Comparison of SRAM and FF PUF in 65nm technology

Mathias Claes[1], Vincent van der Leest[2], and An Braeken[1]

[1] Industrial Sciences and Technology, Erasmus University College, Nijverheidskaai 170, 1070 Brussels, Belgium
[2] Intrinsic-ID, High Tech Campus 9, Eindhoven, The Netherlands

Abstract. Hardware security is an essential tool in the prevention of cloning, theft of service and tampering. This security is often based on cryptographic primitives, which use a key that is securely stored somewhere in the hardware. The strength of the security is therefore dependent upon the effort required from an attacker to compromise this key. Since the tools used to carry out attacks on hardware have increased significantly over the years, the protection provided by simply storing a key in memory has decreased to a minimum. In order to protect devices against attacks on their keys, Hardware Intrinsic Security (HIS) can be used. One of the best known types of HIS primitives are Physically Unclonable Functions (PUFs). PUFs are primitives that extract secrets from physical characteristics of integrated circuits (ICs) and can be used, amongst others, in secure key storage implementations. This paper describes the results of our study on two important types of intrinsic PUFs, based on SRAM and D flip-flops. Both memory types present a specific start-up pattern (when powered up), which can be used as a PUF. For secure practical applications, a PUF should possess enough reliability for a single device and enough randomness between different devices. In this paper, a general test framework is proposed for measuring this reliability and randomness of both PUF types. Based on this framework, tests have been performed on PUFs in 65nm ICs and results are presented and compared between PUF types. From these results it can be concluded that SRAMs are slightly outperforming D flip-flop memories when it comes to usage for PUF implementations.

Keywords: Physical Unclonable Function, Hardware Security, SRAM, D flip-flop

1 Introduction

One of the main assumptions in cryptography is that participants possess a secret key, in order to differentiate themselves from potential attackers. As a consequence, encrypted messages can only be read by the person knowing the key. However it is not easy to store a secret key. Many devices operate in environments where physical attacks can be applied. By opening the unit, an attacker is able to "easily" read the digital key.

To guarantee the secrecy of keys, even if an attackers has physical access to a system, a promising technique called Physical Unclonable Function (PUF), has been introduced by Pappu in 2001[1]. PUFs are based on the internal randomness present in physical systems. The basic idea is that the keys are not stored in the system, but can be dynamically generated as the response on a physical stimulus, called the challenge. Even if an attacker knows all the details of the system, it is impossible to generate the same key or to clone the device. When an attacker tries to intercept the key, he will destroy with high probability the PUF during the physical attack. Another advantage of using a PUF is that additional physical security is achieved without any special manufacturing steps. Moreover, since the process variations are beyond the control of manufacturers, no two systems are equal.

1.1 Related work

Many different PUF instances are known today. A large class of PUFs consists of the delay based PUFs, like the ring oscillator PUF described by Gassend et al. [2] in 2002 and the Arbiter PUF described by Lee et al. in 2004 [3]. In 2007 SRAM based PUFs were introduced by Guajardo et al. [4], followed by Butterfly PUFs introduced in 2008 by Kumar et al. [5], and finally D flip-flop PUFs in 2008 by Maes et al. [6]. Implementations exist for dedicated Integrated Circuit (ICs), programmable logic devices such as Field Programmable Gate Arrays (FPGAs), and also for programmable ICs such as microcontrollers. Besides these examples, there are also a number of other constructions, some of which are purely theoretical.

Since 2002, the concept of PUF has received lots of interest in literature, especially with respect to aspects related to design and applications. We refer to [7] for an overview of the latest evolutions in these areas.

1.2 Our contribution

In this paper, we focus on Static Random Access Memory (SRAM) and Data-flip-flop (FF) PUFs implemented on application-specific integrated circuits (ASICs). D-FF, or shortly FF, PUFs have a real security advantage over SRAM PUFs against invasive attacks such as probing attacks, since they can be distributed across an integrated circuit. It is much harder for an attacker to locate them. SRAM and FF PUFs consist of standard Complementary Metal Oxide Semiconductor (CMOS) components, and thus do not require an extra fabrication process. The choice for ASIC implementations follows from the fact that these are more secure than implementations in reconfigurable logic.

In order to be able to exploit these PUFs for practical purposes, they should possess high reliability and uniqueness/randomness. In order to test both properties, we present a general framework. For measuring the reliability, we describe the behavior of our devices under varying environmental conditions. Measurements are taken from 20 different devices, fabricated in 65nm technology. An estimate of the uniqueness of the device is obtained by two different tests.

The physical strength of the PUF was mostly only theoretically proven or very limited tested, i.e. without evaluation under external stress conditions and over time. The most complete test framework on PUFs has been presented in [8]. These tests were performed on a 90nm SRAM PUF. In this paper, we extend the list of tests and evaluate them on SRAM and FF PUFs in a 65nm technology. It is the first time that both types of PUFs are compared using the same test setting on the same chip.

1.3 Organization

In Section 2 we provide a description of the test framework. A brief system description is given, followed by the testing strategy. In Section 3, we show the results of the different reliability tests and in Section 4 the uniqueness tests. We end in Section 5 with conclusions and future work.

