Ultra-low Power Current-Based PUF

Mehrdad Majzoobi, Golsa Ghiaasi, Farinaz Koushanfar
Electrical and Computer Engineering Department
Rice University
Houston, Texas 77005
Email: {mm7,golsa,farinaz}@rice.edu

Sani R. Nassif
Tools and Technology Department
IBM Austin Research Labratory
Austin, Texas 78758-3493
Email: nassif@us.ibm.com

***Abstract*— In this paper, the first class of low power current-based physically unclonable functions (PUFs) is introduced. The new PUF circuit is able to convert the analog variations present in device leakage currents to a unique digital quantity at high speed and low power. Robust digital responses are achieved with the new architecture in presence of fluctuations in operational conditions such as temperature and supply voltage. The experimental results suggest 3% response error rate under extreme temperature variations from -55^{o}C to 125^{o}C and 20% fluctuations in supply voltage. The PUF consumes 150 μWatt for a duration of 250 ps per each response bit (37.5 femto joules of energy per response bit).**

I. Introduction

In the modern semiconductor business model, the increasing cost of maintaining in-house foundries is forcing many companies to ship their designs to overseas destinations for fabrication. Throughout this process, designs and IPs might be subject to piracy and potential overbuilding for black market sale. The need for controlling and monitoring the genuine hardware is more crucial than ever before. Additionally, on-chip storage of digital keys and secret for use in many cryptographic algorithms has been recently subject to many types of side-channel attacks which have undermined the security of embedded systems. In face of the aforementioned security challenges, novel alternative security mechanisms that extend the capacity of state-of-the-art cryptography are highly desirable.

To address these issues, the concept of physically unclonable functions (PUF) was introduced to exploit secrets that are physically bound to the hardware [1]. PUFs use the unclonable intrinsic analog manufacturing variability of silicon devices to provide a unique mapping from a set of digital inputs (challenges) to a set of digital outputs (responses). Imperfections and uncertainties in fabrication technology prevent duplication of the same hardware with identical device variability, hence the term unclonable. Moreover, the PUF responses must be prohibitively hard to simulate, emulate, or predict [1]. PUFs have found diverse applications including securing data processing [2], IP protection [3], [4], RFIDs, secure key generation [5], [6], remote activation and IC enablement [7].

The first class of silicon PUFs was introduced in [1], which uses digital logic delay variability to extract the secret. The existing classes of PUFs suffer from a multitude of problems including a relatively small challenge space, high implementation overhead, high power consumption, and low reliability of responses. For instance, the linear delay-based PUF in [1], [8] uses a series of 2-input/2-output switches to enable selection of different subpaths for each challenge input. Each switch is made of two 2-to-1 multiplexer and one inverter, which may consist of almost twenty transistors in total. A practical implementation often required 128 switches and long interconnects in between pairs of switches to allow for large enough variations sensible by the arbiter (see Section II for more details of operation). In addition, the generated responses suffer from high degree of instability in presence of fluctuations in environmental conditions. Other types of PUFs such as ring oscillator (RO) PUF [5] work by measuring and comparing the unique oscillation frequency across a group of ROs. The ROs need to be kept oscillating for a given amount time in order to measure their frequency and as a result they have a relatively high power consumption. Also the possibilities of different unique pairings of the ROs for comparison are polynomially bounded by the number of ROs. SRAM PUF [9], [10] and Butterfly PUF [4] suffer from the similar issue of small challenge/response pair space.

In this paper, we introduce the first current-based PUF architecture that enables linear combination of individual currents. The proposed PUF converts the analog variations present in device leakage current to unique digital responses based on the input challenges at high speed and low power. The PUF automatically cuts off the current flow once the responses are generated. We investigate the optimal point of operation for maximal robustness and uniqueness of responses. Our contributions are as follows:

- We introduce the first circuit architecture for building low-power current-based PUFs.
- The introduced PUF structure delivers ultra-low power consumption by operating on subthreshold leakage currents and employing an automatic cutoff mechanism to stop the current flow after response evolution.
- We investigate the optimal operational parameters to achieve highest level of robustness in presence of extreme variations in operational conditions such as temperature and supply voltage.

