Low Power SRAM-PUF with Improved Reliability & Uniformity Utilizing Aging Impact for Security Improvement

ACHIRANSHU GARG

School of Electrical & Electronic Engineering

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

Master of Engineering

2013

Acknowledgements

The research work presented in this thesis is result of a roller-coaster ride of two years in world of research. Like a roller-coaster ride with its exhilarating journey ending in a jubilant finish, research also has its ups and downs. I did not take me much time to realise that this work could be materialised, mainly due to the immense support and guidance of a number of people.

First, my sincere gratitude towards my supervisor Asst. Prof. Kim Tae Hyoung, Tony for reposing trust in me, willing to introduce me to a variety of multidisciplinary works and guiding me throughout my research. His friendly attitude has always been an important factor while working with him. This work is result of his constant guidance and encouragement.

I would like to thanks my fellow group mates who helped me learn and make progress every time I was in need.

I would also like to thank the Economic Development Board (EDB), Singapore for providing me with the IC design Postgraduate Scholarship (ICPS) so as to fund my study at Nanyang Technological University.

I whole-heartedly thank my parents for always supporting me in my endeavours.

Last but not the least; I sincerely appreciate my NTU friends for hearing out my ramblings, guiding me, inspiring me & helping me editing my manuscript.

Table of Contents

Abstract 1

Acknowledgements i

List of Figures v

List of Tables xi

Outline xii

Chapter 1: Introduction 1

1.1 Background 1

1.2 Information Security 5

1.3 Memories based Security System 6

1.4 Challenge-Response Pair 7

1.5 Physical Unclonable Function (PUF) 7

1.5.1 Classification of PUF 8

Chapter 2: Literature Review 16

2.1 SRAM PUF and associated advantages over other PUF 16

2.2 Previous SRAM-PUF Work 17

2.3 Security Parameters 22

2.3.1 Uniqueness 22

2.3.2 Reliability 23

2.3.3 Uniformity 23

2.4 Energy Efficiency dependence on SRAM Array Structure 26

2.4.1 Supply Voltage (VDD) Scaling Effect on SRAM- 26

2.4.2 Sub-array based SRAM architecture 27

2.4.3 Hierarchical Word Line (WL) & Bit Line (BL) 28

2.4.4 Self-timing 29

2.4.5 Data Retention Voltage 30

2.4.6 SRAM array structure influence on energy minimization 31

Chapter 3: Implementation & Simulation Results 33

Section I: SRAM-PUF 33

3.1 SRAM Operations & Power-up Value Based SRAM PUF 33

3.1.1 Read operation 34

3.1.2 Write operation 34

3.1.3 Partially skewed 35

3.1.4 Fully-skewed 35

3.2 SRAM-PUF power-up variations due to environmental fluctuations (Monte Carlo Simulations) 37

3.3 Negative Bias Temperature Instability (NBTI) 42

3.3.1 Impact on Threshold voltage (V_{TH}) 42

3.3.2 Impact on Static Noise Margin (SNM) curve 44

3.4 Optimum Uniformity Methodology 47

3.5 Reliability (Skew) Improvement Methodology 49

3.6 Cell flipping setup 51

3.6.1 SRAM Cell flip output 52

3.7 Proposed reliability and uniformity improvement methodology 54

3.8 Cell Flipping due to NBTI aging effects 55

3.9 Layout using 65nm Global foundry technology library 60

Section II: Proposed SRAM Energy Minimization Methodology 62

3.10 SRAM Energy Model 62

3.12 Energy Efficiency Dependency on SRAM Array Structure. 66

3.13 Impact of Device Variations on SRAM Array Structures for Energy Minimization... 70

Chapter 4: Summary & Future Work 74

4.1 Summary 74

4.2 Future Work 76

References 77

Summary

Physical Unclonable Functions (PUFs) are the latest secure key generation circuits that are analogous to human DNA. Just like each human has a different DNA map that persists throughout his life, silicon devices also exhibit unique and reproducible patterns based on intrinsic properties of silicon. These can be signified as the signature of that device. Currently various kinds of PUFs are under development, namely - Optical PUF, Coating PUF, Delay based PUF, Butterfly PUF. But presently, the most popular and reliable amongst all is the SRAM-PUF. SRAMs are an integral part of many System-on-Chip (SoC) designs. Using them to incorporate security features does not affect area overhead much. Moreover once the security functionality is completed the same SRAM can be used for storage as well.

Hardware Intrinsic Security (HIS) is currently a very crucial aspect of the electronics industry aimed towards protecting hardware IPs from infringement. Also, Wireless Sensor Nodes (WSNs) that are increasingly acting as a backbone to the information channels need a secure and low-power encryption system to protect them from malicious attacks. Traditional electronic devices store encrypted keys in battery-powered volatile memories or use Non-Volatile Memories for permanent storage of security keys. These are not very secure since the security key is exposed and relatively easy to hack from Non-Volatile Memory (NVM). Additionally, it comparatively consumes more power for operating over a long period of time. Thus, SRAM-PUF can provide a viable solution to both the problems - secure encryption & minimal power consumption.

SRAM-PUF makes use of inefficient silicon fabrication process (diffraction of usable light for lithography masks where device dimensions are already reaching wavelength of same light), due to which the production of exact replica devices from the same design is difficult.

If the same SRAM design is used for fabrication of two devices, it would lead to an uncertainty in their physical properties. Doping and channel length ambiguity gives rise to threshold voltage variation in transistors. These unique physical properties variation give rise to a unique SRAM power-up pattern. This pattern can be used as a security key which can be generated only when required; thus giving very less time to any hacker for tampering. Additionally, no constant battery power is required, thus making the system more power efficient.

One of the major issues with SRAM PUF is the variations of power-up pattern with environmental fluctuations. Our aim is to design and develop low power SRAM-PUF with a uniform output (distribution of 1's & 0's) and minimal power-up variations utilising the aging effects (mainly NBTI) to make it more reliable and secure. Negative Bias Temperature Instability (NBTI) is considered as a disadvantage since it leads to skew in SRAM-cells. We plan to provide a desirable skew to SRAM-cells so that each time they power-up, a reliable bit-pattern is generated. Each cell then produces the same bit at every subsequent power-up ensuring minimized Hamming Distance for two different power-ups. Also, the number of 1’s and 0’s in the final output pattern should be equalized to maintain maximum uniformity. The aim of this thesis is to make use of two negative factors (process variation & aging) in a positive way to our advantage and make a secure and reliable key generator.

WSNs are majorly installed in inaccessible locations, thus regular change of battery is not practically possible. There is a need for more energy efficient WSN to ensure battery longevity. A lot of work has been done previously on various SRAM energy optimizations at the circuit level. Hence in this thesis we investigate, a new aspect of SRAM energy minimization, the role of array structures in determining the total energy for SRAMs operating near sub-threshold voltages. Previous research on array structures shows that taller array structures are more energy efficient as associated capacitance with precharge devices

would be less for taller array structures at higher supply voltages. We found that, in contrast to prevalent hypothesis - fat array structures (fewer rows, more columns) are more energy efficient than tall array structures (more rows, less columns) near sub-threshold operating voltage region as static energy plays an important role in determining total energy. The static energy of any SRAM array structure depends on leakage current & read latency for any fixed supply voltage. Our analysis reveals that read delay (which is less in case of fat array) affects the static energy significantly. Energy efficiency can be improved up to 38% (64kb), 10% (8kb) by using fat array structures in SRAM compared to tall arrays. Statistical analysis also reveals that wide array structure should be preferred over tall array structure at sub-threshold voltage operation ensuring less read failures.

List of Figures

Figure 1: Dubious chips as reported by ERAI [1] 1

Figure 2: An example of counterfeit chips traced in US [1] 3

Figure 3: Fundamentals of memory based security system 6

Figure 4: Challenge-Response pair definition for PUF 8

Figure 5: Various type of PUF 9

Figure 6: Optical PUF[9] 10

Figure 7: Coating PUF [9] 11

Figure 8: Arbiter PUF [9] 12

Figure 9: Ring-oscillator PUF[9] 13

Figure 10: SRAM PUF [9] 14

Figure 11: Butterfly PUF[9] 15

Figure 12: Segregation of UF (Useful) and NUF (Not Useful) Cells 18

Figure 13: SRAM cell with additional voltage source and current source 19

Figure 14: MECCA PUF, Memory block with peripheral circuitry and programmable delay circuit 20

Figure 15: (a) block diagram of circuit (b) flow of ID-generation. 21

Figure 16: Parameters of PUF measurement 24

Figure 17: SRAM array divided into sub-arrays with same density 28

Figure 18: Hierarchical word-line (WL) scheme 28

Figure 20: Retention scheme for saving leakage energy 30

Figure 21: SRAM with same density with different array configuration 32

Figure 22: (a) SRAM array (b) Symmetrical 6T cell 33

Figure 23: Hamming distance Vs Temperature 38

Figure 24: Hamming distance Vs Supply ramp-up time 39

Figure 25: Hamming distance Vs Supply voltage 39

Figure 26: Hamming distance Vs Seed variation 40

Figure 27: Hamming Distance variations with temperature and ramp-up time. 41

Figure 28: Schematic description showing the generation of interface traps when a PMOS transistor is biased in inversion[33] 43

Figure 29: (a) Statistical output behaviour of 1-skewed cell (b) A partially-skewed cell which can sway to any direction under influence of noise [31] 45

Figure 30: (a) A partially-skewed cell (b) 0-skewed cell [10] 45

Figure 31: Aging impact on SRAM cell 47

Figure 32: Proposed methodology to improve Uniformity...48

Figure 33: Skew improvement methodology...49

Figure 34: Proposed technique to improve Reliability 50

Figure 36: Cell flipping output 52

In the next section, the proposed methodology is evaluated using statistical simulations for a SRAM cell array to see the impact on uniformity and reliability of SRAM-PUF. 55

Figure 39: Cell flipping due to increased V_{TH} 56

Figure 40: SRAM cell setup for statistical simulation..57

Figure 41: Impact of aging in maintaining Uniformity 58

Figure 42: Impact of aging on reliability 59

Figure 43: SRAM PUF testchip layout...60

Figure 45: (a) An 8kb SRAM sub-array for energy analysis. Note that k X j is 8kb. (b) Bitline structure of the SRAM sub-array in (a). (c) Schematic of the conventional 8T SRAM cell used in this work 63

Figure 46: SRAM modelling for energy estimation 64

Figure 49: Percentage change in the energy using optimal rows over 128 rows.....................68

Figure 50: Read delay variation with rows at various supply voltages 69

Figure 50: Statistical distribution of Total Energy at VDD = 0.4V 71

Figure 51: Statistical distribution of Total Energy at VDD = 0.6V 71

Figure 52: Statistical distribution of Total Energy at VDD = 1.2V 72

Figure 53: Corner Simulations for SRAM 73

List of Tables

Table 1: SRAM cell skew 36

Table 2: Impact of increased threshold Voltage on output pattern 58

Table 3: Impact of aging on cell reliability. 59

Table 4: Optimized energy structure configuration(s) Vs Supply voltage-. 67

Outline

Chapter 1 gives a brief Introduction about the necessity of PUF, SRAM-PUF and Energy efficiency enhancement in SRAMs using different array structures.

Chapter 2 entails the Literature Review of previous work done on PUF mentioning the expectations from an Ideal SRAM-PUF which lays the foundation of our work. Also, it contains a brief idea about energy efficiency dependency on SRAM array structures.

Chapter 3 contains Implementation and Simulation results divided in two sections. Section I shows our implementation strategy for ideal SRAM-PUF. Hamming distance variations due to different environmental conditions show the necessity of a robust SRAM-PUF. Also, it contains the result of SRAM cell flipping setup. Section II contains the implementation details for SRAM energy modelling and simulation results proving the point those wider arrays are more energy efficient for smaller supply voltages. It also contains the statistical analysis for SRAM array structure based energy minimization.

Chapter 4 includes the Summary & Future work.

Chapter 1: Introduction

1.1 Background

Nowadays, electronic devices are increasingly becoming an indispensable part of our lives. From ATM cards, to credit cards, to access cards, to military equipment - a lot of confidential information is being stored & handled. Also, since physical IP design companies who outsource their manufacturing want to prevent their IPs from infringement, an inexpensive solution for the security of these devices is a big challenge for designers today.

Electronics Resellers Association International (ERAI) tracked the counterfeit electronics over a period of 5 years and reported data as in Figure 1. The data reveals the number of counterfeit cases that are increasing every year. Hence these need to be checked as semiconductor chips are becoming integral parts of any system. The failure of counterfeit chips at critical places could have very dangerous consequences.

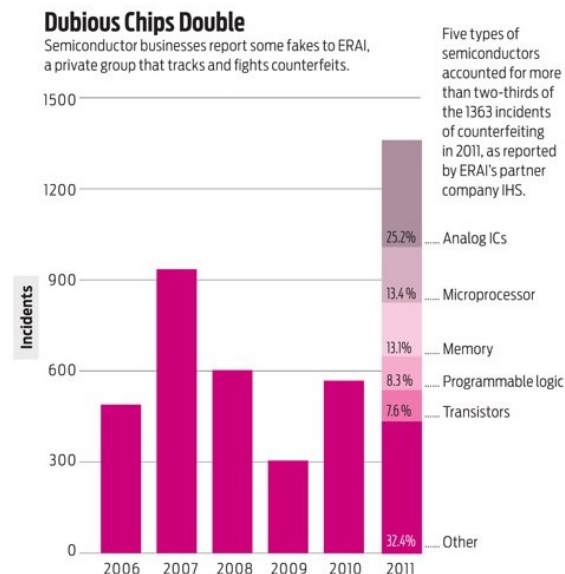

Figure 1: Dubious chips as reported by ERAI [1]

Digital devices or IPs need a robust security mechanism, which should ideally consume low power & area. Conventional systems (e.g. RFID tags, smart cards) use Non-Volatile Memory (NVM) based security system in which a binary encrypted key is stored and authenticated each time to access stored secret information. With the development of new (invasive & non-invasive) tampering methods such as micro-probing, laser cutting, glitch attacks and power analysis it is possible for attackers to steal the binary key. To prevent such physical attacks on ICs, researchers developed a tamper-sensing method in which a sensor mesh is used to detect any tampering with the IC. The limitation of the sensor-mesh is that it cannot detect intrusion when circuit power is off and the hardwired information can be stolen without much difficulty [2]. Thus, HIS is an area of much interest to various researchers for improvement of hardware security.

Conventionally, security key is stored in Non Volatile Memory (NVM) in form of fuses as in Electrically Erasable Programmable Read-Only Memory (EEPROM). But the conventional approach has a shortcoming i.e. the difficulty and expenses to safely manage the security keys as they need to be stored all the time. Also, to provide secure tampering sensing circuitry is expensive in terms of resources on already constrained chip e.g. – RFID chip [3].

Due to technological constraints and financial reasons various big semiconductor companies are also finding it difficult to stop counterfeiting. Figure 2 talks about the number of counterfeit chips (including commercial gear falsely labelled as military grade) from big companies like Intel, Motorola, Cypress, Altera and National Semiconductors used in US military equipments during a year. The criticality of these chips in military products can be gauged from the usage of these components as described in Figure 2. The failure of a counterfeit chip can fail the whole product, which is highly unacceptable.

The US government realized the seriousness of problem and passed a new legislation, the National Defence Authorisation Act to check dubious chips, which holds responsible suppliers for accountability and replacement cost.

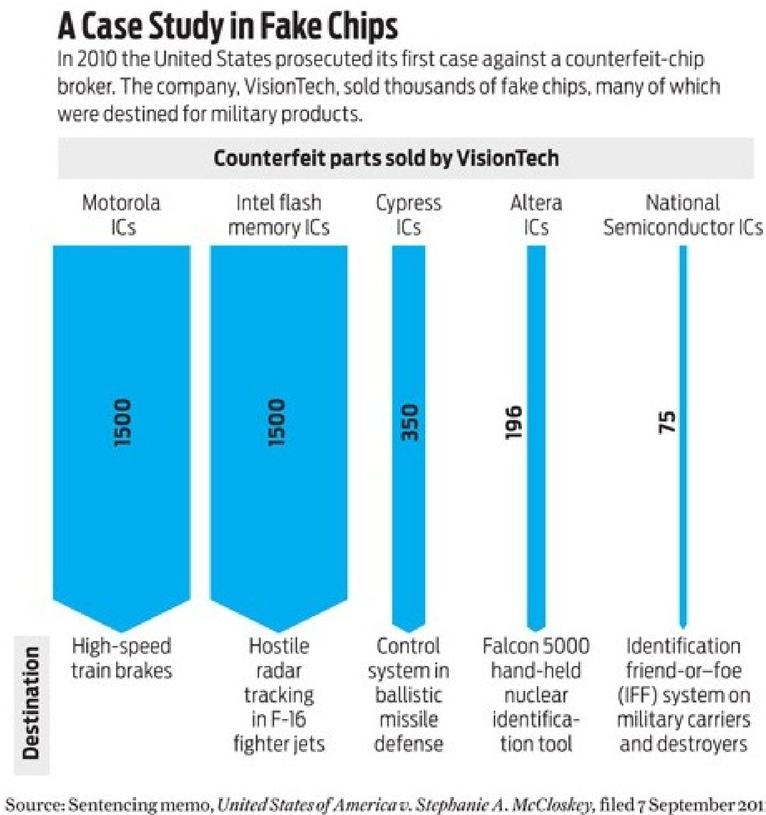

Figure 2: An example of counterfeit chips traced in US [1]

In addition to stringent laws, need of the hour is to design a cost effective security solution which can help provide unbreakable security with minimal area overhead to prevent counterfeiting.

Physical Unclonable Functions is the answer to all these problems. SRAM-PUF is a breakthrough technology in which a secret key is generated using the SRAM. It is secure and unclonable as no two devices can generate the same secret key due to different physical

characteristics. If a manufacturer himself wants to replicate the key, it is extremely difficult and expensive as the keys are random and cannot be controlled.

To ensure the high strength of security key, it should have high uniformity (equal number of 1's & 0's). Also generating same pattern at every power-up is crucial to obtain same security key every time. Due to the initial V_T mismatch the SRAM cells are expected to show skewed behaviour, which means they will show 0 or 1 at power-up of SRAM. We are trying to make use of the aging (NBTI stress) on SRAM cells to maximize the uniformity and reliability so that it gives a very secure, consistent power-up pattern every time.

In addition to the security feature, a SRAM-PUF needs to be energy efficient for applications like WSNs which operate for short time-periods and near sub-threshold region. The SRAMs play a key role in energy consumption due to the high cell density required for computations. In the sub-threshold region there are many design constraints which necessitates the use of innovative energy minimization methods. Various energy minimization circuit techniques were proposed by researchers over the years. Many of these techniques are used in contemporary SRAM designs such as sub-arrays with bussed address lines, divided word line architecture, hierarchical word decoder architecture and self-timing to tackle timing variations. The optimization techniques of SRAM timing can also be seen as energy minimization since their aim is to reduce the capacitance as well.

SRAM array structure also plays role in deciding total energy. From simulations it is observed that the SRAM array structure plays an important role in improving the energy efficiency up to 38% (64kb), 10% (8kb) for the fixed density SRAM at same operating voltage by just changing the array structure from tall structure to wide array structure. The reason for this can be attributed to the fact that at sub-threshold operating voltage the static energy becomes an important part of the total energy and wide array structures shows less read latency (proportional to static energy) due to less number of bit-cells per bitline resulting

in less capacitance per bitline. Since, sub-90nm CMOS technologies have considerable leakage current which makes leakage energy a deciding factor in total energy, a minimization of leakage energy is required. Changing SRAM array structure to a wider configuration (as quoted previously) can have considerable energy savings, also which improves with more dense SRAMs. In addition to it, statistical simulation reveals that wider array structures have less read failures compared to taller structures.

1.2 Information Security

As Wikipedia states "information security means protecting information and **information systems** from unauthorized access, use, disclosure, disruption, modification, perusal, inspection, recording or destruction" [4]. To ensure that the Information is secured we should have robust security systems.

A very fundamental Kerchoffs' principle regarding security systems states: "A system should be secure even if everything about the system, except the key, is public knowledge" [5].

A general view (Figure 3) of key-based security system will help in visualizing the role of security systems in protecting electronic devices.

1.3 Memories based Security System

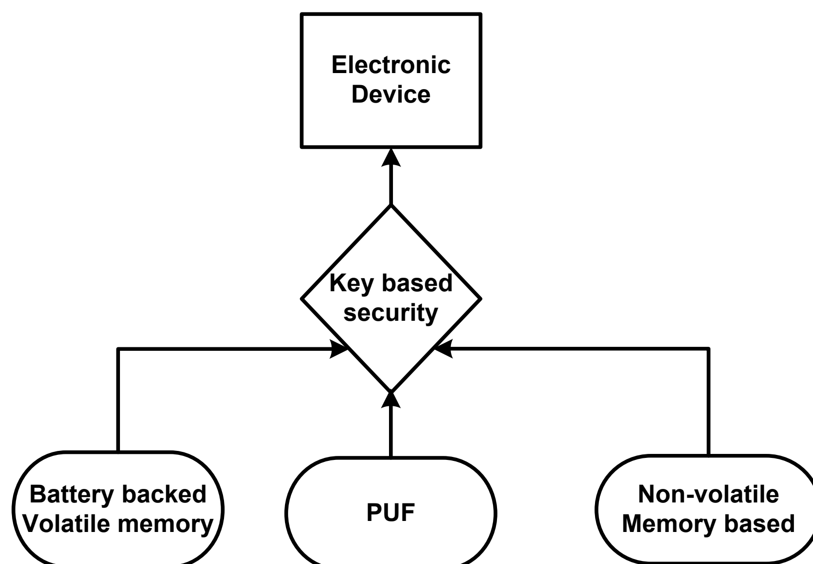

Figure 3: Fundamentals of memory based security system

Interpreting Kerchoffs' principle, the most important thing in any security based system is the security key which should not be visible to anyone at any cost. So SRAM-PUF fits best in this definition as key is only generated when it is required and scores above NVM based security systems wherein the key is stored permanently and exposed to potential hackers without any proof of tampering [6].

