University of New Mexico UNM Digital Repository

Electrical and Computer Engineering ETDs

Engineering ETDs

Spring 4-14-2017

THE EFFECT OF POWER SUPPLY RAMP TIME ON SRAM PUF's

Abdelrahman T. Elshafiey Mr. University of New Mexico - Main Campus

Follow this and additional works at: https://digitalrepository.unm.edu/ece_etds Part of the <u>Electrical and Electronics Commons</u>, <u>Electronic Devices and Semiconductor</u> <u>Manufacturing Commons</u>, and the <u>VLSI and Circuits</u>, <u>Embedded and Hardware Systems Commons</u>

Recommended Citation

Elshafiey, Abdelrahman T. Mr.. "THE EFFECT OF POWER SUPPLY RAMP TIME ON SRAM PUF's." (2017). https://digitalrepository.unm.edu/ece_etds/342

This Thesis is brought to you for free and open access by the Engineering ETDs at UNM Digital Repository. It has been accepted for inclusion in Electrical and Computer Engineering ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Abdelrahman T. Elshafiey

Electrical and Computer Engineering Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Dr. Payman Zarkesh-Ha, Chairperson

Dr. James Plusquellic

Dr. James Aarestad

THE EFFECT OF POWER SUPPLY RAMP TIME ON SRAM PUF's

BY

Abdelrahman T. Elshafiey

B.S., Communications Engineering, The Higher Institute of Engineering at El – Shorouk, 2013

THESIS

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

Electrical Engineering

The University of New Mexico Albuquerque, New Mexico

May, 2017

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

فَمَنْ يَعْمَلْ مِثْقَالَ ذَرَّةٍ خَيْرًا يَرَهُ وَمَنْ يَعْمَلْ مِثْقَالَ ذَرَّةٍ شَرًّا يَرَهُ

In the name of Allah, the Gracious, the Merciful. Whoever has done an atom's weight of good will see it. And whoever has done an atom's weight of evil will see it.

DEDICATION

To my mother and father

To my brothers

To my grandmother and uncles

ACKNOWLEDGMENTS

I heartily acknowledge Dr. Payman Zarkesh-Ha, my advisor, for continuing to encourage me through the time of classroom teaching and the research.

I would like to thank, Dr. James Plusquellic, for the great two courses I took with him.

I also thank my committee member, Dr. James Aarestad, for his valuable recommendation pertaining to this study and assistance in my professional development.

I also thank, Dr. Jane Lehr, and Dr. Mark Gilmore, for the scholarship and for the acceptance to UNM.

Finally, I appreciate the support from my friend, Amr Mostafa.

THE EFFECT OF POWER SUPPLY RAMP TIME ON SRAM PUF's

by

Abdelrahman T. Elshafiey

B.S., Communications Engineering, The Higher Institute of Engineering at El – Shorouk, 2013
M.S., Electrical Engineering, University of New Mexico, 2017

ABSTRACT

Physical unclonable functions (PUFs) are security primitives that exploit the device mismatches. PUFs are a promising solution for hardware cryptography and key storage. They are used in many security applications including identification, authentication and key generation. SRAM is one of the popular implementations of PUFs. SRAM PUFs offer the advantage, over other PUF constructions, of reusing resources (memories) that already exist in many designs.

In this thesis, for the first time, it is demonstrated that the start-up value of an SRAM PUF could be different depending on the SRAM power supply rising time. An analytical model has been developed to determine the range for the power supply ramp time that affects the SRAM PUF start-up value. It has been found that there are two regions of operation. The generated key could possibly be different from one region to another. An SRAM test chip was designed and fabricated using Tower Jazz's 180 nanometer Silicon Germanium (SiGe) Bipolar/CMOS (BiCMOS) process. Based on our

measured data, using the appropriate rising time can decrease the number of flipping bits by 5%. Both simulation and silicon results confirms the analytical model.

TABLE OF CONTENTS

LIST OF FIGURES			
LIST OF TABLESxii			
CHAPTER 1 INTRODUCTION 1			
1.1 Physical Unclonable Functions: Definition & Applications 1			
Physical Unclonable Function Definition 1			
PUF Applications2			
Advantages of PUFs 2			
1.2 Physical Unclonable Functions: Classifications			
Weak and strong PUFs 3			
1.3 Physical Unclonable Functions: Metrics 4			
Uniqueness			
Randomness 5			
Reliability5			
1.4 Physical Unclonable Functions: Silicon PUF Implementations 6			
Arbiter PUF 6			
Ring-Oscillator PUF7			
SRAM PUF7			
1.5 Physical Unclonable Functions: Challenges & Problems 7			
1.6 Objective and Scope of this Thesis 8			
CHAPTER 2 Detail SRAM PUF Chip Design & Simulation9			
2.1 Pad Ring Structure 9			
2.2 SRAM PUF Structure 12			

2.3 SRAM cell simulation	14
CHAPTER 3 Testing Procedures & Setup	
3.1 Chip PCB layout	18
2.3 Testing PCB layout	19
3.3 SRAM PUF Test Procedures	21
CHAPTER 4 Evaluation of 180 nm SRAM test chip	
CHAPTER 5 The effect of power supply ramp time on SRAM PUF's	
5.1 Introduction	
5.2 Analytical Model	
5.3 Simulation	35
5.4 Silicon results	
5.5 Conclusion	

