Computer Engineering
Mekelweg 4,
2628 CD Delft
The Netherlands
http://ce.et.tudelft.nl/

2011

MSc THESIS

Modeling SRAM Start-up Characteristics For Physical Unclonable Functions

Apurva Dargar

Abstract

The security of electronic devices is of crucial importance to companies as well as to users. Moreover, companies that develop Intellectual Property also want to protect them from counterfeiting and overbuilding. Company profits, brand reputations and personal information of the users are at stake if there is a breach in the security of these electronic devices. In the classical approach, a system is secured by storing the cryptographic keys permanently in the non-volatile memories that are present in the security devices. However, this permanent storage of the key makes them easy targets for physical attacks; hence compromising the security of the system. A more secure, cost-effective and elegant solution to this permanent key storage is the use of Physical Unclonable Functions (PUFs). PUF is a method of producing a signature from a physical object, such as an Integrated Circuit, by relying on the non-reproducible physical attributes of a device. These signatures are unique because fabricated circuits exhibit slightly different electrical behavior from one another even if their design, mask and manufacturing process are identical. Various kinds of PUFs exist; examples are Optical PUF, Butterfly PUF and SRAM PUF. However, the start-up value based SRAM PUFs appear to be the most promising ones for usage in ICs.

Although the SRAM PUFs are becoming very popular, only a little has been published in the field of modeling and analysis of their start-up behavior. Reproducing the same start-up behavior, every time the chip is powered-on, is very crucial in order to produce the same cryptographic key. This thesis proposes an analytical model for the start-up value based SRAM PUFs; it helps in understanding the impact of both non-technological parameters (such as supply voltage and temperature) as well as technological parameters (such as the geometry of the transistors and threshold voltage) on the behavior of the start-up values of an SRAM. Various experiments have been performed to analyze and quantify their impact. The results obtained indicate a major impact of the non-technology parameters. The reproducibility of start-up values becomes more likely with slower ramp-ups and lower temperatures. For example, the percentage of reproducible bits increase from 93.5% at $1\mu s$ ramp-up to 96% at $10ms$ ramp-up. Amongst the technology parameters, it is observed that a small mismatch of 1.6% in the threshold voltage is enough to flip the start-up value of the cell for $65nm$ technology. These results have been validated by comparing them with actual silicon data measured at Intrinsic ID. The validation of the results proves the correctness of the analytical model proposed and gives a proof of robustness of the start-up values.

CE-MS-2011-11

Modeling SRAM Start-up Characteristics For Physical Unclonable Functions

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

COMPUTER ENGINEERING

by

Apurva Dargar
born in Bhilwara, India

Computer Engineering
Department of Electrical Engineering
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

Modeling SRAM Start-up Characteristics For Physical Unclonable Functions

by Apurva Dargar

Abstract

The security of electronic devices is of crucial importance to companies as well as to users. Moreover, companies that develop Intellectual Property also want to protect them from counterfeiting and overbuilding. Company profits, brand reputations and personal information of the users are at stake if there is a breach in the security of these electronic devices. In the classical approach, a system is secured by storing the cryptographic keys permanently in the non-volatile memories that are present in the security devices. However, this permanent storage of the key makes them easy targets for physical attacks; hence compromising the security of the system. A more secure, cost-effective and elegant solution to this permanent key storage is the use of Physical Unclonable Functions (PUFs). PUF is a method of producing a signature from a physical object, such as an Integrated Circuit, by relying on the non-reproducible physical attributes of a device. These signatures are unique because fabricated circuits exhibit slightly different electrical behavior from one another even if their design, mask and manufacturing process are identical. Various kinds of PUFs exist; examples are Optical PUF, Butterfly PUF and SRAM PUF. However, the start-up value based SRAM PUFs appear to be the most promising ones for usage in ICs.

Although the SRAM PUFs are becoming very popular, only a little has been published in the field of modeling and analysis of their start-up behavior. Reproducing the same start-up behavior, every time the chip is powered-on, is very crucial in order to produce the same cryptographic key. This thesis proposes an analytical model for the start-up value based SRAM PUFs; it helps in understanding the impact of both non-technological parameters (such as supply voltage and temperature) as well as technological parameters (such as the geometry of the transistors and threshold voltage) on the behavior of the start-up values of an SRAM. Various experiments have been performed to analyze and quantify their impact. The results obtained indicate a major impact of the non-technology parameters. The reproducibility of start-up values becomes more likely with slower ramp-ups and lower temperatures. For example, the percentage of reproducible bits increase from 93.5% at $1\mu s$ ramp-up to 96% at $10ms$ ramp-up. Amongst the technology parameters, it is observed that a small mismatch of 1.6% in the threshold voltage is enough to flip the start-up value of the cell for $65nm$ technology. These results have been validated by comparing them with actual silicon data measured at Intrinsic ID. The validation of the results proves the correctness of the analytical model proposed and gives a proof of robustness of the start-up values.

Laboratory : Computer Engineering
Codenumber : CE-MS-2011-11

Committee Members :

Advisor:	Dr.Ir. Said Hamdioui, CE, TU Delft, The Netherlands
Chairperson:	Dr.Ir. Koen Bertels, CE, TU Delft, The Netherlands
Member:	Ir. Geert-Jan Schrijen, Intrinsic ID, The Netherlands
Member:	Dr.Ir. Jaap Hoekstra, TU Delft, The Netherlands
Member:	Dr.Ir. Zaid Al-Ars, CE, TU Delft, The Netherlands

To my parents for believing in me and my brother for all the encouragement

Contents

List of Figures

List of Tables

Acknowledgments

The fact that you are now holding this report is a testament to the hours that I have put in performing simulations and reading books and research papers. But reaching this stage would not have been possible without the support and counseling of several people who helped me, directly or indirectly, in completing it.

First and foremost, I would like to thank my supervisor, Dr.Ir. Said Hamdioui for his expert guidance. His eye for detail is amazing which helped me keep focused in the right direction. I would also like to thank him for giving me the opportunity to do a thesis in a rather new and interesting area. I would also like to thank all the people associated with this project starting with Mafalda Cortez, my mentor, for always having my back and for helping me in reviewing this thesis. Furthermore, I would like to thank Mathias Claes, for performing the measurements on the SRAM devices which helped me verify my hypothesis. A special thanks to all the people in the CE department for providing a congenial environment for me to work in.

I would like to thank Rahul and Shishir for teaching me that any problem can be solved simply by writing proper equations, Rachit for helping me write those equations and Minni for always being there. The hours spent in playing cards, watching movies, having dinner parties made my life in Delft quite memorable and enjoyable. I greatly acknowledge Shekhar for his encouragement to finish my thesis on time. I appreciate my friends back home specially Khyati, Ritika, Rohan, Harshal, Vineet and Saransh for being so supportive and encouraging.

Lastly but definitely not the least, I would like to thank my parents and brother for believing in me and trusting me enough to allow me to study at TU Delft. Without their love, trust and constant support, none of this would have been possible.

A popular African philosophy says that "I am what I am because of who we all are". This could not be truer in my case.

Apurva Dargar
Delft, The Netherlands
July 26, 2011

1 Motivation

This chapter introduces some basics about security systems using cryptographic keys, motivates the work, stresses on the main contributions of the work and provides the outline of the thesis.

Section 1.1 discusses the need of various kinds of cryptographic key based security solutions and concludes that the use of Physical Unclonable Functions (PUFs) is the best. Section 1.2 elaborates the concept of SRAM PUFs which are the topic of this thesis. The major contributions of this thesis are highlighted in Section 1.3 and the chapter concludes by describing the outline for this thesis in Section 1.4.

1.1 Introduction to Information Security

The digital revolution that has occurred over the past few decades has facilitated transfer of huge amount of data at a very high speed across the globe. This revolution has necessitated the use of applications that require a large number of on-line transactions. From the indispensable on-line banking to the zestfulness of on-line gaming, from limiting the access to a government facility to providing access to a research lab, almost every application used nowadays involves such interactions. The data that is exchanged during these interactions can be anything ranging from personal information of a user to huge amount of money. Therefore, to protect the confidentiality and authenticity of these interactions, it is required to have some degree of security incorporated in the designed systems. The objective of these systems is to protect the information and property from theft while allowing them to remain accessible and productive to its intended users. The field that comprises all the sciences, tools and techniques dedicated to protect information and information systems is the field of Information Security.

To protect the information while being stored or transported over a communications link, it is a common practice to use cryptographic techniques such as encryption and signing algorithms. The cryptographic algorithms used for this purpose are often publicly available, but the secret key used by them is stored securely in the system. It is crucial that this key remains completely secret to ensure the security of the system. Most of the traditional security systems that we see around are based on storing this key or information in a *Non-Volatile Memory* (NVM). Examples are Smart cards used to access universities, Set-top boxes used by various television companies to provide access to TV channels, credit cards used for monetary transactions, etc. The major domains of application of these security systems are preventing theft of service, denial of service, cloning and overbuilding [31].

All these solutions, although secure, are prone to attacks by adversaries. Therefore, one cannot talk about security without talking about security breach. Security breach occurs when an unintended individual or a group of individuals circumvent the security system and get access to the information or property protected by the system. The security of a system is determined by the asymmetry in the effort that is required to protect the information and information systems and the effort that is required to break its protection [26]. A breach in any of the modern day security systems can mostly either be in the form of a cryptographic attack, implementation attack or a physical attack [25]. Since the cryptographic algorithms used nowadays are very secure, physical attacks and implementation attacks become easier alternatives for the adversaries [25]. Thus, it is important to analyze the impact of physical and implementation attacks on a security system. These attacks, if successful, will not only provide the adversary with confidential information of the users but also may result in huge monetary losses. For example, in 2006, TXJ Companies Inc. found out that 45.6 million credit and debit card numbers were stolen from one of its systems over a period of more than 18 months by an unknown number of intruders [38]. The personal information provided by the users in return for these merchandise were also stolen along with it. The main flaw here was not just the data theft but the fact that it took the company 18 months to realize that there has been a security breach. Such examples emphasize the fact that irrespective of the level of security of the system, there is a possibility to break into the system. The existing security systems based on NVMs permanently store the key *somewhere* in the system. Any individual who has an access to the system and knows about its implementation can breach the security of the system. Therefore, more secure and tamper-evident security solutions need to be proposed.

One such solution is the use of *Physical Unclonable Functions* (PUFs) which were first proposed in [26]. These security solutions are based on the physical properties of a system which are non-reproducible even for the manufacturer. The key which is used to provide the access is never stored in the system, rather it is generated whenever required. These properties make PUFs an ideal candidate for protecting the assets. Amongst the various kinds of PUFs that have been proposed in literature, SRAM PUFs look the most promising and are becoming very popular [2][22][23]. However, very little has been published in the field of modeling and analyzing of their start-up behavior; this is the topic for this thesis.

1.2 Start-up value based SRAM PUFs

SRAM PUFs are based on the technology called Hardware Intrinsic Security which is in turn based on PUFs. Since, SRAMs are one of the most widely used components in modern day System-on-Chip, no extra effort and money needs to be invested in implementing the start-up value based SRAM PUFs. Therefore it improves time-to-market which not only makes the SRAM based systems cost-effective but also improves their time-to-market. The key component which determines the start-up value of an SRAM PUF is the SRAM cell which consists of two cross-coupled inverters along with two access transistors. However, the role of access transistors in determining the start-up value can

be neglected because they are disconnected from the cell during the start-up. Therefore, it is just the mismatch between the cross-coupled inverters that decides the start-up value (Section 4.2.2). In order to be used as a cryptographic security key, SRAM PUFs need to have certain characteristics. For example, the key generated by every SRAM should be reliable and unique. In other words, PUFs' signature needs to be reproducible for the same device whereas there should be no correlation between the signature from one device and another.

As already mentioned, SRAM PUFs provide a better security solution and are becoming very popular. Nevertheless, there is nothing in the literature about their start-up value analysis. Understanding the different parameters that impact this value and quantifying their impact is very important for designing reliable and robust SRAM PUF based security systems. This describes the motivation for this thesis. The goal of this thesis is to model the SRAM in terms of its start-up behavior, such that the influence of technology parameters and the non-technology parameters can be predicted and investigated. This modeling is crucial for evaluating the robustness of the SRAM PUFs and for assessing the products like SRAM-PUF based secure key storage systems.

1.3 Contributions of the thesis

To the best of my knowledge, no physical modeling of the SRAM PUFs has been published so far. In this respect, this thesis makes the following contributions:

1. An *analytical model* to analyze the stability of the start-up value of SRAM PUFs.
2. A *classification* of the SRAM cells based on the reproducibility of its start-up value.
3. *Analysis* and *quantification* of the impact of the technology and non-technology parameters on the start-up value of SRAM PUFs.
4. Experimental results to prove the robustness and uniqueness of the start-up values of the SRAM PUFs.
5. *Validation* of the model by comparing the experimental results with the actual silicon data obtained from our industry partner.

1.4 Outline of the thesis

This thesis is organized as follows:

Chapter 1 describes the motivation for pursuing this thesis along with the contributions of the work.

Chapter 2 presents an introduction to the field of information security along with a survey of the existing techniques in this field, their applications, advantages and their shortcomings. The survey helps in understanding the requirements of a security system in general, and an SRAM PUF based security system in particular.

Chapter 3 discusses the start-up value based SRAM PUFs and the SRAM cell which forms the key component for these PUFs. It also discusses process variation which impart the properties of unclonability and randomness to the SRAM PUFs. A classification of SRAM cells based on the extent of the process variation is proposed as well.

Chapter 4 details the analytical model developed; the model is based on the PUF Static Noise Margin, a metric used to measure the stability of SRAM start-up value. Moreover, an in-depth analysis of the analytical model is done to determine the stability parameters that impact the start-up value of an SRAM cell.

Chapter 5 presents the experimental results for $65nm$ technology node and quantifies the impact of technology and non-technology parameters on the start-up values of SRAM PUFs. A proof of robustness of the start-up values is also provided in the chapter.

Chapter 6 focuses on validating the experimental results obtained using the industrial results.

Chapter 7 concludes this study by listing a number of conclusions; it also provides few recommendations for future work.

An Overview of Key-Based Security Systems

2

With increasing need of key-based security systems to protect the digital information, Information Security is becoming a field of extensive research. Most of the modern systems are based on storing the key in a Non-Volatile Memory; this makes them prone to attacks. Therefore, a new field of Physical Unclonable Functions has evolved. They rely on using process variations in a physical system rather than storing the key in the system itself; hence reducing the chances of attacks. This chapter provides an overview of the most commonly used security systems and highlights their pros and cons. The aim of this chapter is to explain the purpose and provide the framework of this work.

Section 2.1 describes a security system in general and provides a list of requirements for a modern security system. The most prevalent security systems used nowadays are described in Section 2.2. Section 2.3 describes the security systems based on Physical Unclonable Functions which form the core of this thesis. Section 2.4 describes the PUFs in which the randomness is explicitly introduced, namely the Coating PUF and the Optical PUF. Section 2.5 elaborates on the PUFs that utilize the intrinsic randomness for their operation like the SRAM PUFs and the Butterfly PUFs. Section 2.6 summarizes the literature study while highlighting the advantages of PUFs in general and SRAM PUFs in particular.

2.1 Introduction

As mentioned in Section 1.1, *Information Security* (IS) is the field that comprises all the sciences, tools and techniques dedicated to protect information and information systems. Over the past few decades, a large number of security primitives have been developed to protect the digital information used for various purposes [36]. The information is protected using *Personal Identification Numbers (PIN)*, passwords, magnetic strips which only the user knows or possesses. In other words, the use of such procedures identify the bearer as the only person authorized to perform a certain action.

Figure 2.1 shows a traditional security system where a user has to enter a set of credentials such as username and password as the first step to login to the system (denoted by step number 1 in the figure). The information entered is then matched with the one stored in the database (DB) via the terminal. If the matching occurs, the user is allowed to access the system whereas access is denied if wrong information is entered. An unfortunate consequence of digital revolution is that copying and tampering information is extremely easy in many cases. In the system shown in Figure 2.1, the adversary can attack the system if he manages to get access to the database storing the information or by eavesdropping the credentials of the user.

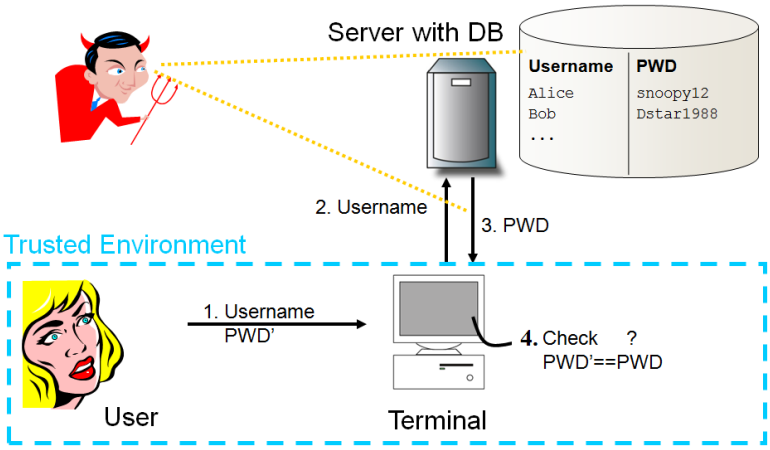

Figure 2.1: Overview of a traditional security system

The traditional way in which a key-based security system works involves key-storage in *Non-Volatile Memories (NVM)*. Since NVMs store the key even when the power supply is removed, a security system using them becomes very prone to attacks. In a security system, physical attacks may occur in various forms, but generally the physical attacks or implementation attacks are performed by the adversaries. There are three main reasons for the successful physical and implementation attacks on a security system based on NVMs. First, the permanent storage of the key in the system even when the key is not needed. Second, the permanent storage of the key even when it is no longer integrated in the rest of the system. Third, the lack of tamper evidence as the key's value is not affected by an attack. In short, most of the security systems utilizing NVMs are not secure enough to prevent data theft through physical attacks in todays world [31].

The biggest challenge that today's security systems are facing is to develop solutions which are more resistant to such physical attacks even if they operate in hostile environments. The basic criterion that any modern security system should fulfill is that its key should be unique and the system should be tamper evident in adverse circumstances. Moreover, it should not be easy for an adversary to predict the key for the security system. The system should be such that in order to break through it, an attacker would need a combination of computational resources and time that are not available. In addition, the resources invested in the development of such systems should be much less than the ones necessary to break them. In summary, paraphrasing Kerckhoffs' principle: "A system should be secure even if everything about the system, except the key, is public knowledge" [18]. This means that the security that a system is able to provide is measured by the ability that a system has to protect its key. For all the rest, such as the system's architecture, its protocols, etc, it should be assumed that the attacker has a knowledge of system's details. In order to understand any security system and its pros and cons, we first need to understand its basic requirements. A security system, in general, should have the following characteristics [36]:

1. An adversary should not be able to predict the key based on invasive and non-

invasive measurements.

2. The storage medium should have read-only access for a user so as to prevent the adversary from tampering with the instruction set of the system.

3. The storage medium in the security system should be bound inseperably to the whole system such that any attempt to attack the storage medium should result in a substantial damage of the system. This helps in making the system tamper-proof.

4. The system should be physically and mathematically unclonable so as to make the key prediction almost impossible.

These characteristics define the requirements of a security system in general and most of the modern-day security applications follow these requirements. One of the best possible security system fulfilling all these requirements is a physical system whose output is totally random and unique for every individual like a human fingerprint, human eye-scan, a Physical Unclonable Function, etc. The following section gives a brief description of few such systems which differ in the way they store their keys. Therefore, various key-storage mechanisms for security systems are discussed next.

2.2 Key-storage mechanisms for security systems

Regardless of the type of methods chosen to protect a system, at the end, the security of a system is defined by the ability of this system to protect its key. Several schemes have been developed to protect or make the access to the key harder but still the key is stored in a NVM or a battery-backed volatile memory for most of the security applications. The last few decades have seen a quantum leap in the field of security. The primitives developed have solved many traditional security problems and achieved a high level of sophistication [36]. The modern day security applications are being developed such that they do not have to rely upon NVMs for key storage, instead the key is generated whenever needed. Such systems are emerging and there is a lot of research that is being carried out in this field. With this research, new developments have been accomplished not only in the field of traditional cryptography but also in areas such as biometrics, anti-counterfeiting, and brand protection which are based on noisy inputs [36]. Figure 2.2 shows the various kinds of key storage mechanisms for security systems and the following sections describe each of them along with their respective pros and cons.

2.2.1 Non-Volatile Memory based security systems

These are the most commonly used security systems wherein the key is permanently stored inside the system in a *Non-Volatile Memory* (NVM). Most of our everyday applications use a NVM based system for security. Examples are: A bank ATM card, a smart-card to access the university or a mobile SIM card, etc.

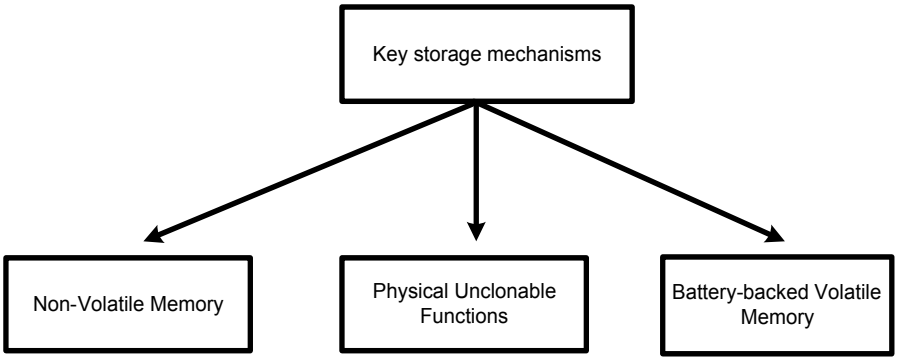

Figure 2.2: Various kinds of key storage mechanisms

***Characteristics and current modalities:**

The basic characteristics of a NVM based security system are similar to any general security system as described in the previous section. Along with this, mostly a smart card is used to embed the NVM in the security system. Therefore, it should be such that it is easy to carry and hence be small in size. Also, it should be able to communicate with external services via card-reading devices such as in ATMs or coffee machines. Nowadays, we see a large number of applications that use such security systems in various modes. Contact smart cards serve as ATM cards, access control cards, public transport cards, etc. Depending upon the level of security required for the application, smart cards may be combined with password protected applications for identification. Also, contact-less smart cards are used in a variety of applications. In such applications, the card communicates with the reader through RF induction technology. Such cards find application in our daily lives in public transportation systems and selected access systems.

***Pros and cons**

Since smart-cards are used for most of the NVM based systems, this section highlights the pros and cons of a smart-card based security system. These systems have revolutionized the modern day payment card industry along with an increased level of security. The traditional methods used for the same purpose were less efficient as they did not use any encryption and authentication technology as used for security nowadays. Another major advantage of a smart card based systems is that a single smart card can be used for authentication in numerous applications. Hence a single card can be used for establishing the identity of an individual in all official or bureaucratic transactions, as is being implemented in India by the name of Multipurpose National Identity Card. Also, since these kinds of systems are embedded in smart cards, they are light weight and easy to carry and this makes them so widely implemented in our everyday lives.

With all the above mentioned advantages of smart card based systems, they also have some disadvantages. The major disadvantage is their susceptibility to chip damage

as they are often carried around in wallets or pockets, which are harsh environmental conditions for the memory or IC used in smart cards. Also, due to the smaller size of the cards that these systems are embedded in, the bearers of such cards can misplace them and if they fall into the wrong hands, they can be misused. Apart from these circumstances, even if we implement a greater level of security by using a combination of passwords along with the card, an adversary can easily steal the card leaving the security measures redundant. Also, such systems can lead to potential identity theft. They can be used by criminals seeking a new identity as they can contain a lot of information about their owner.

2.2.2 Battery-backed volatile memory based systems

This is the least frequently used storage mechanism for security systems and hence will be talked about in brief. The basic setting in which a battery-backed volatile memory is used involves using an SRAM attached to a battery for its power supply requirements. The SRAM here is used to store the key. However, there are additional costs for having a battery on a printed circuit board and therefore such systems have higher costs. Moreover, these systems are less reliable due to the fact that even if the contact to the battery is lost for a brief moment, the key stored inside the SRAM can be lost. In addition, the battery should be able to keep its life for years which is not always possible. These disadvantages make the battery-backed volatile memory based systems less frequently used.