A PUF is basically expected to give a response at various input excitations. It is very difficult to read the contents of SRAM when it is not powered-up, also the tampering process is costly and time consuming during which infringement can be traced.

The basic principle behind the excitation based PUF operation is Challenge Response Pair (CRP).

1.4 Challenge-Response Pair

In [6], Pappu described the idea of "Physical One Way Functions (POWF)" using challenge-response pair criterion under which an object is subjected to a large number of challenges and it produces unique output corresponding to every challenge, called as response.

He laid down four fundamental requirements for an ideal authentication system [7].

1. Easy to fabricate - The security system should be easy and inexpensive to fabricate. In real world, the security system will be employed in large numbers and it should be practically feasible to produce in large numbers security key based system inexpensively.

2. Easy to probe - The system should have a simple and easy probe setup to obtain the output without many complications otherwise it will hinder the practical usability of authentication system by increasing the cost of reader.

3. Hard to clone - The authentication system should be such that it is difficult to re-fabricate a clone of same device. The requirements of mass production and hard to clone when combined together can be interpreted as producing devices in large amount from same design but no devices should produce same token or security key.

4. Structurally stable - The structure of authentication system should be physically stable. It should also be able to handle environmental variations over a long period of time.

1.5 Physical Unclonable Function (PUF)

Gassend et. al.[8] defines Physical Unclonable Function as a physical function which produces a set of responses from a set of input challenges based on complex untraceable

physical interactions between the physical system and challenge inputs. In simpler terms, it can be a method of producing fingerprints of a physical object based on its manufacturing variations.

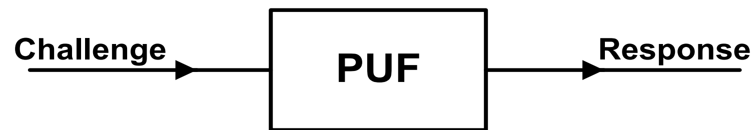

Figure 4: Challenge-Response pair definition for PUF

Physical Unclonable Function can be explained as [9]:

1. Physical - A PUF is a physical embedded system. By physical system we mean it cannot be any mathematical function but the function outcomes are generated only after the physical interactions.

2. Unclonable - An unclonable function is such that it is difficult to make a clone function which gives the same output when given the same input.

3. Function - It is a kind of function but not a mathematical function in which we call input as challenge and output as response, which follows a relation based on device properties.

1.5.1 Classification of PUF

PUFs can be divided into two major categories based on their random physical characteristics.

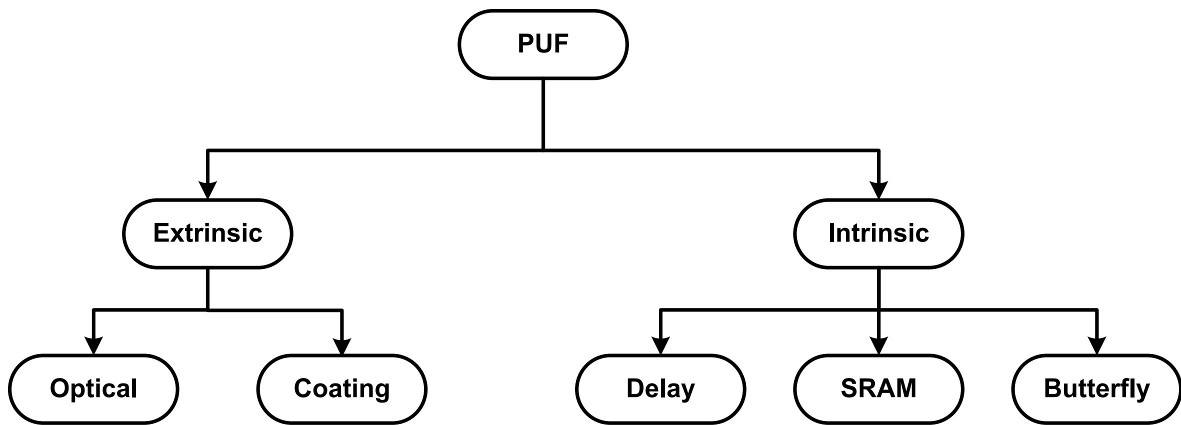

Figure 5: Various type of PUF

(A) Extrinsic PUF

These kinds of PUFs are called extrinsic PUFs because the manufacturer can introduce the randomness in these devices by controlling the parameters of disturbance but the distribution of output still remains random. Therefore, the process is still random and output is unpredictable. The advantage of these devices is that you can optimize and control the extrinsic disturbance and parameters which improves the distinguish-ability between various PUFs. The two extrinsic PUFs are explained as below [9].

1. Optical PUF
2. Coating PUF

Optical PUFs

These were originally proposed by Pappu [7] and consist of a transparent medium (such as glass or plastic) doped with light scattering particles. When coherent lasers beam strikes the transparent medium it generates speckle pattern which is dependent on many factors including wavelength of striking beam and angle of incidence but the most important is the way random particles scatter the light beam. The interaction of

laser beam with random particles is very complex and difficult to replicate for another similar PUF instantiation. Thus, optical PUF are physically unclonable as such.

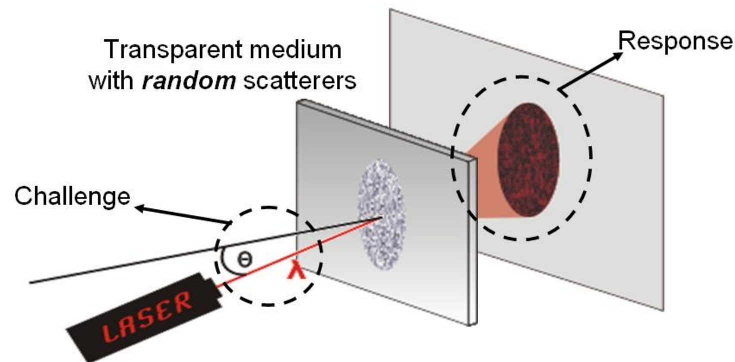

Figure 6: Optical PUF[9]

Also, the difficulty in prediction of complex interaction behaviour makes them mathematically unclonable. Thus for optical PUF, *set of input parameters* is the **challenge** and resulting *speckle pattern* is **response**.

Coating PUF

As the optical PUF was a separate system, coating PUF was conceptualized to be a part of silicon chip itself. The idea was to spray a protective coating of particles of randomly distributed size and dielectric constant on the chip. Below this protective coating, is a metal layer containing the comb shaped sensors to measure the associated capacitance of that particular part of coating. As the size and dielectric of particles are random, so replicating the exact pattern is difficult for similar kind of PUF.

Thus selecting a particular particle amongst various *random particles* becomes a **challenge** for PUF and *capacitance value* gives the **response** value.

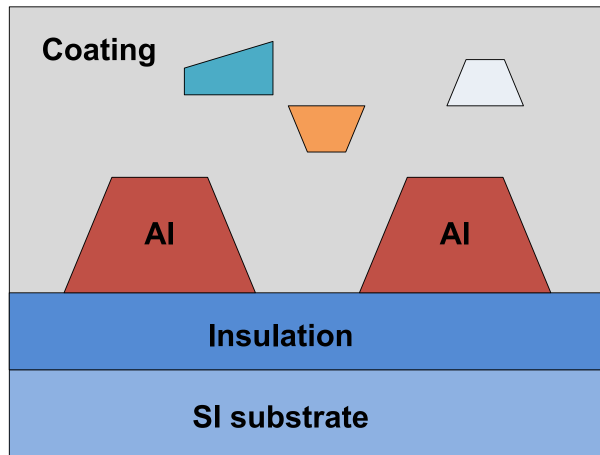

Figure 7: Coating PUF [9]

(B) Intrinsic PUF

As the name suggests, Intrinsic PUFs are based on internal randomness of silicon devices. The Intrinsic randomness is introduced in these devices due to process variations during manufacturing process. Process variations are attributed to ambiguities in the lengths, widths and oxide thicknesses of silicon devices at Nano-meter geometries. Thus, it is the inefficiency of the silicon fabrication process which doesn't allow fabrication of two replica devices irrespective of same mask design. The major advantage of Intrinsic PUF is that it gives digital output so the need to quantization also gets removed. The three kinds of PUFs are [9]:

1. Delay based PUF
2. SRAM PUF
3. Butterfly PUF

Delay based PUF

Delay based PUF utilizes the intrinsic process variation that results in random variations in gate and interconnects delays. The circuit is provided with a challenge and the dedicated delay measurement circuitry measures the input-output through various paths.

Arbiter PUF is one example of delay based PUF, as shown in Figure 8 two symmetrical paths are equally placed in a chip and triggered from a common input i.e. since input source is same, the input mismatch between the two signals is zero, so the mismatch we observe at the output is solely due to the mismatch between two interconnects or the gates in two paths without any difference in overall functionality.

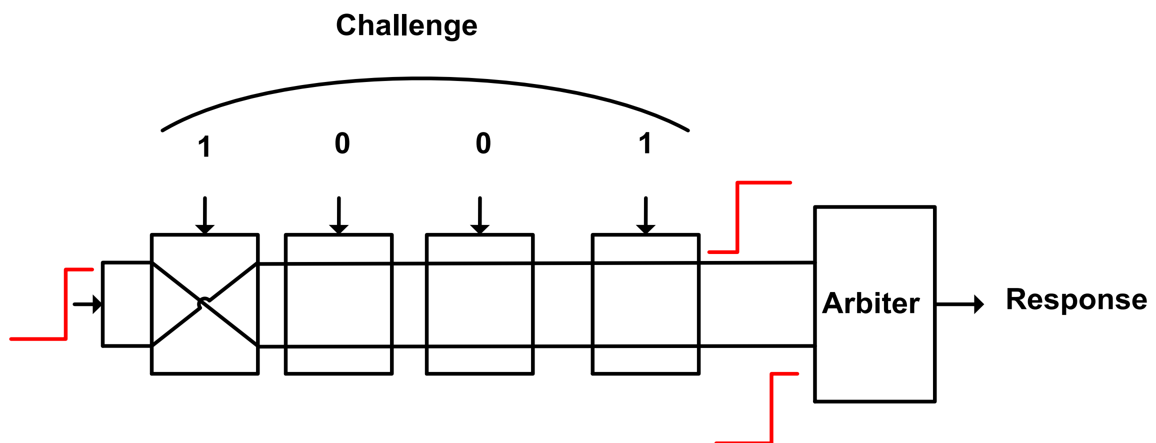

Figure 8: Arbiter PUF [9]

Now after the signal transverses through two similar paths on the basis of the position of two signals, an arbiter will decide the output bit. In this case if top path signal arrives early arbiter will give '1', or else '0' as output. Thus, the *outputs on different configurations of digital circuits act* as **challenge** in the PUF and *corresponding output* as a **response**.

Ring oscillator PUF is another kind of delay based PUF which utilizes the delay variations (intrinsic randomness) in digital circuits due to manufacturing process.

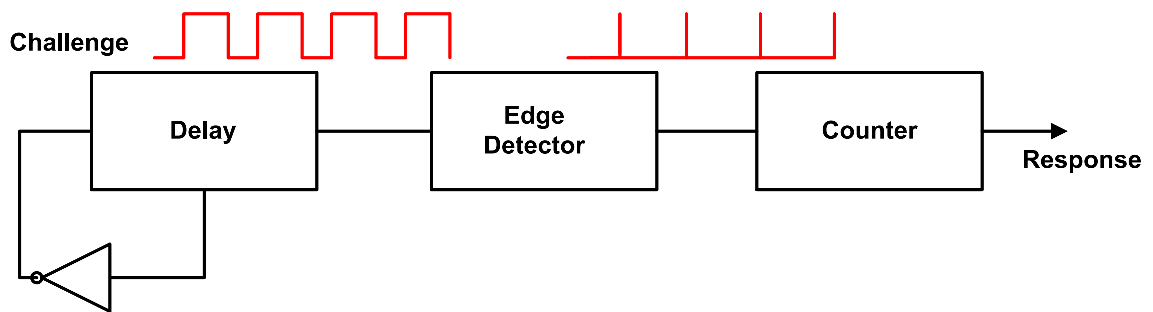

Figure 9: Ring-oscillator PUF[9]

As shown in Figure 9, ring oscillator PUF doesn't measure the delay but transforms the digital delay path into feedback path and puts the inverted output back to input. An AND gate can be utilized to turn the oscillator - 'on', 'off'. An edge detector is used to detect the positive/negative edge in the signal and counter will make the number of counts in predetermined time calculating the frequency of oscillator. As the delay is random and the number of counted pulse will remain random which will depend on the specific PUF device. The delay in this can be adjusted and parameterized. This *parameter* will act as **challenge** and *number of counts* will act as **response**.

SRAM PUF

The most popular intrinsic PUF is SRAM PUF. These also use the random variability introduced by inefficient manufacturing procedure. Unlike delay based PUF these don't use the delay based measurement but instead utilize the internal mismatch between two cross-coupled inverters.

As shown in Figure 10, an SRAM bitcell is a volatile digital memory component consisting of two cross coupled inverters bearing opposite stable values depending on resolving powers of respective inverters. If we assume right side of SRAM bitcell as the state of the cell then it is really difficult to predict the start-up value of the cell due to

symmetric structure. Though the structure looks symmetric in terms of functionality but due to inefficiencies in the manufacturing process there appear differences in the physical parameters of transistors (length, width, oxide thickness) which result in skew in the cell. Thus, the cell will show a biased tendency towards a particular value '1' or '0' every time it is powered up. More about skew will be discussed in section 3.6.

The above description is sufficient to understand the challenge & response pair for SRAM-PUF. Thus *powering-up* is the **challenge** in SRAM PUF & *resulting bit pattern* is **response.**

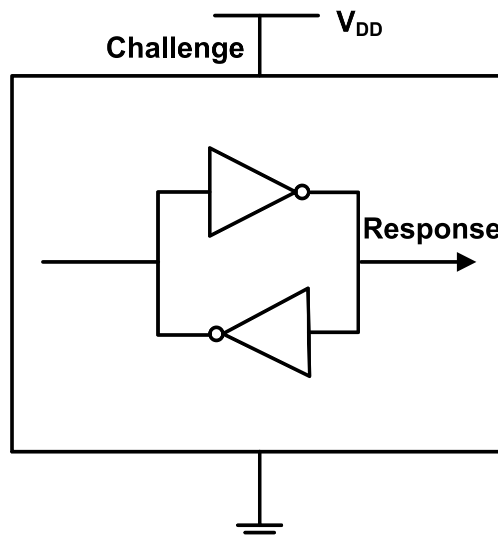

Figure 10: SRAM PUF [9]

Butterfly PUF

As shown in Figure 11, butterfly PUF has a construction similar to that of SRAM PUF but instead of inverters, two cross coupled latches are present. In this kind of PUF, one latch is preset to '1' and another is reset to '0' from a common external signal. Enabling both the latches simultaneously will make the whole circuit unstable and after the external signal is removed the circuit will try to settle down towards a particular value '1' or '0'.

This settling towards '1' or '0' will depend on the mismatch between the two latches and is random.

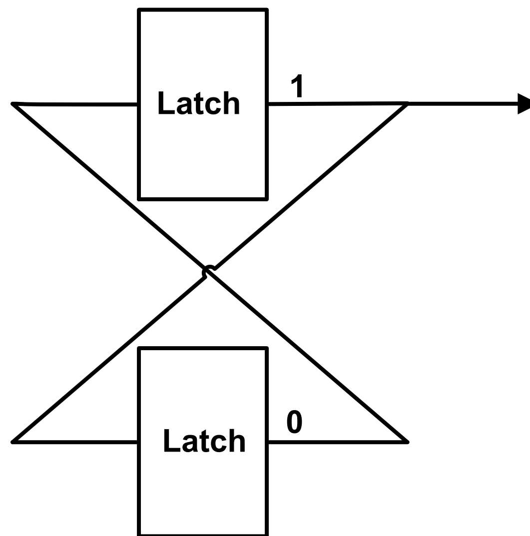

Figure 11: Butterfly PUF[9]

Choosing a latch pair is the **challenge** in this PUF and the random *settling output* is **response**.

Chapter 2: Literature Review

2.1 SRAM PUF and associated advantages over other PUF

We have chosen SRAM-PUF for our research; will try to match SRAM-PUF to the criterion laid down by Pappu [7]:

1. Easy to fabricate - SRAMs are currently produced in mass numbers currently with high cost efficiency, so SRAM-PUF can be easily fabricated as required.
2. Easy to probe - SRAMs can be conveniently probed; much work has been done in past investing of SRAM improvements.
3. Hard to clone - SRAM PUF is hard to clone as device variation is random and even the manufacturer is unable to control and reproduce it.
4. Structurally stable - Past research has shown that SRAM is the most stable memory structure till date.

On being powered up, every SRAM will generate a unique start up pattern due to device variations. These unique patterns are tough to replicate. Hence, it can be inferred that the start-up state of the SRAM is a result of mismatch between the devices representing the manufacturing variability in the fabrication process.

The major advantage of using SRAM-PUF over other PUFs is that SRAMs are already present in many electronic devices which can additionally be used as a security device with minimal additional effort.

2.2 Previous SRAM-PUF Work

Before the intrinsic PUFs were discovered, researchers used custom built circuits or the modification of the IC manufacturing process to generate a reliable PUF. Guajardo et al. [10] introduced the concept of SRAM-PUF. They identified Intrinsic PUF which can be defined as PUF already present in the device and that don't require any modifications to satisfy security goals. They took the results from[11], which shows that microscopic variations in the dopant atoms in the channel region can induce considerable differences in the V_T of the transistors of an SRAM cell.

A number of SRAM-PUF implementations are described in the literature.

To solve the noisy data from random bits of SRAM-PUF, Bosch et al.[12] proposed the Helper Data Algorithm (HDA) key extractor for SRAM-PUF. Their work is focused mainly on study and implementation of fuzzy extractors on FPGA. This work concentrated on methods to reconstruct the sane key from the noisy data. They never took into consideration the hardware cost of HDA or tedious hardware constructions. They tried to make it as a final block necessary to generate cryptographic keys.

Maes et al. [13] were the first to use the soft decision information Helper Data Algorithm (HDA) for extraction of the secure ID. Even though soft data information reduced the number of SRAM cells required for key-generation, the area-overhead due to Error Correction Code (ECC) is not an efficient implementation for SRAM-PUF. The main focus of their work is to reduce the number of unreliable bits and hence reducing the need of Error Correction Code Algorithm complexity.

Hofer et al. [14] demonstrated a pre-processing technique of segregating usable, not usable or not reliable SRAM cells on the basis of threshold voltage (V_{TH}) mismatch between the cross-coupled transistors. As shown in Figure 12 SRAM cells are divided into NUF (Not Useful) if

the V_{TH} difference between the transistors is small and external noise can sway the SRAM cells output in random direction at every power-up. UF (Useful) cells which have considerable difference of threshold voltage in cross-coupled transistors and external noise has no or very less effect on the cells. Figure 13 gives an estimate of this technique to calculate the mismatch between two NMOS transistors. They assumed PMOS to have minimal variations. This can be an effective technique but the area-overhead for segregation circuit and pre-processing time makes it an un-optimized solution. Also assumption made by authors may not stand true.

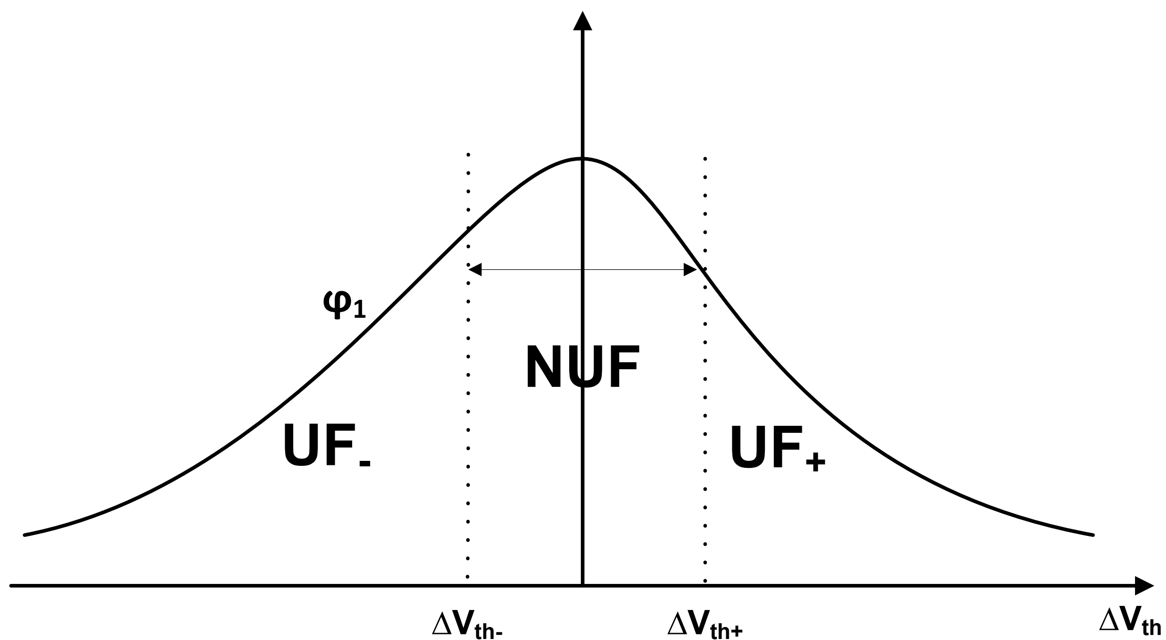

Figure 12: Segregation of UF (Useful) and NUF (Not Useful) Cells

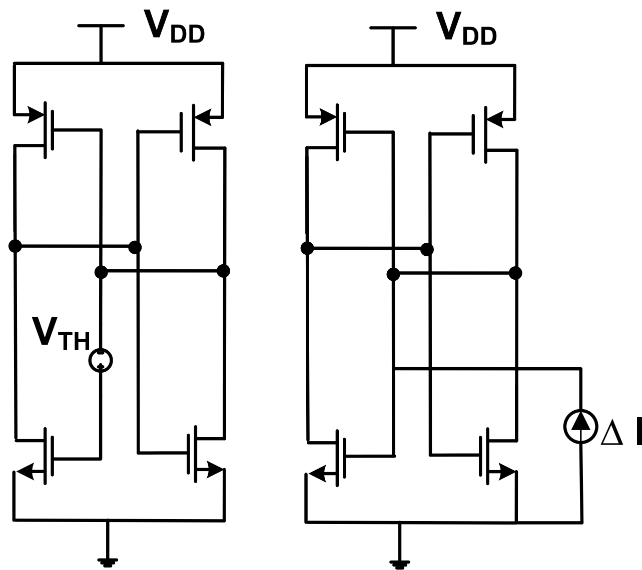

Figure 13: SRAM cell with additional voltage source and current source

Krishna *et al* [15] came up with the idea of MECCA: PUF (MEmory Cell Characterization based Authentication PUF) making use of the different word line (WL) pulse duration for generating unique signatures. It tries to remove area overhead due to HDA or fuzzy extractor but using WL pulse duration as a critical parameter in evaluation of secret key may get affected due to temperature variations. As shown in Figure 14 a programmable delay block is attached to the row decoder to control the WL pulse width. This delay inducing circuitry can be used for controlling WL is responsible for increasing challenge-response samples for the SRAM-PUF. Environmental conditions vulnerability and area overhead can be two limiting factors for this technique.

