Secure and Reliable In-Memory Computing for Edge AI

M.Tech. Thesis

By

SHIVAM VAISH

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

MAY 2025

Secure and Reliable In-Memory Computing for Edge AI

A Thesis

Submitted in partial fulfillment of the

requirements for the award of the degree

of

M.Tech. Thesis

by

SHIVAM VAISH

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

MAY 2025

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Secure and Reliable In-Memory Computing for Edge AI** in the partial fulfillment of the requirements for the award of the degree of **Master of Technology - VLSI Design and Nanoelectronics** and submitted in the **Department of Electrical Engineering, Indian Institute of Technology Indore,** is an authentic record of my own work carried out during the time period from July 2023 to May 2025 under the supervision of Dr. Santosh Kumar Vishvakarma Professor, Indian Institute of Technology Indore, Indore, India.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Shivam

(19/05/2025)

Signature of the Student with Date

(Shivam Vaish)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

19/05/2025

Signature of the Supervisor of M.Tech. Thesis with Date

(Prof. Santosh Kumar Vishvakarma)

Shivam Vaish has successfully given her M.Tech. Oral Examination held on **5th May 2025**

Signature of Supervisor of M.Tech. Thesis

Date: 19/05/2025

Saptarshi Ghosh

Signature of Convener, DPGC

Date: 19-05-2025

ACKNOWLEDGEMENTS

I am immensely grateful to my M.Tech thesis supervisor and mentor, Prof. Santosh Kumar Vishvakarma, for consistently encouraging and supporting me in both my research and personal growth. His unwavering belief in my abilities and his invaluable guidance have served as constant motivation, pushing me to exceed my own limits. I owe him a debt of gratitude for granting me the freedom to explore my research interests and allowing my novel ideas to flourish.

I would also like to extend my sincere appreciation to all of my thesis evaluation committee. Their impartial evaluations and thought-provoking questions have contributed significantly to expanding my research perspective.

My family has played a major role in supporting my research work throughout the course of my master's. They have always boosted my confidence and always motivated me to push my limits. I will always be grateful to them for all their guidance, love and sacrifices. Their faith in me has brought me this far, and it will drive me further, as well, to achieve greater things. I deeply appreciate the Nanoscale Devices, VLSI Circuit System Design Lab (NSDCS) research group, especially Mrs. Neha Maheshwari for their continuous support and guidance. I am also grateful to my friends and labmates Mr. Vikash Vishwakarma, Mr. Sonu Kumar, Mr. Shashank Singh Rawat, Mr. Mukul Lokhande, Mr. Ankit Tenwar, Mr. Akash Pandey, Mr. Akash Sankhe, whose camaraderie and encouragement made my time at the institute truly memorable.

Shivam Vaish

This Thesis is Dedicated

to

My Parents, My Sister, My Grandparents
and the Almighty God

ABSTRACT

Security measures that are portable, energy-efficient, and attack-resistant are desperately needed given the explosive growth of Edge Artificial Intelligence (Edge AI) systems. Conventional encryption techniques are not appropriate for edge devices with limited resources since they frequently need a significant computational cost. Using the inherent characteristics of 8T Static Random Access Memory (SRAM) cells and the bitline discharge technique, this thesis explores a secure and reliable In-memory computing architecture to implement Physical Unclonable Functions (PUF) and True Random Number Generators (TRNG) directly within memory arrays.

The major goal of this work is to create and evaluate a comprehensive 8T SRAM-based framework that can handle secure key generation based on PUF with little overhead. A single 8T SRAM cell is designed and characterized at the start of the study, which then moves on to a 64×32 array layout. In order to ensure proper operation and performance optimization of the memory array and for PUF implementation, extensive efforts were made to create, integrate, and simulate crucial peripheral circuits such as write driver, precharge circuit, and sense amplifier.