2 Test framework

Biometric measurements and PUFs share the same property, exhibiting noise. As a result, PUF responses and biometric measurements are not fully uniformly distributed, which is undesirable for security applications. In order to use PUF responses in cryptographic applications like secure key storage mechanisms as described in [9], processing by means of a fuzzy extractor or helper data algorithm needs to be applied. We do not address these algorithms in this paper as this can be achieved using well-known methods based on secure extractors [10, 9]. It is clear that smaller noise percentages in the PUF responses allow the use of more efficient error correcting codes requiring less redundant information.

The strength of a PUF is expressed in two basic properties, reliability and uniqueness. We explain both concepts more into detail, together with the corresponding tests that give an estimate of their strength. But first, we shortly give the system description.

2.1 System description

For each IC, we evaluate two commercial SRAM memories and one FF memory that are integrated in 65nm CMOS technology. The so called PUFPUF ICs were designed by IMEC The Netherlands and produced on a Multi-Project-Wafer (MPW) at TSMC. The commercial SRAM memories, NXP (nxp_mem1kx64) and TSMC (TS1N65LPA1024X64M8), are organized as 1024×64 bits, while the FF memory is organized as 256×64 bits. Consequently, we examine two types of memory based PUFs without fuzzy extractor and other processing steps.

2.2 Reliability tests

The first important property for PUFs to be studied, is reliability. It measures the consistency or stability when the environment (such as ambient temperature,

supply voltage, etc.) varies. Environmental changes will contribute to temporary or permanent variations in the desired properties. These variations are determined by the main parameters of transistors such as threshold voltage, leakage current, delay, etc. The effect of these variations should be minimized as much as possible because of two reasons. The device should be in the first place resistant since it can be naturally subject to environmental changes. On the other hand, it should not be possible for an attacker to leak information from the device by simply changing, for instance, the temperature.

We have identified six different tests for evaluating the reliability of the PUF. First of all, the behavior under varying temperature should be studied. As chips operate at higher frequencies, the temperature of the die rises. Higher operating temperatures degrade the performance of transistors and inter-connections. High temperatures and temperature gradients can cause delays to change, which may cause transient or permanent failure.

Secondly, the effect of voltage variation is studied. It is well known that a decrease in supply will slow down a circuit. Moreover, the performance loss is not linear, which affects different parts of the circuit and therefore the reliability of the PUF.

To see the effect of power dips on the initial state of the PUFs, the data retention test was carried out. If a dip in the power voltage occurs, a threshold should be set, so the state of the PUFs are not influenced.

The fourth test is a voltage ramp-up test, which is performed at different temperatures. When different ramp-ups are applied, the stability and the present randomness could change.

In the fifth test, called the voltage dip test, the required dip time for properly resetting the memories is studied. It is well known that data remanence gets steadily longer at lower temperatures. For instance static RAM contents below -20° C can persist from seconds to minutes after the power supply is removed. Therefore this test is also performed at different temperatures.

Finally, the last test for measuring the reliability of the PUF is called the ageing test. Silicon will gradually degrade over time, which will have repercussions on the PUF. Several mechanisms stand out: time-dependent dielectric breakdown (TDDB), hot carriers, negative bias temperature instability (NTBI), electro migration, stress migration and soft errors. Some of these failure mechanisms target transistors, while others come from interconnect.

2.3 Uniqueness tests

The other important security parameter for PUFs is uniqueness. This entails the following two aspects:

- Each device should be unique, meaning that the probability for two devices having a PUF response close to each other is negligible.
- Each PUF response is random and unpredictable, meaning that bits in a PUF response can only be predicted with negligible probability.

Two important measurement distances are respectively related with these two property.

- Intra class distance (within-class distribution) is the Hamming distance (HD) between the responses from the same challenge of one PUF instance.
- Inter class distance (between class distribution) is the HD between the responses from the same challenge of two different PUF instances.

In our measurements, we mainly use the fractional Hamming distance (FHD), instead of the HD, which is the HD divided by the total length. As μ_{intra} represents the average noise of one PUF, it is clear that μ_{intra} should be as small as possible. On the other hand, μ_{inter} measures the average distinguishability (how well are we able to distinguish two different devices) of two systems based on their PUF responses. Consequently, μ_{inter} should be ideally equal to 50%.

We have distinguished two different tests for evaluating the uniqueness of PUFs. The first test is called the between-class uniqueness test in which μ_{inter} values are measured, which give a good first indication of the randomness of our PUFs.

The second test is an entropy test which estimates the present entropy in the PUF responses. Although μ_{inter} is a good indication of the uniqueness of the response, it can not be used to assess the true independent entropy. In order to find the independent entropy, we will check the ability to compress the response strings and calculate the min-entropy.

3 Reliability test results

We here describe the test set-up, together with the observations and the conclusions that can be drawn from the six reliability tests, as described above. Since the two SRAM memories, NXP and TSMC, give approximately the same results, we do not show the exact results for both but instead restrict to one of them. We refer to [11] for a detailed description of both.

