Reliable Strong PUF Enrollment and Operation with Temperature and Voltage Optimization

Kleber Hugo Stangherlin
ECE Department
University of Waterloo
Waterloo, Canada
khstangh@uwaterloo.ca

Manoj Sachdev
ECE Department
University of Waterloo
Waterloo, Canada
msachdev@uwaterloo.ca

***Abstract*—Strong PUFs provide low-cost authentication primitive for resource constrained devices. They use inherent process variation as basis to generate a unique fingerprint, which often lacks the required reliability. Environmental factors, and time varying aging mechanisms can further compromise reliability. In this paper, we investigate the impact of power supply voltage and temperature screens to improve strong PUF performance metrics. Our simulation and measurement results in 65 nm show reliability of 97.4% when operating at 0.6V, with 50.1% uniformity and 46.7% uniqueness. A new double arbitration circuit is proposed to assist in detecting unstable challenges. When compared with Majority Voting, the proposed Double Arbiter circuit achieves comparable reliability performance of 99.6% with only half the evaluations.**

***Index Terms*—puf, arbiter, reliability, low-voltage**

I. Introduction

Internet of Things (IoT) is enabling networking of billions of devices world-wide. With limited computing resources, such devices often lack mechanisms for secure authentication. The traditional solution adopted by manufacturers is to use secret IDs programmed during test. The IDs are typically implemented with one-time programmable fuses, or non-volatile memories, making the device susceptible to external tampering attacks and counterfeiting. The so-called Physical Unclonable Functions (PUFs) are a class of low area/energy circuits that harvest intra/inter-die process variations to generate a device fingerprint. A subclass of PUF circuits known as strong PUF is the focus of this paper. Strong PUFs generate chip-unique responses to externally provided challenges. They require an enrollment phase which is typically run during test, or at a later stage in the product assembly line. The enrollment consists of applying a randomly selected subset of challenges and reading the associated responses, storing the challenge-response pairs (CRPs) in a secure database. After deployment, the PUF is inquired with a subset of the stored challenges, and the responses must match the stored values (within a system defined error threshold).

Ideally strong PUFs should not leak any information about its internal characteristics. In real implementations however, it's been widely shown that PUF responses carry significant information about its internal entropy source. Using a small subset of challenge-response pairs, machine learning attacks are able to accurately predict the PUF output [11]. Modeling resistant PUF architectures are an active field of research, where numerous tactics are explored. In particular, researchers observed that machine learning resistance can be significantly improved by using several smaller PUFs to generate intermediate responses, which are then used as input challenge to a second layer PUF that produces the final response [14]. Such architectures are often limited by reliability performance of every individual PUF in it. When PUF decisions are made, they insert quantization errors that propagate in a compound fashion to the next level PUF. This reasoning can be expanded to other architectures such as lightweight PUF and XOR arbiter PUF with respect to the number of forward loops and the number of parallel arbiter PUF instances XORed [10], [12].

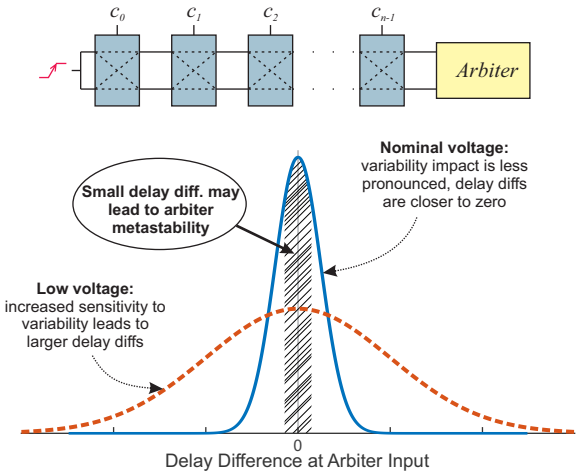

Fig. 1. Low voltage exploration to improve arbiter PUF reliability. The increased sensitivity to process variations flattens the delay difference distribution when transistors operate at lower voltages.

