14.2 A Physically Unclonable Function with BER <10^{-8} for Robust Chip Authentication Using Oscillator Collapse in 40nm CMOS

Kaiyuan Yang, Qing Dong, David Blaauw, Dennis Sylvester

University of Michigan, Ann Arbor, MI

Security is a key concern in today's mobile devices and a number of hardware implementations of security primitives have been proposed, including true random number generators, differential power attack avoidance, and chip-ID generators [1-4]. Recently, physically unclonable functions (PUFs) were proposed as a secure method for chip authentication in unsecure environments [5-7]. A PUF is a function that maps an input code (*"challenge"*) to an output code (*"response"*) in a manner that is unique for every chip. PUFs are increasingly used for IC authentication to offer protection against identity theft, cloning, and counterfeit components [2-4].

Conventional authentication methods using secret keys, digital signatures, and encryption have high computational expense and are vulnerable to tampering [7]. PUFs address this by generating their secret, unique challenge/response pairs (CRPs) using process variation, thus eliminating the expense of programming a secret key and the risk of compromising the stored data [5-7]. The PUF authentication protocol consists of two phases (Fig. 14.2.1): Enrollment occurs in a secure environment where a known chip (#2435 in our example) is interrogated with a large number of random challenges and the resulting CRPs are stored. Later, during verification in an untrusted environment, a chip claiming to be #2435 is interrogated with a small subset of the stored challenges and the chip's response is verified. If chip #2435 was swapped with a fraudulent one, its response would not match and authentication would fail. It is critical that each verification attempt uses a new subset of stored challenges. Hence, illicit observation of previously used CRPs is harmless since each CRP is used only once. This is a key differentiator from chip IDs, which are considered PUFs with only a single CRP and as a result are much less secure than high-CRP PUFs. In such a case, any observation in an untrusted environment of chip #2435's chip ID allows a malicious chip to impersonate chip #2435 by storing and reporting the observed chip ID.

Silicon PUFs have been proposed based on variations in gate and interconnect delay [2], ring oscillator frequency [3], and inverter maximum gain point [4]. Due to the sensitivity of device parameters to operating conditions (temperature, voltage, wearout), the PUF output may change between interrogations, manifesting as BER and possible authentication failures. Stabilizing approaches to address this includes majority voting, burn-in, ECC, and masking [4-7]. However, these all require additional testing and calibration efforts for each chip. This paper presents a PUF based on multi-edge oscillation collapse in a ring-oscillator (RO). The design translates physical variation to a digital output by injecting 2 edges into an even-stage RO (Fig. 14.2.1). The two injected edges travel entirely different paths and hence accumulate delay cell mismatch, causing one edge to overtake the other and collapsing the oscillation. Depending on which path is faster, the output settles to either 0 or 1. Noise averaging along the paths of the two edges aids stability, which is further enhanced by simple dynamic thresholding based on cycles to collapse. The PUF is validated in 40nm CMOS, featuring: (1) BER remains <10^{-8} across -25 to 125°C and 0.7 to 1.2V; (2) average inter-chip hamming distance is 0.5007; (3) the all-digital design occupies 845μm^2 and requires no calibration.

For the proposed structure to respond to a large set of challenges, each stage in the RO is selected from 8 identical delay cells (Fig. 14.2.2). By selecting from 8 cells in each stage (3b of the challenge), rather than 2 cells (1b), the RO length is shortened, making the difference between the 2 paths larger (less averaging) and the output value more robust. To further increase response stability, we add a footer to each delay cell and bias it in near-threshold to ensure that the variation of the footer NMOS dominates the total delay. A CTAT is used to generate the bias voltage on-chip, which performs a first-order temperature compensation of the footer current to reduce the PUF temperature sensitivity. The RO and control logic is reset during the positive phase of the clock (CLK) with the PUF output generated during the negative phase.

Bit stability is an essential property for a reliable PUF. Measurement reveals that the number of cycles to collapse follows a lognormal distribution (Fig. 14.2.3). Also, responses that collapse slower have much larger average BER. This is expected since a slow collapse results from the two paths having nearly matched delay, making it more likely that the response is determined by noise, not process variation. Therefore, we implement a simple yet effective dynamic thresholding technique based on the number of cycles to collapse. A 9b counter in the PUF control logic (Fig. 14.2.2) records the number of cycles to collapse. A counter bit is then used to determine if the count exceeds the set threshold (output OVERFLOW), in which case the PUF output is discarded.

Figure 14.2.3 shows the overall authentication protocol. During enrollment, a chip is interrogated; only CRPs with a collapse cycle smaller than the threshold are recorded. During verification, the chip is interrogated with a stored challenge. If the collapse count is larger than the threshold, the CRP is skipped and the next stored CRP is used. If the collapse count is smaller than threshold, the response is checked against the stored golden CRP. The process is repeated until either a response does not match and the authentication is rejected, or a sufficient number of CRPs match and the authentication is approved.