Also, no comparison is drawn by authors with contemporary SRAM-PUF to prove the exact advantage.

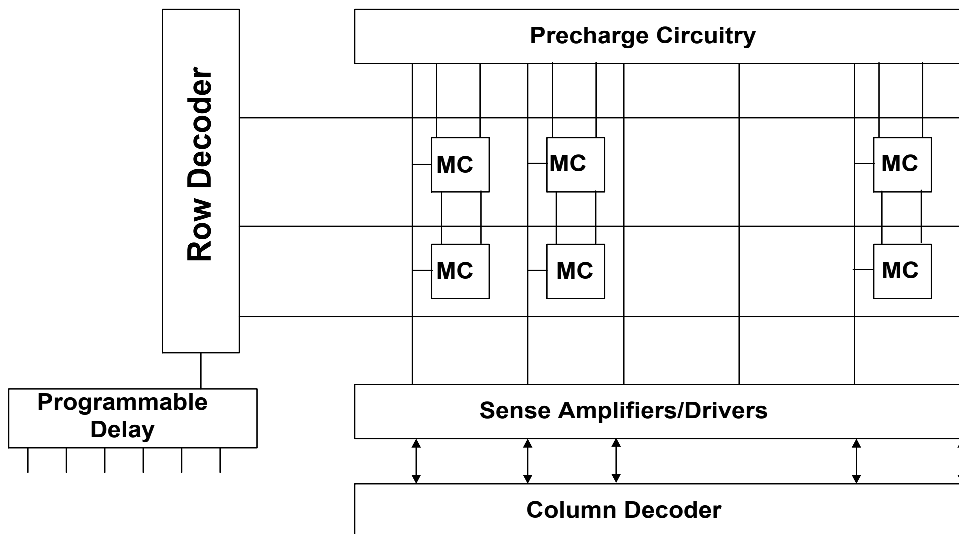

Figure 14: MECCA PUF, Memory block with peripheral circuitry and programmable delay circuit

Fujiwara et al. [16] introduced the concept of extracting unique finger print by using random failure bits in an SRAM using Memory Built-In Self-Test (MBIST) for detecting failure bits. The main idea is to make use of failed random bits to generate secret key or chip ID.

Actually, operating margin of each cell on a SRAM chip is different. So if we operate the SRAM at such a voltage where SNM of SRAM cells become worse, the fail-bits will appear randomly on a chip and their physical locations are unique on a chip. These failed bits can be used to generate random secure-ID. As shown in Figure 15 (a), the circuit block diagram used to generate unique ID. Figure 15(b) describes the workflow of MBIST based SRAM-PUF, initially the voltage is set and failed bits are calculated with MBIST. If failed bits come in the required range, multiple tests are run on same SRAM. Finally, a unique ID is generated from the output of failed bits.

The main drawback in this technique is the translation time required to generate secure ID from random bits. Also, there are chances of degradation of uniqueness as mentioned by authors in this paper.

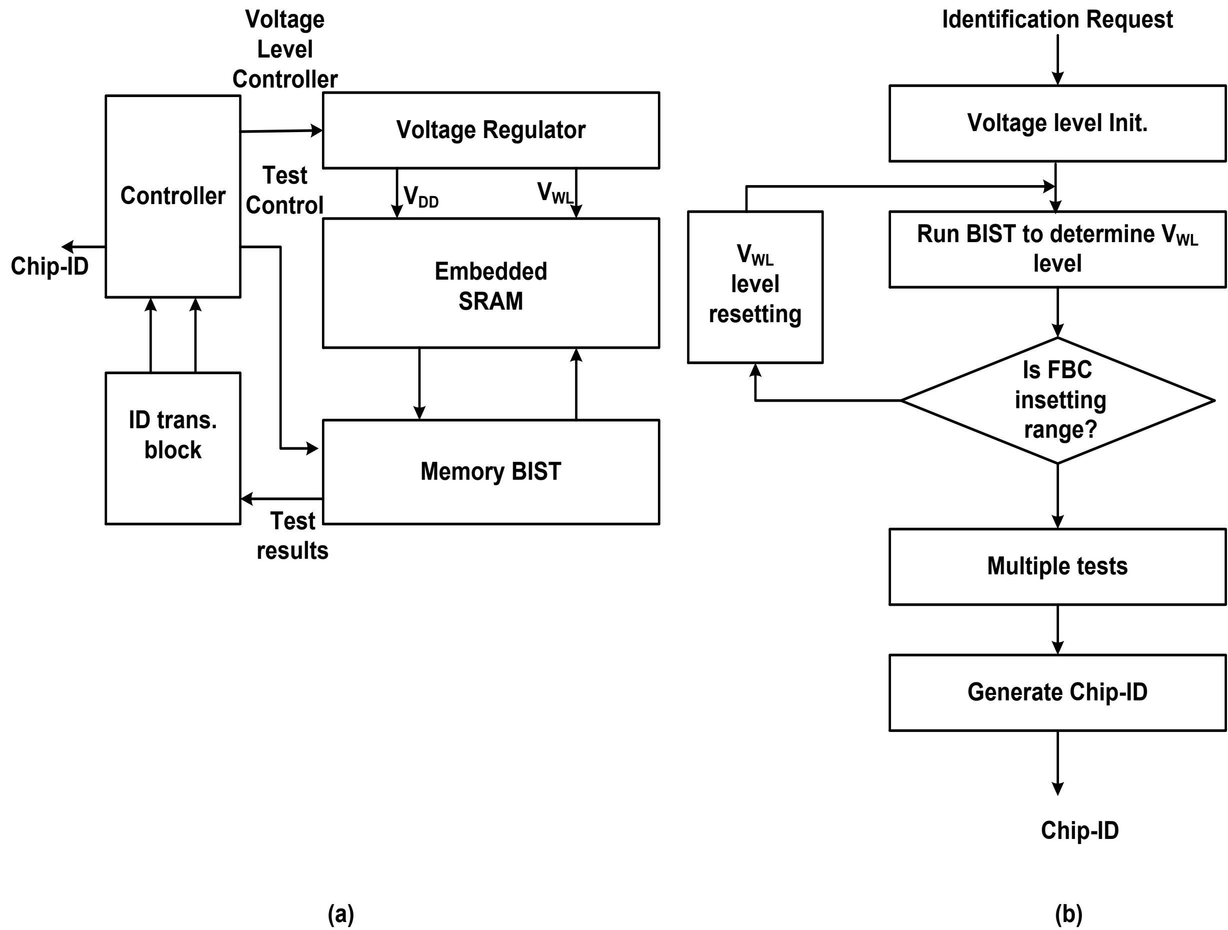

Figure 15: (a) block diagram of circuit (b) flow of ID-generation.

Holocomb et al. [17] introduced the concept of DRV-Fingerprinting, i.e. generating silicon fingerprints at Data Retention Voltage which looks promising due to 28% improvement in reliable SRAM cells compared to power-up SRAM-PUF but authors still believe that further reliability tests are required before comparing DRV based SRAM-PUF. Also, characterization of DRV based PUF is time consuming. Taking all these works into consideration, we propose an aging based SRAM-PUF with improved security (reliability and uniformity of its signature key).

2.3 Security Parameters

Before looking into ideal requirements from a PUF, we should look into 2 important parameters which define the strengths of a PUF [18, 19].

(a) Hamming Distance - Hamming distance measures the parity of two bit-strings or it is defined as parameter that gives an idea about number of bit-mismatch between two strings. The important thing is, the two bit-streams which need to be matched must have equal length.

Example - Bo'o't & Bo'a't has a hamming distance of 1.

(b) Fractional Hamming Distance - It measures the relative bit disparity among two bit-streams. It is defined as

$$Fractional\ Hamming\ Distance = \frac{Hamming\ distance}{Total\ number\ of\ bits}$$

The most important security parameters [18, 19] for the various types of PUFs used to quantify their robustness are - (i) Uniqueness (ii) Reliability and (iii) Uniformity. As depicted in Figure 16 three dimensions of PUF strength are aligned on the 3-axis of security function. Inter-chip variation is captured using device axis, the other two-axis are used to capture the intra-chip variation with space & time axis.

2.3.1 Uniqueness

A security system should be unique and is expected to generate a unique key which any other similar system should not be able to generate or cannot be cloned by using any other method. As explained, since PUF key is generated from process variations which are random, so PUF output key should be random & unique to the IC generating it. In [19], a quantitative method

involving Hamming Distance (HD) between a pair of PUF is used to evaluate uniqueness. Two chips *(i,j)* with *m-bit* responses, R_i and R_j respectively for challenge *C*, the average inter-chip *HD* between *k* chips is -

$$Uniqueness = \frac{2}{k(K-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \frac{HD(R_i,R_j)}{n} \times 100\% \qquad (2.3.1)$$

2.3.2 Reliability

A reliable security system is the one which gives same the output pattern irrespective of environmental variations e.g. Temperature. A hacker can change the surrounding temperature which may affect output pattern, so the system should be robust enough to handle all environmental fluctuations.

Reliability can be evaluated by calculating intra-chip HD among several samples of PUF response bits. At first, using normal operating conditions (at room temperature, normal supply voltage) an *n*-bit reference response (R_i) from chip *i* is obtained. Afterwards, same *n*-bit output, *m* samples are generated using different operating conditions (different temperature or supply voltage) namely R'_i.

For the chip *i,* average intra-chip HD is -

$$HD_{INTRA} = \frac{1}{m} \sum_{t=1}^{m} \frac{HD(R'_{i,t}, R_i)}{n} \times 100\% \qquad (2.3.2)$$

where $R'_{i,t}$ is the t-th sample of R'_i.

$$Reliability = 100\% - HD_{INTRA} \qquad (2.3.2.1)$$

2.3.3 Uniformity

It estimates the proportion of 0's and 1's in the response bits of a PUF. For a random PUF, this factor should reach 50% value. The Uniformity, is defined of an *n-bit* PUF identifier is defined as -

$$(Uniformity)_i = \frac{1}{n}\sum_{l=1}^{n} r_{i,l} \times 100\% \qquad (2.3.3)$$

where $r_{i,l}$ is the lth binary bit of an n-bit response from a chip i.

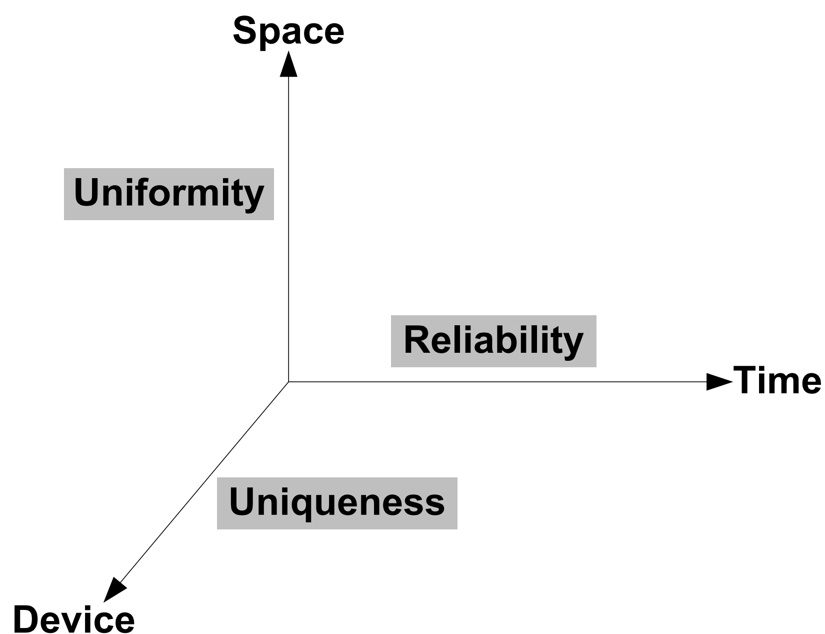

Figure 16: Parameters of PUF measurement

The **Ideal requirements** from a SRAM-PUF are [20]:

1. SRAM cells should maintain same state at every power-up voltage which means hamming distance between different power-up states should be zero to maintain consistency.
2. The power-up states of different SRAMs should give different output string, i.e. hamming distance of two different SRAMs shouldn't be zero at any cost.

3. Uniformity in output should be associated with PUF, which means for SRAM-PUF the number of 0's & 1's should be equal so that the probability of guessing is reduced.

2.4 Energy Efficiency dependence on SRAM Array Structure

High energy efficiency is a paramount design constraint in many ultra-low power applications such as portable electronic devices, wireless sensor nodes, and implantable biomedical devices [21]. In these applications, SRAMs play a key role in energy consumption due to the high cell density for computational power improvements. One of the most popular ways of obtaining minimum energy consumption is to lower the supply voltage around or below the device threshold voltage [22]. However, lowering supply voltage generates various design issues. Degradation in cell stability, noise margin, on-current to off-current ratio, and strong sensitivity to Process-Voltage-Temperature (PVT) variations have to be carefully handled for reliable operation. As we observed, design of SRAMs in this operation region is more challenging due to additional design constraints compared to generic digital logic, various circuit techniques have been published with successful hardware measurements[23-25]. Decoupled SRAM cells have been popularly deployed for improving cell stability. Write margin issues have been tackled through several techniques using positively or negatively boosted voltage, strengthening the write access transistors utilizing channel length modulation, and collapsed supply voltage [24, 25].

2.4.1 Supply Voltage (VDD) Scaling Effect on SRAM.

Supply voltage is a critical parameter in minimizing the SRAM energy. SRAMs for ultra-low energy consumption have been explored for various recently emerging applications where performance can be mitigated for higher energy efficiency. Studies have demonstrated that sub-threshold or near-threshold circuits achieve minimum energy consumption [23]. Thus, SRAM design techniques for low operating voltage have been explored, generally following the traditional SRAM organizing practice of having more rows than columns [26]. Research

works on optimal SRAM array structures for energy minimization have rarely been conducted. Considering the increased SRAM density in ultra-low energy applications, it is highly necessary to revisit SRAM array structures for better energy efficiency. As CMOS technology development is advancing the scope of voltage scaling which is simple and widely used technique for energy efficiency enhancement. Both the dynamic energy associated with the accessed wordline and bitline, and the static leakage energy is strongly affected by the supply voltage [24]. In the supply voltage region where dynamic energy is a dominant component, lowering supply voltage decreases the total SRAM energy. However, as the supply voltage comes around or below the threshold voltage, lowering supply voltage is not much effective in the energy minimization due to the increase in the static energy. This is caused by the exponentially increased delay. Consequently, the minimum energy point is found where the supply voltage is around the device threshold voltage.

2.4.2 Sub-array based SRAM architecture

Splitting SRAM array into sub-arrays which effectively reduce the word-line (WL) and bitline (BL) capacitance, also help implementing local energy minimization[27]. As shown in Figure 17 a fixed density SRAM array can be split into 4 sub-array(s). The total SRAM memory density will remain same after splitting with minimal area-overhead.

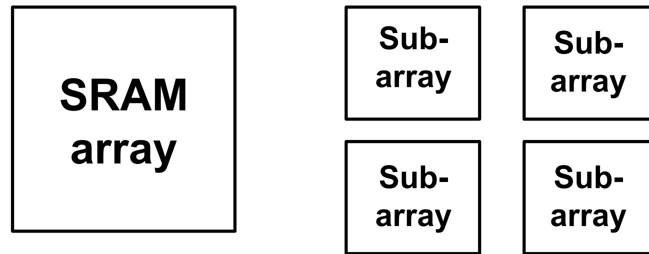

Figure 17: SRAM array divided into sub-arrays with same density

The main disadvantage of this technique is the increased complexity of control circuitry and area-overhead due to extra logic but the trade-off between energy and area is still positive.

2.4.3 Hierarchical Word Line (WL) & Bit Line (BL)

Splitting WLs & BLs into sub hierarchies also helps in reducing total metal line capacitances and hence enhances the speed which can help improving the static energy [28]. RC time delays are directly proportional to the length of the metal wire. Longer the metal line, more is the associated parasitic. Figure 18 shows the division of one word line into hierarchical word lines for sub-arrays. Similar division of bitlines can also help reducing the parasitic.

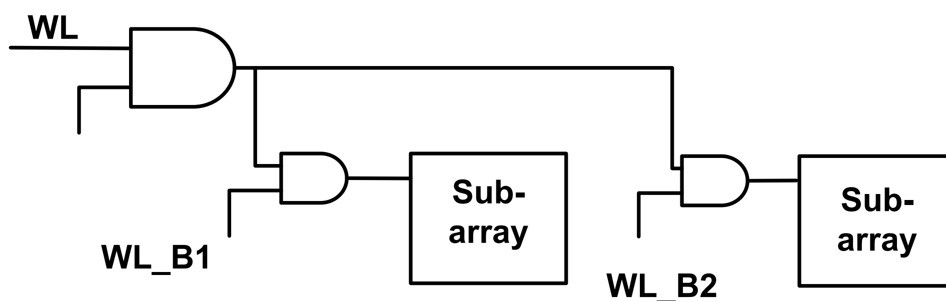

Figure 18: Hierarchical word-line (WL) scheme

2.4.4 Self-timing

This technique [29] mainly used to optimize timing in CMOS SRAMs for sub-90nm technology, where the process variations are quite high. Under different operating conditions, the sensing margin (minimum differential voltage) changes so that it is difficult to fix the timing of Sense Enable signal. To solve this, a column of dummy cells are inserted and pre-charged as the normal SRAM column. The discharge of dummy-column will then estimate the optimum discharging time and enable the sense amplifier accordingly as shown in Figure 19.

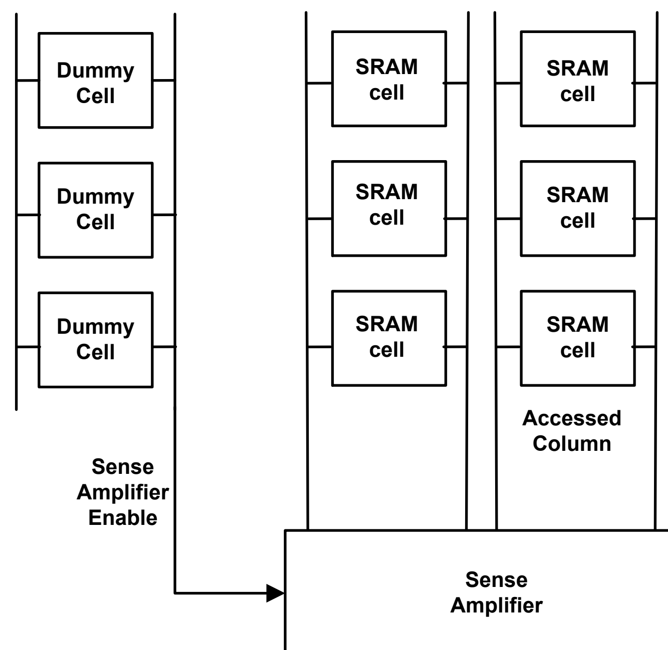

Figure 19: Self-timing timing optimization scheme

2.4.5 Data Retention Voltage

To reduce leakage current when SRAM is not in direct read/write operation, a retention mode can be introduced where the supply voltage of SRAM array can be reduced to a certain level where it is just sufficient to retain the data itself. To reduce the leakage further body bias of transistors can be increased resulting in increased threshold voltage[30]. Figure 20 shows a sleep signal which when asserted, the supply voltage decreases from VDD to some lower value. The reduced supply voltage thus helps in reduced leakage current for the constant resistive path of a SRAM cell.