LIST OF FIGURES

Figure 1. High-level schematic of the three PUF circuits
Figure 2. Two types of pad libraries: io160u5V with 160 μ m pitch and io80u5V with
80μm pitch10
Figure 3. List of all pads in the library and highlighted pads are used in our design10
Figure 4. Choice of I/O pads used in our design and highlighted in Figure 211
Figure 5. Initial pad ring arrangements for the test chip12
Figure 6. The schematic and layout of an SRAM PUF unit cell14
Figure 7. The schematic of a row or column address select using a shift register15
Figure 8. Test schematic layout of a 2×2 SRAM PUF
Figure 9. Simulation results on the extracted circuit from the SRAM PUF layout in Figure
617
Figure 10. Floorplan and placement of the SRAM PUF17
Figure 11. Ball grid array (BGA), through hole Package. Package side view on the left
and plan view on the right18
Figure 12. Chip pcb layout
Figure 13. Testing pcb layout20
Figure 14. Testing setup
Figure 15. Timing diagram for testing the SRAM PUF
Figure 16. Inter-chip hamming distance (HD) statistical distribution23

Figure 17. Intra-chip bit-error-rate (BER), for 4 chips24
Figure 18. Ratio of 1's in the SRAM PUF response for the four test chips25
Figure 19. SRAM response as a function of fast and slow power supply rising time26
Figure 20. Two different regions, with different dominant fabrication process variations.
Figure 21. CMOS SRAM cell including the gate capacitances
Figure 22. Supper position between NMOS N_2 current and V_{dd} power supply, at node Q .
Figure 23. The outputs of the SRAM increasing with the power supply ramp time at a
slower rate
Figure 24. SRAM cell transient simulation, using two different rising times
Figure 25. Threshold voltage variation versus Power supply rising time T, for both the
simulation and the model, at ΔC_1 , ΔC_2 and ΔC_3
Figure 26. 125×150 bits SRAM PUF response. Black dots are 1's, white dots are 0's and
red dots are flipping bits
Figure 27. Extract data from SRAM test chip. Number of flipping bits versus supply
voltage rising time

LIST OF TABLES

Table 1.	Parameters fo	r 180 nm	SBC18H3 model	
----------	---------------	----------	---------------	--

Chapter 1

Introduction

1.1 Physical Unclonable Functions: Definition & Applications

Physical Unclonable Function Definition

A physical unclonable function is an entity that uses production variability to generate a device-specific output which usually is a binary number. This output can be seen as the *fingerprint* of a device [1]. A PUF is made of several components defined by local parameter variations [1]. The differences between the components are called local mismatches [1]. Furthermore, A PUF is a function that generates a set of responses (secrets), when it is stimulated by a set of challenges (Challenge-Response pairs). It is a physical function because the challenge-response relation is defined by complex properties of a physical material, such as the manufacturing variability of CMOS devices. Its unclonability is attributed to the fact that due to the manufacturing variability that defines the secret, one cannot manufacture two identical chips, even with full knowledge of the chip design, and since the variation of the components cannot be controlled from the outside, a PUF cannot be replicated [1]. From a theoretical perspective, the single words in PUF have the following meaning [1]:

Physical means a physical entity, in contrast to an algorithm or a similar function. If physically is used, the meaning changes since now it becomes an adverb to unclonable which means that the function is cloneable in general but not in a physical way.

Unclonable means that a thing cannot be replicated. For PUFs this is true in practice. Theoretically, PUFs are cloneable.

Function means in terms of mathematics that an input value is associated with one specific output value. Since the output of a PUF is usually noisy, it happens that an input produces different outputs. And often PUFs are used without any input in literature, for example, if it is used for key generation. Thus, in general a PUF is not a function in the mathematical meaning. Therefore, we define a PUF as follows: *A PUF is a physical entity which produces an output value at least in dependence of physical structures, which are hard to clone.*

PUF Applications

Physical Unclonable Functions (PUFs) started by the idea of taking advantage of the random physical variations that can be found in various objects [2], [3]. The core concept put forward by PUFs is to take advantage of the submicron variations introduced during fabrication, which is unique to each device, and use it as an identity rather than assigning an arbitrary identity to it upon creation [2]. Furthermore, PUFs were used in storing/generating cryptographic keys and authentication [2]. PUFs become a promising solution to security issues like intellectual property (IP) protection, device authentication, and user data privacy.

Advantages of PUFs

Since non-volatile memories are vulnerable to attacks [4], PUFs provide an alternate to storing secret keys in non-volatile memories (NVMs), by generate the key whenever it is needed only. NVMs used to store the IDs in systems that requires identification [1]. A system without NVM is usually cheaper to produce, because additional processing steps are required [1]. Moreover, PUFs are very hard to attack by reverse engineering methods [1]. Moreover, SRAM PUFs offer the advantage, over other PUF constructions, of reusing resources (memories) that already exist in many designs [5].

1.2 Physical Unclonable Functions: Classifications

Generally, there are more than one category that PUFs can fall into. For example, there are Electronic and Non-Electronic PUFs, Silicon and Non-silicon PUFs, Intrinsic and Extrinsic PUFs, and finally Weak and Strong PUFs. Only Silicon intrinsic PUFs will be discussed in this thesis.

Weak and Strong PUFs

One of the important categories from the application prospective is weak and strong PUFs [4]. This classification is based on the number of challenge-response pairs (CRPs). Strong PUFs have many CRPs which qualifies them for authentication applications. On the other hand, weak PUFs have one CRP up to a few CRPs, which make them suitable for storing and generating secure keys [6].

Explicitly stated, weak PUFs have the following properties [6]:

• a small number of CRPs (linearly related to the number of components whose behavior depends on manufacturing variation);

- response is stable and robust to environmental conditions and multiple readings so that a challenge always yields the same response;
- responses are unpredictable and depend strongly on the innate manufacturing variability of the device;
- it is impractical to manufacture two devices with the same physical fingerprint;
- since weak PUFs in general have only a small number of CRPs, these pairs must be kept secret. If a weak PUF only has one CRP, and it is revealed, then any device can emulate the PUF. For this reason, weak PUFs are well suited for use in key derivation processes.

The requirements for a strong PUF are [6]:

- large enough challenge–response space such that an adversary cannot enumerate all CRPs within a certain fixed time (ideally, exponential in the number of challenge bits);
- responses stable to environment, multiple readings;
- an adversary given a polynomial-sized sample of adaptively chosen CRPs cannot predict the response to a new, randomly chosen challenge;
- not feasible to manufacture two PUFs with the same responses.

