8.3 A 553F² 2-Transistor Amplifier-Based Physically Unclonable Function (PUF) with 1.67% Native Instability

Kaiyuan Yang, Qing Dong, David Blaauw, Dennis Sylvester

University of Michigan, Ann Arbor, MI

Physically Unclonable Functions (PUFs) are among the most promising security primitives for low cost solutions of key storage, chip authentication, and supply chain protection. Two types of PUFs exist in literature [1-6], a "strong" PUF with a large challenge-response space [6] and a "weak" PUF providing a limited length key (chip ID) [1-5]. While the former provides better security theoretically, existing implementations are prone to modeling attacks. So-called "weak" PUFs typically have an array of identically designed PUF cells that leverage device mismatch in fabrication as static entropy source, and serve as a low-cost and more secure alternative to non-volatile-memory-based key storage. Output stability across PVT variations and area are two critical metrics directly related to security and cost of a PUF. Recent works have presented custom PUFs based on NAND gates [1], current mirrors [2], PTAT [3], and cross-coupled inverters [4-5]. These outperform conventional SRAM-based PUFs, but sacrifice other metrics, e.g., [2, 4] are large, [3, 5] has lower native stability and energy efficiency, while [1] is sensitive to supply voltage and may experience large short circuit current. Finally, IoT and wireless sensor nodes tend to use older technologies for lower cost and standby power, which is challenging for PUF design because of smaller process variations.

This work presents a PUF cell based on a simple sub-threshold 2-transistor (2T) amplifier implemented in 180nm CMOS featuring: (1) a small 553F² PUF cell, integrated in an array with all peripheral circuits; (2) excellent stability: 1.65% native unstable bits, reaching 0.05% unstable bits with 11b temporal majority voting (TMV), and 3.16% and 2.01% flipping bits across wide temperature (-40-120°C) and voltage (0.8-1.8V) ranges; (3) high energy efficiency of 11.3fJ/b at nominal 1.2V and 1.5fJ/b at 0.8V; (4) high throughput (4.8Gb/s) via highly parallel operation, despite using an older technology. A masking technique using body bias is employed to find unstable bits without costly temperature sweeps.

The PUF cell is based on a 2T structure that acts in two different ways: an amplifier and a voltage generator (Fig. 8.3.1). A core NMOS with V_{gs}=0 sets a sub-threshold current. When the PMOS gate is used as an input port, the 2T structure is a common-source amplifier with pseudo-resistor loading. Thanks to the large g_m in sub-threshold, gain larger than 40 is provided by minimum-sized transistors. When the input and output of the amplifier are shorted (Fig. 8.3.1 top right) the 2T generates an output voltage that is equal to the "switching" voltage of the amplifier. This output tracks V_{DD} and solely depends on the threshold voltage differences between top and bottom devices (equation in Fig. 8.3.1), assuming both devices have $|V_{ds}|$>200mV, allowing sub-threshold drain-current dependence on V_{ds} to be ignored. When a 2T amplifier is connected to the output of an identically sized 2T voltage generator with the same switching voltage (neglecting mismatch), the amplifier output voltage equals its input voltage. However, mismatch will induce a small difference in the switching voltages of the 2 structures, which will then be amplified by the large amplifier gain. The switching voltage follows a normal distribution (Fig. 8.3.1) and therefore the difference also follows a normal distribution. Four amplifier stages are employed to amplify the voltage difference to full rail in >99.9% cases. This full-rail signal is then used as the digitized PUF output.

The switching voltage linearly tracks V_{DD} and varies across process corners and temperatures, but the PUF output is not affected because all stages are in the same PVT conditions and only mismatch dictates the result. In order to ensure $|V_{ds}|$>200mV for the PMOS transistor, the V_{th} of the NMOS must be significantly lower than that of the PMOS. This is accomplished by two different versions of the PUF: one with low V_{th} NMOS (LVT version) and the other with normal V_{th} NMOS using forward body biasing (DNW version). The LVT version has larger a cell due to design rules. The DNW version uses an on-chip body-bias generator and analog buffer to drive the p-well. We find that the DNW version has slightly better PUF stability when a CTAT p-well bias is applied (Fig. 8.3.1). Also, the bias generator inherently compensates for global PMOS/NMOS mismatch resulting in a more stable switching voltage ensuring $|V_{ds}|$>200mV across PVT conditions without calibration.

By embedding high gain in the PUF cell, we significantly improve stability and reduce complexity compared to shared amplifiers with noise and offset cancellation. It also enables arrangement of the PUF cells in a 16 by 64 crossbar array (Fig. 8.3.2) with the addition of 2 access transistors (Fig. 8.3.1). This improves area and energy efficiency compared to using scan chains [2, 4] or multiplexers [1] to read the PUF outputs. Another advantage of the crossbar configuration is that most SRAM read-assist techniques can be applied to improve read robustness and performance.

