Design and implementation of enhanced lightweight memory-based and monostable physical unclonable functions

Zhong, Hui Shu

2019

Zhong, H. S. (2019). Design and implementation of enhanced lightweight memory-based and monostable physical unclonable functions. Master's thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/105323>

<https://doi.org/10.32657/10220/49517>

Design and Implementation of Enhanced Lightweight Memory-based and Monostable Physical Unclonable Functions

Zhong Hui Shu

**SCHOOL OF ELECTRICAL & ELECTRONIC
ENGINEERING**

2019

**Design and Implementation of
Enhanced Lightweight Memory-
based and Monostable Physical
Unclonable Functions**

Zhong Hui Shu

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Master of Engineering

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarized materials, and has not been submitted for a higher degree to any other university of Institution.

...30/June/2019.....

Date

Zhong Hui Shu

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledge in the author attribution statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang technological University and that the research data are presented honestly and without prejudice.

30/June/2019

.....

Date

.....

Chang Chip Hong

Authorization Attribution Statement

*(B) This thesis contains material from one paper submitted in the following journal in which I am listed as an author.

Chapter 4 is submitted as H. S. Zhong, C. Q. Liu, C. H. Chang “A Robust Monostable Strong PUF with Improved Uniformity and Enhanced Security” 2019 International Symposium on Integrated Circuits and Systems (ISICAS 2019, a CAS journal track symposium), Venetia, Italy, September 29-30, 2019 (submitted).

The contributions of the co-authors are as follows:

- A/Prof Chang Chip Hong provided the initial project direction and edited the manuscript drafts.
- I prepared the manuscript drafts. The manuscript was revised by Mr Chao Qun Liu.
- The schematic design and chip testing were conducted by me in the computer and facility for analysis, characterization, testing and simulation.
- I performed all the laboratory work at the School of Electrical and Electronic Engineering.

30/June/2019

Date

_____...

Zhong Hui Shu

Acknowledgements

This research was carried out at VIRTUS, the IC Design Center of Excellence at Nanyang Technological University, Singapore.

First of all, I would like to express my sincere gratitude to my thesis advisor, Professor Chang Chip Hong for his constant support throughout my Master of Engineering degree program. He not only encouraged me when I was lost in research, but also praised me when I made good progress in my studies. His enthusiasm and passion in research really motivate me a lot and enhanced my research experience. Besides, he shared with me a lot of deep philosophy of life, which is valuable and continue to inspire me in all aspects of life. It is a wonderful and precious experience for me to work and study with him.

I would especially like to thank my senior Mr. Liu Chao Qun for teaching me the knowledge and tools of hardware security. He also helped to review my research work and technical papers. After he graduated from the school, he still patiently answered my questions and discussed research ideas with me through phone. He provided me a lot of valuable suggestions during my studies.

Furthermore, I would also like to thank my other team mates, Wang Si, Ho Truong Phu Truan, Zheng Yue, Liu Wen Ye, and Zhang Yu Fei, who inspired and accompanied me during these two years of study. I very much like to remember this wonderful and fruitful experience with their supports always. Without them, I would not have made many improvements and achievements.

I also appreciate MediaTek Singapore Pte Ltd. for the financial support for my Master degree program and provision of chip tape-out training and opportunity, which equipped me with the essential skills to be a good IC designer.

Last but not the least, I would like to express my gratitude to my family for their support and dedication of my study. They encouraged me a lot and shared my enthusiasm for academic pursuit.

Table of Contents

Statement of Originality	i
Supervisor Declaration Statement.....	ii
Authorization Attribution Statement.....	iii
Acknowledgements	iv
Summary.....	vii
List of Figures.....	ix
List of Tables	xi
Chapter 1	1
1.1 Motivation	1
1.2 Objectives	4
1.3 Major Contributions	5
1.3.1 Design and Implementation of Novel Lightweight Multi-Response SRAM Physical Unclonable.....	5
1.3.2 A Monostable Strong PUF with Enhanced Security and Reduced System Bias.....	6
1.4 Thesis organization.....	6
Chapter 2	8
2.1 Physical Unclonable Function	8
2.2 PUF Quality Metrics	10
2.3 PUF Classification	12
2.3.1 Weak PUF	13
2.3.3 Strong PUF.....	15
2.4 PUF Applications.....	17
2.4.1 Cryptography Key Generation	17
2.4.2 Low-Cost Authentication.....	18
2.6 Monostable PUF	19
2.6.1 Static Weak Monostable PUF for Secure Chip Identification	19
2.6.2 Strong Monostable PUF for Lightweight Authentication	20
2.7 Machine Learning Attack	23
2.7.1 Support Vector Machine	23
2.7.2 Evolution Strategies	24
Chapter 3	26

3.1 Introduction	26
3.1.1 Power Consumption and Security Issues of Data Buffering	27
3.1.2 Design Conflicts for Dual-Mode Applications and Low Entropy of SRAM Cell	28
3.2 8T Dual-Port SRAM PUF	29
3.2.1 Access Mode and Issues of DP-SRAM	30
3.2.2 Working Principle of 8T DP-SPUF	32
3.2.3 Different combinations of Bit-Lines in memory and PUF modes	35
3.3 1024 Bits Dual Port SRAM Circuit Implementation	36
3.4 Result and Discussion	39
3.4.1 Correlation	39
3.4.2 Reliability	40
3.4.3 Uniqueness	41
3.4.4 Randomness	42
3.5 Conclusion	43
Chapter 4	44
4.1 Introduction	44
4.2 Enhancement of Monostable Strong PUF	45
4.3 Results and Discussions	51
4.3.1 Uniformity	51
4.3.2 Uniqueness	52
4.3.3 Reliability	53
4.3.4 Randomness	54
4.3.5 Security Against Machine Learning Attacks	54
4.3.6 Comparison with Other Strong PUFs	55
4.4 Conclusion	56
Chapter 5	57
5.1 Conclusion	57
5.2 Future Work	59
5.2.1 Multi-port SRAM PUF	59
5.2.2 Chip measurements of monostable PUF	59
Author's Publication	1
References	2

Summary

With the rise of internet of things, mobile devices leveraging on the omnipresence of the wireless connectivity have grown exponentially. To enable device identification and authentication, hardware security modules are embedded into the sensor, wireless and mix-signal integrated circuits. Silicon Physical Unclonable Function (PUF) is an emerging technology to augment or substitute existing cryptographic modules for more robust and lightweight hardware security and privacy protection. Over the last decade, it has gained a strong foothold in cryptographic key generation, device identification and authentication applications. Its security rests in the intrinsic complexity and irreproducibility of a random physical disorder system instead of hard-to-solved mathematical problems. Device signature generated by PUF cannot be physically replicated even by the original manufacturer with the same photolithography masks due to the uncontrollable nature of manufacturing process variations. Even though PUFs have been well recognized today as an efficient primitive for key preservation/generation and authentication protocols, some designs have also been criticized for several drawbacks. The research in this dissertation will review these problems and propose the implementations of a higher entropy memory-based weak PUF and an enhanced diode-clamped inverter monostable strong PUF.

Compare to other weak PUFs, the most important advantage of memory PUF is that it is almost cost free as the existing memory circuit in an integrated system can be reused for PUF response generation without requiring additional circuitry. The proposed dual port static random access memory (SRAM) PUF addresses two major limitations of existing SRAM PUFs. Firstly, a typical SRAM PUF needs an external memory to temporarily buffer the originally stored data before it is reset to generate the PUF response if the PUF response is needed in the midst of computations. Secondly, traditional SRAM PUF can only generate one response bit per cell. The proposed dual-port SRAM weak PUF could generate up to a maximum of four uncorrelated response bits per cell. Response bits are generated by launching two pairs of complementary data into the dual ports in write-write operation mode without resetting the entire memory array, which reduced the area-power overhead and potential security leakage caused by data transfer to buffer storage and data-retention attacks. Post-layout simulation results based on UMC65nm CMOS model has corroborated its superior correlation, reliability, uniqueness and randomness properties.

A novel monostable inverter-based strong PUF is also realized in this thesis. This proposed PUF is a new strong PUF structure featuring low power consumption and high area efficiency. The process variability of this proposed PUF can be enlarged by challenge-driven parallel inverter stage before the monostable output is amplified to a distinct binary output. The proposed enhanced lightweight monostable strong PUF harnesses the robust and reliable feedback stabilized entropy source of diode-clamped inverters (DC-INV) in subthreshold region. Due to manufacturing process variations,

the noise-immune monostable output of parallelly connected end-to-end DC-INVs is Gaussian distributed around half V_{dd} . The spread of the distribution is dispersed symmetrically by a challenge-driven parallel inverter stage before the output is buffered into a distinct binary state. The DC-INVs in these two stages are shared by a demultiplexer-multiplexer (DeMUX-MUX) pair in each block and the following parallel inverter buffer is also challenge-controlled to eliminate the systemic bias and improve the response uniformity. The inverter connections to the DeMUX-MUX in each block are also permuted to increase the modelling complexity against machine learning attacks. The 128-bit monostable PUF prototype chip was fabricated in 40nm CMOS technology. It consumes $14400 \mu\text{m}^2$ and 17 pJ per bit. The raw worst-case reliability is 94.1% over -40°C to 120°C and 0.9V to 1.3V . The average and worst-case uniformities are 50% and 54%, respectively. The prediction accuracies of support vector machine and covariance matrix adaptation – evolution strategy algorithms with more than 3000 training challenge-response pairs converge at 49% and 48.5%, respectively.

List of Figures

Figure 1.1 Internet of Things [1].....	2
Figure 2.1 Manufacturing variations of semiconductor devices as entropy source for generation of random challenge-response pairs for physical unclonable function. [9] ..	9
Figure 2.2 A simple PUF functional model.	9
Figure 2.3 Different challenges to a PUF instance produce different responses.	10
Figure 2.4 The same challenges to different PUF instances produce different responses.	11
Figure 2.5(a) Architectural of an SRAM PUF; (b) Transistor level implementation of a single bit SRAM cell; (c) Generation of response bit from SRAM PUF during power-up process.....	14
Figure 2.6 An arbiter PUF [9].....	16
Figure 2.7 Cryptograph key generation by PUF.	17
Figure 2.8 Low-cost device authentication. [50]	18
Figure 2.9(a) Proposed PUF bitcell; (b) PUF bitcell adds a sense amplifier for improved output stability.	20
Figure 2.10(a) Inverter Voltage Transfer Curve (VTC) and Trip point; (b) Schematic of diode-clamped inverter.	21
Figure 2.11 Architecture of the monostable strong PUF design.....	22
Figure 2.13 The working mechanism of SVM.	24
Figure 3.1 (a) 6T-SP-SRAM Cell (1RW); (b) 8T-DP-SRAM Cell (1R1W); (c) 10T-DP-SRAM cell (1R1W); (d) 8T-DP-SRAM (2RW).[13]	27
Figure 3.2 Working mechanism of a conventional memory based PUF system.[29] ..	28
Figure 3.3 8T DP-SRAM cell.	29
Figure 3.4 Four access modes of Dual-Port SRAM.	30
Figure 3.5(a) Word-line enable mode; (b) Response bit determined mainly by the mismatches of access transistors.....	33
Figure 3.6(a) Word-line disable mode; (b) Response bit determined mainly by mismatches of cross-couple inverters.	34
Figure 3.7 8T DP-SRAM operation in PUF mode (a) BLL ₁ = 0, BLL ₂ =1, BLR ₁ = 0, BLR ₂ = 1; (b) BLL ₁ = 0, BLL ₂ =1, BLR ₁ =1, BLR ₂ =0; (c) BLL ₁ =1, BLL ₂ =0, BLR ₁ =0, BLR ₂ =1; (d) BLL ₁ =1, BLL ₂ =0, BLR ₁ =1, BLR ₂ =0.	36
Figure 3.8 Architecture of the proposed 1024-bit DP-SRAM array.....	37
Figure 3.9 Layout of SRAM cells and array.....	38
Figure 3.10 Response bits generated from the same cell for 4 different access modes.	39
Figure 3.11(a) Reliability of PUF over voltage range from 1.08V to 1.32V; (b) Reliability of PUF over temperature range from -40°C to 120°C.	40
Figure 3.1 Interdie fractional HD for the proposed 8T DP-SRAM PUF.....	41
Figure 4.1(a) Basic inverter with feedback; (b) Diode-clamped inverter with feedback; (c) Feedback stabilized voltage V _M of diode-clamped inverter; (d) Widening V _M deviation from half V _{dd}	46
Figure 4.2(a) Enhanced monostable strong PUF circuit; (b) Buffer circuit and its gain; (c) Filter circuit and VTC curves of normal, low and high V _{trip} inverters.....	47

Figure 4.3 Monte Carlo simulation results of output voltage distributions at (a) Stage 1, (b) Stage 2, (c) Stage 3, and (d) final V_{out}	50
Figure 4.4 Layouts of (a) the 128-bit monostable PUF, and (b) a DC-INV.	51
Figure 4.5 Distribution of inter-die fractional harming distance of 5000 responses from 80 PUF instances.....	52
Figure 4.6 Reliability against (a) temperature and (b) voltage variations.	53
Figure 4.7 The prediction accuracy by SVM and CMA-ES for the proposed 128-bit enhanced PUF and the original 128-bit DC-INV PUF [12].	556

List of Tables

Table 3.1 Different access conditions.	31
Table 3.2 Conflict of different access condition.	31
Table 3.3 Two different combinations of bit-line in memory mode.....	35
Table 3.4 Four different combinations of bit-line in PUF mode.	35
Table 3.5 Correlation coefficients of 4 access cases.....	40
Table 3.6 Results of ten main NIST tests.	42
Table 3.7 Percentage of occurrences of all combinations of output patterns for the four different cases.	43
Table 4.1 Connection sequence of the shared inverters between the two DeMUX-MUX pairs in different blocks.	49
Table 4.2 NIST test of bit sequences generated by the proposed PUF.....	54
Table 4.3 Comparison with state-of-art strong PUFs.	56

Chapter 1

Introduction

1.1 Motivation

The proliferation of mobile electronic devices benefits the ubiquity of internet of things (IoT) and drives tremendous growth of economic value. The term “Internet of Things” according to Wikipedia, was proposed by Kevin Ashton, the cofounder of the Auto-ID Center at the Massachusetts Institute of Technology, which created a global standard system for RFID and other sensors. His concept of IoT is illustrated in Figure.1, where the rapid growth and convergence of wired and wireless data communication technologies are making network connections more relevant and valuable via ubiquitous sensors that connect the physical world to the cyber space [1]. With the large-scale deployment of IoT devices across various fields, the privacy of sensitive data is jeopardized. Building secure features for identification, authentication and counterfeit detection of IoT devices have become increasing desirable and urgent. With various reported vulnerabilities and limitations of their implementation by traditional cryptographic methods, there is a dire need to explore new approaches to privacy protection and security in IoT communications by embedding lightweight hardware intrinsic security primitives into IoT devices.

Software security modules are easier, more convenient and flexible to implement, update and maintain. However, software security suffers from a fundamental problem, which is the heavy reliance on the security and trustworthiness of its underlying physical layer that executes the program. Most of the time, the assumption of a well-protected operating system or physical processing unit cannot be assured in practice, especially for IoT devices and edge computing systems which are publicly and physically accessible. Furthermore, the source codes stored in the memory of edge computer are exposed to a greater chance of tampering. In general, reverse engineering of software to recover the confidential information is less costly and easier even by the low class adversaries with limited resources. Lastly, the known timing and power consumption of instructions of general-purpose digital signal processors, microprocessors or microcontrollers used for the execution of these codes can be easily exploited for delay- and power-based channel attacks [2]. As the chain is only as strong as its weakest link, the incentives of attacking software cryptographic modules are high.