2.2.3 Physical Unclonable Functions for security systems

Humans have features which are unique and are difficult to clone and hence such features can be used for the purpose of identification. This concept of biometrics for humans when applied to silicon or other objects brings us to what is called Physical Unclonable Function (PUFs) which are based on biometrics of objects. Traditional use of PUFs traces back to 1960 when Wienser proposed an anti-counterfeiting method [36] that makes use of a very important property of quantum physics which states that it is impossible to duplicate a quantum state with high probability of success. By equipping an object with a quantum state, it is ensured that adversaries cannot clone the object but the authorized user can validate his/her identity.

*Characteristics and current modalities

PUFs basically exploit the unavoidable IC fabrication process variations for their operation. The system based on PUF cannot be reproduced even by the manufacturer. One more important aspect of PUF is its ability to give the same response for a given device under different environmental conditions. PUFs will be discussed in detail in the next section but a general idea about their applications as well as their pros and cons is given below.

PUFs are physical structures embedded, typically, in Integrated Circuits. They are characterized by Challenge-Response Pairs (CRPs). The challenge, from CRP, is a physical stimulus applied to the PUF, e.g. voltage applied to an IC device. As a consequence of this stimulus, the PUF will generate a value that is measured, e.g. the time that the IC takes to produce a result. This value is the response of the PUF. Since, due to process variation, the parameters of each PUF are unique, each device has a unique and unpredictable collection of CRPs, which is same as saying that each PUF is unique and unpredictable. This unique characteristics of PUFs makes them powerful allies in the security field since they can identify ICs unequivocally.

An example of system using PUF would be start-up value based SRAM PUF which can be used for anti-counterfeiting measures relevant to an FPGA [2]. The procedure here is that the manufacturer notes the unique identifier of every FPGA based on the SRAM initial signature. For every FPGA that is shipped, the initial signature found on the product at the point-of-sale is compared with the one recorded before shipping to validate the authenticity of the FPGA.

***Pros and cons**

PUFs represent a promising approach to meet the requirements of high-security applications as they carry numerous advantages. The foremost of these advantages being that they are highly secure and robust due to the fact that they do not store any secret keys like the conventional security systems. Also, characterization of a PUF structure needs a lot of technical expertise and hence it would be very difficult to characterize it, if not impossible. Another important point to note here is that characterization of a single PUF would not reveal any information about another similar PUF. Moreover, PUFs do not contain any information sensitive to the user as is the case in biometrics based systems. Since PUFs are based on the process variations that are unavoidable and random, it is impossible to manufacture an identical copy of a PUF.

Like all other security systems, the ones based on PUFs also have some limitations and disadvantages. Extreme environmental conditions may affect the PUF based systems;factors like humidity, radiation and temperature can have an impact on the expected behavior of an IC. It has been shown that it is possible to change the behavior of an IC by exposing it to very low temperatures [31]. Additional processing is required to correct such influences. However, since efficient Error Correction Codes exist and can be applied , and this makes PUFs a very promising approach in the field of security. Given the kind of advantages that PUFs offer, it indeed looks very promising and the rest of the chapter focuses on discussing the various kinds of PUFs in detail.

2.3 Physical Unclonable Functions

PUF is basically a method of producing a signature from a physical object and hence its definitions may vary according to the purpose it is used for. In general, a PUF can be characterized by defining the words it consists of as follows [22]:

- **Physical:** A PUF is by definition embedded in a physical system, therefore it should not be (is not) a purely mathematical function, but its outcome arises out of a physical interaction.
- **Unclonable:** Given an instance of a particular PUF, it should be hard to (physically) reproduce it such that the exact functionality is preserved.
- **Function:** A PUF behaves like a function, i.e. it can take an input which is called a challenge and produces an output which is called a response. However, it is not a function in a strictly mathematical sense, since each challenge can map to more than one response.

PUF technology is the latest breakthrough in the field of semiconductor security. As we know, semiconductor manufacturing process has unavoidable variations. Any circuit design, when fabricated in silicon, exhibits slightly different electrical behavior from one chip to another even if the design, mask and fab are identical. This forms the basis of the PUF technology which proves highly beneficial in elevating the security level of a number of IC based systems used nowadays.

The normal way in which these physical objects are used is to apply a stimulus or challenge and measure the output which is called response. For every challenge, there must be a valid response for an object to be authenticated. This physical system of challenge and response is referred to as PUF. In 2001, Pappu [26] proposed the idea of Physical One Way Functions (POWFs) in a challenge-response setting such that the object can be subjected to a large number of challenges all of which produce an unpredictable unique response. The same applies to modern day PUFs which have the same functionality and characteristics. The basic characteristics of the response produced by such systems should be the following:

- Unpredictable yet persistent,
- Impossible to model or replicate and
- Beyond manufacturer's control

As already mentioned, amongst the various security primitives that have been discussed in Section 2.2, PUFs are the most suitable ones for high-security applications. A brief idea about their working principle and their basic characteristics has already been explained in Section 2.3. PUFs can be categorized into two classes depending upon the source of randomness which makes them unique. In the case when PUFs utilize the inherent process variations for their functioning, they are called the *Intrinsic randomness based PUFs*. On the other hand, if randomness is extrinsically introduced in the system to provide it randomness, they are termed as *Extrinsic randomness based PUFs*. Figure 2.3 presents a classification of various kinds of PUFs available; they are discussed in the next two sections.

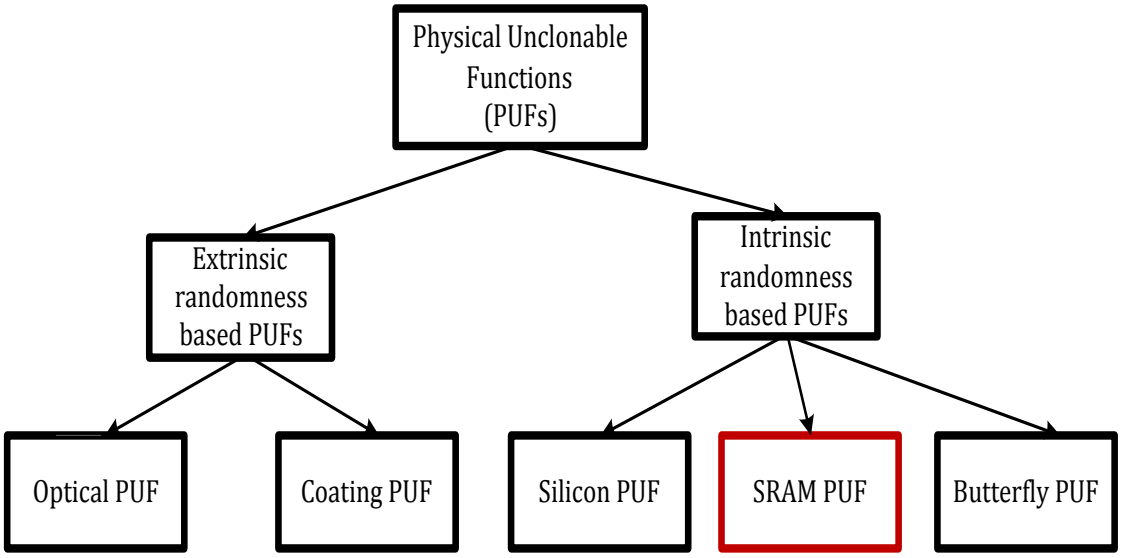

Figure 2.3: Classification of PUFs

2.4 Extrinsic randomness based PUFs

These types of PUFs utilize the randomness which is explicitly introduced in the system. The manufacturer or the user can choose the material or size of particles to introduce this randomness; however, their end location and distribution cannot be controlled. Therefore, the process is still random and the response still unpredictable. The foremost advantage that these kinds of PUFs provide is that since the randomness is extrinsic, its parameters can be optimized and controlled in a much better way and hence they have much greater ability to be distinguished from each other. There are two kinds of Extrinsic randomness based PUFs; see Figure 2.3. They are explained in next two subsections.

2.4.1 Optical PUFs

An Optical PUF consists of a transparent medium such as glass, in which light scattering particles are explicitly introduced. The placement of light scattering particles in the medium is totally random. When such a medium is hit by a coherent laser beam; a speckle pattern is observed which not only depends on the characteristics of the laser beam like wavelength, incident angle (θ), incident location, etc. but also on the position of the scattering particles. Since, the interaction between the laser and the particle is very complex, it is almost impossible to duplicate the Optical PUF such that the same speckle pattern is reproduced. Figure 2.4 shows an Optical PUF where a laser beam hits a transparent medium and results into a unique speckle pattern. These PUFs were first proposed in [26].

The benefit of using Optical PUFs is that it can be challenged in numerous ways as changing the angle (θ) as shown in Figure 2.4) of the laser beam results into a completely

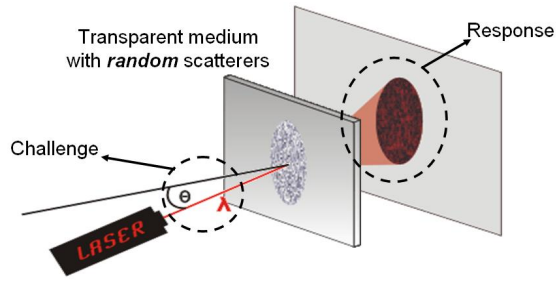

Figure 2.4: Optical PUF [22]

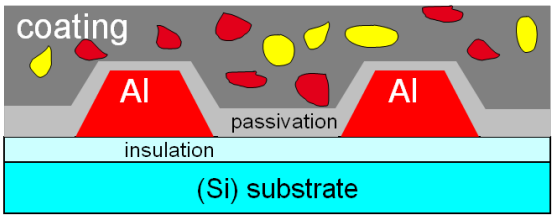

Figure 2.5: Coating PUF [37]

new response hence creating a different challenge-response pair. A major disadvantage of Optical PUFs is that its response to the challenges is not binary whereas most the contemporary security primitives use computers which are based on binary numbers. Hence, the output from optical PUFs has to be quantized before it can be used for such applications. Moreover, the readout equipment required for Optical PUFs is very expensive and hence this limits their usage.

2.4.2 Coating PUFs

A coating PUF is *prima facie* an approach of building a PUF on the top layer of an IC. This layer consists of a protective opaque coating which contains random sized dielectric particles. The randomness in this PUF comes from the random size and placement of these dielectric particles. The top layer of the metal of the chip underneath the coating contains comb shaped sensors. Dedicated circuits on the chip use the sensors to accurately measure the capacitance of a particular part of the coating. Due to the randomness of the coating particles, this capacitance will be random and unique for every sensor on every chip. This property makes coating PUFs unclonable and its response random yet persistent. Titanium dioxide (TiO_2) and Titanium Nitride (TiN) are the most commonly used dielectrics for this purpose [22].

Figure 2.5 shows a Coating PUF [37]. Here, the protective coating not only acts as a PUF but also prevents the circuit below it to be inspected by the attacker due to lack of visibility of the circuit underneath. Moreover, when an attacker tries to remove the coating or a part of it, the capacitance is bound to change and the original unique identifier will be destroyed. This is one of the most important advantages that coating PUFs provide. The problem with using Coating PUFs is to design a dedicated measurement circuit and the coating itself has to be done to every circuit which in itself is a difficult task.

2.5 Intrinsic randomness based PUFs

Unlike the Extrinsic randomness based PUFs, Intrinsic randomness based PUFs use randomness intrinsic in them for their operation. This intrinsic randomness occurs due

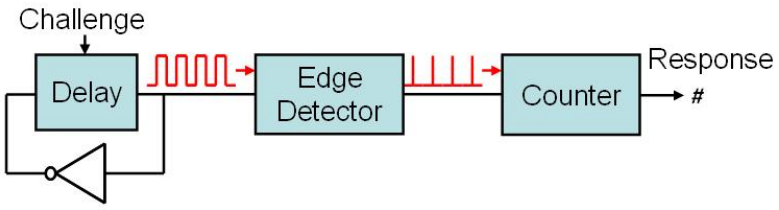

Figure 2.6: Ring-oscillator PUF [22]

to the process variation during the manufacturing process which is unavoidable and random in nature. Such PUFs are highly attractive because they can be included in any design without any modifications to the manufacturing process. An added advantage of intrinsic PUFs is that the majority of the proposed PUFs based on them give a digital response and hence need no quantization to be used for security purposes. There are three kinds of Intrinsic randomness based PUFs [22] as shown in Figure 2.3; they are described next.

2.5.1 Silicon PUFs

Silicon PUFs utilize the random variations in gate and wire delays in a circuit for their operation. Given an input challenge, dedicated delay circuitry in the circuit measures the propagation delays of this input to the output via different paths. Depending on the path through which the signal reaches first, an arbiter assigns a 1 or a 0. The Silicon PUF also uses the process variations for the randomness. The basic idea here is that even a circuit with the same layout mask has some random variation in the delays of different paths when fabricated on different chips.

Two PUFs using this principle have already been implemented namely Ring Oscillator PUF [22], which is based on a delay loop, and a Multiplexer-based PUF [22], which combined with an RF interface is used for RFID anti-counterfeiting applications.

Figure 2.6 shows a Ring Oscillator PUF which is a prime example of silicon PUF. Instead of directly measuring the delay, it transforms a digital delay path into a ring oscillator by feeding back the inverted output to its input. An edge detector generates a pulse every time a rising edge occurs in the oscillation. By counting the number of pulses over a predetermined time period, one gets a measure of the frequency of the ring oscillator, and hence of the delay of the circuit. Because of random variations of this delay, the counted number of pulses will contain a random, PUF-specific, component. This pulse-count is the response of the ring oscillator PUF [34][22].

The advantage that these PUFs provide as compared to extrinsic randomness based PUFs is that they do not need any special treatment to generate the randomness. Also, an added advantage of these PUFs is their small size and since they can be spread everywhere on the chip, this makes it difficult to detect and tamper with. The major disadvantage of Silicon PUFs is the design effort needed and the extra circuitry to measure the response [22].

2.5.2 Butterfly PUFs

Butterfly PUFs appear as a solution for the protection of Intellectual Property in FPGAs that do not use integrated memories or for FPGAs which force their memories to be initialized to a 0 state on start-up citebutterfly. The basic principle behind the working of Butterfly PUF is the utilization of the bi-stable structure of two cross coupled inverters. The circuit can be easily driven from an unstable state to a stable state by applying an external signal to the input and due to slight differences in the inverters that are used to build the circuit. We use this fact to build a PUF where the circuit is initially at the unstable operating point and is left to attain one of the two stable operating points without any external excitation. We find the high probability of the circuit going to a particular state due to process variations. One pair of cross-coupled inverters provide one bit of information and hence a large number of such circuits put together act as a PUF.

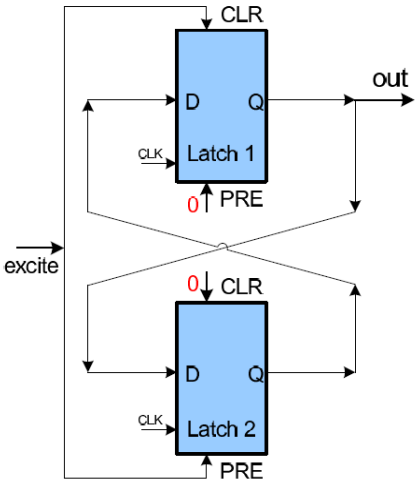

Figure 2.7: Butterfly PUF [20]

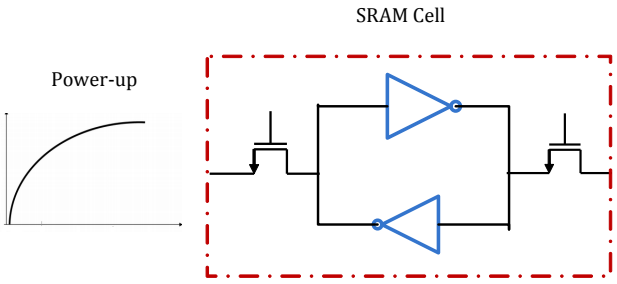

Figure 2.8: SRAM PUF

Figure 2.7 shows a butterfly PUF with two latches which are cross-coupled. Each of the latches has a preset (PRE) signal (which turns output Q to 1 on high) and a clear (CLR) signal (which turns output Q to 0 on high). The data D is transferred to the output Q when the CLK is high. We set CLK in both latches to always high, effectively simulating a combinational loop. To start the PUF operation, the excite signal is set to high. This brings the PUF circuit to an unstable operating point due to opposite signals at inputs and outputs of both latches. After a few clock cycles, the excite signal is set to low. This starts the process wherein the PUF circuit attains either one of the two possible stable states, 0 or 1, on the out signal. The stable state depends on the slight differences in the delays of the connecting wires which are designed using symmetrical paths on the FPGA matrix. Hence, these slight variations are only based on the intrinsic characteristics of the integrated circuit and vary from device to device and the position on the FPGA [20].

The advantage of butterfly PUFs as compared to SRAM PUFs is that it can be put anywhere on the IC thus making it more difficult for an attacker to hack into the system.

However, the cross-coupled latch system needs to be tested thus adding to the cost. Another major disadvantage of a butterfly PUF is that the length of the wires connecting the output of the latch to the input of another latch needs to be exactly equal for both latches. This can however be a problem when the Butterfly PUFs are fabricated in silicon. Their other disadvantages are the same as shared by all PUFs - sensitivity to variation in temperature, humidity, etc.

2.5.3 SRAM PUFs

The SRAM PUF is also an intrinsic PUF which uses the random fluctuations of the silicon components caused by process variations as a source of randomness. This randomness is utilized for producing unique signatures of the embedded IC. However, contrary to Silicon PUFs, the SRAM PUF is not based on a delay measurement; but on the intrinsic mismatch present between the two inverters of an SRAM cell to determine the start-up value of the SRAM cell. Ideally both the inverters should be identical, but due to manufacturing variability, there is always some random offset between the two inverters. Hence, the start-up state of an SRAM cell is determined by this mismatch which is random as well as unclonable. It is a collection of start-up values of SRAM cells that is used as a PUF, where each cell contributes one bit.

Figure 2.8 shows an SRAM cell which when powered-up takes a random initial state depending upon the intrinsic mismatch. The challenge in case of SRAM PUF is the power-up of the memory and the state taken by the SRAM cell acts as the response. The major advantage of SRAM PUFs is that SRAMs nowadays are standard components in most of the integrated circuits and hence can be used for their protection without any additional effort. Moreover, SRAMs are a known technology and their reliability can be ensured. This project deals with SRAM PUFs. Next chapter will discuss all aspects of such PUFs.

2.6 Summary

A Physical Unclonable Function or PUF is a function that is embodied in a physical structure and is easy to evaluate but hard to predict. They can be categorized into intrinsic and extrinsic PUFs depending upon the source of randomness that they have. Each PUF is unique due to its randomness and hence can be used in security. Although there is no bullet proof system, systems using PUFs to derive the secret key are a very secure, cost-effective alternative to the classic method of storing a key in a NVM. Although there are several types of PUFs, SRAM PUFs and Butterfly PUFs are the best candidates for most applications because they require no extra processing steps to be manufactured and to be read out. SRAM PUFs are nevertheless the most popular PUFs because they are reliable and most of the todays IC contain embedded memories. Hence, it is a mature technology and cost-effective. The chapters that follow deal with various aspects of SRAM PUFs, their working and modeling in detail.

Start-up Value Based SRAM PUF

3

Start-up value based SRAM PUFs appear to be the most promising approach for secure key-storage applications as discussed in the previous chapter. For the collection of start-up values of SRAM cells to be used as a PUF, we need to understand the basic underlying principles. This chapter gives a more detailed explanation of the SRAM PUF, its architecture and various other aspects of information security. Moreover, an overview of the process variation, which makes the randomness of SRAM PUF feasible, is also provided.

Section 3.1 describes the basic principle on which SRAM PUF is based. Section 3.2 explains the SRAM PUF and its requirements in brief. Section 3.3 discusses an SRAM cell which is the key component for SRAM PUFs. Section 3.4 covers the phenomenon of process variations along with few of the most prevalent process variations.

3.1 Introduction

With fabricated device dimensions approaching the limits of process technology capabilities, a number of process variation, systematic and random, get introduced in the design [6]. Systematic variations are more predictable in nature and depend on factors like layout structure whereas the random variations depend on the microscopic variations in the position and number of dopant atoms in the channel region of the device and variation in the line edge width and length [6]. The random microscopic variations induce increasingly limiting electrical deviations in the device characteristics [3]. Although the process variation is an undesired phenomenon in most of the applications, it was the genesis of a new field, the field of Physical Unclonable Functions (PUFs). The increasing share of memory in SoCs when combined with the new field of Physical Unclonable Functions leads us to use of SRAM PUFs as an ideal choice for IP protection of SoC and FPGAs.

Start-up values of SRAM cells can be used as a PUF because they utilize the random process mismatches and non-identical device features to generate a fingerprint and also satisfy the basic characteristics of a PUF as described in Chapter 2. It has been proved that the power-up memory state of each SRAM cell repeats itself every time it is powered up to the extent that process variation is more significant than the noise present in the SRAM [15]. This chapter gives a detailed description of the SRAM PUF, its architecture and the phenomenon of process variation which plays a vital role in realizing the SRAM PUF.

3.2 Start-up value based SRAM PUF

Start-up value based SRAM PUFs are a new break-through in semiconductor security which intend to use the inherent, unavoidable and random process variation in the SRAMs for their operation. The idea is to use the collection of the power-up state of the SRAM cells as a key. A security system based on PUFs, in general, works on the principle of a challenge-response pair. For SRAM PUFs, the challenge is the power-up of the supply voltage and the response is the initial state assumed by the SRAM cells as shown in Figure 3.1. Every cell in the SRAM array provides 1-bit key and a collection of such SRAM cells can be used as a PUF. The key obtained in this manner is different for every IC and is fuzzy in the sense that it contains noise. In other words, repeated measurements of the same device will be slightly different. The key obtained should have few characteristics like high entropy, Low Fractional Hamming Distance, etc. In Information Security, several metrics exist but the primordial one is entropy [11]. Another interesting concept which is frequently used is the Hamming distance. These concepts are briefly discussed below.

1. **Entropy:**

   Entropy is an important concept while studying information security and is a measure of the uncertainty associated with a variable. In other words, it measures the average information being missed when the value of a certain variable is not known [11]. Entropy is expressed by equation 3.1.

   $$H(X) = -\sum_{N} P_x(x_i) log_2(P_x(x_i)) \tag{3.1}$$

   where

   X is a random variable with N outcomes $(x_i : i = 1, 2, 3..n)$,
   P(x) is the probability of the occurrence of the random variable x.

   Example: Consider the tossing of a fair coin. Here, the probability of obtaining a heads and tails is 50% which implies that the uncertainty for this event is 50%. Entropy can be calculated as follows.

   $$H(X) = -2 * \frac{1}{2} log_2(\frac{1}{2}) = 1bit \tag{3.2}$$

   On the other hand, in case of a biased coin, with probability of heads as 70% and tails as 30%. The entropy of the system is as follows.

   $$H(X) = -(\frac{7}{10} log_2(\frac{7}{10}) + \frac{3}{10} log_2(\frac{3}{10})) = 0.88bit \tag{3.3}$$

   Ideally, if the key of a security system is X bits, then the entropy of the system should be X bits as well. Therefore a system with higher entropy is required.

2. **Hamming Distance:**

   The *Hamming Distance* between two strings of equal length is the number of positions at which the corresponding symbols are different. In other words, it measures the minimum number of substitutions required to change one set of strings into the other, or the number of errors that can transform one string into the other [11].

   Example: Hamming Distance between *butter* and *batter* is 1. For a successful authentication in a security system with the password *'x'*, the hamming distance between the password entered in every login should be ideally 0. However, while using biometrics and PUFs, two measurements from the same source are never the same, although very close. In other words, the hamming distance between these measurements is not 0 but a finite integer. *Error Correction Codes* (ECC) are therefore used to correct the small deviations from the original key and hence match the two keys.

3. **Fractional Hamming Distance:**

   It is defined by the number of bit differences between two bit strings (Hamming Distance) divided by the length of the bit strings. For this thesis, it is an important metric to evaluate the error and uniqueness of the start-up pattern in two different SRAM devices.

   Example: Fractional Hamming distance between *butter* and *batter* is 1/6 as out of the total 6 bits in the strings, 1 bit differs.

Having understood these basic concepts, we now define the basic requirements for the SRAM cells to be used as a security key. These are:

- Ideally, all of the cells belonging to a single SRAM should assume a fixed state in every power-up cycle i.e. Hamming distance between two consecutive power-ups should be ideally 0.

- The power-up state of cells of two different SRAMs should be significantly different from each other [17]; i.e. Hamming distance between the power-up states of two SRAMs should be high enough such that the ECC will not correct the key for one SRAM to the key for another SRAM.

