A DRAM based Physical Unclonable Function Capable of Generating >10^{32} Challenge Response Pairs per 1Kbit Array for Secure Chip Authentication

Qianying Tang, Chen Zhou, *Woong Choi, *Gyuseong Kang, *Jongsun Park, Keshab K. Parhi, and Chris. H. Kim
*Korea University, Seoul, Korea, University of Minnesota, Minneapolis, MN 55455 USA

Abstract- **A DRAM based Physical Unclonable Function (PUF) utilizing the location of weak retention cells is demonstrated in 65nm CMOS. A new authentication scheme is proposed for the DRAM PUF where a random pattern is written to a small section of the DRAM and then retention failures are induced. To further increase the number of Challenge Response Pairs (CPRs), the data pattern including retention failures is transferred to a different memory location where additional retention failures are induced. This scheme enables more than 10^{32} unique CRPs from a 1Kbit array. To improve the stability of the PUF response, a zero-overhead repetitive write-back technique along with bit-masking was utilized. Voltage and temperature induced instabilities were mitigated by adjusting the read reference voltage and refresh time before each authentication operation. The proposed DRAM PUF has a bit cell area of 0.68μm^2.**

I. INTRODUCTION

Physical Unclonable Functions (PUFs) have emerged as an effective tool for protecting semiconductor devices against physical attacks such as counterfeiting and secret key theft [1]-[3]. PUFs utilize manufacturing variability inherent in circuits as their entropy source. Therefore, when presented with a set of randomly selected challenges, each PUF generates the corresponding set of responses that is unique and almost impossible to duplicate. The secret information stored in a PUF circuit is only available when the device is powered on, and therefore PUF based systems are highly immune to offline attacks.

SRAM or DRAM are attractive candidates for PUF [4][5] as they are readily available in most processors, requiring almost no modifications to the underlying hardware. Moreover, the array based structures provide a large set of independent entropy sources. However, one shortcoming of memory based PUFs compared to delay based PUFs [6][7] is that the number of Challenge Response Pairs (CRPs) is linearly proportional to the number of circuit units. Memory PUFs are therefore categorized as "weak" PUFs [1] and are not suitable direct chip authentication applications. To address this shortcoming, we present a novel DRAM based "strong" PUF capable of generating >10^{32} CRPs from a 1Kbit array. The main highlights of this work are: 1) a local encrypting scheme that enhances the authentication security and allows a DRAM to serve as a strong PUF; 2) a repetitive write-back scheme based on existing DRAM refresh circuits for enhancing PUF stability; and 3) a simple calibration routine to suppress voltage and temperature variation effects.

II. SRAM PUF VS. DRAM PUF

Fig. 1 compares the properties of SRAM PUF and DRAM PUF. Unlike SRAM PUFs where the supply voltage is turned off and turned on to generate a response, a DRAM PUF can be accessed anytime during normal operation by writing a '0' or '1' and checking whether the data has flipped or not after a certain retention time. This unique feature allows us to generate an exponentially high number of CRPs. The schematic of the 2T DRAM [8] used in this work is shown in Fig. 2 (a). Compared to a 1T1C DRAM cell, this type of DRAM cell does not require a dedicated trench or stacked capacitor process, and has decoupled read and write paths enabling good low voltage margin. The data retention time depends on the storage node capacitance and the leakage current surrounding the storage node. The read reference voltage (Vref) can be adjusted such that a certain number of cells fail for a given retention time. Fig. 2 (b) shows several retention time failure scenarios. A cell with strong pull down leakage will not hold a data '1' value very well, and vice versa. Locations of weak retention cells are unique to each chip.

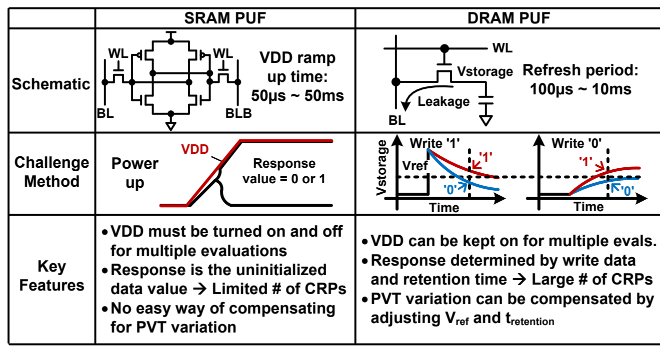


Fig. 1. Qualitative comparison between SRAM PUF and DRAM PUF.

