototyped using 64×64 crossbar circuits with passively integrated Al(70)/TiN(45)/Al_2O_3(1.5)/TiO_{2-x}(30)/Ti(15)/Al(90)/TiN(80) devices, with thicknesses shown in nm (Fig. 2). The bi-layer (Al_2O_3/TiO_{2-x}) insulating material in such crossbars is similar to Ref. 1. However, the main difference from prior work is new etch-down patterning process allowing for higher aspect ratio and smoother electrodes, which in turn helped with improving I-V uniformity and scaling up the complexity of the crossbar circuit. Furthermore, the low-temperature budget makes the developed fabrication process suitable for BEOL CMOS integration.

The major novelty of VRPUF design is utilization of post-fabrication virgin-state crossbar circuits, whose conductance distribution (Fig. 3) provides a rich entropy source for cryptographic secret key generation. This randomness is completely intrinsic, unlike the previous ReRAM-based PUFs [1,5] in which the devices must be formed and tuned to a narrow conductance distribution. The elimination of programming/forming steps has two advantages: (1) it simplifies peripheral circuits resulting in up to a ~4× improvement in area efficiency and (2) greatly limits the options for information leakage, because adversaries are not physically capable of measuring the state of each device, individually. Furthermore, the statistical correlation between the pristine conductance states of crossbar circuits fabricated in the same process is negligible (Fig. 4).

Similar to our previous work [1,5], the proposed PUF design takes advantage of the nonlinear I-V characteristics and

device-to-device variations in resistive memories to implement Shannon's confusion paradigm. Fig. 5 shows I-V characteristics of virgin-state devices. The I-V nonlinearity and its variations are higher at larger biases (Fig. 6), with the average nonlinearity at 0.4 V is ~80%, which is ~20× larger than that of a formed 50 kΩ device [1].

One of the most serious concerns in ReRAM-based strong PUFs are outlier devices with large conductance, which create a bias in the response and strong input-output correlations. For example, the conductance distribution range in [6] is 5 orders of magnitude. Highly conductive devices, even if unselected, dominate the response. In [1], a novel tunning procedure is proposed to mitigate this issue and in [5], an auxiliary line of devices is considered to remove the correlations. Fig. 7 shows how this problem (e.g., due to unwanted forming of the devices by electrostatic discharge during wire-bonding or integration) is addressed in VRPUF's architecture. Specifically, a 192-bit length challenge is applied to the crossbar, from which 64 bits are used for row selection and the rest is utilized for column selection in two cycles. Here, a single-bit response is generated after 4 cycles. In the first two cycles, bottom (top) lines are operated as columns (rows), two bits are generated and XORed. This provides us with the opportunity to read the current from both sides and possibly generates more entropy in the output. In the last two cycles, top (bottom) lines are operated as rows (columns) and the resultant two bits are XORed likewise. The final response bit is the result of XORing the two generated bits.

III. Experimental Results

The distribution of the common-mode and differential-mode currents sensed by the amplifier are shown in Fig. 8. Sub-μA common-mode currents enable extremely low-power consumption at 0.4 V, which is quite unusual for such large-scale crossbar circuit. In fact, power consumption is ~ 100× less compared to previous designs adjusted to similar crossbar circuit complexity [5]. The tight differential-mode current distribution stems from a uniform fabrication process which has resulted in a very narrow conductance distribution. The speckle pattern depicted in Fig. 9 for 1000 128-bit keys also confirms very uniform and unbiased bit generation.

In general, a robust PUF should have negligible correlations between the input and output bits. Fig. 10 confirms that such correlations are negligible in our design, which is in part due to utilization of mutli-step selection scheme in which the readout is performed on both top and bottom electrodes.