Figure 1 Internet of Things [1].

To enhance security while preserving flexibility, cryptographic algorithms are partially migrated from software to hardware implementation to enjoy the best of both worlds. The private key (shared secret) for encryption and decryption is typically stored in a non-volatile memory (NVM) or battery-backed static random-access memory (SRAM) and protected by cryptographic hash [3]. Unfortunately, using hardware security primitive to provide the root of trust to software execution is also not completely free from attacks. Recently, attackers have successfully hack into hardware system by manipulating hardware settings to evade the protection mechanism [4]. Safely managing secrets in memory is difficult and expensive, which often incurs non-negligible overhead in area and power consumption. Secondly, NVM technologies, such as Electrical and Erasable Programmable Read Only Memory (EEPROM), and even polysilicon fuses and antifuses used for hardcoded device identities, are vulnerable to invasive and semi-invasive attacks [5]–[7]. Secret can be eavesdropped by utilizing advanced probing techniques or dismantling the storage device to measure the side channel signal without incurring much cost on time and money. Battery-backed SRAMs can be read even after the secret information has been stored for a sufficiently long time by data remanence attacks [8]. Once the protected content of the memory has been extracted, analysis can be performed offline by powerful computer to recover the original confidential information.

One of the most representative hardware solutions to reliably manage privacy information is using the physical characteristics of hardware to generate secret keys [9]. This solution is proposed by Bauder and Simmons in 1980s and is broadly accepted as a primitive trial to exploit physical properties of hardware for security purpose. A

Physical Unclonable Function (PUF) is a hardware security primitive that is designed to specifically generate mismatched parametric performance through the random and unpredictable manufacturing process variation of nano-scale devices. The first silicon based PUF was introduced by Gassend *et al.* in 2002 [10], where identically designed circuits were used to produce different outputs (responses) when the same input (challenge) was applied as a stimulus. Since then, PUF has been extensively studied and explored as a hardware root of trust for many security applications, notably for device identification and authentication in IoT environment and unforgeable IDs for supply chain management. Within one year, approximately 300 articles on PUF have been published. Some of the unique features of PUF as a hardware security primitive are summarized as follows [2]:

- As manufacturing process variation is uncontrollable, it is unlikely to produce two identical chips with perfectly matched transistors even with the same mask. Consequently, the challenge-response pairs (CRP) generated by each PUF instance of a well-designed PUF are random, unique and unpredictable. The responses to input challenges are easy to measure but hard to clone. The unclonability arises from the hardness of reproducing the multitude of uncontrollable manufacturing process parameters.
- The manufacturing process variation is unmodifiable, so costly anti-tamper mechanisms are not required. It is difficult for physical attack to reverse engineer the PUF circuit to produce an exact replicate without changing the physical characteristic [3]. The signatures produced by the PUF cannot be easily removed, copied or compromised, as the secrets are inherent in the physical structure of the PUF. Any invasive or semi-invasive attack on the chip can easily destroy the original secrets. This is in stark contrast to digital IDs or conventional hardcoded electronic tag, digital signature or cryptographic hash, that need to be preserved persistently in a digital form.
- The challenge-response interface can be locked down or obfuscated to prevent them from externally accessible after enrolment. Using PUF technology, the device signature becomes volatile as it will only be generated on request and vanish after use. Physical attack to probe into the internal nodes to extract the CRPs must be performed when the chip is powered up. Hence, the cost of invasive attacks increase significantly with negligible chance of success as any subtle change can easily destruct the challenge-response mapping of the PUF.
- No non-volatile memory is required to store the secret as it is embodied in the circuit structure. Power-down the circuit will not erase the secret.

- It takes only a small footprint and can be monolithically integrated as part of an existing integrated circuit or system.

One important application of PUF in cryptography is authentication. To use PUF for authentication of devices, it is necessary to prevent the adversary from exploiting the communication protocol to masquerading as an authentic device. To avoid replay and man-in-the-middle attack, the challenge should not be reused. A PUF that is capable of supporting an exponential number of challenges with a bounded number of basic bitcells is classified as a strong PUF. Strong PUF designs are mostly delay-based. Arbiter PUF (APUF) is a typical representative of low-power and high-speed delay-based strong PUF. APUF architecture exploits the cumulation of delay mismatches of symmetrical wires and transistors of each bit slice controlled by a challenge bit to generate a unique response bit to an input challenge. However, not all CRPs are equally reliable due to the metastability of arbitration between two symmetric racing paths. APUF is also vulnerable to machine learning based modelling attacks as its structure is built by linear additive blocks. Weak PUF, on the contrary, can only support a limited number of challenges per bitcell. Hence weak PUF is usually used for device identification, unforgeable electronic tag, seed to pseudo random number generator and secret key generator. Memory-based PUF is a typical example of weak PUF. The advantage of memory-based PUF is the existing memory of a device can be directly reused to generate PUF response for device identification. Unfortunately, it has limited entropy of only one response bit per cell. The memory reset required for response generation has introduced extra area-power overhead as well as potential security leakage due to the requirement to temporarily buffer or mirror the currently stored data in the memory array if the PUF response is required in the midst of a computation.

Motivated by the shortcomings of existing strong and weak PUFs, new designs of volatile memory-based weak PUF with greater entropy, and monostable strong PUF with enhanced response uniformity and machine learning resistance are developed and efficiently implemented in CMOS integrated circuits.

1.2 Objectives

The research work in this thesis was inspired by prior research work published in [11] [12] [13]. The objective of this Master of Engineering program is to propose enhanced designs of a memory-based weak PUF with multiple response bits per cell and a highly secure and good uniformity monostable strong PUF.

The most attractive advantage of SRAM PUF is to extend the ubiquity of memory for both security and storage applications without introducing additional circuit to conserve chip area. However, the conventional 6T SRAM PUF suffers from several shortcomings. Firstly, response bit generation on 6T SRAM PUF relies on the power-up process. The entire SRAM array needs to be powered off and on to generate response bits even though only a small fraction of cells is addressed by the challenge, which consumes spurious power. Secondly, the data stored in memory will be overwritten by random bits during power-up process. To prevent the useful data from being overwritten, an external buffer storage is required to temporarily store these original data before switching to PUF operation. This additional buffer memory not only introduces area and power overheads, but also creates potential security vulnerability to fault-injection and data remanence attacks. Thirdly, this 6T SRAM can only store 1 bit per cell, it limits the entropy of response bit generation to at most 1 bit per cell. To this end, a SRAM-based PUF is designed to address these problems and constraints.

The diode-clamped inverter PUF is a new strong PUF structure. The elementary entropy is derived from the process variations of diode-clamped single inverter ring working in the subthreshold region. Due to manufacturing process variation, the output voltage of an end-to-end parallel inverter ring is Gaussian distributed at half V_{dd} with a narrow spread. This process variability is enlarged by additional challenge-driven parallel inverter stage before the metastable output to an input challenge is amplified to a distinct binary output. The chip testing results show that the original work proposed in [12] has a poor uniformity and less ideal uniqueness. The objective of this research is to improve the uniformity and uniqueness. Moreover, the new design also improves the complexity of mathematically modelling the PUF and reduce the predication accuracy of machine learning based modelling attack.

1.3 Major Contributions

During this Master of Engineering research program, several contributions have been made and they are summarized below.

1.3.1 Design and Implementation of Novel Lightweight Multi-Response SRAM Physical Unclonable Functions

A 1024 bits dual-port SRAM PUF is designed. To minimise the potential security vulnerability to fault-injection attack, the proposed SRAM PUF is designed to be buffer-free. The output response bits are generated by writing complementary voltages into an

addressed SRAM cell instead of resetting the entire memory array. It has eliminated the use of buffer memory, which also reduced the area overhead. Each memory cell can generate a maximum of four output response bits due to four different combinations of input voltages. Compare with [13], it provides a higher entropy for response generation. The proposed DP-SRAM PUF is evaluated on UMC65nm CMOS technology and the post-layout simulation results are presented in Chapter 3.

1.3.2 A Monostable Strong PUF with Enhanced Security and Reduced System Bias

An improved monostable clamp-diode inverter strong PUF over the design of [12] is proposed and fabricated. The uniformity of the ‘0’ and ‘1’ response bits is improved by merging the first two challenge-controlled stages of [12] and the inverters between these two stages are shared to eliminate the systemic bias. The proposed circuit explores the nonlinear relationship between the inverter’s trip point voltage and the physical process variations to significantly increase the complexity of modelling. The inverters are connected differently in different blocks to reduce the prediction accuracy of the existing powerful black-box modelling by Support Vector Machine (CMA) and Covariant Matrix Adaptation Evolution Strategy (CMA-ES) machine learning algorithms. The proposed monostable strong PUF is evaluated on TSMC 40nm CMOS technology. The post-layout simulation and physically fabricated chip measurement results are presented in Chapter 4.

1.4 Thesis organization

The thesis is organized into five chapters:

Chapter 1 introduces the research motivation, objectives and research direction, and summarizes the major contributions of this thesis.

Chapter 2 presents a brief review on the background and related research works. In this chapter, the definition of PUF is introduced. The metrics used for assessing the quality of PUF are presented. Additionally, two main categories of PUF (strong PUF and weak PUF) are compared in terms of working mechanisms and applications. Arbiter and memory PUFs are used as typical examples of strong and weak PUFs for illustrating their principles of challenge-response generation. Moreover, the original designs of two emerging monostable PUFs are reviewed. Besides, multiple port SRAM is reviewed. Finally, SVM is singled out for review as an example of machine learning attack.

Chapter 3 is devoted to the dual-port memory based PUF design. The chapter begins with a brief review of conventional 6T SRAM PUF. The advantages and disadvantages are analyzed. Three different types of 8T SRAM access mode and one forbidden mode are introduced. Then the working mechanism and advantages of 8T SRAM PUF are explained. Next, the circuit implementation of 1024-bit 8T dual-port SRAM PUF is presented. Lastly, the post-layout simulation results of the proposed circuit and the conventional approaches are compared and discussed.

Chapter 4 presents a monostable strong PUF with enhanced security and reduced system bias. In this chapter, the shortcomings of the existing monostable strong PUF are introduced. The circuit structure is analyzed for improvement. The new design to optimize the uniformity of proposed PUF circuit is introduced. The circuit are also designed with consideration of increasing modelling complexity to reduce the machine learning prediction accuracy. Lastly, chip measurement results and comparison with existing state-of-the-art strong PUF designs are provided.

Chapter 5 draws the conclusion and proposes prospective future works based on the current findings.

Chapter 2

Literature Review

2.1 Physical Unclonable Function

The term Physical Unclonable Function (PUF) was initially introduced by Pappu [14] in 2002 as a physical one-way function. At around the same time, this notion was physically demonstrated and popularized by the MIT researchers Gassend et al. [15] who built two variants of CMOS integrated circuits (ICs) to harness the fabrication process variations of CMOS integrated circuit (IC) components for authentication applications. Research work in PUF has since then mushroomed. With the scale of semiconductor fabrication process went down into the deep submicron and nanometer regime, it became possible to leverage on the previously undesirable process variations in IC fabrication to design even better quality PUFs [16]-[22]. Today, PUF has become well known as the device ‘biometric’ or IC ‘fingerprint’ that can be monolithically integrated into or embodied by a physical system to provide a lightweight hardware-intrinsic root-of-trust for security-critical applications.

A pictorial illustration of the underlying principle of PUF is shown in Figure 2.1, where the mismatch in electrical parameters due to the manufacturing process variations of different transistors fabricated on the same wafer is extracted by identically designed PUF circuits to generate a random binary response upon stimulated by a binary input bitstream (also known as challenge). Because such manufacturing process variations are unavoidable and unpredictable, the random binary string produced by applying the same challenge is unique to each PUF circuit instance and cannot be reproduced even with the full knowledge of its circuit design. Moreover, it is relatively easy to evaluate the CRPs to authenticate a hardware with an embedded PUF but significantly harder, if not impossible, to reverse engineer the circuit to uncover the exact device parameters for reproducing the same challenge-response behavior [9].

Figure 2.1 Manufacturing variations of semiconductor devices as entropy source for generation of random challenge-response pairs for physical unclonable function [9].

As depicted in Figure 2.2, a PUF can be formally modeled as a physical system that probabilistically maps a stimulating challenge c selected from a finite discrete space $C = \{0,1\}^k$ to the output response r from the binary space $R = \{0,1\}^l$ according to its intrinsic stochastic properties χ arising from the manufacturing process variations of its components. The challenge-response mapping of the PUF can be expressed as:

$$PUF_{\chi} : C \rightarrow R \quad (2.1)$$

The operation of generating the response $r \in R$ given a challenge $c \in C$ can be expressed as:

$$r \leftarrow PUF_{\alpha_p}(c) \quad (2.2)$$

where $\alpha_p \in \chi$ represents the set of disorder physical parameters of the specific PUF instance P .

Figure 2.2 A simple PUF functional model.

2.2 PUF Quality Metrics

Although there is no common consensus to assess the performance of PUF, the following criteria are widely used in academic and research communities to evaluate the quality of PUFs.

2.2.1 Uniformity

Uniformity measures the percentage of ‘0’ bits or ‘1’ bits among all the response bits that are generated from the same PUF. As shown in Figure 2.3, when different challenges are fed into a PUF instance, response bits of ‘0’s and ‘1’s are produced randomly. The uniformity is defined as:

$$Uniformity = \frac{1}{n} \sum_{i=1}^n r_i \times 100\% \quad (2.3)$$

where r_i is the i -th binary bit of n responses. The ideal value of uniformity is 50%.

Figure 2.3 Different challenges to a PUF instance produce different responses.

2.2.2 Uniqueness

The response of a PUF circuit should be unique compared with the response generated by other identically designed PUF circuits in the same or different manufacturing lots when the same challenge is applied. Uniqueness indicates the distinguishability of a PUF instance from other PUF instances. As shown in Figure 2.4, the same challenge applies to different PUFs will produce different responses. Ideally, the responses

generated from two PUF instances by applying the same challenges should have half of the bits different. Thus, the ideal uniqueness should be 50%. The uniqueness can be quantified by measuring the fractional Hamming distance (HD) among the responses. For a l -bit response, the uniqueness is expressed as follows:

$$Uniqueness = \frac{2}{m \times (m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m \frac{HD(R_i, R_j)}{l} \times 100\% \quad (2.4)$$

where m is the number of sampled response strings, R_i and R_j are the two l -bit responses from two different PUF instances generated by the same challenge. $HD(R_i, R_j)$ is a function that evaluates the Hamming distance between two binary strings, R_i and R_j .

Figure 2.4 The same challenges to different PUF instances produce different responses.

2.2.3 Reliability

Reliability measures the capability of a PUF to reproduce the same response from the same challenge in different environments. In practice, flipping of response bits may occur due to the fluctuation of ambient temperature, supply voltage and aging effect. The reliability can be evaluated by the Bit Error Rate (BER), which can in turn be calculated from the intra-die Hamming distance. For a l -bit response, the reliability is expressed as follows:

$$Reliability = 1 - BER = 1 - \frac{1}{m} \sum_{t=1}^m \frac{HD(R_i, R_{i,t})}{l} \quad (2.5)$$

where m is the number of sampled response strings, R_i is the response of the i -th PUF instance under the nominal condition, and $R_{i,t}$ is the response of the i -th PUF instance at time t where the operating condition may have changed and deviated from the nominal condition. The ideal value of reliability is ‘1’.