- The power-up states of a single SRAM should have almost equal amount of 0's and 1's spread throughout the memory to make the key prediction very difficult; i.e. the entropy of the system should be equal to the number of bits in the SRAM.

These requirements define the ideal characteristics of SRAM PUF but to understand the properties of SRAM as a PUF at physical level, we first need to understand the working of an SRAM cell which is the key element for SRAM PUFs. The following section describes the architecture, functioning and various other aspects of an SRAM cell.

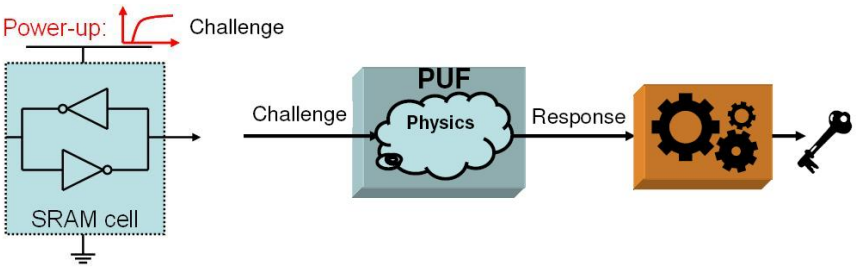

Figure 3.1: SRAM PUF shown as challenge-response pair [20]

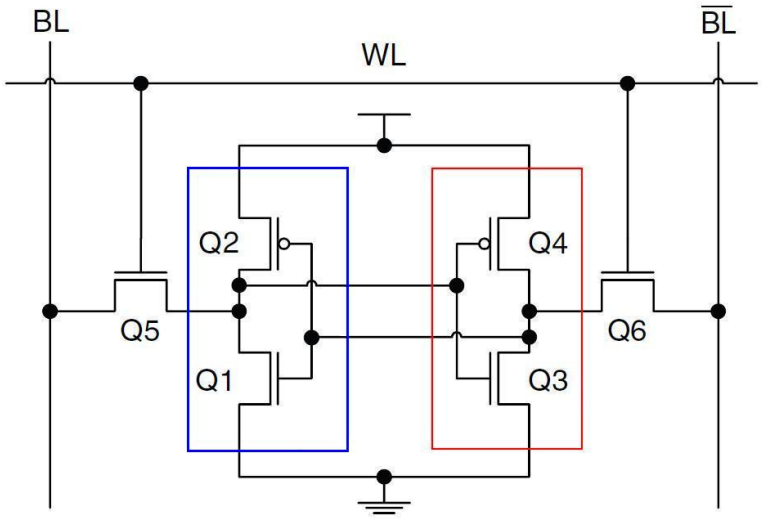

Figure 3.2: A 6-T CMOS SRAM cell

3.3 SRAM Cell

An SRAM cell is the key SRAM component storing the binary information. The mainstream six-transistor (6-T) CMOS SRAM cell is shown in the Figure 3.2. Four transistors (Q1-Q4) comprise the cross-coupled CMOS inverters (formed by Q1, Q2 and Q3, Q4) which means that the output of one inverter is the input of the other and vice-versa. This implies that not only the inverters will always have opposite values at their outputs but also that the output value of one inverter will reinforce the output value of the other inverter. The access transistors (Q5 and Q6), just like the name suggests, are used to access the cell every time it is needed to perform a read or a write operation in the cell. The bitline (BL), the compliment bitline ($\overline{BL}$) and the wordline (WL) are used to select a cell in the cell array among other functions [27].

SRAM yield is an important metric from an economic view point due to the critical and the ubiquitous nature of memory in modern FPGAs and SoCs [27]. Therefore chip area

is an important metric economically and hence SRAM cells use the smallest possible device sizes in any given technology. An important concern while designing SRAM cells is that the contents of the cell should not get altered during the read operation and the cell should be able to write the value quickly during the write operation. These conflicting read and write requirements are satisfied by balancing the relative strengths of the devices in the design [27].

A 6-T CMOS SRAM cell is the most popular SRAM cell due to its superior robustness, low-power and low-voltage operation [27]. Nevertheless, because PUF technology does not use SRAM in a traditional way, one question that still needs to be answered is "What state does an SRAM cell take when it is powered-up and what influences it? The logical understanding of an SRAM cell does not answer this question. Ideally, a manufacturer would like to implement the two inverters of the SRAM cell as identically as possible, since this improves the power and speed characteristics of the memory [22]. However, in reality, due to process variation they present small differences. These small differences introduce a mismatch in the cell which means that the two inverters will not exactly have the same behavior. It is this mismatch which determines the value of the power-up state of an SRAM cell. Depending on the sign of the mismatch, the power-up state of a cell will be biased towards 0 or 1. Hence, every SRAM cell provides a 1-bit digital fingerprint. An SRAM array when considered as a whole can provide a large fingerprint upon its start-up and this system is called the Start-up Value based SRAM PUF.

Figure 3.2 shows an SRAM cell which can be used as the SRAM PUF cell. In SRAM PUF, the output is decided by the start-up state of the SRAM cells. As mentioned earlier, it is required that the start-up state should not vary with time but if the mismatch between the transistors is too small or negligible, the output state may depend upon external factors like noise, temperature, etc. Therefore, we define a new classification of SRAM cells depending upon their start-up behavior which in turn depends on the intrinsic mismatch between the inverters. This terminology will be followed throughout the thesis. The SRAM cells can be classified in three categories:

1. **Non-skewed cell:** A non-skewed cell is basically an SRAM cell wherein the impact of process variations does not cause any mismatch between the two inverters. As the technology scales down, the probability of finding a perfectly matched cell decreases because of the unavoidable process variations. A non-skewed cell will produce a 0 or 1 at the output depending upon the noise present in the system.

2. **Partially-skewed cell:** A partially-skewed cell is the one with a little mismatch between the respective inverters. This kind of SRAM cell will be susceptible to noise, temperature and voltage fluctuations. The cell will have a tendency to assume a value 0 or 1 (depending upon the nature of mismatch) but the cell can flip its value upon variation in external parameters.

3. **Fully-skewed cell:** A fully-skewed cell is a highly mismatched SRAM cell such that the cell always takes a fixed value of 1 or 0 independent of the noise, temperature and voltage fluctuations on power-up. However, this mismatch does not affect the normal storage functionality of SRAM cell. For SRAMs to be used as PUF,

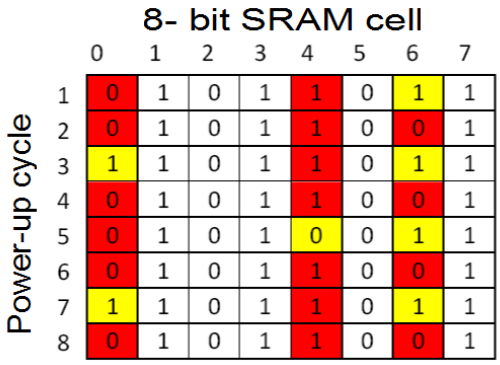

Figure 3.3: Power-up state of 8-bit SRAM cell in eight cycles

the ideal scenario will have a majority of fully-skewed cells in the system. These cells will determine the security key of the system and the partially-skewed cells will provide randomness to the system which will make it difficult for an adversary to attack the system.

Figure 3.3 shows an 8-bit SRAM and the states of its eight cells in eight consecutive power-up cycles. It can be observed that the cells 1,2,3,5 and 7 represent fully-skewed cells which take a fixed state in all the power-up cycles. Cell 6 represents a non-skewed cell which takes a random state in every power-up. The remaining two cells (cell 0 and 4) represent partially-skewed cells which have a tendency to take a particular value but may flip for certain conditions.

The classification described above discriminates the SRAM cells based upon the mismatch in the two inverters. But the question that arises is "How does this mismatch occur in the SRAM cell?" The answer to this is the phenomenon of *Process variation*. The following section describes this phenomenon in more detail followed by few of the most prevalent process variations.

3.4 Process variations

Moore's law has driven the semiconductor industry over last 45 years and has improved the VLSI performance by five orders of magnitude [13]. As the technology scaling continues, one challenge that industry faces today is to manage the process variation. Process variation refers to a set of undesired alterations introduced in the design due to the process of fabrication. Process variations have been existent and critical for fabrication purposes since many years but their role has become increasingly significant in modern-day technologies with feature sizes below 65nm and below.

As mentioned earlier, process variation can either be systematic of random. Random process variations tend to have no spatial or layout correlation which may result in different transistor characteristics even for neighboring devices. In contrast, systematic

variations show similar behaviors in transistor characteristics for the devices that are close to each other in a die or have similar layout characteristics [32]. Also, the impact of random variations is more severe in SRAM cells where minimum sized transistors are often used as mentioned in Section 3.3 [27]. Therefore, random process variations are of more interest for this thesis. Random Dopant Fluctuation, line-edge and line-width roughness, variations in gate dielectric, etc are the most prevalent random process variations [13]. These sources of random process variation are described in brief below.

3.4.1 Random Dopant Fluctuation

Random Dopant Fluctuation is a source of random process variation which results from the discreteness of dopant atoms in the channel of the MOSFET. It is influenced by the position and number of dopant atoms and has a direct impact on the threshold voltage. With technology scaling, a small change in the number of dopant atoms has a significant impact. For example, for $180\mu m$ technology, there are thousands of dopant atoms in the channel whereas this number has reduced to about 100 atoms for 32nm technology. Stolk's equation [33] describes the relation between the threshold voltage variation and the physical parameters of a MOSFET:

$$\sigma_{V_T} = (\frac{\sqrt[4]{4q^3\varepsilon_{si}\phi_B}}{2}).\frac{t_{ox}}{\varepsilon_{ox}}.(\frac{\sqrt[4]{N}}{\sqrt{W_{eff}.L_{eff}}})(\text{Stolk's formula [33]}) \tag{3.4}$$

where $\phi_B = 2\kappa_B T ln(\frac{N}{\eta_i})$

κ_B is Boltzmanns constant measured in $\frac{eV}{K}$,

T the absolute temperature measured in K,

η_i the intrinsic carrier concentration measured in cm^{-3},

q the elementary charge measured in C,

ε_{si} and ε_{ox} are the permitivity of the silicon and the oxide respectively,

t_{ox} is the thickness of the oxide measured in m,

ϵ_{ox} is the oxide permitivity measured in $\frac{F}{m}$,

W_{eff} and L_{eff} are the effective channel width and channel length respectively, both measured in m,

N is the number of dopant atoms.

It is known that other effects (such as dopant segregation, poly depletion) besides the *Random Doping Fluctuation* (RDF) have an influence on the σ_{V_T} [35]. Nevertheless, measurements show that RDF represents 65% of the total variance for a 65nm NMOS or 60% for a 45nm PMOS [13].

3.4.2 Line Edge and Line Width Roughness

Line edge roughness implies a condition wherein the gate of a transistor does not have a constant length, because the edges of the gate are not straight but rough lines [16]. The deviation of the edges from the mean straight line is termed Line Edge Roughness (LER), while the deviation from the mean gate length is termed Line Width Roughness (LWR). Measurements and simulations have shown that LER and LWR start to have impact in technology nodes smaller than 65nm [19]. Further, Asenov et al. [4] have demonstrated that LER induced subthreshold current has a strong dependence on the MOSFET channel length and therefore it is expected that the impact of LER will increase as the technology scales beyond 45nm.

3.4.3 Variations in gate dielectric

The high-k gate dielectric used in technologies like 45nm is highly susceptible to variations in gate dielectric i.e. gate oxide thickness, oxide charges and interface traps. These physical changes in the dielectric result in parametric variations in drive current, gate tunneling current, or threshold voltage. With technology scaling, the thickness of gate oxide (t_{ox}) decreases such that the thickness of silicon dioxide is of order thickness of 4 or 5 atoms. The roughness introduced by process variation, although small between silicon and silicon dioxide, can be of one or two atomic layers, meaning around 50% of the of the t_{ox} (when t_{ox} is around 1nm) [30].

The above mentioned sources of process variations (i.e., RDF, LER/LWR and Variation in gate dielectric) are the most common and dominant ones that play a significant role in the nano-scaled technologies. There are various other effects such as proximity effects, variations associated with strain, etc. that can be observed due to the phenomenon of process variation but they are outside the scope for this thesis [13].

3.5 Conclusion

It is well known that process variation is a challenge to overcome for a few applications; nevertheless they have given rise to a new field of hardware intrinsic security which encashes on the presence of these process variation in nano-scaled technologies. Start-up values based SRAM PUF is one such application which utilizes the mismatch between the two inverters of an SRAM cell to generate a secret key. As mentioned in [17], SRAM cells in general are not immune to variation in external factors like temperature or supply voltage. The amount of mismatch which can make a cell fully-skewed or partially-skewed needs to be quantified. Therefore, it is required to have a metric to evaluate the resistance of SRAM PUFs to the variations of external factors. In other words, it is required to understand the factors that influence the stability of SRAM PUFs. As the PUFs that we intend to use are based on the start-up values of the SRAM cells, analyzing their stability is equivalent to analyzing the stability of the start-up

values formed by the array of SRAM cells. The metric widely used for this purpose is the Static Noise Margin (SNM). SNM analysis of an SRAM cell determines the amount of external factors' variations that a cell can tolerate without changing the actual state of the cell. The next chapter describes the SNM and the parameters affecting the SNM of an SRAM cell.

Modeling The Stability Parameters For SRAM PUF

4

SRAM PUFs utilize the start-up state of an SRAM cell for generation of secret key as discussed in previous chapter. The stability and reliability of this start-up state is an important issue that needs to be ensured before putting it to use. This chapter aims at developing an analytical model to analyze the parameters affecting the stability of start-up value based SRAM PUFs.

This chapter is organized as follows. *Section 4.1 gives a general idea of the factors that can impact the SRAM cell stability. Section 4.2 describes a model for the stability of an SRAM cell in general; this model is further modified to model SRAM PUFs in Section 4.3. Section 4.4 analyzes the model developed, deduces the stability parameters and classifies them to determine the factors that can impact the stability of SRAM PUFs.*

4.1 SRAM cell stability and noise margin

Modern SRAMs strive to increase bit counts while maintaining low chip area and high performance [27]. The cell area consumes about two-thirds of the total memory chip area whereas the cell stability determines the sensitivity of the memory to process tolerances and operating conditions. The two aspects are interdependent since designing a cell for improved stability invariably requires a larger cell area [29].

The metric used to determine SRAM cell stability is called *Static Noise Margin* (SNM). It represents the maximum amount of noise that an SRAM cell can tolerate while retaining in its state [27]. Due to technology scaling and low power consumption requirements in todays applications, supply voltage is scaling down continuously and this poses a great challenge in front of the designers who strive to achieve reliability in storing the bits of an SRAM cell. SRAM stability margin was projected to reduce by 4X as scaling progresses from $250nm$ CMOS technology down to $50nm$ technology in [6],[21]. Since the stability of SRAM cells is reducing with the technology scaling, accurate estimation of SRAM data storage stability in pre-silicon design stage and verification of SRAM stability in the post-silicon testing stage are increasingly important steps in SRAM design and test flows [27].

As mentioned in [17], SRAM PUFs in general are not immune to variation in factors like temperature or supply voltage. Therefore, it is required to have a metric to evaluate the resistance of SRAM PUFs to the variations of such factors. In other words, it is required to understand the factors that influence the stability of SRAM PUFs and then analyze their impact on the start-up values of SRAM cell. This work deals with PUFs that are based on the start-up values of the SRAM cells; therefore analyzing their stability is equivalent to analyzing the stability of the start-up values formed by the array of SRAM

cells. An important task here is to understand the various parameters that can impact the stability of SRAM cells. Hence, we need a model to determine these parameters. This is the purpose of this chapter wherein we will develop a model for SNM of SRAM PUF and then determine its stability parameters.

4.2 Modeling Static Noise Margin for SRAM

As mentioned in Chapter 2, this work focuses on using start-up value based SRAM PUF for security applications. Due to the scaling technology and increasing sensitivity of SRAMs to process variations, modeling of SRAM cell stability is an important exercise. Refreshing the basic architecture of a 6-T SRAM cell from Chapter 3, we know that a basic SRAM cell is composed of two access transistors (Q_3 and Q_4) that are connected to the bit lines along with two cross-coupled inverters (formed by Q_1, Q_5 and Q_2, Q_6) as shown in Figure 4.2. Therefore, in order to understand the SNM for an SRAM cell, we first need to understand the working and input-output voltage characteristics of an inverter. This will be discussed in the section that follows and the concept is then extended to calculate SNM for an SRAM cell. The following sections give a detailed description of the voltage transfer characteristics of a basic inverter and then focus on developing an analytical model for SNM of the SRAM cell.

4.2.1 Input-output characteristics of inverters

Noise Margin is the maximum amount of noise that can be accepted by the device in a system while maintaining the proper operation [28]. Noise Margin can be calculated by using the input to output voltage characteristic curve. A general assumption is that the noise is present for enough time in the circuit for it to react which implies that the noise is DC or static.

Figure 4.1a shows the *Voltage Transfer Characteristics* (VTC) of an inverter wherein the input voltage is varied from 0 to V_{DD} measuring the output voltage characteristics [27]. The critical points for a VTC are the points where the modulus of derivative of V_{out} with respect to V_{in} has a value 1. In other words, an ideal inverter can bear a change of V_{in} without any change in output voltage V_{out} until the input voltage reaches the switching point. At the switching point, $|\frac{dV_{out}}{dV_{in}}| = 1$; this is represented in the figure as point A. For an ideal inverter, at transition point, the change in output is abrupt and the slope is infinite, implying $|\frac{dV_{out}}{dV_{in}}| = \infty$.

In case of a real inverter, there is a transition region instead of a switching point as shown in Figure 4.1b; the slope in this region is finite such that $1 < \frac{dV_{out}}{dV_{in}} < \infty$. As the slope ($\frac{dV_{out}}{dV_{in}}$) in the transition region approaches infinity, the real inverter approaches the ideal one. Figure 4.1b shows the two points where $|\frac{dV_{out}}{dV_{in}}| = 1$ for a real inverter, we will use the coordinates of these points to calculate noise margins for an SRAM cell in the following sub sections.

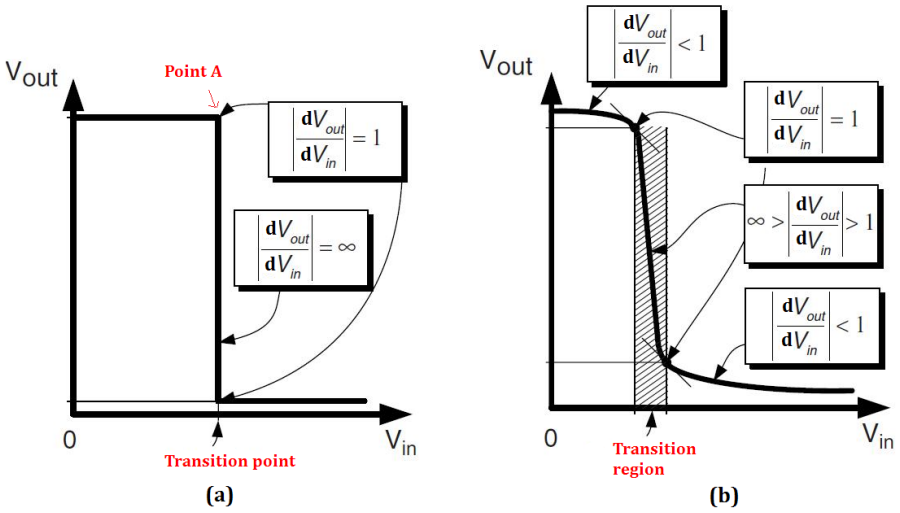

Figure 4.1: Voltage Transfer Characteristics of (a) An ideal inverter (b) A real inverter [27]

4.2.2 Static Noise Margin for SRAM cell

A number of definitions have been proposed for SNM and there is none which has been accepted universally. As mentioned in Section 4.2.1, we assume that the noise present in the circuit is DC, and hence the definition of Noise Margin can be applied to define SNM. SNM is the maximum amount of DC noise that can be tolerated by a system while maintaining its proper operation [27]. The system in our case is a 6-T CMOS SRAM cell as shown in Figure 4.2. Furthermore, it has been shown in [8] that a cell has worst case Noise Margins in read-access mode; therefore, the rest of this section deals with describing the steps involved in calculation of SNM for an SRAM cell in read-access mode as described in [27].

The value of SNM for SRAM cell can be analytically determined by plotting the VTC of cross-coupled inverters as shown in Figure 4.3, where the blue line (dotted) represents the VTC of inverter 1 and the red line (solid) represents the VTC of inverter 2 of the SRAM cell. An important thing here is that since the two inverters are cross-coupled, VTCs of two inverters can be plotted on the same axis by plotting the output of inverter 2 on the input axis of inverter 1.

In the approach that follows, an SRAM cell can be described as two identical inverters with two voltage noise sources (V_n) inserted in between the corresponding inputs and outputs. The noise sources are such that they represent the worst case noise margins. Figure 4.4 shows an SRAM cell in read-access mode i.e. with activated word line and noise sources inserted. The figure shows a read-0 operation as the node connected to the true bitline (BL) has a value 0. Our task here is to determine the SNM (maximum value of V_n) such that the cell retains its value. Assuming that the value of $V_n > V_{TH}$, the modes of operation of various transistors in read-access mode can be written as follows.

Table 4.1: MOSFETs operation regions

Operating Region	NMOS	PMOS
Cutoff	$V_{GS} \leq V_{TH}$	$V_{GS} \geq V_{TH}$
Linear	$V_{GS} \geq V_{TH}$ $V_{DS} \leq V_{GS} - V_{TH}$	$V_{GS} \leq V_{TH}$ $V_{DS} \geq V_{GS} - V_{TH}$
Saturation	$V_{GS} \geq V_{TH}$ $V_{DS} \geq V_{GS} - V_{TH}$	$V_{GS} \leq V_{TH}$ $V_{DS} \leq V_{GS} - V_{TH}$

Table 4.2: Current - Voltage relationships of ideal NMOS and PMOS

Operating Region	$I_D for NMOS$	$I_D for PMOS$
Cutoff	0	0
Linear	$\frac{1}{2}\beta[2(V_{GS} - V_{TH})V_{DS} - (V_{DS})^2]$	$\frac{1}{2}\beta[2(V_{GS} - V_{TH})V_{DS} - (V_{DS})^2]$
Saturation	$\frac{1}{2}\beta(V_{GS} - V_{TH})^2$	$\frac{1}{2}\beta(V_{GS} - V_{TH})^2$
Saturation (short-channel)	$\frac{1}{2}\beta(V_{GS} - V_{TH})^2(1 + \lambda V_{DS})$	$\frac{1}{2}\beta(V_{GS} - V_{TH})^2(1 + \lambda V_{DS})$

- Q_3 and Q_6 are in cutoff mode
- Q_1 and Q_4 are in saturation mode
- Q_2 and Q_5 are in linear mode

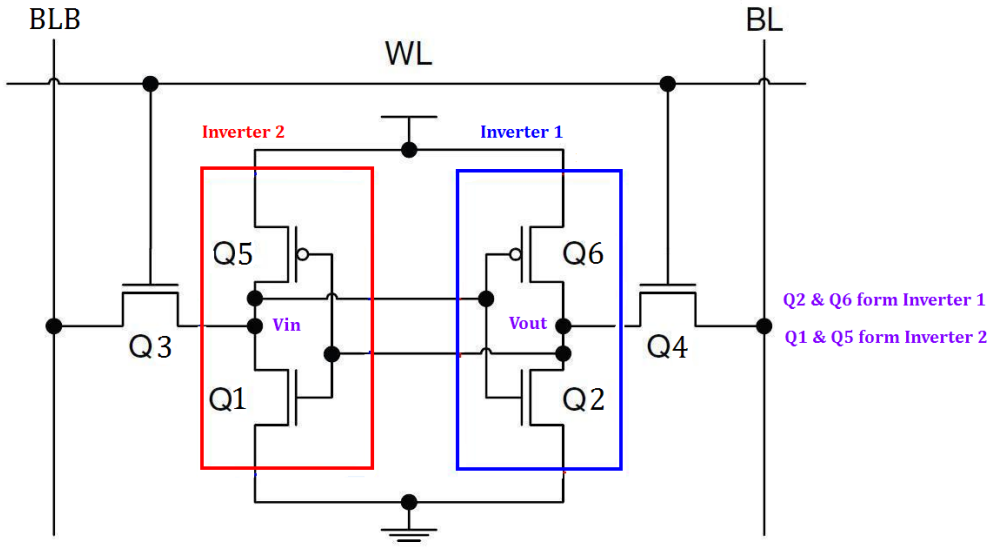

Figure 4.2: 6-T CMOS SRAM cell

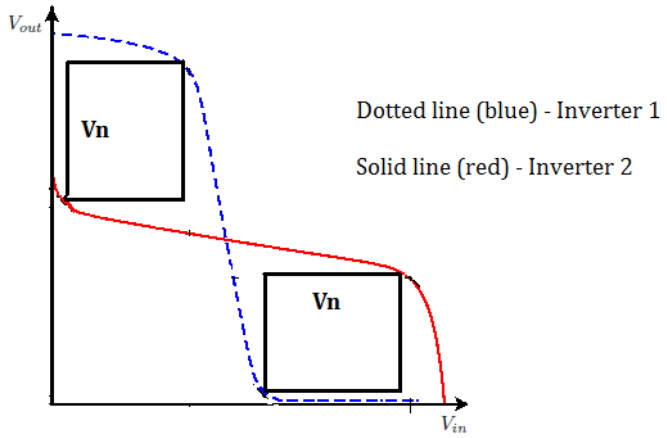

Figure 4.3: Voltage Transfer Characteristics of an SRAM cell

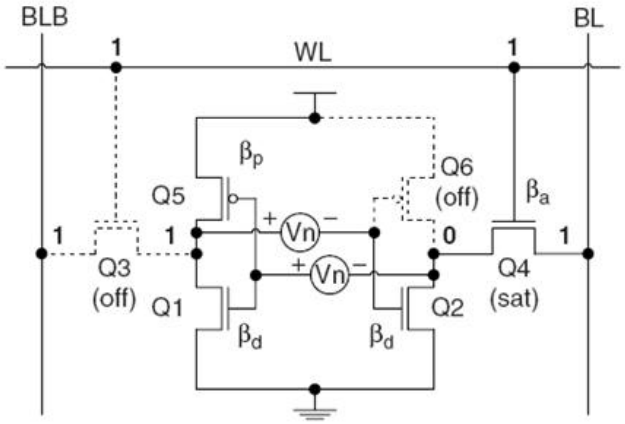

Figure 4.4: SRAM cell in read-access mode with DC noise sources (V_n) [27]

Table 4.1 and 4.2 show the voltage and drain current equations and operating conditions for transistors in various modes of operation as described in [24]. The nomenclature that has been followed to write these equations and throughout this thesis is as follows:

- I_D denotes the drain current
- G denotes the Gate node, D denotes the Drain node, S denotes the Source node
- V_{TH_n} denotes the threshold voltage of transistor number 'n'
- V_{DD} denotes the supply voltage
- β_n denotes the transconductance of transistor number 'n'

Using these equations for various transistors in their respective modes of operation, we will now calculate the SNM for the read-access SRAM cell.

