A Physically Unclonable Function with 0% BER Using Soft Oxide Breakdown in 40nm CMOS

Kai-Hsin Chuang*†, Erik Bury†, Robin Degraeve†, Ben Kaczer†, Dimitri Linten† and Ingrid Verbauwhede*
*imec-COSIC, KU Leuven, Belgium
Email: kai.hsin.chuang@imec.be, ingrid.verbauwhede@esat.kuleuven.be
†imec, Belgium

***Abstract*—A physically unclonable function (PUF) utilizing the randomness of soft oxide breakdown (BD) locations in MOSFETs is presented. The so-called *soft-BD PUF* features a self-limiting mechanism to generate one single soft-BD spot in a pair of MOSFETs; the subsequent BD location is used as the source of entropy to generate a highly stable "0" or "1" bit with an equal probability of 0.5. The soft-BD PUF comprising all the essential periphery circuits are fabricated in a 40nm CMOS process. Experiments show that the PUF has no instability in most of the operating conditions using the proposed readout scheme. The native bit error rate remains zero from V_{DD}=0.8V to 1.5V at room temperature and from -20°C to 120°C at nominal V_{DD}=0.9V. The throughput is shown to be at least 40 Mb/s and the PUF readout consumes only 51.8 fJ/bit. The randomness and uniqueness of the PUF are close to an ideal case, and no spatial correlation was observed.**

I. Introduction

A physically unclonable function (PUF) in modern silicon technologies is an essential circuit primitive for on-chip security and cryptographic applications. As one of the most common applications, the PUF-based *key generation* procedure [1] harvests the entropy from uncontrollable process variations of a PUF, as illustrated in Fig. 1. Typically, the data obtained from a PUF, namely *raw* PUF data, do not have full-entropy and good stability. In order to generate a cryptographic key which meets certain standards, the raw data need to be post-processed, including entropy extraction and error correction [1]. An error correcting code (ECC), or called *helper data*, is stored in a non-volatile memory (NVM) to assist the post-processing circuits during the key-generation procedure.

A. Stability of PUF

In recent PUF works [2]–[7], the stability is considered as a major concern, since it usually takes more resource to be optimized. On the other hand, a PUF with an excellent *native* stability, i.e. the stability of raw PUF data, is beneficial, especially for an application in which no NVM is accessible. Since the ECC cannot be stored, it necessitates a bit-error-rate (BER) of 0%, which is considered as the *ideal* stability.

Temporary majority voting (TMV) [1] and dark-bit masking [2] methods are widely used to improve the bit stability of PUFs, but both cases have their own limitations. The TMV can reduce BER, but is insufficient for elimination. Using a dark-bit mask to filter out the unstable PUF bits was reported having 0% BER from the unmasked PUF bits [2], it however requires more testing time and/or additional circuitry to identify the unstable bits. Moreover, once a NVM is required to store the masking information locally, there is not much distinction between the traditional error correction methods.

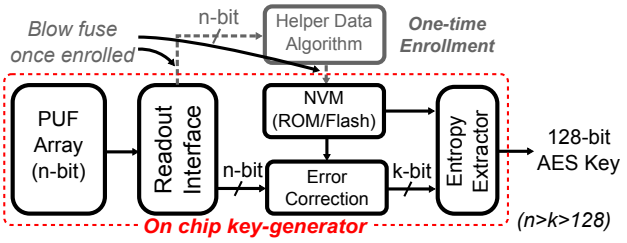

Fig. 1. An example block diagram of a PUF-based AES-128 key generator.

B. Approaches towards ideal stability

Recently, two major approaches were proposed to achieve an ideal native bit stability. The first one is to exploit the variability of emerging memory devices, such as the resistive-RAM (RRAM) based PUFs [8]. The bit stability in such case is usually good but is out of scope for this work, since it is linked to the specific device physics. The other approach is based on *aging* effects that can be electrically induced, such as biased-temperature instability (BTI) [3], hot-carrier injection (HCI) [4] and oxide breakdown (BD) [5]–[7]. In this paper, we will focus on the stability aspects of our proposed PUF using *soft* oxide breakdown (SBD), namely soft-BD PUF, with detailed experiments on the chips fabricated in 40nm.