II. Background and Related Work

The first implementation of PUFs on silicon is introduced in [1], where the delay variations of CMOS logic components are used to produce unique responses. In the delay-based PUF,

the analog delay difference between two structurally identical parallel paths is compared. Due to manufacturing variations, the delay of these two paths are slightly different. The architecture of the arbiter-based PUF with two racing parallel paths is demonstrated in Figure 1. A step input simultaneously triggers the two paths. At the end of the two parallel paths, an arbiter is used to convert the analog delay difference between the paths to a digital value. The arbiter can be implemented by a D-flip flip in practice. The two paths can be divided into several smaller subpaths by inserting path swapping switches. Each set of inputs to the switches act as a challenge set (denoted by C_i), defining a new pair of racing paths whose delays can be compared by the arbiter to generate a one-bit response. A working implementation of the delay-based PUF on FPGA is shown in [8] using high resolution programmable delay lines and symmetric switch structures. The work in [11] shows an implementation of the delay-based PUF on ASICs that operates in sub-threshold conditions and uses a symmetric arbiter circuit to achieve higher stability of responses. Athors in [5] take advantage of the unique oscillation frequency of a group ring oscillator along with a pairing mechanism to compare the measured frequencies. The impact of mismatch and manufacture variability on the SRAM power-on states is utilized in [9], [4] to extract secret digital bits. A similar work in [10] implements the same concept with nonstandard custom SRAM cells. A digital ID extraction system based on device mismatch using an auto-zeroing comparator was introduced in [12]. The major difference between static ID generation using physical device variations and PUFs is the lack of challenges in the former to select and combine the analog variations before quantization/digitization. Finally in order to safeguard linear PUFs against machine learning and reverse engineering attacks, addition of input/output XOR networks and feedforward arbiters are proposed in [6], [13]. A comprehensive study of machine learning attacks on PUFs and their complexity is provided in [14].

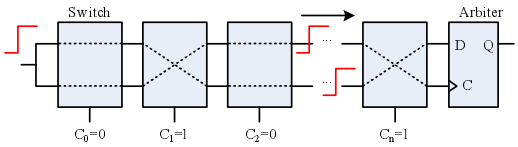

Fig. 1. Delay-based PUF.

III. Concept and Circuit Realization

In this section, we present the concept and circuit architecture of the new low power current-based PUF. Figure 2 depicts the conceptual architecture of our new PUF circuit. First, process sensitive (PV) voltages/currents are generated. These quantities should ideally be as much sensitive to process parameters as possible but highly insensitive to environmental parameters such as temperature to achieve high levels of response stability and robustness. Next, based on a given input challenge, a subset of these voltages/currents are selected and combined. The combined quantities are compared and converted to digital responses. The comparator maybe tuned for maximum accuracy and reliability based on the predicted statistics of the compared signals. The circuit implementation

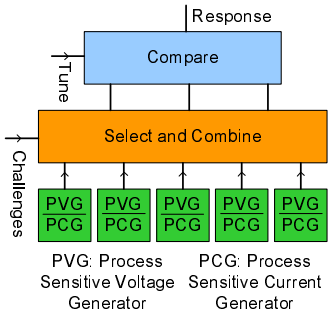

Fig. 2. The conceptual block diagram of the proposed PUF structure.

of the proposed PUF concept is shown in Figure 3. In this implementation, process sensitive currents are generated by using individual FETs whose gate voltages are tied to a fixed voltage source. Next, based on the input challenge that drive the differential current switches, a subset of currents are selected and combined. In other words, by connecting the outputs of the current switches as illustrated in Figure 3 and controlling the inputs to the current switches, we can select and add up a subset of currents into either left and right side of the circuit which accordingly flows into the left and right inputs of the sense amplifier. Note that if both input challenges to a current switch are set to '0' (ground) no current will flow to either left or right sides. Additionally if both input challenge bits are set to '1' (V_{DD}), then half of the total current that enters the current switch will flow through each side. If input challenge bit on one side is '0' and '1' on the other side, the total current that enters the current switch from the bottom single FET current generator will be stirred to the latter side. Equations 1 and 2 formally express each current in terms of the inputs to the current switch.

$$I^a[i] = \begin{cases} I[i], & \text{if } C^a[i] = 1 \text{ and } C^b[i] = 0; \\ 0.5I[i], & \text{if } C^a[i] = 1 \text{ and } C^b[i] = 1; \\ 0, & \text{if } C^a[i] = 0 \text{ and } C^b[i] = X; \end{cases} \quad (1)$$

$$I^b[i] = \begin{cases} I[i], & \text{if } C^a[i] = 0 \text{ and } C^b[i] = 1; \\ 0.5I[i], & \text{if } C^a[i] = 1 \text{ and } C^b[i] = 1; \\ 0, & \text{if } C^a[i] = X \text{ and } C^b[i] = 0; \end{cases} \quad (2)$$