From Figure 4.4, we can say that

$$I_{D_{Q_1}} = I_{D_{Q_5}} \tag{4.1}$$

$$I_{D_{Q_4}} = I_{D_{Q_2}} \tag{4.2}$$

The current equations for the circuit (ignoring the short-channel effects) assuming Q_1 and Q_4 in saturation and Q_2 and Q_5 in linear mode can be written as follows:

$$\frac{1}{2}\beta_1(V_{GS_1} - V_{TH_1})^2 = \frac{1}{2}\beta_5[2(V_{GS_5} - V_{TH_5})V_{DS_5} - (V_{DS_5})^2] \tag{4.3}$$

$$\frac{1}{2}\beta_4(V_{GS_4} - V_{TH_4})^2 = \frac{1}{2}\beta_2[2(V_{GS_2} - V_{TH_2})V_{DS_2} - (V_{DS_2})^2] \tag{4.4}$$

Applying Kirchhoffs law for the SRAM cell,

$$V_{GS_1} = V_n + V_{DS_2} \tag{4.5}$$

$$V_{DS_5} = V_n + V_{GS_2} - V_{DD} \tag{4.6}$$

$$V_{GS_5} = V_n + V_{DS_2} - V_{DD} \tag{4.7}$$

$$V_{GS_4} = V_{DD} - V_{DS_2} \tag{4.8}$$

To simplify the equations further, we can replace $\frac{\beta_5}{\beta_1} = q$ and $\frac{\beta_2}{\beta_4} = r$ and for simplicity sake, we assume $|V_{TH_{1,2,4,5}}| = |V_{TH}|$. Substituting the equations 4.5 to 4.8 in equation 4.3 and 4.4 and then eliminating V_{GS_2} and V_{DS_2} from the equations, we obtain the following equation:

$$X^2\left(1 + 2k + \frac{r}{q}k^2\right) + 2X\left(\frac{r}{q}kA + A + V_{TH} - V_s\right) + \frac{r}{q}A^2 = 0 \tag{4.9}$$

Where,

$$V_s = V_{DD} - V_{TH} \tag{4.10}$$

$$V_r = V_s - \left(\frac{r}{r+1}\right)V_{TH} \tag{4.11}$$

$$k = \frac{r}{(r+1)}\sqrt{\frac{(r+1)}{(r+1-\frac{V_s^2}{V_r^2}-1}} \tag{4.12}$$

$$V_0 = kV_s + \left(\frac{1+r}{1+r+\frac{r}{k}}\right)V_r \tag{4.13}$$

$$X = V_{DD} - V_n - V_{GS_2} \tag{4.14}$$

$$V_{DS_2} = V_0 - kV_{GS_2} \tag{4.15}$$

$$A = V_0 + (k+1)V_n - kV_{DD} - V_{TH} \tag{4.16}$$

Equation 4.9 can be understood in terms of a basic quadratic equation $aX^2+bX+c=0$, the roots of which are given by:

$$X_1 = \frac{-b+\sqrt{b^2-4ac}}{2a} \tag{4.17}$$

$$X_2 = \frac{-b-\sqrt{b^2-4ac}}{2a} \tag{4.18}$$

The two roots of the basic equation (X_1 and X_2) are equal if $b^2-4ac=0$. Now, since the two inverters are identical with equal noise sources, the value of SNM can be calculated by equating the two roots of the equation 4.9. Applying this condition and then calculating the worst case noise margin represented by V_n gives us the value of SNM (V_n)(see Figure 4.3) which can be written as follows:

$$SNM = V_{TH} - \left(\frac{1}{k+1}\right)\left(\frac{V_{DD} - (\frac{2r+1}{r+1})V_{TH}}{1+\frac{r}{k(r+1)}} - \frac{V_{DD}-2V_{TH}}{1+\frac{kr}{q}+\sqrt{\frac{r}{q}(1+2k+\frac{r}{q}k^2)}}\right) \tag{4.19}$$

From the above equation, we can derive the factors that may impact the SNM of an SRAM cell. These factors are also depicted in Figure 4.5. They are:

- Supply Voltage V_{DD}
- Threshold Voltage V_{TH}
- Transconductance of the transistors β

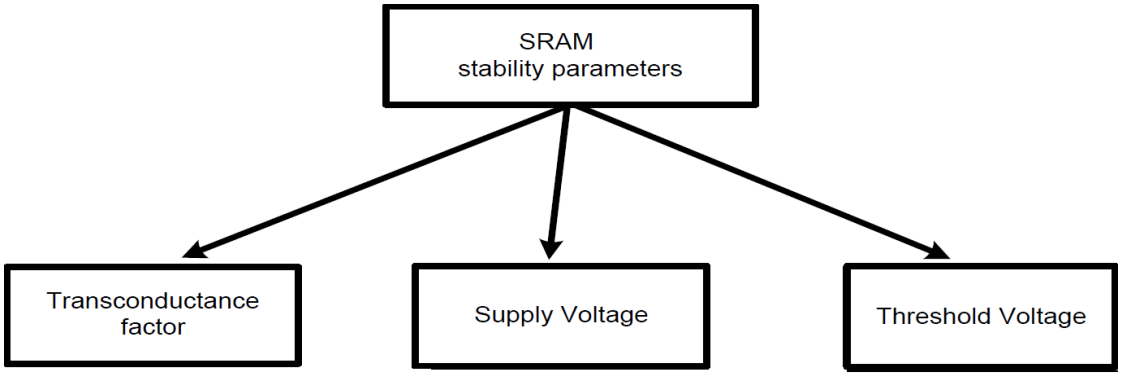

Figure 4.5: Factors affecting SNM of an SRAM cell

This is the classical model for calculating SNM of an SRAM cell which takes both the transistors forming the inverters and the access transistors into account as all of them affect the stability of the SRAM cell in read-access mode. It is worth noting that this model cannot be used for SRAM PUFs for many reasons. Because our goal is to determine the factors which can impact the generation of start-up value of SRAM cell and not reading of the start-up value, the classical model developed above cannot be used as it considers the cell in read-access mode and not in the start-up mode. In addition, the transistors that play a major role in determining the start-up value of SRAM cell are the ones forming the cross-coupled inverters and not the access transistors. Moreover the classical model of SNM assumes the circuit to be symmetric and all the parameters for both inverters are identical; this is not the case while using SRAM as PUFs because of the existence of process variations. Many simplifications have been done assuming this symmetry and hence this model does not suit our purpose. Therefore, a new analytical model has to be developed to analyze the SNM for SRAM PUFs. This is the topic for the next section.

4.3 Static Noise Margin for SRAM PUFs

From now on, we will refer to the model for SNM of SRAM PUF as PSNM. As described in Section 4.2.2, SNM for an SRAM cell is defined as the maximum amount of noise that can be tolerated by any cell such that it retains its value without flipping. It can be graphically calculated by plotting the VTC for both the inverters and then using length of the largest square that can fit in either of the 'eyes' of this VTC curve as proposed in [29]. The size of this square depends on coordinates of four critical points denoted by Point A, B, C and D; these are the points with $|\frac{dV_{out}}{dV_{in}}|= 1$ as shown in Figure 4.6. The coordinates of the points A, B, C, D are given by (V_{IL}, V_{OH}), $(V_{O'L}, V_{I'H})$, (V_{IH}, V_{OL}), $(V_{O'H}, V_{I'L})$ respectively where,

- V_{IL} and V'_{IL} represent the maximum input voltage recognized as logical 0
- V_{IH} and V'_{IH} represent the minimum input level recognized as logical 1

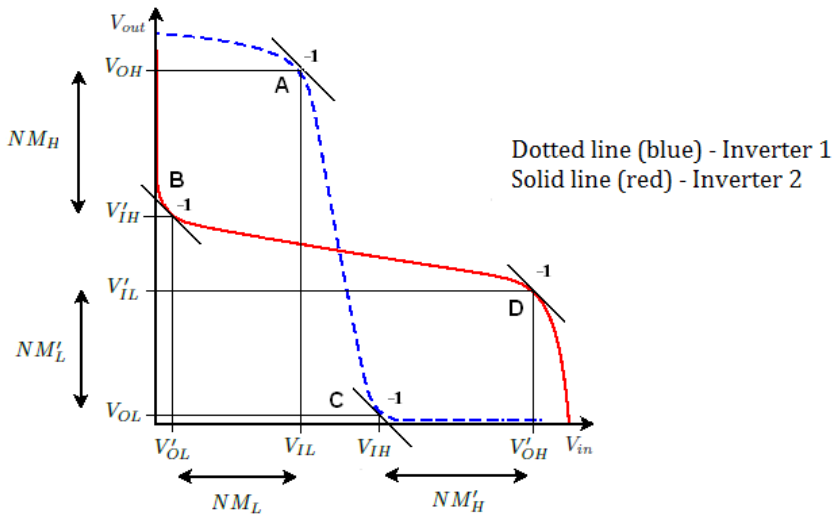

Figure 4.6: VTC of an SRAM cell describing the noise margins

- V_{OL} and V'_{OL} represent the maximum logical output voltage recognized as 0
- V_{OH} and V'_{OH} represent the minimum logical output voltage recognized as 1

Figure 4.6 shows the VTC of the cross-coupled inverters such that V_{OL} and V_{OH} are the logic stable output points of inverter 1 (dotted line) satisfying $\frac{dV_{out}}{dV_{in}} = -1$ with input coordinate points as V_{IH} (see point C in the Figure 4.6) and V_{IL} (see point A in the Figure 4.6) respectively. The corresponding points for inverter 2 (solid line) are also shown. Having the coordinates of all the critical points, we will then calculate the noise margins for the inverters. In [27], Noise Margin High (NM_H) and Noise Margin Low (NM_L) have been defined as follows (see Figure 4.6).

- $NM_H = V_{OH} - V'_{IH}$
- $NM_L = V_{IL} - V'_{OL}$
- $NM'_H = V'_{OH} - V_{IH}$
- $NM'_L = V'_{IL} - V_{OL}$

Therefore, in order to determine the PSNM, we first need to determine the coordinates of four critical points described above; this is done in next subsections.

4.3.1 Calculation of critical points

At each of these four points, the transistors involved are in a certain mode of operation, which implies being in cutoff, saturation or linear mode. Table 4.1 and 4.2 show the

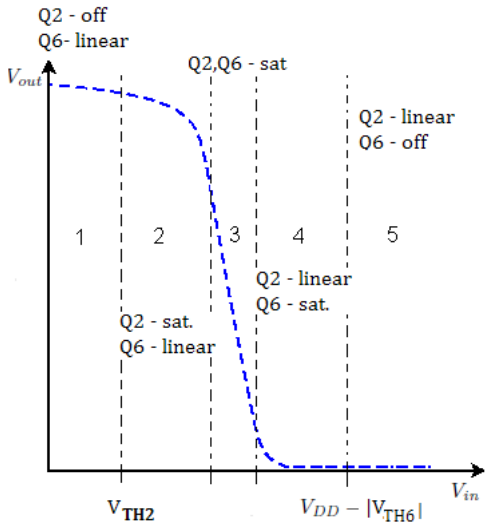

Figure 4.7: Operating regions for inverter 1

operation conditions and current equations in each of these regions. As mentioned earlier in Section 4.2.2, the coordinates for critical points of inverter 1 (i.e. V_{IL}, V_{IH}, V_{OL} and V_{OH}) correspond to those of inverter 2 (i.e. V'_{IL}, V'_{IH}, V'_{OL} and V'_{OH}). Therefore, we first focus on calculating the critical points for inverter 1 and then derive the corresponding points for inverter 2.

As described earlier in this section, we only consider the transistors forming the inverters to calculate the PSNM. Figure 4.7 exhibits the modes of operation of these four transistors by dividing the VTC of an inverter into five regions, which are:

- Region 1 - Q_2 is in cutoff mode and Q_6 in linear mode
- Region 2 - Q_2 is in saturation mode and Q_6 in linear mode
- Region 3 - Both Q_2 and Q_6 are in saturation mode
- Region 4 - Q_2 is in linear mode and Q_6 in saturation mode
- Region 5 - Q_2 in linear mode and Q_6 in cutoff mode

Having known the operating conditions, we now calculate the coordinates of each of the critical points. The steps followed for calculation of critical points are:

1. Write the current equations for transistors in their respective modes of operation.
2. Differentiate the equations obtained in step 1 with respect to the input voltage and then replace the derivative with $\frac{dV_{out}}{dV_{in}} = -1$.
3. Utilize the equations obtained in step 1 and 2 to obtain an expression for the coordinates of critical points.

Coordinates of point A

The coordinates of point A are (V_{IL}, V_{OH}). As can be seen from Figure 4.7, for inverter 1, the coordinates of point A can be calculated by taking into account the operating mode of transistors at this point. It is evident from the figure that point A lies in region 2 and hence transistor Q_2 is in saturation mode and Q_6 is in linear mode. One more thing worth noting is that we are dealing with nano-scaled SRAM PUF and therefore we need to consider the short channel effects. The effect of channel length modulation parameter λ has been introduced in the current equations described in Table 4.2.

Referring back to Figure4.6 and Figure4.7, we can equate the drain currents for the transistors forming the inverter.

$$I_{D_{Q_2}} = I_{D_{Q_6}} \tag{4.20}$$

Substituting the values of these currents in terms of input and output voltages and other known parameters, equation 4.20 becomes:

$$\frac{1}{2}\beta_2(V_{GS_2} - V_{TH_2})^2(1 + \lambda_2 V_{DS_2}) = \frac{1}{2}\beta_6[2(V_{GS_6} - V_{TH_6})V_{DS_6} - V_{DS_6}^2] \tag{4.21}$$

Where,

$$V_{G_2} = V_{G_6} = V_{Gate} = V_{in} \tag{4.22}$$

$$V_{S_2} = GND = 0 \tag{4.23}$$

$$V_{D_2} = V_{D_6} = V_{Drain} = V_{out} \tag{4.24}$$

$$V_{S_6} = V_{DD} \tag{4.25}$$

Substituting equation 4.22 to 4.25 in 4.21 results in the following equation:

$$\beta_2(V_{in}-V_{TH_2})^2(1+\lambda_2 V_{out}) = \beta_6[2(V_{in}-V_{DD}-V_{TH_6})(V_{out}-V_{DD})-(V_{out}-V_{DD})^2] \tag{4.26}$$

Since we know that at point A, $\frac{dV_{out}}{dV_{in}} = -1$, we differentiate equation 4.26 with respect to V_{in}. The differential equation can be written as follows:

$$\begin{aligned}\beta_2\left[2(V_{in} - V_{TH_2})(1 + \lambda_2 V_{out}) + \lambda_2(V_{in} - V_{TH_2})^2\frac{dV_{out}}{dV_{in}}\right] &= \beta_6[2(V_{out} - V_{DD}) \\ &\quad + 2(V_{in} - V_{DD} - V_{TH_6})\frac{dV_{out}}{dV_{in}} \\ &\quad - 2(V_{out} - V_{DD})\frac{dV_{out}}{dV_{in}}]\end{aligned} \tag{4.27}$$

As mentioned earlier, $\frac{dV_{out}}{dV_{in}} = -1$, $V_{in} = V_{IL}$ and $V_{out} = V_{OH}$ at point A. Hence substituting these values in equation 4.27 gives:

$$\begin{aligned}\beta_2 \left(2(V_{IL} - V_{TH_2})(1 + \lambda_2 V_{OH}) - \lambda_2 (V_{IL} - V_{TH_2})^2\right) &= 2\beta_6((V_{OH} - V_{DD}) - (V_{IL} - V_{DD} - V_{TH_6}) \\ &\quad + (V_{OH} - V_{DD})) \\ &= 2\beta_6 \left(2V_{OH} - V_{IL} + V_{TH_6} - V_{DD}\right)\end{aligned} \tag{4.28}$$

Equation 4.28 gives us the y coordinate (V_{OH}) as follows (see Figure 4.6):

$$\begin{aligned}V_{OH} &= \frac{\frac{1}{2}\frac{\beta_2}{\beta_6}(V_{IL} - V_{TH_2}) - \frac{1}{4}\frac{\beta_2}{\beta_6}\lambda_2 (V_{IL} - V_{TH_2})^2 + \frac{V_{IL} - V_{TH_6} + V_{DD}}{2}}{1 - \frac{1}{2}\frac{\beta_2}{\beta_6}\lambda_2 V_{IL} + \frac{1}{2}\frac{\beta_2}{\beta_6}\lambda_2 V_{TH_2}} \\ &= \frac{\frac{\beta_2}{\beta_6}(V_{IL} - V_{TH_2}) - \frac{1}{2}\frac{\beta_2}{\beta_6}\lambda_2 (V_{IL} - V_{TH_2})^2 + V_{IL} - V_{TH_6} + V_{DD}}{2 - \frac{\beta_2}{\beta_6}\lambda_2 (V_{IL} - V_{TH_2})}\end{aligned} \tag{4.29}$$

In order to calculate V_{IL}, we use $V_{in} = V_{IL}$ and $V_{out} = V_{OH}$ (as determined in equation 4.29) in equation 4.26 and since the equations are complex, we can make use of tools like Maple to obtain a solution for V_{IL}. Since, point D represents the point similar to point A but for Inverter 2, one can determine the coordinates of point D (see Figure 4.7) in a similar way. This can also be calculated straightforward by replacing V_{IL} by V'_{IL}, V_{OH} by V'_{OH}, the variables with index 6 with index 5 and the variables with index 2 with index 1 in the equations obtained for point A.

Coordinates of Point C

The coordinates of point C are (V_{IH}, V_{OL}). This point for inverter 1 corresponds to Q_2 being in linear mode and Q_6 in saturation mode, which lies in region 4 as shown in Figure 4.7. The basic idea of the equations remain the same as in the case of point A. Hence,

$$I_{D_{Q_2}} = I_{D_{Q_6}} \tag{4.30}$$

Substituting the values of these currents in terms of the input, the output voltages and other parameters results in following equation:

$$\frac{1}{2}\beta_2 [2(V_{GS_2} - V_{TH_2})V_{DS_2} - V_{DS_2}^2] = \frac{1}{2}\beta_6 (V_{GS_6} - V_{TH_6})^2 (1 + \lambda_6 V_{DS_6}) \tag{4.31}$$

where all the parameters have the same meaning as they had for point A.

Substituting the corresponding values for voltages in equation 4.31 gives the following:

$$\beta_2[2(V_{in} - V_{TH_2})V_{out} - V_{out}^2] = \beta_6(V_{in} - V_{DD} - V_{TH_6})^2(1 + \lambda_6(V_{out} - V_{DD})). \quad (4.32)$$

Similar to the procedure followed for calculation of coordinates in point A, we now differentiate the equation 4.32 with respect to V_{in} which gives us the following relation:

$$\begin{aligned}\beta_2\left[2V_{out} + 2(V_{in} - V_{TH_2})\frac{dV_{out}}{dV_{in}} - 2V_{out}\frac{dV_{out}}{dV_{in}}\right] &= \beta_6[\lambda_6(V_{in} - V_{DD} - V_{TH_6})^2\frac{dV_{out}}{dV_{in}} \\ &\quad + 2(1 + \lambda_6(V_{out} - V_{DD}))(V_{in} - V_{DD} - V_{TH_6})] \end{aligned} \quad (4.33)$$

At point C, $\frac{dV_{out}}{dV_{in}} = -1$, $V_{in} = V_{IH}$ and $V_{out} = V_{OL}$. Substituting these values in equation 4.33 gives

$$\begin{aligned}2\beta_2\left[V_{OL} - (V_{IH} - V_{TH_2}) + V_{OL}\right] &= \beta_6[2(1 + \lambda_6(V_{OL} - V_{DD}))(V_{IH} - V_{DD} - V_{TH_6}) \\ &\quad - \lambda_6(V_{IH} - V_{DD} - V_{TH_6})^2]. \end{aligned} \quad (4.34)$$

Equation 4.34 gives us the y coordinate of point C as required (see Figure 4.6). The coordinate V_{OL} can be expressed as follows:

$$\begin{aligned}V_{OL} &= \frac{2\frac{\beta_6}{\beta_2}\left[(1 - \lambda_6 V_{DD})(V_{IH} - V_{DD} - V_{TH_6})\right] - \frac{\beta_6}{\beta_2}\lambda_6(V_{IH} - V_{DD} - V_{TH_6})^2 + 2(V_{IH} - V_{TH_2})}{4 - 2\frac{\beta_6}{\beta_2}\lambda_6(V_{IH} - V_{DD} - V_{TH_6})} \\ &= \frac{\frac{\beta_6}{\beta_2}\left[(1 - \lambda_6 V_{DD})(V_{IH} - V_{DD} - V_{TH_6})\right] - \frac{1}{2}\frac{\beta_6}{\beta_2}\lambda_6(V_{IH} - V_{DD} - V_{TH_6})^2 + (V_{IH} - V_{TH_2})}{2 - \frac{\beta_6}{\beta_2}\lambda_6(V_{IH} - V_{DD} - V_{TH_6})} \end{aligned} \quad (4.35)$$

To calculate the x coordinate of point C (V_{IH}), we can substitute $V_{in} = V_{IH}$ and $V_{out} = V_{OL}$ in equation 4.32 and since the equations are complex, we can use a mathematical tool to obtain a solution as mentioned in calculation for point A. Furthermore, point C for inverter 1 corresponds to point B for inverter 2; therefore one can determine the coordinates of point B (see Figure 4.7 defined for inverter 2 where $\frac{dV_{out}}{dV_{in}} = -1$) in a similar way. This can also be calculated straightforward by replacing V_{IH} by V'_{IH}, V_{OL} by V'_{OL}, the variables with index 6 with index 5 and the variables with index 2 with index 1 in the same equations.

After calculating the coordinates of all the critical points, we will now calculate the PSNM in the next section.