The study included extensive pre-layout and post-layout simulations for PUF implementation as a crucial component. In order to examine the effects of parasitic elements included during layout synthesis, the PUF operation such as delay and power consumption metrics were carefully retrieved at both stages. The performance differences between schematic-level (pre-layout) and extracted-level (post-layout) findings were illustrated through comparative analysis.

Particular attention was paid to locating and reducing the sources of bit errors and instability that impair PUF performance. To improve the system's resilience to changes like supply noise, temperature swings, and process corners, methods like majority voting, stable bit filtering, and voltage tuning were investigated and simulated.

Contents

List of Figures

List of Tables

List of Abbrevations

PUF	-	Physically Unclonable Function
TRNG	-	True Random Number Generator
AI	-	Artificial Intelligence
IC	-	Integrated Circuit
SRAM	-	Static Random Access Memory
CRPs	-	Challenge-Response Pairs
BER	-	Bit Error Rate
PVT	-	Process Voltage Temperature
WWL	-	Write Word Line
RWL	-	Read Word Line
RBL	-	Read Bit Line
TDC	-	Time To Digital Concerter
SoC	-	System-on-Chip
ASIC	-	Application Specific Integrated Circuit
IC	-	Integrated Circuits

Chapter 1

Introduction

1.1 Need for secure and reliable random number generation

With the growth of the internet of things and embedded devices, integrated circuit security has emerged as a crucial issue in contemporary computer systems. Secure key generation and authentication are made possible by hardware primitives known as Physically Unclonable Functions (PUFs) and True Random Number Generators (TRNGs) [3]. SRAM PUFs are one of the most studied hardware-based security primitives because of their straightforward design and simplicity of integration with current systems. Systems that are both intelligent and safe are required due to the growing integration of AI at the network edge, where data is created and consumed [4,5]. Conventional computing architectures result in inefficiencies in terms of speed and energy since data is constantly moved between memory and processors. By allowing data processing inside memory arrays, in-memory computing overcomes this bottleneck and significantly lowers latency and power usage.

But security is still a major issue, particularly in dispersed AI and IoT contexts. Because hardware-based security primitives like PUFs and TRNGs rely on noise and process fluctuations, which are hard to predict or duplicate, they have inherent advantages. The widespread use of SRAM-based systems and their compatibility

with current CMOS processes make them especially appealing.

Process variances happen when ICs are being manufactured. Each transistor in an integrated circuit (IC) has unique physical characteristics as a result of these variances, which further alters the transistors behavior [7]. In particular, this leads to minor variations in electrical characteristics. Variations in the manufacturing process of an IC cannot be fully controlled. Therefore, it is impossible to replicate or clone the qualities that result from process changes [8]. A huge number of SRAM unit cells make up an SRAM memory. Two coupled inverters, each composed of PMOS and NMOS transistors, make up each SRAM cell. The relationship between the threshold voltages of the PMOS transistors found in the associated inverters determines the state of the SRAM cell during power-up. A condition of high or low is determined by the transistor that conducts first. Figure 1.1 shows the true random number are crucial for security.

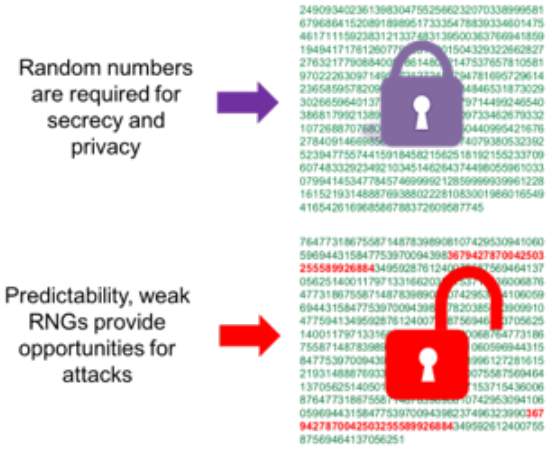

Figure 1.1: Need for secure random number [1]