1.3 Physical Unclonable Functions: Metrics.

In this section, some parameters will be presented, which will define the quality and requirements of the PUF responses. These parameters indicate if it is suitable to use the PUF in identification, authentication and security applications or not.

Uniqueness

Uniqueness measures how the response of two PUF instances are different (Interchip). Uniqueness can be accessed by the Hamming Distance (HD). For example, the two binary numbers "1101" & "1001" have a HD of 1 (25%). The ideal HD for two 128bit response PUFs should be 64, which corresponds to a uniqueness of 50%.

Reliability

Reliability measures how the responses of the same PUF instance are different using the same challenge (Intra-chip). In other words, Reliability of a PUF measures its ability to produce the same responses under varying environmental conditions, e.g. temperature and supply voltage [7]. It is also accessed by the Hamming Distance (HD). The ideal HD for the same PUF instance is 0. Reliability can also be measured using bit-error-rate (BER).

$$BER = \frac{HD}{Total \, response \, bits} \, x \, 100 \, (\%)$$

Randomness

Randomness (intra-chip randomness) is a measure of the unpredictability of the response [7]. This implies (i) unpredictability of a response for a new challenge despite the prior knowledge of a large number of challenge-response pairs (CRPs) as well as (ii) unpredictability of every bit in the response even with a knowledge of all other response bits [8]. Therefore, it is required to have a ratio of 0's and 1's very close to 50%.

1.4 Physical Unclonable Functions: Silicon PUF Implementations.

In this section, Three PUF implementations will be presented as shown in Figure 1. Arbiter PUF is an example of strong PUFs, while both Ring-Oscillator and SRAM PUFs are examples of weak PUFs.

Arbiter PUF

Arbiter PUF is composed of two identically configured delay paths that are stimulated by a triggering signal. The difference in the propagation delay of the signal in the two delay paths is measured by an edge triggered flip-flop known as the arbiter. The delay



Figure 1: High-level schematic of the three PUF circuits.

difference is a function of the manufacturing process variation present in the delay paths. Several PUF response bits can be generated by configuring the delay paths in multiple ways using the challenge inputs [9].

Ring-Oscillator PUF

Ring Oscillator PUF exploits the variations in frequencies of several identically laid out ring oscillators to build the PUF [10]. The RO frequencies are captured in a counter, and are subsequently transformed into binary outputs by a simple comparison method.

SRAM PUF

SRAM PUF employs an SRAM cell (two cross-coupled inverters), and exploits the random assignment of a stable state from an initial unstable state. The final state of the cell is determined by the random mismatches in the pair of inverters [11].

1.5 Physical Unclonable Functions: Challenges & Problems

It is needless to say that PUFs are not perfect [1]. PUFs do not provide the same output each time [1]. PUFs show bit errors. These errors appear randomly and deterministically. The random errors are generated by circuit noise [1]. The deterministic errors are generated by mismatch between parameters of the involved components, e.g., mismatches of temperature coefficients or aging effects. Thus, if no errors are allowed by the application (e.g., key generation), error correction data must be generated and stored inside NVMs on the chip or outside the chip in databases [1]. Both are costly and reduce the advantages of PUFs. For identification purposes, errors may be no problem as long as the distance between the IDs is large enough. Accordingly, even if errors occur, the devices can be identified correctly [1].

1.6 Objective and Scope of this Thesis

The objectives of this thesis are:

- Testing and evaluating a 180nm SRAM PUF test chip.
- Modeling the start-up value (SUV) of the SRAM cell as a function of power supply rising time.

In chapter 2, the detailed SRAM chip design will be discussed, a 2x2 SRAM PUF will be simulated. Chapter 3 will show the testing setup, and the procedures to extract data from the SRAM test chip. Chapter 4, the 180 SRAM PUF will be tested and evaluated. Finally, in chapter 5 an analytical model will be discussed for the start-up value of the SRAM PUF as a function of power supply rising time, both simulation and silicon data will be shown to support the model.

Chapter 2

Detail SRAM PUF Chip Design

This chapter covers the detail circuit and layout design, as well as SPICE simulations and verifications for the SRAM PUF structure.

2.1 Pad Ring Structure

The pad structures are chosen from the Tower Jazz I/O standard cell library. An example of pads in the two libraries are illustrated in Figure 2. We have followed the Jazz digital I/O usage guidelines in this design. Jazz I/O pads are designed to be used in a contiguous pad ring with the supplied corner cells. ESD protection circuits are included in individual pads and work in conjunction with the corner cells, which contain large diodes. All pads have 8-11 horizontal metal busses that abut together to form the various power, ground, and ESD rings. There can be up to 8 nets which are named VDD, VDDO, VDDP, VGG, VSS, VSSO, and VESD (internal net). Many of these can be shorted in the pad ring using various power pads supplied in the library, reducing the number of electrical nets. In the simplest case, three power supply connections will be needed: core power, I/O power, and a common ground.

After all the functional I/O pads have been selected a number of power pads, 4 corner cells, and one power-up sequencing pad need to be added. There are a number of power pads to choose from depending on which supplies need to be kept separate or can be shorted together. If the chip core and I/O operate at different voltages, then 3 supply nets – VDD,

VDDO, and VSS – will be required at a minimum. List of I/O pads used in our design is highlighted in Figure 3 and their layout are illustrated in Figure 4.