To fully characterize the PUF across process, we tested both TT dies and skewed corner dies (FF, SS, FS, SF). Uniqueness between PUF instances and uniformity inside a single PUF are fundamental requirements for PUFs. Fig. 8.3.3 shows inter-chip Hamming Distance (HD), with both versions exhibiting close to ideal distribution. Intra-chip HD indicates PUF repeatability and shows only 0.0008/0.0007b differences for the 2 versions, providing >600/700× separation between inter- and intra-chip average HDs (identifiability of PUF), compared to a previous reported value of 143× [2]. PUF uniformity is demonstrated by spatial autocorrelations with 0.0173/0.0167 bounds.

Output stability across PVT variations is the most critical metric for PUFs. Bit error rates (BERs) and unstable bits (# of bit locations with at least one error across all evaluations) are common metrics for stability [4], shown in Fig. 8.3.4. Stability of the 2T PUF can be further improved with 5b or 11b temporal majority voting (TMV). To characterize bit stability across V/T variations, additional bit error rates and flipping bits (unstable bits due to environmental factors after removing the impact of noise via majority voting) are measured (Fig. 8.3.4). Average results of all corner chips show as few as 0.2% additional flipping bits per 10°C change and 0.2% per 0.1V change across –40-120°C and 0.8-1.8V. The DNW version flipping bit data at 4 corners is also plotted to show that stability is not significantly affected by global process variation.

Masking is an efficient way to filter out unstable bits (dark bits). Conventional approaches include: 1) finding unstable bits by many evaluations at room temperature [4], which often misses bit flips due to temperature variations; and 2) finding unstable bits by sweeping temperature [1], which incurs high testing cost. This work uses external control of the PMOS n-well voltage (connected to V_{DD} during normal operation) to generate threshold voltage shifts and mimic temperature changes. The PMOS impacts both amplifier gain and the switching voltage, and as the input transistor, offers a larger impact than NMOS body biasing. The n-well body bias is swept at room temperature, avoiding the high costs of temperature sweeping. Without n-well body biasing, masking only improves BER at -40 and 120°C by 15.6%, while the new technique improves BER by up to 60% with ±0.3V body bias during testing (Fig. 8.3.5).

The 180nm DNW PUF delivers 4.8Gb/s with 64b wide outputs while consuming 11.3fJ/b for the PUF core. This energy includes PUF cell static power, bias generator/analog buffer static power, and bitline driver dynamic power and allows a fair comparison with prior works using scan chain to read PUFs (Fig. 8.3.5). PUF cell static power is 26pW/b while the shared bias generator/analog buffer consumes 185pW. Total array energy including all peripherals (timing generation, decoder, WL driver, and BL latches) is 91.1fJ/b. Best energy efficiency occurs at 0.8V with 1.5fJ/b core power and 52fJ/b total power. Fig. 8.3.6 summarizes measurement results and comparisons to prior PUFs. The die micrograph and PUF cell layouts are shown in Fig. 8.3.7.

References:

[1] B. Karpinskyy, et al., "Physically Unclonable Function for Secure Key Generation with a Key Error Rate Of 2E-38 in 45nm Smart-Card Chips," *ISSCC*, pp. 158–160, 2016.

[2] A. Alvarez, et al., "15fJ/b Static Physically Unclonable Functions for Secure Chip Identification with <2% Native Bit Instability and 140× Inter/Intra PUF Hamming Distance Separation in 65nm," *ISSCC*, pp. 256–257, 2015.

[3] J. Li and M. Seok, "A 3.07 μm²/bitcell Physically Unclonable Function with 3.5% and 1% Bit-Instability Across 0 to 80°C and 0.6 to 1.2V in a 65nm CMOS," *IEEE Symp. VLSI Circuits*, pp. 250–251, 2015.

[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*, pp. 278-279, 2014.

[5] Y. Su, et al., "A Digital 1.6 Pj/Bit Chip Identification Circuit Using Process Variations," *JSSC*, vol. 43, no. 1, pp.69–77, Jan. 2008.

[6] K. Yang, et al., "A Physically Unclonable Function with BER <10^{-8} for Robust Chip Authentication Using Oscillator Collapse in 40nm CMOS," *ISSCC,* pp. 254–255, 2014.