3.1 Temperature test

In order to test the effect of varying the temperature 20 PUFPUF ICs were placed in a test set-up. These ICs were powered up repeatedly and after each power up the contents of the memories were read (and stored into a file). During the test, each IC was read 20 times at 5 different ambient temperatures (-40°C, -10°C, 20°C, 50°C, 80°C), resulting in 100 files per memory per device. One measurement at an ambient temperature of 20°C is used as enrollment, to which all other measurements are compared. Comparison between measurements is based on the FHD between the start-up patterns of the chip.

The resulting FHD values of the NXP SRAM are plotted in Fig. 1 on the top side. The number of measurements per device is set to the horizontal axis, while the vertical axis present the FHD between start-up patterns and enrollment of the chip. At the top of each graph, the current condition (in this case: the

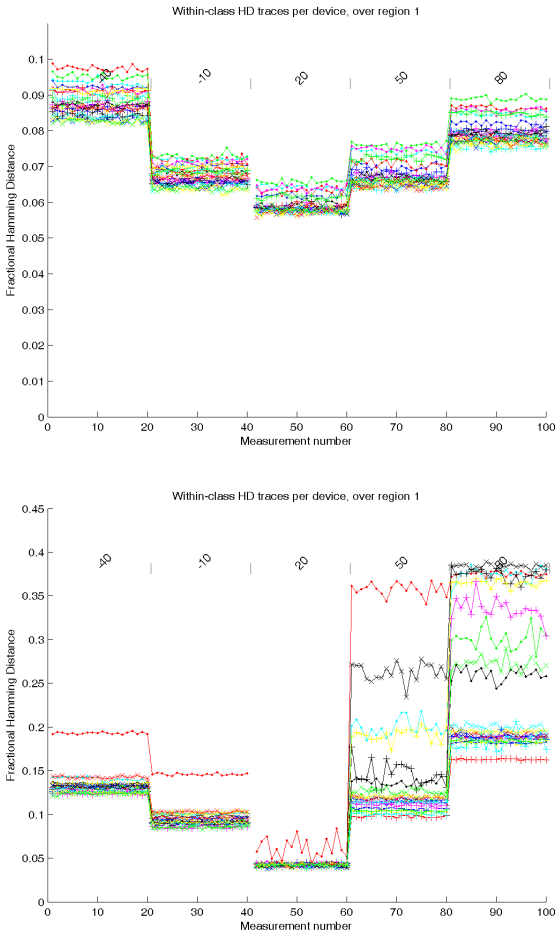

Fig. 1. FHD vs. temperature for 20 SRAM memories (top) and FF memories (bottom)

different temperatures of the measurements) is specified. Various ICs are marked with colored lines. A similar representation for all the other test results is used in this paper.

From this graph we deduce that the noise levels steadily remain below 10% (in comparison to enrollment at 20°C), no matter at what temperature the measurements are taken. This means that the reconstructed values are extremely stable and consequently allow very efficient error correction codes in the fuzzy extractor.

The FHD of the FF memories are shown in Fig. 1 on the bottom side. Variation in temperature shows a maximum deviation of 0.4 measured for the FFs, which will require additional processing in the fuzzy extractor for these FF mem-

ories to be usable as PUFs. Hence the fuzzy extractor used for the FFs will be more complex (and therefore require more hardware resources) than the one used for SRAM. Furthermore, results of FF memories vary considerably from chip to chip. Consequently, we can conclude that SRAM memories have a better performance in this test than the FF memories, since the FHD in regard to enrollment is low.

3.2 Voltage variation test

In order to find out the consistency of the start-up values of the memories under slight variations of the power voltages, 10 PUFPUF ICs were placed in a test set-up. These ICs were powered up repeatedly and after each power up the contents of the SRAM and FF memories were read (and stored into a file). During the test, each memory was read 10 times at 5 different core voltages (90% of Vdd, 95%, 100%, 105% and 110%), resulting in 50 measurement files per memory per device. One measurement (at 100% of Vdd) per memory per IC is used as enrollment, to which all other measurements are compared. Comparison between measurements is based on FHD between start-up patterns of the memories.

From this test, we conclude that there are no remarkable variations in FHD between the different core voltages. The FHDs are approximately constant over all supply voltages. A maximum deviation of 0.07 is measured for both types of memories, which is very good.

3.3 Data retention test

To investigate the effect of power dips on the initial state of the PUFs, the Hamming Weight HW (number of bits in a string with value 1) is used, more specifically the fractional FHW (HW divided by the string length).

At the beginning of the test, the memories of the ICs were filled with 0xFF (FHW = 1). Next, the supply was lowered to a certain percentage (from 100% to 10% in steps of 10%) of Vdd for 1 second. Then, the supply was set to Vdd again and the contents of the memories are read out. For each supply value, this test was performed 10 times. During the test, the FHW of the measurement of three PUFPUF ICs were monitored. If the FHW drops below 1, the memories lose their content.

When the supply voltage lowered to 20% of Vdd, some bits flip to zero. However, the FHW still remains approximately 1. At 10% of Vdd, we measured a FHW of 0.2 for the FFs and 0.5 for the SRAMs. Consequently, the results from this test are very good since the voltage must be very low (20% of Vdd) in order for the memories to lose their values. Together with the results of the voltage variation test, we conclude that the devices are very resistant to variations in supply voltage.