Dynamic thresholding dramatically reduces the measured BER in worst-case operating conditions (-25°C, 0.7V V_{DD}) from 9% to 0.002% with a threshold value of 32, or down to 0 with a threshold of 16. Smaller threshold values give lower BER at the cost of more discarded CRPs and more evaluations during enrollment and verification. However, even at a threshold of 16, the total discarded CRPs is an acceptable 34%, of which 80% is discarded during the enrollment phase. Also, PUF throughput is not necessarily reduced with a smaller threshold because the global clock can run faster due to faster collapse. Note also that unlike other stabilization methods [4, 7], the proposed dynamic thresholding does not require any extra testing effort before each authentication.

Inter-chip and intra-chip Hamming distances (HD) are also important metrics for a PUF, quantifying spatial uniqueness and temporal stability, respectively. Inter-chip HDs are measured across 20 dies with 1000 challenges (Fig. 14.2.4). Outputs are grouped into 100b keys; keys from different dies but the same challenge sets are compared in all possible pairs. The average normalized HD is 0.5007 (σ=0.0627), very close to the ideal 0.5 value. Intra-chip Hamming distance is measured using 5000 challenges with each challenge evaluated 1000 times. The average intra-chip HD is 0.0101 (σ=0.0635) at nominal conditions without dynamic thresholding. After applying dynamic thresholding with a threshold value 16, the BER and intra-die HD remain 0. Even without thresholding, a 50× mean value separation between inter and intra-die HD is sufficient to ensure the PUF uniqueness with a failure probability of 1.2×10^{-30} (assume 10^9 chips, 256b responses and a tolerance of 15 error bits) [5]. In this case, the false alarm rate (FAR) and false detection rate (FDR) are 1.16×0^{-39} and 2×10^{-73}, respectively. Moreover, average HD between responses of different challenges on same die is 0.4722 (σ=0.0496), showing good uniqueness among CRPs (Fig. 14.2.5).

A PUF will experience significant operating condition variations in a hosting device. For secure authentication, bit flipping due to varying environments should be minimized. With the current-starved delay cells, CTAT bias voltage, and dynamic thresholding, the PUF can maintain a <10^{-8} BER across –25 to 125°C and 0.7 to 1.2V ranges with the golden CRPs generated at nominal 25°C and 0.9V (Fig. 14.2.5). Fig. 14.2.6 summarizes measurement results and compares to prior work. In 40nm CMOS, the PUF generates response bits at ~1.6Mb/s while consuming 28.4μW and 845μm^2 with excellent BER across wide operating conditions (die photo in Fig. 14.2.7).

Acknowledgements:
The authors thank the TSMC University Shuttle Program for chip fabrication and NSF for research support.

References:

[1] Y. Su, *et a*l., "A 1.6pJ/bit 96% Stable Chip-ID Generating Circuit Using Process Variations," *ISSCC Dig. Tech. Papers*, pp. 406-407, Feb. 2007.

[2] N. Liu, *et al.*, "OxID: On-chip One-Time Random ID Generation Using Oxide Breakdown," *IEEE Symp. VLSI Circuits*, pp. 231-232, 2010.

[3] K. Yang, *et al.*, "A 23Mb/s 23pJ/b Fully Synthesized True-Random-Number Generator in 28nm and 65nm CMOS," *ISSCC Dig. Tech. Papers*, pp. 280-281, Feb. 2014.

[4] S. Mathew, *et al.*, "A 0.19pJ/b PVT-Variation-Tolerant Hybrid Physically Unclonable Function Circuit for 100% Stable Secure Key Generation in 22nm CMOS," *ISSCC Dig. Tech. Papers*, pp. 278-279, Feb. 2014.

[5] J. W. Lee, *et al.*, "A Technique to Build a Secret Key in ICs for ID and Authentication Applications," *IEEE Symp. VLSI Circuits*, pp. 176-179, 2004.

[6] G. E. Suh and S. Devadas, "Physical Unclonable Functions for Device Authentication and Secret Key Generation," *IEEE/ACM Design Automation Conf.*, pp. 9–14, 2007.

[7] S. Stanzione, *et al.*, "CMOS Silicon Physical Unclonable Functions Based on Intrinsic Process Variability," *IEEE J. Solid-State Circuits*, vol.46, no.6, pp. 1456-1463, 2011.