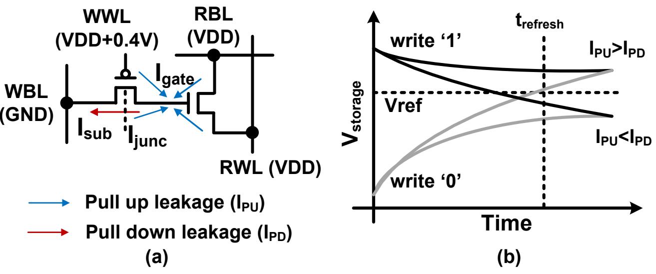


Fig. 2. (a) 2T DRAM cell schematic and leakage components in hold mode. (b) A DRAM cell generates a different response depending on the write data and retention time.

III. PROPOSED DRAM PUF DESIGN

The proposed authentication scheme illustrated in Fig. 3 comprises of four steps: (1) writing a random pattern (=challenge from server) to a small portion of the DRAM array; (2) letting 10% of the cells to fail by reading the data after a certain retention time; (3) transferring the data stored in the array to a different location in a random bitwise mapping fashion; and finally (4) repeating step (2). The second step

provides local encryption which generates a new random pattern that is hidden from the outside world, providing an added level of security. The local encryption operation cannot be implemented in an SRAM PUF because it requires a random data pattern to be written to the memory array. Access to the local encrypted pattern is only allowed during chip enrollment phase and will be permanently disabled thereafter using fuses. Randomly transferring the array data to a different location, combined with the initial random pattern from the server, enables an exponentially higher number of CRPs. The total number of CRPs for a 10% retention failure probability can be calculated as follows.

$$N_{CRP} = 2 \cdot n \cdot P(n, i \times 10\%)$$

Here, the pre-factor '2' represents the two values that can be written to a DRAM cell, P(.) is the permutation function, 'n' is the half array size, and 'i' is the number of response bits. According to this equation, the total number of CRPs attainable from a 1Kbit DRAM array for a 128-bit response output is greater than 10^{32}.

The enrollment and authentication procedures of the proposed DRAM PUF are shown in Fig. 4. Compared to conventional strong PUFs which require an exhaustive test to collect a large number of CRPs, the proposed PUF only needs to store whether a retention failure occurs or not for data '0' and data '1' under a certain retention time. So for a 1Kbit array, the unique PUF information can be stored in just two 1Kbit maps, one for data '1' and one for data '0'. To generate the bit maps, we first write all '1's to the 1Kbit DRAM array, let retention failures occur, and then read the pattern including retention failures. The same procedure is repeated for data '0'. The bit maps are stored on the server as reference key. Fig. 4 illustrates the authentication flow for generating a 128-bit response from a 1Kbit array. During authentication, the initial 128-bit random pattern along with the 128 x 10 bit random mapping information is generated by the server and sent to the chip as challenge bits. Based on the reference key, the server computes the expected response and compares it with the response from the chip. If the Hamming Distance (HD) between the two responses satisfies the match criterion, access permission is granted to the user.

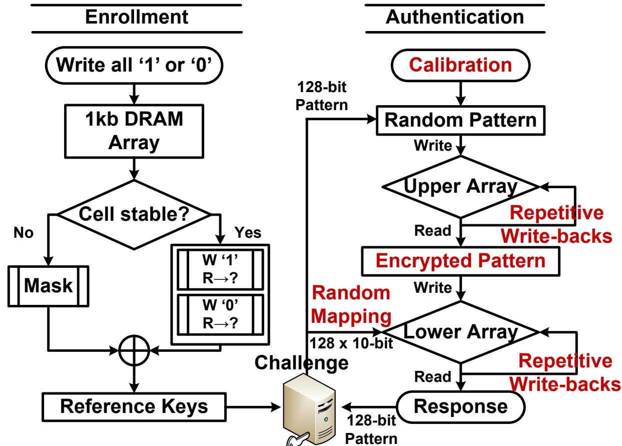


Fig. 4. Overall enrollment and authentication flow of the proposed DRAM PUF. New techniques proposed in this work are highlighted in red.