The 'X's in Equation 1 and 2 represent 'don't-care'. $I^a[i]$ and $I^b[i]$ denote the left and right output currents of the i-th current switch respectively. Also $C^a[i]$ and $C^b[i]$ respectively represent the left and right inputs to the i-th current switch. Therefore the total current on the left side, i.e. I^a in Figure 3, and on the right side, i.e. I^b in Figure 3, can be written as the sum of each individual current on each side,

$$I^a = \sum_{i=1}^{N} I^a[i], \qquad I^b = \sum_{i=1}^{N} I^b[i] \quad (3)$$

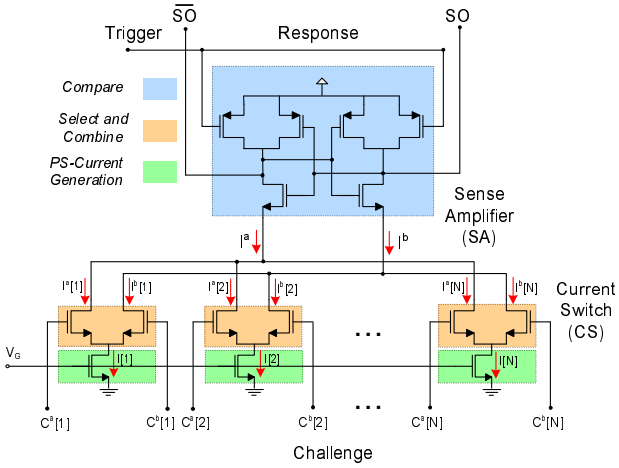

Fig. 3. The proposed current based PUF system.

where N is the total number of PV current generator FETs (or current switches). Now, the total current on both sides flows into a latch-based sense amplifier. The sense amplifier, based on which current is larger, will produce a zero or one digital response, i.e.,

$$\text{response} = \begin{cases} 1, & \text{if } I^a > I^b; \\ 0, & \text{if otherwise;} \end{cases} \tag{4}$$

The latch-based sense amplifier used in the PUF system in effect consists of a pair of back-to-back connected inverters. Initially, the output of the inverters are pulled up to V_{DD} by the trigger signal, charging the output node capacitances. Once the challenges are applied, trigger signal goes to zero releasing the output nodes. Soon after the currents start flowing though both sides of the sense amplifier, the output capacitances begin discharging. The discharge pace of the node capacitances is a function of each current magnitude; i.e., the larger the current, the faster the discharge. Whichever node voltage drops first by V_{th} turns on the top inverter transistor and establishes a positive feedback which settles to a response. After the sense amplifier settles, one of the transistors in each inverters turns off and the current flow stops automatically.

In order to avoid any bias and predictability of the output responses and to achieve maximum randomness in responses, the mean/nominal value of the compared currents must be the same. Meeting such property requires the number of combined currents on each side or equivalently the number of ones in the right and left input challenges to be equal, i.e.,

$$\sum_{i=1}^{N} C^a[i] = \sum_{i=1}^{N} C^b[i]. \tag{5}$$

In case of existence of any bias in sense amplifier operation, calibration and tuning can be performed by introducing imbalances in Equality 5 to have more the number of ones in challenges (larger the nominal current value) on the desired side. This degree of freedom can also be used to sift and distinguish the robust challenges from unstable challenges [8].

IV. Experimental Results

In this section, the evaluation results for the new PUF architecture are presented. We simulate the system with N=64 current generators (and current switches) using IBM 90nm technology models. To achieve maximum level of variability, the device sizes are set to the technology minimum of W/L = 120nm/100nm. Using Monte Carlo simulation guided by the IBM statistical models, 100 circuit instances are generated. Next, we apply 100 challenges to each PUF circuit instance at frequency of 100MHz and the responses to the applied challenges are evaluated and stored. We refer to this setup as the base experiment. In what follows, we run multiple instances of the basic experiment under different scenarios and operational conditions. Figure 4 shows the sense amplifier response waveform to a series of random input challenges.

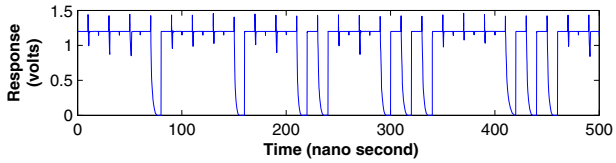

Fig. 4. The sense amplifier output response waveform to a set of random challenges.