2.2.4 Unpredictability

Unpredictability is a probabilistic measure of the generated responses to different challenges. It is useful in formally evaluating the semantic security of a PUF system under a given adversarial scenario. Sometimes, “randomness”, measured by “entropy”, can be found in the literature and is equivalently used to represent a statistical measure of unpredictability. As the theoretic entropy of a PUF is often difficult to determinate, some well accepted cryptographic randomness test suites are often used to evaluate if the sequence of responses generated by applying random challenges to a PUF circuit fulfill the criterion of unpredictability statistically.

Intuitively, it is difficult to predict the random responses of a PUF to unknown challenges as the challenge-response mapping characteristics is a consequence of uncontrollable physical disordered properties. This implies that even the foundry that manufactures the PUF chips has no idea about its physical properties, hence it is impossible to deliberately fabricate a clone of it. Under some circumstances, however, an adversary may gain access to the internal challenge and/or response interface, or even the physical layer of the PUF chip to directly read or obtain some responses to specific challenges of the PUF and use them to predict the responses generated from the same PUF device to some unused challenges. This attack is a generic prediction attack where the actual form of attack can be of any kind that makes the successive prediction following a challenge-response collection successful. Therefore, a stronger assertion of unpredictability is unclonability, which refers to the resilience of a PUF instance to shield itself against mathematical approaches to create a clone of it. A specific example of this attack is the modeling attack where the output responses of a PUF are collected and trained with a PUF model by using machine learning techniques. The model will then be used to predict responses to other possible challenges by using the learning algorithm.

2.3 PUF Classification

With the rapid evolution of PUF during the last few years, the number of different types of PUF is growing continuously. It is difficult to categorize the vast variety of PUFs

exactly from any single perspective. In this report, a widely accepted broad classification of weak and strong PUFs according to the response access restriction and modelling complexity is considered. The qualifier weak or strong should not be interpreted as related to the security strength or quality of the PUF. Ideally, all PUFs, weak or strong, should share the common characteristic that they are both impossible to be physically cloned, since their challenge-response mappings are stem from the subtle manufacturing process variations of nano-scaled devices, which cannot be exactly duplicated. The main characteristic differences of weak and strong PUFs are presented below.

2.3.1 Weak PUF

Weak PUFs are also historically known as physically obfuscated keys (POKs) [23]. Examples are memory PUF, butterfly PUF and coating PUF [1]. The number of CRPs are linearly or polynomial proportional to the number of bit-slice components abutted to complete the PUF circuit. The unpredictability of weak PUF is very high since response bits are generated with dedicated challenges, minimizing the correlation among CRPs. Due to finite physical size of CRPs, it is crucial to protect the CRP interface from being accessed by limiting the direct access to the internal input challenge and/or raw output response. In the view of the limited number of challenges of weak PUF, it is more suitable for secret key generation in cryptographic applications, such as data encryption and decryption. For these applications, error correction of noisy response is indispensable to ensure the bit accuracy and reliability. Hence, it is important to assure a high raw reliability to minimize the cost of error correction hardware.

Memory-based PUF is the representative of weak PUF. Memory-based PUF exploits the metastable states of latch structures (e.g., two cross-coupled inverters) to generate random bits determined by threshold voltage mismatch of the cross-coupled devices. Variants of memory PUFs have been developed in the past few years, such as latch PUF [24], butterfly PUF [25], D flip-flop PUF [26], buskeeper PUF [27] and SRAM PUF [28]. Due to its ubiquity, SRAM is also one of the most popular building blocks used for the construction of PUF to protect the integrity of data storage. An architectural diagram of SRAM PUF is given in Figure 2.5(a). It consists of a control unit, a row/column decoder, a SRAM array, column sense amplifiers and write drivers. A specific SRAM cell is addressed by a set of column-row indicators (challenges) to obtain a bit string (responses) [29]. A fuzzy extractor combined with a complex Error Correction Code (ECC) is used to postprocess the response before it is produced as a cryptographic key to ensure a good reliability of SRAM PUF [30].

6T SRAM cell is the most widely used bit cell in a SRAM array. The schematic diagram of a 6T SRAM cell is shown in Figure 2.5(b). The circuit is designed symmetrically with

a pair of pull-up PMOS transistors (M_2 and M_3), a pair of pull-down NMOS transistors (M_1 and M_4) and two access transistors (M_5 and M_6). The random bit generation relies on the mismatch of the cross-coupled inverters within the SRAM cell during power-up process as shown in Figure 2.5(c). The output will be stabilized by the positive feedback loop formed by these cross-coupled inverters. The mismatch is predominantly ascribed to the variation in the transistor geometries, such as the channel length (L), the channel width (W), the thickness of oxide (T_{ox}), the threshold voltage (V_{th}), and some physical phenomena, such as edge-line roughness and fluctuations of the doped impurities. As the transistor size shrinks with the advance of CMOS technology, V_{th} variation becomes the most significant entropy for SRAM cell [31]. The threshold voltage is dependent on the transistor geometry and its process variation is determined by the following formula [32]:

$$\sigma_{V_{th}} = \frac{\Lambda \cdot \sigma_{V_{th,max}}}{\sqrt{(W \cdot L)}} \quad (2.6)$$

where Λ is a technology-dependent parameter, $\sigma_{V_{th,max}}$ is the maximum standard deviation of V_{th} , W and L are the transistor width and length, respectively. Typically, for Λ of around 5nm, $\sigma_{V_{th,max}}$ is about 15% of the nominal value of $\sigma_{V_{th}}$.

Figure 2.5(a) Architectural of an SRAM PUF; (b) Transistor level implementation of a single bit SRAM cell; (c) Generation of response bit from SRAM PUF during power-up process.

During power-up reset operation, all the six transistors within the SRAM cell are working in the subthreshold region. The drain-source current can be expressed as:

$$I_{SUB} = I_S e^{\frac{V_{gs} - V_{th}}{nV_T}} \left(1 - e^{-\frac{V_{ds}}{V_T}} \right) \quad (2.7)$$

where I_{SUB} is the subthreshold current, $n = 1 + \frac{C_d}{C_{ox}}$ is the subthreshold slope factor,

$V_T = \frac{KT}{q}$, k , T and q are the Boltzmann constant, temperature and elementary charge, respectively.

When the subthreshold current flows through transistor M_1 and M_4 , the voltages of Q and QB will be pulled up simultaneously. Due to the mismatch of the cross-coupled inverters, either Q or QB will reach the threshold voltage earlier than the other. Consequently, one of the two nodes will be pushed to either '0' or '1' stable state by the positive feedback loop. For instance, after the voltage of node Q ramps up to the threshold voltage of inverter B (INVB), the voltage of QB, which is the output node of INVB, will be pulled down to the ground. As result, M_2 will be completely turned on with descendance of voltage at node QB and the voltage of node Q will be pulled up to V_{dd} . The mismatches of these two cross-coupled inverters among the different cells of the same SRAM array and among the same cells of different SRAM arrays are random and unpredictable. Hence, the responses generated during the power-up operation of each SRAM array can be used as a unique signature for device identification.

2.3.3 Strong PUF

In contrast to weak PUFs, strong PUFs have a huge number of challenges. The number of CRPs are exponentially proportional to the number of PUF bit-slice components, such as arbiter PUF [33] and lightweight secure PUF [34]. Due to the exponentially large number of CRPs, it is very hard to access all the challenges in a practical amount of time. Hence, strong PUFs do not have the same constraint of depleting all challenges and can avoid replay attack by restricting each challenge to be used only once. It can in principle be authenticated directly with an open and unprotected challenge-response interface. Consequently, strong PUFs are more suitable for low cost authentication. However, due to the advance in machine learning, the used CRPs can be learnt by machine learning algorithms to mathematically model the strong PUF. Because of this, some protection, for example by challenge or response obfuscation, is required to prevent the raw CRPs from being collected by the adversary. Most of the strong PUFs are delay-based. Delay-based PUFs generate the CRPs by racing pulse through replicated delay chains of identically designed components and comparing the

difference in propagation delays of two timing-paths as a consequence of uncontrollable fabrication process variability to generate a random response bit.

Figure 2.6 An arbiter PUF [9].

Arbiter PUF (APUF), as shown in Figure 2.6, is a typical and most popular delay-based strong PUF architecture proposed by Gassend as Physical Random Function in [10]. APUF consists of several multiplexers and a single latch. Two multiplexer daisy chains are controlled by the challenge string, which work as crossbar switches. In principle, the nominal delay of these two paths should be identical when the input pulse launched into the upper and lower multiplexer arrays simultaneously upon application of the input challenge. However, due to the intra-die variations, the delays of identically designed multiplexer switches are slightly different, and these differences cumulate over the number of cascaded switches, resulting in different time delays between the two paths. Finally, the response bit is generated by the final arbiter, which is determined by which timing path is faster. If the top delay path is faster, the response is ‘1’. Otherwise, the response is ‘0’. Since the response generation entirely relies on the difference of the propagation delay between the two symmetric signal race paths, more stages are required to provide a sufficient time delay difference to ensure the response stability, which result in higher power and area consumptions. In addition, it is hard to ensure the reliability of APUF as the delay of each stage will be affected significantly by the dynamic environmental conditions such as supply voltage, ambient temperature, random noise and circuit aging. Finally, APUF is vulnerable to modelling attack as its structure are built by linear additive delay blocks. To thwart machine learning attacks, some “nonlinear” effects are added to increase the complexity of APUF such as XOR-APUF [22], and feed-forward APUF [35].

2.4 PUF Applications

PUF can be deployed in many applications encountered in daily life. For example, RFID is often used to grant access to building or identify products in supermarket. Based on the number of CRPs, there are two main applications, namely cryptography key generation and low-cost device identification for weak PUF, and authentication for strong PUF.

2.4.1 Cryptography Key Generation

Figure 2.7 Cryptograph key generation by PUF.

A cryptography key generation mechanism is shown in Figure 2.7. There are two main steps involved in this application, which are initialization and re-generation. These two steps ensure that the same output could be produced consistently under different environmental conditions. In the initialization step, random response bits will be generated. In case the response bits are flipped by noise or are not independently and identically distributed, they will be post-processed by a fuzzy extractor (FE) to correct the flipped bits and boost the key entropy. The helper data is generated as a public information and can be saved in a non-volatile memory (NVM) for the re-generation step. In the re-generation step, a potentially noisy response will be generated from the PUF circuit. The FE uses the helper data from the initialization step to correct the errors in the response. This mechanism can be expressed mathematically as follows [2]:

$$r \leftarrow \text{Ext}_{\alpha_e}(\text{PUF}(C) + e, h) \quad (2.8)$$

where r is the output response, $\text{Ext}()$ is the fuzzy extractor, e is the error bit string introduced by noise, and $h \in H$ is the helper data that is utilized by an Error Correction

Code to correct bit errors and universal hash function to randomize the output response. $\alpha_e \in E$ is an attribute of the Error Correction Code and hash function used in $Ext()$.

With FE, the challenge-response mapping becomes:

$$PUF_{K,E} : C \leftarrow R \times H \quad (2.9)$$

In this way, a reliable shared key is produced. As the cryptographic key is produced by the PUF when it is requested through the applied challenge. It needs not be stored in non-volatile memory, which minimizes the risk of attack.

2.4.2 Low-Cost Authentication

Figure 2.8 Low-cost device authentication. [50]

Strong PUFs are used for low-cost device authentication as shown in Figure 2.8. Strong PUFs can produce exponential number of CRPs. A sufficiently large subset of CRPs will be randomly measured and stored in the database by the server during the enrolment phase of the device for future interrogation. During deployment, the authenticity of a client device A embedded with the PUF in an untrusted supply chain or environment can be examined by the server. The server selects a challenge C from its database and sends it to the client for authentication operation. The client applies the challenge C into the device A to generate a new response R . If R generated by the device matches with the previous response recorded in the database for challenge C with a tolerable threshold of fractional Hamming distance, the device is authenticated successfully. Otherwise, the authentication fails. For security reason, the used challenge will be discarded to prevent replay attack. If the device is intercepted by an adversary, it is very hard for the attacker

to measure the complete set of CRPs within a reasonable time and the probability of overlapping CRPs between the subsets of CRPs measured by the attacker and the server during enrolment is low if the CRP space is huge. Therefore, if the CRPs of the PUF is externally accessible or unprotected after enrolment, the PUF device A should be transported to the client through a secure physical channel. The client should send the acknowledgement of receipt of device A to the server within a stipulated time through a private binary channel before use. This practice efficiently prevents the device from falling into the hands of adversaries.

2.6 Monostable PUF

Besides conventional PUFs, a group of emerging PUFs are sprouted in the literature, such as reconfigurable PUF [36], public PUF [37], emerging memory PUF [38] [39], aging resilient RO-PUF [40], sensor PUF [41], extreme learning machine PUF [42] etc. Of which, a new type of CMOS-based strong PUFs, called monostable PUF [12], which is highly nonlinear and is not based on linear additive delay bitslice structure, has attracted the attention in this study. The basic structures of weak and strong monostable PUFs will be reviewed and presented here. Further enhancements to the monostable strong PUF will be presented in Chapter 4.

2.6.1 Static Weak Monostable PUF for Secure Chip Identification

The concept of this weak monostable PUF relies on the mismatch among the transistors of two complementary current mirrors as shown in Figure 2.9. The schematic of an INV_PUF Bitcell consisting of cascode current mirrors M_1 - M_4 and M_5 - M_8 is shown in Figure 2.9(a). The same amount of current carried by the current mirror flows through their input transistors M_3 and M_5 . The current ratio of the current mirror circuit is 1:1, the equivalent amount of current of M_3 and M_5 will be mirrored to M_4 and M_6 , respectively. Low power consumption can be realized by making the transistors, M_1 , M_3 , M_5 and M_7 , conduct in subthreshold region because the supply voltage V_{DD} is lower than the sum of the transistors' threshold voltages.

Ideally, due to the symmetry of the circuit topology shown in Figure 2.9(a), M_4 and M_6 would conduct the same amount of saturation current (I_{4sat} , I_{6sat}), and the voltage at node Y (V_Y) is equal to the voltage at node X (V_X) ($\approx V_{DD}/2$). In practice, the transistors are manufactured under deep submicron technology which is inevitably affected by the manufacturing process variations. These random mismatches among transistors M_1 - M_2 and M_7 - M_8 will induce slight mismatch between I_{M4} and I_{M6} . The small signal impedance R_Y at node Y is large enough to convert this current difference to either a

high output voltage (V_{DD}) or a low output voltage (GND). When $(I_{M4} - I_{M6}) > 0$, V_Y is swung to V_{DD} . When $(I_{M4} - I_{M6}) < 0$, V_Y is swung to GND. As a result, a digital output is derived from the random mismatch current that is essentially caused by the manufacturing process variation among transistors. In some statistically infrequent cases, the extremely small difference of $(I_{M4} - I_{M6})$ is not capable of making a full swing of V_Y , resulting in unstable output response. In order to reduce the unstable response, one sense amplifier formed by transistor M_9 - M_{13} is inserted to add gain and introduce more random mismatch through the sense amplifier offset as shown in Figure 2.9(b), thus reducing the number of unstable response bits. However, this proposed circuit is a weak PUF that is more suitable for cryptography key generation due to the limit number of CRPs.

Figure 2.9(a) Proposed PUF bitcell; (b) PUF bitcell adds a sense amplifier for improved output stability.

2.6.2 Strong Monostable PUF for Lightweight Authentication

The fundamental concept of response bit generation mechanism of the monostable PUF in [12] is to extract the reliable response bits by comparing the voltage trip point of different diode-clamped inverters. V_{trip} is also known as the switching threshold voltage of inverter. It can be obtained at $V_{out} = V_{in}$ on the Voltage Transfer Curve (VTC), as shown in Figure 2.10(a).