IV. IMPROVING DRAM PUF RELIABILITY

Fig. 5 shows the soft response distribution measured from a 1Kbit DRAM array for 500 trials. Soft response is defined as the average of 500 response values for a particular DRAM cell. For example, if the response is '1' for 90% of the time and '0' for 10% of the time, then the soft response value is 0.9. The left-most and right-most bars represent the stable cells with 0% and 100% retention failures for the entire 500 trial period. The bars in the middle represent unstable cells that generate some retention time failures. Experimental data shows that the percentage of unstable cells (i.e. 0 < soft response < 1) are 8.8% and 7.2% for data '1' and data '0', respectively.

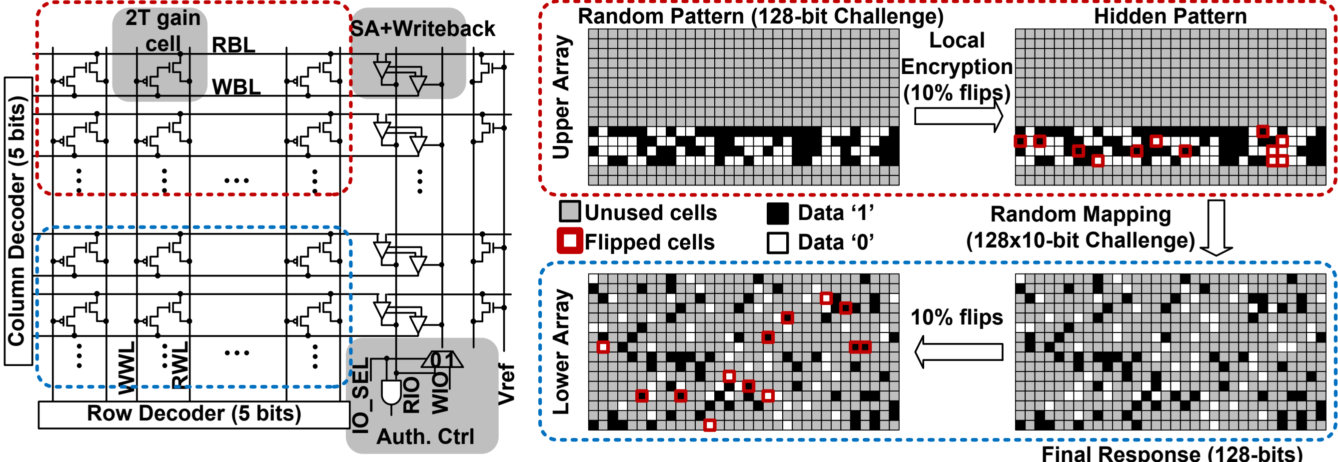


Fig. 3. The proposed authentication scheme consists of four steps: (1) write random 128 bit challenge to DRAM upper array, (2) allow 10% of bits to flip due to retention failure, (3) transfer data to lower array according to random mapping info from server, and finally (4) repeat step (2). The inherent DRAM retention failure rate is utilized for generating a unique and secure response. For the chip demonstration, we chose a 128-bit random input pattern, a 128 x 10 bit random address mapping info (=128+128x10=1,408 total challenge bits) and a 128-bit response.

To reduce the percentage of unstable cells, many PUF designs employ Temporal Majority Voting (TMV). TMV is a technique in which a PUF is evaluated multiple times using the same challenge and the majority output value is taken as the final response. The main drawback of TMV is the large area and delay overhead for storing and processing the PUF outputs from each TMV trial. For example, to perform a 15 trial TMV, a 4 bit counter is required for each accessed cell. Sharing a single TMV counter for the entire array will reduce the area overhead, but the authentication time becomes prohibitively long. As an alternative to TMV, we propose a repetitive write-back scheme that can be implemented using existing DRAM refresh circuitry with no hardware overhead.