3.4 Voltage ramp-up test

The test set-up consisted of 10 PUFPUFs ICs. These ICs were powered up repeatedly at 8 different ramp-up times. After each power up, the contents of

the memories were read (and stored into a file), resulting in 80 files per memory per device. One measurement with a ramp-up time of 1μs at 20°C is used as enrollment. Comparison between the measurements is based on FHD between the start-up patterns of the memories.

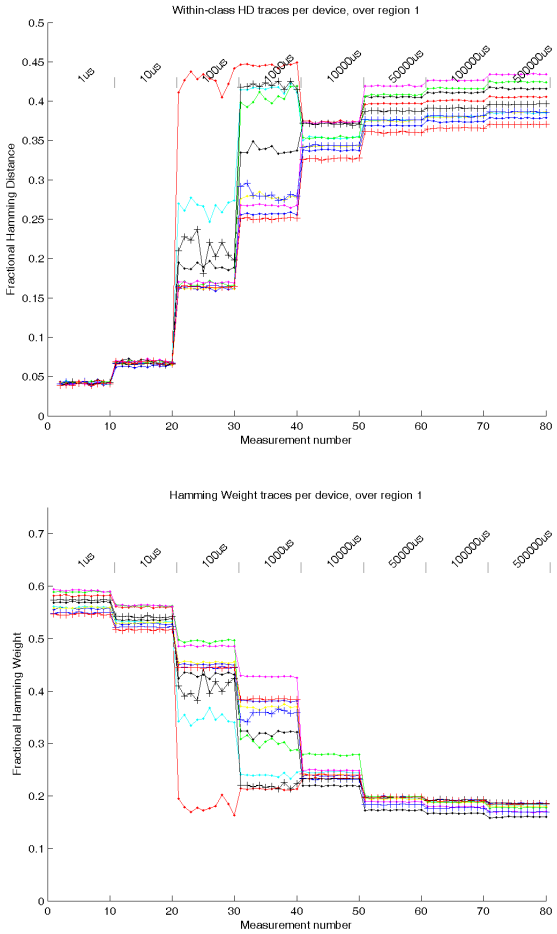

Fig. 2. FHD (top) and FHW (bottom) vs. ramp-up time for 10 FF memories at 20°C

When the ramp-up time of the supply becomes longer, the FHD with regard to enrollment becomes larger. In Fig. 2 (top) it can be seen that the response of the FF memories change rapidly (unstable) when the ramp-up time becomes longer (10μs to 100μs). At 500ms the FHD of the SRAMs is less than 0.2, as can be seen in Fig. 3 (top), while the FHD of the FF is almost 0.45 (Fig. 2 top). When we look at the FHW of the FF in Fig. 2 (bottom), we see a strong biasing

towards zero at slow ramp-ups. This is not the case for SRAM memories, as can be seen in Fig. 3 (bottom).

Therefore, a ramp-up time of the supply should always be kept sufficiently short in any set-up. When the ramp-up time is kept below 100μs, this will not cause problems (FHD $<$ 0.1). Based on this observation, a ramp-up time of 25μs at 20°C is used as enrollment for the voltage ramp-up test at different temperatures (-40°C to +80°C).

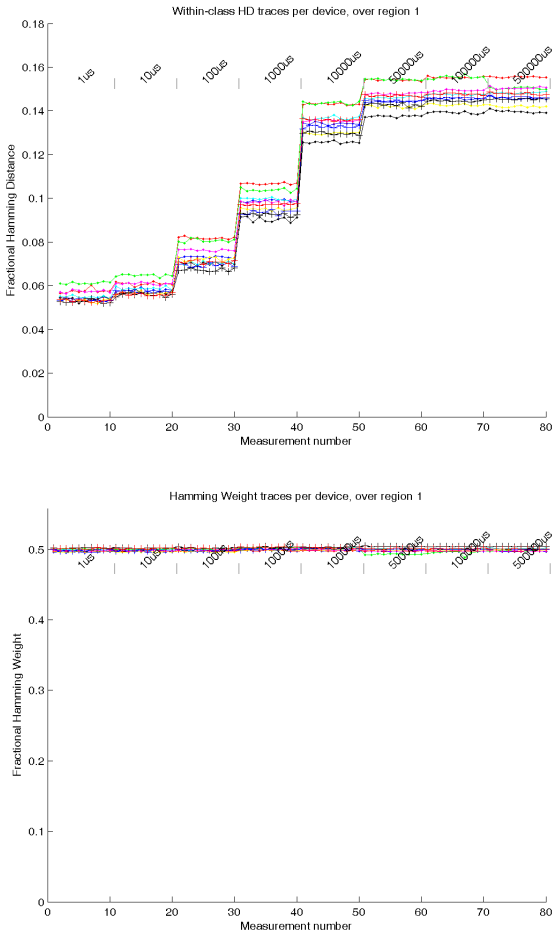

Fig. 3. FHD (top) and FHW (bottom) vs. ramp-up for 10 SRAM memories at 20°C

The FHD graphs at different temperatures (like -40°C, represented in Fig. 4) show that the SRAM memories do not experience a significant impact when combing temperature and ramp-up variation. The FF memories behave normal

over ramp-up times at low temperatures, but change rapidly at high temperatures. If the ramp-up time at low temperatures is less than 100μs, the distance is kept below 15%. Less steep ramp-ups at low temperature seem to be closer to enrollment, due to decrease in propagation delay with operating temperature.