Figure 20: Retention scheme for saving leakage energy

2.4.6 SRAM array structure influence on energy minimization

In addition to all the above techniques, SRAM array structures also influence energy consumption. As we can see in Figure 21, same density SRAM array structure can take different array shapes just by changing the number of rows and columns keeping the total density constant. Evans et al[26] conducted initial investigations of SRAM array structures for optimum energy consumption. In their work, the optimum SRAM array structures for minimized energy consumption were found to be non-square, more rows than columns, while the optimum array structures for minimizing the memory access time were squarer than those for the minimum energy consumption. To support their energy optimized taller array structure they claimed that energy cost of precharge section is more costly in n direction (m - number of rows, n - number of columns, m + n = constant). They used simulation based model for their study employing MOSIS 2.0 micron process. For older process nodes or long channel devices (>90nm) dynamic energy is much higher than static energy and is the only dominant component in total energy calculations.

In sub-90nm technologies, we suspect assumptions made by Evans may not be true for SRAMs operating at sub-threshold or near-to sub-threshold voltage. With increase in leakage current & read delay, static energy also starts increasing and plays an important role in the total energy at lower supply voltages. This increase of leakage current & read delay amounts to more static power for taller arrays in-comparison to fat arrays due to more capacitance/bitline. Figure 21 shows that square SRAM array which can be modified into tall or wide array with same memory density.

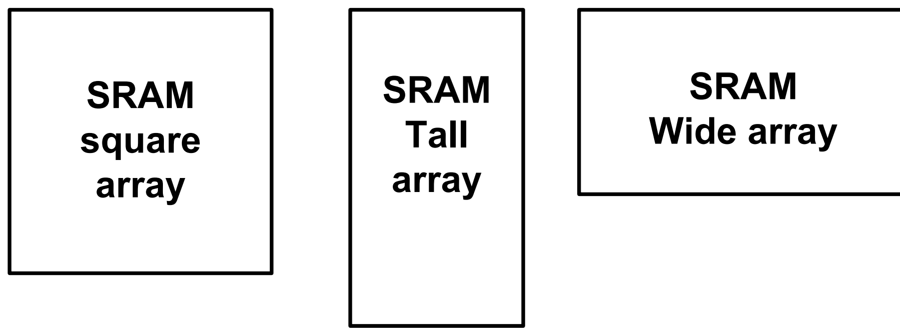

Figure 21: SRAM with same density with different array configuration

Chapter 3: Implementation & Simulation Results

Section I: SRAM-PUF

This chapter will explain the implementation strategy used for this thesis, before that which it is necessary to understand basic SRAM operation which will form the foundation of this work.

Simulation Setup - 65nm Process (TT), Voltage (1.2V), Temperature (25 °C)

3.1 SRAM Operations & Power-up Value Based SRAM PUF

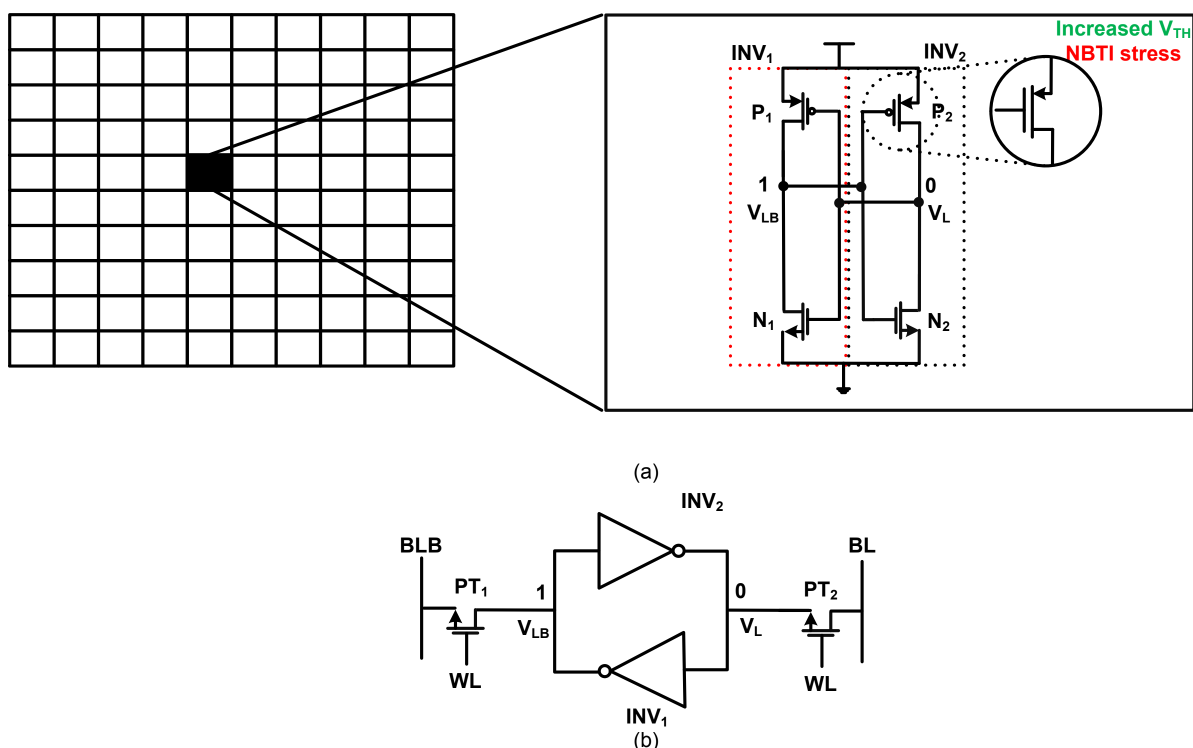

Figure 22: (a) SRAM array (b) Symmetrical 6T cell

As described in the previous section SRAM PUF has many advantages over other PUFs. Figure 22(a) shows a basic SRAM array consisting of number cell, figure 22(b) consist of a SRAM cell ,ade up of two inverters (INV1 & INV2) connected back-to-back. Also, it has two pass transistors (PT1 & PT2) controlled by wordline (WL), which insulates the cell from external bitline (BL) and bitline bar (BLB). The read and write operation of SRAM cell is explained as below.

3.1.1 Read operation

For read, initially the bitlines are precharged to VDD and wordline (WL=0) is switched off. After wordline is switched ON (WL=1), the two inverters try to resolve internally by providing positive feedback to each other. The internal nodes (VL & VLB) now have complimentary values on the basis of physical parameters of corresponding inverters (INV2 & INV1) respectively. As soon as wordline is switched ON (WL=1), both the bitlines (BL & BLB) try to read the internal value of the cell. The bitline corresponding to node storing '0' will discharge while the other bitline will maintain its VDD value in absence of discharge path. For example - if both the inverters resolved the internal state and node VL stores '0' whereas node VLB stores '1'. Now as soon as WL=1, BL & BLB try to discharge through PT1 & PT2. For BL there is potential difference across PT2 as VL is '0'. Also transistor N2 is ON (due to VLB = 1). So BL will discharge through the discharging path (shown in Figure 12) and BLB maintains VDD. The cell is said to have bit-value as '0'.

3.1.2 Write operation

For write, initially cell is insulated from bitlines by switching wordline OFF (WL=0). Bitline & bitline bar are charged according to the value expected to be write in the cell. If we want to write '1' in cell, then precharge only bitline (BL=1), bitline_bar is discharged

(BLB=0). Now as soon as we switch wordline ON (WL=1), the BL value will be written inside the cell.

The power-up value of SRAM cell can be defined as **read** value of SRAM bitcell without any previous **write.** It can also be explained as value which a SRAM bitcell exhibits when powered-up from rest. The following paragraph will explain the reason why this power-up value is unique to individual devices.

This back-to-back inverter system seems to be symmetric and unstable at power-up without any value stored in it. But in reality both the inverters are not exactly same and have different physical parameters (length, width) due to process variations during fabrication process. This process variation is random and cannot be controlled and hence there is variation in physical parameters of the devices.

These physical parameters will decide the power-up value of the particular SRAM bit-cell. Since it is difficult to predict the variations in physical parameters, it is difficult to predict the power-up values.

After understanding the power-up behaviour of SRAM cells, classification of SRAM cells can be done on the basis of their power-up behaviour. The two major classifications of SRAM cell are:

3.1.3 Partially skewed

Partially skewed cells are cells that show little mismatch between two back-to-back inverters. They show skew in one particular direction (0 or 1) but under varying environmental conditions, they can flip to opposite directions.

3.1.4 Fully-skewed

Fully skewed cells are those cells which have high mismatch value i.e. irrespective of environmental variations they power-up to a particular value. Normal SRAM operation is not affected by fully-skewed cells as external write operation can force cell to behave as per requirement.

Table 1: SRAM cell skew

SRAM Cells

Power-ups	1	2	3	4	5	6	7	8	9	10
	1	1	0	0	1	1	0	1	0	1
	1	0	0	0	1	1	1	0	0	1
	1	1	0	0	1	1	0	1	0	1
	1	1	0	0	1	0	0	1	0	1

Cells 2,6,7,8 changing their power-up pattern and can be categorized as partially skewed cells.

Cells 1,3,4,5,9,10 show consistent behaviour and can be categorized under fully skewed cells.

As stated in the ideal requirements of SRAM PUF:

1. To achieve minimum Hamming distance between different power-up states, an Ideal SRAM-PUF should have a minimum number of partially skewed cells. But since it is difficult to control skew in manufacturing process, there is need to control it using post-fabrication techniques only.
2. The simulation results show that the hamming distance between the two SRAM devices is nearly 50% of total number of bits. So this requirement is fulfilled automatically due to different device mismatch variation for different

devices. Hamming distance between two devices is also well explained in [31].

Simulation results show that the number of 1's and 0's in power-up state is not equal. To equalize the number of 1's & 0's, it is necessary to maximize the uniformity by skewing some the majority cells in opposite direction. Also, to improve the reliability of PUF, number of partially skewed cells should be reduced.

To see the effect of environmental conditions on partially skewed cells simulations on SRAM under various environmental conditions were performed.

3.2 SRAM-PUF power-up variations due to environmental fluctuations (Monte Carlo Simulations)

To evaluate the impact of various environmental factors on an SRAM or to show the importance for an SRAM to show a consistent behaviour at every power-up. An analysis on the power-up value of SRAM using commercial 65nm technology library was conducted. The analysis of SRAM bit cell for 8000 Monte-Carlo runs were performed which is equivalent to analysis of a 8kb SRAM memory chip. The results are explained as -

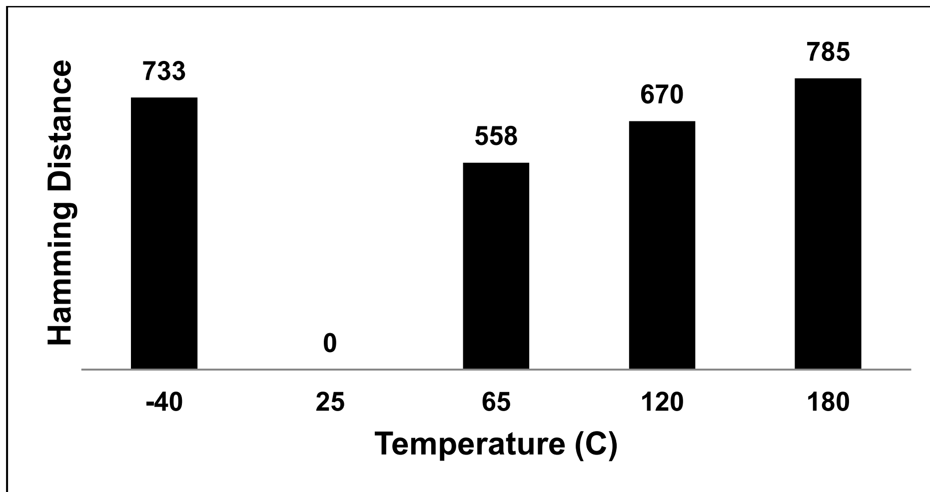

Figure 23: Hamming distance Vs Temperature

For the above simulation, standard operating conditions were fixed as (Process = TT (typical-typical), Voltage = 600mV, Temp = 25˚C) and power-up bit-pattern was compared under various temperature conditions keeping all other values same. Figure 23 shows variation of Hamming distance at temperatures (-40, 65, 120, 180) with respect to bit pattern at temp = 25.

The results of simulations are as expected and are random, since power-up pattern is random and skew of SRAM cells is difficult to predict thus the output of above simulation is difficult to predict and should follow random behaviour.

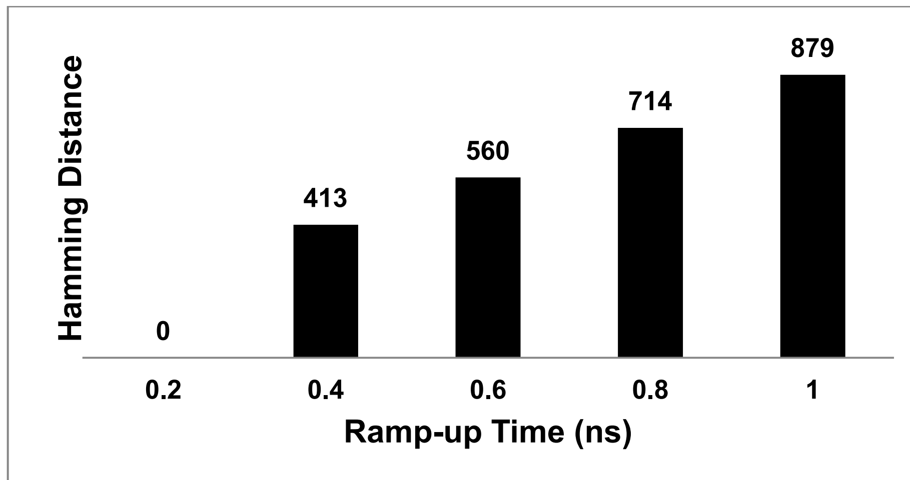

Figure 24: Hamming distance Vs Supply ramp-up time

Environmental variations can also result in variations of ramp-up voltage timings. This variation in voltage ramp-up timings can result in variations in power-up value of SRAM cell. From the Figure 24, linear variation in ramp-up supply voltage results in output behaviour (power-up value variation) that is not linear and does not show any particular variation pattern. Hence it can be concluded that the variation due to ramp-up supply voltage is also random for SRAM power-up values.

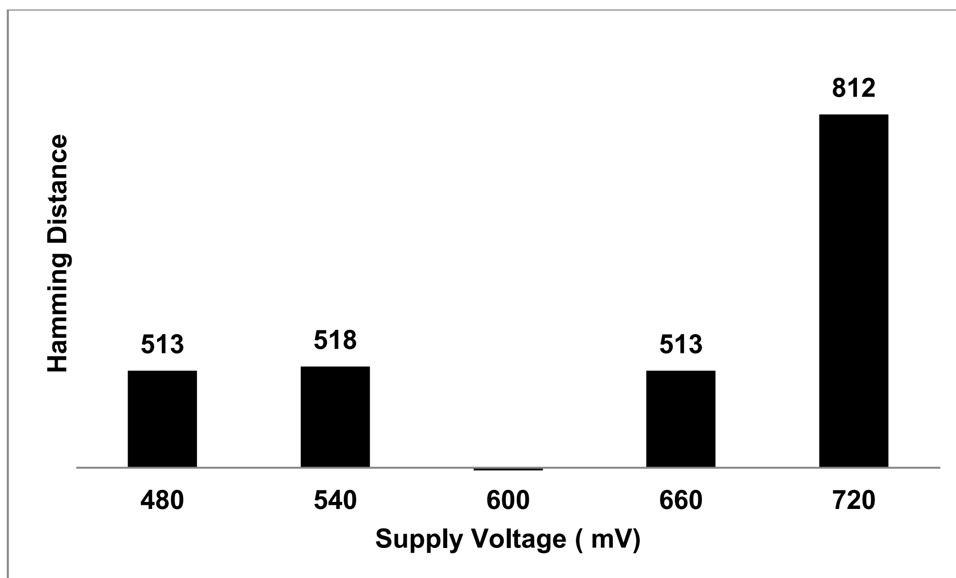

Figure 25: Hamming distance Vs Supply voltage

Figure 25 describes the variations in the SRAM power-up patterns with changing supply voltage. The start-up pattern is random and even if variation from mean value is same, the variation in output pattern is entirely random. As shown in the figure above, the +/-120mV variation in supply voltage from mean value (600mV), the output variation is also different (812, 513 respectively). This shows the need of methodology to improve the consistency of power-up pattern of SRAM-PUF.

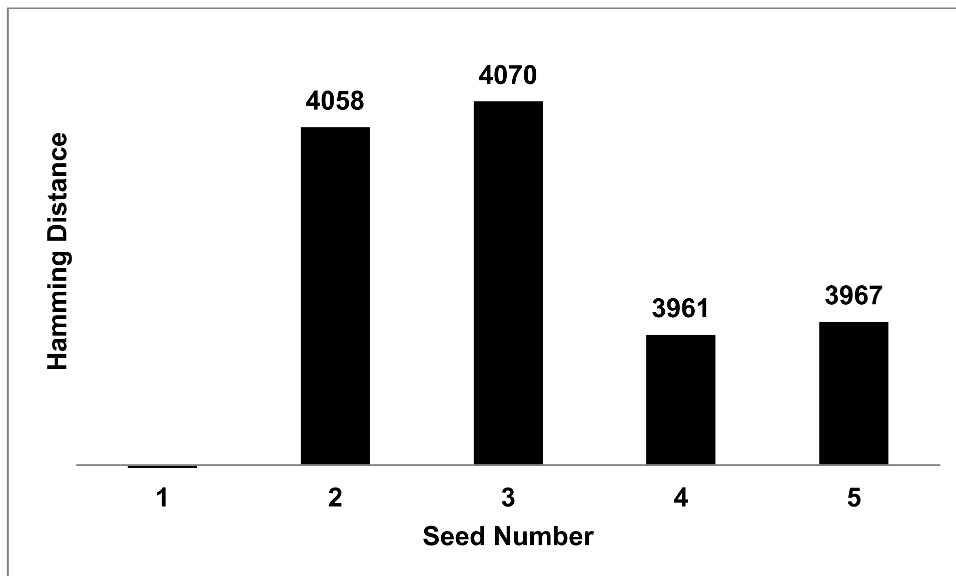

Figure 26: Hamming distance Vs Seed variation

Figure 26 reveals the difference in bit-pattern for different devices which are nearly 50% of total number of bits. Hence the various SRAM devices having the same design can still be differentiated on the basis of their power-up pattern. This is a very important result to see that even after considering variation which is nearly 10%, still two different devices cannot generate same power-up pattern as the variation between devices is much larger.

Figure (23-26) describe the variation in power-up pattern of SRAM PUF with environmental conditions. In these, we assumed that only one variation is affecting the PUF at a time but in reality it is not possible to control variations individually. To check the impact of various parameters, we simultaneously simulated the impact of Temperature and Ramp-up time on the Hamming distance assuming (temp=25 & ramp-up time = 0.8ns) as the standard bit pattern. The result shows that Hamming Distance increases as the number of variations increasing.

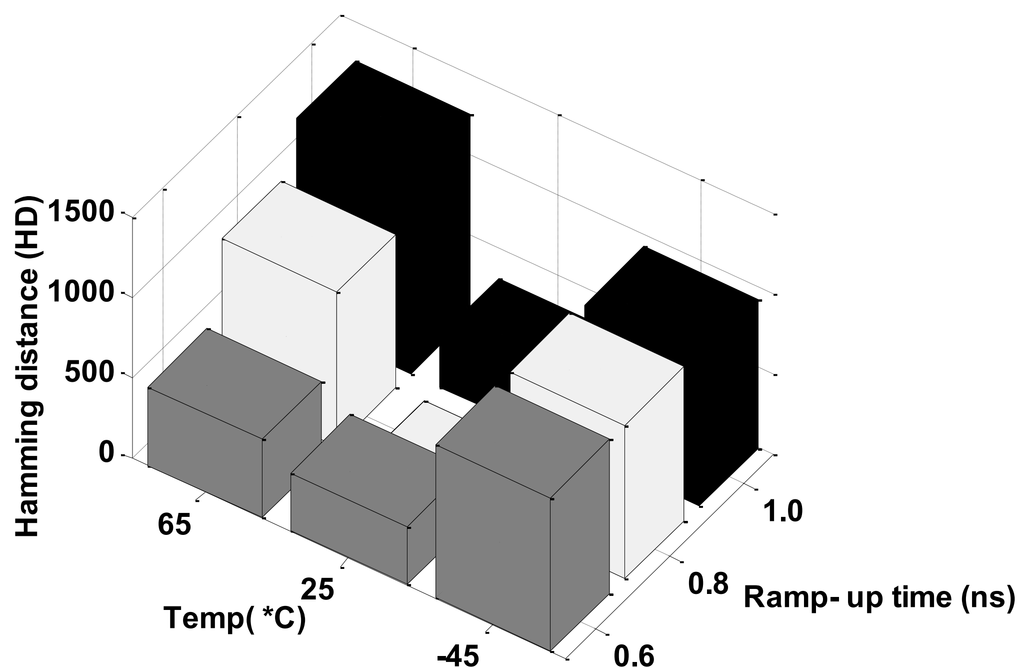

Figure 27: Hamming Distance variations with temperature and ramp-up time.

For, 8000 Monte-Carlo runs the above results show the hamming distance of 1251 for (temp= 65 & ramp-up time = 1ns). Additionally, if we include variation of supply voltage also Hamming distance is expected to increase.

In order to design an ideal SRAM-PUF a skew has to be introduced, that requires the knowledge of Negative Bias Temperature Instability (NBTI) stress phenomenon. NBTI is

considered as degrading factor for normal SRAM operation as it introduces unwanted skew, but in the proposed methodology for SRAM-PUF we're making use of skew to our advantage. The downside of introducing NBTI is reducing the life of SRAM-PUF but since the SRAM-PUF is powered-up for very small time during secret key generation, the overall age of circuits will not be much affected.