The first experiment consists of twelve base experiment runs under the combination of the following two sets of scenarios; In the first set of scenarios, the number of active currents on each side is set to 8, 16, 32 (the number of ones in each challenge vector, i.e., $K = \sum_{i=1}^{N} C^a[i] = \sum_{i=1}^{N} C^b[i]$ where $K = \{8, 16, 32\}$). In the second set of seniors, the gate voltage of the current-generator FETs (V_{gate}) is set at different ratios of V_{DD}, i.e. $V_{gate} = \{0.1, 0.3, 0.5, 0.7\} \times V_{DD}$. All of the base experiments are performed under normal operating conditions i.e. temperature of 25^{o}C and V_{DD}=1.2 volts. Therefore, if we define E_1 as the set of scenarios for the first experiment, then $E_1 = \{S_{k,v} \mid (k, v) \in K \times V_{gate}\}$, where $S_{k,v}$ is the experiment scenario under a given k and v.

For each experiment the number of '1's in 100 responses is counted and normalized to 100. Ideally, we would like to have equal number of ones and zeros in responses for highest level of randomness (see [15]). Figure 5 shows the distribution of this value across the 100 PUF instances versus different gate voltages for different number of active currents using boxplots. The central mark on the boxplot denotes the median, the edges of the boxes correspond to the 25th and 75th percentiles, the whiskers extent to the most extreme data points and the red plus signs show the outlier points. As it can be observed on the plots, for V_{gate}/V_{DD} = 0.1 the responses are highly biased toward '1'. A closer investigation reveals that the for this gate voltage, the generated currents are too small to provoke any response from the sense amplifier.

The goal of the second and third experiments is to find the operation parameters which achieves the highest level of robustness against the fluctuations in temperature and supply voltage. Similar to the previous experiment, we first define a

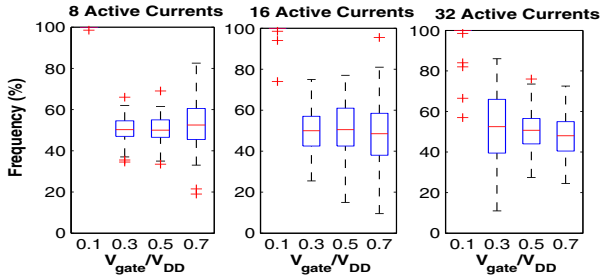

Fig. 5. The distribution of number of ones in responses to 100 challenges over 100 PUFs.

set of scenarios under which we run the base experiment. Let us denote the second set of experiments by E_2 such that $E_2 = \{S_{k,v,t} \mid (k,v,t) \in K \times V_{gate} \times T\}$ where $T = \{-55, 125\}$ are the operating temperatures in Celsius degrees, and K and V_{gate} are the sets defined previously. Next the responses from experiments in scenarios $\{S_{k,v,t} \mid t = -55\}$ are compared to the responses from $\{S_{k,v,t} \mid t = 125\}$ for all k, v and the discrepancies and differences are counted and normalized to the total number of responses (=100). Note that the same challenges are applied to the PUF in each experiment. These two low and high temperatures correspond to standard military operational conditions. The same experiment is repeated for $T = \{-40, 85\}$ and $T = \{0, 75\}$ each corresponding to industrial and commercial operational conditions respectively. The plots on the top row of Figure 6 depict the results of this experiment. The 'y' axis on each plot shows the error rate in the responses averaged over the 100 PUF instances and the 'x' axis corresponds to the gate voltage of the current generator FETs. The lines on each plot marked by stars, circles, and dots correspond to commercial, industrial and military operational conditions respectively. The columns in the plot from left to right correspond to the cases where 8, 16, and 32 currents are activated, combined, and compared. As it can be observed, increasing the gate voltage of the current generator FETs raises the response error rates and thus reduces the level of robustness in responses. Moreover, the results suggest that as larger number of currents are combined, the error rate also increases. Note that the error rates for $V_{gate}/V_{DD} = 0.1$ are invalid due to the large bias in the responses as shown in Figure 5. The plots in the bottom row of of Figure 6 present the same results, however, this time the temperature is fixed to 25^{o}C and supply voltage is varied in three intervals of $V_{DD} = \{1.1, 1.2\}$, $V_{DD} = \{1.1, 1.3\}$, and $V_{DD} = \{1, 1.4\}$. The same conclusions apply to these results as well. Finally, note that the lowest error rate can be achieved for the smallest sub-threshold currents that are large enough to drive the sense amplifier. The PUF consumes 150 μWatt for a duration of 250 ps per each response bit.