This work is driven by our desire to improve the overall reliability of delay based strong PUFs, in particular, the arbiter PUF. We investigate how reliability can be improved by means of two different approaches: i) low-voltage operation; and ii) challenge selection during enrollment. MOSFET drain current has increased sensitivity to process variations when supply voltage approaches the threshold voltage [19]. Fig. 1 plots a flattened delay difference distribution for delay based strong PUFs operating at low/nominal voltages. With early

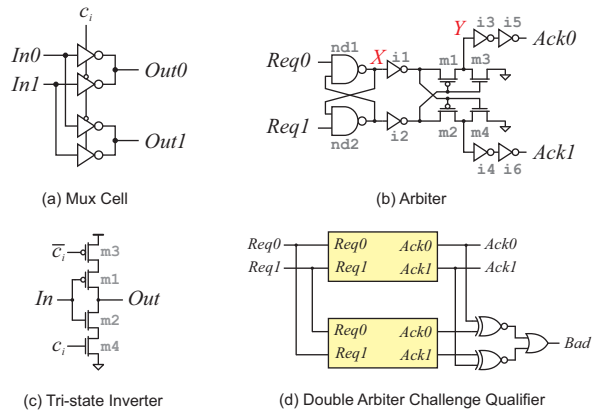

Fig. 2. Circuit level implementation of the multiple components used in this work. The chip was fabricated in 65nm technology.

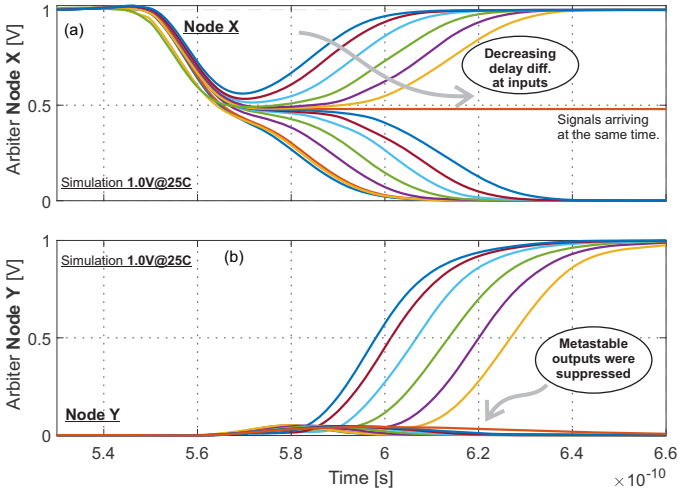

Fig. 3. Arbiter outputs at (a) node X, and (b) node Y. The glitch suppression circuit avoids propagation of metastability.

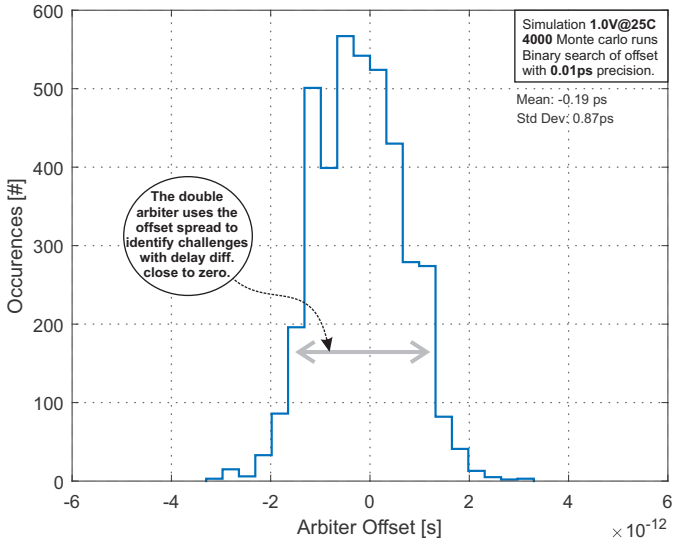

Fig. 4. Histogram for a single arbiter offset simulated with 4000 Monte Carlo runs, and a binary search with resolution of 0.01 ps (mean of -0.19 ps and standard deviation of 0.87 ps).

selection of challenges during enrollment, we seek to exclude unstable challenges from the pool used for authentication. The technique is referred to as challenge qualification. The main idea is to add extra circuitry for early detection of unstable challenges.

II. Previous Work

The topic of low-voltage PUF design was initially discussed in [13], where the author simulates ring oscillator PUFs from 0.2V to 1V. In [6], [7] a 64-bit arbiter PUF is simulated and validated in 45nm CMOS process. Authors reported that at low-voltage many challenge-response pairs (CPRs) were noisy and unreliable, and were excluded. In [4], the authors advocated better temperature reliability by operating delay based PUFs in a supply voltage known as zero temperature coefficient. In [9] and [5], the high sensitivity of leakage current to process variations is investigated; the first used an analog sense amplifier as quantizer for different currents, while the second implemented two arrays of half-latches, and uses an arbiter to evaluate which half-latch toggles first. In [3] a sub-threshold current array is implemented, and two voltages produced by nominally identical arrays structures are compared. Among the issues faced by this architecture are the small voltage from which the decision is derived, and also the overall low-reliability when compared to typical arbiter PUF values. In [2], a pool of oscillators runs at low-voltages, but the authors add extra circuitry to compensate for variability induced effects.