3.3 Negative Bias Temperature Instability (NBTI)

3.3.1 Impact on Threshold voltage (V_{TH})

As the name suggests NBTI occurs when a p-channel MOS device is subject to negative bias (i.e. $V_{GS} = -V_{DD}$) on gate under high operating temperatures. The impact is decrease in absolute drain current (I_{Dsat}) & transconductance (g_m) whereas 'off' current I_{off} & threshold voltage (V_T) increases. The typical stress temperature lies in the range of 100-250 ^oC and oxide field below 6MV/cm. similar fields and temperature are encountered during high-performance applications in ICs [32].

Following description will reveal the physics behind the increase in V_T in a transistor:

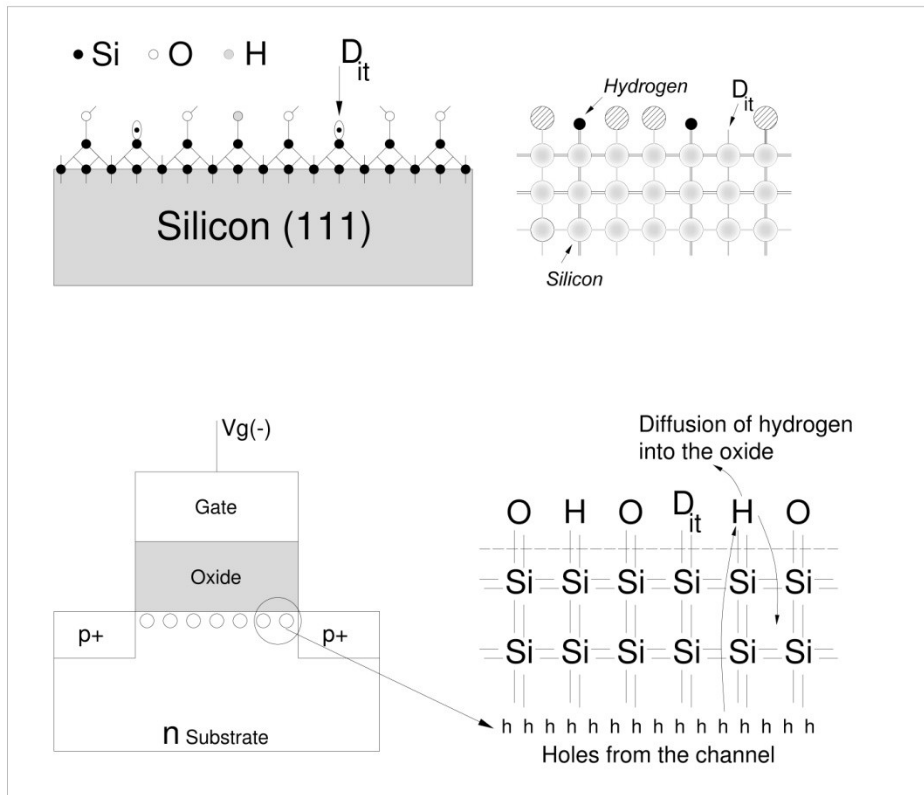

Figure 28: Schematic description showing the generation of interface traps when a PMOS transistor is biased in inversion[33]

The basis of NBTI degradation is attributed to continuos trap generation at the interface of Si-SiO_2 due to structural mismatch. During oxidation of Si (tetrahedron valency) atoms prefer to bond with oxygen but some atoms bond with hydrogen to form weak Si-H bonds. As the PMOS transistor is reverse biased, the holes in channel disassociate the weak Si-H bonds resulting in generation of interface traps.

Let's review the fundamental equation of V_T and see what factors impact V_T -

$$V_T = V_{FB} - 2\varphi_F - \frac{|Q_B|}{C_{ox}} \qquad (3.3.1)$$

where V_{FB} = flat-band voltage,

$\varphi_F = (kT/q)\ln(N_D/n_i)$

$|Q_B| = (4qK_S\varepsilon\varphi_F N_D)^{1/2}$

C_{ox} = Oxide capacitance per unit area.

Here, V_{FB} is

$$V_{FB} = \emptyset_{\varphi MS} - Q_f/C_{ox} - Q_{it}(\varphi\emptyset_s)/C_{ox} \quad (3.3.2)$$

where Q_f = fixed charge density

Q_{it} = interface trap density.

If we assume that, charge density (N_D) and oxide thickness are not changing then only Q_f and Q_{it} are responsible for change in V_T. As Q_{it} depends on φ_s which means interface trapped charge occupancy depends on surface potential. Any positive change in value of Q_{it} & Q_f will result in more negative threshold voltage (V_T) for PMOS device [33].

3.3.2 Impact on Static Noise Margin (SNM) curve

Before the detail of impact of NBTI on SNM curve is explained, the classification of Skew in SRAM cell should be made clear.

Skew in SRAM cell:

As previously defined, SRAM cells can be categorized under *partially skewed* and *fully skewed* cells. To introduce the further classification under fully skewed cells as 0-skewed cells & 1-skewed cells.

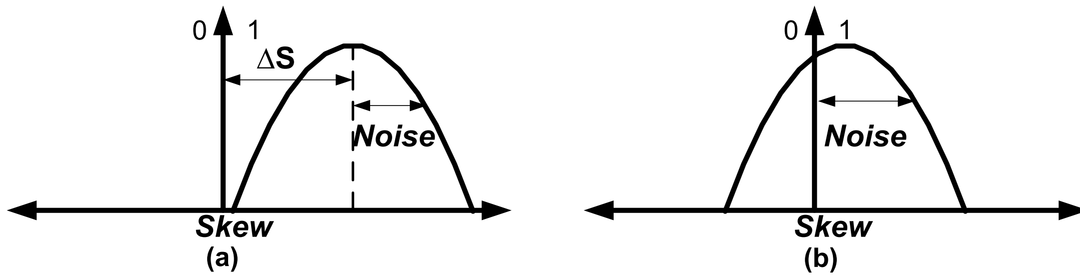

Figure 29: (a) Statistical output behaviour of 1-skewed cell (b) A partially-skewed cell which can sway to any direction under influence of noise [31]

As shown in Figure 29(a), if a cell is 1-skewed, it will generate output as 1 and minor noise fluctuations won't be able to flip the cell. In case of partially skewed cells (Figure 29(b)), a small noise fluctuation can flip the cell in any random direction.

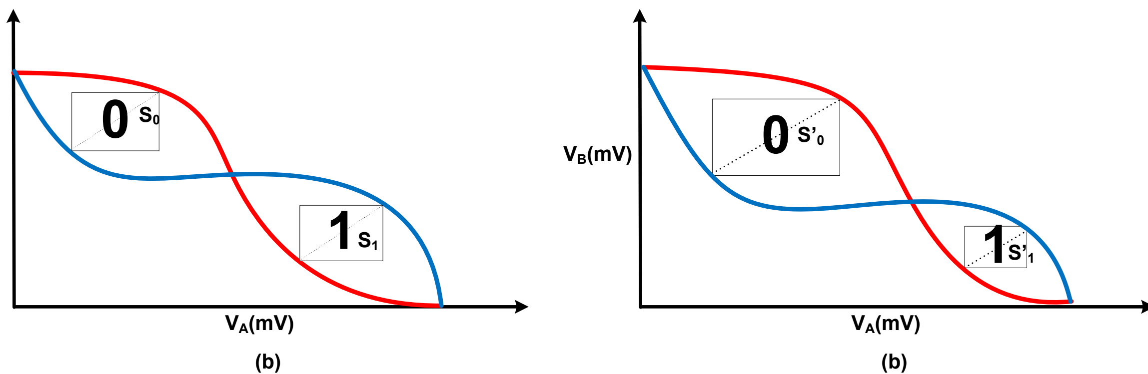

Figure 30: (a) A partially-skewed cell (b) 0-skewed cell [10]

Figure 30 depicts the impact of NBTI stress on a partially skewed cell which initially (Figure 30(a)) shows almost balanced butterfly curve or the probability of generating output '1' or '0' is equal and depends on the environmental fluctuations. This random output behaviour violates the Ideal requirement (1), which says "Hamming distance between various power-ups should be minimum". So, if the skew is enhanced in one particular direction while maintaining the uniformity, the SRAM cell will give same power-up pattern every time it is

power-up. This can be done by increasing the skew in one particular direction, as in this case cell is 0-skewed i.e. powering-up probability of cell is predominantly '0'.

The detailed setup & methodology for improving uniformity, reliability is explained in the following section.

3.4 Optimum Uniformity Methodology

Optimum Uniformity - For an SRAM-PUF, optimum uniformity is defined as equal number of **1's** and **0's** in the output pattern maximizing the strength of the generated security key. Higher the uniformity of security key, the more difficult it is to predict the key. To maximise this uniformity, we propose the method as shown below. In this methodology the number of **1's** and **0's** initially are checked and if statistics are found to be skewed in a particular direction, some of the majority cells are flipped in opposite direction.

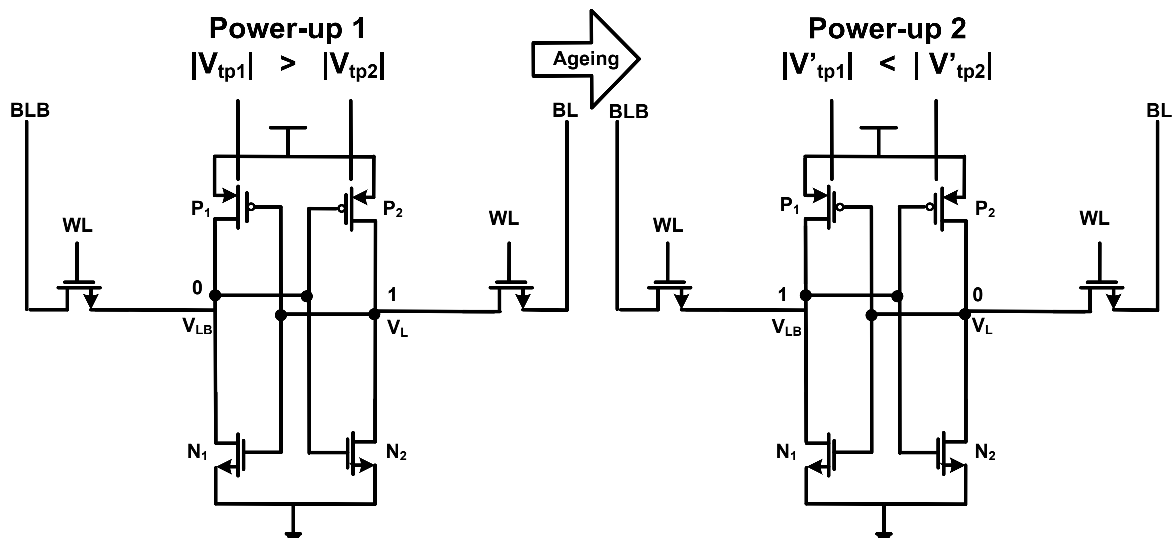

Figure 31: Aging impact on SRAM cell

As described in Figure 31, if initially a SRAM cell is biased towards '1', i.e. $|V_{tp1}| > |V_{tp2}|$ due to device mismatch, at power-up it can be safely assumed that P_2 will pull-up node V_L to '1'. If majority of cells in an SRAM array are found to be demonstrating similar behaviour, which means number of 0's are less compared to number of 1's, there is loss in uniformity. To improve the uniformity, there is a need to skew some of the cells in opposite direction which can be done by applying NBTI stress on the bit-cell. Since, PMOS (P_2) has $V_{gs} = -V_{dd}$ and

under high operating temperature it will experience NBTI stress which will lead to a change in the threshold voltage of transistor P_2.

After subsequent power-up, some of the cells will experience reversal in skew and the power-up pattern corresponding to these cells will flip showing opposite values. The only assumption here is that majority flipping cells are those cells which were storing '1' initially which is apparently true as aging impact will be uniform throughout the SRAM array. Now, after nearly equalizing the number of 1's & 0's, the focus should be to increase the existing skew in SRAM bit-cells so that subsequent power-ups give same pattern.

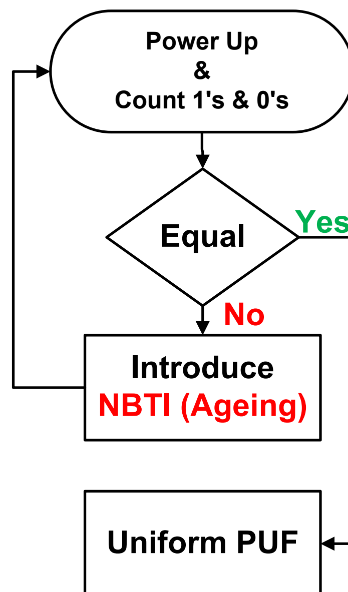

Figure 32: Proposed methodology to improve Uniformity

3.5 Reliability (Skew) Improvement Methodology

The 1st Ideal requirement says, "SRAM cells should maintain same power-up state at every power-up voltage which means hamming distance between different power-up states should be zero to maintain consistency". This difference cannot be reduced to absolute zero but it can be minimized, so that a robust functionality can be obtained. The following description will give an idea about our methodology to increase the skew in an SRAM cell and hence reducing the variations in subsequent power-ups.

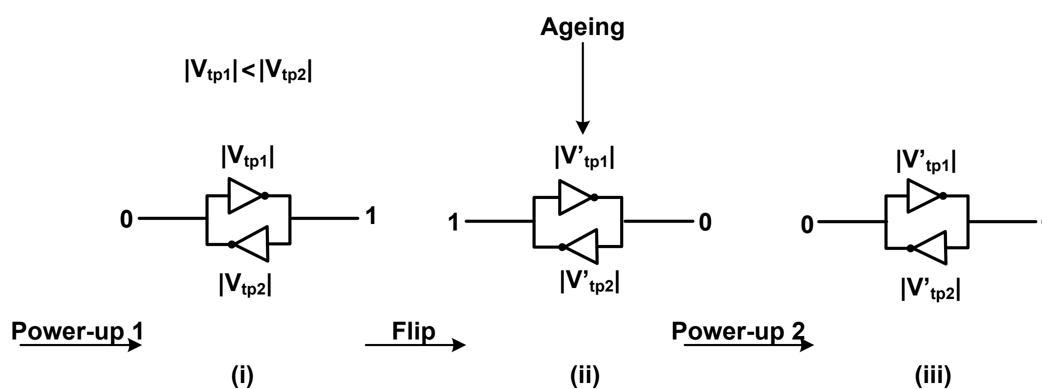

Figure 33: Skew improvement methodology

Basically, power-up pattern of SRAM cell is determined by V_T mismatch between two back-to-back inverters forming the cell. If V_T mismatch between the inverters is nominally small, the power-up bit-cell pattern will depend on environmental factors and cell can flip to any direction depending on the external conditions.

This inconsistency in power-up behaviour is not desired from SRAM-PUF, so if some more skew can be provided to already skewed cells, then the power-up variations can be reduced. Figure 33 explains the methodology we're planning to undertake, as described if the initial threshold voltage values of two devices ($|V_{tp1}|$ & $|V_{tp2}|$) are nearly same and if $|V_{tp2}|$ is slightly larger in comparison to $|V_{tp1}|$, it is expected to observe '1' on right side of cell and '0'

on left side of cell. Now, if we intentionally flip the cell by writing '0' in the cell and put some NBTI stress on the cell, it will result in change in the V_T of transistors to $|V'_{tp1}|$ & $|V'_{tp2}|$ and difference between two V_T(s) increases, making bit-cell more skewed in right direction. On subsequent power-up, the cell is expected to show '1' on right side with less ambiguity.

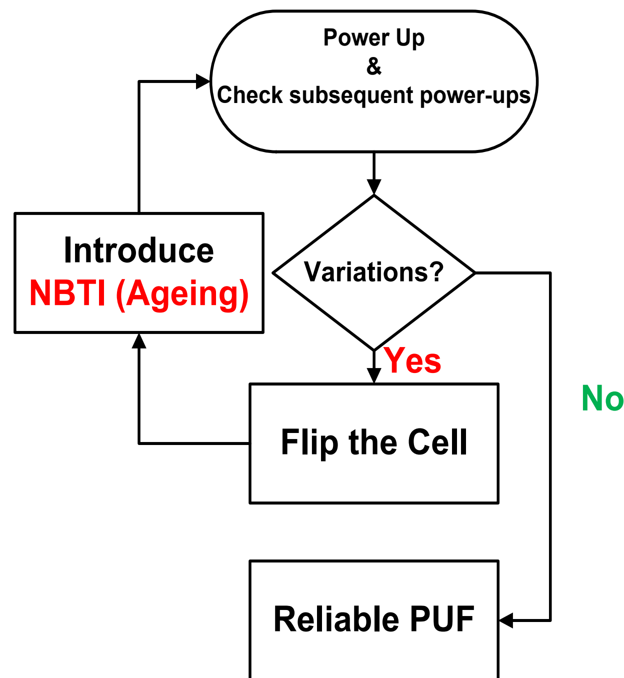

Figure 34: Proposed technique to improve Reliability

To implement flipping of SRAM bits, a cell flipping setup based circuit is described as shown in Figure 35. It involves an address counter which increments address sequentially. For every address, a data is read-out and after few clock cycles is flipped and written back at the same address location in the next clock cycle. More details of circuit functionality is explained in next section.

3.6 Cell flipping setup

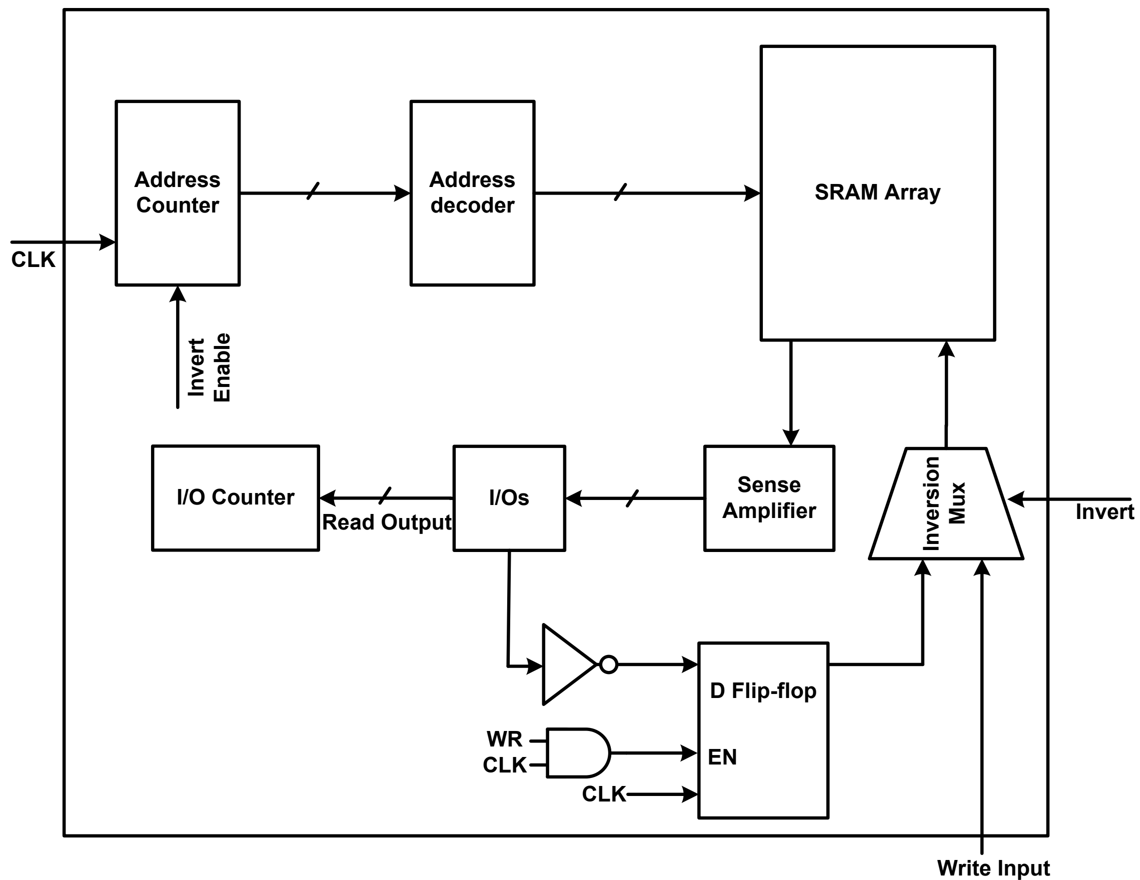

Figure 35: Cell-flipping setup

Steps to follow using this methodology:

1. Check number of 1's & 0's.
2. If unequal, expose SRAM to NBTI stress, it will flip some of the majority bits.
3. Check number of 1's & 0's again, if uniformity improves, go to next step, else repeat step 2.
4. Check with successive power-ups if some cells show partially skewed behaviour.
5. If there are partially skewed cells, flip the cells by writing opposite content and put aging which will improve the already present skew in cells.
6. Check the number of 1's & 0's after several power-ups.

The major advantage of this setup is that the same SRAM can be used for security functionality as well as for normal storage operation. The ***Invert (In)*** & ***Invert Enable (IE)*** are employed. When flipping of SRAM cells is required, ***IE*** will disable the address counter in alternate cycles allowing the flipping of cell value at the same address. Also, *In* will block external ***Write Input*** allowing the flipped content of cell to be re-written inside the same location where previous data was stored. When storage functionality is desired, external value can be written into the cell by passing the value to ***Write Input*** and read using the ***Read Output***.