Figure 2: Two types of pad libraries: io160u5V with 160µm pitch and io80u5V with 80µm pitch.

pcx00n	32KHz Xtal Oscillator				
pcx01n	Medium Speed Xtal Oscillator (1 MHz to 25 MHz)				
plb0ln	5V Tolerant (level shifter to core) Bidirectional 120ohm/2mA				
plb02n	5V Tolerant (level shifter to core) Bidirectional 80ohm/4mA				
plb03n	5V Tolerant (level shifter to core) Bidirectional 50ohm/8mA				
plb04n	5V Tolerant (level shifter to core) Bidirectional 32ohm/12mA				
pld00n	5V Tolerant (level shifter to core) Input				
pld00r	5V Tolerant (level shifter to core) Input w/Repeater				
pld10n	5V Tolerant (level shifter to core) Input w/Hysteresis				
plt0ln	5V Tolerant (level shifter to core) 3-State 120ohm/2mA				
plt02n	5V Tolerant (level shifter to core) 3-State 80ohm/4mA				
plt03n	5V Tolerant (level shifter to core) 3-State 50ohm/8mA				
plt04n	5V Tolerant (level shifter to core) 3-State 32ohm/12mA				
ptgcor	Padring Corner				
pybbcpo	Breaks VDD, VDDP, VDDO, VSSO				
pybdcnn	Breaks VDD				
pvbdnpo	Breaks VDDO, VDDP				
pybgcpo	Breaks VGG, VDDO, VDDP, VDD				
pvbgnnn	Breaks VGG				
pvdcnn	Connects VDD				
pvdnnn	Unconnected / Diode-Protected				
pvdnno	Connects VDDO				
pvdnpn	Connects VDDP				
pvdnpo	Connects VDDP VDDO				
pvgnnn	Connects VGG (ESD/Clamping Bus)				
pvgnno	Connects VDDO VGG				
pvgnpo	Connects VDDP VDDO VGG				
pyscnn	Connects VSS				
pysenns	Alternate Core VSS Pad with Power Sequencing Protection				
pvscnnu	Alternate Core VSS Pad with Pull-up to Disable Power-sequence				
pvscno	Connects VSS, VSSO				
pvsnnn	Unconnected / Diode-Protected				
pysnno	Connects VSSO				

Figure 3: List of all pads in the library and highlighted pads are used in our design.

Power-up sequencing refers to the order in which the I/O and core supplies are powered up, and the period between powering up each of these supplies. When the I/O supply (VDDP/VDDO) is powered up first, the output drivers can be in an indeterminate state until the core supply (VDD) is powered up. If the delay in the power sequence is long enough (several milliseconds), the unknown state of the output drivers could cause system problems.

Although the library includes a feature that can prevent the issue with power-up sequencing, we decided to disable this feature, and instead observe the power-up sequencing process, because the power sequencing protection feature draws a static current, which causes additional static power dissipation. To disable power sequencing protection, we will need to use pvscnnu pad in our design, as highlighted in Figure 3.

Regardless of whether pvscnnu or pvscnns is used, the recommended power-up sequence is:

- 5.0V (VGG)
- 3.3V (VDDP & VDDO)
- 1.8V (VDD)



Figure 4: Choice of I/O pads used in our design and highlighted in Figure 3.



Figure 5: Initial pad ring arrangements for the test chip.

Initial pad ring arrangements for the test chip is shown in Figure 5, where the overall layout edge dimension in about 4.9mm, which meets the initial criteria of < 5mm. This pad ring contains 26 pads on each edge, a total of 104 pads. The breakdown list of the pads is also shown in Figure 5, where there are 68 I/O pads, 16 ground pads, 12 core VDD pads, and 8 high voltage power pads for I/O. The core area inside the pad ring is 4.47×4.47 mm2.

2.2 SRAM PUF Structure

The circuit diagram of an SRAM PUF unit cell is shown in Figure 6. Like a standard SRAM cell, the unit cell in Figure 6 consists of two back-to-back inverters, where the random bit is generated. The output from this section of the circuit then goes to a pair of tri-state inverters, one on each side of the first pair of inverters. These tri-state inverters send their outputs to a low bit line and a high bit line. Only one output is necessary to get the PUF data,

but both outputs are used to better balance the circuit. The inputs to the tri-state inverters come from the AND gates, which is activated when the unit cell is selected. Both AND gates get their inputs from the "Row_Select" and "Column_Select" lines. These select lines come from the Row Select and Column Select circuits that is used to readout the SRAM content.

It is critical to layout the SRAM PUF precisely symmetric to eliminate any systematic bias into the PUF circuit. Only the process variation should determine the output of the unit cell. The layout of the SRAM PUF unit cell is also shown in Figure 6, where it has carefully been designed to ensure a symmetric layout. Symmetry is not necessary in typical SRAM memories, therefore in this project we decided not to use a standard SRAM compiler to design the SRAM PUF. Instead, a carefully hand drawn layout was designed in this project. The SRAM PUF unit cell size $20.46 \times 20.46 \mu m2$.

SRAM PUF generates the data just at the start up, when its power supply rises to V_{dd} . Every time SRAM PUF is tested, it is required to rise the core V_{dd} up and then down, which may damage other part of the test chip. Moreover, the core V_{dd} rise is normally slow (several millisecond), which could affect the performance of SRAM PUF. Therefore, we decided to separate the power supply for SRAM PUF unit cell from the core V_{dd} . This was necessary to create a more reliable testing process for SRAM PUF. The SRAM PUF supply is connected to an "Unconnected Pad", which has only ESD protection devices.

Figure 7 shows the schematic of "Row_Select" or "Column_Select" circuit. After a reset, this shift register shifts a "1" (i.e. selected column/row) from left to right. As shown, only one column/row is selected each time. Since the tri-state output of all SRAM PUF unit

cells are connected, only the selected SRAM PUF unit cell associate to that column and row will get out and will be read.

2.3 2 × 2 SRAM PUF simulation

To ensure the accuracy of the layout, a test layout of a 2×2 SRAM PUF is constructed from the actual design as illustrated in Figure 8. The layout is then extracted and simulated using TSPICE tool.