To further verify PUF robustness, we have performed extensive sets of simulations of resiliency towards machine learning attacks. Particularly, we have used a 192×100×100×100×1 multilayer perceptron (MLP) shown in Fig. 11a, and online packages utilized in previous works [4] such as LIBSVM, and LIBLINEAR. Fig. 11b shows the predictability of the trained models verifying the machine learning attack resiliency of our PUF. Training is performed on 95% of the data among which 20% is used for cross validation and the remaining mutually exclusive 5% CRPs are used for testing. The test accuracy for all networks is close to the ideal 50% predication accuracy. Finally, we tried a long short-term memory (LSTM) consisting of two 128-unit LSTM layers (with rectified linear activation function) and two fully connected layers (with sigmoid activation function) based on the encoding scheme in [1]. We observed that the predictability of the implemented network is ~50% for 32, 64, and 128-bit input sequences. To evaluate the statistical properties of 100kb measured output data, we used the common National Institute of Standards and Technology (NIST) test suite. The bitstreams generated by VRPUF pass all relevant tests. The average p-values of each subtest is provided in Fig. 12.

We have also measured the reliability of the proposed PUF across a wide range of temperature and voltage deviations from the nominal condition. Due to the extremely low-power operation of the circuit, the differential-mode currents are sometimes comparable to the noise floor. This results in 3.73% unstable bits at the nominal conditions (Fig. 13), which is still much better than the 16% in SRAM, ~34% in latch-based PUFs, and 30% in the hybrid circuit weak PUFs reported in [3]. Moreover, the proposed design can be classified as strong PUF and given its large CRP capacity (~10^{25}), discarding the unstable bits should not impact the security metrics. Masking unstable bits will reduce the worst-case bit-error rate at +10% voltage deviation by ~1.5×. Fig. 14 shows the reliability with respect to the temperature variations assuming golden ("reference") values are defined at room temperature. The worst-case BER at 85°C is 9.5%.

The uniformity distributions of measured 64 and 128-bit keys indicate near ideal response bias of VRPUF (Fig. 15). As illustrated in Fig. 16 and Fig. 17, the average diffuseness and uniqueness of 64-bit keys generated by VRPUF instances are 49.96% and 50.03%, respectively. Finally, Table I shows a favorable quantitative comparison between VRPUF and the state-of-the-art implementations.

IV. Summary

We presented VRPUF, a strong PUF design based on crossbar circuits with simple (nonlinear, fixed I-V) crosspoint devices. VRPUF operation was experimentally verified using unformed 64×64 crossbar circuits with passively-integrated ReRAM devices, fabricated with a novel CMOS-compatible etch-down patterning process. The results showed that the proposed circuits are ~4× denser, consume ~100× less power, and feature more robust security metrics compared to the previous work. Notably, VRPUF design does not require forming/ switching of crossbar devices, and hence using resistive memory is not essential (though practical for fabrication flows with BEOL ReRAM process). The simpler device functionality broadens implementation options, which could ultimately lead to more cost-efficient CMOS integration.

References

[1] H. Nili et al. *Nat. Electron*. **1** 293 (2018). [2] S. K. Mathew et al. *ISSCC* 278 (2014). [3] S. K. Alvarez et al. *ISSCC* 146 (2015). [4] S. Jeloka et al. *VLSI Symp.* C270 (2017). [5] M. R. Mahmoodi et al. *VLSI Symp*. 176 (2018). [6] L. Gao et al. *TED* **63** 3109 (2016). [7] M. R. Mahmoodi et al. *DAC* 137 (2019)

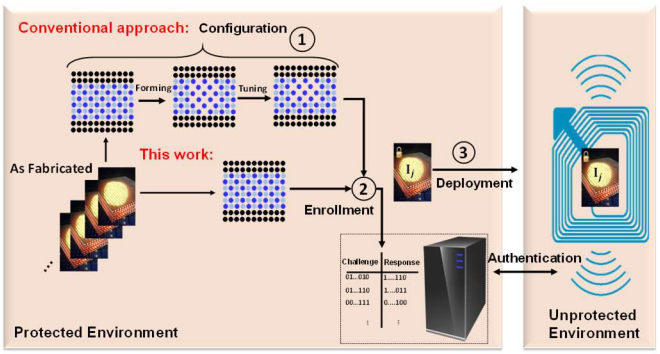


Fig. 1: Conventional three-step ReRAM-based PUF application scenario. Each instance is deployed in field after passing through pre-reconfiguration and enrolment phases. In the proposed design, the first configuration phase is skipped which improves the security, significantly reduces the power consumption and area, and speeds up the production time.