4.3.2 Calculation of PUF Static Noise Margin

This is the analytical model for calculating PSNM which takes only the transistors forming the inverters into account and ignores the access transistors. The model considers that both inverters are affected by process variations and hence none of their parameters are necessarily identical. From the critical points obtained in the previous sections, we can now calculate the PSNM for an SRAM cell. This can be done in a similar way as it was done for the SNM SRAM classical model.

- $PNM_H = V_{OH} - V'_{IH}$
- $PNM_L = V_{IL} - V'_{OL}$
- $PNM'_H = V'_{OH} - V_{IH}$
- $PNM'_L = V'_{IL} - V_{OL}$

Here, each of the input and output voltages can be calculated by using equations from Section 4.3.1. The mathematical expressions for each of the voltages on the right hand side of the equations cannot be obtained manually because of the high complexity and hence we use mathematical tools such as MATLAB and Maple to calculate the values of noise margins. The smallest of these noise margins gives us the PSNM of the cell as this will be the maximum value of noise that can be tolerated by the cell without flipping its value.

One important information that we can derive from the PSNM expression obtained is the factors that influence the PSNM of the SRAM cell. From the equations involved in the calculation, we conclude that the factors influencing SNM are the Supply Voltage V_{DD}, Threshold Voltage V_{TH}, Transconductance factor β, and the Channel length modulation factor λ. Figure 4.8 shows all the factors influencing the PSNM and hence the start-up values of an SRAM cell which would be used as a Physical Unclonable Function. Since the SRAM cell stability is determined by these factors, we call them the stability parameters for SRAM PUF. The following section gives a more detailed description of these factors and their dependencies on other parameters.

4.4 Stability parameters for SRAM PUF

The equations mentioned in the previous section leads us to important information regarding the parameters which can impact the PSNM for SRAM PUF. From the equations involved, we conclude that the factors influencing SNM are:

- Supply Voltage (V_{DD}): It means that V_{DD} scaling for lower technologies will impact PSNM.
- Threshold Voltage (V_{TH}): It means that V_{TH} scaling for lower technologies and transistor mismatch will impact PSNM.

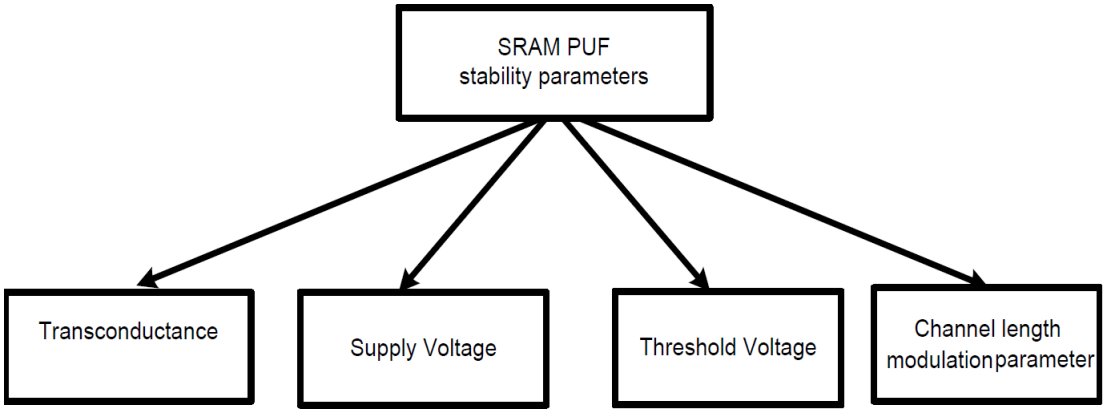

Figure 4.8: Stability parameters for SRAM PUF

Table 4.3: SRAM PUF stability parameters and the factors they are dependent on

Stability parameter	Factors affecting the stability parameter
Supply Voltage V_{DD}	Ramp-up speed Supply voltage value
Threshold Voltage V_{TH}	Gate oxide thickness t_{ox} Doping concentration of the substrate N_a, N_d Temperature T
Transconductance β	Gate oxide thickness t_{ox} Oxide charge density N_{oc} Width of the MOSFETs $W_{n,p}$ Length of the MOSFETs $L_{n,p}$
Channel length modulation parameter λ	Early Voltage V'_A Length of the channel L

- Transconductance (β): It means that geometry of transistor will impact PSNM.
- Channel length modulation parameter (λ): This parameter has more impact for smaller technologies as the short-channel effects become dominant [12].

The parameters mentioned above are the ones to which the PSNM is sensitive to and will be referred to as stability parameters for the rest of our discussion. Inspecting the above mentioned stability parameters reveals that supply voltage can be considered as an external parameter that can be controlled by the user whereas the rest of the factors are dependent on the process/technology node used. This will become more clear in the next section where we will discuss each of the stability parameter in detail. In the rest of this section, each of the stability parameter will be analyzed in order to explore all the factors that can affect the PSNM. The analysis is done using the standard equations for parameters as described in [24]. A summary of the results obtained is given in Table 4.3

4.4.1 Supply Voltage V_{DD}

This is one of the external factors influencing the PSNM. There is no such physical factor that has an impact on the supply voltage assuming that the voltage that is being supplied to SRAM is as constant as coming from a 'DC' source. But since we are using the power-up state of SRAM which is a transient event during which the supply voltage is raised from 0 to V_{DD}, the ramp-up speed of the power-up is expected to have an impact on PSNM. The nature of this impact will be studied in the next chapter.

4.4.2 Threshold voltage V_{TH}

The threshold voltage (V_{TH}) is usually defined as the gate voltage where an inversion layer forms at the interface between the oxide layer and the body of the transistor [12]. Threshold voltage for a MOSFET with p-type substrate (NMOS) and n-type substrate (PMOS) can be expressed in terms of various parameters of MOS capacitor as follows [24]:

$$V_{TH_{NMOS}} = V_{FB} + 2|\phi_F| + \gamma\sqrt{2|\phi_F|} \tag{4.36}$$

$$V_{TH_{PMOS}} = V_{FB} - 2\phi_F - \gamma\sqrt{2\phi_F} \tag{4.37}$$

The terms involved in the equations are:

1. Flat band voltage V_{FB}

   This is affected by the presence of charge in the oxide or at the oxide-semiconductor interface [39]. The flat band voltage corresponds to the voltage which when applied to the gate electrode yields a flat energy band in the semiconductor. The V_{FB} is described as follows:

   $$V_{FB} = \phi_{ms} - \left(\frac{q.N_{oc}}{C_{ox}}\right) \tag{4.38}$$

   where,

   ϕ_{ms} is the difference in the work function of the gate and the substrate, measured in V,

   q is the charge of an electron, measured in C,

   N_{oc} is the oxide charge density, measured in cm^{-3},

   C_{ox} is the capacitance of the gate oxide per unit area, measured in $\frac{F}{m^2}$. Here,

   $$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \tag{4.39}$$

   where t_{ox} represents the oxide thickness in meters and ϵ_{ox} is the oxide permitivity in $\frac{F}{m}$.

2. Surface potential ϕ_F

   This factor depends upon the substrate doping, the intrinsic carrier concentration and the temperature as shown by the equations below [24]:

$$\phi_{F_{NMOS}} = -\frac{kT}{q} ln\left(\frac{N_a}{n_i}\right) \tag{4.40}$$

$$\phi_{F_{PMOS}} = \frac{kT}{q} ln\left(\frac{N_d}{n_i}\right) \tag{4.41}$$

   where,

   k is the Boltzman constant measured in $\frac{eV}{K}$,

   T is the temperature measured in K,

   q is the charge of an electron measured in C,

   N_a is the concentration of acceptors in substrate for NMOS measured in cm^{-3},

   N_d is the concentration of acceptors in substrate for PMOS measured in cm^{-3},

   n_i is the intrinsic carrier concentration measured in cm^{-3}.

3. Body effect parameter γ

   This is a dimensionless quantity and is given by the following equation:

$$\gamma_{NMOS} = \frac{\sqrt{2\epsilon_s q N_a}}{C_{ox}} \tag{4.42}$$

$$\gamma_{PMOS} = \frac{\sqrt{2\epsilon_s q N_d}}{C_{ox}} \tag{4.43}$$

   where,

   ϵ_s is the silicon permitivity measured in $\frac{F}{m}$,

   all other parameters have the same meaning as mentioned above.

   It can be concluded by inspecting each of the above factors that parameters affecting threshold voltage and hence the PSNM are :

   - Oxide thickness (t_{ox})
   - Temperature (T)
   - Substrate doping concentration (N_a, N_d)
   - Oxide charge density N_{oc}

4.4.3 Transconductance β

The transconductance coefficient β is given by the following equation as described in [17]:

$$\beta_{n,p} = \mu_{n,p} C_{ox} \frac{W_{n,p}}{L_{n,p}} \tag{4.44}$$

where,

$\mu_{n,p}$ is the mobility of electrons or holes depending upon the type of MOSFET measured in $\frac{m^2}{V-s}$,

C_{ox} is the capacitance per unit area of the gate oxide measured in $\frac{F}{m^2}$,

$W_{n,p}$ is the width of the NMOS or PMOS respectively measured in m,

$L_{n,p}$ is the length of the NMOS or PMOS respectively measured in m.

As it can be seen, all factors affecting the transconductance are related to the geometry of the transistors involved; hence, it will be referred to as transistor geometry parameter.

It can be concluded that the factors affecting transconductance and hence the PSNM are:

- Oxide thickness (t_{ox})
- Width of the MOSFETs ($W_{n,p}$)
- Length of the MOSFETs ($L_{n,p}$)

4.4.4 Channel Length Modulation parameter λ

The Channel Length Modulation (CLM) parameter is one of short channel effects in MOSFET scaling. CLM is a shortening of the length of the inverted channel region with increase in drain bias for large drain biases [12]. The equation for λ can be written as [12]:

$$\lambda = \frac{1}{V'_A L} \tag{4.45}$$

where,

V'_A is the early voltage that depends on the process technology used; it is measured in V,

L is the actual length of the channel measured in m.

It is obvious that there is only one factor that affects channel length modulation parameter, that is:

- Length of the channel (L)

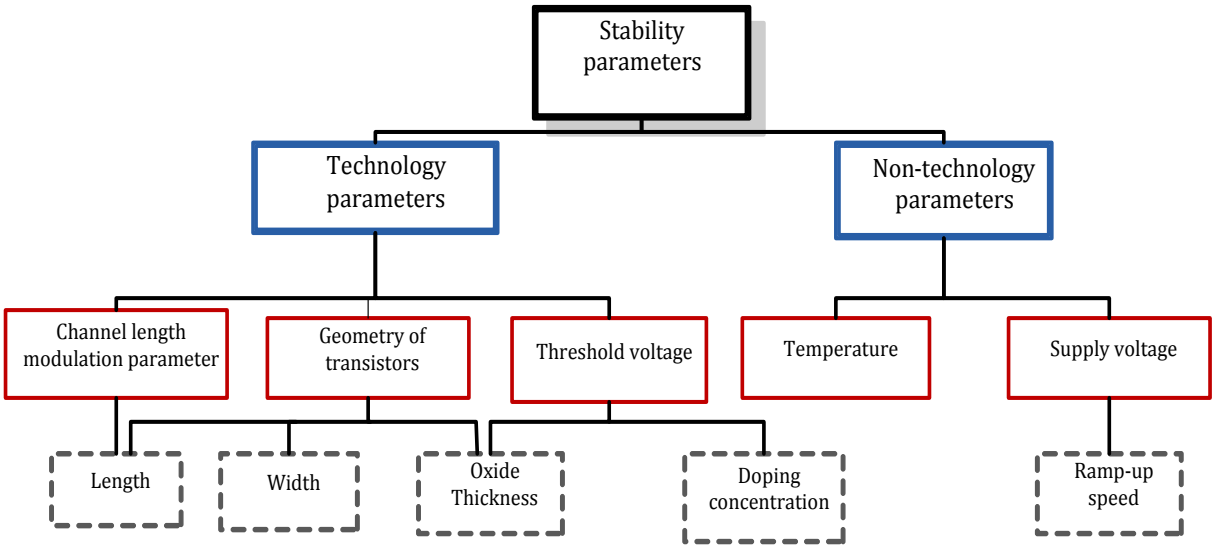

Figure 4.9: Classification of stability parameters for SRAM PUFs

4.4.5 Stability parameters classification

Table 4.3 summarizes all the stability parameters and the associated factors which can impact each of these parameters. Inspecting the list of stability parameters, it can be concluded that the supply voltage can be controlled externally by the user whereas the other stability parameters are dependent on the technology node used. This leads us to the following classification of stability parameters:

- Technology parameters: They consist of Threshold voltage, Channel length modulation parameter and Transistor geometry
- Non-technology parameters: They consist of Supply voltage and Temperature

One more important thing to notice is that temperature (T) is an orthogonal parameter which can impact a number of parameters in different ways and since it is not technology dependent; it can be considered to be a non-technology parameter. Furthermore, environmental factors like humidity and radiation can impact the stability of SRAM cell as well[14]. These parameters are not involved in the equations considered but their role in impacting the stability of SRAM cell cannot be ignored. Figure 4.9 gives the classification of the different stability parameters of SRAM PUF.

4.5 Summary

Process variations have a significant impact on the SRAM cell stability in nano-scaled CMOS technologies. Static Noise Margin is considered to be the metric to determine the SRAM cell stability. This chapter proposes an analytical model for the calculation

of SNM for SRAM cell in general and SRAM PUF in particular. The stability parameters impacting the SNM were analyzed; Figure 4.9 summarizes all these parameters, their classification and their dependencies. The stability parameters are classified as Technology parameters (Geometry of transistors, Threshold Voltage (V_{TH}) and Channel length modulation parameter (λ)) and Non-technology parameters (Supply Voltage (V_{DD}), Temperature (T)). The dependency of these parameters on physical factors is as follows:

- Technology parameters depend on - t_{ox}, $L_{n,p}$, $W_{n,p}$, $N_{a,d}$
- Non-technology parameters depend on - Voltage ramp-up

Any variation in any of these parameters can impact the PSNM and thus reduce the cell stability. The impact of these parameters on start-up values and the PSNM will be the topic of discussion in the next chapter.

Experimental Results For 65nm Technology

5

The analytical model developed in the previous chapter was instrumental in determining the technology as well as the non-technology parameters that impact the start-up value of an SRAM cell. This chapter focuses on quantifying this impact and determining the most dominant parameter amongst the set of stability parameters using circuit simulation.

Section 5.1 describes the objectives of the experiments performed. The impact of the non-technology parameters like supply voltage and temperature are presented in Section 5.2. The combined impact of technology and non-technology parameters parameters is discussed in Section 5.4. Section 5.6 concludes the chapter by summarizing the experimental results obtained.

5.1 Objective and simulation model

The objective of the experiments performed is to emulate the impact of variation in technology and non-technology parameters on the start-up value of an SRAM cell. To do so, we vary one of the technology or non-technology parameters for one of the MOSFETs and we experimentally quantify this impact using circuit simulation. By varying the technology parameters, we effectively introduce a mismatch between the two inverters of the cell. The mismatch introduced for experimental purpose is such that variation in the parameter represents the worst case scenario as predicted by [41]. Variation in non-technology parameters helps in determining the behavior of the SRAM PUFs under stress conditions. The objective of the experiment is therefore:

1. Determine the impact of non-technology parameters on the start-up value of the SRAM cell.
2. Determine the impact of technology parameters on the start-up value of the SRAM cell.
3. Determine the combined impact of the technology and non-technology parameters on the start-up value of SRAM cell.
4. Validate the results obtained using analytical model by comparing them with actual silicon results.

The objectives of the experiment and the experiment itself are unique and to the best of out knowledge, nothing is published about the stability parameters of SRAM PUF and their impact on the start-up values of SRAM cells till now. Most of the literature

Table 5.1: Parameters for SRAM cell in for $65nm$ BSIM4 model

Parameter	NMOS	PMOS
Temperature (in ° C)	20	20
Supply voltage (in Volts)	1.2	1.2
Length (in nm)	65	65
Width (in nm)	195	130
Threshold Voltage (in V)	0.423	0.365
Gate Oxide Thickness (in nm)	1.85	1.95

[7][8][27][29] analyzes the impact of process variation on the traditional SRAM cell during read or write operation. However, the same results cannot be applied for SRAM PUFs as the access transistors play a negligible role in determining the start-up value of the SRAM cell as discussed in Section 4.2.2. Therefore, the rest of the chapter focuses on each of the objectives mentioned above and summarizes the results obtained for each of them. The validation of the results obtained, however, will be done in Chapter 6. The experiments have been performed using BSIM4 models for $65nm$ technology node [1][5][9][40]. The analytical model proposed in Chapter 4 has been mathematically modeled using Maple. Table 5.1 provides the nominal values of the various parameters which impact the start-up value for $65nm$ technology.

5.2 Impact of the non-technology parameters

The reliability of the SRAM PUFs' start-up value depends upon how sensitive the start-up value is to the environment where the PUF will be used. As mentioned in Section 4.1, PSNM is used as a metric to measure the stability of SRAM PUF and can be graphically calculated using the VTC of the SRAM cell. This section explores the potential influence of environmental or the non-technology factors such as supply voltage and the ambient temperature on the PSNM and hence on the start-up value of an SRAM cell. One of the non-technology parameters is varied for an SRAM cell and its impact has been analyzed for skewed as well as cell non-skewed cell. The experiment is performed using the technology parameters for $65nm$ technology and the nominal values of all the parameters are indicated in Table 5.1.

5.2.1 Impact of supply voltage

The start-up value of an SRAM cell depends mainly on the mismatch between the inverters, noise present in the environment and the supply voltage as well. On varying the supply voltage, it is observed that the PSNM for an SRAM cell linearly increases with the increase in supply voltage as shown in Figure 5.3. For a $\pm 10\%$ change in supply voltage, PSNM changes by 1%. In other words, lower supply voltage makes a cell more

susceptible to noise and hence the probability of a cell to flip its state is higher at low supply voltages.

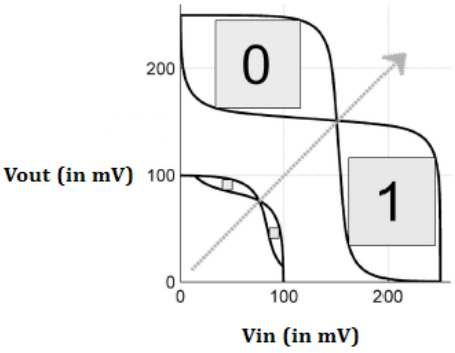

Figure 5.1: VTC of a non-skewed SRAM cell for different voltages [17]

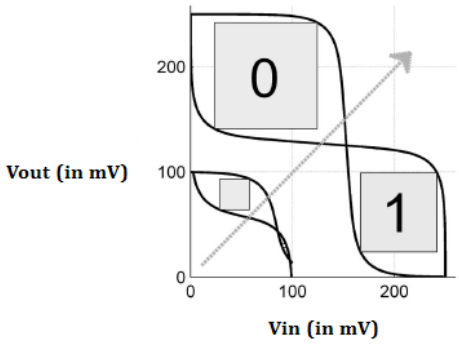

Figure 5.2: VTC of a skewed SRAM cell for different voltages [17]

Figure 5.4 show the PSNM for a non-skewed cell for different supply voltages. As can be seen from the figure, the two eyes of the VTC curve reduce equally for this non-skewed cell with a decrease in supply voltage. However, for a skewed cell, lowering the supply voltage would reduce the SNM for one state significantly as compared to the other as shown in Figure 5.1 and Figure 5.2. It can be deduced from the Figure 5.2 that for every skewed cell, there is a supply voltage for which the cell takes its preferred state. In the cell shown in the figure, this transition occurs at 100 mV. The transition voltage may vary for different cells depending upon the amount of mismatch present in the cell.

5.2.2 Impact of temperature

The impact of temperature on the SRAM cell has been studied for temperatures ranging from $-40°C$ to $80°C$ which represents the commercial temperature range for SRAMs.

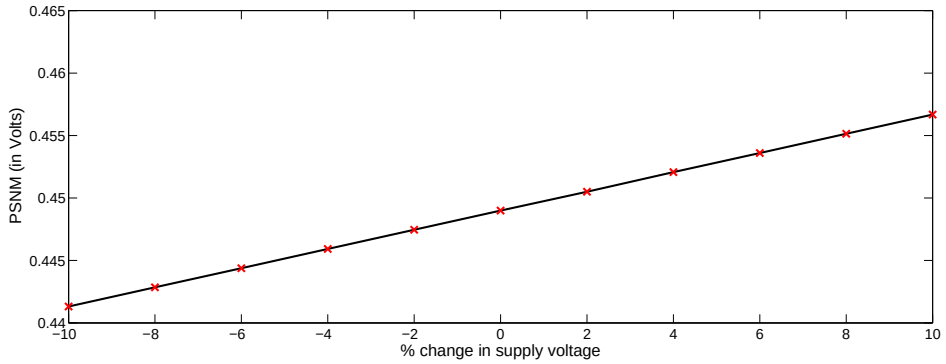

Figure 5.3: PSNM of an SRAM cell for different supply voltages

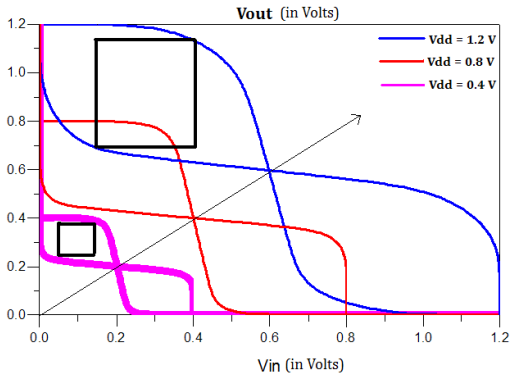

Figure 5.4: 'Eyes' of the PSNM curve of an SRAM cell for different supply voltages

PSNM is observed to linearly decrease with the increase in temperature as shown in Figure 5.5. For a ±10% change in temperature, PSNM changes by 0.5%. However, the impact of temperature becomes more pronounced for large variations in temperature.

This observation can be explained as follows: The change in temperature of the environment has an impact on the threshold voltages of the devices, the electron and hole mobilities and the thermal noise in the devices. The threshold voltage for the devices decrease [10] whereas the mobility of the electrons and the holes increase with the increase in temperature [24]. The thermal noise increases with the increase in temperature. For a non-skewed cell, the first two effects impact both the inverters similarly. Hence, the PSNM is impacted only by thermal noise and therefore it decreases with the increase in temperature.

For a skewed cell, due to decrease in threshold voltage at higher temperatures, the gain of the inverters is lowered and hence one of the 'eyes' of the PSNM curve become smaller. Moreover, the thermal noise in the system still increases. Therefore, a combination of both these factors in turn reduce the PSNM.

Therefore, it can be concluded that the higher temperatures are more prone to noise and hence the tendency of a cell to flip at higher temperatures would be more.

5.3 Impact of variation in technology parameters

Assuming that the interconnects are ideal, it can be said that the non-technology parameters impact all the MOSFETs in the SRAM cell in a similar way unlike the process variation. Due to the process variation, different technology parameters belonging to one or more MOSFETs vary and hence may make a cell skewed. To simulate such a condition and find its impact on the PSNM as well as the start-up value of SRAM cell, we have varied one of the technology parameters in one of the MOSFETs at a time.

This experiment utilizes the analytical model proposed in the previous chapter and determines the impact of process variation in the analyzed parameter on the PSNM. The

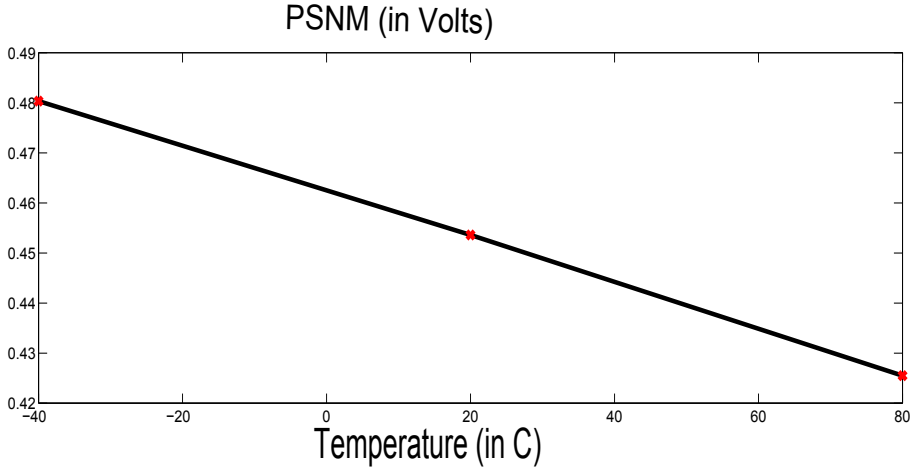

Figure 5.5: Variation in PSNM due to variation in temperature for $65nm$

analysis has been done using Maple as the tool. The experiment helps in determining the noise tolerance of the cell due to variation in the analyzed parameter. Table 5.1 shows the nominal values of the various parameters.