The results of SPICE simulations are illustrated in Figure 9. In this figure, "Col_CLK", "Col_Reset", "Row_CLK", and "Row_Reset" are the input signals. Also "SRAM_VDD" is the separate power supply for SRAM. "Bit_Line" and "Bit_Line_bar" are the output of the SRAM PUF that show the content of each unit cell when it is readout. "Co" and "C1" are the internal test points for two unit cells to verify the functionality of the readout



Figure 6: The schematic and layout of an SRAM PUF unit cell.



Figure 7: The schematic of a row or column address select using a shift register.

circuit. The results of Figure 9, shows that the behavior in the SPICE simulation matches up to the behavior in the simulation from the actual circuit that was laid out. The encouraging results from these simulations increases the confidence that the final chip will perform as expected.

Figure 10 illustrates the floorplan and placement of the SRAM PUFs on the test chip. The SRAM PUF contains $150 \times 125 = 18,750$ cells and occupies 2.6×3.1 mm2.



Figure 8: Test schematic layout of a 2×2 SRAM PUF.



Figure 9: Simulation results on the extracted circuit from the SRAM PUF layout in Figure 6.



Figure 10: Floorplan and placement of the SRAM PUF.

Chapter 3





Figure 11: Ball grid array (BGA), through hole Package. Package side view on the left and plan view on the right.

3.1 Chip PCB

To be able to access the chip pins, a printed circuit board (PCB) for the chip package shown if Figure 11, will be designed. In Figure 12, the chip PCB layout is presented. Each PUF type has been separated for easy access, and a hardwired ID pins were added before each PUF for testing purposes. On the left side of the board, all the power connections are routed. A tantalum capacitor has been added between each PAD ring supply (+3.3v) pin, core V_{dd} (+1.8v) pin and ground (GND) pins, to eliminate A.C noise from D.C power supplies as possible.

3.2 Testing PCB

To provide the proper supply power voltages needed for the chip to operate, and the control signals to extract the data from the chip, a testing PCB has been designed in Figure 13. Texas Instruments Tiva C development board has been chosen for providing the control signals and for the testing. As shown in Figure 13, all the development board pins have been routed to the right side of the board for easy access. Three voltage regulators have been used to provide +3.3v for the PAD ring supply of the chip, +1.8v for the core V_{dd} of the chip and +1.8v for SRAM V_{dd} . All the recommended capacitors were added to the voltage regulators, where each voltage regulator was chosen to provide a minimum supply current of 150 *mA*. The enable pin of each voltage regulator was routed near the development board, to be able to control the timing and the power up sequence of the chip.



Figure 12: Chip pcb layout.



Figure 13: Testing pcb layout.



Figure 14: Testing setup.



Figure 15: Timing diagram for testing the SRAM PUF.

3.3 SRAM PUF Test Procedure

Figure 15 illustrates the timing diagram for testing the SRAM PUF designed in this test chip. The SRAM PUF test starts with applying "Col_Reset" and "Row_Reset" active low pulses to reset column and row. Then the SRAM power supply (SRAM_VDD) is applied to initialize the SRAM PUF. It is recommended to wait for ~1ms to make sure the SRAM PUF unit cells are all settled down to their final values, before the readout starts.

The readout then starts by applying one pulse to "Row_CLK", which activates the first row of the SRAM PUF unit cells. Then "Col_CLK" is applied and the data ("Bit" and "Bit_bar") signals are readout. The data will appear at the output pin at the falling edge of the "Col_CLK" pulses. Based on the design specs, we recommend the period of 1µs for "Col_CLK". The "Col_CLK" is applied until all 150 columns in the first row is read, from (1, 1) to (1, 150), as shown in Figure 14.

Then the second row of the SRAM PUF is activated by giving a pulse at the "Row_CLK". At the same time a pulse is given to "Col_Reset" so that the readout starts from the first column again. As shown in Figure 14, "Col_Reset" is an active low input and "Row_CLK" is high-to-low edge triggered input. Since the row select and column select are independent circuits, "Col_Reset" and "Row_CLK" can be applied simultaneously. In this phase the second row is read, from (2, 1) to (2, 150), as shown in Figure 14.

This readout process continues until the last row (i.e. row 125) is read, where the values of SRAM PUF at locations (125,1) to (125, 150) are read. The readout process can end at this point, or it can be repeated to verify the values of the unit cells. It is a good practice to repeat the test several times (with and without SRAM power up, SRAM_VDD) and readout the data to test the proper functionality of the chip.

The readout physically starts from top left SRAM PUF unit cell and continues to the right, then it goes to the lower row and continues until the lowest row at the bottom of the structure. It will be valuable information to create a map for the SRAM PUF data, where more analysis of variation span in the SRAM PUF may be extracted.

Based on the design specs, we recommend the width of $\sim 1\mu s$ for all input pulses. Also, there are 3 SRAM_VDD pins on the test chip. It is recommended that the SRAM_VDD pulse comes from a source that can deliver at least 50mA of current at +1.8V without any voltage drop.

22

Chapter 4

Evaluation of 180 nm SRAM test chip

In this chapter, four 180 nm SRAM PUF chips will be tested and evaluated. The nominal operating condition of the evaluation is at room temperature with a 1.8 V supply voltage. The four chips will be evaluated based on the three-metrics discussed previously (Uniqueness, Reliability and Randomness). Since testing four chips is not enough to evaluate the SRAM PUF, therefore up to 80 chips will be tested and evaluated in the future work.

1) Uniqueness

From testing 4 chips of the 125x150-bit (18750 bit) SRAM PUF, a total of six comparison are used to obtain the HD. As shown in Figure 16, the mean HD is 9452, corresponding to a uniqueness of 50.4%. Therefore, the four SRAM PUF chips are distinguishable and unique.



Figure 16: Inter-chip hamming distance (HD) statistical distribution.



Figure 17: Intra-chip bit-error-rate (BER), for 4 chips.