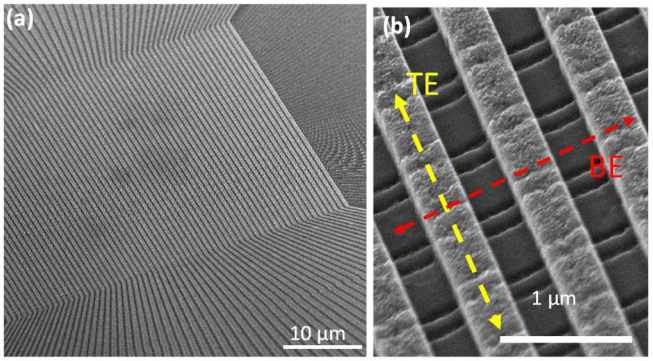


Fig. 2: Memristive array: (a) SEM image of the fabricated 4096 ReRAM crossbar array, (b) zoom-in on the portion of the crossbar.

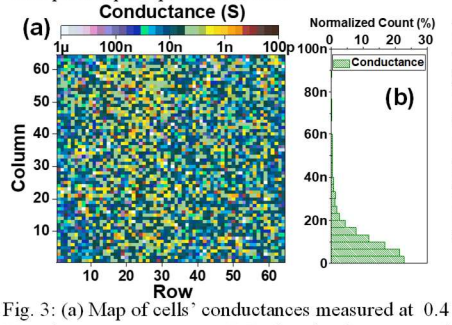


Fig. 3: (a) Map of cells' conductances measured at 0.4 V and room temperature (RT) for the demonstrated PUF and (b) its corresponding distribution.

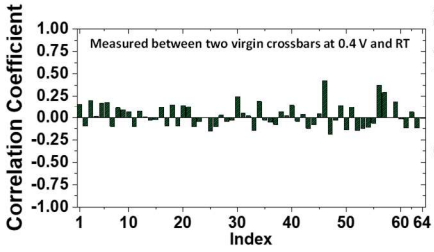


Fig. 4: Pearson's Linear Correlation Coefficient between the conductance state of two virgin crossbars. (±1 show strong correlations, 0 means no correlation.)

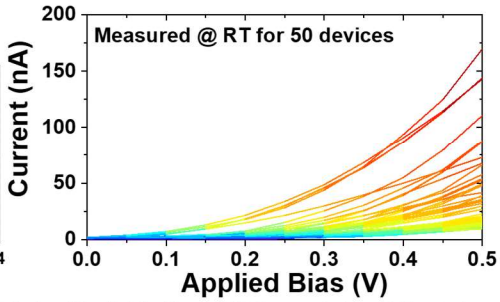


Fig. 5: Nonlinear *I-V* characteristics of 50 pristine devices in the demonstrated PUF. The currents also include electrode-to-electrode leakage.

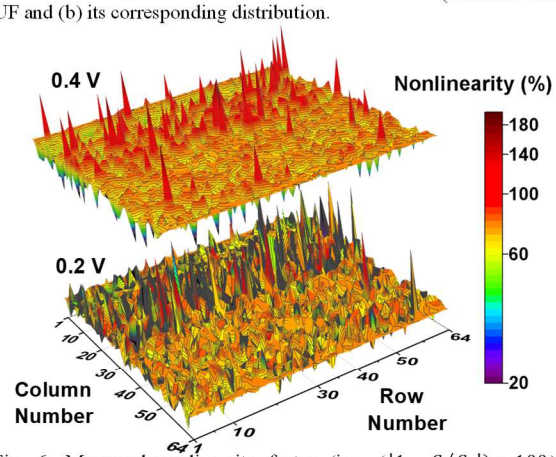


Fig. 6: Measured nonlinearity factor (i.e., $(|1 - G/G_0|) \times 100)$) of the demonstrated PUF for two biasing conditions at room temperature corroborating the presence of strongly nonlinear computing elements in the present design. (G_0 is defined at 0.1 V.)