II. Circuit design and operation

The unit cell and array of the soft-BD PUF (Fig. 2), was first introduced in [6], without any periphery circuit. The three transistor (3T) unit cell consists of two minimum-sized NMOS transistors, which will be stressed to generate random soft-breakdown, and one PMOS transistor serves as the word line (WL) selector. The proposed PUF circuit in Fig. 2 is designed as a 32-by-32 array, with periphery circuits including sense amplifiers (SA) and other control logic. The operating concept of the PUF cell and array will be first introduced; followed by the readout scheme utilizing the proposed reference-free sense-amplifiers.

A. PUF cell and array

As illustrated in Fig. 2, a soft-breakdown path will occur in one of the two NMOS transistors, and the detailed procedure towarding this result can be found in [6]. Briefly, once a high

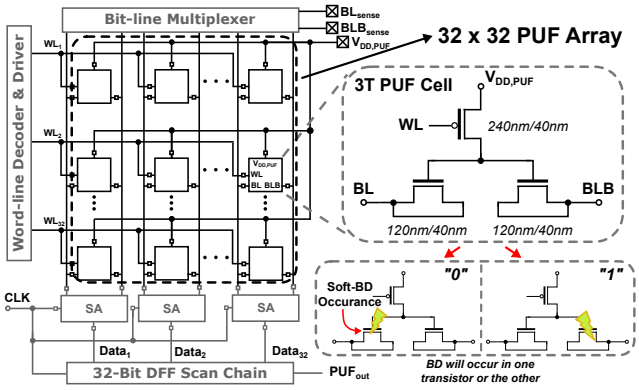

Fig. 2. The schematic of the PUF array including the periphery circuits and the 3T PUF cell.

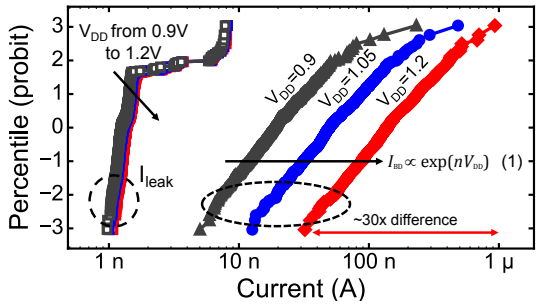

Fig. 3. The experimental statistics of the two current components from the soft-BD PUF cells. The BD current, which flows through the soft-BD spot is widely distributed and has an exponential voltage dependence, as described in (1), in which n is a scaling parameter. The leakage current, which flows through the unbroken gate oxides, is also widely distributed but has no strong voltage dependency. A more detailed explanation can be found in [6].

voltage stress is applied to $V_{\mathrm{DD,\,PUF}}$ and a WL is enabled, a dielectric breakdown will eventually occur in a NMOS transistor, and the stress voltage and current will be both limited by the PMOS transistor as soon as this event occurs. The subsequent voltage stress will be too small to trigger another breakdown, and the breakdown path will not be able to grow towards the *"hard"* breakdown (HBD).

The main reason of aiming soft breakdown is to have less *visibility*, making it more secure against *invasive attacks*. Note that the HBD spots in nanoscale anti-fuse devices are not detectable by the *scanning electron microscope* (SEM) [9], so as the PUF in [7]. It can be, however, detected by the *transmission electron microscope* (TEM), as discussed in [10], while a SBD spot is much less obvious in this case. Even though using a powerful tool like TEM to attack a PUF is unrealistic, it is still reasonable to consider using SBD, as a precaution for the rapid advancing of attacking techniques.

The PUF cells are organized as a typical array with shared word lines and bit lines (BLs), the layout of the PUF cell and array are shown in Fig. 5. Note that for experiment and modeling purposes, the current flows through individual BLs can be directly multiplexed to sensing pads for external DC measurements (similar to [6]). The locations of BD spots obtained based on these DC measurements, exactly matches the PUF data generated by the SAs as expected (not shown).