With respect to challenge selection on enrollment, the work in [15] uses a machine learning model to compute predicted delay differences that are likely to be unreliable for a given PUF (excluding those challenges from the pool used for authentication). In [17], [18], the authors use multiple flip-flops to design arbiters that can detect unstable states (encoded by multiple valued outputs). The work in [16] proposed a reliability checker circuit that uses signals from various stages to produce an internal bit for sanity check.

III. Circuit Level Implementation

To investigate the impact of low-voltage operation on the reliability of delay based PUFs, we designed a full-custom, 64-bit arbiter PUF in a 65nm technology. The relevant cells are shown in Fig. 2. Our implemented mux cell uses tri-state inverters and requires an inverted challenge input (which is derived from the register that stores the challenge).

The design of the arbiter circuit (Fig. 2 (b)) is crucial to ensure reliable performance of the PUF. We carefully laid out arbiter circuit ensuring symmetrical implementation. The cross-coupled NAND gates can generate metastable outputs when the delay difference of the signals at the input is too small (see Fig. 3 (a)). To enhance the reliability of our implemented arbiter, we added a glitch suppression circuit after the cross coupled NANDs, shown in Fig. 2 (b). The

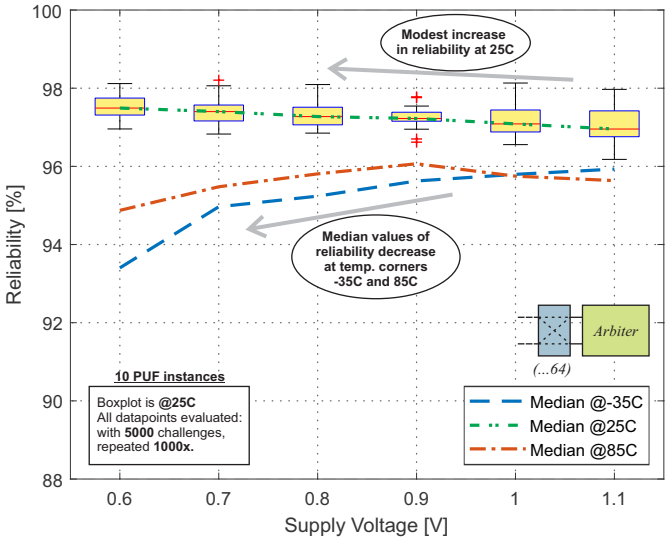

Fig. 5. Reliability of arbiter PUF in different supply voltages and temperature corners. The boxplot shows reliability at 25C, while the dashed series plot the median reliability at -35C, and 85C (responses obtained at 25C are used as reference).

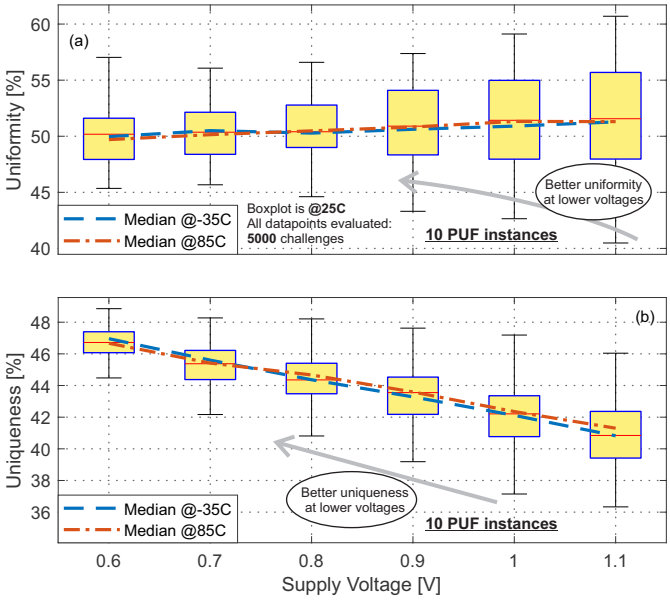

Fig. 6. Uniformity (a) and uniqueness (b) of an arbiter PUF in different supply voltages. Low voltage operation suggests a trend of improvement in uniformity and uniqueness.

circuit consists of four MOSFETs, from M1 to M4, that keep the arbiter output low until one of the NAND outputs differs from the other by more than one V_{th}, as shown in Fig. 3 (b). Although not represented on the schematic diagram, both request signals arrive at the same topological input on its corresponding NAND gate, to ensure identical loads on both signal paths.