Figure 2.10(b) depicts the diode-clamp inverter employed in the monostable strong PUF circuit. Instead of using normal inverter, the current of diode-clamped inverter is greatly limited by the two diode-connected transistors, MP_2 and MN_2 , which forces MP_3 and MN_3 to work in the subthreshold region. The subthreshold current is:

$$I_D = \mu C_{ox} \left(\frac{W}{L} \right) (m-1) V_T^2 e^{\frac{V_{gs}-V_{th}}{mV_T}} \left(1 - e^{-\frac{V_{ds}}{V_T}} \right) \quad (2.10)$$

where μ , C_{ox} , W , L , m , and V_T are the charge carrier's mobility, gate oxide capacitance, channel width and length, subthreshold slope factor and thermal voltage, respectively.

Figure 2.10(a) Inverter Voltage Transfer Curve (VTC) and Trip point; (b) Schematic of diode-clamped inverter.

The term $e^{-\frac{V_{ds}}{V_T}}$ in (2.10) is small enough to be negligible when $V_{ds} > 150\text{mV}$. So the current can be simplified to:

$$I_D = \mu_n C_{ox_n} \left(\frac{W_n}{L_n} \right) (m_n - 1) V_T^2 e^{\frac{V_{trip} - V_{thn}}{m_n V_T}} = \mu_p C_{ox_p} \left(\frac{W_p}{L_p} \right) (m_p - 1) V_T^2 e^{\frac{V_{dd} - V_{trip} - V_{thp}}{m_p V_T}} \quad (2.11)$$

V_{trip} can be solved by:

$$V_{trip} = \frac{m_n V_{dd} + 2m_n V_{thp} + 2m_p V_{thn} - m_n m_p V_T A}{m_n + m_p} \quad (2.12)$$

where $A = \ln \left(\frac{\mu_n C_{ox_n} \left(\frac{W_n}{L_n} \right) (m_n - 1)}{\mu_p C_{ox_p} \left(\frac{W_p}{L_p} \right) (m_p - 1)} \right)$.

The inverter size can be designed to obtain the ideal value of $V_{trip} = \frac{1}{2}V_{dd}$. However, as μ , C_{ox} , W , L , m , and V_{th} are susceptible to manufacturing process variations, V_{trip} is also sensitive to process variations. Different inverters have different V_{trip} 's due to the inevitable and unpredictable process variations, even though the transistors of the inverters are sized equally.

The structure of the monostable strong PUF is shown in Figure 2.11. The two independent stages are connected in series. Stage 1 and Stage 2 both consist of 16 multiplexer blocks with a pool of 16 inverters each. However, the inverters in Stage 1 has a feedback loop but not those of Stage 2. As a single-pole inverter ring is inadequate to meet the oscillation criteria, the output of Stage 1 is monostable with a stable output voltage centred at V_{trip} of the “large” inverter. The variations of switching threshold voltages V_{trip} due to the process variation of inverters in Stage 1 and Stage 2 constitute the main and secondary entropy sources. Unlike traditional PUFs, monostable PUF has the advantage that it only has one stable state. This monostable state can tolerate certain amount of noise fluctuations because of the negative feedback. Any small increase in V_{in} reduces V_{out} while the feedback line works against it (positive slope) to retard such a change. This monostable strong PUF exploits the nonlinear relationship between the trip point of diode-clamped inverter and fabrication process dependent physical parameters to significantly increase the complexity of modeling a factorially large number of trip-points of single inverter rings composed by different combinations of parallelly connected diode-clamped inverters. The low power/energy consumption makes it feasible for integration into resource-constrained designs, e.g., passive RF tags.

Figure 2.11 Architecture of the monostable strong PUF design.

2.7 Machine Learning Attack

PUFs have great promise as hardware authentication primitives due to their physical unclonability and high resistance to reverse engineer the circuit to recover the challenge-response mapping mechanism. Although it is difficult to hack the PUF directly, the PUF structure can be mathematically modelled and cracked by machine learning (ML) algorithms. Machine learning based modelling attack targets strong PUFs. Strong PUFs are distinguished by an exponential large number of CRPs. The secrecy of strong PUFs entirely relies on the unpredictability of the CRP behaviour. Machine learning based modelling attacks have been the most promising attack methods for compromising a strong PUF, which can be trained with a subset of known CRPs to predict the output response. Various machine learning algorithms have been tested to replicate the results and investigate the predictability of the CRPs to evaluate the secrecy of PUF. Support Vector Machine (SVM) and Evolution Strategies (ES) are two simple, efficient and powerful machine learning algorithms that have been widely used to evaluate the modelling complexity of strong PUF.

2.7.1 Support Vector Machine

SVM is proposed by Vladimir N. Vapnik [43] in 1963. Since then, it has undergone extensive research. Until 1990s, SVM has become practically useful in many real applications, and have popularly used for binary classification. SVM model is built based on the separation of data points in space. SVM constructs a separating hyperplane which is able to divide all collected data points into two classes [44]. However, there are infinite possible hyperplanes to do the data classification, as shown in Figure 2.12. SVM will choose the hyperplane with the largest margin of the minimum distance to the data points. This hyperplane is called the decision boundary, which can be expressed by the following equation:

$$\beta^T \cdot x + \beta_0 = 0 \quad (2.13)$$

where β is the vector parameter which is perpendicular to the hyperplane, x is the input dataset and β_0 is the bias. The top boundary is the positive hyperplane ($\beta^T \cdot x + \beta_0 = 1$) and the bottom boundary is the negative hyperplane ($\beta^T \cdot x + \beta_0 = -1$).

The margin size M or the distance between the top and bottom hyperplanes can be calculated by using the equation below:

$$M = \frac{2}{\|\beta\|} \quad (2.14)$$

Figure 2.12 The working mechanism of SVM.

To maximize the margin size M , the problem can be further simplified to:

$$\begin{aligned} \min_{\beta, \beta_0} J(\beta) &= \frac{1}{2} \|\beta\| \\ \text{subject to } y_i(\beta^T \cdot x_i + \beta_0) &\geq 1 \quad \forall i \end{aligned} \quad (2.15)$$

where y_i is the labels or the output of the training datasets.

This equation can be solved by using Lagrange multipliers to obtain the vector parameter β and the bias β_0 .

2.7.2 Evolution Strategies

ES are probabilistic optimization methods which emulate natural evolution. They were developed in 1970s by Rechenberg [45] and Schwefel [46]. ES are inspired by the mechanism of biological evolution, such as reproduction, mutation, recombination and selection. A set of candidate solutions is created randomly to maximize the quality function. The fitness function is determined and used as the basis for the process of selection. Some better candidates are selected from next generation based on the fitness function. The individuals are determined depend on their quality by parent selection such that good quality individuals become parent of the next generation. If an individual is selected as a parent, it undergoes variation to create offspring. This is achieved by applying the technique of recombination and/or mutation. Recombination is represented by a binary operator. This operator can be applied to two or more selected parents and it generates one or more new children as a result. After executing this recombination or mutation, it will generate a set of new candidates. This iterative process repeats until no further improvement can be achieved based on the fitness function or a maximum number of generations is reached. ES is proved to be more efficient compared to other iterative methods on a large number of test samples. It is capable of solving high dimensional and multimodal problems subject to linear and nonlinear constraints [47]-[49]. Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is the de-facto evolutionary optimization strategy. It works by iteratively updating the covariance

matrix of a multivariate normal distribution of the mutated population. CMA-ES is especially successful in solving badly conditioned, multimodal and noisy problem with high-dimensional rugged search landscape. CMA-ES is employed as a machine learning attack method to the proposed PUF in the experiments of Chapter 4.

Chapter 3

1024-Bit Dual Port SRAM-Based PUF

3.1 Introduction

PUF is believed to be a promising candidate to efficiently tackle the safety issues that long persecuted the EEPROM based key storage component in computers. Memory subsystems have been widely used in computing devices and systems ranging from lightweight RFID to complicated server data storage system. For this reason, memory is among the most popular embedded blocks in computing systems that have been used for PUF construction to generate device-specific keys for security applications against counterfeiting, ID theft and tampering. SRAM PUF is a hardware security primitive that can leverage on existing memory resources instead of introducing dedicated cells.

Most embedded memories are designed with 6T SRAM as shown in Figure 2.12(a). It has one shared access port for either read or write operation. This access mode of data is designated as one read-write (1RW). As supply voltage scales with technology, single port access limits access speed and handicaps memory management to offer sufficient flexibility for high speed System-on-Chip (SoC) operation. Some optimized multiple-port SRAM structures are explored to support simultaneous read and write accesses as shown in Figure 3.1. An eight transistors (8T) SRAM cell and a ten transistor (10T) SRAM cell are shown in **Error! Reference source not found.**(b) and **Error! Reference source not found.**(c), respectively. Both designs employ one additional port for read operation, so that the write and read operations could be carried out simultaneously without interference. This is known as 1R1W. Another 8T DP-SRAM is depicted in **Error! Reference source not found.**(d). It has two additional transistors to provide two separate ports that are capable of two concurrent read-write (2WR) operations. Compared with the conventional 6T SRAM structure, 8T DP-SRAM has one additional pair of access transistors (M_7 and M_8), one additional word-line (WL_A) and one additional pair of complementary bit-lines (BL_A and BL_{AB}). Both ports of A and B are available to separate reading and writing operations, it is designated as 2RW. DP-SRAM has more capability and flexibility for high speed data access which is widely used as an interface between macro-blocks in SoC.

Figure 3.1 (a) 6T-SP-SRAM Cell (1RW); (b) 8T-DP-SRAM Cell (1R1W); (c) 10T-DP-SRAM cell (1R1W); (d) 8T-DP-SRAM (2RW) [13].

Variants of other volatile and non-volatile memory exploit the same principle of metastable power-up state of two cross-coupled inverters to generate process variation induced random bits. The advantages of this approach are (1) low design complexity, (2) low area cost, (3) high reliability and (4) low mutual information. To aid the reliable management of sensitive data in electronic computing without introducing additional dedicated cells, it is desirable to extend the ubiquity of memory for both security and storage applications [17]. However, reusing the memory for PUF application is not without problems. One major drawback of using conventional memory circuitry (e.g., SRAM) for key generation purpose is it may help adversaries to exploit data remanence effect on SRAM cells to perform fault injection attack so as to obtain the secret bits preserved in the memory based PUF. Another problem is while process variations of devices is welcoming for PUF construction, they increase the yield loss when the devices are used as ordinary storage elements.

3.1.1 Power Consumption and Security Issues of Data Buffering

The working mechanism of a conventional dual-mode memory-based PUF system is shown in Figure 3.2.2. Most of the memory-based PUFs are constructed based on either volatile or non-volatile memory, the key generation process in the conventional memory-based PUFs will inevitably overwrite temporary or permanent useful data that are already stored in the memory due to the power-up reset operation. To preserve the

original useful data that are stored in the SRAM cell, a buffer storage is required to store the data temporarily until the end of the key generation, and the buffered data need to be written back to the original memory cell eventually [29]. This additional data buffering and write back operations not only introduce undesired area/performance/power overheads, but also create potential security vulnerability to fault-injection attacks. Such attack, as introduced in [51], utilizes the remanence decay side channel information by iteratively generating and collecting responses from the SRAM cells with different power-off time to exhaustively find the key. Finally, for design simplicity and cost consideration, most SRAM arrays are powered by one single power supply. The entire SRAM block need to be powered on and off to generate the response bits from a small number of cells that are addressed by the challenges, it will consume a huge amount of spurious power and reduce the useful lifespan of the memory. In recent years, some new structures of memory circuit are designed to improve the functionality of PUF. For example, a ten-transistor subthreshold SRAM PUF is designed in [52] to achieve low power consumption. Nevertheless, it still needs power-up reset to generate the random response bits.

Figure 3.2. Working mechanism of a conventional memory based PUF system [30].

3.1.2 Design Conflicts for Dual-Mode Applications and Low Entropy of SRAM Cell

Random variations can cause a significant mismatch that are largely responsible for the poor yield of the SRAM arrays in scaled technologies [53-55]. On the contrary, random variations are desired for PUF application to produce reliable and random response bits.

Some new optimization strategies in circuit level are investigated in [55] to adequately satisfy the quality requirements for both memory and PUF modes of operation, such as Work-Line Voltage Modulation (WLM), Dynamic Voltage Scaling (DVS), Negative Bit-line (NBL) and Adaptive Body Bias (ABB). Unfortunately, these optimization strategies suffer from increased complexity of the design and added circuit redundancy like the negative voltage generation circuit. It not only increases the level of difficulty on circuit design, but also incurs more power and area overheads and may introduce potential security vulnerabilities.

A 6T SRAM PUF can generate only one response bit per memory cell, which limits the entropy of response generation to at most one independent random response bit per cell. An original 8T dual port SRAM (DP-SRAM) PUF is proposed in [13], but it only generates two response bits per cell. The analyses are based on the simulation results from one single cell, and there is no study on the correlation of the two response bits within and across cells. A current-based PUF is proposed in [31] to produce multiple response bits per cell, but the hardware and power consumptions increase due to the requirement of additional current sensors to read the cells, which is not desirable for SRAM circuit design.

3.2 8T Dual-Port SRAM PUF

Compared with the conventional 6T SRAM structure, an 8T DP-SRAM as shown in Figure 3.3 has one additional pair of access transistors (M_7 and M_8), one additional word-line (WL_A) and one additional pair of complementary bit-lines (BL_A and BL_{AB}). Both ports A and B can be used independently for reading and writing operations. This feature is designated as two read-write (2RW). DP-SRAM optimize memory access and adds flexibility for high speed data access, which is widely used as an interface between macro-blocks in many System-on-Chips (SoCs) to overcome the memory access bottleneck and increase the system throughput.

Figure 3.3 8T DP-SRAM cell.

3.2.1 Access Mode and Issues of DP-SRAM

When both ports A and B of the 8T DP-SRAM are initiated simultaneously, the four different access modes are displayed in Figure 3.4. For simplicity and illustrative purpose, a 2×2 8T DP-SRAM array with two rows (Row1 and Row0) and two columns (Column 1 and Column0) is investigated. The two ports of the SRAM cell are presented as two square boxes: A and B. Each column comprises two pairs of bit-lines: BL_A(B) and BLB_A(B) that are held by ports A and B, respectively. Each row comprises two word-lines: WL_A(B). The active signal is indicated in black color, and deactivated signal is indicated in grey color. The selected memory cells are indicated by solid black boxes. The half selected (only row selected) memory cells are indicated by dotted black boxes. The unselected memory cells are indicated by grey boxes.

Figure 3.4 Four access modes of Dual-Port SRAM.

3.2.1.1 Different Row Access Mode

Figure 3.4(a) and (b) depict two access modes where the word-lines are selected from different rows for port A and port B, and the bit-lines are selected from either different columns or the same column. In these two situations, simultaneous reading and writing of both ports can be carried out without interference and access conflict. Since either WL_A or WL_B of the selected memory cell is active, the two ports will behave like single port access. Simultaneous writing or/and reading for both ports are possible as shown in Table 3.1.

Table 3.1 Different access conditions.

A & B	Different Columns	Same Column
Reading — Reading	No Interference	No Interference
Reading — Writing	No Interference	No Interference
Writing — Reading	No Interference	No Interference
Writing — Writing	No Interference	No Interference

3.2.1.2 Common Row Access Mode

Figure 3.4(c) and (d) depict two access modes where the word-lines are selected from common row for port A and port B, and the bit-lines are selected from either two different columns or the same column. Common row and common column accesses are inhibited due to the abnormal leakage current flowing through the selected memory cell when complementary data is written to the DP-SRAM. This conflict will not occur when the same data is written to the DP-SRAM cell as show in Table 3.2 [55].

Table 3.2 Conflict of different access condition.