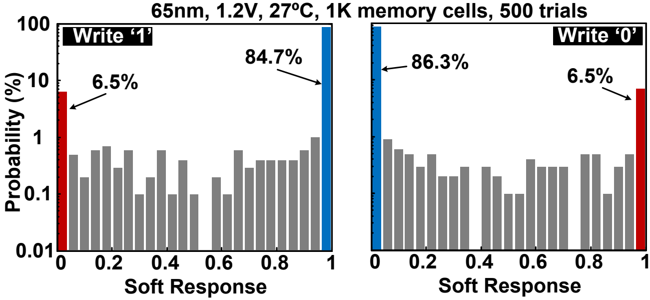


Fig. 5. DRAM PUF soft response distribution for data '1' and data '0'. Soft response is defined as the average response value over 500 trials. For example, if the output for a particular memory cell is 1 for 90% of the time and 0 for 10% of the time, the soft response for this memory cell is 0.9.

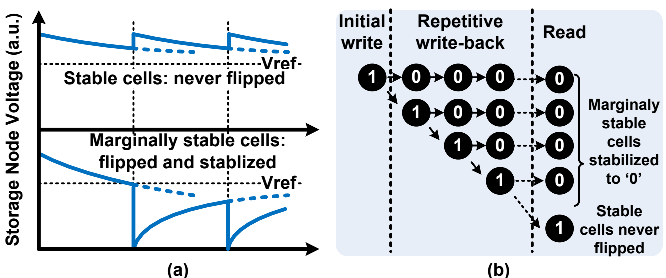
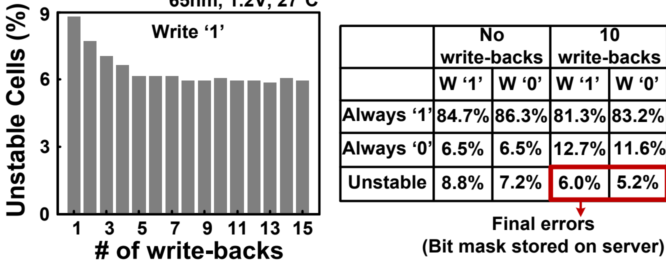


Fig. 6. Repetitive write back scheme for improving DRAM PUF stability. (a) Waveforms of DRAM cell storage voltage with repetitive write back. (b) Marginally stable bits can be stabilized to the opposite value with repetitive write-back.



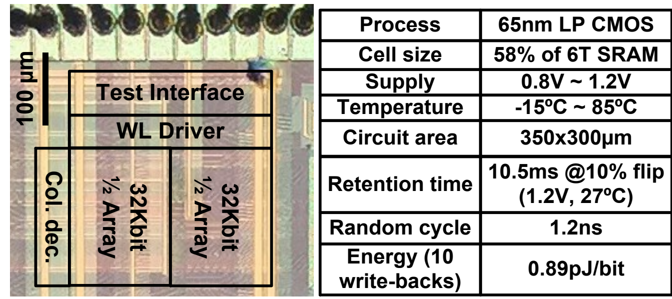
	No write-backs		10 write-backs	
	W '1'	W '0'	W '1'	W '0'
Always '1'	84.7%	86.3%	81.3%	83.2%
Always '0'	6.5%	6.5%	12.7%	11.6%
Unstable	8.8%	7.2%	6.0%	5.2%

Fig. 7. Percentage of unstable cells decreases with more write-backs. Cells that remain unstable after 10 write-backs will be flagged and masked by the server during chip authentication.

The idea is based on the fact that a cell with a small read margin for data '1', generally has a large read margin for data '0', and vice versa. Based on this unique characteristic, we propose the repetitive write-back scheme shown in Fig. 6 where DRAM cells are written with the data read from the previous cycle. Measurement results in Fig. 7 verify that the percentage of unstable cells decreases, although after 5 cycles it levels out. After 10 write-back cycles, the percentage of unstable cells reduces from 8.8% to 6% for data '1', and from 7.2% to 5.2% for data '0'. Although the improvement is not significant, the repetitive write-back scheme is still useful as it incurs no hardware overhead. DRAM cells that remain unstable after the repetitive write-back will be masked by the server.