3.6.1 SRAM Cell flip output

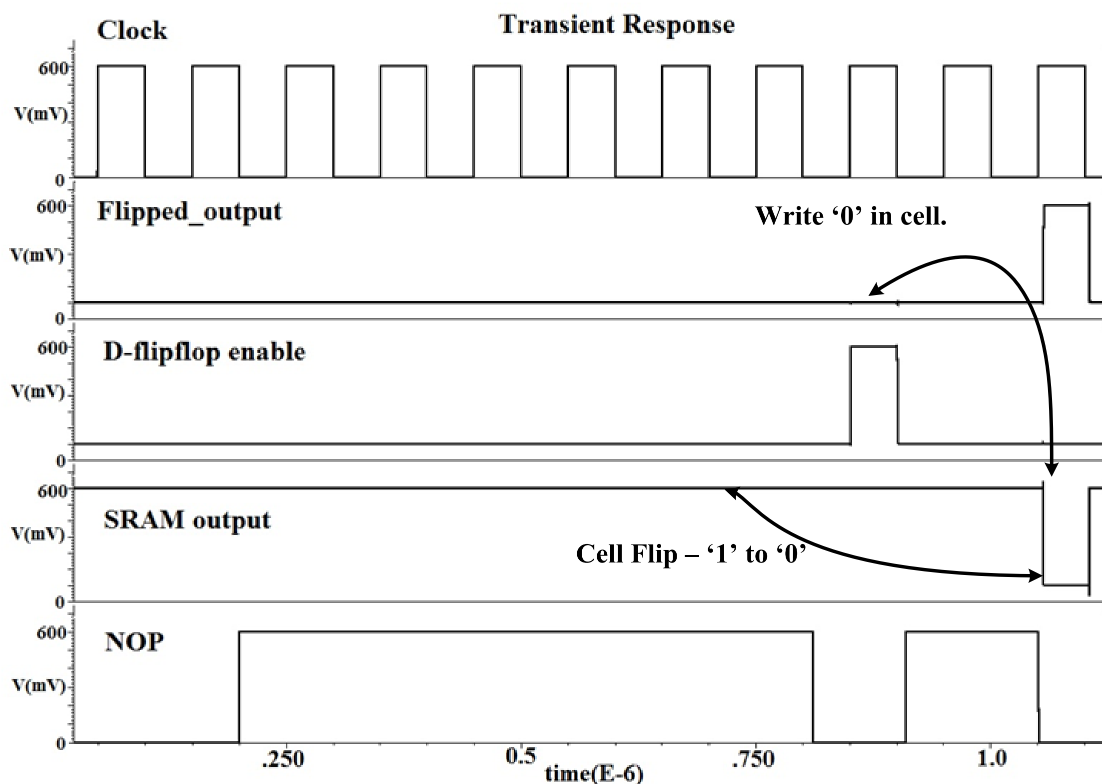

Figure 36: Cell flipping output

The Cell flipping setup (as explained above) is used to increase the skew in the SRAM cells so as to remove the variations at various power-ups. Figure 36 shows initially if SRAM cell is powering-up at '1', the cell can be flipped to '0' by using the setup described (in Figure 35).

SRAM is initially read out by giving an arbitrary address and the cell read out value was found to be '1'. This read out value is inverted and delayed using a Delay flip-flop. This flop enable is controlled by 'CLK' & 'WR' signals. So at the next clock edge when 'WR=1' enables the D flip-flop, an inverted value is written at the same address location. The same location is again read out and contents are found to be flipped.

In this case as pointed by arrows **SRAM Output** is flipping from '1' to '0'.

3.7 Proposed reliability and uniformity improvement methodology

In a nutshell, the proposed uniformity and reliability enhancement methodology for improved SRAM-PUF is as below:

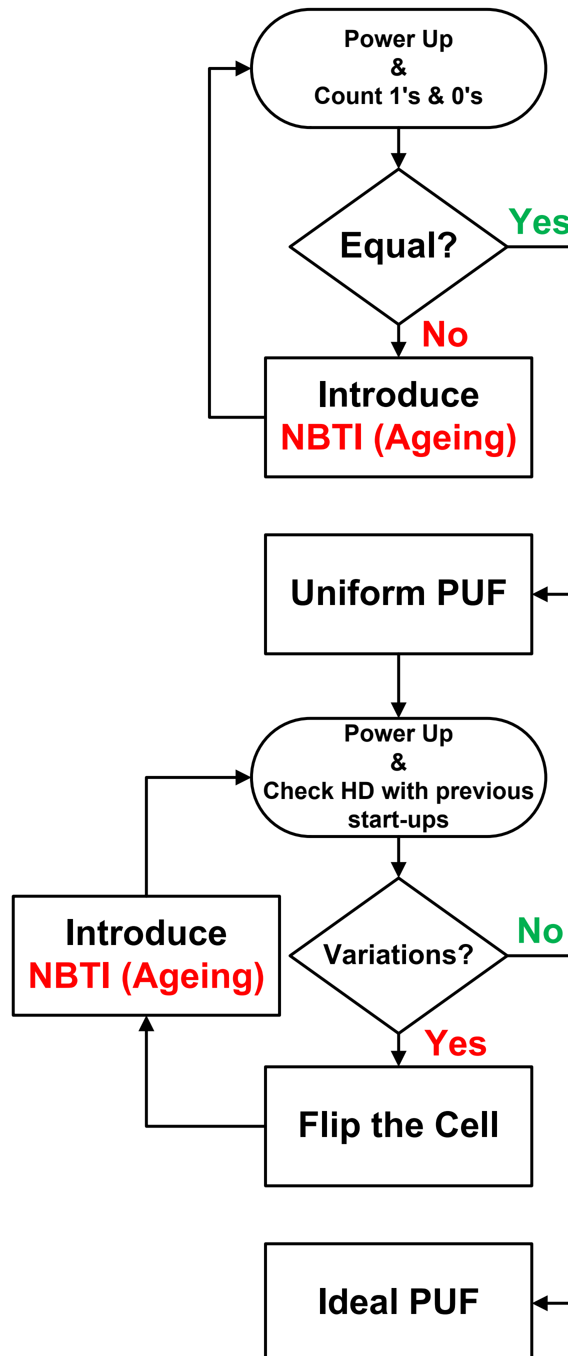

Figure 37: Proposed Reliability & Uniformity enhancing Methodology

In the next section, the proposed methodology is evaluated using statistical simulations for a SRAM cell array to see the impact on uniformity and reliability of SRAM-PUF.

3.8 Cell Flipping due to NBTI aging effects

The impact of NBTI aging on power-up values of SRAM-PUF is explained in this section. A single 6T- SRAM bit cell is used for simulations and results of this bit-cell can be extended to whole SRAM array as all cells will exhibit similar behaviour under the impact of aging. To simulate the effects of aging without changing the physical dimensions of transistors, we make use of increase in body-bias voltage to compensate the increase in threshold voltage for the transistor with same dimensions (length, width). The equation mentioned below gives the relationship between the body-bias voltage (V_{SB}) and change in threshold voltage of a transistor.

$$\Delta V_T = \gamma\left(\sqrt{(2\varphi + V_{SB})} - \sqrt{2\varphi}\right) \qquad (3.8.1)$$

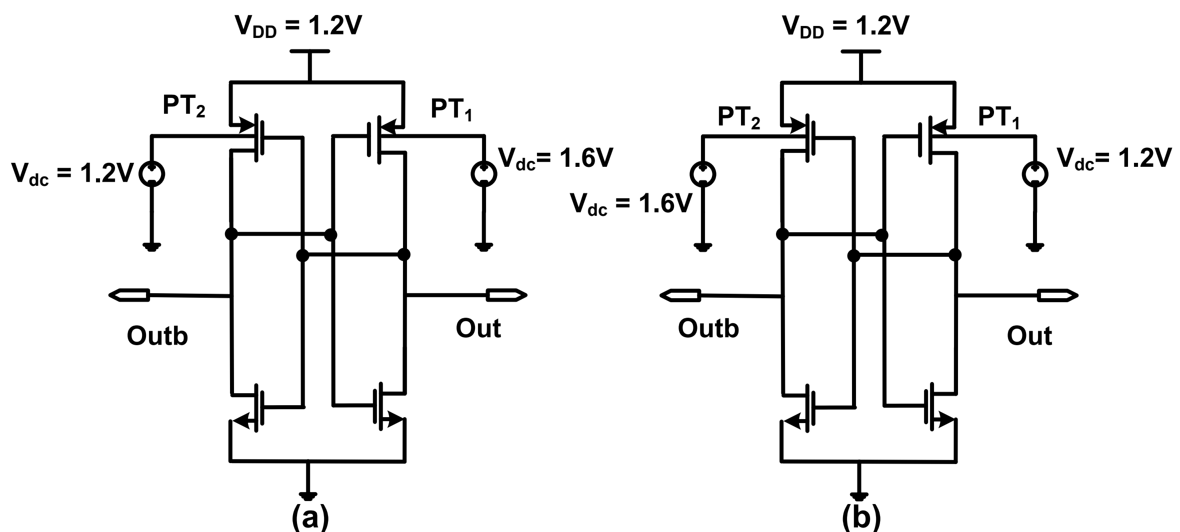

Figure 38: SRAM power-up value flipping due to aging

Figure 38, shows the simulation setup of raised threshold voltage corresponding to PMOS, PT_1, a body-bias of 1.6V is applied whereas for other PMOS, PT_2 the body terminal is tied to V_{DD}. Corresponding to Figure 38(a) the power-up pattern is as shown in Figure 39(a). The result can be explained as - with increase in body-bias voltage for PT_1, threshold voltage corresponding to PT increases making PT_1 weaker as compared to PT_2. So, at the power-up PT_2 will pull node Outb to '1' and node Out will get '0'. So, if we change the body bias voltage to opposite side, output behaviour at start-up will also flip because now PT_1 is stronger PMOS amongst two cross-coupled PMOS.

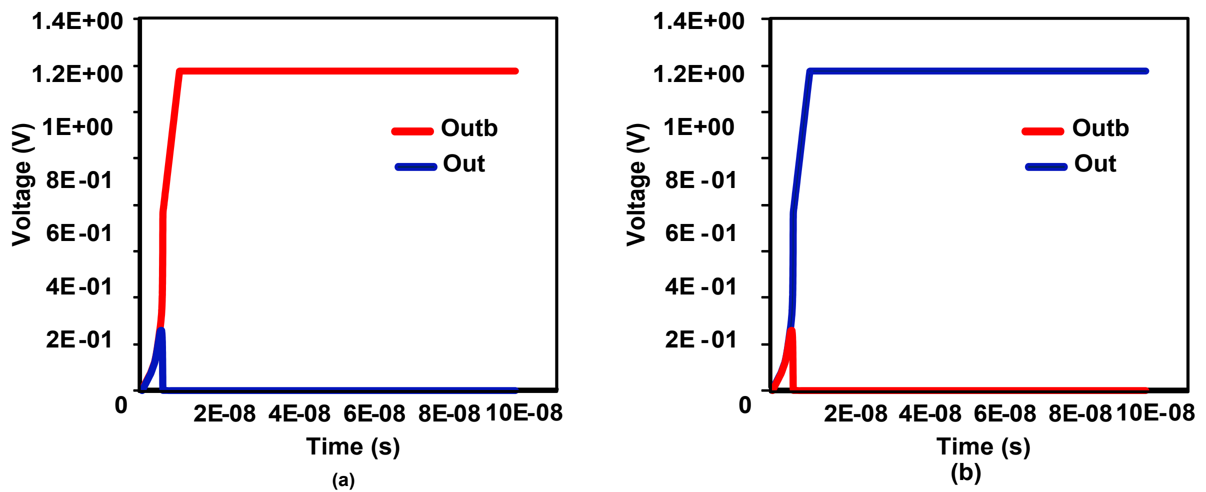

Figure 39: Cell flipping due to increased V_{TH}

From the above figure, we can safely conclude that power-up value of SRAM cell can be flipped using the action of aging (or increased V_{TH}). This result will be used further to make simulations for proposed methodologies.

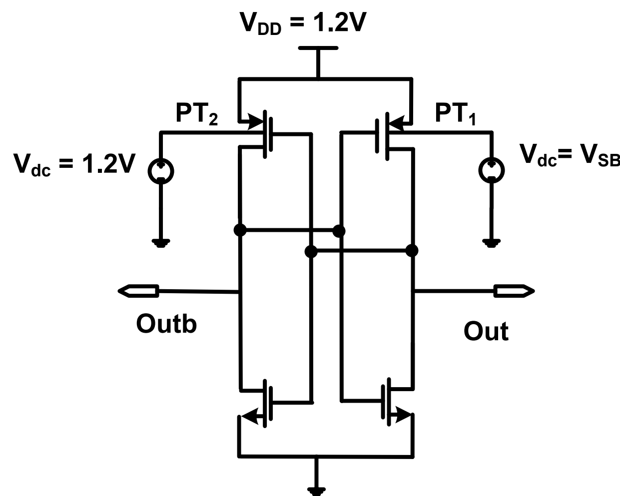

Figure 40: SRAM cell setup for statistical simulation

Figure 40, shows the SRAM cell used for the statistical simulation to see the impact of aging on uniformity and reliability. As already described in figure 39, if a SRAM cell is subjected to aging the power-up pattern of SRAM cell array may flip.

If we assume the initial output pattern for SRAM array is skewed to a particular direction then it can be made uniform by applying NBTI aging. Aging or change in sub-threshold voltage (V_{TH}) is proportional to the body-bias voltage (from 3.8.1). So, increase in the body-bias voltage is equivalent to aging a skewed SRAM cell(s).

Table 2, gives the statistical data value for skewed SRAM cells which power-up at 1 and 0 for 10000 monte-carlo statistical simulations. As, we can see initially SRAM output is highly skewed and after aging there is net change in the majority cells flipping to minority side. If aging continues, at certain point we will see a nearly balanced output (in terms of number of 1s and 0s in final output). Using the definition of Uniformity (Section 2.3.3), for aging in the above case.

Uniformity = 49.57 (~50%)

Table 2: Impact of increased threshold Voltage on output pattern

Aging (mV)	Cells powering up to '1'	Cells powering up to '0'
0	7253	2745
13.9	6615	3385
25.9	6021	3979
36.5	5518	4482
46	4957	5043

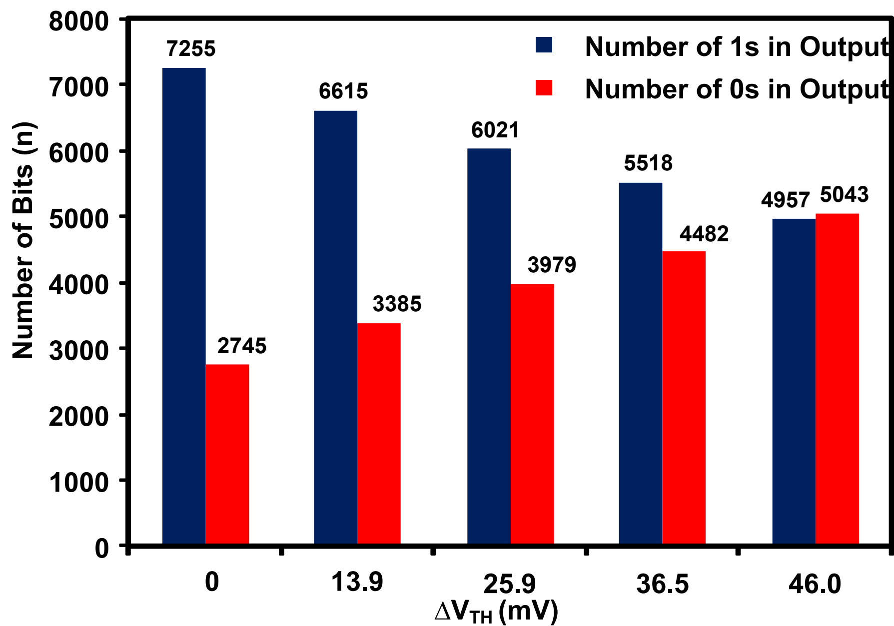

Figure 41: Impact of aging in maintaining Uniformity

Figure 41 gives the impact of aging on Uniformity of SRAM PUF output, aging helps in improving uniformity and results in a true random PUF.

In the proposed methodology after making the distribution nearly equal, the next step would be to reduce the variations at subsequent power-ups. To reduce the variations, skew should be increased using aging. For that, the inverted value of a cell(s) is written in the same cell

subsequent to the read cycle. The inverted value is stored temporarily in delay flip-flop and written back to the same cell at the next clock edge. After all the cells in the SRAM array are inverted, aging is applied to the cell(s) which will increase the skew in the desired direction as explained previously. Table 3 & Figure 42 give the trend of powering up of a SRAM cell to a particular value with increase in skew (ΔV_{TH}) value of corresponding PMOS transistors.

Table 3: Impact of aging on cell reliability

Body- Bias (V_{SB})	Out	Outb	PT_2 Threshold Voltage(mV)	PT_1 Threshold Voltage (mV)	Skew, ΔV_{TH} (mV)
1.3	4624	5376	446.5	456.3	9.8
1.6	3553	6446	446.5	475.2	28.7
1.8	3043	6955	446.5	485.2	38.7
2.0	2706	7292	446.5	493.3	46.8
2.4	2308	7689	446.5	504.5	58.0

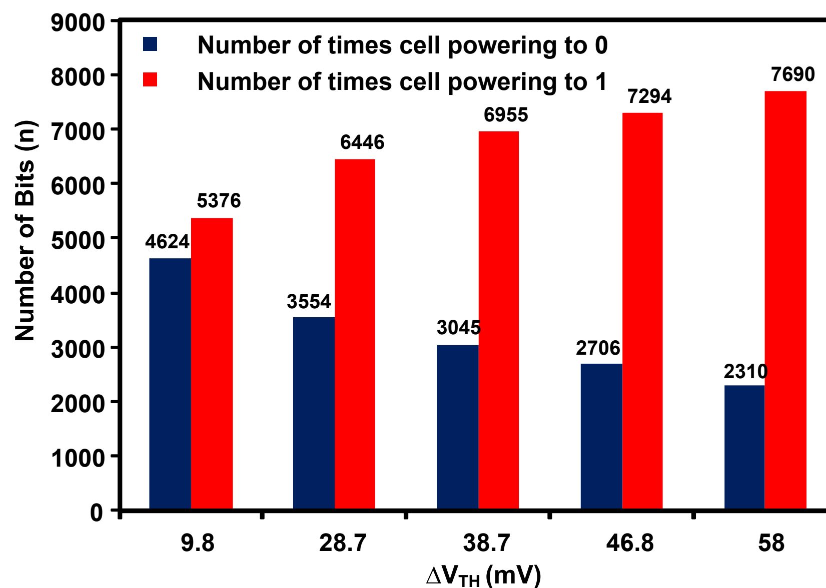

Figure 42: Impact of aging on reliability

From the above figure, it can be interpreted that with increased skew the reliability of an SRAM cell(s) powering-up to a particular value can be increased. For this particular case reliability can be enhanced by ~23%.

3.9 Layout using 65nm Global foundry technology library

Layout plays an important role in any efficient design under the given lithographic constraints and capabilities. We used standard design methods & techniques in our layout as shown in Figure 43 which gives block level view of SRAM PUF testchip.

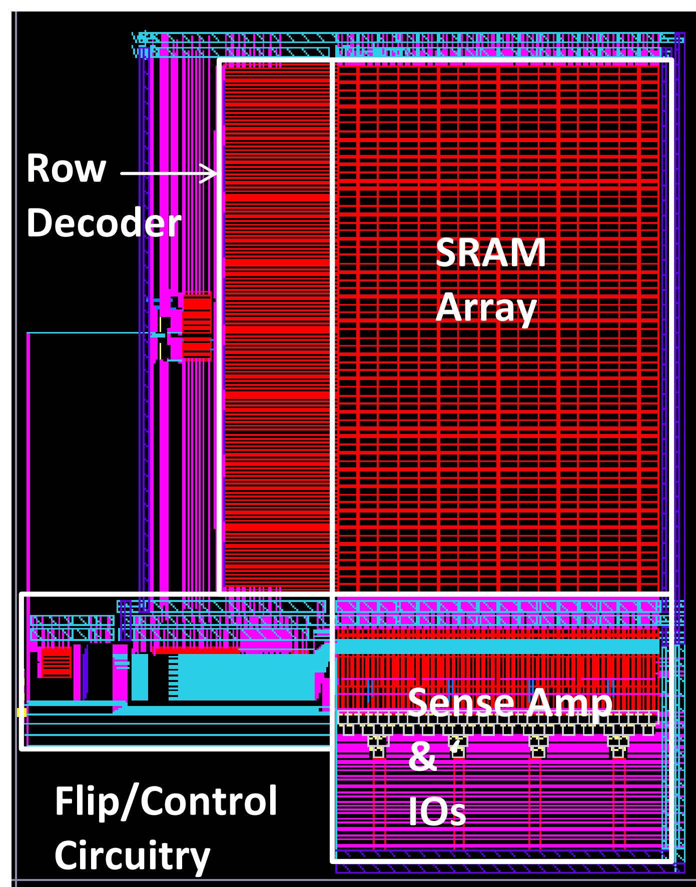

Figure 43: SRAM PUF testchip layout

SRAM-PUF layout is divided mainly into following blocks:

1. SRAM Array

2. Sense Amplifier & IOs
3. Flip/Control Circuitry
4. Row Decoder

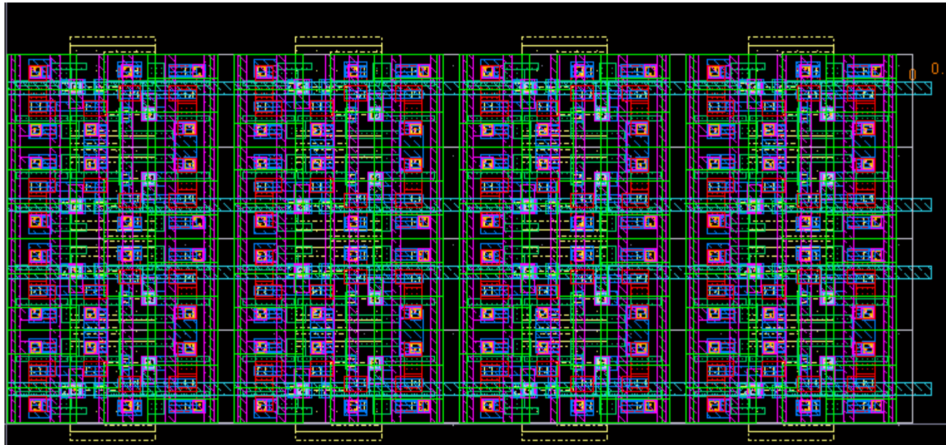

Figure 44: SRAM-PUF Cell layout 4x4

As shown in Figure 44, for SRAM cell layout wide-cell layout technique is implemented, which provides uniform orientation for the all the transistors, this provides better pattern reproducibility. Also, the wide cell layout can offer less word line resistance and shorter bitline per cell.