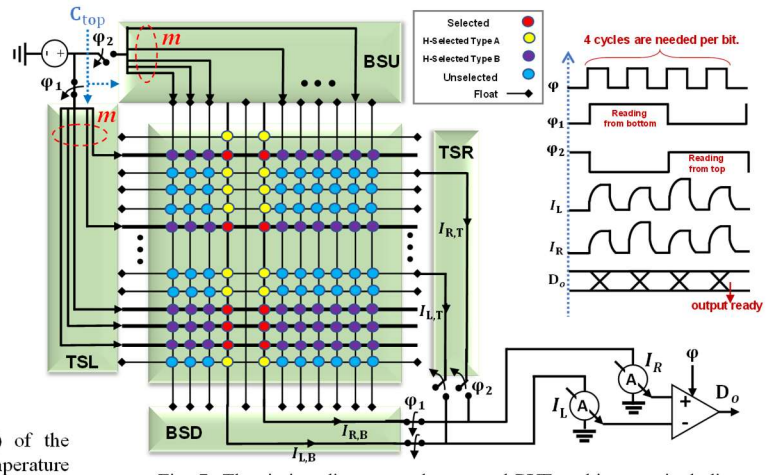


Fig. 7: The timing diagram and proposed PUF architecture including a 64×64 crossbar, selectors (top, bottom, right, left), and a comparator. After 4 cycles, the final response is generated by XORing the previously readout bits from bottom (2 bits) and top (2 bits).

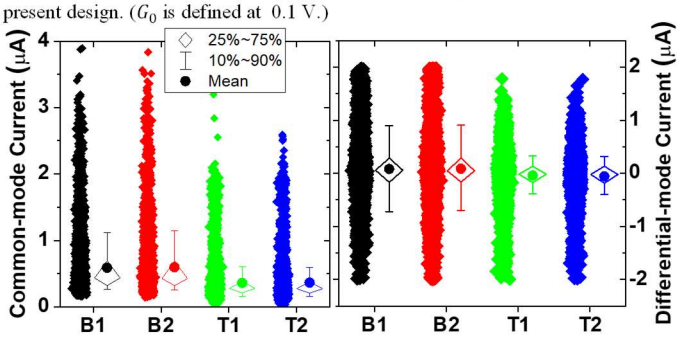


Fig. 8: The measured common-mode and differential current distributions of readout currents from top and bottom electrodes in the first (B1,T1) and second (B2,T2) cycles for 2k challenges @ 0.4V and RT.

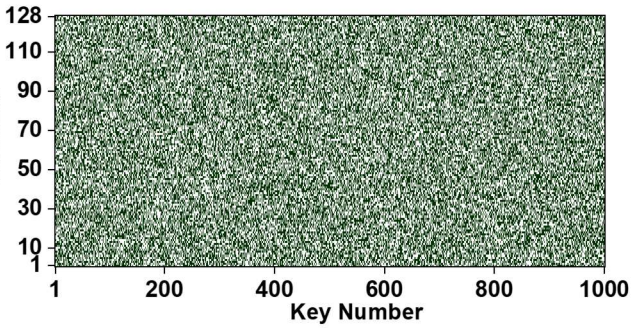


Fig. 9: The speckle pattern for 1k measured 128-bit keys.