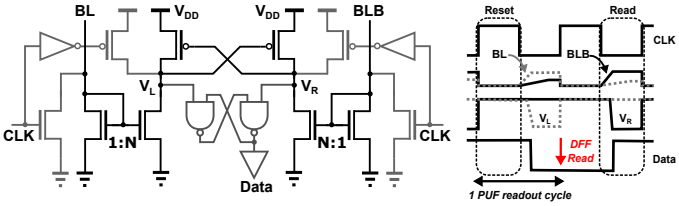

Fig. 4. The schematic of the proposed reference-free sense-amplifier and the corresponding timing diagram.

B. Reference-free sense-amplifier

The current flow through the soft breakdown spot exhibits approximately an exponential voltage dependence, as discussed in [6]. It leads to very low BD currents at low V_{DD}, in addition to a strong variation between devices (from 5nA to 200nA at V_{DD}=0.9V), as shown in Fig. 3. Consequently, the sense-amplifier has to sense the current in the nA range at nominal V_{DD} and below, which is not a typical current sensing range of a SA used for SRAMs. Note that in [7], a single-ended SA with reference voltage works well on sensing currents from the ruptured spots with equivalent resistance $< 100k\Omega$. This technique, however, cannot be directly adapted to fit the SBD spots, since it is rather difficult to define an optimized reference voltage or current.

In order to solve this issue, a dedicated reference-free SA is designed, as shown in Fig. 4, consisting of a current-mirror input stage and a cross-coupled pair second stage (in darker lines). As illustrated by the timing diagram, a readout cycle starts with resetting the SA when *CLK* is set to 1. Current sensing starts when *CLK* is set to 0, the soft-BD current from either *BL* or *BLB* will be amplified to cause a fast discharging on V_{L} or V_{R}. The second stage will be latched once sensing a certain difference between V_{L} and V_{R}. The state (*Data*) of the final SR-latch will be updated accordingly, which is ready for DFF registering at the next clock rising edge.

Each bit-line pair consists of 32 PUF cells is connected to an individual sense amplifier, in order to avoid the additional series resistance from a multiplexer, which degrades the sensing resolution. Having multiple SAs increases the overall throughput as well, since multiple bits can be readout in parallel. Note that only a small amount of area overhead is introduced by placing multiple SAs, as shown in Fig. 5.

III. Experiment Description and Results

For experimental characterization of the soft-BD PUF, the test chips are packaged and measured on a PCB; a packaged chip is shown in Fig. 5. An FPGA is being used to generate the control signals; it also receives and synchronizes the digital output signals from the PUF chips.

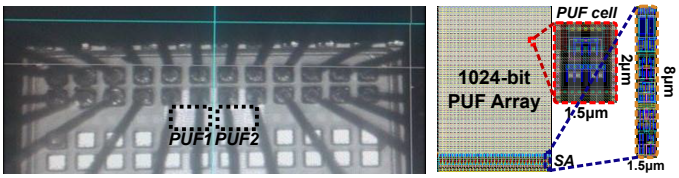

Fig. 5. Layout of the 1024-bit PUF array with sub-circuits and the die photo.

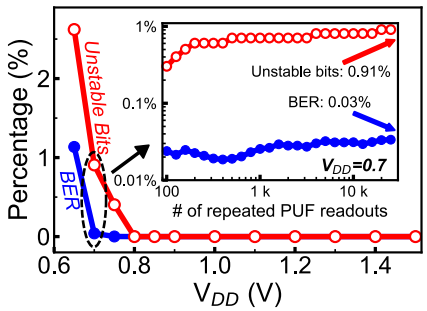

Fig. 6. The percentage of unstable bits and BER v.s. V_{DD} and v.s. repeating readout cycles at V_{DD}=0.7V (inset).