 978-1-5090-3758-2/17/$31.00

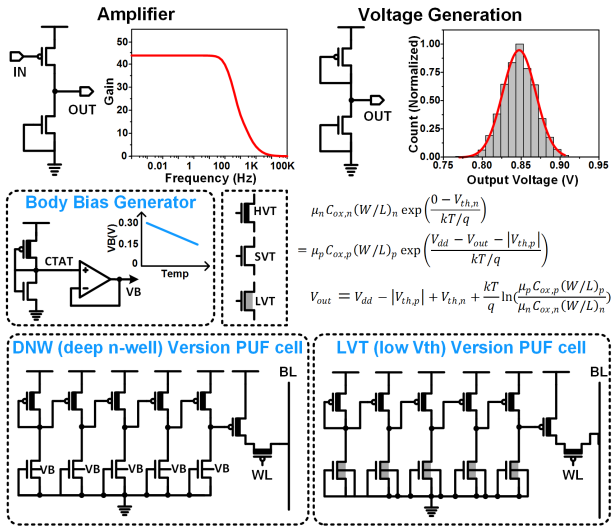

Figure 8.3.1: PUF cell using 2 transistor structure as switching voltage generator and amplifier, along with schematics of two PUF versions.

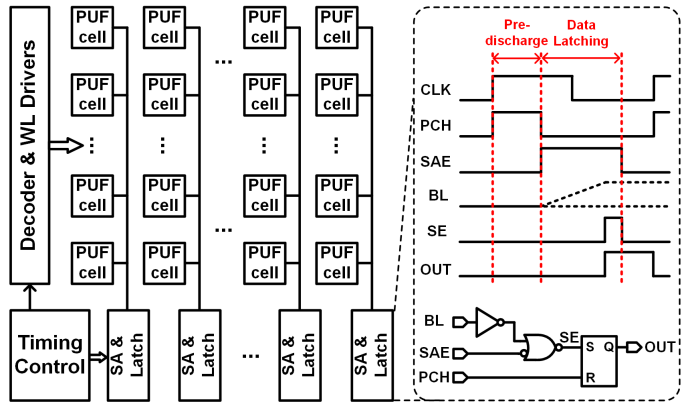

Figure 8.3.2: PUF crossbar array diagram along with read out circuits and waveforms, which are similar to that of SRAM.

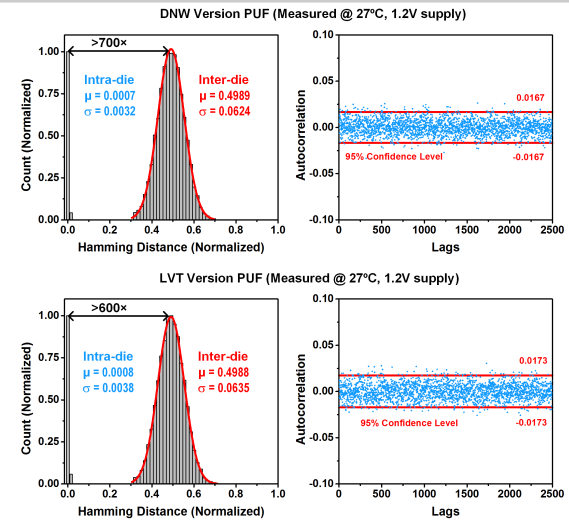

Figure 8.3.3: Measured intra-die/inter-die Hamming Distances, spatial autocorrelation function across 14 chips (6TT, 2FF, 2SS, 2FS and 2SF).

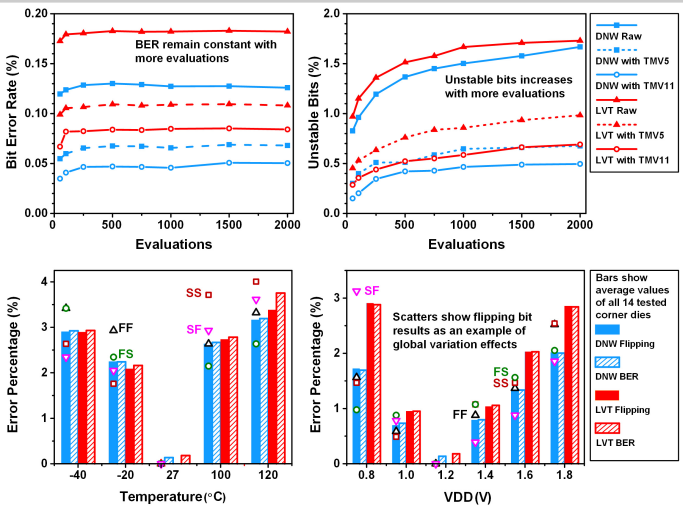

Figure 8.3.4: Measured BER and unstable bits vs. # of PUF readings, with and without TMV (top); BER and flipping bits across V_{DD} and temperature variations (bottom).