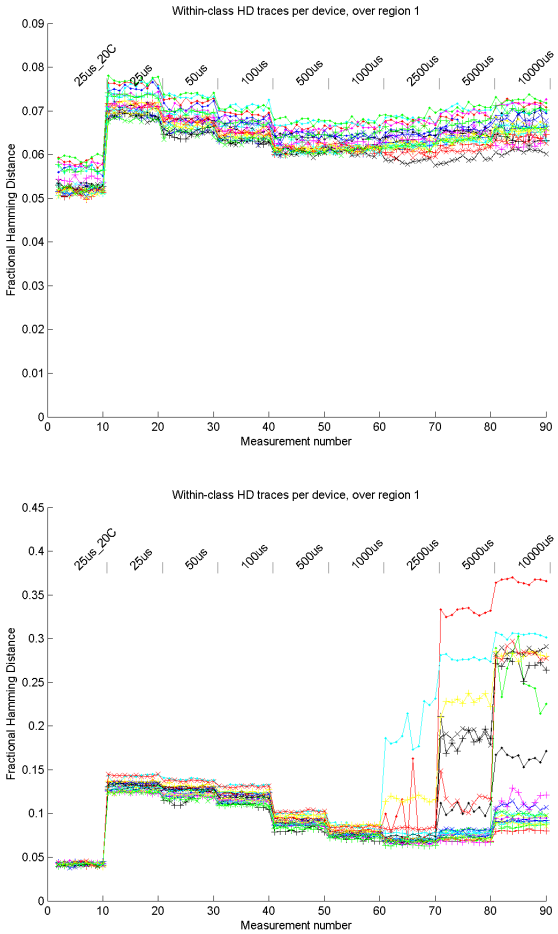

Fig. 4. FHD vs. ramp-up time for 20 SRAM (top) and FF (bottom) memories at -40°C

3.5 Voltage dip test

Remanence is tested by placing PUFPUF ICs in a test set-up, which is suitable for asserting a dip on the core voltage of the IC. At the beginning of the test the memories of 10 ICs were filled with 0xFF (all 1s). Then the ICs were powered

down for a certain amount of time (voltage dip). After the ICs were powered up, the contents of the memories were read. As data remanence gets steadily longer at low temperatures, the dip test is performed at 20°C and -40°C. Each memory was read 10 times at 8 different dip times and compared to enrollment where dip time of 1s was used, which is long enough for a proper reset of all the memories.

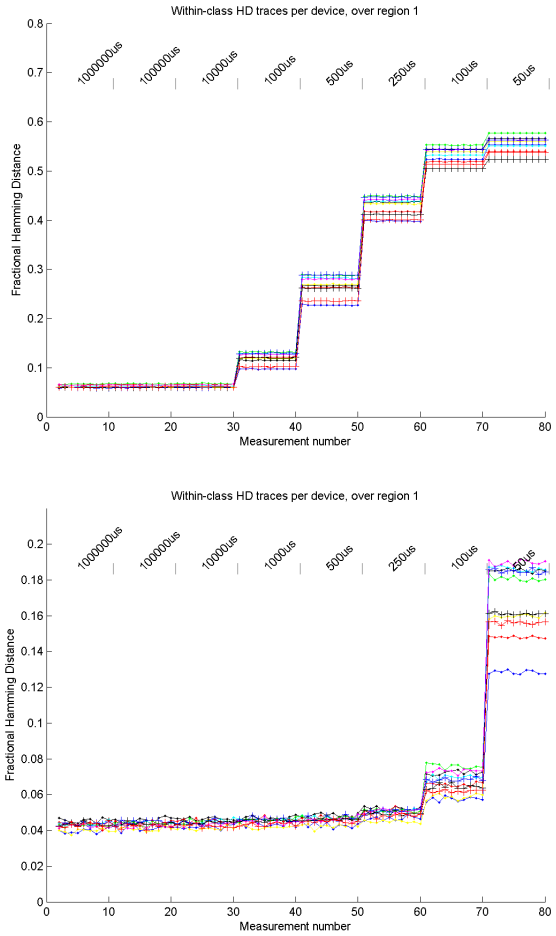

Fig. 5. FHD vs. dip time for 10 SRAM memories (top) and FF memories (bottom)

From Fig. 5 we conclude that when the dip time of the supply voltage becomes shorter at 20°C, the FHD with regard to enrollment becomes larger. In order to have a FHD below 0.15, a proper reset should take at least 1ms for the SRAM memories and only 100μs for the FF memory. At -40°C, as shown in Fig. 6,

the reset should take at least 1s for the SRAM memories and 50ms for the FF memory.