2) Reliability

The four SRAM test chips are tested at two different power supply rising times (30 ms & 10 ns). As shown in Figure 17, it is noticeable that the bit-error-rate (BER) increased by almost the double from 30 ms to 10 ns. The effect of the power supply rising time on the BER will be discussed on chapter 5. In general, the BER of the four chips are in the acceptable range.



Figure 18: Ratio of 1's in the SRAM PUF response for the four test chips.

3) Randomness

As shown in Figure 18, the ratio of 1's in the SRAM PUF response for all of the four chips is just above 30%. As discussed before the ratio of 1's to 0's should be very close to 50%, which is not the case here. This non-uniformity could be due to the large systematic process variations, which will cause the number of 1's in an SRAM array to be much greater than number of 0's or vice versa after power-up [12]. A proposed solution to this problem is utilizing aging effects (mainly NBTI) [12].

Chapter 5



The Effect of Power Supply Ramp Time on SRAM PUFs [13]

Figure 19: SRAM response as a function of fast and slow power supply rising time.

5.1 Introduction

SRAM is one of the popular implementations of PUF. SRAM PUF employs an SRAM cell (two cross-coupled inverters), and exploits the random assignment of a stable state from an initial unstable state. The final state of the cell is determined by the random mismatches in the pair of inverters [11]. The mismatches in the SRAM cells produced during fabrication process could vary from cell to cell. We can classify the SRAM cells based on the mismatches degree, non-skewed cells, partially skewed cells and fully skewed cells [14]. All the work in this paper will be focused on partially skewed cells, since non-skewed cells will always cause bit flipping under any conditions, and fully skewed cells will always produce a stable output under normal conditions.



Figure 20: Two different regions, with different dominant fabrication process variations.

Only few publications have discussed and modeled the start-up value of SRAM PUF. Previous publications focused on the behavior of the start-up value as a function of the supply voltage and temperature only [15]. In this paper, we will present An analytical model for the start-up value as a function of the power supply rising time.

The initial concept demonstrated in this paper is shown in Figure 19, where two different power supply ramps could lead to two different outputs, depending on the cell variations and the power supply ramp time. However, if the variations are very small, or if the cell is symmetrical, the outputs will reach a metastable state, this metastable point does not hold for long, i.e. any small deviation from the metastable point is immediately amplified by the positive feedback and the circuit moves away from the metastable point towards one of both stable points. Since electronic circuits are constantly affected by small deviations due to random noise, the non-skewed SRAM cell will never stay in its metastable state very long instead it will quickly end up in one of both stable states (randomly) [9].



Figure 21: CMOS SRAM cell including the gate capacitances.

There are several mismatches in the cells during the fabrication process, however capacitance and threshold variations are the most crucial mismatches in determining the final SRAM cell state. As shown in Figure 20, there are two regions of operation. Where T is the power supply rising time, which separate between the two regions. If the rising time of the power supply is faster than time *T*, the SRAM cell will be operating in region I, where capacitance and threshold variations decide the final state. If the rising time is slower than *T*, the SRAM will operate in region II, where the threshold variations alone will decide the final SRAM state. To understand how capacitance and threshold mismatches play a key role in determining the SRAM cell outputs, an analytical model will be discussed in section II. Simulation will be performed and presented on an SRAM cell in section III. Experimental data will be shown and discussed in section IV. Finally, the paper will be concluded in section V.

5.2 Analytical model



Figure 22: Supper position between NMOS N_2 current and V_{dd} power supply, at node Q.

In this section an analytical model will be derived for the rising time of the power supply *T*, which will distinguish and separate the two regions, region I and II. The model will also explain how the rising time of the power supply will determine which process variations will decide the final state of the output in both region I and II. The SRAM cell here can be explained using a simple equivalent circuit where the gate capacitances are depicted in Figure 21. Assuming three capacitances are equal to *C*, and the gate capacitance of transistor NMOS N_2 is equal to $C + \Delta C$, where ΔC is the capacitance variation. Since the rising time of the power supply will be gradually increasing, the four transistors will not turn on immediately, and they will be conducting in the subthreshold region. Also, assuming the threshold voltage of transistor N_1 is $V_{thn} - \Delta V_{th}$, where ΔV_{th} is the threshold voltage variation, notice that lower threshold voltage means higher leakage current. For simplification, it can be assumed that the threshold voltage of NMOS transistors is less than that of PMOS, therefore the current of NMOS transistors will dominate, and we can ignore the PMOS currents in our analysis.

A. Derivation

Equation (1) is the power supply ramp function $V_{DD}(t)$, where V_{dd} is the supply voltage, *T* is the power supply rising time, and t is the time domain. As shown in Figure 21(b), the two outputs Q and \overline{Q} , will build capacitive voltage dividers, and will follow $V_{DD}(t)$ based on the ratio between the gate capacitances. At the same time the outputs Q and \overline{Q} are following the supply voltage, the two NMOS transistors are also discharging Q and \overline{Q} . Therefore, the two outputs will increase with V_{dd} at a slower rate. In Figure 21(b), at node Q and \overline{Q} , there is a superposition between the power supply current, and the NMOS leakage current. To derive an equation for *T*, an equation for V_Q and $V_{\overline{Q}}$ will be obtained as a function of V_{dd} , *T* and NMOS leakage current.

$$V_{DD}(t) = V_{dd} * \frac{t}{r} \tag{1}$$

1) At node *Q*, as shown in Figure 22 there is a superposition between the current supplied from the power supply, and the current drawn by the NMOS transistor (subthreshold current)

a) $V_I(t)$ is the voltage supplied by $V_{DD}(t)$ at node Q, and will be calculated by voltage division between the two capacitors as shown in Figure 22(a):



Figure 23: The outputs of the SRAM increasing with the power supply ramp time at a slower rate.