V. Conclusion and Future Work

The first class of low power current-based physically unclonable functions was introduced and studied for optimal operation parameters. The PUF yields the highest responses

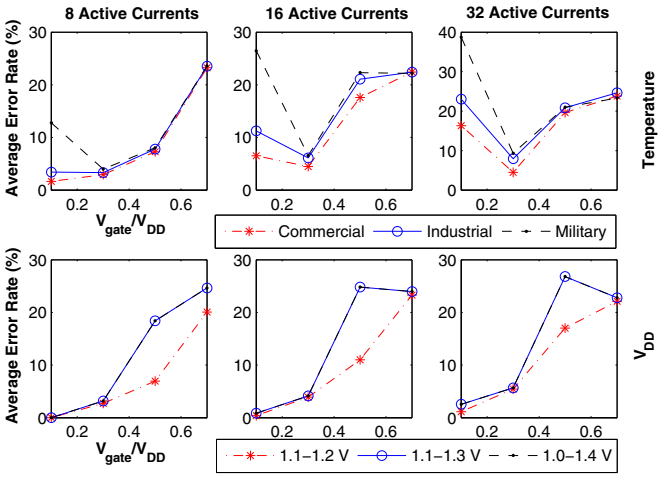

Fig. 6. The average response error rate as a function of the current generator transistor gate voltage.

by using the lowest gate voltage under which the sense amp still function. Our study revealed that the more currents are combined, the lower the stability and robustness of responses against variations in operational conditions. The experimental results suggest 3% response error rate under extreme temperature variations from -55^{o}C to 125^{o}C and 20% fluctuations in supply voltage.

References

[1] B. Gassend, D. Clarke, M. van Dijk, and S. Devadas, "Silicon physical random functions," in *CCS*, 2002, pp. 148–160.

[2] G. Suh, C. O'Donnell, I. Sachdev, and S. Devadas, "Design and implementation of the AEGIS single-chip secure processor using physical random functions," in *ISCA*, 2005, pp. 25–36.

[3] J. Guajardo, S. Kumar, G. Schrijen, and P. Tuyls, "FPGA intrinsic PUFs and their use for IP protection," in *CHES*, 2007, pp. 63–80.

[4] S. Kumar, J. Guajardo, R. Maes, G. Schrijen, and P. Tuyls, "The butterfly PUF protecting IP on every FPGA," in *HOST*, 2008, pp. 67–70.

[5] G. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *DAC*, 2007, pp. 9–14.

[6] M. Majzoobi, F. Koushanfar, and M. Potkonjak, "Techniques for design and implementation of secure reconfigurable PUFs," *ACM TRETS*, vol. 2, no. 1, pp. 1–33, 2009.

[7] Y. M. Alkabani and F. Koushanfar, "Active hardware metering for intellectual property protection and security," in *USENIX Security Symposium*, 2007, pp. 1–16.

[8] M. Majzoobi, F. Koushanfar, and S. Devadas, "FPGA PUF using programmable delay lines," in *WIFS*, 2010.

[9] D. Holcomb, W. Burleson, and K. Fu, "Power-up SRAM state as an identifying fingerprint and source of true random numbers," *IEEE Trans. Computers*, 2009.

[10] S. Ying, J. Holleman, and B. Otis, "A digital 1.6 pj/bit chip identification circuit using process variations," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 1, pp. 69 –77, jan. 2008.

[11] L. Lin, D. Holcomb, D. Krishnappa, P. Shabadi, and W. Burleson, "Low-power sub-threshold design of secure physical unclonable functions," in *ISLPED*, 2010, pp. 43–48.

[12] K. Lofstrom, W. Daasch, and D. Taylor, "IC identification circuit using device mismatch," 2000, pp. 372 –373.

[13] L. Daihyun, J. Lee, B. Gassend, G. Suh, M. van Dijk, and S. Devadas, "Extracting secret keys from integrated circuits," *IEEE Transactions on VLSI Systems*, vol. 13, no. 10, pp. 1200 – 1205, 2005.

[14] U. Rhrmair, F. Sehnke, J. Slter, G. Dror, S. Devadas, and J. Schmidhuber, "Modeling attacks on physical unclonable functions," in *Conference on Computer and Communications Security*, 2010.

[15] M. Majzoobi, F. Koushanfar, and M. Potkonjak, "Testing techniques for hardware security," in *ITC*, 2008, pp. 1–10.