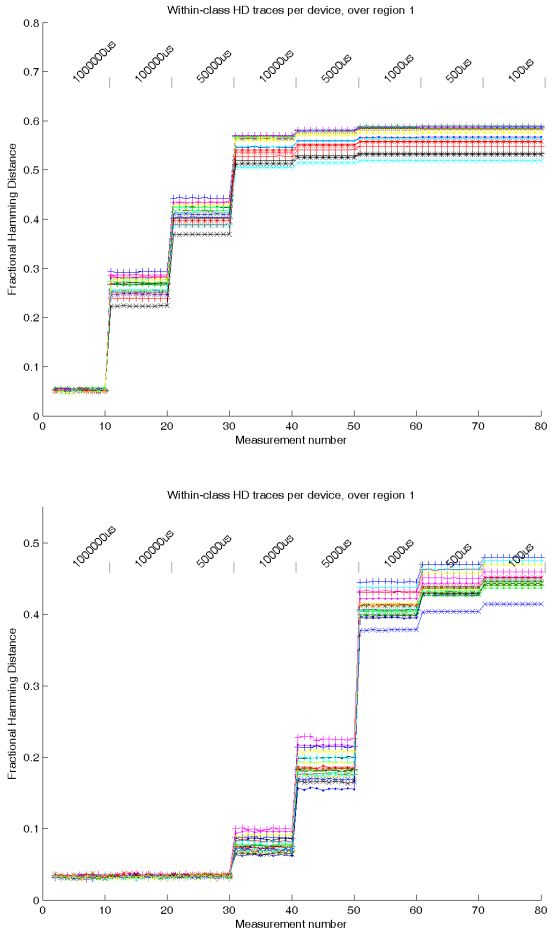

Fig. 6. FHD vs. dip for 10 SRAM memories (top) and FF memories (bottom) at -40°C

3.6 Ageing Test

For the ageing tests, one PUFPUF IC was placed in an oven at 80°C with a supply voltage of 110% Vdd (1.32V). Under these conditions, we accelerate the ageing effect of a chip. The total acceleration factor [12] is computed as the product of the thermal acceleration factor (TAF) and the voltage acceleration factor (VAF), which are computed as:

$$TAF = e^{\frac{E_a}{k}(\frac{1}{T_{op}} - \frac{1}{T_{stress}})}$$

$$VAF = e^{\gamma(V_{stress} - V_{op})}$$

The factor E_a (0.5 eV) is the activation energy, k (8,62 10^{-5} eV/°K) is Boltzmann's constant, T_{op} (313°K (40°C)) is the normal operating temperature, T_{stress} (353°K (80°C)) is the temperature used in the stress test, γ (2.6) is the voltage exponent factor, V_{stress} (1.32V) is the core voltage under stress conditions and V_{op} (1.2V) is the core voltage under normal operating conditions. This results in a total estimated acceleration factor of TAF×VAF = 8.17×1.37 = 11.2.

Every few days the ambient temperature was lowered to +20°C and the SRAM start-up values were measured (and stored in a file). Afterwards, the temperature was increased back to +80°C. One measurement at an ambient temperature of 20°C before starting the ageing test was used as enrollment, to which all other measurements are compared. Comparison between measurements is based on the FHD between the start-up patterns of the memories.

The ageing test has been running for 111 days . With the estimated acceleration factor of 11.2, we simulate an effective ageing of around 41 months , hence almost 3.5 years. The results show that within this time frame the ageing is quite limited. The maximum FHD remains below 10%. The results furthermore show that the SRAM memories experience less influence from ageing than the FF memories. Hence SRAM is more resistant to ageing than FFs.

3.7 Summary of the reliability tests

Table 1 summarizes the reliability tests. The notation used in the table represents mainly the relative strength between the different memories.

Table 1. Summarization of test results for measuring PUF reliability

Memory	Temperature	Voltage	Retention	Ramp-up	Dip time	Ageing
NXP SRAM	++	++	++	+	+/−	+
TSMC SRAM	++	++	++	+	+/−	+
FF	+/−	++	++	+/−	++	+/−

The results with respect to the ramp-up and dip time tests can be used for defining the system parameters. The ramp-up time is stricter for a FF than for an SRAM memory. On the other hand the required dip time is smaller for a FF than for an SRAM memory.

However, the results of the temperature test will have the largest consequences on the required efficiency of the fuzzy extractor. For the SRAM memories, there is only a 10% deviation for the different temperature measurements. This number is far below the acceptable boundaries (approximately 25% errors) for efficient error correction within the fuzzy extractor, where the efficiency is measured in terms of required hardware resources [13]. The FF reaches a maximum deviation of 40%, which will require extra processing and therefore more complex fuzzy extractors.

4 Uniqueness test results

We here describe the test set-up, together with the observations and the conclusions that can be drawn for the two uniqueness tests, as described earlier.

4.1 Between-class uniqueness test

When performing uniqueness tests, we are interested in finding out whether it is possible to distinguish between different devices given their PUF responses. This is mandatory when considering PUFs for authentication purposes or applications requiring unique identifiers. In order to create a between-class distribution, the response on one specific challenge of a particular device is compared to responses on the same challenge from different devices. The intra-class distribution is computed by calculating the FHD for different responses on a specific challenge from one particular device. Both histograms can be approximated by a Gaussian distribution and are summarized by providing their means, respectively, μ_{inter} and μ_{intra}, and their standard deviations, respectively, σ_{inter} and σ_{intra}.

In other words, the between-class uniqueness test measures the average distinguishability of two systems based on their PUF responses, i.e. μ_{inter}. For this reason μ_{inter} should be close to 50%. The calculation of μ_{inter} is based on 20 different ICs. It can be concluded that μ_{inter} of the 3 different memories are concentrated around 0.5. We refer to Table 2 for the exact values of μ_{inter} and σ_{inter}.