A & B	Different Columns	Common Column (Forbidden Mode)
Reading – Reading	No Interference	No Interference
Reading – Writing	No Interference	No Interference
Writing – Reading	No Interference	No Interference
Writing – Writing	No Interference	No Interference (Same data is written into BL and BLB)
		Conflict Occur (Complementary data is written into BL and BLB)

It should be alerted on the stability because the static noise margin of selected cells will be degraded by activating both word-lines together during the read operation. The read-stability of half-selected memory cell needs to be taken care, since this situation can cause read failure similar to the regular read operation [56]. In reality, the write conflict mode is forbidden for memory operation, this contention can nevertheless be utilized for PUF operation to generate the response bits without interfering with or increasing the failure rates of the normal legitimate read and write operations of the cell.

3.2.2 Working Principle of 8T DP-SPUF

Typically, the 6T SRAM PUF suffers from limited extractable entropy and poor robustness of signature. Because the generation of response is dependent on power-up process to generate one response bit per cell and the generated response bit could be altered by the residual charge and minor environmental fluctuation. The proposed 8T DP-SPUF utilizes the forbidden write conflict mode of DP-SRAM to generate a larger population of unique, robust and random response bits from the same number of cells. In PUF operation, WL_A and WL_B are activated concurrently, i.e., WL_A = WL_B = 1. Complementary data are written into the 4 bit-lines of 8T DP-SRAM PUF, i.e., BL_A = 1, BL_B = 0, BLB_A = 0 and BLB_B = 1. The drain-source current flows from the bit-line of port A (BL_A=1), through M₇ and M₅ to the bit-line of port B (BL_B=0). The dotted lines with arrow indicate the paths and directions of the drain-source current. During this process, M₇ is working in the saturation region. The drain-source current can be expressed as follows:

$$I_{DS} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{th})^2 (1 + \lambda V_{DS}) \quad (3.1)$$

where μ_n is the effective charge-carrier mobility, C_{ox} is the gate oxide capacitance per unit area, λ is the channel-length modulation parameter, V_{GS} is the gate-to-source bias, V_{DS} is the drain-to-source bias and V_{th} is the threshold voltage.

M₅ is working in triode region. Its drain-source current can be expressed as follows:

$$I_{DS} = \mu_n C_{ox} \frac{W}{L} \left[(V_{GS} - V_{th}) V_{DS} - \frac{V_{DS}^2}{2} \right] (1 + \lambda V_{DS}) \quad (3.2)$$

Similarly, M₆ is working in the saturation region, with drain-source current given by Equation (3.1). M₈ is working in the triode region, with drain-source current given by Equation (3.2).

3.2.2.1 Word-Line Enabled (Access Transistors)

In the first situation when the word-lines are enabled, the drain-source current flows through nodes Q and QB as shown in Figure 3.5(a). The voltages at Q and QB are identical due to the symmetry of the circuit. Unfortunately, variations between the voltages at nodes Q and QB are always present due to the different strength among the access transistors. The voltage of Q is determined by the relative strength between M_5 and M_7 , whereas the voltage of QB is determined by the relative strength between M_6 and M_8 . If the difference between V_Q and V_{QB} is sufficiently large to make the circuit stable, the generation of response bits relies on the mismatches among transistors M_5 to M_8 . This phase is shown in Figure 3.5(b), where the red dotted line represents V_Q and the blue dotted line represent V_{QB} .

Figure 3.5(a) Word-line enable mode; (b) Response bit determined mainly by the mismatches of access transistors.

3.2.2.2 Word-Line Disabled (Cross-Couple Inverters)

In the second situation when both word-lines are deactivated, the access transistors are turned off. The positive feedback loops formed by INVA and INVB will push the cell towards to a stable state ('0' or '1') depending on the voltage values of nodes Q and QB, as shown in Figure 3.6(a). If the voltages at Q and QB are sufficiently close to $\frac{1}{2}V_{dd}$, the cell is unstable, and will settle to a stable state. In actual implementation, one feedback loop is always stronger than the other due to the mismatch of transistors within the cross-coupled inverters. So, the voltage at node Q will be stabilized because of the different strength of positive feedback loops. As shown in Figure 3.6(b), the red color dotted line represents V_Q and the blue color dotted line represents V_{QB} . If the mismatch of V_Q and V_{QB} is sufficiently small which can be neglected, the positive feedback loop formed by the cross-coupled inverters will stabilize the output state to either '0' or '1' without the activation of word-line. The entropy of DP-SRAM PUF is enriched by the process variability of all the transistors inside the memory cell during these two steps of response generation process.

(a)

(b)

Figure 3.6(a) Word-line disable mode; (b) Response bit determined mainly by mismatches of cross-couple inverters.

3.2.3 Different combinations of Bit-Lines in memory and PUF modes

The four bit-lines (BL_A, BLB_A, BL_B and BLB_B) inside each memory cell are connected to ports A and B. In the write conflict access mode, there are two possible combinations of bit-lines data, as shown in Table 3.3.

Table 3.3 Two different combinations of bit-line in memory mode.

Bit-line	BL_A	BL_B	BLB_A	BLB_B
Combination1	0	1	1	0
Combination2	1	0	0	1

In PUF mode, complementary bits need to be written into the bit-lines located at the left and right sides of the memory cell to produce device signatures. For simplicity and ease of exposition, BL_A/B and BLB_A/B are renamed as BLL_1/2 (located at the left side of the memory cell) and BLR_1/2 (located at the right side of the memory cell). When the word-lines are activated, two pairs of complementary data are written into BLL_1/2 and BLR_1/2. The four possible combinations of complementary data are shown in Table 3.4, which allow a maximum of four response bits per cell to be generated. The corresponding bit generation processes of 8T DP-SPUF are illustrated in Figure 3.7.

Table 3.4 Four different combinations of bit-line in PUF mode.

Bit-line	BLL_1	BLL_2	BLR_1	BLR_2
Combination1	0	1	0	1
Combination2	0	1	1	0
Combination3	1	0	0	1
Combination4	1	0	1	0

Figure 3.7 8T DP-SRAM operation in PUF mode (a) $BLL_1=0, BLL_2=1, BLR_1=0, BLR_2=1$; (b) $BLL_1=0, BLL_2=1, BLR_1=1, BLR_2=0$; (c) $BLL_1=1, BLL_2=0, BLR_1=0, BLR_2=1$; (d) $BLL_1=1, BLL_2=0, BLR_1=1, BLR_2=0$.

3.3 1024 Bits Dual Port SRAM Circuit Implementation

The proposed 1024-bit DP-SRAM array makes use of the write conflict mode to generate unique, distinguishable and random response bits. Reduction of power consumption and elimination of the data remanence decay attacks are achieved by eliminating the power-up operation and requirement of temporary buffer. Each cell has a potential to generate a maximum of four response bits as four different combinations of complimentary data can be used for bit-line conflict writing. The architecture of the proposed 1024-bit 8T DP-SRAM is depicted in Figure 3.8. The memory array containing 1024 memory cells are arranged into a square matrix of 32 rows and 32 columns. Each row consists of two word-lines and each column consists of four bit-lines. A 10-bit maximum length Linear Feedback Shift Register (LFSR) with feedback polynomial of $X^9 + X^6 + 1$ is used to generate the random challenge bits C9-C0 to address different locations of memory cell from an input challenge. In the write conflict access mode, C9-C5 will be launched into the two Word-Line MUX simultaneously to achieve common row access. C4-C0 will be launched into the four Bit-Line MUX simultaneously to achieve common column access. Complementary bits are written into the bit-lines to generate random response bits when the word-lines are activated. The final state of stabilized response bit is unpredictable but reproducible. The layout of

1024 DP-SRAM array is shown in Figure 3.9. It is constructed as symmetrical as possible by manually routing the signal wires to minimize the cumulated time delay differences due to the wire routings in the IC chips.

Figure 3.8 Architecture of the proposed 1024-bit DP-SRAM array.

Figure 3.9 Layout of SRAM cells and array.

3.4 Result and Discussion

In this section, the proposed 1024-bit DP-SRAM PUF is evaluated in terms of correlation, reliability, uniqueness and randomness by Monte Carlo simulation results based on the commercial UMC 65-nm CMOS process technology. The collected PUF responses are analyzed by Matlab program.

3.4.1 Correlation

To achieve multiple bits per cell, the response bits from the four different access modes of the same cell shown in Figure 3.7 should be uncorrelated. Figure 3.10 depicts the Monte Carlo simulation results of twenty cells. No obvious correlation is displayed among these response bits.

Figure 3.10 Response bits generated from the same cell for 4 different access modes.

For further verification, correlation coefficients are calculated by making pairwise comparisons among 256 responses that are produced by writing four different complementary bits (BL1, BL2, BL1B, BL2B=0101, 0110, 1001, 1010) into the same cell. The correlation coefficients are very closed to the ideal value 0 as shown in Table 3.5. It can be concluded that the responses of the four access cases are uncorrelated and each cell is capable of generating up to a maximum of four uncorrelated and distinguishable response bits.

Table 3.5 Correlation coefficients of 4 access cases.

BL1,BL2,BL1B,BL2B	0101	0110	1001	1010
0101	1	0.097	-0.062	-0.017
0110	0.097	1	0.273	0.103
1001	-0.062	0.273	1	-0.094
1010	-0.017	0.103	-0.094	1

3.4.2 Reliability

The reliability is evaluated by calculating the Hamming distance of the responses that are generated from the same instance to a given challenge under different environmental conditions. Voltage fluctuations and ambient temperature variations are the main environmental factors that can cause a reliability degradation. Let R_i be the response of the PUF instance i under the nominal condition. $R_{i,j}$ is the response generated by writing the same set of challenges into the same instance k times under different environmental conditions, where $j = 1, 2, \dots, k$.

$$S = 1 - BER = 1 - \frac{1}{k} \sum_{j=1}^k \frac{HD(R_i, R_{i,j})}{n} \times 100\% \quad (3.3)$$

Figure 3.11(a) Reliability of PUF over voltage range from 1.08V to 1.32V;
(b) Reliability of PUF over temperature range from -40°C to 120°C.

To evaluate the reliability of the PUF against temperature and voltage variations, the reference response is taken from a nominal condition of 27°C and 1.2V. Figure 3.11 delineates the reliability results obtained from 2,300 CRPs when both 6T SRAM and 8T DP-SRAM PUFs are operating at different temperatures and supply voltages. Evidently,

the proposed DP-SRAM PUF exhibits higher reliability than the SP SRAM PUF over a wide range of temperatures (from -40°C to $+120^{\circ}\text{C}$) and variations of supply voltage (from 1.08V to 1.32V).

3.4.3 Uniqueness

Uniqueness is crucial to device identification as it measures how difference are the response bits produced by different PUF instances with identical circuit design to the same challenges. It can be quantified by calculating the inter-die Hamming distance (HD). To achieve the maximum discriminability, when the same challenges are applied to any two PUF instances with the same circuit design, half of the random response bits should be different. Thus, the ideal value of uniqueness is 50%. Let R_u and R_v be an n -bit response generated by the same challenge from two PUF instances. The uniqueness (U) for m different chips is expressed as follows:

$$U = \frac{2}{m(m-1)} \sum_{u=1}^{m-1} \sum_{v=u+1}^m \frac{HD(R_u, R_v)}{n} \times 100\% \quad (3.4)$$

The uniqueness of the 1024-bit DP-SRAM PUF is evaluated by making the pairwise comparison of responses among 50 instances. Figure 3.12 shows the distribution of the fractional HD. The best-fit Gaussian to the histogram has a mean value (μ) of 49.6% and a standard deviation (σ) of 13.6%, which indicates that the uniqueness of the proposed DP-SRAM PUF is close to the ideal uniqueness of 50%.

Figure 3.12 Interdie fractional HD for the proposed 8T DP-SRAM PUF.

3.4.4 Randomness

3.4.4.1 National Institute for Science and Technology statistical tests

Randomness evaluates the capability of a PUF to generate a sequence of uncorrelated response bits with even probability of occurrences of ‘0’s and ‘1’s. The National Institute for Science and Technology (NIST) statistical test suite is used to evaluate the randomness of binary bitstream. 10,000 response bits obtained from the PUF are split into 10 blocks of 1000 bits each. As shown in Table 3.6, ten main statistical tests were performed. To pass each test, the P-value obtained by applying a chi-square test at 1% confidence level should be larger than 0.01 and the minimum proportion of binary sequence that passed the test is 8/10. The results indicate that the response bits generated by the DP-SRAM PUF passed these NIST randomness tests.

Table 3.6 Results of ten main NIST tests.

Statistical	Stream Length	Runs	Potion	PASS?
Frequency	1000	10	9/10	Yes
Block Frequency	1000	10	9/10	Yes
Cumulative Sums(forward)	1000	10	8/10	Yes
Cumulative Sums(backward)	1000	10	8/10	Yes
Runs	1000	10	10/10	Yes
LongestRun	1000	10	10/10	Yes
FFT	1000	10	9/10	Yes
Serial(forward)	1000	10	8/10	Yes
Serial(backword)	1000	10	10/10	Yes

3.4.4.2 Entropy

The randomness of memory PUF can also be evaluated theoretically by entropy, which is defined as follows:

$$\hat{E} = -\sum_{x \in \mathcal{X}} \log(P_r(X))P_r(X) \quad (3.5)$$

where X is a random event from the set $\mathcal{X}$. There are 16 possible combinations of output pattern as shown in Table 3.7.

Table 3.7 Percentage of occurrences of all combinations of output patterns for the four different cases.

BL	Bit Pattern															
0101	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
0110	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1010	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1001	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
%	6.19	6.28	6.22	6.17	6.18	6.25	6.29	6.34	6.36	6.23	6.19	6.3	6.25	6.2	6.3	6.25

The average entropy of this DP-SRAM PUF for the four different input cases was estimated directly by Eq. (3.5) as follows:

$$\begin{aligned}
\hat{E} = & -[\log \Pr(0000) \times \Pr(0000) + \log \Pr(0001) \times \Pr(0001) + \log \Pr(0010) \times \Pr(0010) \\
& + \log \Pr(0011) \times \Pr(0011) + \log \Pr(0100) \times \Pr(0100) + \log \Pr(0101) \times \Pr(0101) \\
& + \log \Pr(0110) \times \Pr(0110) + \log \Pr(0111) \times \Pr(0111) + \log \Pr(1000) \times \Pr(1000) \\
& + \log \Pr(1001) \times \Pr(1001) + \log \Pr(1010) \times \Pr(1010) + \log \Pr(1011) \times \Pr(1011) \\
& + \log \Pr(1100) \times \Pr(1100) + \log \Pr(1101) \times \Pr(1101) + \log \Pr(1110) \times \Pr(1110) \\
& + \log \Pr(1111) \times \Pr(1111)] = 3.99998
\end{aligned}$$

which turns out to be equal to 3.99998, which is almost equal to the ideal four independent response bits. In other words, the dynamic predictability of the cell is $2^{-3.99998}$ which indicates a very high randomness.

3.5 Conclusion

This chapter has presented a new method to generate multiple response bits per cell from an 8T-DP-SRAM. Simulation results using UMC 65nm process show that the proposed 8T-DP-SPUF has better reliability than the traditional 6T-SP-SPUF. The pairwise correlation coefficient of the 4 responses is evaluated to be very close to the ideal value 0, indicating that there is no correlation among the four response bits per cell. The theoretical entropy is also elaborated. The result shows that entropy is equal to 4 with accuracy up to 4 decimal places, indicating that the four response bits per cell are independent. As the DP-SRAM does not need to be reset to generate the response, data buffering during mode switching is limited to only those cells that are addressed by the challenge, which greatly minimizes its risk of side-channel attacks over the power-up reset based memory PUFs. Such features are important to multiple access memory management for enhanced data integrity.