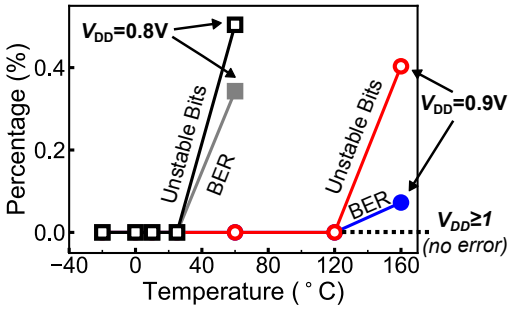

Fig. 7. The percentage of unstable bits and BER under different temperature, operating at V_{DD}=0.8V and 0.9V.

Using this setup, the maximum measured throughput is 40Mb/s (not the actual limit) for all tested conditions and the average energy consumption per PUF bit is 51.8fJ. The stability and the other properties are examined as follows.

A. Bit stability

As the voltage dependence is a major concern for the soft-BD PUF, we first examine the stability at different V_{DD}, as shown in Fig. 6 at the room temperature. The PUF is well functioning down to V_{DD}=0.7V; the bit-error-rate (BER) is below 0.1% at 0.7V and is 0% (ideal) for $V_{DD}\geq$0.8V. Note that the error-free region well covers the $\pm$10% of the nominal V_{DD} (0.9V for this 40nm technology).

The stability at reduced and elevated temperature has been tested using a temperature chamber. Note that an on-chip poly-heater, like the one characterized in [6], was used as an alternative heat source for 60 °C and above. The temperature inside the chamber cannot be further increased due to the temperature limit of the connecting cables.

As shown in Fig. 7, both the ratio of unstable bits and BER remain 0% at the room temperature and below. When operating at V_{DD}=0.9V, this excellent stability holds until 120°C. Once the V_{DD} is lowered to 0.8V, some of the PUF bits become unstable at 60°C. This result is in contrast to the temperature dependence of the current ratio observed in [6], where a better stability is expected at an elevated temperature. This effect can be attributed to the performance degradation of the SAs at higher temperature, and is verified by simulation, as shown in Fig. 8. The difference betweenV_L and $V_R$becomes smaller when the circuit is heated up, which makes it more sensitive to noise, and hence the output becomes less stable.

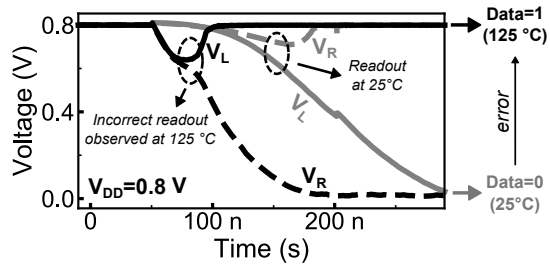

Fig. 8. Simulated voltage of SA nodes V_L and V_R (see Fig. 4) at V_{DD}=0.8V and 25°C/125°C. Here shows one of the error cycles observed at 125 °C.

B. Randomness and Uniqueness

The quality of PUF data is typically checked by three indices: randomness (bias), uniqueness (hamming distance) and spatial correlation (auto-correlation function). The normalized hamming weight (number of "1"s) distribution of 128-bit words from 20 measured PUF arrays shows no bias, as shown in Fig. 9 (a). The auto-correlation function (ACF) of PUF data, as the example shown in Fig. 9 (b), is also computed and plotted in Fig. 10. All the data sequences have passed the requirement of the auto-correlation test specified in AIS31 (T5) [11], which checks the bitwise correlation. As the result, we conclude that the PUF bits are uncorrelated, and hence no spatial correlation within PUF chips.

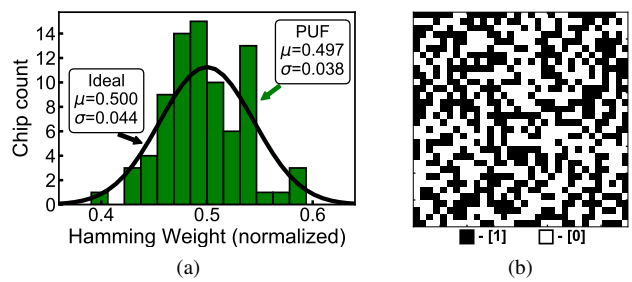

Fig. 9. (a) Normalized hamming weight distribution of the 128-bit PUF words generated from 20 PUF arrays and (b) an example PUF data.