As μ_{intra} can be considered as the average noise on the response, it should be close to 0. The results of the tests are also very good for all three memories, as can be seen in Table 2.

Table 2. Summary of estimated means and standard deviations

Memory	μ_{inter}	σ_{inter}	μ_{intra}	σ_{intra}
NXP SRAM	0.4927	0.0035	0.0597	0.00270
TSMC SRAM	0.4970	0.0029	0.0536	0.00259
FF	0.4992	0.0039	0.0434	0.00512

From the between-class results, it can be concluded that it is possible to distinguish between different devices given their PUF responses.

4.2 Entropy

To estimate the entropy we use two compression algorithms (to estimate an upper bound for the entropy) and calculate the min-entropy, which leads to a lower bound on the entropy. Context-Tree Weighting (CTW) [14] is an optimal compression method for stationary sources and shows a good estimator of the available entropy. ZIP compression is the most common compression method.

Both algorithms are used to check the ability to compress response strings, as shown in [15]. The amount of compression will give an estimation of the upper bound of the entropy from our PUF responses. When the algorithm is capable of compressing the PUF responses, the responses do not have full entropy. This test was carried out by first concatenating all PUF responses into one string of 163840 bits respectively (8192×20) for the SRAM (NXP and TSMC) and 40960 bits (20×2048) for the FF. As can be seen in Table 3, the three types of memories turn out to have good compression resistance.

Table 3. Compression results of a concatenated string of 20 different devices.

Memory	Response	CTW	ZIP	CTW ratio	ZIP ratio
NXP	163840	162525	163207	99,1%	99,6%
TSMC	163840	164171	164002	100%	100%
FF	40960	41173	41087	100%	100%

Besides the compression factor, it is also possible to estimate the min-entropy of these memories. Min-entropy is the worst-case (i.e., the greatest lower bound) measure of uncertainty for a random variable. For this purpose we will be using the method that is described in NIST specification 800-90 [16] for binary sources. The output values of these sources have a probability of occurring p_0 and p_1 respectively (the sum of these two probabilities is 1). When p_{max} is the maximum value of these two probabilities, the definition for min-entropy of a binary source is:

$$H_{min} = -log_2(p_{max})$$

Assuming that all bits from the PUF start-up pattern are independent, each bit of the pattern can be viewed as an individual binary source. For n independent sources (in this case n is the length of the start-up pattern) the definition below holds, which is a summation of the entropy from each individual bit.

$$(H_{min})_{total} = \sum_{i=1}^{n} -log_2(p_{i\ max})$$

For our calculations we take the enrollment patterns that we have used during the temperature test. These patterns are bitwise added together to calculate a HW per bit, which can have a value between 0 and the number of enrollment

patterns (m). Based on this HW, p_{max} can be calculated for each individual bit of the start-up pattern:

$$\begin{aligned} \text{if } \mathrm{HW}_i > m/2 : p_{i\ max} &= \mathrm{HW}_i/m, \\ \text{else: } p_{i\ max} &= (m - \mathrm{HW}_i)/m \end{aligned}$$

Based on these values for p_{max}, the min-entropy of each individual bit (source) as well as the total min-entropy of the start-up pattern can be calculated using the formulas above. Finally, the average min-entropy per bit of a memory is calculated by dividing $(H_{min})_{total}$ by the length of the pattern n.

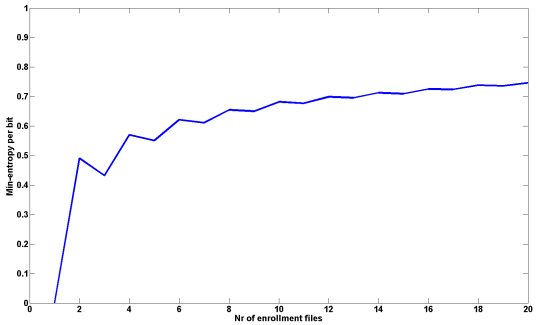

Fig. 7. Min-entropy development over the number of enrollment files (m).

Fig. 7 displays how the average min-entropy per bit of the NXP memory develops over an increasing m. It can be seen that after using 20 devices for this min-entropy test (the total number of chips measured for this paper), the average min-entropy per bit is still rising. This means that the values found by this test for the different memories are still conservative estimates, since these values would increase with more devices. Hence, the min-entropies from Table 4 are a conservative lower bound of the total entropy per bit for these memories.

Table 4. Conservative min-entropy estimate per bit based on 20 enrollment files.

Memory	Min-entropy
NXP	0.75
TSMC	0.76
FF	0.77

Based on the results from this section, it can be concluded that the entropy per bit for all tested memories is a value somewhere between the 0.75 (from min-entropy) and 1 (based on the compression test). This is a very high entropy,

especially considering the fact that the lower threshold is based on a very conservative estimate. We therefore conclude that the amount of entropy indicates that these memories are sufficiently unique to be used as PUFs.

4.3 Conclusions of the uniqueness tests

Table 5 summarizes the results from the uniqueness tests.