To improve reliability of the arbiter PUF, we added extra circuitry to enable the early detection of unstable challenges. We propose an enrollment process where "bad" challenges, are removed from the pool and not used for chip authentication. To qualify *good* and *bad* challenges, we introduce a circuit topology called Double Arbiter (Fig. 2 (d)). The newly added arbiter component does not change the circuit critical path, its decision is only used for error detection. The Double Arbiter explores the offset spread of two arbiter instances to detect challenges that generate small delay differences. For a correct decision, the two arbiters should always present complementary outputs. Fig. 4 (a) plots the distribution of a single arbiter offset, obtained with 4000 Monte Carlo simulation runs ($\mu = -0.19$ ps and $\sigma = 0.87$ ps). In order to demonstrate the Double Arbiter's effectiveness to remove unreliable responses, we performed Monte Carlo simulations with 4,000 instances and 5,000 challenges each. Each instance had a unique Double Arbiter circuit, which was able to was able detect approximately 2.50% responses as unreliable.

IV. Testchip and Measurement Results

We fabricated a 65nm chip with 10 instances of a 64-stages arbiter PUF using single arbiter circuit, and 10 other instances of 64-stages arbiter PUF using Double Arbiter circuit. Furthermore, we also added one arbiter PUF instance without the arbiter circuit, and instead, two level converters that connect the mux path outputs to IO pads (details are discussed in section V). The responses of each instance was measured across different temperature corners to evaluate uniformity, uniqueness, and reliability. Measurements were performed with supply voltages from 0.6V, up to 1.1V (steps of 0.1V). All PUF instances were evaluated with 5000 challenges, where every one of those challenges was applied 1000 times – at aforementioned supply voltages and temperature corners of -35C, 25C, and 80C. To ensure correctness of results, chip reset was applied after every challenge (clearing the state of all sequential elements). The automated measurement software iterates over all unique challenges before starting a re-evaluation loop, i.e., repeating the challenges.

The results in Fig. 5 show the boxplot for 32-bit response measurements across supply voltages at 25C of 10 PUF instances. The Y-axis represent the PUF reliability, it evaluates the response consistency of each of 5,000 challenges which are repeated 1,000 times. Dashed lines refer median reliability values at temperature corners, and outliers represent a single PUF instance. For example, at 0.8V the measured reliability median is 97.4% for 25C, 95.4% for -35C, and 95.9% for 80C. We observe a trend of modest improvement for median values of reliability at lower supply voltages and at 25C. Possible reasons for the limited gains when operating at low voltage relate to increased noise components and are discussed in section V. At the temperature corners, the arbiter PUF reliability values suggest a downturn trend, which might be explained as an effect of the transition from strong to weak inversion. In this region, the current shifts towards an increasingly exponential relationship between device current and temperature, given by $I \propto \exp(V_{GS}/V_T)$, with $V_T = kT/q$. The term k denotes the Boltzmann constant, q the electron charge, and T the absolute

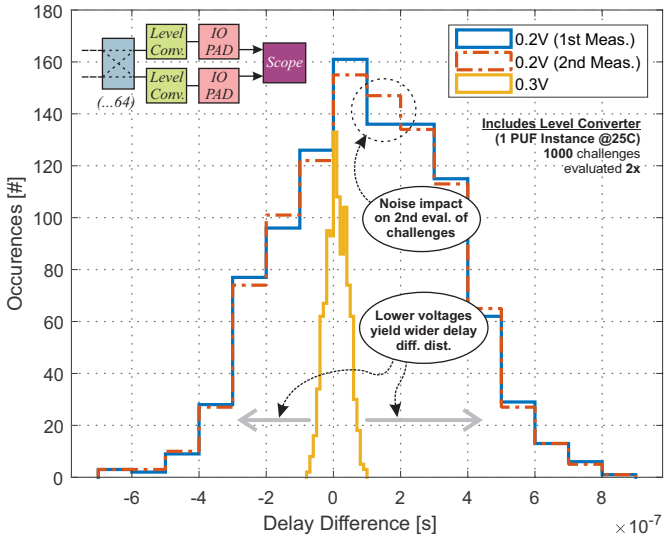

Fig. 7. Histogram of measured delay differences using dedicated outputs to IO PADs. Lower voltages yield wider delay difference distributions. The impact of noise is seen at 0.2V when the same set of challenges is measured a second time.