V. Test Chip Measurement

A 64Kb DRAM PUF array was fabricated in a 1.2V, 65nm process for concept verification. Fig. 8 shows the die photo and key features. We selected a 32x32 (=1Kbit) DRAM subarray to test the proposed PUF authentication scheme. The measured intra-chip HD and inter-chip HD distributions are shown in Fig. 9. The former represents the reproducibility of the responses while the latter represents the uniqueness of the responses. The intra-chip HD was obtained by applying the same challenge 500 times. The intra-chip HD measured under a nominal condition (1.2V and 27˚C) has an average of 2.2% and a standard deviation of 1.03%. After masking the unstable responses, the average and standard deviation of intra-chip HD improves to 0.39% and 0.19%, respectively. Note that bit masking was performed on the server side which obviates the need for an Error Correcting Code (ECC) unit. The inter-chip HD distribution was obtained by applying 10k random challenges to 15 different chips. The distribution has an average of 35.9% and a standard deviation of 6.35%. The reason why the average inter-chip HD is not centered around 50% is because we deliberately chose a retention failure rate of 10% (and not 50%) to speed up the authentication process. Moving the inter-chip HD distribution to the center is possible but at the expense of a longer authentication time. The margin between the intra-chip and inter-chip HD distributions is 10.2%.



Process	65nm LP CMOS
Cell size	58% of 6T SRAM
Supply	0.8V ~ 1.2V
Temperature	-15ºC ~ 85ºC
Circuit area	350x300μm
Retention time	10.5ms @10% flip (1.2V, 27ºC)
Random cycle	1.2ns
Energy (10 write-backs)	0.89pJ/bit

Fig. 8. 65nm DRAM PUF chip micrograph and summary table.

Maintaining a narrow intra-chip HD distribution across different voltages and temperatures is imperative for PUFs in real products. As shown in Fig. 10, the retention failure probability varies significantly under different voltage and temperature conditions. To mitigate V and T effects, a calibration scheme is proposed based on the observation that the order of failure locations remains almost the same under different V and T conditions. This is because V and T affect each DRAM cell in the same way. Basic operation of the calibration scheme is given in Fig. 11. First, a checker board

pattern is written into the array. The calibration circuit measures the ratio between '1's and '0's after a certain retention period by counting the number of '1's in the array pattern. If the percentage of '1's does not fall within the desired range (e.g. 49%-51%), Vref is adjusted accordingly. Finally, the calibration circuit adjusts the refresh time $t_{refresh}$ to ensure that the retention failure probability P_{flip} is close to the target (e.g. 10%). Once the calibration is complete for a specific V and T condition, the authentication procedure depicted in Fig. 4 (right) can start.

Fig. 11 shows the intra-chip HDs obtained under different supply voltages and temperatures. Using the calibration scheme, the average intra-chip HD is reduced from 15% to 2% for a voltage range of 0.8V to 1.2V. Similarly, the average intra-chip HD is reduced to 1.5% for a temperature range of -15°C to 85°C.

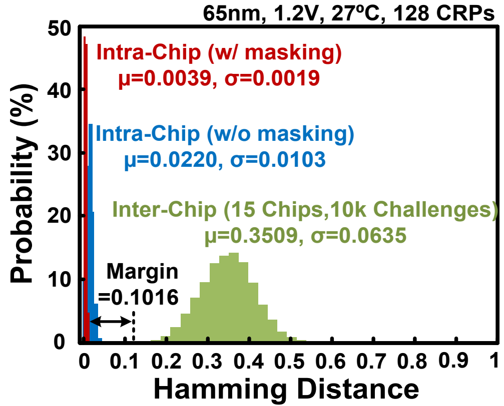


Fig. 9. Measured inter-chip (15 chips and 10,000 different challenges) and intra-chip Hamming Distance with and without bit masking.