$$V_1(t) = V_{DD}(t) * \frac{C}{2 C + \Delta C}$$
 (2)

b) Substituting (1) into (2):

$$V_1(t) = V_{dd} * \frac{t}{T} * \frac{C}{2C + \Delta C}$$
 (3)

c) As shown in Figure 22(b) $V_2(t)$ is the voltage driven by the NMOS current source at node Q, and will be calculated from the current capacitance relation:

$$I_{leak} = C_{equ} * \frac{\Delta V}{\Delta t} \tag{4}$$

Where I_{leak} is the NMOS leakage current, C_{equ} is the equivalent capacitance which is C in parallel with $C + \Delta C$, and ΔV is $V_2(t)$.

$$V_2(t) = \frac{l_{leak} * \Delta t}{2 C + \Delta C}$$
(5)

d) Therefore, $V_Q(t)$ equals $V_1(t) - V_2(t)$:

$$V_Q(t) = V_{dd} * \frac{t}{T} * \frac{C}{2C + \Delta C} - \frac{I_{leak} * \Delta t}{2C + \Delta C} \quad (6)$$

2) Repeating the previous steps to find $V_{\overline{Q}}$, considering that both capacitances are equal to C, and there is a variation ΔV_{th} in the threshold of the NMOS transistor, which will lead to having I^{Δ}_{leak} instead of I_{leak} .

$$V_{\bar{Q}}(t) = V_{dd} * \frac{t}{T} * \frac{1}{2} - \frac{I^{\Delta}_{leak} * \Delta t}{2C}$$
(7)

3) From Figure 23, it shows that both $V_Q(t)$ and $V_{\overline{Q}}(t)$ are equal, and both will reach V_{thn} exactly at the same time t_{on} . Which means that the capacitance variation ΔC exists in $V_Q(t)$, and the threshold voltage variation ΔV_{th} exists in $V_{\overline{Q}}(t)$, will compensate each other and both outputs will reach the metastable point described previously. Therefore, we can equate (6) and (7) at time t_{on} , and then get an expression for the power supply rising time T, which is in this case the borderline between region I and II.

$$\frac{V_{dd} * t_{on}}{T} * \frac{C}{2C + \Delta C} - \frac{I_{leak} * t_{on}}{2C + \Delta C} = \frac{V_{dd} * t_{on}}{T * 2} - \frac{I^{\Delta}_{leak} * t_{on}}{2C}$$
(8)

4) After some simplifications, and assumptions we can get the final expression for *T*:a) After simplifications:

$$T = \frac{V_{dd} * \Delta C/2}{\left[I^{\Delta}_{leak} * \left(1 + \frac{\Delta C}{2C}\right)\right] - I_{leak}}$$
(9)

b)Assuming, $\frac{\Delta C}{2c} \ll 1$:

$$T = \frac{V_{dd} * \Delta C/2}{I^{\Delta}_{leak} - I_{leak}}$$
(10)

c) The equations of the leakage currents are:

$$I_{leak} = I_o * e^{\frac{V_{GS} - V_{thn}}{n * V_T}} * (1 - e^{\frac{-V_{DS}}{V_T}}) (11)$$

$$I^{\Delta}_{leak} = I_{o} * e^{\frac{V_{GS} - V_{thn} + \Delta V_{th}}{n * V_{T}}} * (1 - e^{\frac{-V_{DS}}{V_{T}}})$$
(12)

d)*From (11) and (12)*:

$$I^{\Delta}_{leak} = I_{leak} * e^{\frac{\Delta V_{th}}{n * V_T}}$$
(13)

Where I_o is the saturation current, V_T is the thermal voltage (25 mv), and n is an empirical parameter (around 1.5).

5) Finally, we can say that *T* is:

$$T = \frac{C * V_{dd}}{I_{leak}} * \frac{\frac{\Delta C}{2 C}}{e^{\frac{\Delta V_{th}}{n * V_T}} - 1}$$
(14)

B. Discussion

From the derived equaiton of *T*, an observation can be made, if there is no capacitance variation in the cell, $\Delta C = 0$, the rising time *T* will be zero, therefore region I disappears. On the other hand, if there is no threshold voltage variation in the cell, $\Delta V_{th} = 0$, accordingly *T* will go to infinity, therefore region II disappears.

Since $V_1(t)$ is the voltage at node Q, which is following the power supply ramp by a ratio of the gate capacitances. Therefore, if the supply ramp is slow enough, $V_1(t)$ can be neglected, and the capacitance variations can be neglected as well, which is the case in region II. While in region I, the supply ramp is fast enough, therefore the capacitance variation cannot be neglected.



Figure 24: SRAM cell transient simulation, using two different rising times.

5.3 Simulation results

Parameter	NMOS	PMOS	
Supply voltage V _{dd} (in V)	1.8		
Length L (in µm)	0.18	0.18	
Width W (in <i>µm</i>)	0.6	0.9	
Threshold voltage inverter 1 V _{th} (in V)	0.3532	- 0.3958	
Threshold voltage inverter 2 V _{th} (in V)	0.335	- 0.3958	
Threshold voltage variation <i>∆V_{th}</i> (in <i>mV</i>)	17		
Gate capacitance C (fF)	150		
Capacitance variation ΔC_1 (<i>fF</i>)	7.5		
Capacitance variation ΔC_2 (<i>fF</i>)	10		
Capacitance variation ΔC_3 (<i>fF</i>)	15		

TABLE 1. PARAMETERS FOR 180NM SBC18H3 MODEL

A. Setup

The start-up value of an SRAM cell is simulated using SPICE and SBC18H3 model. The CMOS parameters used in the simulation are listed in Table. 1. The difference between NMOS V_{thn1} and NMOS V_{thn2} is 5%. A 7.5 *fF* capacitance is added between node \overline{Q} and ground to simulate the capacitance variations, and to balance the threshold variation at node Q.



Figure 25: Threshold voltage variation versus Power supply rising time T, for both the simulation and the model, at ΔC_1 , ΔC_2 and ΔC_3 .