Chapter 4

A Monostable Strong PUF With Enhanced Uniformity and Reduced System Bias

4.1 Introduction

With increasing number of sensors being embedded into electronic devices, wearables, biomedical patches, smart home appliances and infotainment systems to gain connectivity to transfer and communicate valuable and sensitive information through the internet, there is a pressing need for semiconductor device and system manufacturers to add security functionalities into their products for a safe and trustworthy Internet-of-Things (IoT). Key management, key provisioning and authentication are three important considerations for the pervasive deployment of security solutions. With unclonability arises from the hardness of reproducing the multitude of uncontrollable manufacturing process parameters, PUF is touted as a promising new technology for building hardware root of trust with lightweight crypto for robust authentication between devices, and establish trust for data being read or written into an IoT endpoint without the need for stored keys, expensive secure memory and anti-tamper technologies. Over years, a vast multifarious of PUFs have been developed [57]-[60]. Some of which have been successfully commercialized, notably for generating the essential part of a self-sovereign identity, the private key, in a tamper-proof manner. Using PUF technology, the private key becomes volatile as it will only be generated upon demand and vanish after use. Self-sovereign identity allows sharing of anonymized information for health, credit or services when needed without having to rely on a third party “identity providers” to store and share personal data. It turns out that the use of real identity like fingerprint or facial features as self-sovereign identity cannot be easily replaced once they are compromised. Alternative biometric solutions [60], [61] that intertwine PUF as the fingerprint of the hardware owned by the user with his traditional biometrics catalyse completely new ways of attestation for human-machine interactions. Although the advent of PUF has broadened the scope of possible applications beyond simple device identification and authentication, weak PUFs such as memory PUFs [17], [58] and the multi-port memory PUF presented in the previous chapter, despite the extended entropy of multiple response bits per cell, are still very much constrained by the possibility of exhaustive read out due to its limited number of CRPs. Strong PUFs with exponentially large number of challenges are more versatile to guarantee challenge freshness against protocol attacks for the expanded scope of authentication and attestation. Unfortunately, most strong PUFs like arbiter PUF [57], [35] are delay based. Arbitration of the race between two

symmetric timing paths suffers from the inherently poor trade-off between uniqueness and reliability. The linear additive delay structure of arbiter PUF is also vulnerable to modelling attack by machine learning algorithms [62], [63]. Although its modelling complexity can be increased by adding “nonlinear” effects such as XOR [23], lightweight secure [35] and feed-forward arbiter [63] PUFs, they tend to sacrifice circuit area, energy consumption or reliability.

In this chapter, a monostable diode-clamped inverter-based strong PUF is proposed. It is an enhancement to the previously proposed design [13] of our group with significantly improved uniformity, raw response reliability and higher modelling complexity. It achieves better worst-case reliability with a smaller fraction of pruned CRPs than [13] over a broader range of environmental variations. Unlike the static weak monostable PUF [64], the proposed design is a strong monostable PUF with exponential number of CRPs even after pruning the potentially unreliable responses. The rest of the chapter is organized as follows. Section 4.2 introduces the basic structure of the monostable PUF with elaboration of the enhancements made. Session 4.3 presents the measurement results and discussions. Section 4.4 concludes the work presented in this chapter.

4.2 Enhancement of Monostable Strong PUF

A monostable state can be obtained by connecting the output of an inverter to its input, as shown in Figure 4.1(a). As the loop delay is insufficient for the signal to make a full transition, the output voltage V_M will rest at an intermediate voltage between V_{dd} and ground given by the intersection of the Voltage Transfer Characteristic (VTC) of the inverter and the line $V_{out} = V_{in}$, as shown in Figure 4.1(c). At this point, both the PMOS and NMOS transistors are saturated since $V_{DS} = V_{GS}$, and the drain current of the inverter is given by:

$$I_D = \frac{\mu C_{ox}}{2} \left(\frac{W}{L} \right) (V_{GS} - V_{th})^2 (1 + \lambda V_{DS}) \quad (4.1)$$

where μ , C_{ox} , W , L , V_{th} , and λ are the charge carrier’s mobility, gate oxide capacitance, channel width and length, threshold voltage and channel modulation parameter, respectively.

Figure 4.1(a) Basic inverter with feedback; (b) Diode-clamped inverter with feedback; (c) Feedback stabilized voltage V_M of diode-clamped inverter; (d) Widening V_M deviation from half V_{dd} .

Ignoring the effect of λ , V_M can be calculated by equating the drain currents of NMOS and PMOS transistors.

$$\frac{1}{2} K_n (V_M - V_{thn})^2 = \frac{1}{2} K_p (V_{DD} - V_M + V_{thp})^2 \quad (4.2)$$

$$V_M = \frac{V_{DD} - V_{thp} + A V_{thn}}{1 + A} \quad (4.3)$$

where $K_{n,p} = \mu_{n,p} C_{ox} \left(\frac{W}{L} \right)$ and $A = \sqrt{\frac{K_n}{K_p}}$.

The transistors can be sized to make V_M equal to half V_{dd} . However, as the PMOS and NMOS transistors are both partially switched on at this voltage, a large crowbar current will flow from V_{dd} to ground. To limit the crowbar current, a diode clamped inverter (DC-INV), as shown in Figure 4.1(b), is used instead to generate this monostable output [12]. The diode-connected NMOS (MN2) and PMOS (MP2) transistors force MP3 and MN3 to conduct in the subthreshold region. The subthreshold current is given by:

$$I_D = K(m-1)V_T^2 e^{\left(\frac{V_{gs} - V_{th}}{mV_T} \right)} \left(1 - e^{-\frac{V_{ds}}{V_T}} \right) \quad (4.4)$$

where m and V_T are the subthreshold slope factor and thermal voltage, respectively. V_M can also be made equal to half V_{dd} by sizing the transistors appropriately.

As V_M lies nearly midway in the undefined logic region between the maximum input low V_{IL} and minimum input high V_{IH} of an inverter, it has high noise immunity. Small temporal noise may change V_M momentarily, but the negative feedback will restore V_M once the noise is removed. By connecting the output V_M of the single inverter ring (INV1) to an inverter (INV2), the output of INV2 will be skewed further away from the monostable voltage V_M . INV2 overcomes the ability of noise to drive V_M to a metastable state that has a higher propensity to resolve to a stable logic '0' or '1' state with time or by further boosting. Due to the manufacturing process variations, each transistor has a slightly different V_{th} . The subtle difference in V_M of the feedback stabilized inverter INV1 and the trip point V_{trip} of the open-loop inverter INV2 will lead to random and unpredictable resolution of the perturbed state, even if the inverters are identically sized.

(a)

(b)

(c)

Figure 4.2(a) Enhanced monostable strong PUF circuit; (b) Buffer circuit and its gain; (c) Filter circuit and VTC curves of normal, low and high V_{trip} inverters.

To extract this robust and reliable entropy source for PUF response, the original monostable strong PUF [12] proposed by our group uses two independent stages of parallelly connected inverters, one with and one without loop back, as described in Chapter 2. The inverters in these two stages are selectable by two disjoint subsets of input challenge bits. The idea of having selectable inverters connected in parallel is to average out the bias of a single fixed inverter so that its output voltage will assume a Gaussian distribution over all combinations of input challenge. An additional stage of fixed parallelly connected inverters and a final buffer stage are used to amplify any unresolved metastable voltage to a distinct binary output.

In [12], the inverters in Stages 1 and 2 are independent. Although the spread of V_M distribution is broadened by two independent Gaussians, their small biases add up, resulting in less ideal uniformity. Its Stage 3 is also independent of the input challenge. This parallel inverter stage merely operates as an open loop amplifier, which will not mitigate the bias or increase the modeling complexity. To improve the uniformity and modeling complexity without unduly increasing the hardware cost, an improved version of the original architecture is presented in this section by sharing the 512 DC-INVs in Stages 1 and 2 and making them selectable by the same 128-bit input challenge without conflict. As shown in Figure 4.2(a), the merged Stages 1 and 2 consists of 32 blocks of DC-INVs. Each block contains a pool of 16 DC-INVs shared between two demultiplexer-multiplexer (DeMUX-MUX) pairs. These two DeMUX-MUX pairs in each block are controlled by the same 4-bit challenge, and different 4-bit challenges for different blocks. A 128-bit maximum length LFSR with feedback polynomial of $x^{128} + x^{127} + x^{126} + x^{121} + 1$ is used to generate n 128-bit internal challenges for an n -bit response to an input challenge (seed).

The first DeMUX-MUX pairs of all blocks share a common input V_{1in} and a common output V_{1out} . Similarly, the second DeMUX-MUX pairs of all blocks also share a common input V_{2in} and a common output V_{2out} . In addition, the input V_{1in} and output V_{1out} are connected together. V_{1out} is also connected to V_{2in} but the input V_{2in} and output V_{2out} of the second DeMUX-MUX pair are not connected. By applying a 128-bit challenge to the 32 inverter blocks, one of the 16 DC-INVs from each block will be selected to form a parallel connection of 32 feedback stabilized inverters for Stage 1. Meantime, a different DC-INV is also selected from each block by the same input challenge to the second DeMUX-MUX pair to make a parallel connection of 32 open-loop DC-INVs for Stage 2. To avoid the same inverter from being selected by the two DeMUX-MUX pairs within the same block, the order of connections of the 16 DC-INVs to the two DeMUX-MUX pairs are reversed. To increase the modelling complexity, the connection sequence is also cyclically shifted across blocks. As shown in Table I, the i -th inverter of block k is connected to the j -th pin of its first DeMUX-MUX pair and the $(17 - j)$ -th pin of its second DeMUX-MUX pair for all $i = 1, 2, \dots, 16$ and $j = (i - k + 1) \bmod 16$. In this way, Stages 1 and 2 share the same pool of inverters to better average out the bias over all

combinations of input challenge. Unlike the linearly accumulated delay parameters obtained from independently measured bit slices before the arbitration for most delay-based strong PUF, this redesigned shared parallel inverter connection scheme reduces spatial correlation of inverters across blocks for the same 4-bit input challenge. It reduces the prediction accuracy and success rate of black-box modelling attacks with existing powerful machine learning algorithms.

Table 4.1 Connection sequence of the shared inverters between the two DeMUX-MUX pairs in different blocks.

	Inv. no.	MUX pin no. Block 1	MUX pin no. Block 2	...	MUX pin no. Block 15	MUX pin no. Block 16
1 st MUX pair	1	1	16	...	3	2
	2	2	1		4	3
	3	3	2		5	4
	⋮	⋮	⋮		⋮	⋮
	15	15	14		1	16
	16	16	15		2	1
2 nd MUX pair	16	16	1	...	14	15
	15	15	16		13	14
	14	14	15		12	13
	⋮	⋮	⋮		⋮	⋮
	2	2	3		16	1
	1	1	2		15	16

Stage 3 is also modified to reduce the bias of fixed parallel inverters by eight parallel blocks of 16 selectable inverters each. One inverter from each block is selected by 32 arbitrary bits of the 128-bit challenge. A static CMOS inverter is used as a buffer to ensure that any metastable response, if present after State 3, will be driven to a distinct logic ‘0’ or ‘1’ state. The inverter buffer has the maximum gain in its narrow transition region as shown in Figure 4.2(b). The response bit generated is post-processed by a filter to determine its reliability, where the output of an inverter with high V_{trip} (V_{trip-H}) is compared against the output of an inverter with low V_{trip} (V_{trip-L}) to produce a ‘Valid’ flag. The inverter with V_{trip-H} is made up of a PMOS with low V_{th} and an NMOS with high V_{th} . The inverter with V_{trip-L} is composed by a PMOS with high V_{th} and an NMOS with low V_{th} . If V_{3out} falls in the region between V_{trip-L} and V_{trip-H} , $Valid = ‘1’$ and V_{out} after the buffer will be passed as a reliable response bit. Otherwise, $Valid = ‘0’$ to flag potentially unreliable response bit. The filter circuit and the VTCs of normal, low and high V_{trip} inverters are shown in Figure 4.2(c).

The output voltage distribution at each stage is obtained by applying 10,000 challenges to the proposed monostable PUF. A narrow Gaussian distribution of V_M centered closed to half V_{dd} ($V_{dd} = 1.1$ V) is shown in Figure 4.3(a) for the output of Stage 1. The distribution is broadened by Stage 2 to Figure 4.3(b) without affecting its means by comparing the average V_M of the parallel feedback stabilized inverters selected by the input challenge against the average V_{trip} of a different set of parallel inverters selected by the same challenge in Stage 2.

Figure 4.3 Monte Carlo simulation results of output voltage distributions at (a) Stage 1, (b) Stage 2, (c) Stage 3, and (d) final V_{out} .

Stage 3 reduces the population in the undefined region. More outputs are pushed towards the weak ‘1’ or ‘0’ regions by amplifying their voltage difference from the average V_{trip} obtained by the parallel inverters selected by a subset of 32 arbitrary challenge bits. The output distribution of Stage 3 is shown in Figure 4.3(c). The final V_{out} after the strong buffer has a symmetric distribution of two discrete logic voltages. This design has removed the systemic bias between Stage 1 and Stage 2 of [12]. As all the inverters shared by Stages 1 and 2 are equally probable to be selected into either stage, there is an equal probability of V_{trip} of Stage 2 falling above or below V_M in Stage 1. This ensures that the means of the broaden Gaussian will not be skewed before the metastable states are resolved by Stage 3 and the buffer. Instead of fixing the trip point of parallel inverters

in Stage 3, the enhanced design makes Stage 3 dependent on a quarter of input challenge bits to minimize the probability of skewed distribution of final response bits as the fixed V_{trip} of original Stage 3 is more likely to be skewed for some chips and become higher or lower than most outputs of Stage 2.

Figure 4.4 Layouts of (a) the 128-bit monostable PUF, and (b) a DC-INV.

4.3 Results and Discussions

A 128-bit monostable PUF circuit was implemented and successfully fabricated in TSMC 40nm CMOS process technology. The layout of the 128-bit monostable PUF and DC-INV are shown in Figure 4.4(a) and (b), respectively. The core size is $480\mu\text{m} \times 30\mu\text{m} = 14400 \mu\text{m}^2$. As this chip is fabricated through the multi-project wafer program shuttle service under the auspices of MediaTek Pte. Ltd. The die area is shared between three projects and the number of good packaged chips given to the author for testing is less than five. The number of samples is insufficient to provide informative statistical evidence of uniqueness across dies. Hence uniqueness is evaluated by post-layout Monte-Carlo simulation. Except uniqueness, the following quality metrics are evaluated by responses measured from the fabricated chips.

4.3.1 Uniformity

Uniformity can be evaluated by the percentage of ‘1’s in the responses generated from the same PUF chip, i.e.,

$$Uniformity = \frac{1}{n} \sum_{i=1}^n r_i \times 100\% \quad (4.5)$$

where r_i is the i -th response bit generated by applying n random challenges to a PUF.

The ideal value of uniformity is 50%. The worst and average uniformity of the proposed monostable PUF are 54.6% and 50.86%, respectively by counting the number of 1's in 80,000 response bits. Comparing with the worst and average uniformity of 73.08% and 54%, respectively of the design in [12], the worst uniformity has been improved significantly by 18.4% and our average uniformity is much closer to the ideal value than that of [12].

4.3.2 Uniqueness

Uniqueness indicates the distinguishability of a PUF from other PUF instances. It can be calculated by the fractional Hamming distance (HD) among the responses generated from different PUF instances. Let R_u and R_v be two n -bit responses generated by two PUF instances with the same challenge and m be the number of distinct challenges. $HD(R_u, R_v)$ is the Hamming distance between two binary strings, R_u and R_v . Then

$$Uniqueness = \frac{2}{m(m-1)} \sum_{u=1}^{m-1} \sum_{v=u+1}^m \frac{HD(R_u, R_v)}{n} \times 100\% \quad (4.6)$$

Figure 4.5 Distribution of inter-die fractional harming distance of 5000 responses from 80 PUF instances.