Each SRAM cell has two stable states, one and zero. However, a certain cell stable state during powerup is unpredictable. The preference to return up as a zero or one is actually caused by random sub-microscopic discrepancies among the transistors that make up the SRAM cells. When the SRAM is powered on, this collection of cells creates a random pattern. It can be used as a silicon fingerprint because it is specific to each IC. A hardware-based cryptographic key can be created using this approach. The primary cause of each SRAM unit cells unique power-on state is the

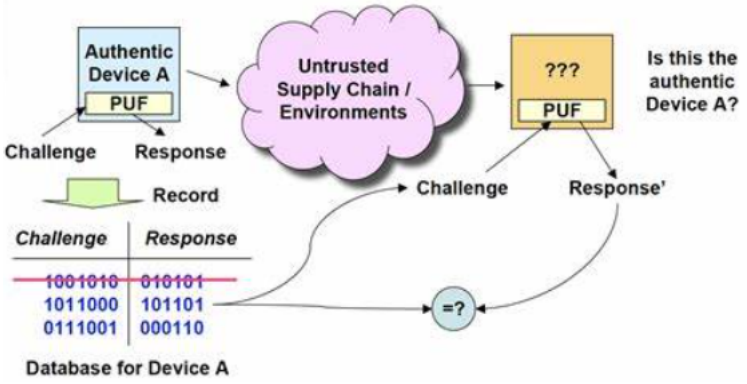

Figure 1.2: Physical unclonable functions [19]

fluctuation in the threshold voltage of the transistors that make up SRAM. This preference is independent of the position of the cell within the chip or the power of the nearby cell. Thus, a new and unknown sequence of zeros and ones is produced by SRAM. This pattern might be referred to as an SRAM fingerprint because it is specific to each SRAM. Thus, it is frequently used as a PUF and function of PUF is shown in figure 1.2 .

1.2 Physical unclonable functions and the role of SRAM in secure hardware design

Hardware security primitives known as physical unclonable functions take advantage of the uncontrollable and intrinsic process variations that are produced during the fabrication of semiconductors. These minute distinctions, including variances in transistor mobility, channel length, or threshold voltage, are specific to each integrated circuit (IC). Utilizing these physical differences, PUF produce Challenge-Response Pairs (CRPs), which are the basis for safe cryptographic key creation, device authentication, and anti-counterfeiting mechanisms, in response to inputs. PUF offer a portable and impenetrable solution for hardware-level security because their physical attributes are very hard to duplicate or anticipate, even by the original manufacturer [9, 10].

Due to their ease of use, low overhead, and the fact that SRAM blocks are present

in the majority of digital systems, SRAM-based PUF have drawn the most interest among the several kinds of PUF. An SRAM cell is made up of two cross-coupled inverters, especially in its typical 6T design. Each SRAM cell settles into a state of either 0 or 1 upon power-up because of intrinsic transistor mismatches brought on by manufacture variances [11]. The SRAM PUF is based on this power-up state, which is generally stable and reproducible under typical operating conditions. Without the need for extra random number generators or dedicated cryptographic memory, the accumulation of these initial values throughout a range of SRAM cells creates a distinct and device-specific fingerprint. SRAM was chosen for PUF implementation due to a number of significant benefits. First, existing SRAM cells can be used directly, saving space and power. No extra circuitry is needed. Second, SRAM cells power-up behavior offers high entropy because of physical randomness, which makes it ideal for cryptographic applications. Thirdly, SRAM PUFs exhibit non-volatile behavior as, under normal circumstances, the same power-up pattern can be consistently replicated over several resets. There are still certain issues, though, such as the power-up values vulnerability to external variations like temperature and voltage, which could cause instability or bit flipping. In order to improve the robustness of the PUF, methods including error correction codes (ECC), auxiliary data algorithms, and modified SRAM layouts such as 8T cells are used. The 8T SRAM cell increases read stability and power-up value dependability by adding two more transistors for read buffer isolation. These improvements are essential when employing SRAM PUFs in settings with fluctuating operating circumstances or where high dependability is needed, like in cryptographic key storage or true random number generators as shown in figure 1.3. All things considered, the combination of high entropy, cheap cost, and simplicity of integration inside current VLSI systems makes the use of SRAM in PUF architecture a realistic and effective solution to hardware security.