5.3.1 Impact of length of the MOSFETs

Since the SRAM cells use the smallest device size possible, the length for all the MOSFETs used for the experimental setup is 65nm; the smallest possible dimension for the $65nm$ technology node. With the technology scaling, the device sizes are shrinking and a very small variation in the device's dimension can impact the cell stability and hence the power-up state of the SRAM. For an SRAM, two kinds of MOSFETs are involved in deciding the power-up state of the SRAM: the PMOS i.e. the load and the NMOS i.e. the driver. Variation in length of each of them will have a different impact on the PSNM. The process variations that leads to variation in SRAM cell length are the LER and the LWR which were discussed in Section 3.4.2. The impact of variation in channel length modulation parameter on PSNM can also be seen in variation in length because of it's dependence on length.

It has been observed that for $65nm$ technology node, the ratio of standard deviation to mean variation (σ) for length due to process variation is $\pm 4\%$ [41]. Therefore, to cover the entire range of $\pm 3\sigma$, we consider a range of $\pm 12\%$ variation in length and see its impact on the PSNM. Figure 5.6 shows the probability distribution function of the process variations across the length of the MOSFETs for $65nm$ technology.

Figure 5.8 shows the impact of variation in length of the two MOSFETs on the PSNM. Three observations can be made:

1. Any deviation from the nominal value decreases the PSNM thus reducing the noise tolerance of the cell.

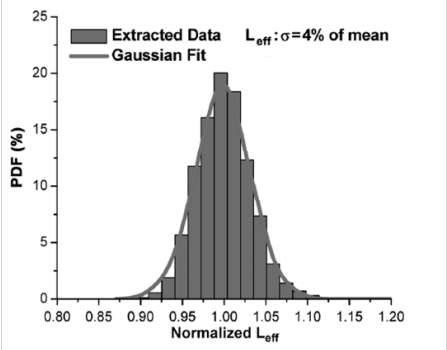

Figure 5.6: Probability Distribution Function of length due to process variation for $65nm$ [41]

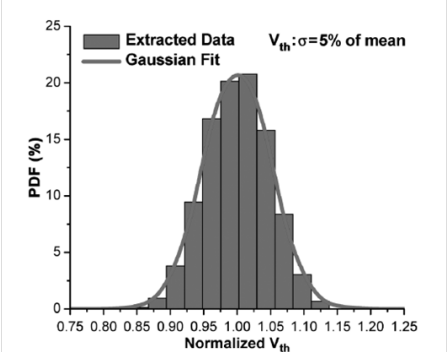

Figure 5.7: Probability Distribution Function of V_{TH} due to process variation for $65nm$ [41]

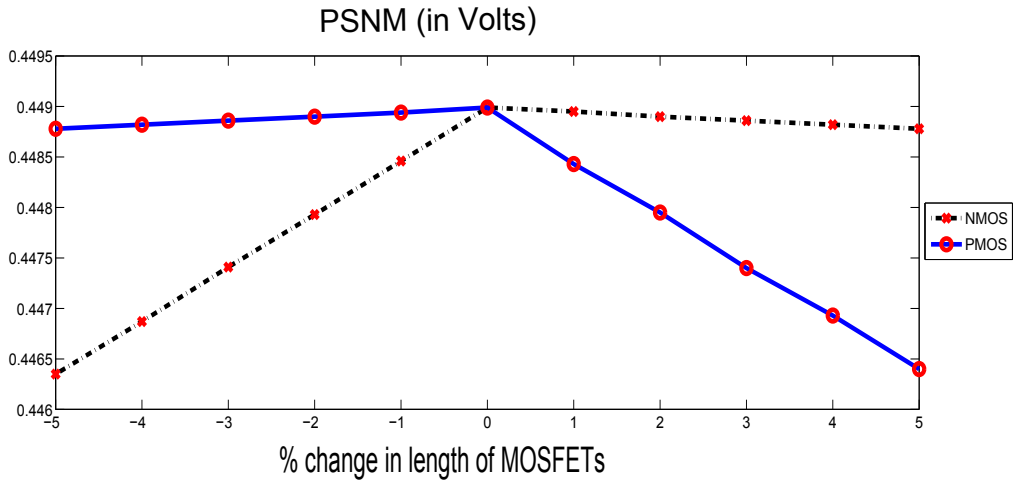

Figure 5.8: Impact on PSNM due to variation in length

Explanation: Increasing or decreasing the length of one of the MOSFETs changes the VTC curve such that one of the 'eyes' of the VTC curve decreases in size and the other one increases/remains the same. Since the PSNM is determined by the minimum side of the square that can fit in the VTC curve as explained in Chapter 4, the overall value of the PSNM decreases.

2. Only decreasing the length of NMOS and increasing the length of PMOS has an impact on the PSNM.
   Explanation: By decreasing the length of the NMOS (of suppose Inverter 1), we effectively decrease the resistance offered by it and hence make the Inverter 1 faster. This makes the VTC curve for Inverter 1 more steep thus reducing the size of one of the 'eyes' of the VTC curve. Thus the PSNM decreases.
   On the other hand, increasing the length of the same NMOS increases the size of

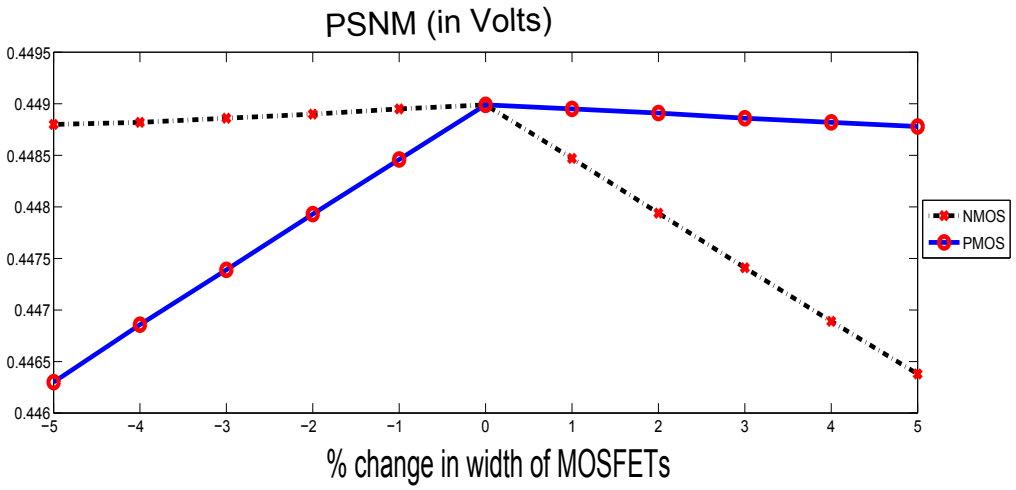

Figure 5.9: Impact on PSNM due to variation in width for $65nm$

one of the 'eyes' whereas the other 'eye' remains the same. Again, since PSNM is determined by the minimum side of the square that can fit in the VTC, the PSNM here remains the same as was for the nominal case.

3. The percentage change in PSNM due to both the load transistor (PMOS) and the driver (NMOS) is similar for same variation in length. For a ±5% variation in length, the PSNM changes by 1%.
   Explanation: Since the length of the PMOS and the NMOS for both the inverters is identical, the change in PSNM is similar for the same variation in length.

5.3.2 Impact of width of the MOSFETs

A perfectly matched SRAM cell has the ratio of $\frac{W_{pmos}}{W_{nmos}} = \frac{2}{3}$ with length being $65nm$ for both of them as shown in Table 5.1. The variations in width of the MOSFETs occur due to LER and LWR which are also the main sources of variation in length of the MOSFETs.

Figure 5.9 shows the impact of variation in length of the two MOSFETs on the PSNM. Three observations can be made:

1. Any deviation from the nominal value decreases the PSNM thus reducing the noise tolerance of the cell.
   Explanation: Increasing or decreasing the width of one of the MOSFETs changes the VTC curve such that one of the 'eyes' of the VTC curve decreases in size and the other one increases/remains the same. Since the PSNM is determined by the minimum side of the square that can fit in the VTC curve as explained in Chapter 4, the PSNM decreases

2. Only decreasing the width of PMOS and increasing the width of NMOS has an impact on the PSNM.
   Explanation: By decreasing the width of the PMOS (of suppose Inverter 1), we effectively increase the resistance offered by it and hence make the Inverter 1 slower. This makes the VTC curve for Inverter 1 less steep thus increasing the size of one of the 'eyes' of the VTC curve. Thus the PSNM here remains the same as was for the nominal case.
   On the other hand, increasing the width of the same NMOS lowers the resistance and the size of one of the 'eyes' decreases whereas the other 'eye' remains the same. Again, since PSNM is determined by the minimum side of the square that can fit in the VTC, the overall PSNM decreases.

3. The percentage change in PSNM due to both the load transistor (PMOS) and the driver (NMOS) is similar for same variation in length. For a $\pm 5\%$ variation in length, the PSNM changes by 0.44%.
   Explanation: The impact of variation in width of PMOS and NMOS is much smaller as compared to variation in length because of the large feature size. Also, since the source of variation is same for length and width, even though there was no given *Probability Distribution Function* (PDF) for width for $65nm$ technology, we assume the worst case scenario and base our results upon PDF given for length as was considered in previous subsection.

5.3.3 Impact of the threshold voltage of the MOSFETs

The threshold voltage (V_{TH}) of the MOSFET depends upon a number of technology and non-technology parameters like the doping concentration, thickness of the oxide, temperature, etc as discussed in Section 4.3.2. Any change due to process variation in one or more of these parameters directly influences the threshold voltage and hence can induce a mismatch between the inverters of the SRAM cell. One of the most dominant sources of change in V_{TH} is Random Dopant Fluctuation which was discussed in Section 3.4.1. Also, the variation in thickness of the gate oxide has an impact on the V_{TH}. Variation of V_{TH} in the driver and the load transistors impacts the PSNM and the start-up value of the SRAM cell differently.

It has been observed that for $65nm$ technology node, the ratio of standard deviation to mean variation (σ) for V_{TH} due to process variation is $\pm 5\%$ [41]. Therefore, to cover the entire range of $\pm 3\sigma$, we consider a range of $\pm 15\%$ variation in the V_{TH} and see its impact in the two scenarios mentioned above. Figure 5.7 shows the probability distribution function of the process variations across the threshold voltage of the MOSFETs for $65nm$ technology. To emulate this variation in the simulation model, we have varied the parameter V_{TH_0} which denotes the threshold voltage of a long channel MOSFET at zero volt substrate bias.

Figure 5.10 shows the impact of variation in V_{TH} of the two MOSFETs on the PSNM. Following observation can be made from the figure:

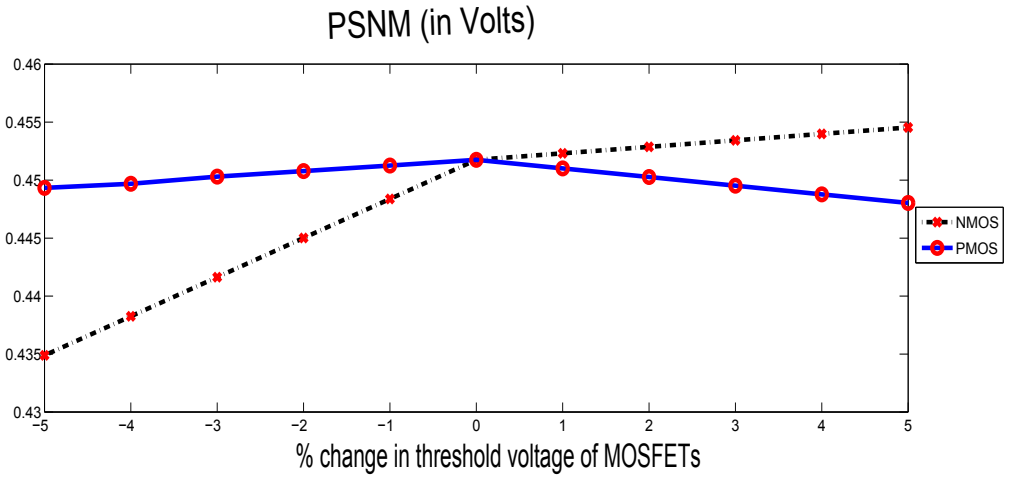

Figure 5.10: Variation in PSNM due to variation in threshold voltage for $65nm$

- The impact of variation in V_{TH} of NMOS dominates and the PSNM changes by 3.33% for a ±5% change in V_{TH} of the NMOS and the change in PSNM due to PMOS is negligible.
  Explanation: The SRAM cell is designed such that the width of the NMOS is greater than that of the PMOS. This enhances the drive capability of the NMOS and a very minor variation in its threshold voltage changes the PSNM significantly. On the other hand, the PMOS is very weak and a small change in its threshold voltage hardly impacts the PSNM.

5.3.4 Impact of oxide thickness of the MOSFETs

The thickness of gate oxide (t_{ox}) for $65nm$ technology node is order of thickness 4 to 5 atoms. As mentioned in Section 3.4.3, the roughness introduced by process variation, although small between silicon and silicon dioxide, can be of one or two atomic layers, meaning around 50% of the of the t_{ox} (when t_{ox} is around 1nm) [30]. For the given technology file, t_{ox} for both the PMOS and the NMOS are indicated in Table 5.1. Since there was no available distribution function for t_{ox} for this technology node, we have assumed the worst case variation of ±30% and analyzed its impact in the two scenarios considered.

Figure 5.11 shows the impact of variation in t_{ox} of the two MOSFETs on the PSNM. The following observation can be made by inspecting the figure:

- The change in PSNM due to the NMOS is much higher than due to the NMOS. The PSNM changes by 4.4% for a ±10% change in t_{ox}.

  Explanation: This can be attributed to the strong driving capability of NMOS and hence a small change in its t_{ox} changes the V_{TH}; this in turn changes the PSNM by a large amount.

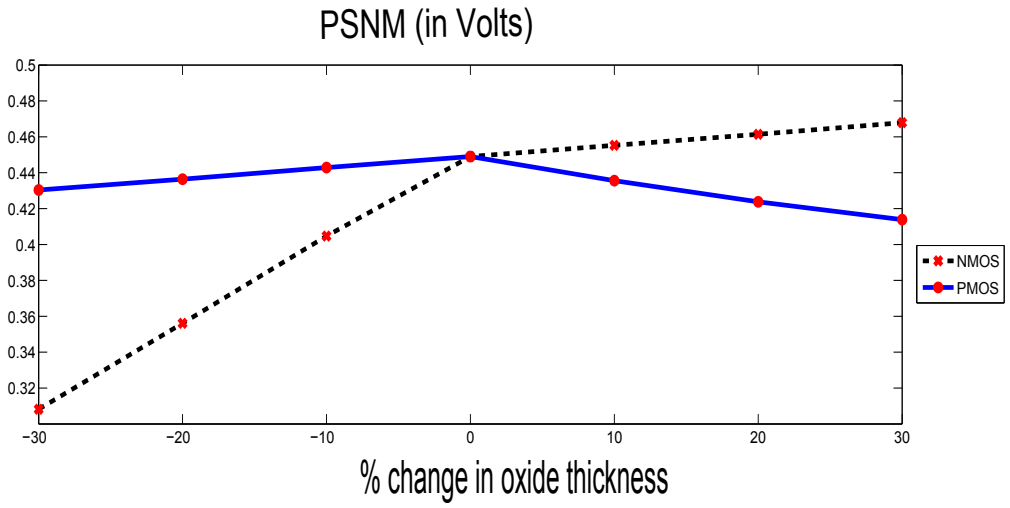

Figure 5.11: Variation in PSNM due to variation in oxide thickness for $65nm$

5.4 Combined impact of variation in stability parameters

The previous two sections dealt with analyzing the impact of technology and non-technology parameters alone on the PSNM of the SRAM cell. However, in reality, a combination of these parameters together impact the PSNM and hence determine the start-up value of the SRAM cell. To emulate this condition, we consider a skewed-cell (has a variation in one of the the technology parameters) and analyze the impact of non-technology parameters (voltage ramp-up and temperature) on this cell.

The objective of this experiment is to determine the impact of the voltage ramp-up on the start-up value of the cell. The reliability of the start-up values needs to be ensured under these conditions. As mentioned earlier, to be used as a security key, it needs to be assured that the probability of the cell to be fully-skewed is very high for the allowed range of variation in the parameter. To refresh the idea of a fully-skewed cell, it is a cell which has enough mismatch between the two inverters such that the cell takes a fixed value (1 or a 0) for all stress conditions. Therefore, this experiment focuses on determining the percentage variation in the parameter that will make the cell fully-skewed. The experiment has been done by implementing the schematic of SRAM cell in Agilent ADS which is a spice based simulation tool. For experimental purposes, three values of ramp-up i.e. $1\mu s$, $10\mu s$ and $100\mu s$ are considered.

The following sections show the impact of variation in each parameter for various voltage ramp-ups. The values of various parameters for a perfectly matched SRAM cell have been shown in Table 5.1.

5.4.1 Impact of variation in length of the MOSFETs for various voltage ramp-ups

For an SRAM, the two kinds of MOSFETs are involved in deciding the power-up state of the SRAM: the PMOS i.e. the load and the NMOS i.e. the driver. To make a cell fully-skewed, the percentage variation in length required for both NMOS and PMOS are different.

Figure 5.12 and Figure 5.13 show the impact of the variation in length of NMOS and PMOS for the three different ramp-up values considered. The variation in length is represented on the x-axis whereas the y-axis represents the output state of the cell for a particular variation. A marker at 1 represents a fully-skewed cell with the output state 1. A marker at 0 represents a fully-skewed cell with the output state 0. A marker at 0.5 represents a partially-skewed or a fully-skewed cell whose state can either be 0 or 1. By inspecting the figure, we see that for a cell to become fully-skewed, a 5% variation in length of NMOS is required whereas for PMOS, only 3% variation is sufficient to do the same. Also, the following observation can be made from the figure:

- As we reduce the ramp-up speed, the cell tends to take its preferred state as predicted by the variation in the technology parameter.
  Explanation: For every cell, there is a certain voltage at which the transition to certain favored state (0 or 1) occurs. Although, the voltage at which the transition to this favored state occurs would differ for every cell depending upon the amount of mismatch. By increasing the supply voltage very slowly, we ideally increase the amount of time for which the voltage is held at any particular value including the one for which the transition occurs. This makes the cell take the preferred state for slow ramp-ups whereas for fast ramp-ups, the change in voltage in much faster than the amount of time it takes to attain the start-up state. A similar argument was given in [17] where the authors used the SRAM cell for random number generation. Therefore, it can be concluded that slower ramp-ups make the cell take their preferred state.

5.4.2 Impact of variation in width of the MOSFETs for various voltage ramp-ups

MOSFETs involved in deciding the power-up state of the SRAM are the PMOS and the NMOS. To make a cell fully-skewed, the percentage variation in width required for both NMOS and PMOS are different. Since, the impact of variation in width has a very less impact on PSNM, the behavior of width variation at different voltage ramp-ups can be ignored. Although the impact of variation in width has a similar impact for different voltage ramp-ups as was for the length variation. Therefore, the observation still remains the same; slower the ramp-up, higher is the probability of cell to take its preferred state. Moreover, this observation can be explained in similar way as was done for length in the previous subsection.

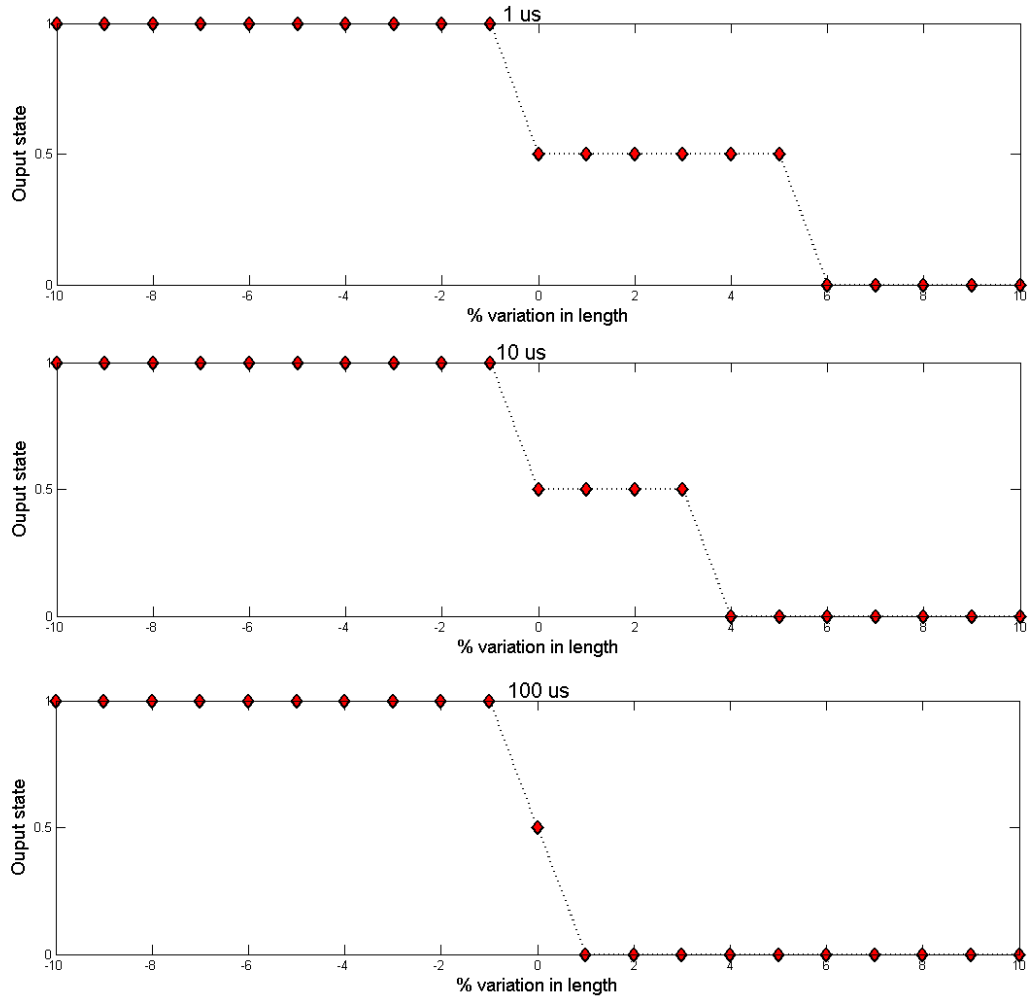

Figure 5.12: Output state of cells for % variation in length of NMOS at various ramp-ups

5.4.3 Impact of variation in threshold voltage of the MOSFETs for various voltage ramp-ups

Figure 5.14 and 5.15 shows the impact of the variation in V_{TH} of NMOS and PMOS respectively, for three different ramp-up values considered. The variation in threshold voltage is represented on the x-axis whereas the y-axis represents the output state of the cell for a particular variation. A marker at 1 represents a fully-skewed cell with the output state 1. A marker at 0 represents a fully-skewed cell with the output state 0. A marker at 0.5 represents a partially-skewed or a non-skewed cell whose state can either be 0 or 1. For a cell to become fully-skewed, a 2% variation in threshold voltage of the PMOS is required whereas for NMOS, only 1.5% variation is sufficient to do the same as can be seen from Figure 5.14 and Figure 5.15. The following observation can also be made by inspecting the figure:

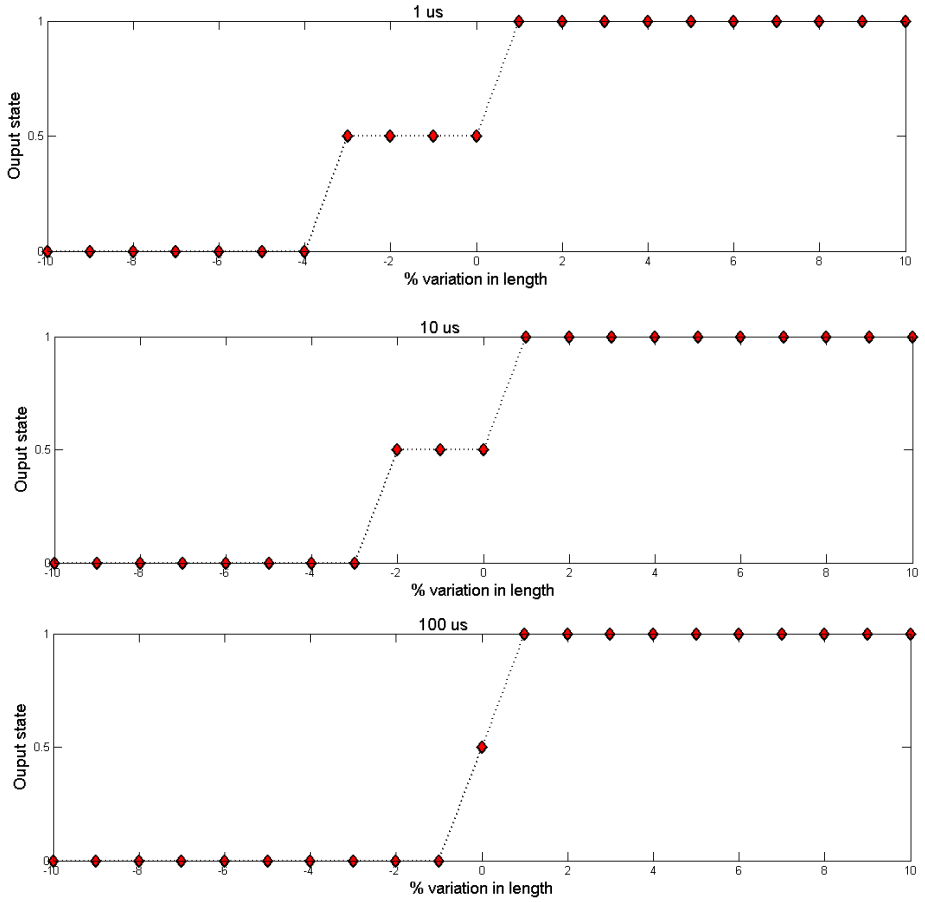

Figure 5.13: Output state of cells for % variation in length of PMOS at various ramp-ups

- As we reduce the ramp-up speed, the cell tends to take its preferred state as predicted by the variation in the technology parameter.
  Explanation: This can be explained using the same logic as was explained for variation in length. The slower ramp-ups make the cell settle to its preferred state easily.