 978-1-4799-6224-2/15/$31.00

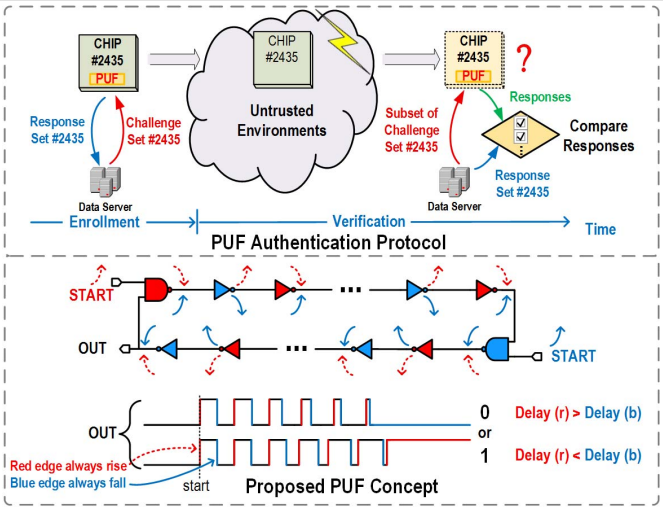

Figure 14.2.1: Basic PUF authentication protocol for resource-constrained devices and generating robust and large number of CRPs by using oscillation collapse in an even stage RO with noise averaged.

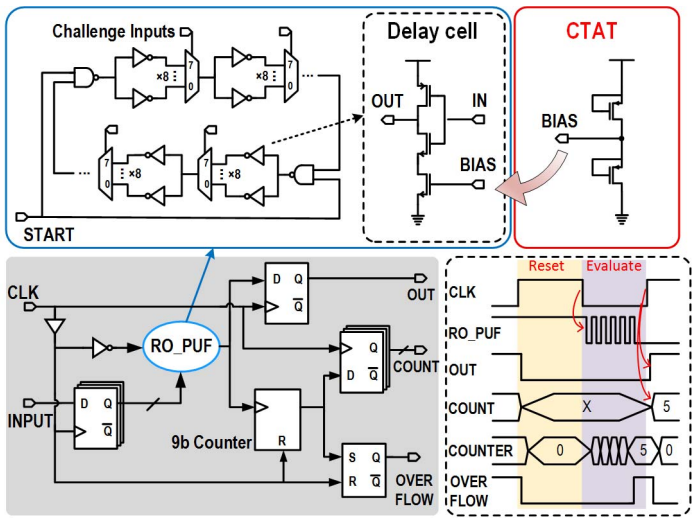

Figure 14.2.2: PUF block diagram with circuit implementations and operation waveforms.

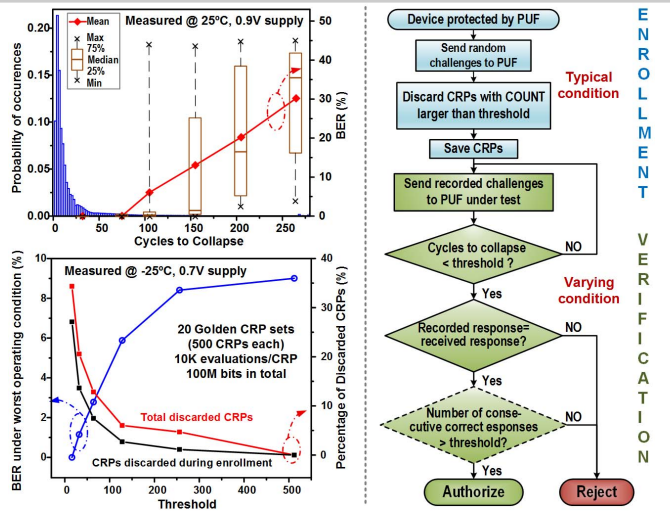

Figure 14.2.3: Measured distribution of cycles to collapse with its impact on BER, impact of threshold value on BER and discarded CRPs and a basic protocol employing the dynamic thresholding technique.

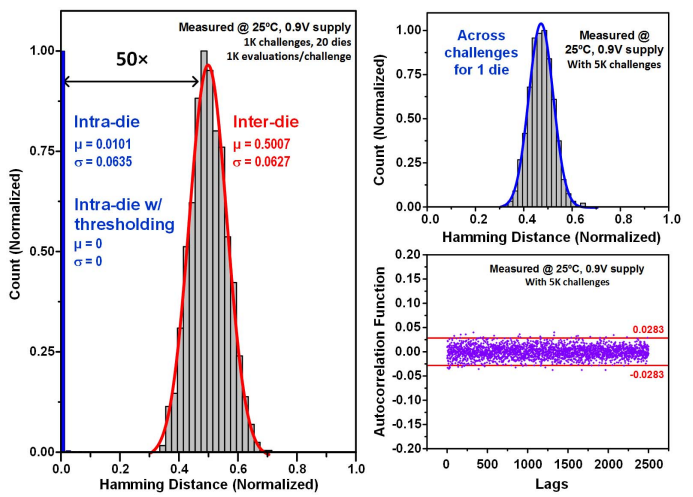

Figure 14.2.4: Measured intra-die and inter-die Hamming distances and autocorrelation function for responses to same challenges across 20 dies and responses to different challenges on a single chip.