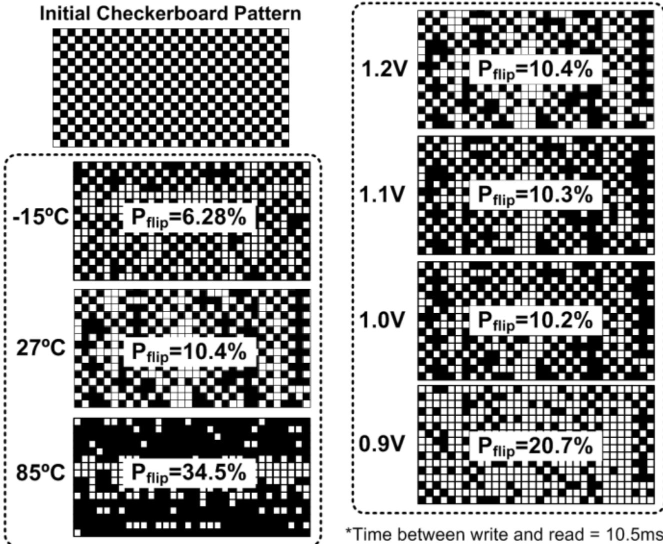


Fig. 10. DRAM retention failure map measured under different supply voltages and temperatures. The failure probability can be kept within the desired range of 9.5%<P<11% before each authentication test using the calibration scheme described in Fig. 11.

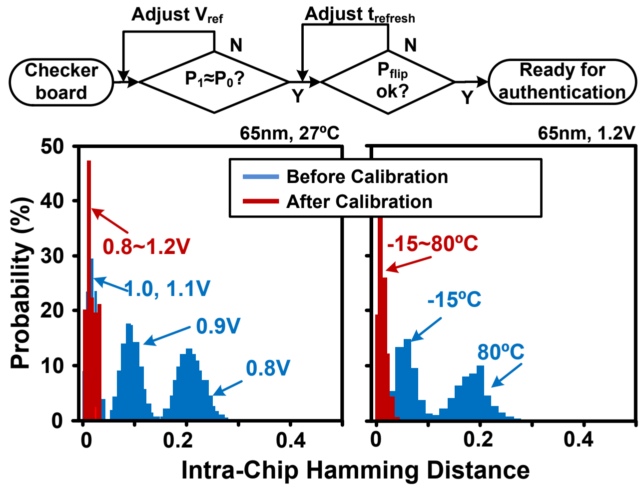


Fig. 11. Intra-chip Hamming distance distributions measured under different supply voltages and temperatures with and without a calibration routine.

VI. Conclusion

We have demonstrated a DRAM PUF utilizing the location of weak retention cells. The proposed PUF can generate more than 10^{32} CRPs from a 1Kbit DRAM array. To improve the consistency of the PUF response, we employed a repetitive write-back scheme along with bit-masking. Intra-chip and inter-chip Hamming distance distributions were measured from a 65nm chip under different supply voltages and temperatures. A calibration routine performed before each authentication operation has shown to effectively suppress voltage and temperature induced instabilities.

VII. Acknowledgement

This work was supported in part by the National Science Foundation under Grant CNS-1441639, and in part by the Semiconductor Research Corporation under Contract 2014-TS-2560.

References

[1] C. Herder, M. D. Yu, F. Koushanfar and S. Devadas, "Physical Unclonable Functions and Applications: A Tutorial," Proceedings of the IEEE, vol. 102, no. 8, pp. 1126-1141, Aug. 2014.

[2] M. Cortez, A. Dargar, S. Hamdioui and G. J. Schrijen, "Modeling SRAM start-up behavior for Physical Unclonable Functions," Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT), pp. 1-6, 2012.

[3] "PUF-Physical Unclonable Functions-Protecting Next-Generation Smart Card ICs with SRAM-based PUFs." NXP Semiconductor N.V. Feb. 2013. Available: http://www.nxp.com/documents/other/75017366.pdf

[4] S. K. Mathew, et al., "16.2 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] S. Rosenblatt, et al., "Field Tolerant Dynamic Intrinsic Chip ID Using 32 nm High-K/Metal Gate SOI Embedded DRAM," JSSC, vol. 47, no. 2, pp. 547-559, Feb. 2012.

[6] M. Bhargava, et al., "Comparison of bi-stable and delay-based Physical Unclonable Functions from measurements in 65nm bulk CMOS," CICC, pp. 1-4, 2012.

[7] C. Herder, et al., "Physical Unclonable Functions and Applications: A Tutorial," Proceedings of the IEEE, vol. 102, no. 8, pp. 1126-1141, Aug. 2014.

[8] K. C. Chun, et al., "A 667 MHz Logic-Compatible Embedded DRAM Featuring an Asymmetric 2T Gain Cell for High Speed On-Die Caches," JSSC, vol. 47, no. 10, pp. 2517-2526, Oct. 2012.