Due to limited number of dies for shared chip area from different projects in this MPW run, the distribution of fractional Hamming distance plotted in Figure 4.5 is obtained by simulation with 100 random challenges applied to 80 different PUF instances. The distribution can be fitted by a Gaussian curve with mean $\mu = 0.5002$ and standard deviation $\sigma = 0.0595$. The uniqueness is 50.02%, which is equal to the ideal uniqueness with accuracy up to one decimal point.

4.3.3 Reliability

Reliability measures the capability of a PUF to reproduce the same response from the same challenge in different environmental conditions. The reliability S can be evaluated by the Bit Error Rate (BER), which can be calculated as follows:

$$S = 1 - BER = 1 - \frac{1}{k} \sum_{t=1}^k \frac{HD(R_0, R_t)}{n} \times 100\% \quad (4.7)$$

where R_0 denotes the referenced n -bit PUF response generated at the nominal condition, and R_t are the n -bit PUF response measured at other time $t = 1, 2, \dots, k$ where the operating condition may have changed.

Figure 4.6 Reliability against (a) temperature and (b) voltage variations.

Response bits may flip due to the fluctuation of ambient temperature and supply voltage. To evaluate the reliability of PUF circuit against temperature and voltage variations, the reference response is taken at 27°C and 1.1 V. Figure 4.6(a) and (b) show the reliability measured from a set of CRPs generated at different temperatures and supply voltages with and without filtering out potentially unreliable responses, respectively. The worst BER of the proposed PUF are 6% with temperature varying from -40 °C to 120 °C and 2% with supply voltage ranges from 0.9V to 1.3V without filtering, which are better than the raw BER of 7% and 4%, respectively reported in [12]. The worst BERs are

further improved to 1% at (1.1V, 120°C) and 0.054% at (0.9V, 27 °C) by discarding 25% and 14.7%, respectively of potentially unreliable response bits. These are again better than the BERs of 1.2% at (1.1V, 90°C) and 0.82% at (1.3V, 27 °C), respectively with 26.9% of discarded response bits in [12].

4.3.4 Randomness

Randomness of PUF responses can be evaluated by the National Institute of Standards and Technology (NIST) statistical test suite [65]. The responses generated by the proposed PUF have passed all the NIST tests, as shown in Table 4.2.

Table 4.2 NIST test of bit sequences generated by the proposed PUF.

Statistical	Length	Runs	Proportion	P-value	PASS?
Frequency	20000	50	48/50	0.0401	Yes
Block Frequency	20000	50	48/50	0.1088	Yes
Cumulative Sums(forward)	20000	50	49/50	0.0589	Yes
Cumulative Sums(backward)	20000	50	50/50	0.2305	Yes
Runs	20000	50	50/50	0.5341	Yes
LongestRun	20000	50	50/50	0.6121	Yes
Rank	20000	50	50/50	0.2407	Yes
FFT	20000	50	48/50	0.0112	Yes
Approximate Entropy	20000	50	49/50	0.1306	Yes
Non-Overlapping Temp.	20000	50	49/50	0.3555	Yes
Overlapping Template	20000	50	48/50	0.2172	Yes
Universal	20000	50	50/50	0.7869	Yes
Serial(forward)	20000	50	50/50	0.4507	Yes
Serial(backward)	20000	50	50/50	0.2139	Yes
Linear Complexity	20000	50	49/50	0.3322	Yes

4.3.5 Security Against Machine Learning Attacks

Two powerful and popular machine learning algorithms, SVM [62] and CMA-ES [63], are used to train the black box PUF model to evaluate the resilience of the proposed PUF against machine learning attacks. 20,000 CRPs collected and grouped into 20 groups with 1,000 CRPs in each subset. 10 groups are used as training data and 10 groups are used for testing the trained model. The predication accuracies of SVM and CMA-ES are shown in Figure 4.7(a) and (b), respectively. The predication accuracies are 50% for SVM and 49.5% for CMA-ES with 1000 training CRPs, which are almost equivalent to the perfect wild guess probability. These probabilities converge to 49% for SVM and 48.5% for CMA-ES with more than 3000 training CRPs. The design of [12] has the

correct prediction rates of 50% for SVM and 47% for CMA-ES with 1000 training CRPs, and they converge to 48% for SVM with more than 9000 training CRPs, and 45% with more than 4000 training CRPs. The complexity of modelling the trip point voltages has been increased significantly with different inverter connections to the DeMUX-MUX pairs in different blocks of the proposed PUF design, making it harder to train the model accurately.

Figure 4.7 The prediction accuracy by SVM and CMA-ES for the proposed 128-bit enhanced PUF and the original 128-bit DC-INV PUF.

4.3.6 Comparison with Other Strong PUFs

Table 4.3 summarizes the figures of merit of the proposed work and other state-of-art strong PUF designs. Overall, the proposed 128-bit monostable PUF has the largest CRP space before and after pruning, better uniformity than [12] and the best raw response reliability over a wider range of operating conditions among the strong PUFs in comparison. The power consumption and energy per bit are comparable to other designs implemented in the same technology node. Its worst uniformity is 54.6%, which is much better than 73.08% of [12].

Table 4.3 Comparison with state-of-art strong PUFs.

Metric	This work [12]		[66]	[67]	[68]
Technology(nm)	40	40	180	130	40
Number of CRPs	3.4×10^{38}	3.4×10^{38}	4.3×10^9	3.7×10^{19}	5.5×10^{28}
Temperature range (°C)	−40 ~ 120	−40 ~ 90	−40 ~ 125	−20 ~ 80	−25 ~ 125
Voltage range(V)	0.9 ~ 1.3	0.9 ~ 1.3	1.5 ~ 2.1	1.08 ~ 1.32	0.7 ~ 1.2
Power (μW)	8.52	7.71	-	0.068	28.4
Bit rate (Mb/s)	0.5	0.5	-	0.006	1.6
Energy per bit (pJ/b)	17.04	15.42	-	11	17.75
Worst uniformity (%)	54.6	73.08	-	-	-
Average Uniformity	50.86	54	-	-	-
Uniqueness (%)	0.500	0.502	0.499	0.499	0.501
Worst BER, raw (%)	5.9	7.9	8.1	9.0	9.0
Worst BER, with pruning (%)	1	1.2	$<10^{-9}$	0.4	$<10^{-8}$
Pruned CRPs (%)	25	29.8	80.4	42	34
Pruning method	Threshold	Threshold	BIST	Threshold	Threshold

4.4 Conclusion

A robust and high entropy DC-INV based strong PUF is proposed in this chapter. The response bit is generated by driving the monostable output voltage of parallelly connected DC-INV's controlled by the input challenge into one of the two stable logic voltages. The design has removed the systemic bias of [12] by reusing the inverters in Stages 1 and 2 and making Stage 3 driver challenge dependent. With asymmetric inverter connections in different blocks, this improved strong PUF is harder to model. Its response predictability by SVM and CMA-ES learning are as good as wild guess. The worst reliability is 94% and improved to 99% after filtering away a smaller fraction of potentially unreliable response bits. Its high modelling complexity coupled with good uniformity and low native response reliability over a broad range of operating conditions are attractive for lightweight and low-cost authentication.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

With the rise of internet of things (IoTs), a super network of huge number of smart devices with unique identifiers and sensing and communication capabilities has been built up to transfer data over interconnected networks into the cyber space for storage, processing, reasoning, etc. of complex heterogenous data cooperatively, as well as controlling mechanical parts, actuators, and machines remotely and in real time. The convergence of wireless technologies, microelectromechanical systems, microservices, emerging memories, artificial intelligence and the internet has torn down the silos between operational and information technologies to enable unstructured machine-generated data to be analysed for driving unprecedented revolutionary advancements and changes in the way people live, work and play. With reducing human supervision, intervention and monitoring, increasing ubiquity of sensors and device mobility, and strong empowerment of these smart devices and machines to interact autonomously and make decisions, attack surface has also expanded rapidly and uncontrollably. Consequently, security and privacy are cited as major concerns in the IoT era. Because IoT devices are closely connected, attackers can exploit poorly secured IoT endpoints and vulnerabilities of any IoT devices to gain access and control to multiple endpoints distributed all over the world. Privacy is endangered as connected devices often request users to input their personal information before authorizing access to services. Beyond leaking personal data, the risk of insecure IoT devices extend to critical infrastructure, including electricity, transportation and financial institutions, as evinced by one of the most notorious distributed denial-of-service (DDoS) attacks happened in September 2016 where Mirai, a botnet that infiltrated domain name server provider Dyn and took down many websites for an extended period of time [69].

On-device data confidentiality and integrity are typically achieved by using non-volatile memory in computing systems to preserve the keys and sensitive information protected by encryption or message digest. Such approach is not always feasible or practical due to the cost and power consumptions in implementing cryptographic algorithms on hardware. Besides, the protection is only as strong as its weakest link, once the key is hacked, the last line of defence is torn down. Due to the ease of access of most IoT devices, storing the key in non-volatile storage has proven to be vulnerable to various forms of non-invasive, semi-invasive and invasive attacks [3] [60]. In this light, silicon PUF emerges as a burgeoning technology for IoT security. It has gained a strong

foothold in cryptographic key generation, device identification and authentication over the last decade. It is essentially a challenge-response interrogation interface derived from a disorder physical system [2]. Such physical disorder is a congenital birthmark of a device that last through its life time. Hence, the signatures generated by a PUF device are analogous to the human biometrics. They are difficult to physically replicate even by the original manufacturer with the same photolithography masks. PUF is more secure than other cryptographic primitives since the device needs to be powered up and stimulated with an input challenge before the signature can be extracted, which is in stark contrast to the hardcoded form of identities statically and permanently present in the device. The introductory chapter of this thesis reviews the definition and figures of merit of PUF, weak and strong PUF classification, their differentiating features and applications. 6T SRAM PUF and arbiter PUF are introduced as representatives of weak PUF and strong PUF, respectively. The working mechanisms of two emerging monostable PUFs are also illustrated, a specific machine learning algorithm widely used for modeling attack of strong PUF is also presented. The design and implementation of a novel dual port SRAM PUFs and a monostable strong PUF with substantially enhanced properties such as increase entropy for response generation, better uniformity, higher raw response reliability under varying and broader range of operating conditions and increased resistance to machine learning attacks, etc. are presented after the introduction chapter. The main contribution and works are summarized as follows.

The idea of developing a dual-port SRAM PUF was motivated by the process variability issues of memory devices and circuits. Inspired by the work of [13], the main drawback of conventional SRAM PUF is investigated to explore new memory structures that can overcome its limitation. In Chapter 3, the design problems of typical memory-based PUF are discussed. A new 8T-DP-SRAM PUF is then proposed to generate multiple response bits per cell by writing complementary bits into two ports in the write conflict mode. Comparing with the 6T SRAM PUF, the 8T-DP-SRAM PUF has more access flexibility as the two ports can perform read and write operations separately. This 8T-DP-SRAM PUF avoids buffering data into and out of the memory when switching between SRAM and PUF operations, which protects the sensitive data from aggressive memory attacks. As it harnesses more entropy from the manufacturing process variations of all eight transistors of a memory cell, it can generate up to a maximum of four independent response bits per cell. A 1024-bit DP-SRAM PUF is implemented on UMC65nm CMOS technology and its performances are evaluated. The simulation results are presented to prove its reliability, uniqueness, randomness and entropy.

The design of an improved monostable Diode-Clamped inverter-based strong PUF is motivated by the measured skewed responses of the work reported in [12]. The response bits of this PUF are generated by comparing the voltage trip points V_{trip} of different DC-INV. It was found that there was systemic bias in the two stages and these biases add up, which account for the poor uniformity of its responses. The original architecture has

been modified to include two demultiplexer-multiplexer pairs to share the 512 DC-INV_s in these two stages and making them selectable by the same input challenge without conflict. This way the systemic bias will be averaged out. The modeling attack complexity has also been increased by the merging of two stages and permutation of inverter connections across blocks. Stage 3 of the monostable strong PUF is also modified to eight parallel blocks of 16 inverters each. They are made to be selectable by a quarter of the input challenge bits to reduce the bias and further increase modeling complexity. In addition, a filter circuit is used to filter out potentially unstable response bits. The enhanced design has been successfully signed off using TSMC 40nm CMOS technology. The measurement results on the prototype chips attested the improved uniformity, reliability, uniqueness, randomness and resistance against machine learning attacks.

5.2 Future Work

5.2.1 Multi-port SRAM PUF

Currently, the quality metrics of the proposed 1024-bit DP-SRAM PUF are verified based on Monte Carlo simulation. The circuit will be fabricated and tested in future to verify its quality physically. Reliability against aging has not been considered. It will be useful to explore techniques to improve reliability against aging. Aging-resistance tests can be carried out more efficiently and accurately by subjecting the physical chips to accelerated aging with over-rated voltages and temperature. In addition, other multi-port SRAM or emerging volatile memory circuits may be explored.

5.2.2 Chip measurements of monostable PUF

Due to limited number of dies for shared chip area from different projects in this industrial sponsored MPW shuttle, the distribution of fractional Hamming distance plotted in Chapter 4 was obtained by simulation with 100 random challenges applied to 80 different PUF instances. Additional chips will be acquired from other groups that share the dies to perform more rigorous tests on actual distribution of inter-die Hamming distance. One potential future work is to explore the use of this strong PUF together with another entropy source for the design of lightweight and high-throughput true random number generator.

Author's Publication

Journal papers

H. S. Zhong, C. Q. Liu, C. H. Chang, "A robust monostable strong PUF with improved uniformity and enhanced security," *2019 International Symposium on Integrated Circuits and Systems (ISICAS 2019, a CAS journal track symposium), Venetia, Italy, September 29-30, 2019* (submitted).

H. S. Zhong, C. Q. Liu, C. H. Chang, "A lightweight buffer free multiple port SRAM PUF with high entropy," *IEEE Transactions on Circuits and Systems-I: Regular Papers* (in preparation).

References

- [1] S. Singh and N. Singh, "Internet of Things (IoT): Security challenges, business opportunities and reference architecture for E-commerce," in *Proc. 2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, Noida, India, Oct. 2015, pp. 1577-1581.
- [2] C. H. Chang, Y. Zheng, and L. Zhang, "A retrospective and a look forward: fifteen years of physical unclonable function advancement," *IEEE Circuits and Systems Magazine*, vol. 17, no. 3, pp. 32–62, Aug. 2017.
- [3] C. Herder et al., "Physical unclonable functions and applications: A tutorial," *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1126-1141, Aug. 2014.
- [4] N. Potlapally, "Hardware security in practice: Challenges and opportunities," in *Proc. 2011 IEEE International Symposium on Hardware-Oriented Security and Trust*, San Diego, CA, USA, Jun. 2011, pp. 93-98.
- [5] F. Cai, G. Bai, H. Liu, X. Hu, "Optical fault injection attacks for flash memory of smartcards," in *Proc. 2016 6th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, Beijing, China, Jun. 2016, pp. 46-50.
- [6] S. E. Quadir, J. Chen, D. Forte, N. Asadizanjani, S. Shahbazmohamadi, L. Wang, J. Chandy, M. Tehranipoor, "A survey on chip to system reverse engineering," *ACM Journal on Emerging Technologies in Computing Systems*, vol. 13, no. 1, pp. 1–34, Jan. 2016.
- [7] J. Breier, D. Jap, C. N. Chen, "Laser profiling for the back-side fault attacks: with a practical laser skip instruction attack on AES," in *Proc. the 1st ACM Workshop Cyber-Physical Syst. Security*, Singapore, Apr. 2015, pp. 99–103.
- [8] N. A. Anagnostopoulos, T. Arul, M. Rosenstihl, A. Schaller, S. Gabmeyer, S. Katzenbeisser, "Low-temperature data remanence attacks against intrinsic SRAM PUFs," in *Proc. 2018 21st Euromicro Conference on Digital System Design (DSD)*, Prague, Aug. 2018, pp. 581-585.
- [9] V. P. Yanambaka, S. P. Mohanty, and E. Kougianos, "Making use of manufacturing process variations: A dopingless transistor based-PUF for hardware-assisted security," *IEEE Transactions on Semiconductor Manufacturing*, vol. 31, no. 2, pp. 285–294, May 2018.
- [10] B. Gassend et al., "Silicon physical random functions," in *Proc ACM Conf. on Computer and Communication Security*, Washiton, DC, USA, Nov. 2002, pp. 149.