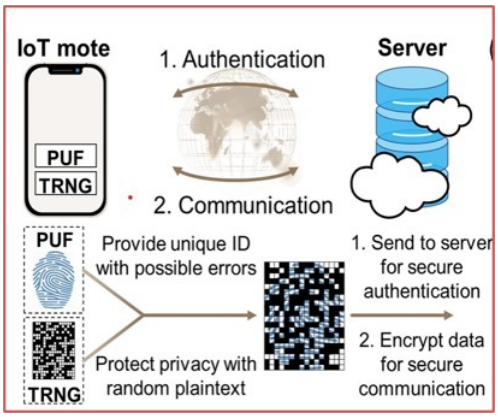

Figure 1.3: Introduction of PUF and TRNG and their applications in e IoT security [4]

1.3 Bitline discharge method for PUF and TRNG

The power up values of memory cells are the main source of randomness in traditional SRAM-based PUF designs. However, environmental changes like temperature, supply voltage fluctuations, and process variances can frequently cause instability in this technology [12]. This thesis proposes and implements a unique method that uses the bitline discharge behavior to address these issues.

The bitline discharge technique measures the variation in bitline discharge rates during a read operation. Depending on the internal state of the memory cell, the precharged bitlines in a typical 6T or 8T SRAM cell start to discharge when a read access is started. The pull-down transistors and access transistors intrinsic manufacturing inconsistencies have an impact on this discharge. A distinct and reproducible signature can be extracted from every SRAM cell by closely observing and comparing the bitlines discharge rates [13].

Compared to traditional power-up-based methods, this approach has the following benefits:

- **Increased stability:** The measurement is less susceptible to noise from the environment because it is based on relative discharge behavior rather than absolute logic values.
- **Improved uniqueness and entropy:** For TRNG and PUF applications,

the analog nature of discharge rates allows for more precise cell distinction, which improves entropy.

- **Decreased Bit Error Rate (BER):** The dependability of the generated keys is improved by concentrating on differential characteristics rather than binary states, which greatly reduces the BER.

This work uses a 8T SRAM cell to implement the bitline discharge method. The methods reliability under process-voltage-temperature (PVT) fluctuations is validated by post-layout simulations. The technique is appropriate for secure key generation and device authentication applications since the results show a potential improvement in both randomness quality and repeatability.

1.4 Objectives

The main goal of this study is to use an 8T SRAM cell to investigate and develop a dual-function circuit design that combines physical unclonable function and true random number generator capabilities. The need for improved hardware-level security primitives in low-power and space-constrained contexts, such embedded systems and the internet of things, is what motivates the research. The following lists the precise objectives:

- **Design of 8T SRAM for PUF generation**- Create an 8T SRAM cell structure that is dependable and strong, with PUF behavior in mind. In order to provide device-unique responses with high entropy and stability, the cell should take use of intrinsic differences in the manufacturing process. Under various operational conditions (temperature, voltage, etc.), the design must provide a minimal bit error rate.