Table 5. Summarization of test results for measuring PUF uniqueness

Memory	between-class	compression	min-entropy
NXP SRAM	++	++	++
TSMC SRAM	++	++	++
FF	++	++	++

From these results, we conclude that there are no significant differences between SRAM and FF memories regarding PUF uniqueness. Both memory types perform very well in the uniqueness tests, since their entropy is high and μ_{inter} is close to 50%. These results show that it is possible to distinguish between different devices given their PUF responses.

5 Conclusions and future work

In this paper, we first defined a test framework for measuring the reliability and uniqueness properties of a PUF. This framework is used for comparing two types of intrinsic PUFs, the SRAM and FF PUFs, in 65nm technology. By means of six reliability tests, we have evaluated the strength of the PUFs under several external stress conditions. The SRAM PUFs turn out to have a shorter ramp-up time but a larger reset time, compared to the FF PUF. However, the most important difference is the resistance against temperature variations which is much better for the SRAM PUFs than for the FF PUFs. This results in a more efficient fuzzy extractor (being an implementation with less hardware resources) required for the SRAM PUF. From the results of the uniqueness tests, we conclude that both PUF types possess a sufficient amount of randomness.

Future work will be the evaluation of other types of PUFs following the proposed test framework of this paper. Using more devices (than the 20 used for this paper) for future tests will result in better statistics which would allow for even more confidence in test results. Secondly, it is also interesting to study the behavior of the PUF in combination with its processing algorithms, like the fuzzy extractor.

Acknowledgements

The authors would like to thank IMEC The Netherlands for supplying the PUF-PUF chips that have been used for the study as described in this publication.

All work preformed by Vincent van der Leest in this study has been supported by the European Commission through the FP7 program under contract 238811 UNIQUE.

References

1. P. S. Ravikanth, *Physical one-way functions*. PhD thesis, 2001. AAI0803255.
2. B. Gassend, D. Clarke, M. van Dijk, and S. Devadas, "Silicon physical random functions," in *ACM Conference on Computer and Communications Security*, (New York, NY, USA), pp. 148–160, ACM Press, 2002.
3. J. W. Lee, D. Lim, B. Gassend, G. E. Suh, M. V. Dijk, and S. Devadas, "A technique to build a secret key in integrated circuits with identification and authentication applications," in *In Proceedings of the IEEE VLSI Circuits Symposium*, pp. 176–179, 2004.
4. J. Guajardo, S. S. Kumar, G.-J. Schrijen, and P. Tuyls, "FPGA intrinsic PUFs and their use for IP protection," in *Proceedings of the 9th international workshop on Cryptographic Hardware and Embedded Systems*, CHES '07, (Berlin, Heidelberg), pp. 63–80, Springer-Verlag, 2007.
5. S. S. Kumar, J. Guajardo, R. Maes, G. J. Schrijen, and P. Tuyls, "The Butterfly PUF: Protecting IP on every FPGA," pp. 67–70, 2008.
6. R. Maes, P. Tuyls, and I. Verbauwhede, "Intrinsic PUFs from Flip-flops on reconfigurable devices," in *3rd Benelux Workshop on Information and System Security (WISSec 2008)*, (Eindhoven,NL), p. 17, 2008.
7. A.-R. Sadeghi and D. Naccache, *Towards Hardware-Intrinsic Security: Foundations and Practice*. New York, NY, USA: Springer-Verlag New York, Inc., 1st ed., 2010.
8. G. N. Selimis, M. Konijnenburg, M. Ashouei, J. Huisken, H. de Groot, V. van der Leest, G. J. Schrijen, M. van Hulst, and P. Tuyls, "Evaluation of 90nm 6T-SRAM as Physical Unclonable Function for secure key generation in wireless sensor nodes," in *ISCAS*, pp. 567–570, 2011.
9. P. Tuyls, B. Skoric, and T. Kevenaar, *Security with Noisy Data: Private Biometrics, Secure Key Storage and Anti-Counterfeiting*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007.
10. Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," *SIAM J. Comput.*, vol. 38, pp. 97–139, March 2008.
11. M. Claes and V. van der Leest, "PUFPUF test results," tech. rep., Intrinsic-ID, 2011.
12. Altera, "Reliability report 49 q1 2010," tech. rep.
13. C. Bösch, J. Guajardo, A.-R. Sadeghi, J. Shokrollahi, and P. Tuyls, "Efficient helper data key extractor on FPGAs," in *Proceeding sof the 10th international workshop on Cryptographic Hardware and Embedded Systems*, CHES '08, (Berlin, Heidelberg), pp. 181–197, Springer-Verlag, 2008.
14. F. Willems, Y. Shtarkov, and T. Tjalkens, "Context-Tree Weighting: Basic properties," *IEEE Transactions on Information Theory*, vol. 41, pp. 653–644, 1995.
15. V. van der Leest, G.-J. Schrijen, H. Handschuh, and P. Tuyls, "Hardware intrinsic security from D flip-flops," in *Proceedings of the fifth ACM workshop on Scalable trusted computing*, STC '10, (New York, NY, USA), pp. 53–62, ACM, 2010.
16. E. Barker and J. Kelsey, "NIST Special Publication 800-90: Recommendation for random number generation using deterministic random bit generators (revised), March 2007 NIST," tech. rep.