5.4.4 Impact of variation in oxide thickness of the MOSFETs for various voltage ramp-ups

Since change in t_{ox} impacts the V_{TH}, the impact of the variation in t_{ox} with different ramp-up values is similar to variation in V_{TH} for same stress condition. Therefore, the slower the ramp-up, the cell tends to take its preferred state for a lesser variation in t_{ox}.

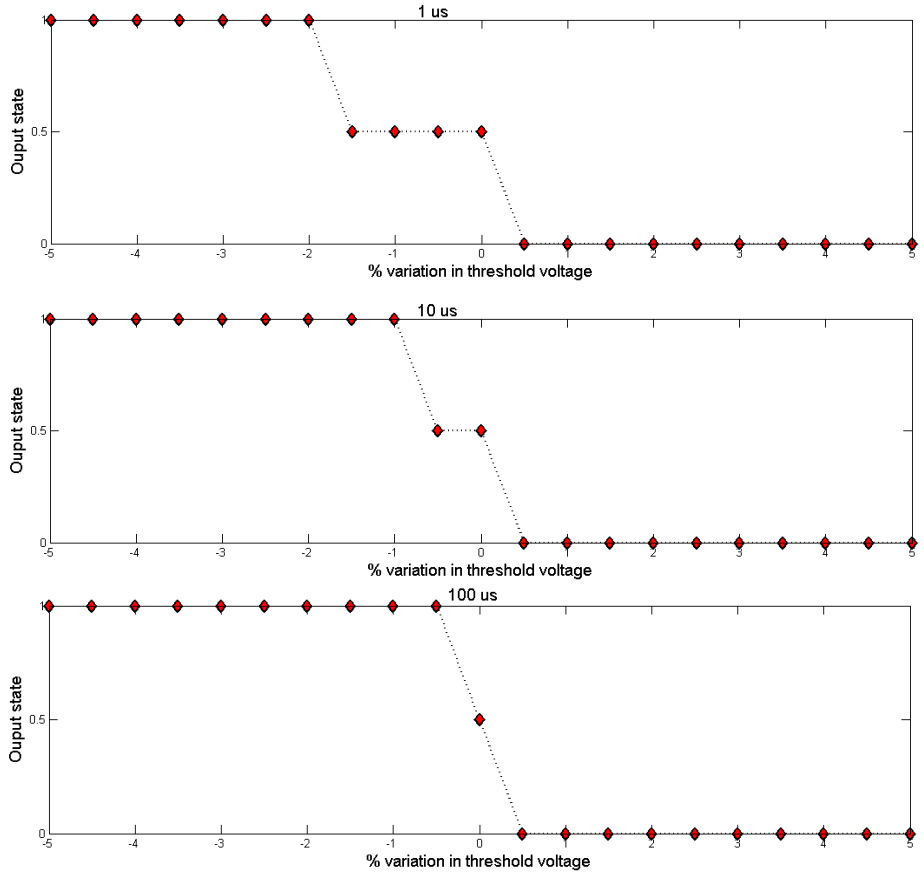

Figure 5.14: Output state of cells for % variation in threshold voltage of NMOS at various ramp-ups

5.5 Proof of robustness

In the previous two sections, it was observed that the variation in V_{TH} of the NMOS has the highest probability of making the cell fully-skewed. To prove the robustness of the start-up values of SRAM PUF, we need to prove the reproducibility of SRAM start-up value for V_{TH} variation. If successful, we can ensure the reproducibility of the SRAM start-up value. This section focuses on mathematically proving the reproducibility of SRAM start-up value for V_{TH} variation and hence the robustness of SRAM PUF.

The PDF for the threshold voltage has been shown in Figure 5.7. The PDF, as can be seen from the figure, is a Gaussian distribution. The probability of a cell being fully-skewed can be calculated by integrating the Gaussian curve for the range which can make the cell fully-skewed. In Section 5.3, we observed that a 1.6% variation in V_{TH} of NMOS makes the cell fully-skewed. Therefore, we integrate the Gaussian distribution for V_{TH} from the mean value (μ) to this range ($\mu + 0.016$). Any variation outside this range will make the cell fully-skewed. Equation 5.1 represents the Gaussian distribution

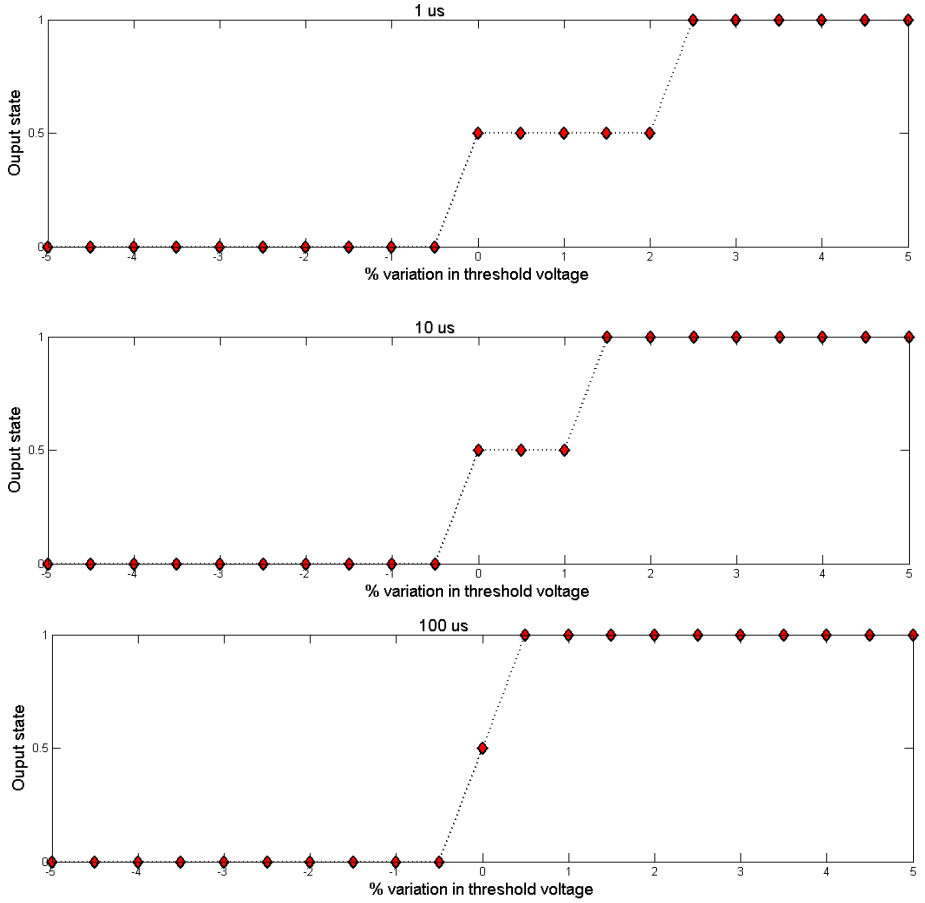

Figure 5.15: Output state of cells for % variation in threshold voltage of PMOS at various ramp-ups

equation.

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{5.1}$$

where σ is the standard deviation of the V_{TH} and

x represents the variation in V_{TH}.

On putting the respective values of the variables for V_{TH} variation, we get the following equation:

$$P(x) = \int_{1}^{1.016} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-1)^2/2\sigma^2} dx \tag{5.2}$$

Substituting $\sigma = 0.05$ in equation 5.2 gives the probability of cell being non-skewed of partially-skewed as 16%. Thus, the probability of cell being fully-skewed becomes 84%.

This calculation is done using variation in one parameter for one of the MOSFETs at a time. The number is bound to increase considering the variation in all the MOSFETs. But this calculation gives a fair idea that the cell has a high probability of being fully-skewed due to threshold voltage variation. Therefore, the reproducibility of the start-up behavior can be ensured.

5.6 Inferences from experimental results

The variation of technology as well as the non-technology parameters has been analyzed in this chapter. Few inferences can be made from the analysis of analytical model. These are:

1. Slower the voltage ramp-up, higher is the probability of cell taking its preferred state.
2. Lower temperatures have a higher probability of making the cell take its preffered state.
3. For a $\pm 5\%$ change in a technology parameter, PSNM changes as follows:

   Length - 1.04%
   Width - 0.44%
   Oxide thickness - 4.44%
   Threshold Voltage - 3.33%
4. Amongst the technology parameters, a 1.6% variation in threshold voltage of the NMOS makes the cell fully-skewed. Due to this variation, cell has a 84% probability of being fully-skewed.
5. Order of impact of the technology parameters: Threshold voltage > Geometry of the transistors.

These results are based on the experiments done with the analytical model. In order to validate these results, measurements on real silicon SRAM devices have been performed. The results are reported in Chapter 6.

Validation Of The Analytical Model For 65nm

6

The impact of the technology and the non-technology parameters was analyzed in the previous chapter. This chapter reports the measurements performed on 40 SRAM devices built in 65nm technology node under various stress conditions and compares them with the experimental results obtained in Chapter 5. The experiments were performed at a company, Intrinsic ID, in Eindhoven, The Netherlands.

Section 6.1 gives a brief overview about the SRAM devices under consideration and the experiments that have been performed; SRAM devices from NXP and TSMC are considered for the experiments. Section 6.2 discusses the results obtained for voltage ramp-up experiment using the SRAM devices. Section 6.3 describes the results obtained for temperature cycle experiment. The industrial results obtained in these two sections have been used to validate the analytical model proposed in this work.

6.1 Experimental setup

In the previous chapter, we have analyzed the impact of the technology and the non-technology parameters on the stability of SRAM PUFs using the analytical model developed in Chapter 4. To evaluate the correctness of the analytical model, actual silicon data is needed. Therefore experiments where measurements on real silicon SRAMs need to be performed. The variation in the technology parameters is a process dependent phenomenon and is hard for us (if not impossible) to control or measure it. Therefore, the results obtained for technology parameters cannot be validated industrially. However, the impact of non-technology parameters (supply voltage and temperature) can be measured by stressing the chip for a range of temperatures and voltage ramp-ups.

The behavior of 40 different SRAM devices have been considered for the experiment at Intrinsic ID; 20 NXP Semiconductor devices and 20 TSMC devices. For test purposes, one SRAM each from NXP and TSMC has been integrated in an IC by IMEC, Holst centre, The Netherlands. All these devices are manufactured using the $65nm$ technology node. The size of memory from both the companies is 65536 bits (8kB). Two different experiments were performed at Intrinsic ID. First the start-up value of the SRAM devices were measured for three voltage ramp-ups namely $1\mu s$, $100\mu s$, and $10ms$. Thereafter, the start-up value of the devices were screened, but now for three different values of temperatures: $-40°C$, $20°C$, and $80°C$. The results and their analysis will be presented in this chapter.

Figure 6.1 shows an example of the start-up state of an SRAM from TSMC at a temperature of $20°C$ and a voltage ramp-up of $1\mu s$. Each dot here represents an SRAM cell bit; a black dot present a 0 value whereas a white dot represents the 1 value. The string

formed by these start-up values of the memory cells of the SRAM form an identifier that identifies the SRAM uniquely. Depending upon the stress conditions that each of these devices is put to, some of the cells might flip their value thus changing the identifier. This variation in the identifier is measured using the concept of *Fractional Hamming Distance* (FHD) as explained in Section 3.2 [25]. The following paragraph explains the concept in more detail with respect to the measurements performed.

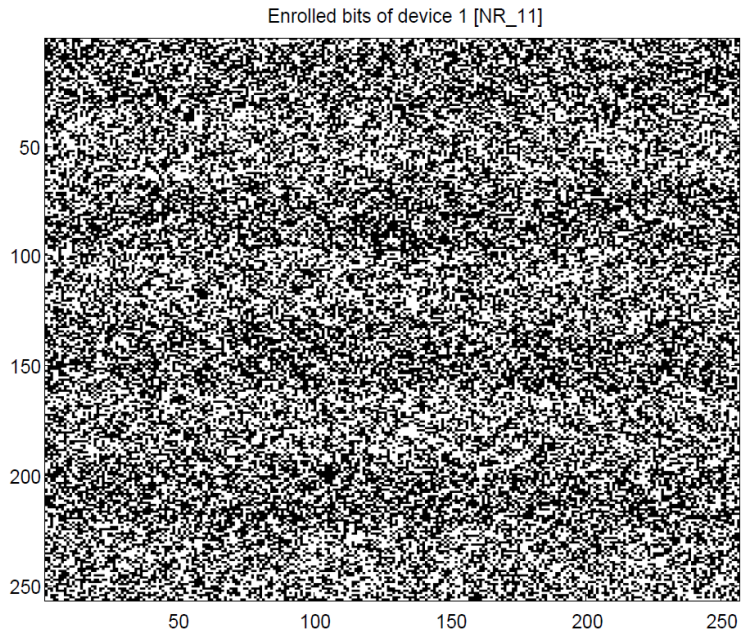

Figure 6.1: Start-up state of a memory

Fractional Hamming Distance (FHD)

Fractional Hamming distance among the two measurements of the same device gives us the amount of noise that is present between these measurements. In other words, it denotes the percentage of bits that flip when compared to the first measurement. As mentioned in Section 1.2, in order to be used for the generation of security key, the start-up behavior of an SRAM needs to be reproducible. In practice, this implies that the FHD distance should be as low as possible. In an ideal case, when all the bits are reproducible, the FHD would be 0. The experiments performed use the FHD as a criteria to estimate the noise resulting from different stresses.

6.2 Voltage ramp-up experiment

From the experiments performed using the analytical model in Section 5.2, we had concluded that the amount of noise in an SRAM cell is proportional to the voltage ramp-up used. In order to validate this result, we analyze each of the ICs for a range

of ramp-ups: $1\mu s$, $100\mu s$, and $10ms$ at room temperature of $20°C$. This has been done both for TSMC devices and NXP devices.

For the experiment, ten ICs containing ten devices each from NXP and TSMC are kept in a test set-up which is efficient for varying voltage ramp-ups of the core voltage of the IC. Each of these IC's is powered-up repeatedly ten times and after each power-up, the start-up states of the memory are read and stored in a binary dump. This binary dump is then analyzed using a MATLAB file. The start-up values of TSMC and NXP devices are analyzed separately.

6.2.1 Results for TSMC devices

Figure 6.2 to Figure 6.4 show the analysis of the measurements performed for ten TSMC devices at different ramp-ups.

Every measurement of a particular device is denoted by a colored line in the plots shown in Figure 6.2 to Figure 6.4. The measurement number is plotted on the x-axis and the FHD on the y-axis. The FHD in each case is calculated with respect to the first measurement at a particular ramp-up. Therefore, the y-axis here denotes the fraction of bits that differ between measurement 1 and the rest of the measurements. Ideally, to be used as a cryptographic key directly, there should be no difference between the consecutive power-ups of the same device. This implies that the FHD should be 0 in the ideal case.

It is observed that the FHD or the percentage error in measurement for the ten devices at $1\mu s$ is centered around 5.75%. However, for measurements at $100\mu s$, the error becomes approximately equal to 5.5%. As the ramp-up speed is further reduced, the percentage error reduces significantly and becomes around 3% for a ramp-up speed of $10ms$. From the figures, it is clear that the FHD decreases with slower ramp-ups for all the devices under consideration. In other words, the percentage of bits that flip decreases with slower ramp-ups.

This can be understood as follows: For every SRAM cell, there is a certain supply voltage at which it tends to take its preferred state (as predicted by the mismatch). By slowly ramping-up the supply voltage, we allow the SRAM to be at a particular voltage for a longer time. This makes more and more number of cells take their preferred state during slow ramp-ups. An important thing to note here is, the supply voltage at which this transition occurs differs for every SRAM cell. Every skewed cell will have a supply voltage where it will take its preferred state, given sufficient time to settle. Therefore, at slow ramp-ups like $10ms$, most of the cells take their preferred state and hence the percentage error in the measurement decreases.

6.2.2 Results for NXP devices

Figure 6.5 to Figure 6.7 show the analysis of the measurements performed for ten NXP devices at different ramp-ups.

Every measurement of a particular device is denoted by a colored line in the plots shown in Figure 6.5 to Figure 6.7. The measurement number is plotted on the x-axis and the FHD on the y-axis. The FHD in each case is calculated with respect to the first measurement at a particular ramp-up. Therefore, the y-axis here denotes the fraction of bits that differ between measurement 1 and the rest of the measurements. Ideally, to be used as a cryptographic key directly, there should be no difference between the consecutive power-ups of the same device. This implies that the FHD should be 0 in the ideal case.

It is observed that the FHD or the percentage error in measurement for the ten devices at $1\mu s$ is centered around 6.5%. however, for measurements at $100\mu s$, the error becomes approximately equal to 6%. As the ramp-up speed is further reduced, the percentage error reduces significantly and becomes around 4% for a ramp-up speed of $10ms$. From the figures, it is clear that the FHD decreases with slower ramp-ups for all the devices under consideration. In other words, the percentage of bits that flip decreases with slower ramp-ups.

For various values of ramp-ups, the observation is similar to that of the TSMC devices and therefore same logic can be applied to explain the results. However, the percentage error differs for NXP and TSMC devices and this can be accredited to the different design and layout of the SRAM for both the companies.

6.3 Temperature cycle experiment

From the simulations performed using the analytical model in Section 5.2, we determined that lower temperatures linearly reduce the amount of noise in an SRAM cell. In order to validate this result, we analyzed 20 ICs (with one NXP and TSMC device each) for a range of temperatures: $-40°C$, $20°C$, and $80°C$ for a ramp-up of $1\mu s$. 20 ICs are kept in a test set-up which is efficient for varying voltage ramp-ups of the core voltage of the IC. Each of these IC's is powered-up repeatedly 20 times and after each power-up, the start-up states of the memory are read and stored in a binary dump. These binary dumps are then analyzed using a MATLAB file.

6.3.1 Results for TSMC devices

Figure 6.8 to Figure 6.10 show the analysis of the measurements performed for ten TSMC devices at different temperatures.

Every measurement of a device and its corresponding measurements are denoted by a colored line in the plots shown in Figure 6.8 to Figure 6.10. The measurement number is plotted on the x-axis and the FHD on the y-axis. The FHD in each case is calculated with respect to the first measurement at a particular ramp-up. Therefore, the y-axis here would denote the fraction of bits that differ between Measurement 1 and the rest of the measurements. Ideally, to be used as a cryptographic key directly, there should be no difference between the consecutive power-ups of the same device. This implies that the FHD should be 0 in the ideal case.

It is observed that the FHD or the percentage error in measurement for twenty devices at $-40°C$ is centered around 5.5%. however, for measurements at $20°C$, the error becomes approximately equal to 6%. As the temperature is further increased, the percentage error increases and becomes around 6.5% for a temperature of $80°C$. From the figures, it is clear that the FHD increases with increase in temperature for all the devices under consideration. In other words, the percentage of bits that flip decreases at lower temperatures.

This can be understood as follows: As the temperature of the SRAM is increased, the thermal noise in the system increases which leads to a more random power-up state. Therefore, more number of cell start flipping at higher temperatures and we observe that the percentage error in measurement increases for higher temperatures like $80°C$.

6.3.2 Results for NXP devices

Figure 6.11 to Figure 6.13 show the analysis of the measurements performed for ten NXP devices at different temperatures.

Every measurement of a device and its corresponding measurements are denoted by a colored line in the plots shown in Figure 6.8 to Figure 6.10. The measurement number is plotted on the x-axis and the FHD on the y-axis. The FHD in each case is calculated with respect to the first measurement at a particular ramp-up. Therefore, the y-axis here would denote the fraction of bits that differ between Measurement 1 and the rest of the measurements. Ideally, to be used as a cryptographic key directly, there should be no difference between the consecutive power-ups of the same device. This implies that the FHD should be 0 in the ideal case.

It is observed that the FHD or the percentage error in measurement for twenty devices at $-40°C$ is centered around 5.75%. however, for measurements at $20°C$, the error becomes approximately equal to 6%. As the temperature is further increased, the percentage error increases and becomes around 6.5% for a temperature of $80°C$. From the figures, it is clear that the FHD increases with increase in temperature for all the devices under consideration. In other words, the percentage of bits that flip decreases at lower temperatures.

For various values of temperatures, the observation for NXP devices is similar to that of the TSMC devices and therefore same logic can be applied to explain the results. However, the percentage error differs for NXP and TSMC devices and this can be accredited to the different design and layout of the SRAM for both the companies.

6.4 Validation of the model

In this section, the results obtained from the analytical model will be compared with the results obtained from the measurements performed.

6.4.1 For voltage ramp-up experiment

Figure 6.14 and Figure 6.15 summarize the results obtained from the analysis of the measurements performed on the chip for TSMC and NXP devices respectively. The figures represent the percentage of bits that flip (error) for a given ramp-up on the y-axis and the voltage ramp-up time is shown on the x-axis. The percentage error decreases from 6% at $1\mu s$ to 3% at $10ms$ for TSMC devices whereas for the NXP devices, it decreases from 6.5% to 4%. Both the figures show the same trend and we can conclude that the error percentage decreases with the increase in ramp-up time. In other words, slower ramp-ups allow the SRAM cells to take their preferred state. By inspecting the simulation results obtained from the analytical model as shown in Figure 6.16, we observe a similar trend as in Figure 6.14 and Figure 6.15. The percentage variation that is required to flip a cell decreases from 5% at $1\mu s$ to almost 0% at $100\mu s$. This implies that slower ramp-ups make the cell take their preferred state.

Note that Figure 6.16 presents the results obtained from the analytical model and since the behavior of the actual silicon data matches with that of the analytical model, the correctness of analytical model with regard to voltage ramp-up experiment can be concluded. However, the industrial results represent a statistical scenario wherein all the bits of the 8kB SRAM are considered and hence to compare these results with the simulation results, probabilistic analysis needs to be done.

6.4.2 For Temperature cycle experiment

Figure 6.17 and Figure 6.18 summarize the results obtained from the analysis of the measurements performed for TSMC and NXP devices respectively. The figures represent the percentage of bits that are stable for different temperatures. From these figures, we can conclude that the percentage of stable bits linearly decrease with the increase in temperature. In other words, the noise tolerance decreases at higher temperatures. For TSMC devices, the percentage of stable bits decrease from 95% at $-40°C$ to 94% at $80°C$. For the NXP devices, it decreases from 94% at $-40°C$ to 93% at $80°C$. Both the graphs show a similar linear decreasing trend with increase in temperature.