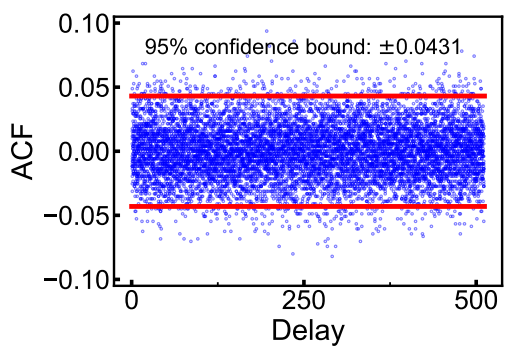

Fig. 10. The auto-correlation function (ACF) of the PUF arrays with the indication of 95% confidence bound, which shows no observable spatial correlation.

The resulting hamming distance from the 128-bit words are plotted in Fig. 11, which is almost identical to the ideal values. Moreover, the identifiability, which is defined as the ratio of the inter and intra hamming distance, is infinite at the nominal operating condition, as a benefit from the ideal stability.

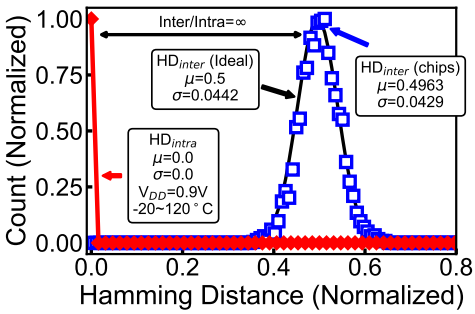

Fig. 11. The inter and intra hamming distance resulting from the 128-bit words, showing an uniqueness nearly indistinguishable from the ideal PUF.

IV. Discussion and Comparison

The comparison with prior works is shown in Table I. The proposed soft-BD PUF shows no downside among all the performance indices. The minimum V_{DD} to obtain 0% BER in the experiments is higher than [7], but the energy consumption is much lower. The relatively low resistance of the "hard" breakdown spot results in a higher current level comparing to the soft-BD current, especially at low V_{DD}. Consequently, the proposed soft-BD PUF is less robust against the extreme operating conditions comparing to [7], but it keeps several advantages, including *good energy efficiency* and *less visibility*. Moreover, we have noticed that the degrading stability at higher temperature is mainly originated from the custom designed sense-amplifier, i.e. not the PUF cell itself. In other words, the SA design can be further improved to obtain better temperature stability.

The SA-PUF with hot-carrier injection based stabilization technique [4] shows BER=0% as well, but it necessitates a relatively long burn-in period (25s) to achieve this target. The uniqueness is also worse as the average $\mathrm{HD}_{\mathrm{inter}}$ is apart from 0.5, which may be attributed to the additional circuitry to enables this HCI burn-in procedure.

For the SRAM-like hybrid PUF [12], it utilize burn-in, TMV and dark-bit masking to reduce the instability while keeping an excellent energy efficiency, but it is insufficient to achieve a BER of 0%. The error correcting logic and NVM is therefore not removable for this case.

Besides all the advantages of the soft-BD PUF, it should be noted that the additional procedure to generate the oxide breakdown spots also occupies resources. It requires either an additional pin or an on-chip high-voltage generator and more testing time. Nevertheless, once considering the trade-off between an embedded NVM for error correction, the proposed soft-BD PUF stays competitive even in terms of cost.

V. Conclusion

A PUF using the randomness from the *soft* oxide breakdown mechanism in MOSFETs has been demonstrated. The reference-free sense amplifier can distinguish the current difference in the nA range, resulting in an excellent bit stability over a wide range of voltage and temperature variations. The soft-BD PUF also shows good randomness and uniqueness; it also consumes less energy per bit comparing to the designs using hard oxide breakdown. The experimental results prove that the soft breakdown mechanism is suitable for high quality PUF implementations, in particular for the on-chip stable key generation without error correction. The less-visible nature of the soft-BD spots also makes it a more viable candidate for the applications with higher security requirements.