- **Leverage bitline discharge technique for TRNG**-In order to provide a source of genuine randomness, incorporate a new bitline discharge technique into the existing 8T SRAM system. This method should produce unbiased,

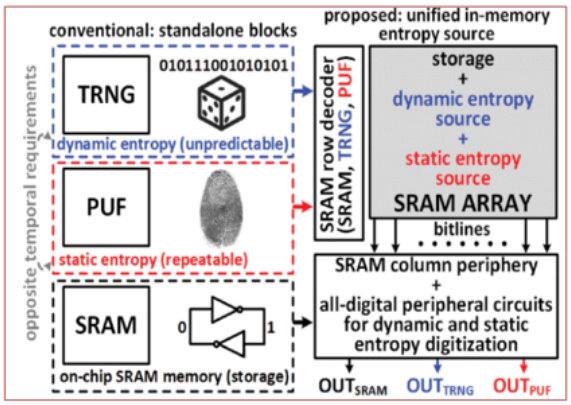

Figure 1.4: In-memory unified entropy source (SRAM with TRNG and PUF) for secure SoCs. [1]

unpredictable bits appropriate for cryptography applications by utilizing transient analog behaviors like metastability and discharge noise during read or write cycles.

- **Validation through simulation and statistical metrics**-Verify the design thorough pre and post-layout simulations at a typical CMOS technology node using UMC 40nm. The design will be examined for:

- **PUF metrics**: Bit error rate, instability.

- **Performance metrics**: Power consumption, area, and delay.

The goal of this study is to show that a single 8T SRAM-based architecture may function as a small, low-power solution for both random number and identity generation, opening the door for future systems to use more secure hardware implementations as shown in figure 1.4.

1.5 Scope and organization of thesis

Scope of the work- In this thesis, an 8T SRAM cell architecture combined with a bitline discharge technique is used to design and build a physical unclonable function and true random number generator. The main objective is to extract

high-quality random bits appropriate for security and cryptography applications by taking advantage of inherent process variability in CMOS devices.

The work scope includes:

- Creating an 8T SRAM cell that has been designed for improved bitline control and stability.
- Using industry-standard tools like cadence virtuoso to do post-layout and schematic-level simulations.
- Applying common PUF metrics to assess the generated bits instability and bit error rate.
- Making suggestions for methods to lower bit error rates and enhance PUF stability in a range of scenarios.

Thesis Organization:

Chapter 1: In the introduction, the purpose of the study is explained, the issue statement is outlined, the objectives are established, and the methodology is briefly introduced.

Chapter 2: Literature Review: Describes earlier research on PUFs and TRNGs, points out the shortcomings of current methods, and outlines the research gap this thesis attempts to fill.

Chapter 3: Gives a summary of SRAM cell architectures, explains the importance of the bitline discharge technique, and lists the main performance indicators for PUF and TRNG.

Chapter 4: Design and methodology of PUF/TRNG: Explains the bitline discharge technique operation, the suggested 8T SRAM-based PUF architecture, and the simulation setup utilized for testing and validation.

Chapter 5: Implementation demonstrates the bitline behavior and initial value extraction for entropy creation, presents the schematic and layout design, and displays simulation results. Examines the results of the simulation, assesses the PUF and TRNG properties, and offers a comparison with alternative approaches.

Chapter 6: Conclusion and future work: Provides a summary of the results, identifies significant contributions, and recommends possible directions for additional study and advancement.

Chapter 2

Literature Review

2.1 Brief on PUF types (Arbiter, Ring oscillator, SRAM)

Physical unclonable functions, are primitives in hardware security that take advantage of intrinsic process variances in integrated circuits (ICs) to produce distinct and unpredictable reactions. PUF of various kinds have been proposed, each of which makes use of unique physical properties of the silicon substrate. Arbiter PUF, ring oscillator PUF, and SRAM PUF are the PUF kinds that are most frequently researched. Figure 2.1 shows different types of PUF.

2.1.1 Arbiter PUF

One of the first and most extensively researched delay-based PUF architectures is the arbiter PUF. It is made up of two identical signal routes formed by a group of multiplexers set up in several stages. These parallel lines are used to simultaneously launch two signals. Chip to chip, the delays in the routes vary slightly due to intrinsic manufacturing variances. The signal that arrives first is determined by an arbiter at the end of the pathways, which outputs a PUF answer of either 0 or 1.