Figure 6.19 shows the simulation results of our model; it presents the PSNM for different temperatures. It is clear from the figure that the PSNM linearly decreases with the increase in temperature. Reduced PSNM means reduced number of stable bits. Therefore, the trend of the Figure 6.19 is same as that of the Figure 6.17 and Figure 6.18.

Since the behavior of the actual silicon data matches with that of the simulation results from the analytical model, the correctness of analytical model with regard to the temperature behavior can be concluded. However, the industrial results represent a statistical scenario wherein all the bits of the 8kB SRAM are considered and hence to compare these results with the simulation results, probabilistic analysis needs to be done.

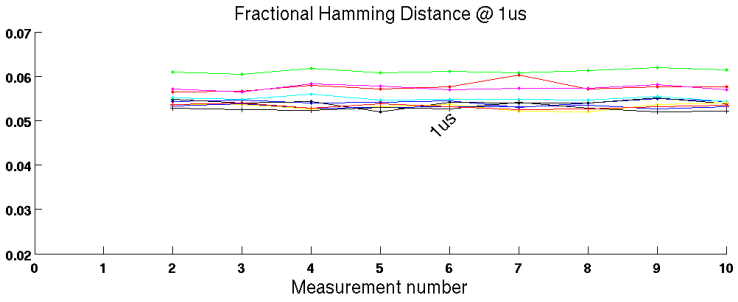

Figure 6.2: Fractional Hamming Distance at $1\mu s$ for ten TSMC devices

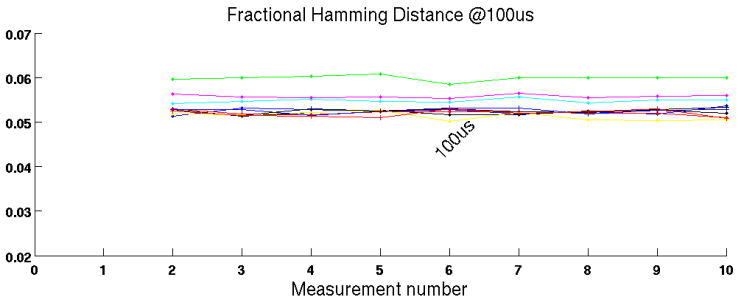

Figure 6.3: Fractional Hamming Distance at $100\mu s$ for ten TSMC devices

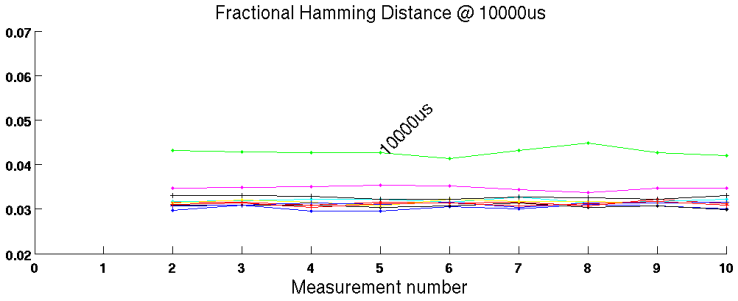

Figure 6.4: Fractional Hamming Distance at $10ms$ for ten TSMC devices

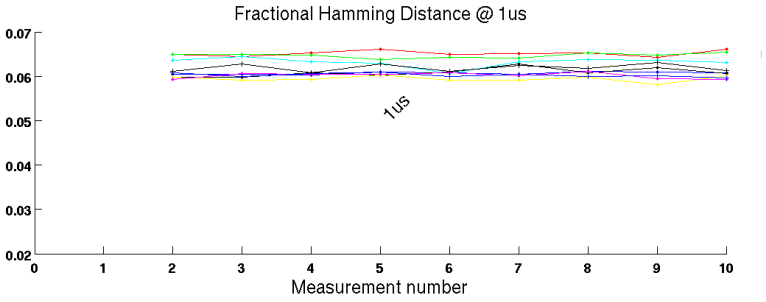

Figure 6.5: Fractional Hamming Distance at $1\mu s$ for ten NXP devices

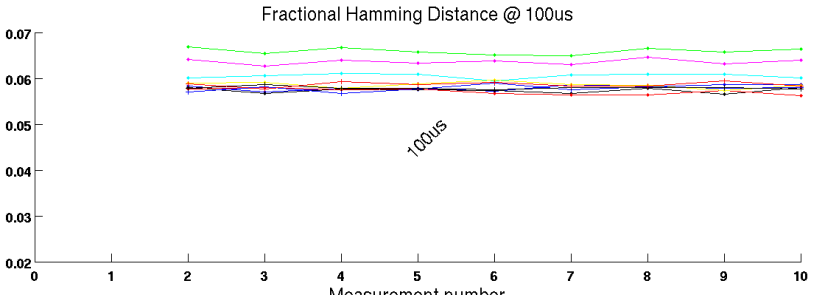

Figure 6.6: Fractional Hamming Distance at $100\mu s$ for ten NXP devices

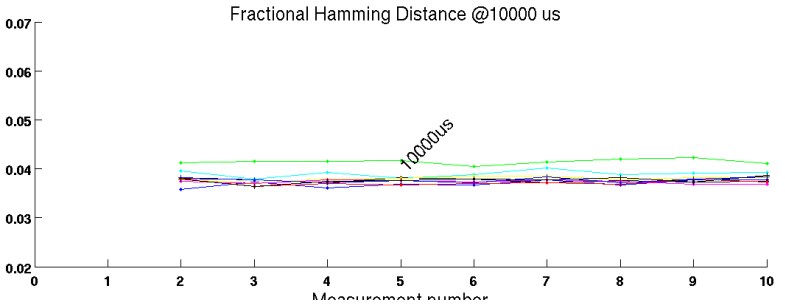

Figure 6.7: Fractional Hamming Distance at $10ms$ for ten NXP devices

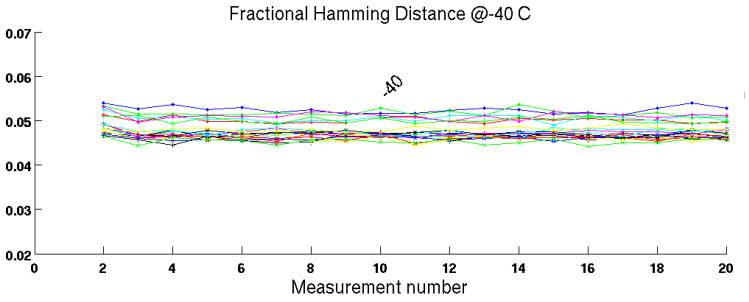

Figure 6.8: Fractional Hamming Distance at $-40°C$ for 20 TSMC devices

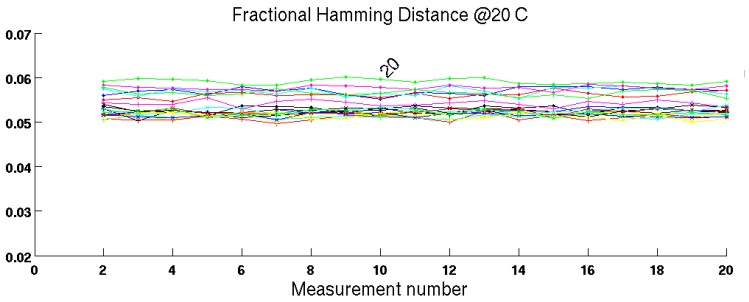

Figure 6.9: Fractional Hamming Distance at $20°C$ for 20 TSMC devices

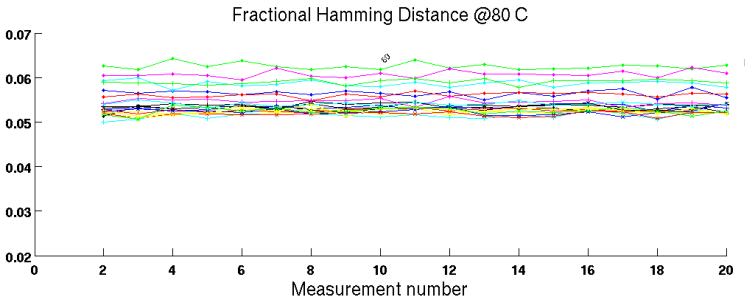

Figure 6.10: Fractional Hamming Distance at $80°C$ for 20 TSMC devices

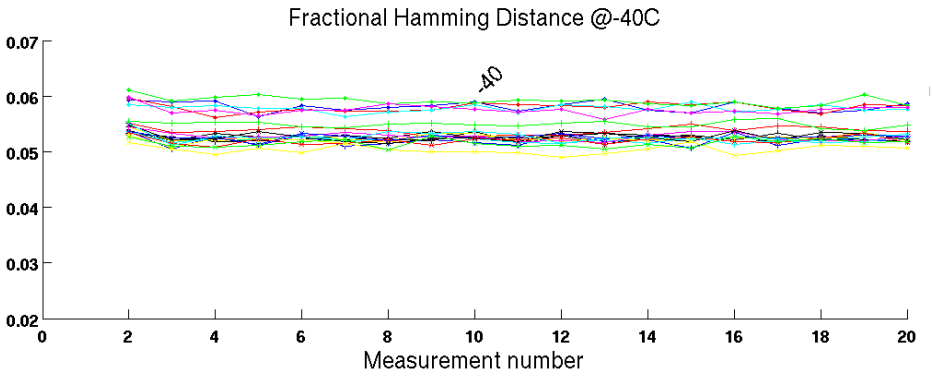

Figure 6.11: Fractional Hamming Distance at $-40°C$ for 20 NXP devices

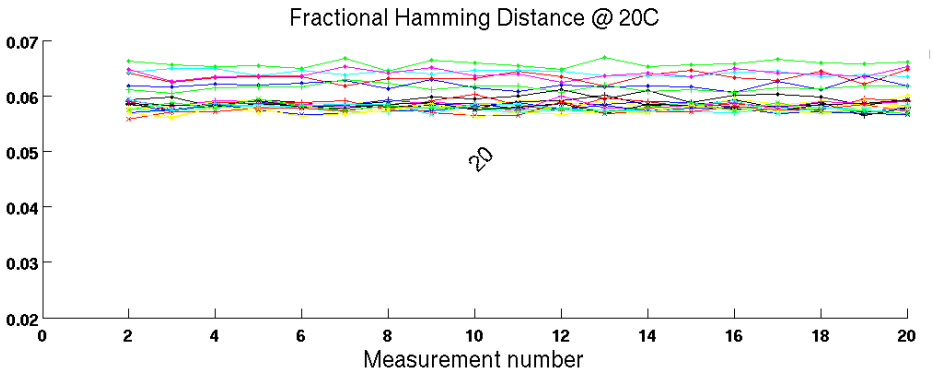

Figure 6.12: Fractional Hamming Distance at $20°C$ for 20 NXP devices

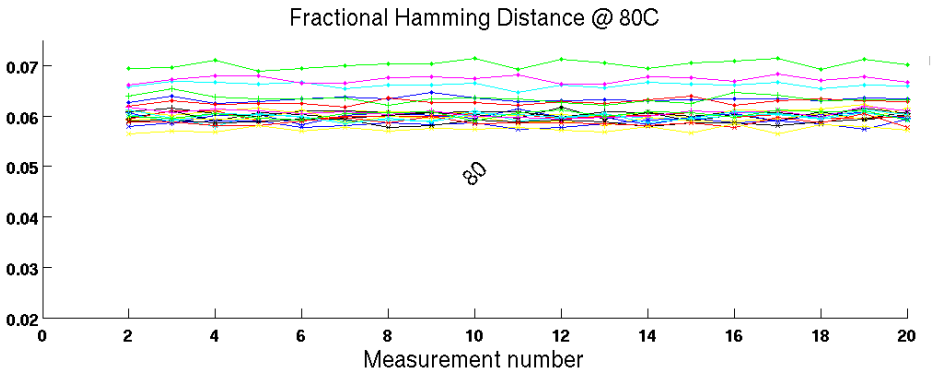

Figure 6.13: Fractional Hamming Distance at $80°C$ for 20 NXP devices

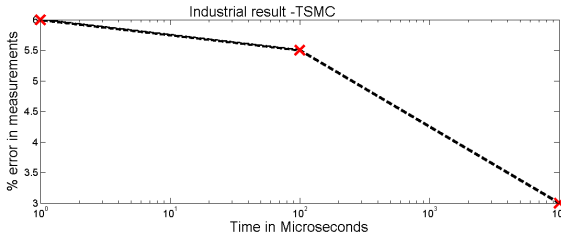

Figure 6.14: Percentage of bits flipping (error) for various voltage ramp-ups for TSMC devices

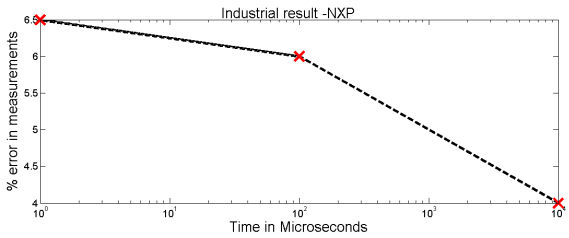

Figure 6.15: Percentage of bits flipping (error) for various voltage ramp-ups for NXP devices

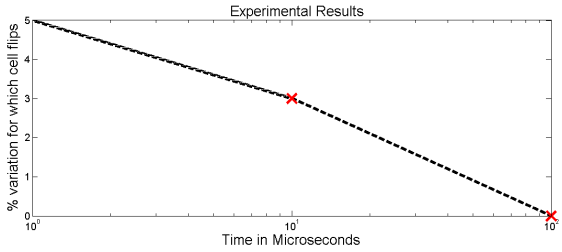

Figure 6.16: Percentage of variation in a length required to make a cell fully-skewed for various voltage ramp-ups

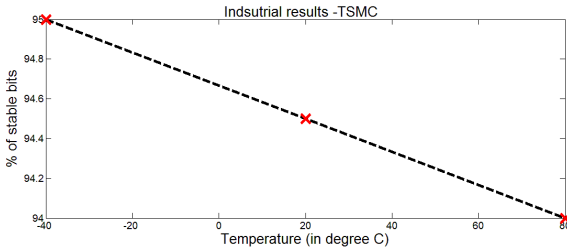

Figure 6.17: Percentage of stable bits for various temperatures for TSMC devices

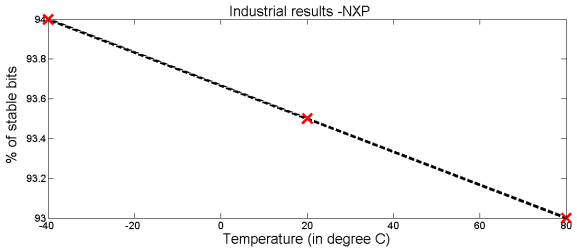

Figure 6.18: Percentage of stable bits for various temperatures for NXP devices

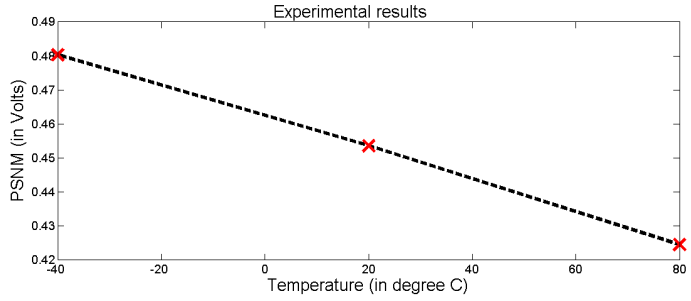

Figure 6.19: PSNM for various temperatures

7 Conclusion & Future Work

This chapter summarizes the results obtained in this work and enumerates few recommendations for future work.

Section 7.1 gives an overview of the conclusions that can be made based on the analysis done in this thesis. Subsequently, Section 7.2 provides a few recommendations for further research.

7.1 Conclusion

In the context of modeling the start-up behavior of SRAM PUF, its reproducibility and its usage as a cryptographic key, this thesis makes the following contributions.

1. A classification of the SRAM cells, based on the amount of mismatch between the two cross-coupled inverters, is proposed. The classification categorizes the cells into three kinds: non-skewed cells, partially-skewed cells and the fully-skewed cells. To be used as cryptographic keys, an ideal scenario would correspond to all the SRAM cells in the memory to be fully-skewed so as to ensure the reproducibility of the key.

2. An analytical model is developed to determine the impact of technology parameters (i.e. threshold voltage, geometry of MOSFETs and channel length modulation factor) and non-technology parameters (i.e. supply voltage and temperature) on the start-up value of the SRAM cell. The start-up behavior is determined based on the calculation of PUF Static Noise Margin.

3. Order of impact of the technology parameters: Threshold voltage > Geometry of the transistors.
   Amongst the technology parameters, threshold voltage variation has the highest impact on the start-up state of an SRAM. A mismatch of only 1.6% between the threshold voltage of two NMOS can make the cell fully-skewed.

4. Amongst the non-technology parameters, it can be concluded that slower voltage ramp-ups and lower temperatures have a higher probability of making the cell take its preferred state. These results are also validated by comparing them with the results obtained from actual silicon data measured at Intrinsic ID.

5. Considering the variation in one MOSFET parameter at a time (threshold voltage of NMOS), it is determined that there is an 84% probability of the cell to be fully-skewed. This ensures the robustness of SRAM PUFs.

7.2 Future Work

Based on the analysis done in this project, few recommendations can be made for further research. Some recommendations are specific to the analytical model. Moreover, few general recommendations can also be made for SRAM PUFs.

1. The analytical model considers only the variation in one parameter in one of the MOSFETs at a time. However, in real scenario, a number of parameters can be influenced by process variation. Although, the variation in threshold voltage will dominate the output as predicted by the analytical model, the model can be extended for variation in more than one parameter.

2. The process/technology node will have an impact on SRAM PUFs. The analytical model developed can be used to explore the impact of technology scaling on start-up value of SRAM cell. It is expected that factors like channel length modulation will play an important role for smaller technology nodes.

3. Entropy of the SRAM PUFs can be analyzed to get more insight into their uniqueness property.

Bibliography

[1] *http://ptm.asu.edu/*, 2011.

[2] *http://www.intrinsic-id.com/*, 2011.

[3] A. Agarwal, B.C. Paul, S. Mukhopadhyay, and K. Roy, *Process variation in embedded memories: failure analysis and variation aware architecture*, IEEE Journal of Solid-State Circuits **40** (2005), no. 9, 1804– 1814.

[4] A. Asenov, S. Kaya, and A.R. Brown, *Intrinsic parameter fluctuations in decananometer mosfets introduced by gate line edge roughness*, IEEE Transactions on Electron Devices **50** (2003), no. 5, 1254–1260.

[5] A. Balijepalli, S. Sinha, and Y. Cao, *Compact modeling of carbon nanotube transistor for early stage process-design exploration*, IEEE International Symposium on Low Power Electronics and Design, 2007, pp. 2–7.

[6] A. Bhavnagarwala, X. Tang, and J. Meindl, *The impact of intrinsic device fluctuations on cmos sram cell stability*, IEEE Journal of Solid-State Circuits **36** (2001), no. 4, 658–665.

[7] B.H. Calhoun and A. Chandrakasan, *Static noise margin variation for sub-threshold sram in 65-nm cmos*, IEEE Journal of Solid-State Circuits **41** (2006), no. 7, 1673–1679.

[8] B.H. Calhoun and A.P. Chandrakasan, *Analyzing static noise margin for sub-threshold sram in 65nm cmos*, European Solid-State Circuits Conference, 2005, pp. 363–366.

[9] Y. Cao, T. Sato, D. Sylvester, M. Orshansky, and C. Hu, *New paradigm of predictive mosfet and interconnect modeling for early circuit design*, Custom Integrated Circuits Conference,, 2000, pp. 201–204.

[10] Y. Cheng and C. Hu, *Mosfet modeling and bsim3 user's guide*, Kluwer Academic Publishers, 1999.

[11] A.J. Cover, M.T.; Thomas, *Elements of information theory*, Wiley, 1991.

[12] S. Dimitirjev, *Understanding semiconductor devices*, Oxford University Press, 2000.

[13] K. Kuhn et al, *Managing process variation in intel's 45nm cmos technology*, Intel Technology Journal **12** (2008), no. 2, 92–110.

[14] P. P. Fastykovsky and A. A. Mogilnitsky, *Effect of air humidity on the metal-oxide-semiconductor tunnel structures' capacitance*, Sensors and Actuators B: Chemical **57** (1999), no. 1-3, 51–55.

[15] F.H. Gebara, *Temperature-profiled device fingerprint generation and authentication from power-up states of static cells*, 2009.

[16] E. Gogolides, V. Constantoudis, G. Patsis, and A.i Tserep, *A review of line edge roughness and surface nanotexture resulting from patterning processes*, Microelectronic Engineering **83** (2006), no. 4-9, 1067–1072.

[17] W.P. Holcomb, D.E.; Burleson and K. Fu, *Power-up sram state as an identifying fingerprint and source of true random numbers*, IEEE Transactions on Computers **58** (2009), no. 9, 1198–1210.

[18] A. Kerckhoffs, *La cryptographie militaire*, Journal des sciences militaires **9** (1883), 161–191.

[19] H-W. et al Kim, *Experimental investigation of the impact of lwr on sub-100-nm device performance*, IEEE Transactions on Electron Devices **51** (2004), no. 12, 1984–1988.

[20] S.S. Kumar, J. Guajardo, R. Maes, G.-J. Schrijen, and P. Tuyls, *Extended abstract: The butterfly puf protecting ip on every fpga*, IEEE International workshop on Hardware-Oriented Security and Trust, 2008, pp. 67–70.

[21] F. Lai and C. Lee, *On-chip voltage down converter to improve sram read-write margin and static power for sub-nano cmos technology*, IEEE Journal of Solid-State Circuits **42** (2007), no. 9, 2061–2070.

[22] R. Maes, *http://homes.esat.kuleuven.be/ rmaes/puf.html*, 2011.

[23] R. Maes, P. Tuyls, and I. Verbauwhede, *A soft decision helper data algorithm for sram pufs*, IEEE International Symposium on Information Theory, 2009, pp. 2101–2105.

[24] D. Neamen, *An introduction to semiconductor devices*, McGraw-Hill, 2006.

[25] C. Paar and J. Pelzl, *Understanding cryptography*, Springer, 2010.

[26] R.S. Pappu, *Physical one-way functions*, Ph.D. thesis, Massachusetts Institute of Technology, 2001.

[27] A. Pavlov and M. Sachdev, *Cmos sram circuit design and parametric test in nano-scaled technologies*, Springer, 2008.

[28] J.M. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital integrated circuits*, Prentice-Hall, 2002.

[29] E. Seevinck, F.J. List, and J. Lohstroh, *Static-noise margin analysis of mos sram cells*, IEEE Journal of Solid-State Circuits **22** (1987), no. 5, 748–754.

[30] G. N. Silva and A. Chandrakasan, *Leakage in nanometer cmos technologies*, Springer, 2006.

[31] S.P. Skorobogatov, *Semi-invasive attacks- a new approach to hardware security analysis*, Tech. report, University of Cambridge, 2005.

[32] F. Stellari, P. Song, A.J. Weger, and D.L. Miles, *Mapping systematic and random process variations using light emission from off-state leakage*, IEEE International Symposium on Reliability Physics (2009), 640–649.

[33] P.A. Stolk, F.P. Widdershoven, and D.B.M. Klaassen, *Modeling statistical dopant fluctuations in mos transistors*, IEEE Transactions on Electron Devices **45** (1998), no. 9, 1960–1971.

[34] G.E. Suh and S. Devadas, *Physical unclonable functions for device authentication and secret key generation*, Design Automation Conference, 2007, pp. 9–14.

[35] K. Takeuchi, T. Fukai, T. Tsunomura, A.T. Putra, A. Nishida, S. Kamohara, and T. Hiramoto, *Understanding random threshold voltage fluctuation by comparing multiple fabs and technologies*, IEEE International Electron Devices Meeting, 2007, pp. 467–470.

[36] P. Tuyls, *Security with noisy data*, Springer, 2007.

[37] P. Tuyls and L. Batina, *Unclonable rfid-tags*, 2011.

[38] Jaikumar Vijayan, *http://www.computerworld.com/s/article/9014782/*, 2007.

[39] Bart J. Van Zeghbroeck, *http://ecee.colorado.edu/ bart/book/flatband.htm*.

[40] W. Zhao and Y. Cao, *New generation of predictive technology model for sub-45nm early design exploration*, International Symposium on Quality Electronic Design, 2006, pp. 585–590.

[41] W. et al Zhao, *Rigorous extraction of process variations for 65nm cmos design*, European Solid State Device Research Conference, 2007, pp. 89–92.