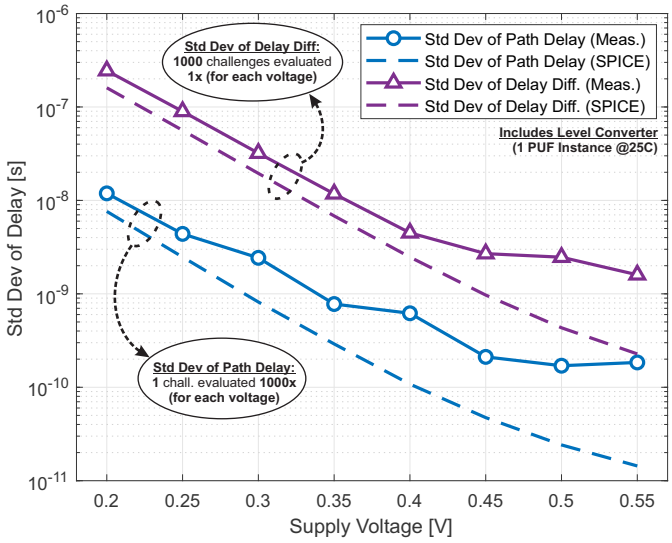

Fig. 8. Oscilloscope measurements and simulation results for: standard deviation of delay differences for 1000 challenges; and standard deviation of path delay for 1000 measurement with the same challenge. Both curves show the same exponential trend, confirmed by SPICE simulation results.

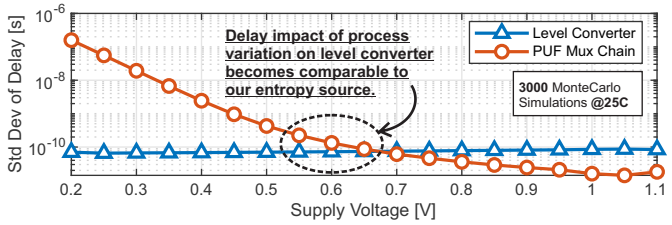

Fig. 9. Delay impact of process variability on level converter, and on a chain of 64 mux cells. The standard deviation of delay for a single level converter overcomes the standard deviation delay of the entire chain of mux cell for voltages near and above 0.6V.

temperature.

Fig. 6 shows the uniformity (a), and uniqueness (b) performance of the PUF instances. Uniformity is calculated with the intra-die hamming-weight of all PUF responses, across instances – outliers in the uniformity plot refer to individual PUF instances. The uniqueness is the inter-die hamming distance, and it is calculated with the method proposed in [8]. The data suggests a trend of improvement in uniformity when operating at low voltages. The uniformity median remains near 50%, but the distribution spread reduces showing smaller values for standard deviation. Uniqueness also shows improvements in both the median values and standard deviation when operating at lower voltages. The difference between the uniqueness and uniformity plots may be explained by the fact that uniformity sets a ceiling for uniqueness. In other words, if there is a bias in the responses of a strong PUF, it will necessarily be more difficult to differentiate between two chip instances. Both uniformity and uniqueness did not present significant variations across temperature corners of -35C, and 85C.

V. Noise Analysis

The arbiter PUF captures process induced variations through the cumulative delay difference of two paths designed to be identical. To further understand and characterize the arbiter PUF behavior at low-voltages, we measured the delay difference distribution between arbiter PUF paths with an oscilloscope. We implemented a PUF instance without an arbiter; the output of delay lines were terminated on level converters. Each level converter output is connected to an IO pad, allowing an external oscilloscope to measure the delay difference between the two racing signals. Fig. 7 shows the measured delay difference distribution at 0.2V, and 0.3V. The measurement at 0.2V was repeated again. The low-voltage operation (at 0.2V) flattens the distribution as expected. Nevertheless, we can see the impact of noise in the histogram plot, where the bin counter for the same set of challenges shows deviations when evaluated a second time.

Fig. 8 provides a summary of delay difference measurements together with SPICE simulation results covering supply voltage from 0.55V down to 0.2V. The standard deviation of delay difference is evaluated for 1000 challenges and is increased exponentially with supply voltage reduction. Fig. 8 also plots the standard deviation of path delay, where the delay of a single mux chain is measured 1000 times, but with the same challenge. Thus, from the data presented in Fig. 8, one may observe that an arbiter PUF, when operating at low-voltage, presents significantly larger delay difference values (being more sensitive to manufacture variations, as expected). Nevertheless, the standard deviation of path delay (for a single challenge) increases nearly at the same rate, suggesting that the noise components are also rising when it operates at low voltages.