TABLE I
Comparison Summary with Prior PUF Work

	[4]	[5]	[7]	[12]	**This work**
Design	SA	Anti-fuse	Anti-fuse	Hybrid cell	**Soft-BD**
Technology	65nm	65nm	55nm	14nm	**40nm**
Stabilizing method	hot-carrier injection	native	native	delay-hardened	**native**
BER (%)	0	0	0	1.46	**0**
V_{DD} (V)	0.8–1.2	1.0–1.2	0.81–1.32	0.55–0.75	**0.9–1.5**
Temp. (°C)	-20–85	0–85	-40–150	25–110	**-20–120**
$\mathrm{HD}_{\mathrm{inter}}$	0.468	0.501	0.50	0.486	**0.496**
Energy/bit	N/A	340 fJ	5200 fJ *	4 fJ	**51.8 fJ**

* Including all peripheral blocks, in which the high-V source and BIST circuit are not implemented in this work. It was not clearly stated if these blocks are *active* during readout phase, hence it cannot be concluded whether this is a fair comparison or not.

Acknowledgment

This work was supported in part by the Research Council KU Leuven: C16/15/058. In addition, this work is supported in part by Cathedral ERC Advanced Grant 695305.

References

[1] J. Delvaux, D. Gu, D. Schellekens, and I. Verbauwhede, "Helper data algorithms for puf-based key generation: Overview and analysis," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, pp. 889–902, 2015.

[2] M. Liu, C. Zhou, Q. Tang *et al.*, "A data remanence based approach to generate 100% stable keys from an sram physical unclonable function," in *2017 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*, July, pp. 1–6.

[3] R. Maes and V. van der Leest, "Countering the effects of silicon aging on sram pufs," in *Hardware-Oriented Security and Trust (HOST), 2014*. IEEE, pp. 148–153.

[4] M. Bhargava and K. Mai, "A high reliability puf using hot carrier injection based response reinforcement," in *International Workshop on Cryptographic Hardware and Embedded Systems*. Springer, 2013, pp. 90–106.

[5] N. Liu, S. Hanson, D. Sylvester, and D. Blaauw, "Oxid: On-chip one-time random id generation using oxide breakdown," in *VLSI Circuits (VLSIC), 2010 IEEE Symposium on*. IEEE, pp. 231–232.

[6] K. H. Chuang, E. Bury, R. Degraeve *et al.*, "Physically unclonable function using cmos breakdown position," in *2017 IEEE International Reliability Physics Symposium (IRPS)*, April, pp. 4C–1.1–4C–1.7.

[7] M. Y. Wu, T. H. Yang, L. C. Chen *et al.*, "A puf scheme using competing oxide rupture with bit error rate approaching zero," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, pp. 130–132.

[8] R. Liu, H. Wu, Y. Pang *et al.*, "Experimental characterization of physical unclonable function based on 1 kb resistive random access memory arrays," *IEEE Electron Device Letters*, vol. 36, pp. 1380–1383, Dec 2015.

[9] N. Chen. (2016) The benefits of antifuse otp. [Online]. Available: http://semiengineering.com/the-benefits-of-antifuse-otp/

[10] K. L. Pey, C. H. Tung, M. K. Radhakrishnan *et al.*, "Dielectric breakdown induced epitaxy in ultrathin gate oxide - a reliability concern," in *Digest. International Electron Devices Meeting,*, Dec 2002, pp. 163–166.

[11] W. Killmann and W. Schindler, "A proposal for: Functionality classes for random number generators," 2011.

[12] S. Satpathy, S. K. Mathew, V. Suresh *et al.*, "A 4-fj/b delay-hardened physically unclonable function circuit with selective bit destabilization in 14-nm trigate cmos," *IEEE Journal of Solid-State Circuits*, vol. 52, pp. 940–949, April 2017.