B. Results

Since the threshold voltage of N_2 is lower, transistor N_2 will discharge output Q at a faster rate than N_1 discharging \overline{Q} . However, the gate capacitance of N_2 is larger by 17 *fF*, which will force \overline{Q} to increase at a rate slower than Q. This balance in charging and discharging between Q and \overline{Q} will be separated by the power supply ramp time. After trial and error, it has been found that the boarder time T, between region I and II is around 1.5 *ms*. Which means that at a supply rising time of T less than 1.5 *ms*, the capacitance variation will dominate and \overline{Q} will eventually go to zero. And if T is more than 1.5 *ms*, the threshold variation will dominate and Q will go to zero. As shown in Figure 24, on the left figure the rising time is less than 1.5 *ms*, and on the right figure the rising time is more than 1.5 *ms*. In Figure 25, the rising time T developed from the model is plotted over the range of different AV_{th} with different AC (ΔC_1 , ΔC_2 and ΔC_3), using the parameters in Table. 1. The SRAM cell is also simulated and plotted over the same range of ΔV_{th} with different ΔC (ΔC_1 , ΔC_2 and

 ΔC_3). It can be clearly seen that both the simulation and the model are matching.

5.4 Silicon results

In Figure 26, one of the tests at a power supply rising time of 15 ms is presented as a

map. It can be shown that the faulty bits are evenly distributed, and they are not accumulated

in a certain area, which concludes that the chip is well designed and well fabricated.

Four SRAM test chips have been tested. Each chip has been tested 20 consecutive times, over the range of 10 ns to 30 ms power supply rising time. From Figure 27, two regions



Figure 26: 125×150 bits SRAM PUF response. Black dots are 1's, white dots are 0's and red dots are flipping bits.

can be seen. the first region which is before 14 ms, where the flipping bits is around 10%. The second region, which is after 14 ms, the bit flipping is almost 5%. The higher bit flipping in the first region compared to the second region, can be explained by the fact that at a rising time less than 14 ms, the capacitance variations start to have an effect, and the capacitance variations balanced the threshold variations, therefore resulting in more symmetric cells or non-skewed cells. Since non-skewed cells reaches the metastable state, an external noise will move the cell to one of the stable states randomly.



Figure 27: Extract data from SRAM test chip. Number of flipping bits versus supply voltage rising time.

5.5 Conclusion

In this thesis, the start-up value of an SRAM cell as a function of power supply rising time has been discussed. It has been found that there are two regions of operation. The starup value of an SRAM cell depends on both capacitance and threshold variations in region I, while it depends on threshold variations only in region II. An analytical model was presented, and the power supply rising time T which separate the two regions, has been found as a function of threshold and capacitance variations. The simulation results support the analytical model. The extracted data from the test chip also show two regions. Per the extracted data, in region I, 90% of the cells are fully skewed, and 10% are non-skewed cells. in region II 95% of the cells are fully skewed, and only 5% are non- skewed. Therefore, we can conclude that the rising time of the power supply can change the skewness of some cells.

References

- Christoph Bohm and Maximilian Hofer, "Physical Unclonable Functions in Theory and Practice", Springer, 2012.
- [2] F. Armknecht, R. Maes, A. Sadeghi, F. Standaert, and C. Wachsmann, "A Formal Foundation for the Security Features of Physical Functions", IEEE Symposium on Security and Privacy, 2011.
- [3] R. S. Pappu, B. Recht, J. Taylor, and N. Gershenfeld, "Physical one-way functions," Science, vol. 297, pp. 2026–2030, 2002.
- [4] Ken Mai, "Introduction to hardware security and trust", Springer, 2012.
- [5] S. Eiroa, J. Castro, M. C. Martínez-Rodríguez, E. Tena, P. Brox, I. Baturone, "Reducing bit flipping problems in SRAM physical unclonable functions for chip identification", 19th IEEE International Conference on Electronics, Circuits and Systems (ICECS), 2012.
- [6] C. Herder, Meng-Day Yu, F. Koushanfar, and S. Devadas, "Physical Unclonable Functions and Applications: A Tutorial", Proceedings of the IEEE, Vol. 102, No. 8, August 2014.
- [7] S. Tao and E. Dubrova, "Ultra-energy-efficient temperature-stable physical unclonable function in 65 nm CMOS", ELECTRONICS LETTERS Vol. 52 No. 10 pp. 805–806, 12th May 2016.
- [8] Mudit Bhargava, "Reliable, Secure, Efficient Physical Unclonable Functions", Ph.D. dissertation, Carnegie Mellon University, May 2013.
- [9] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in Proceedings of the 44th annual Design Automation Conference, ser. DAC '07. New York, NY, USA: ACM, 2007, pp. 9–14.

- [10] Lim, D., Lee, J.W., Gassend, B., Suh, G. E., Van Dijk, M., and Devadas, S., "Extracting secret keys from integrated circuits", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2005.
- [11] D. Holocomb, Wayne P. Burleson and Kevin Fu, "Power-up SRAM State as an Identifying Fingerprint and Source of True Random Numbers". IEEE Transactions on Computers. Vol. 57, No. 11, November, 2008.
- [12] Achiranshu Garg and Tony T. Kim, "Design of SRAM PUF with Improved Uniformity and Reliability Utilizing Device Aging Effect", IEEE International Symposium on Circuits and Systems (ISCAS), 2014.
- [13] Abdelrahman T. Elshafiey, Payman Zarkesh-Ha and Joshua Trujillo, "The Effect of Power Supply Ramp Time on SRAM PUFs", Submitted to 60th IEEE International Midwest Symposium on Circuits and Systems, 2017.
- [14] M. Cortez, A. Dargar and S. Hamdioui, "Modeling SRAM Start-Up Behavior for Physical Unclonable Functions", Int. Symp. on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT), 2012.
- [15] R. Maes, "Physically Unclonable Functions: Constructions, Properties and Applications", Springer, 2013.