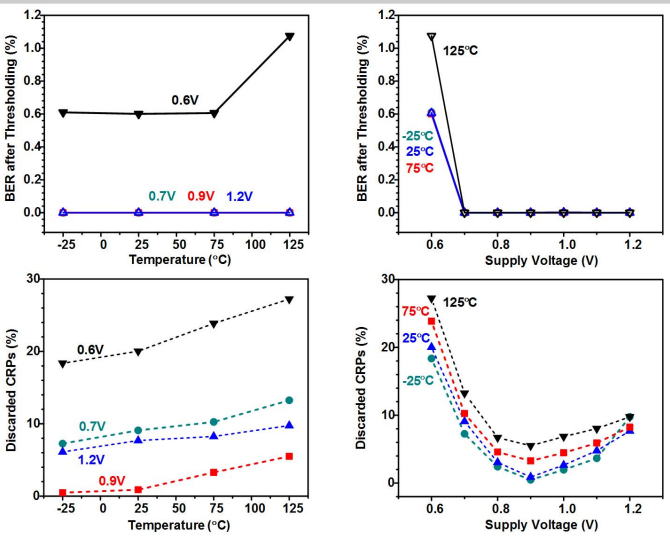

Figure 14.2.5: Measured BER and discarded CRPs over temperature and supply voltage variations, using threshold value of 16.

	This work (25°C, 0.9V core supply)	JSSC' 11 [7]	VLSI' 04 [5]	DAC' 07 [6]	ISSCC' 14 [4]	VLSI' 10 [2]	ISSCC' 07 [1]
Technology	40nm	90nm	0.18μm	FPGA	22nm	65nm	0.13μm
Architecture	Digital	Analog	Digital	Synthesized	Digital	Digital	Digital
Number of possible CRPs	~5.5×10^{28}	10^{25}	1.4×10^{20}	523776	1 (Chip ID)	1(Chip ID)	1(Chip ID)
BER in Typical Case	0	0.009%	0.7%	-	-	0	3.04%
Tested Operating Conditions: Temp(°C)	-25~125	25~125	27~67	-20~120	25~50	0~85	-
Tested Operating Conditions: Supply	0.7~1.2V	±10%	±2%	±10%	0.7~0.9V	±10%	-
BER in Worst Case	< 1×10^{-8} *	0.1%	4.8%	0.48%	0.97%**	0	-
Bit Rate (Mb/s)	1.6***	0.00625	20	-	2000	625	1
Core Area (μm²)	845	35000	-	-	4.66*256b =1193	1242	15288
Power (μW)	28.4	38	-	-	25	212.5	0.137
Efficiency (pJ/bit)	17.75	6080	-	-	0.19	0.34	1.6
Stabilizing Method	Dynamic thresholding	Mask	Voting	-	Voting, Burn-in, Mask, ECC	-	-

* Threshold is 16 for this BER measurement and BER is 0 in 100M bits tested.
** After ECC with BCH code, BER is 0.
*** Effective throughput = Clock frequency × (1 - Percentage of CRPs discarded during evaluation in worst condition).

Figure 14.2.6: Summary of measurement results and a comparison with state-of-the-art silicon PUF and chip ID designs.

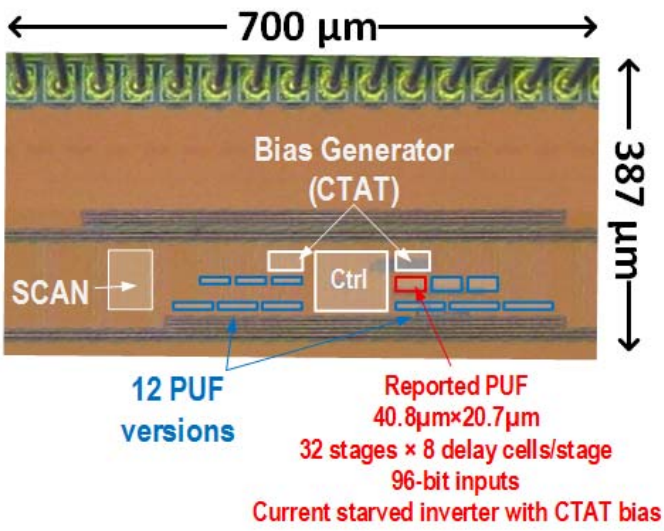

Figure 14.2.7: Die micrograph of 40nm CMOS PUF test chip.

978-1-4799-6224-2/15/$31.00