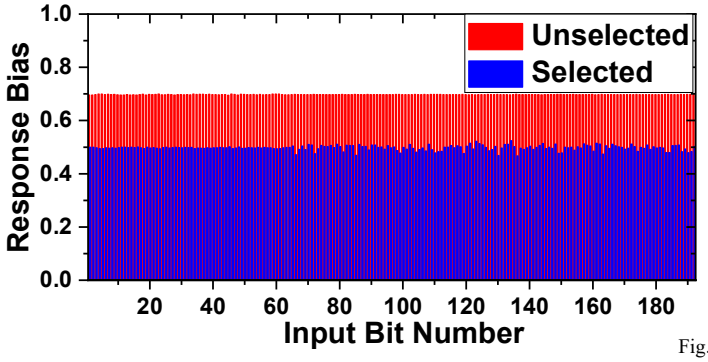


Fig. 10: Measured input-output correlation between 110k CRPs at 0.4V and RT. Unselected data are shifted by +0.2 for clarity.

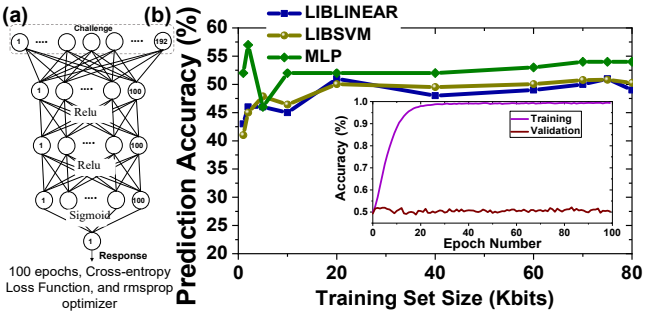
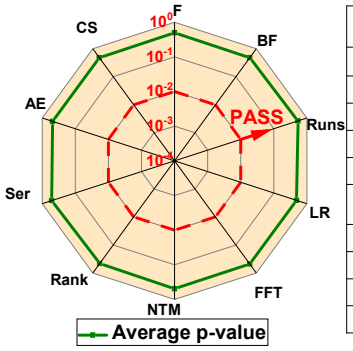


Fig. 11: ML attack results: (a) the structure and hyperparameters of the MLP network and (b) prediction accuracy on unseen data for three models. Inset shows the accuracy during training for the 10k training size.



Test	Result
Frequency (F)	Pass
Block Frequency (BF)	Pass
Runs	Pass
Longest Runs (LR)	Pass
Fast Frequency Trans (FFT)	Pass
Non-overlap Template Matching (NTM)	Pass
Rank	Pass
Serial (S)	Pass
Approximate Entropy (AE)	Pass
Cumulative Sum (CS)	Pass

Fig. 12: Result of NIST randomness test based on 52 2k-length bitstreams measured at 0.4 V and RT.

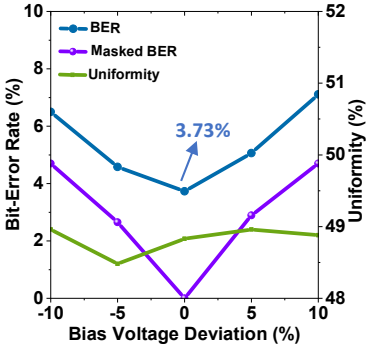


Fig. 13: Measured BER versus applied voltage with reference values measured at 0.4 V and RT.

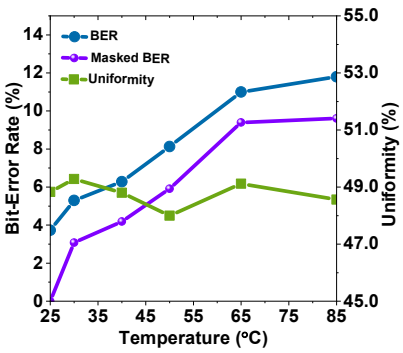


Fig. 14: Measured BER versus temperature with reference values measured at 0.4 V and RT.