- [11] A. Alvarez, W. Zhao, and M. Alioto, "Static physically unclonable functions for secure chip identification with $<2\%$ native bit instability and $140\times$ inter/intra PUF Hamming distance separation in 65nm," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2015, pp. 1–3.
- [12] Y. Cao, C. Q. Liu, and C. H. Chang, "A low power diode-clamped inverter-based strong physical unclonable function for robust and lightweight authentication," *IEEE Transactions in Circuit and System I: Regular Papers*, vol. 65, no. 11, pp. 3864-3873, Nov. 2018.
- [13] C. Q. Liu, Y. Zheng, and C. H. Chang, "A new write-contention based dual-port SRAM PUF with multiple response bits per cell," in *Proc. 2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, Baltimore, MD, USA, May 2017, pp.1-4.
- [14] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld. "Physical one-way functions," *Science*, vol. 297, no. 5589, pp. 2026–2030, 2002.
- [15] B. Gassend, D. Clarke, M. V. Dijk, and S. Devadas, "Controlled physical random functions," in *Proc. 18th Annual Computer Security Applications Conference*, Las Vegas, NV, USA, Dec. 2002, pp. 149-160.
- [16] J. Guajardo et al., "FPGA intrinsic PUFs and their use for IP protection," in *Proc. Workshop on Cryptographic Hardware and Embedded Syst.*, Vienna, Austria, Sept. 2007, pp. 63–80.
- [17] D. E. Holcomb, W. P. Burleson, and K. Fu., "Power-up SRAM state as an identifying fingerprint and source of true random numbers," *IEEE Transsctions on Computers*, vol. 58, no. 9, pp. 1198–1210, Sep. 2009.
- [18] D. Lim, J. W. Lee, B. Gassend, G. E. Suh, M. van Dijk and S. Devadas, "Extracting secret keys from integrated circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) System.*, vol. 13, no. 10, pp. 1200–1205, Oct. 2005.
- [19] K. Lofstrom, W. R. Daasch, and D. Taylor, "IC identification circuit using device mismatch," in *Proc. 2000 IEEE International Solid-State Circuits Conference. Digest of Technical Paper (Cat. No.00CH37056)*, San Francisco, CA, USA, Feb. 2000, pp.372-373
- [20] R. Maes, V. Rozic, I. Verbaauwhede, P. Koeberl, E. Van der Sluis, and V. van der Leest, "Experimental evaluation of physically unclonable functions in 65 nm CMOS," in *Proc. 2012 Proceedings of the ESSCIRC (ESSCIC)*, Brodeaux, Sep. 2012, France, pp.486-489.

- [21] Y. Su, J. Holleman, and B. P. 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.
- [22] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *Proc. 2007 44th ACM/IEEE Design Automation Conference*, San Diego, CA, USA, Jun. 2007, pp. 9–14.
- [23] M. van Dijk and U. Rührmair, "Physical unclonable functions in cryptographic protocols: Security proofs and impossibility results," *IACR Cryptology ePrint Archive*, vol. 2012, p. 228, 2012.
- [24] N. Torii, D. Yamamoto and T. Matsumoto, "Evaluation of latch-based PUFs implemented on 40 nm ASICs," in *Proc. 2016 Fourth International Symposium on Computing and Networking (CANDAR)*, Hiroshima, Japan, Nov. 2016, pp. 642-648.
- [25] S. S. Kumar, J. Guajardo, R. Maes, G. Schrijen and P. Tuyls, "Extended abstract: The butterfly PUF protecting IP on every FPGA," in *Proc. 2008 IEEE International Workshop on Hardware-Oriented Security and Trust*, Anaheim, CA, Jun. 2008, pp. 67-70.
- [26] P. Simons, E. van der Sluis and V. van der Leest, "Buskeeper PUFs, a promising alternative to D Flip-Flop PUFs," in *Proc. 2012 IEEE International Symposium on Hardware-Oriented Security and Trust*, San Francisco, CA, USA, Jun. 2012, pp. 7-12.
- [27] Y. Ogasahara, Y. Hori and H. Koike, "Standard cell implementation of buskeeper PUF with symmetric inverters and neighboring cells for passing randomness tests," in *Proc. 2015 IEEE 4th Global Conference on Consumer Electronics (GCCE)*, Osaka, Japan, Oct. 2015, pp. 550-551.
- [28] J. Jang and S. Ghosh, "Design and analysis of novel SRAM PUFs with embedded latch for robustness," in *Proc. Sixteenth International Symposium on Quality Electronic Design*, Santa Clara, CA, Mar. USA, 2015, pp. 298-302.
- [29] L. Zhang, X. Fong, C. H. Chang, Z. H. Kong, and K. Roy. "Optimizing emerging nonvolatile memories for dual-mode applications: data storage and key generator," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, no. 7, pp.1176-1187, Jul. 2015.
- [30] E. I. Vatajelu, G. D. Natale and P. Prinetto, "Towards a highly reliable SRAM-based PUFs," in *Proc. 2016 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Dresden, Germany, Mar. 2016, pp.273-276

- [31] F. Zhang, S. Yang and J. Plusquellic, "Current based PUF exploiting random variations in SRAM cells," in *Proc. Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Dresden, Germany, Mar. 2016, pp. 277-280
- [32] M. Orshansky, S. Nassif, D. Boning, Design for Manufacturability and Statistical Design: A constructive Approach, New York: Springer Science & Business Media, 2007.
- [33] K. Suzuki, K. Miura and K. Nakamae, "NBTI/PBTI tolerant arbiter PUF circuits," in *Proc. 2017 IEEE 23rd International Symposium on On-Line Testing and Robust System Design (IOLTS)*, Thessaloniki, Greece, Jul. 2017, pp. 80-84.
- [34] T. Idriss and M. Bayoumi, "Lightweight highly secure PUF protocol for mutual authentication and secret message exchange," in *Proc. 2017 IEEE International Conference on RFID Technology & Application (RFID-TA)*, Warsaw, Poland, Sept. 2017, pp. 214-219
- [35] J. W. Lee, D. Lim, B. Gassend, G. E. Suh, M. van Dijk and S. Devadas, "A technique to build a secret key in integrated circuits for identification and authentication applications," in *Proc. 2004 Symposium on VLSI Circuits. Digest of Technical Papers (IEEE Cat. No. 04CH37525)*, Honolulu, HI, USA, Jun. 2014, pp.176-179
- [36] S. Gehrler and G. Sigl, "Reconfigurable PUFs for FPGA-based SoCs," in *Proc. 2014 International Symposium on Integrated Circuits (ISIC)*, Singapore, Dec. 2014, pp. 140-143.
- [37] M. Potkonjak and V. Goudar, "Public Physical Unclonable Functions," *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1142-1156, Aug. 2014.
- [38] L. Zhang, Z. H. Kong, C. H. Chang, A. Cabrini, and G. Torelli, "Exploiting process variations and programming sensitivity of phase change memory for reconfigurable physical unclonable functions," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 6, pp. 921-932, Jun. 2014.
- [39] L. Zhang, X. Fong, C. H. Chang, Z. H. Kong, and K. Roy, "Highly reliable spin-transfer torque magnetic RAM based physical unclonable function with multi-response-bits per cell," *IEEE Transcation on Information Forensics and Security*, vol. 10, no. 8, pp. 1630-1642, Aug. 2015.
- [40] C. Q. Liu, Y. Cao, and C. H. Chang, "ACRO-PUF: A low-power, reliable and aging-resilient current starved inverter-based ring oscillator physical unclonable function," *IEEE Transactions on Circuits and System I: Regular Paper*, vol. 64, no.12, pp. 3138-3149, Dec. 2017

- [41] Y. Cao, L. Zhang, S. S. Zalivaka, C. H. Chang and S. Chen "CMOS image sensor based physical unclonable function for coherent sensor-level authentication," *IEEE Transaction on Circuit and System I: Regular Papers*, vol. 62, no. 11, pp. 2629-2640, Nov. 2015
- [42] Z. Wang et al., "Current mirror array: a novel circuit topology for combining physical unclonable function and machine learning," *IEEE Transactions on Circuits and System I: Reg. Papers*, vol. 65, no. 4, pp. 1314-1326, Apr. 2018.
- [43] U. Rührmair et al., "PUF modeling attacks on simulated and silicon data," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 11, pp. 1876–1891, Nov. 2013.
- [44] M. Ross, C. A. Graves, J. W. Campbell and J. H. Kim, "Using Support Vector Machines to Classify Student Attentiveness for the Development of Personalized Learning Systems," *2013 12th International Conference on Machine Learning and Applications*, Miami, FL, 2013, pp. 325-328
- [45] I. Rechenberg, *Evolutionsstrategie-Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Stuttgart, Germany: Frommann-Holzboog, 1973.
- [46] H. P. Schwefel, "Evolutionsstrategie und numerische Optimierung," Ph.D. dissertation, Technische Universität Berlin, Berlin, Germany, 1975
- [47] Q. Fan and X. Yan, "Self-Adaptive Differential Evolution Algorithm With Zoning Evolution of Control Parameters and Adaptive Mutation Strategies," in *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 219-232, Jan. 2016.
- [48] J. K. Zao, M. Hornansky and Pei-lun Diao, "Design of optimal short-length LT codes using evolution strategies," *2012 IEEE Congress on Evolutionary Computation*, Brisbane, Australia, Jun, 2012, pp. 1-9.
- [49] Karthik Sindhya, "An Introduction to Evolutionary Algorithms", Industrial Optimization Group, Department of Mathematical Information Technology, University of Tyvaskyla.
- [50] S. Zeitouni, Y. Oren, C. Wachsmann, P. Koeberl and A. Sadeghi, "Remanence decay side-channel: the PUF case," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 6, pp. 1106-1116, Jun. 2016.
- [51] M. Kassem, M. Mansour and A. Chehab, and A. Kayssi, "A sub-threshold SRAM based PUF," in *Proc. 2010 International Conference on Energy Aware Computing*, Cairo, Egypt, Dec. 2010, pp.1-4.

- [52] D. Burnett, K. Erington, C. Subramanian, and K. Baker, "Implications of fundamental threshold voltage variations for high-density SRAM and logic circuits," in *Proc. 1994 VLSI Technology Symposium*, Honolulu, HI, USA, Jun. 1994, pp. 15-16.
- [53] B. Cheng, S. Roy, and A. Asenov, "The impact of random doping effects on CMOS SRAM cell," in *Proc. The 30th European Solid-State Circuits Conference*, Leuven, Belgium, Sept. 2004, pp. 219-222
- [54] C. H. Chang, C. Q. Liu, L. Zhang, and Z. H. Kong, "Sizing of SRAM cell with voltage biasing techniques for reliability enhancement of memory and PUF functions," *Journal of Low Power Electronics and Applications*, vol. 6, no. 3, p. 16, Aug. 2016
- [55] M. S. M. Siddiqui, S. Srivastav, D. R. Wanjul, M. Suthar, and S. Kumar, "A 7-Nm dual port 8T SRAM with duplicated inter-port write data to mitigate write disturbance," in *Proc. 31st International Conference on VLSI Design (VLSI)*, Pune, India, Jan. 2018, pp. 266-270.
- [56] K. Nii et al., "Synchronous ultra-high-density 2RW dual-port 8T-SRAM with circumvention of simultaneous common-row-access," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 3, pp. 977-986, Mar. 2009
- [57] S. Devadas et al., "Design and implementation of PUF-based "unclonable" RFID ICs for anti-counterfeiting and security applications," in *Proc. 2008 IEEE International Conference on RFID*, Las Vegas, NV, USA, Apr. 2008, pp. 58-64.
- [58] I. Baturone, M. A. Prada-Delgado, and S. Eiroa, "Improved generation of identifiers, secret keys, and random numbers from SRAMs," *IEEE Transactions Information Forensics Security*, vol. 10, no. 12, pp. 2653-2668, Dec. 2015.
- [59] D. Lim, J. W. Lee, B. Gassend, G. E. Suh, M. van Dijk and S. Devadas, "Extracting secret keys from integrated circuits," *IEEE Transaction on Very Large Scale Integration (VLSI) System*, vol. 13, no. 10, pp. 1200-1205, Oct. 2005.
- [60] R. Arjona, M. A. Prada-Delgado, J. Arcenegui, and I. Baturone, "A PUF- and biometric-based lightweight hardware solution to increase security at sensor nodes," *Sensors*, vol. 18, no. 8, pp.2429, Jul. 2018.
- [61] Y. Zheng, Y. Cao, and C. H. Chang, "UDhashing: Physical unclonable function based user-device hash for endpoint authentication," *IEEE Transactions on Industrial Electronics*, Jan. 2019.
- [62] T. Joachims, "Learning to classify text using support vector machines," Ph.D. dissertation, Univ. Dortmund, Germany, 1999.

- [63] U. Rührmair *et al.*, “PUF modeling attacks on simulated and silicon data,” *IEEE Transactions on Information Forensics Secur.*, vol. 8, no. 11, pp. 1876–1891, Nov. 2013.
- [64] A. B. Alvarez, W. Zhao and M. Alioto, "Static physically unclonable functions for secure chip identification with 1.9–5.8V and 15 fJ/bit in 65 nm," *IEEE J. Solid-State Circuit*, vol. 51, no. 3, pp. 763–775, Mar. 2016.
- [65] A. Rukhin *et al.*, “A statistical test suite for random and pseudo-random number generators for cryptographic applications,” *NIST Special Publication 800-22*, 2010.
- [66] Z. He, M. Wan, J. Deng, C. Bai, and K. Dai, “A reliable strong PUF based on switched-capacitor circuit,” *IEEE Transactions on Very Large Scale Integration (VLSI) System*, vol. 26, no. 6, pp. 1073–1083, Jun. 2018.
- [67] X. Xi, H. Zhuang, N. Sun, and M. Orshansky, “Strong subthreshold current array PUF with 2^{65} challenge-response pairs resilient to machine learning attacks in 130nm CMOS,” in *Proc. 2017 Symposium on VLSI Circuits*, Kyoto, Japan, Jun. 2017, pp. C268–C269.
- [68] K. Yang, Q. Dong, D. Blaauw, and D. Sylvester, “A physically unclonable function with BER $<10^{-8}$ for robust chip authentication using oscillator collapse in 40nm CMOS,” in *Proc. 2015 IEEE International Solid-State Circuit Conference-(ISSCC)Digest of Technical Paper*, San Francisco, CA, USA, Jan. 2015, pp. 1–3.
- [69] L. Constantin, “Armies of hacked IoT devices launch unprecedented DDoS attacks,” News release, Sept. 26, 2016.
<https://www.csoonline.com/article/3124344/armies-of-hacked-iot-devices-launch-unprecedented-ddos-attacks.html>