Section II: Proposed SRAM Energy Minimization Methodology

To conduct energy analysis on various SRAM array structures, it is very difficult to design and simulate for every SRAM array size. The proposed method to estimate energy for various arrays is to parameterize and model SRAM array which gives equivalent time delay, parasitic and associated energy results. Peripheral circuitry and input drivers are also parameterized in proportion to the load value to imitate the actual SRAM loading effects.

An 8kb SRAM sub-array structure for energy analysis is shown in Figure 45(a). The number of rows (k) and that of columns (j) can be changed while their product remains constant. In high performance applications, they have been mainly selected to meet the system performance requirement. However, in the SRAMs for ultra-low power applications, the array structures are limited by additional design parameter such as cell stability, read bitline sensing margin, and leakage current. A bitline structure employing the conventional 8T SRAM cell (Figure 45(c)) and including related parameters in the wordlines and the bitlines is illustrated in Figure 45(b). 8T cells are used for energy estimation because of immunity of 8T cells over 6T cells in the sub-threshold region. Near to sub-threshold or below threshold voltage 6T cells face many read failures, so to avoid them we preferred to use 8T cell. However, the results are valid for 6T cells also.

3.10 SRAM Energy Model

In this sub-section, we will model the energy components in the SRAM sub-array structure in Figure 45[34]. Using the parameters in Figure 45(b), the read energy (E_{total_read}) and the write energy (E_{total_write}) of the SRAM sub-array can be written as

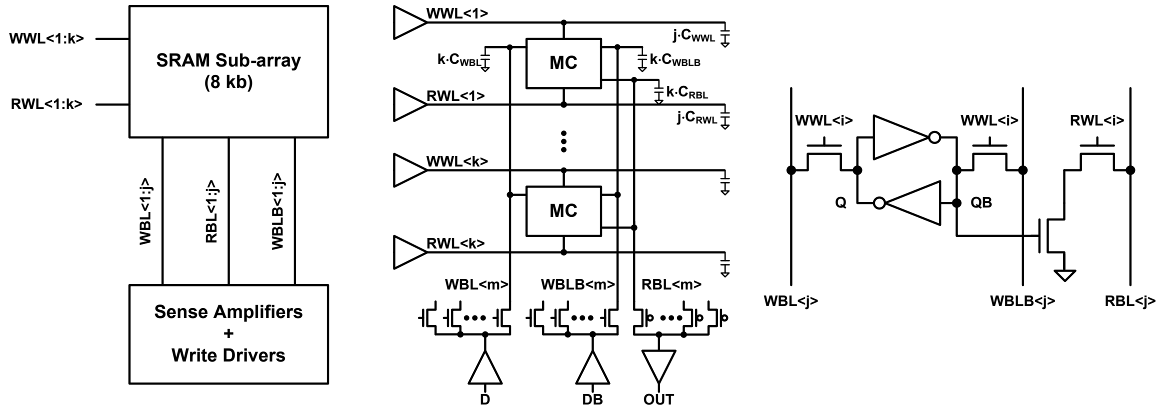

Figure 45: (a) An 8kb SRAM sub-array for energy analysis. Note that k X j is 8kb. (b) Bitline structure of the SRAM sub-array in (a). (c) Schematic of the conventional 8T SRAM cell used in this work

$$E_{total_read} = j \times k \times I_{l_cell} \times t \times V_{DD} + (j \times C_{RWL} \times V_{DD}^2 + k \times 0.5 C_{RBL} \times V_{DD}^2) \quad 3.10.1$$

and

$$E_{total_write} = j \times k \times I_{l_cell} \times t \times V_{DD} + (j \times C_{WWL} \times V_{DD}^2 + k \times C_{WBL} \times V_{DD}^2). \quad 3.10.2$$

Here, k is the number of rows, j is the number of columns, I_{l_cell} is the leakage current in an SRAM cell, t is the cycle time (T_{cyc}) of the SRAM, C_{RWL} is the read wordline capacitance per cell, C_{WBL} is the write bitline capacitance per cell, and V_{DD} is the supply voltage. In the read energy equation, it is assumed that the probabilities of data ‘1’ and that of data ‘0’ are equal and are 0.5. The dynamic energy component is mainly determined by the wordline capacitance and the bitline capacitance while the static energy component coming from the leakage current is determined by the memory density. The effects of the read and write operations on the leakage current of the accessed row and column are insignificant. Thus, they are neglected in the energy estimation. E_{total_read} and E_{total_write} can be merged into the total SRAM energy (E_{total}) by including the probability of the read operation (P_r) and that of the write operation (P_w), which is given by

$$
\begin{aligned}
E_{total} &= j \times k \times I_{l_cell} \times t \times V_{DD} \\
&+ P_r(j \times C_{RWL} \times V_{DD}{}^{2} + k \times 0.5C_{RBL} \times V_{DD}{}^{2}) \\
&+ P_w(j \times C_{WWL} \times V_{DD}{}^{2} + k \times C_{WBL} \times V_{DD}{}^{2}).
\end{aligned} \qquad 3.10.3
$$

As shown in (3.10.3), SRAM energy is a function of multiple variables such as supply voltage, capacitance, performance, temperature, workload, and organization. SRAM energy minimization has to be conducted while considering all the above components carefully.

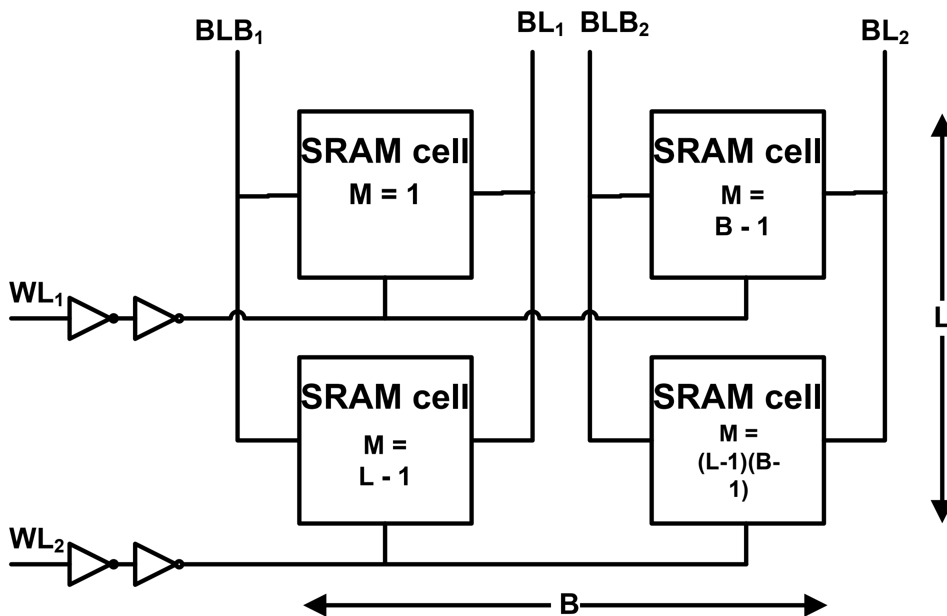

Figure 46: SRAM modelling for energy estimation

Setup for SRAM energy estimation is shown in Figure 46. The above setup will give an estimation of total energy consumed during SRAM read operation. For simplicity in calculations, SRAM array is modelled using SRAM cells having different multiplication (m-

factor in SPICE) factor. If we assume length of array (= number of rows) as **L** and breadth of array (= number of columns) as **B,** then an equivalent of SRAM array of **L**x**B** cells can be simulated using this setup. Basically, the loading capacitance will be almost same and can be used for energy estimation using different combinations of rows and columns (L and B), keeping cell density (=LxB) constant. Total Energy calculations are made using equations mentioned below using values from simulation.

$$Total\ Energy = Static\ Energy + Dynamic\ Energy \quad 3.10.4$$

$$where, Static\ Energy = V_{dd} * I_{leakage} * T_{total} \quad 3.10.5$$

$$Dynamic\ Energy = C_{Total} * V_{dd}^{2} \quad 3.10.6$$

For *Static Energy,* calculations at a fixed V_{dd}, $I_{leakage}$ can be calculated by switching all the signals 'off' and by estimating the leakage current from the voltage source. Read delay (T_{total}) can be calculated using the delay between charging 50% of V_{dd} wordline (WL_1, WL_2) and Bitline (BL_1 & BL_2) discharging by 50% of V_{dd}.

For *Dynamic Energy* at fixed supply voltage, C_{Total} can be calculated using summation of bitline capacitance & wordline capacitance using HSPICE simulations.

3.12 Energy Efficiency Dependency on SRAM Array Structure.

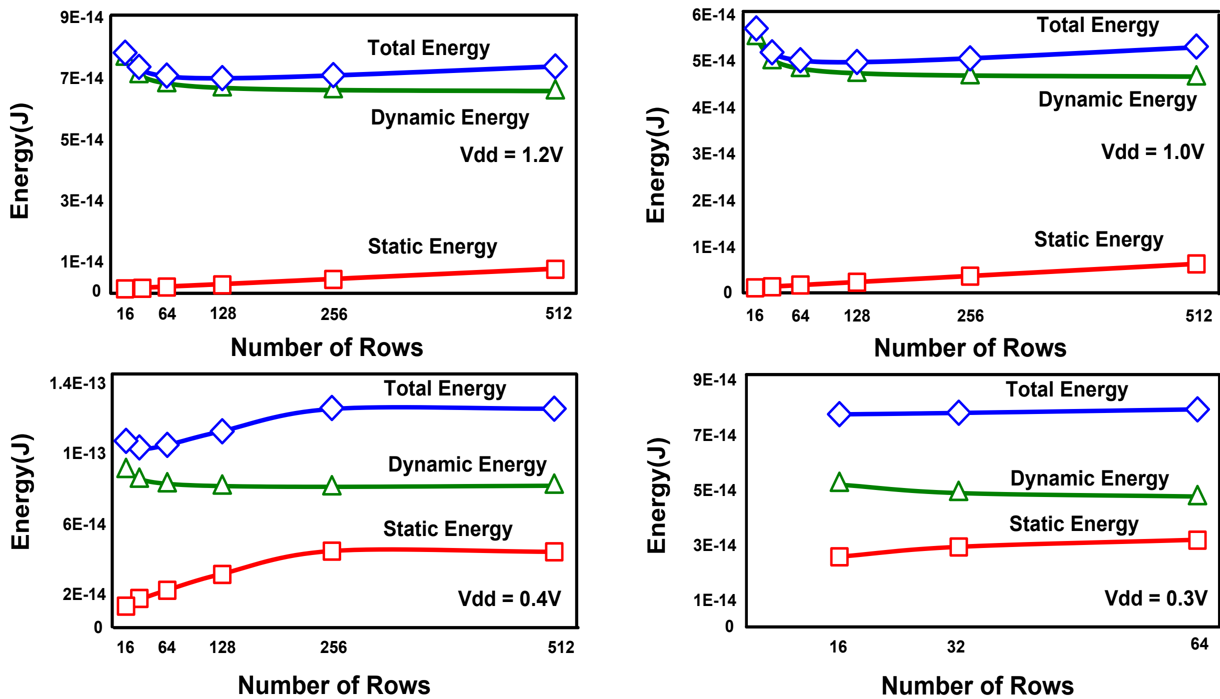

Figure 47: Energy consumption of a 8kb SRAM sub-array over various number of rows. Simulation results show that energy consumption is substantially affected by SRAM topologies. (a) VDD = 1.2 V (b) VDD = 1.0 V (c) VDD = 0.4 V (d) VDD = 0.3 V

Previous studies reveal that a SRAM array can be energy optimized by making the number of rows more than the number of columns i.e. tall array structures[26]. In contrast, our experiment reveals that the traditional (taller array) hypothesis is valid for SRAM operating at higher supply voltages (> 0.7V) and for SRAMs operating at smaller supply voltages, wide arrays (less rows, more columns) give the most energy-optimized results. The above Figure shows the impacts on total energy of SRAM array with changing supply voltage and array structure. As shown, Figure 47(a), (b), (c), (d) gives the total energy results for supply voltage values = 1.2V, 1.0V, 0.4V, 0.3V respectively having SRAM array structure with constant cell density for all cases. It can be seen that the difference in dynamic energy & static energy at

higher supply voltage is in orders of 10 whereas at lower supply voltage the difference is reduced to comparable values This reduced difference value enables static energy to play an important role in determining total energy of SRAM. Since static energy very much depends upon read latency & leakage current, array structure having less read latency & leakage current should be preferred. Also, at ultra-low voltages only bitlines in wider structures are able to discharge for obvious reason of less number of cells per bitline. As can be seen from Figure 46 (d) at 0.3V only the wider structures are valid for a fixed density SRAM, array structures having more rows than 64 at this supply voltage fail to discharge the bitline in the specific read time cycle. The following table show us the optimized array structure parameters for given supply voltage values as shown in Figure 48.

Table 4: Optimized energy structure configuration(s) Vs Supply voltage.

Supply Voltage (V)	Rows (N)	Columns (N)
1.2	128	64
1.0	128	64
0.4	32	256
0.3	16	512

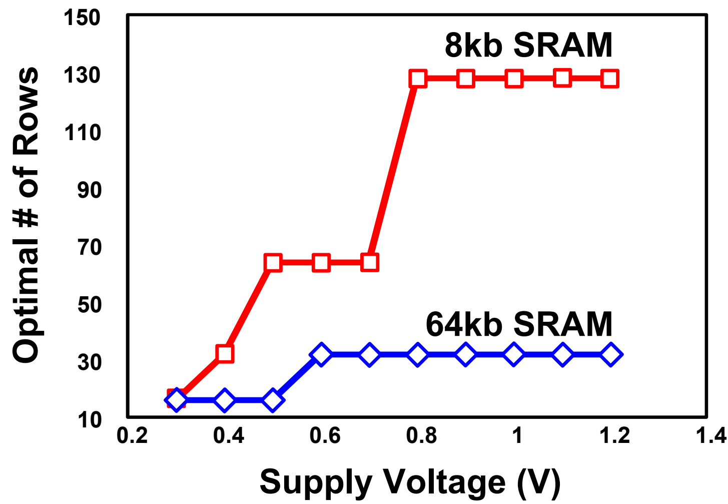

Figure 48: Optimal numbers of rows VS supply voltage for Energy Efficient array structure

The above figure describes the trend for energy-optimized array structure at different supply voltage values for a fixed density SRAM. As shown for higher supply voltage values, optimized structure corresponds to 128 rows (64 columns, for 8kb SRAM). For lower supply voltages, the trend is reversed and energy optimized structure corresponds to 32 rows (256 columns, 8kb SRAM) at V_{dd} = 0.4V, 64 rows (128 columns, 8kb SRAM) at V_{dd} = 0.5, 0.6, 0.7V.

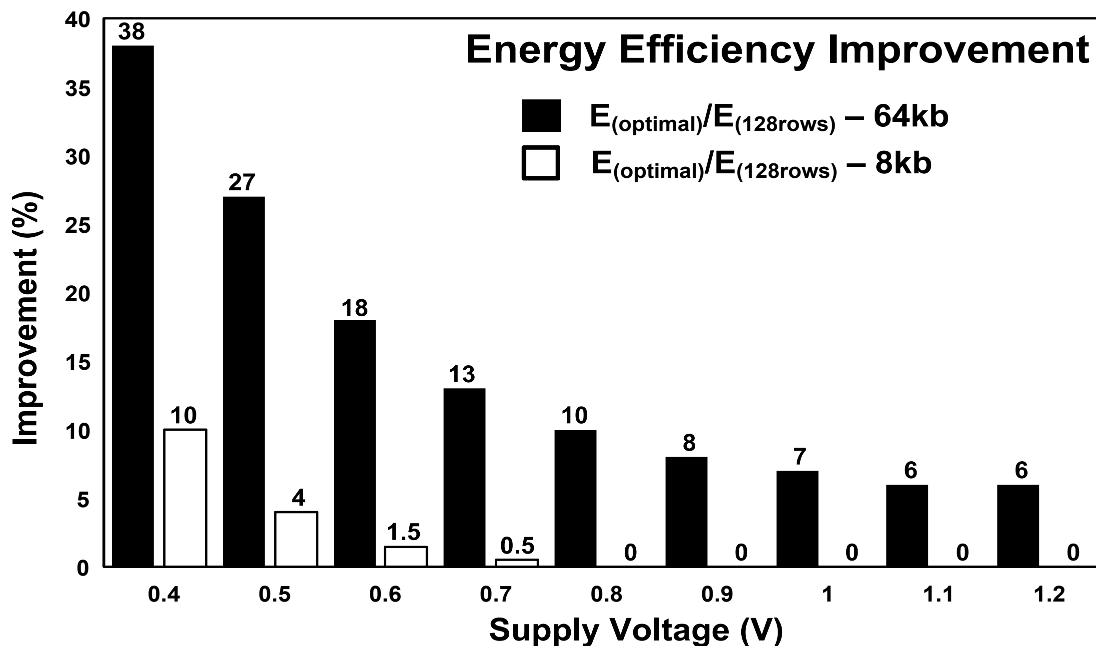

Figure 49: Percentage change in the energy using optimal rows over 128 rows

Figure 49 depicts percentage change in total energy for array structure using optimum rows and minimum rows (=16) with respect to maximum rows (=512) for a fixed density SRAM. Optimum rows are the number of rows corresponding to the array structures showing minimum total energy at the particular supply voltage as shown in Figure 48 for different supply voltages. From the bar graph it is clear that energy gain up to 23% can be achieved using optimal number of rows in an array structure operating at near to sub-threshold region.

Similarly the energy efficiency enhancement corresponding to tallest (rows = 512) and widest (rows =16) array structures over the range of supply voltage values. The results show the same behaviour as expected. At lower supply voltages wider structure can be 18% more energy efficient than the tallest structure. Also, at higher supply voltages taller structure is more energy efficient following the explanation from traditional guide.

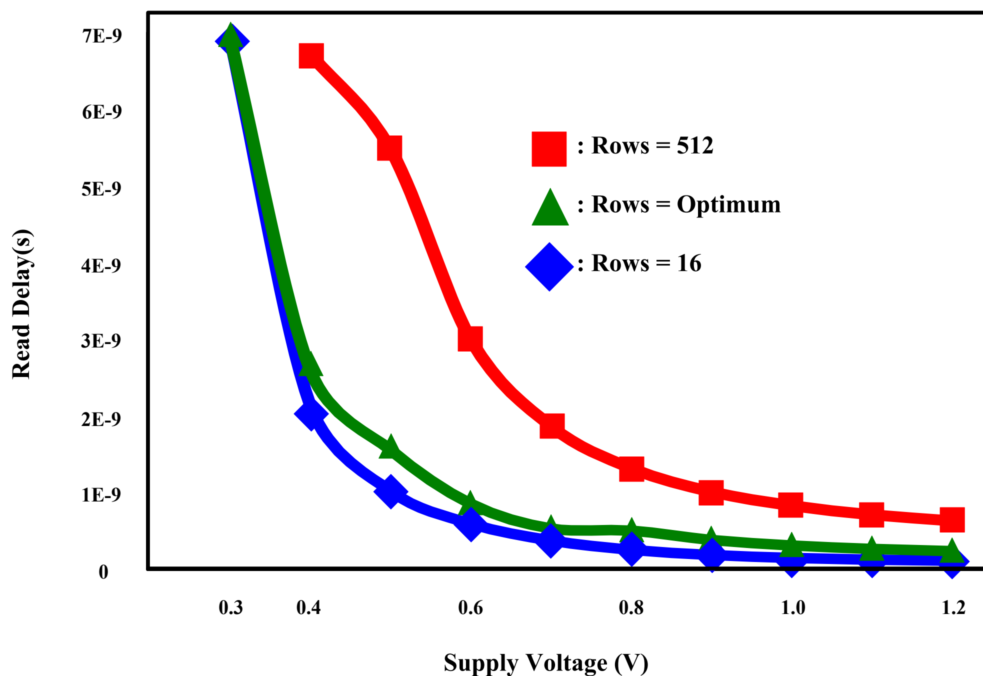

Figure 50: Read delay variation with rows at various supply voltages

Figure 50 reveals the exact reason for minimum energy corresponding to the minimum number of rows. For an 8kb SRAM we used the optimum number of rows from Figure 48 and plot the read delay corresponding to those rows.

Read delay is actually the time required for bitline to discharge from VDD to VDD/2.

For a particular bitline, parasitic are due to inter-connect length and due to loading of pass-transistors which connects SRAM cell to the bitline. So, if the number of devices on the bitline is reduced, there would be fewer devices per bitline which means reduction in the corresponding total bitline capacitance. Hence, the delay corresponding to array structure with less number of rows is less in comparison to taller arrays.