To validate this result, we performed circuit level simulation of mismatch and noise. Current state of the art commercial

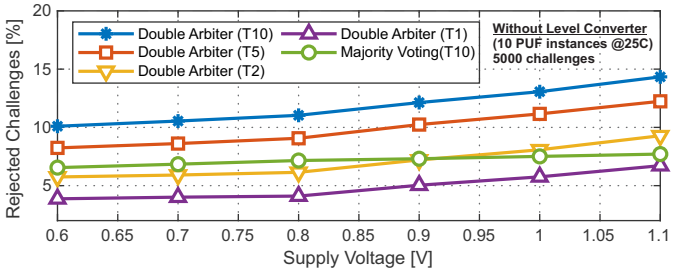

Fig. 10. Percentage of rejected challenges during enrollment by different challenge qualification methods. The label `TXX` refer to the number of repeated evaluations during enrollment. The *Majority Voting* data represents no extra circuit for challenge qualification, but repeated evaluations on a single arbiter are used.

simulators are not able to evaluate transient noise and mismatch in the same simulation run. To overcome this limitation, we adopted a characterization based methodology where the delay and noise of a single mux cell are simulated using two different runs, one for parameter mismatch, and another for transient noise analysis. The electrical simulation tool used is HSPICE with BSIM 4.5 models. The used MOSFET model was obtained from the foundry, it calculates the noise components with options `fnoimod=1` and `tnoimod=0`. These options select a unified physical model for flicker noise, and a charge based model for channel thermal noise, respectively. Once the mean and standard deviation values are characterized for the mismatch and noise at every voltage, a custom script samples the mismatch and noise delay contribution for each mux cell in the path (assuming a normal distribution). The computed path delay and delay difference are shown by the dashed lines in Fig. 8. The simulation results track the measurement results but do not exactly match due to model, and measurement limitations. The data suggests that the increased sensitivity to manufacture variability achieved in lower voltages, might not translate into substantial reliability gains due to the increasing contribution of noise components.

As mentioned before, for measurement results shown in Figs. 7 and 8, delay lines are terminated on a level converter. The level converters operating at nominal supply add delay uncertainties in measurements. Fig. 9 illustrates standard deviations of the level converter and delay chain as a function of supply voltage. At supply voltage below 0.6 V, the level converter's variability contribution is relatively small compared to that of the delay chain. On the other hand, as supply voltage of the delay chain is increased, it results in corresponding decrease in its standard deviation while that of the level converter is flat. Therefore the delay data captured by the oscilloscope for voltages higher than 0.6 V in Fig. 8 may not reflect the standard deviation of the delay chain, and is not presented.

From a physical/device perspective, the strong impact of noise in the arbiter PUF operating at low-voltages can be explained using the work reported in [1]. Authors carried out phase noise analysis in ring oscillators, and deriving expressions to calculate jitter and phase noise in strong inversion.