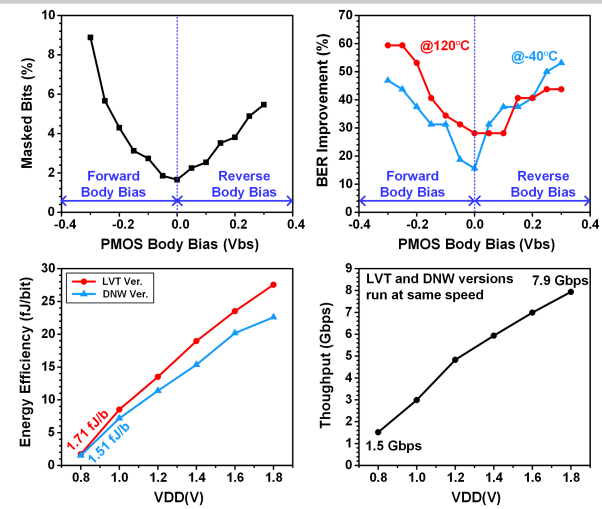

Figure 8.3.5: Measured masked bits percentage and BER improvement at extreme temperatures after proposed masking technique; measured PUF throughput and energy efficiency across V_{DD} (bottom).

		This work (LVT Version)	This work (DNW Version)	ISSCC' 16 [1]	ISSCC' 15 [2]	VLSI' 15 [3]	ISSCC' 14 [4]	JSSC' 08 [5]
Technology		180nm		45nm	65nm	65nm	22nm	130nm
PUF Cell Area/Bit (F^2)		782	553	2613	6036	756	9628	1092
Total Area/Bit (F^2)		1082	843	-	~36450	1756	-	1767
Native Unstable Bits (# of evaluations)		1.73% (2000)	1.67% (2000)	-	1.73% (400)	6.54% (500)	30% (5000)	-
Native Unstable Bits (# of Majority voting)		0.69% (TMV11)	0.50% (TMV11)	-	-	2% (TMV11)	3%[b] (TMV15)	-
Bit Error Rates (nominal condition)		0.18% 0.08% (TMV11)	0.13% 0.05% (TMV11)	0.1%[a]	-	-	8.3% 0.97%[d]	3.04%
Tested Operating Conditions	Temp (°C)	-40~120		-25~85	25~85	0~80	25~50	0~80
	Supply (V)	0.8~1.8		-	0.7~1	0.6~1.2	0.7~0.9	0.9~1.2
Bit Errors per 10°C		0.21%	0.2%	0.15%	0.47%	0.44%[b]	-	0.68%
Bit Errors per 0.1V		0.29%	0.2%	-	1.27%	0.17%[c]	0.49%[d]	1.82%
Bit Rate (Mb/s)		4832 @1.2V	4832 @1.2V	1.92	-	10.2	2000	1
PUF Core Energy (fJ/bit)		13.5 @1.2V 1.71 @0.8V	11.3 @1.2V 1.51 @0.8V	-	15	548	13	930
Norm. Inter-PUF Hamming Distance		0.499	0.499	0.498	0.5014	0.5001	~0.49	0.506

a. With 2-bit glitch detection
b. Comparator is re-calibrated manually at each temperature
c. Using off-chip ADC
d. After burn-in, 15-bit temporal majority voting (TMV15) and dark bits masking

Figure 8.3.6: Summary of measurement results and a comparison with state-of-the-art silicon PUFs.

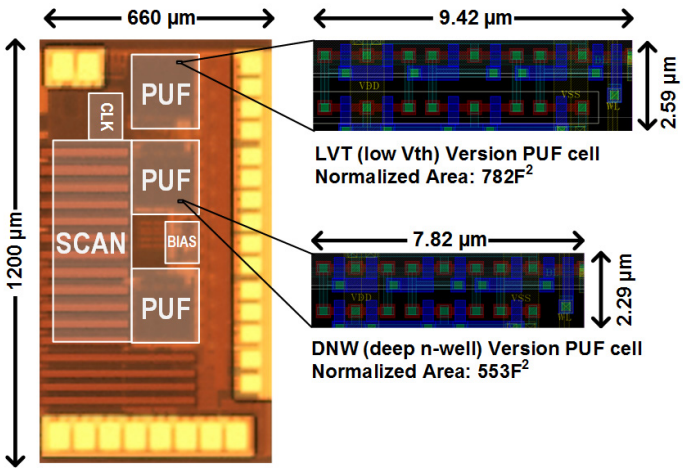

Figure 8.3.7: Die micrograph of 180nm PUF test chip and PUF cell layouts.

978-1-5090-3758-2/17/$31.00