3.13 Impact of Device Variations on SRAM Array Structures for Energy Minimization

The adoption of minimum or near-minimum devices aggravates the device's current deviation along with various design parameters including energy consumption. In this sub-section, we will investigate the impact of device variations on the minimum-energy-driven SRAM array structures. We performed statistical simulation with 1000 sample runs for 8kb SRAM. Figure 51-53 illustrate the statistical distribution of the SRAM total energy at the supply voltage of 0.4 V, 0.6 V and 1.2 V respectively using the 8kb SRAM. To verify the effectiveness of the proposed idea, various array structures are evaluated. The numbers of rows explained in Figure 48 are used as the optimal array structures to be compared other array structures voltage (e.g. 32 rows at 0.4 V and 64 rows at 0.6 V). Simulation results reveal that wider array structures provide higher energy efficiencies at low supply voltage (Figure 49 and Figure 50), which corresponds to the results in Figure 47 and 48. At the same time, the proposed optimal structures produce the lowest sigma values at 0.4 V and 0.6 V which mean less spread of energy values from the mean value of energy for wide array structure.

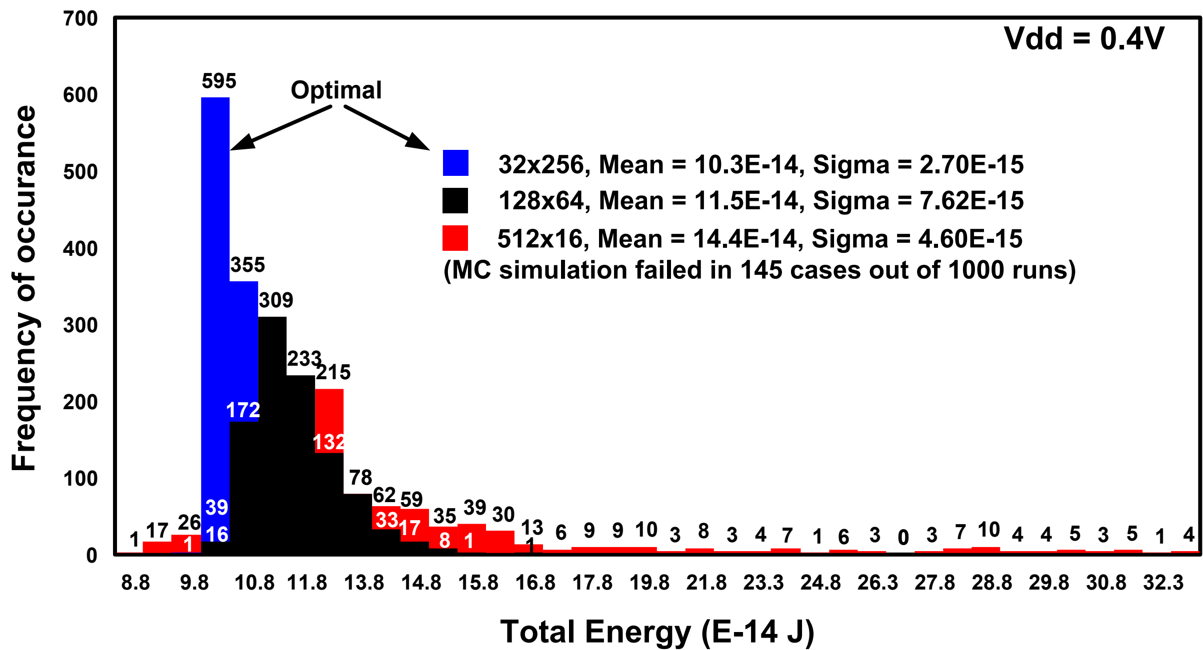

Figure 51: Statistical distribution of Total Energy at VDD = 0.4V

Note that read failures occurred at 0.4 V when 512 rows are used. At VDD = 1.2 V (Figure 53), the optimal structure shows the smallest mean energy value. However, the energy variation of the structure is not the smallest. Smaller energy variations can be obtained by lowering the number of cells per bitline beyond the optimal value, which will result in higher mean energy as illustrated by the red graph in Figure 53.

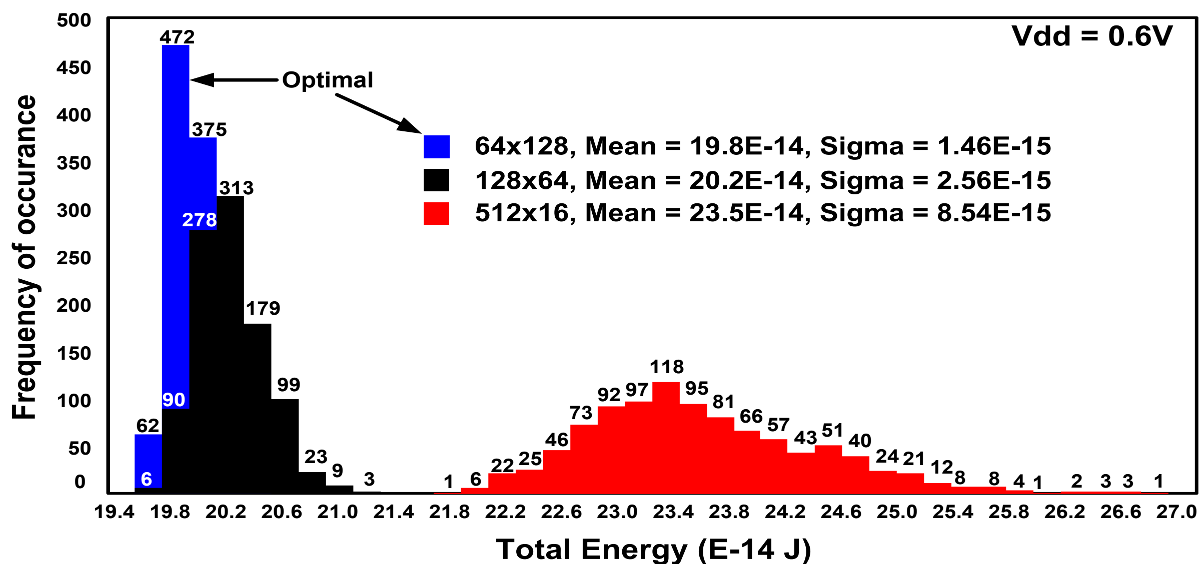

Figure 52: Statistical distribution of Total Energy at VDD = 0.6V

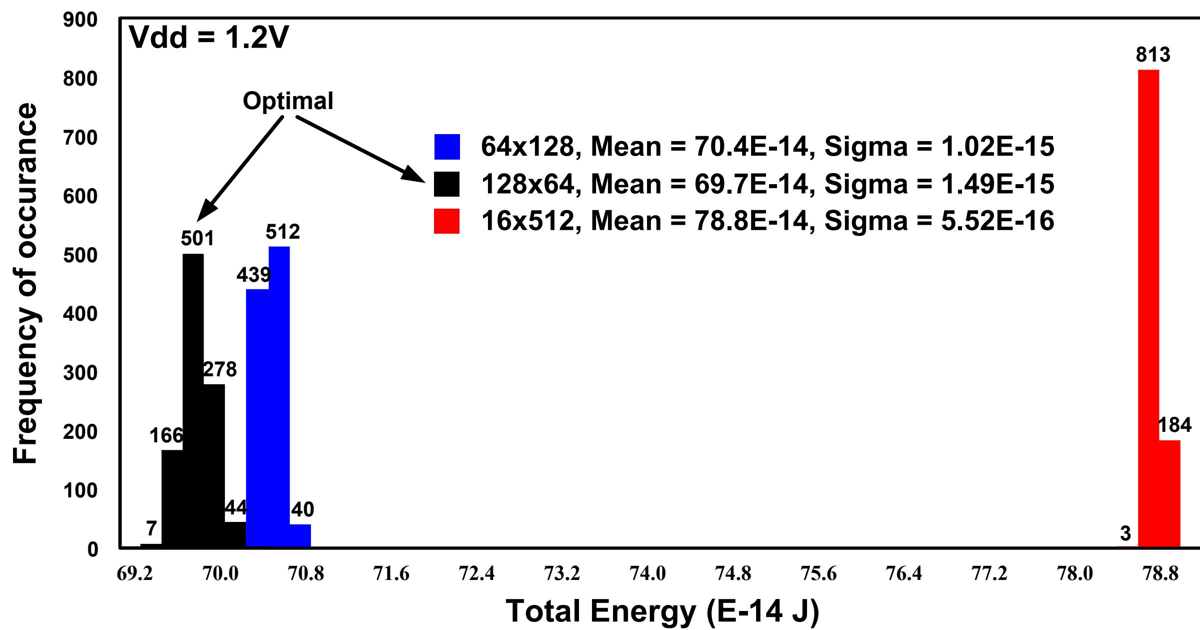

Figure 53: Statistical distribution of Total Energy at VDD = 1.2V

The above three figures reveals the results corresponding to statistical simulation of SRAM array energy. Mean value represents the most dominant value out of 1000 values for that particular array configuration and sigma represents the spread of value around that mean. From the results, we can inferthat sigma spread corresponding to wider array structures is also small as we can expect since the number of devices are less so the total variations also follow the same behaviour. Also, near to sub-threshold voltage for taller arrays some of the monte carlo runs fail to converge because of read failures.

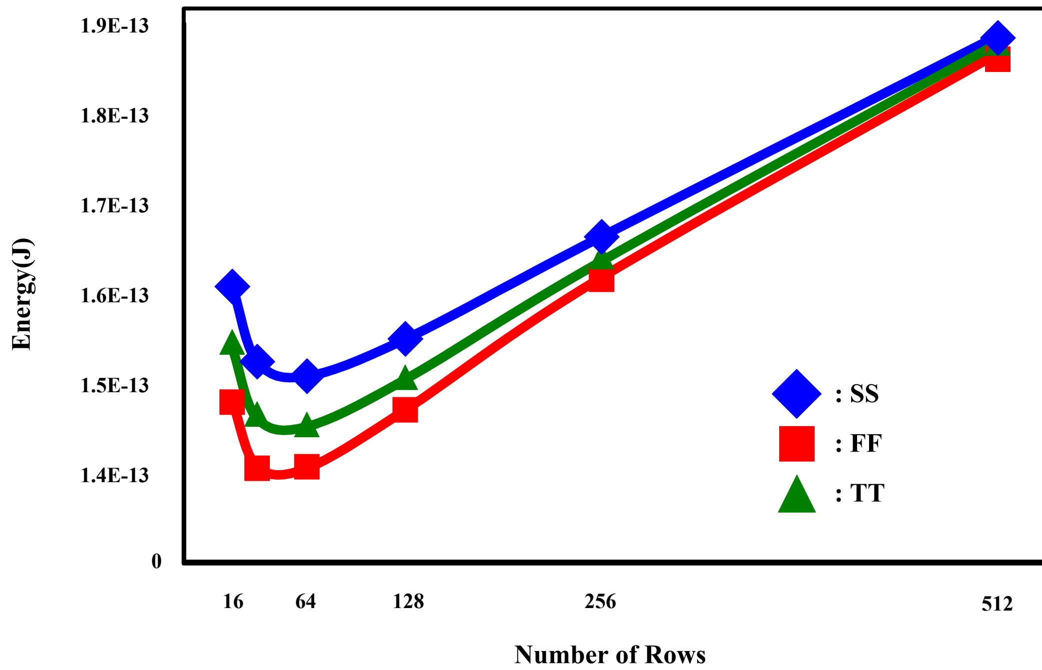

Figure 54: Corner Simulations for SRAM

Figure 54 depicts the corner simulation for energy minimization. Fast-NMOS, fast-PMOS devices give the most favourable energy results when plotted against the number of rows in comparison to slow-slow and typical-typical devices. The explanation for this can be attributed to the minimum read delay corresponding to FF devices in comparison to their SS and TT counterparts for the same supply voltage.

Chapter 4: Summary & Future Work

4.1 Summary

In the current age of technological advancements, electronics are playing a crucial part in our daily lives. They carry a lot of personal information that needs to be protected from misuse. Also, semiconductor IP industry that is under constant threat of infringement, stealing of IP designs and over-production needs some sort of security ring that is difficult to break. In addition to having a strong security feature, the system is expected to be affordable, stable under varying environmental conditions and can be produced in large quantities.

Physical Unclonable Functions (PUF) is the latest circuit based security systems which hold the potential to qualify for all the expectations mentioned above. These systems generate unique and reproducible keys based on the intrinsic properties of silicon where input-output pair is called as Challenge-Response Pair only when keys are required. Cloning these devices is also very difficult, even for the manufacturer of PUF it is difficult to produce exact replica of any device. The unclonability of PUFs is guaranteed by the fact that the interaction between input challenge and silicon device is totally random and difficult to be replicated. Additionally, physical properties (length, width & oxide layer width) of the silicon devices are different for different devices and are random due to inefficient fabrication process.

There are many kinds of PUF devices such as - Optical PUF, Coating PUF, Delay based PUF, SRAM PUF and Butterfly PUF. SRAM PUF is selected as our candidate for ideal PUF since it is the most reliable and stable PUF. It uses power-up as input challenge and give out power-up value as output response. On the basis of skew values, cells are classified in two categories - partially skewed & fully skewed. Fully skewed are those cells that show consistent power-up value irrespective of environmental variations whereas partially skewed

cells can flip under environmental variations. So the roadblock for SRAM-PUF to generate consistent pattern each time are these partially skewed cells.

This thesis proposes to make the use of Aging (NBTI stress) to imp rove the uniformity and reliability. There are two important concepts about robust security system that need to be understood - **Uniformity** and **Reliability**. Uniformity is defined as fair distribution of 1's and 0's in the power-up pattern, to ensure the robustness of security system it should have high uniformity. For a truly random PUF, a high degree of uniformity is required. The SRAM PUF power-up pattern is expected to be consistent at every power-up so as to remove any variations in the security key. To make a cell more consistent, it is required to increase the skew in the cell. To increase the skew in the cell, a cell-flipping methodology is required after which cell is exposed to NBTI stress. This post fabrication technique can help in designing robust SRAM PUF with less variation due to environmental fluctuations. Also with technology advancements, we expect the deviation from Ideal requirements will increase. Hence, this post fabrication modification can help in improving robustness of SRAM-PUF to a large extent. The proposed methodology can help improving reliability by 23% and uniformity to nearly 50%.

In addition to security modern SoCs are expected to be energy efficient at near sub threshold region. We tried to investigate the role of array structures in determining the total energy of SRAM. From our analysis we can conclude safely that if SRAM is operating near the sub-threshold region, wider array structures (less number of cells/bitline than columns) are more energy efficient compared to taller array structures. Also, for wider array structures the impact of device variations is minimal as compared to the taller structure. Finally, for low power SRAMs operating near the threshold voltage; energy efficiency can be improved up to 23% by changing the array structure from taller array to wider array.

4.2 Future Work

Looking forward, we plan to do test chip analysis of 6T SRAM-PUF chip with in-built flip mechanism using 65nm technology library. As explained we will be using NBTI stress to change the skew in SRAM cells. A real time setup and calculations will be made using logic analyzer.

Based on learning experience from this work, we would like to make few recommendations for further research in this area -

1. On-chip monitor for keeping track of uniformity and reliability for automatic aging control mechanism.

2. Impact of proposed reliability and uniformity improvement methodology on advanced CMOS technologies (45nm, 28nm and beyond).

References

[1] I. Spectrum. [Online]. Available: http://spectrum.ieee.org/computing/hardware/counterfeit-chips-on-the-rise.

[2] D. Lim, J. W. Lee, B. Gassend, G. E. Suh, M. Van Dijk, and S. Devadas, "Extracting secret keys from integrated circuits," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on,* vol. 13, pp. 1200-1205, 2005.

[3] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *Proceedings of the 44th annual Design Automation Conference*, 2007, pp. 9-14.

[4] W.-I. Security. [Online]. Available: http://en.wikipedia.org/wiki/Information_security

[5] A. Ollier, *La cryptographie militaire avant la guerre de 1914*: Lavauzelle, 2002.

[6] S. P. Skorobogatov, "Semi-invasive attacks-a new approach to hardware security analysis," *Technical report, University of Cambridge, Computer Laboratory,* 2005.

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

[8] B. L. Gassend, "Physical random functions," Massachusetts Institute of Technology, 2003.

[9] R. Maes. [Online]. Available: http://homes.esat.kuleuven.be/~rmaes/puf.html

[10] J. Guajardo, S. S. Kumar, G.-J. Schrijen, and P. Tuyls, "Physical unclonable functions and public-key crypto for FPGA IP protection," in *Field Programmable Logic and Applications, 2007. FPL 2007. International Conference on*, 2007, pp. 189-195.

[11] A. J. Bhavnagarwala, X. Tang, and J. D. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," *Solid-State Circuits, IEEE Journal of,* vol. 36, pp. 658-665, 2001.

[12] C. Bösch, J. Guajardo, A.-R. Sadeghi, J. Shokrollahi, and P. Tuyls, "Efficient helper data key extractor on FPGAs," in *Cryptographic Hardware and Embedded Systems–CHES 2008*, ed: Springer, 2008, pp. 181-197.

[13] R. Maes, P. Tuyls, and I. Verbauwhede, "A soft decision helper data algorithm for SRAM PUFs," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, 2009, pp. 2101-2105.

[14] M. Hofer and C. Boehm, "An alternative to error correction for sram-like pufs," in *Cryptographic Hardware and Embedded Systems, CHES 2010*, ed: Springer, 2010, pp. 335-350.

[15] A. R. Krishna, S. Narasimhan, X. Wang, and S. Bhunia, "MECCA: a robust low-overhead PUF using embedded memory array," in *Cryptographic Hardware and Embedded Systems–CHES 2011*, ed: Springer, 2011, pp. 407-420.

[16] H. Fujiwara, M. Yabuuchi, H. Nakano, H. Kawai, K. Nii, and K. Arimoto, "A chip-ID generating circuit for dependable LSI using random address errors on embedded SRAM and on-chip memory BIST," in *VLSI Circuits (VLSIC), 2011 Symposium on*, 2011, pp. 76-77.

[17] D. E. Holcomb, A. Rahmati, M. Salajegheh, W. P. Burleson, and K. Fu, "DRV-Fingerprinting: using data retention voltage of SRAM cells for chip identification," in *Radio Frequency Identification. Security and Privacy Issues*, ed: Springer, 2013, pp. 165-179.

[18] A. Maiti, V. Gunreddy, and P. Schaumont, "A systematic method to evaluate and compare the performance of physical unclonable functions," in *Embedded Systems Design with FPGAs*, ed: Springer, 2013, pp. 245-267.

[19] A. J. Cover and M. T. Thomas, *Elements of information theory*: Wiley, 1991.

[20] A. Dagar, "Modeling SRAM Power-up Characteristics For Physical Unclonable Functions," MSc Thesis, Delft university of Technology, 2011.

[21] C.-I. Kim, H. Soeleman, and K. Roy, "Ultra-low-power DLMS adaptive filter for hearing aid applications," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on,* vol. 11, pp. 1058-1067, 2003.

[22] S. Cserveny, L. Sumanen, J.-M. Masgonty, and C. Piguet, "Locally switched and limited source-body bias and other leakage reduction techniques for a low-power embedded SRAM," *Circuits and Systems II: Express Briefs, IEEE Transactions on,* vol. 52, pp. 636-640, 2005.

[23] B. H. Calhoun and A. P. Chandrakasan, "A 256-kb 65-nm sub-threshold SRAM design for ultra-low-voltage operation," *Solid-State Circuits, IEEE Journal of,* vol. 42, pp. 680-688, 2007.

[24] M. Yamaoka, N. Maeda, Y. Shinozaki, Y. Shimazaki, K. Nii, S. Shimada*, et al.*, "90-nm process-variation adaptive embedded SRAM modules with power-line-floating write technique," *Solid-State Circuits, IEEE Journal of,* vol. 41, pp. 705-711, 2006.

[25] T.-H. Kim, J. Liu, J. Keane, and C. H. Kim, "A 0.2 V, 480 kb subthreshold SRAM with 1 k cells per bitline for ultra-low-voltage computing," *Solid-State Circuits, IEEE Journal of,* vol. 43, pp. 518-529, 2008.

[26] R. J. Evans and P. D. Franzon, "Energy consumption modeling and optimization for SRAM's," *Solid-State Circuits, IEEE Journal of,* vol. 30, pp. 571-579, 1995.

[27] B. S. Amrutur, "Design and analysis of fast low power SRAMs," Stanford University, 1999.

[28] A. Karandikar and K. K. Parhi, "Low power SRAM design using hierarchical divided bit-line approach," in *Computer Design: VLSI in Computers and Processors, 1998. ICCD'98. Proceedings. International Conference on*, 1998, pp. 82-88.

[29] A. Pavlov and M. Sachdev, *CMOS SRAM circuit design and parametric test in nano-scaled technologies: process-aware SRAM design and test* vol. 40: Springer, 2008.

[30] H. Qin, Y. Cao, D. Markovic, A. Vladimirescu, and J. Rabaey, "SRAM leakage suppression by minimizing standby supply voltage," in *5th International Symposium on Quality Electronic Design, 2004. Proceedings.* , 2004, pp. 55-60.

[31] D. E. Holcomb, W. P. Burleson, and K. Fu, "Power-up SRAM state as an identifying fingerprint and source of true random numbers," *Computers, IEEE Transactions on,* vol. 58, pp. 1198-1210, 2009.

[32] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *Journal of Applied Physics,* vol. 94, pp. 1-18, 2003.

[33] S. V. Kumar, K. Kim, and S. S. Sapatnekar, "Impact of NBTI on SRAM read stability and design for reliability," in *Quality Electronic Design, 2006. ISQED'06. 7th International Symposium on*, 2006, pp. 6 pp.-218.

[34] A. Garg and T. T. H. Kim, "SRAM Array Structures for Energy Efficiency Enhancement," *Circuits and Systems II: Express Briefs, IEEE Transactions on,* vol. 60, pp. 351-355, 2013.