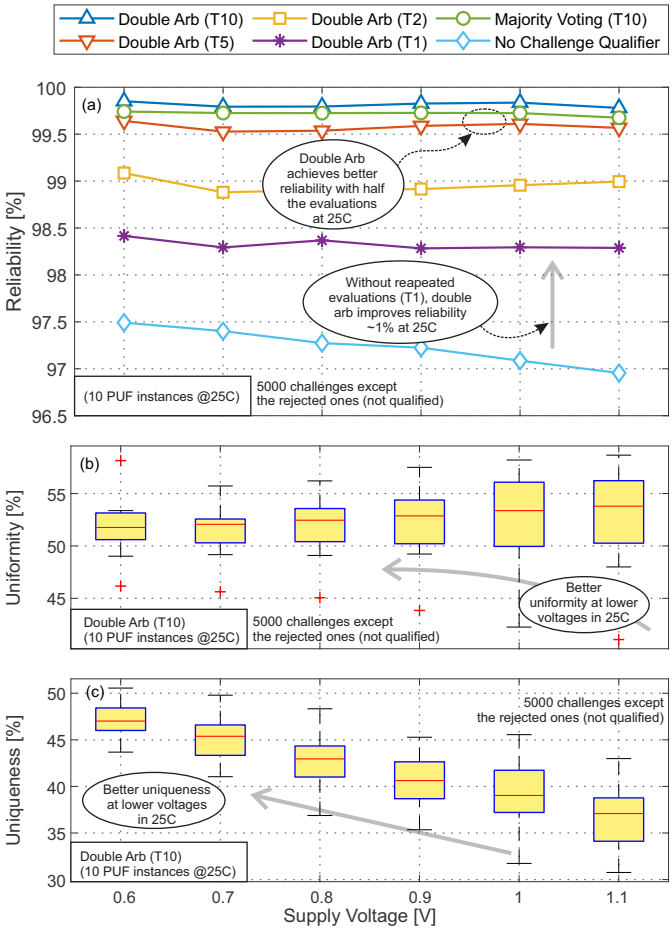

Fig. 11. At 25C, (a) reliability, (b) uniformity and (c) uniqueness performance metrics for different challenge qualification processes. Without repeated evaluations, the Double Arbiter challenge qualifier improves the reliability near 1%. Nevertheless, the rejection of certain challenges caused an impact in the uniformity metric. Uniformity and uniqueness are plotted for T10 (ten repeated Double Arbiter evaluations during enrollment).

In one of the intermediate results, they show that the mean-square value of the integrated voltage noise, $\langle v_n^2 \rangle$, is given by

$$\langle v_n^2 \rangle = \frac{S_{i_n}}{2C^2} t_d \propto t_d, \tag{1}$$

where S_{i_n} is the spectral density of the noise current, C is the loading stage capacitance, and t_d the interval over which the noise is integrated (propagation delay of the driving inverter). Eq. 1 depicts that noise is proportional to the inverter delay, and as the supply voltage is reduced the inverter delay is increased which results in higher noise voltage. If the supply voltage is further reduced to weak inversion, the noise voltage related delay variation becomes comparable to the delay mismatch caused by the process variation which results in poor PUF reliability.

VI. Challenge Selection with Qualifiers

We also investigated other techniques to improve PUF reliability by an improved arbiter circuit that is capable of detecting

unreliable CRPs during the enrollment phase. Needless to say, these unreliable responses are excluded from authentication pool. If the evaluations do not yield all the same response, the challenge is rejected; this approach does not require extra circuitry, and we refer to it as *Majority Voting*. We propose a new circuit to support the challenge qualification process during enrollment, the *Double Arbiter Qualifier*, described in section III and shown in Fig. 2 (d), provides an additional output that identifies unstable (bad) challenges, increasing the number of unstable challenges detected during enrollment, and consequently, removed from the authentication pool to improve reliability.

The measurements in Fig. 10 show the percentage of challenges that have failed the qualification test during enrollment (at 25C) for different challenge qualifiers, across several voltages. Data is labeled in the format `TXX`, where the `XX` corresponds to the number of repeated evaluations a challenge has been submitted during enrollment. The Double Arbiter Qualifier rejects more challenges than Majority Voting in any supply voltage. In particular, the Double Arbiter Qualifier data suggests a trend of increasingly more challenges being rejected at higher supply voltages; e.g. at 1.1V, a single evaluation with Double Arbiter rejects nearly the same number of challenges as 10 evaluations of Majority Voting.

Fig. 11 (a) plots the reliability measurements for Double Arbiter Qualifier and Majority Voting. The data labeled as *No Challenge Qualifier* refers to the usage of a single arbiter, without any repeated evaluations (no challenges were excluded from the authentication pool). If no reevaluations are performed (`T1`), the Double Arbiter Qualifier achieves a 1% improvement in reliability compared to the No Qualifier option; when the Double Arbiter is evaluated 5 times (`T5`), it achieves reliability levels comparable to Majority Voting with 10 evaluations (`T10`). The uniformity and uniqueness distributions for Double Arbiter Qualifier with 10 evaluations (`T10`) are plotted in Fig. 11 (b) and (c); both metrics show a trend of improvement towards lower voltages.

VII. Conclusion

Delay-based strong PUFs are widely used as building blocks for composite PUF architectures, which are known to be limited by reliability constraints. We presented circuit simulation and measurement results of two techniques to improve reliability of arbiter PUFs, low-voltage operation and challenge selection during enrollment.

Running an arbiter PUF at lower supply voltages, to benefit from increased sensitivity to process variations, showed modest improvements in reliability at 25C. Noise components are analyzed and suggested as possible reason. Other performance metrics such as uniformity and uniqueness show a trend of improvement at lower supplies, with smaller standard deviation values and medians that approach the ideal value of 50%. At 0.6 V, the measured median reliability value is 97.4%, with 50.1% uniformity and 46.7% uniqueness.