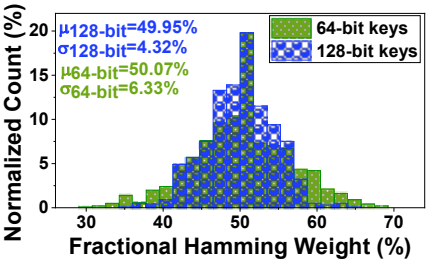


Fig. 14: Uniformity: fractional Hamming weight distribution computed based on 1k randomly 64-bit and 128-bit generated keys at 0.4 V and RT.

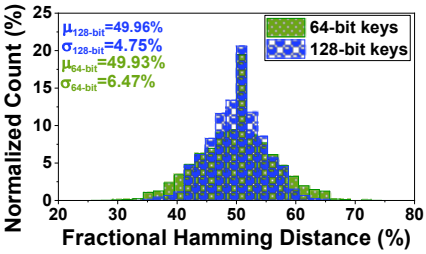


Fig. 15: Diffuseness: fractional Hamming distance distribution computed based on 1k randomly generated keys measured at 0.4 V and RT.

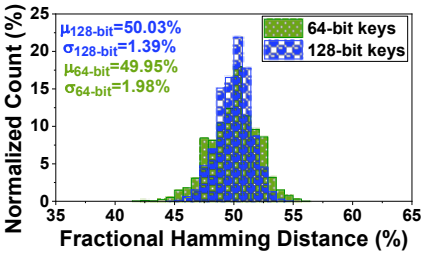


Fig. 16: Uniqueness: fractional Hamming distance distribution computed based on 1k randomly generated keys of two PUF instances measured at 0.4 V and RT.

Table I: Comparison with previous works.

		Nature '18 [1]	VLSI'18 [5]	DAC'19 [7]	TED'16 [6]	ISSCC'15 [2]	ISSCC'15 [3]	VLSI'17 [4]	This work
Technology		Passive ReRAM	Passive ReRAM	55 nm LPe CMOS	Passive ReRAM	22 nm CMOS	65 nm CMOS	FDSOI 28 nm CMOS	Passive ReRAM
PUF Type		Strong*	Strong*	Strong	Strong*	Weak	Weak	Strong*	Strong
Demo Complexity		3D 2×10×10 ~$4 \times 4\ F^2$ 0T1R[xx]	20×20 ~$4 \times 4\ F^2$ 0T1R	TDM 10×10 ~ $5 \times 120\ F^2$ eFlash	12×12 ~$4 \times 4\ F^2$ 0T1R	Crosscouple 9581F^2	SA PUF 12000 F^2	64×64 SRAM ~ 194 F^2	64×64 $4\ F^2$ 0T1R
Pre-configuration		Forming/ Preprogramming	Forming/ Preprogramming	Preprogramming	Forming/ Preprogramming	Intrinsic	Intrinsic	Intrinsic	Intrinsic
Capacity		~700×10^6	~40×10^6	~10^{211}	~1K	-	~3×10^3	~10^{11}	~7×10^{25}
Uncorrected BER @ 85 °C		1.5 %	4.2 %	~5 %	-	~4.5 %	5 %	~12.5 %	11.5%
Power/ Energy	Average Common-mode Current	~53 μA	~35 μA × 2	~30 μA × 42	~ 22 μA	-	-	-	~0.5 μA × 4
	Energy/bit	-	~45 fJ	560 fJ	-	13 fJ[#]	163 fJ	97 fJ	4 fJ
NIST Test		PASS	PASS	PASS	Not Tested	Fail	PASS	Not Tested	PASS
ML Attack/Prediction Error		~50%	~50%	~50%	Predictable[&]	-	-	~ 40%	~50%

* Strong PUF, by definition, though the demo's complexity is small. [&] Based on the very non-ideal conductance map. [xx] A factor of 4 is multiplied to account for the required programming switching circuits. [#] Excluding nonnoticeable power in peripheral blocks