The proposed Double Arbiter challenge qualifier showed promising results for early detection of unstable challenges during the enrollment phase. The Double Arbiter circuit requires only 5 repeated evaluations to achieve comparable reliability performance to Majority Voting with 10 evaluations which translates to faster enrollment time. At 0.6 V, the Double Arbiter obtained reliability of 99.85% when 10 repeated evaluations were used (`T10`). In addition to that, even when repeated evaluations are not used (`T1` and *No Challenge Qualifier*) the Double Arbiter improves reliability in nearly 1% for all supply voltages evaluated.

References

[1] AA Abidi. Phase noise and jitter in cmos ring oscillators. *IEEE JSSC*, 41(8):1803–1816, 2006.

[2] Y Cao, L Zhang, C-H Chang, and S Chen. A low-power hybrid ro puf with improved thermal stability for lightweight applications. *IEEE Trans on Comp-aided design of integrated circuits and systems*, 34(7):1143–1147, 2015.

[3] M Kalyanaraman and M Orshansky. Novel strong puf based on nonlinearity of mosfet subthreshold operation. In *IEEE HOST*, pages 13–18, 2013.

[4] R Kumar, HK Chandrikakutty, and S Kundu. On improving reliability of delay based physically unclonable functions under temperature variations. In *IEEE HOST*, pages 142–147, 2011.

[5] J Lee, D Lee, Y Lee, and Y Lee. A 445f 2 leakage-based physically unclonable function with lossless stabilization through remapping for iot security. In *IEEE ISSCC*, pages 132–134, 2018.

[6] L Lin, D Holcomb, DK Krishnappa, P Shabadi, and W Burleson. Low-power sub-threshold design of secure physical unclonable functions. In *IEEE Int Symp on Low power Elec and design*, pages 43–48, 2010.

[7] L Lin, S Srivathsa, DK Krishnappa, P Shabadi, and W Burleson. Design and validation of arbiter-based pufs for sub-45-nm low-power security applications. *IEEE Trans on Inf Forensics and Security*, 7(4):1394–1403, 2012.

[8] A Maiti, V Gunreddy, and P Schaumont. A systematic method to evaluate and compare the performance of physical unclonable functions. In *Embedded systems design with FPGAs*, pages 245–267. Springer, 2013.

[9] M Majzoobi, G Ghiaasi, F Koushanfar, and SR Nassif. Ultra-low power current-based puf. In *IEEE ISCAS*, pages 2071–2074, 2011.

[10] M Majzoobi, F Koushanfar, and M Potkonjak. Lightweight secure pufs. In *IEEE/ACM Int Conference on Comp-Aided Design*, pages 670–673, 2008.

[11] U Rührmair, J Sölter, 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 on Inf forensics and security*, 8(11):1876–1891, 2013.

[12] GE Suh and S Devadas. Physical unclonable functions for device authentication and secret key generation. In *ACM/IEEE Design Automation Conf*, pages 9–14, 2007.

[13] V Vivekraja and L Nazhandali. Circuit-level techniques for reliable physically uncloneable functions. In *IEEE HOST*, pages 30–35, 2009.

[14] Z Wu, HD Patel, M Sachdev, and MV Tripunitara. Strengthening pufs using composition. In *ICCAD*, pages 1–8, 2019.

[15] X Xu, W Burleson, and DE Holcomb. Using statistical models to improve the reliability of delay-based pufs. In *IEEE ISVLSI*, pages 547–552, 2016.

[16] J Ye, Y Hu, and X Li. Vpuf: Voter based physical unclonable function with high reliability and modeling attack resistance. In *IEEE IOLTS*, pages 74–79, 2017.

[17] SS Zalivaka, AA Ivaniuk, and C-H Chang. Reliable and modeling attack resistant authentication of arbiter puf in fpga implementation with trinary quadruple response. *IEEE Trans on Inf Forensics and Security*, 14(4):1109–1123, 2018.

[18] SS Zalivaka, AV Puchkov, VP Klybik, AA Ivaniuk, and C-H Chang. Multi-valued arbiters for quality enhancement of puf responses on fpga implementation. In *IEEE ASP-DAC*, pages 533–538, 2016.

[19] B Zhai, S Hanson, D Blaauw, and D Sylvester. Analysis and mitigation of variability in subthreshold design. In *Int Symp on Low Power Elec and Design*, pages 20–25, 2005.