- **Challenge**: The pathways are configured via a binary input vector.

- **Response**: A single-bit output based on the difference in the signal arrival time.
- **Benefits**: Scalable to several levels and having a straightforward framework.
- **Constraints**: Needs exact synchronization, sensitive to changes in the environment, such as temperature and voltage.

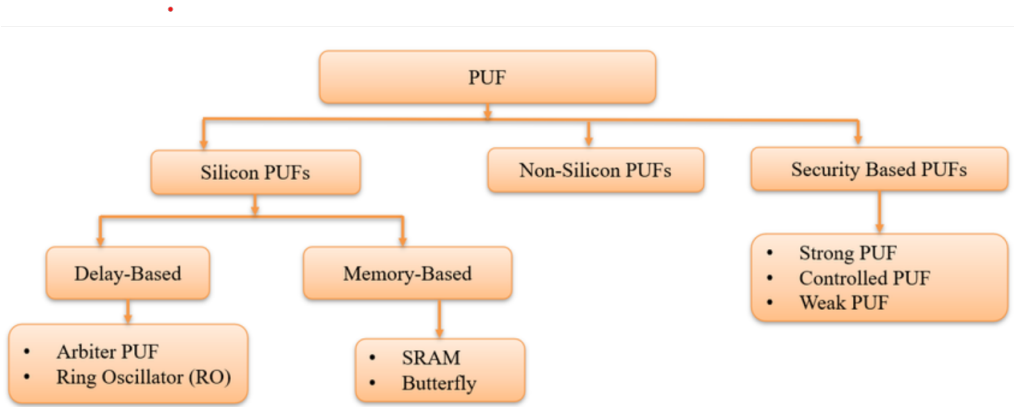

Figure 2.1: Different types of PUF [14].

2.1.2 Ring Oscillator (RO) PUF

The ring oscillator PUF takes advantage of on-chip ring oscillator frequency fluctuation. An odd number of inverters in a closed loop that oscillates at a frequency set by the inverter and interconnects delays is called a ring oscillator. The frequency of identically designed ring oscillators vary slightly due to manufacturing variances. A piece of the PUF response is formed by the comparison of several pairs of ring oscillators in a typical RO PUF architecture.

- **Challenge**: Choosing oscillator pairs for frequency comparison.
- **Response**: A bit sequence determined by the results of a frequency comparison.
- **Benefits**: Easy integration into digital design flow and high entropy.
- Requires frequency measuring equipment or on-chip counters; power consumption may be an issue.

2.1.3 SRAM PUF

The SRAM PUF takes advantage of the random power-up state of SRAM cells, which is impacted by tiny discrepancies between the pull-up and pull-down transistors. An SRAM cell settles into a stable state (0 or 1) based on these mismatches when it is first turned on (or following a complete reset) [14,15]. Under regulated environmental circumstances, this activity can be repeated for the same cell, offering a reliable and inherent source of entropy.

- **Challenge**: Not relevant (no input challenge is needed; inherent PUF).
- **Reaction**: SRAM cells power up and generate a binary sequence.
- **Benefits**: Include quick reaction, high-density entropy source, and no additional hardware.
- **Limitations**: Requires error correction and reliability enhancement approaches; susceptible to external conditions like temperature and voltage.

Regarding uniqueness, dependability, area, and power consumption, each form of PUF has pros and downsides of its own. The choice of the best PUF architecture depends on the application and is frequently impacted by the target reliability requirements and the resources that are available.

2.2 Existing SRAM PUF methods

Because of their inherent qualities, high-density entropy source, and ease of integration, SRAM based PUF have drawn a lot of interest. SRAM PUF take advantage of the unpredictable power-up behavior of conventional SRAM cells brought on by process-induced mismatches, in contrast to delay-based PUF that call for particular circuit designs. The basic function of PUF is shown in figure 2.2. The research has suggested a number of techniques and improvements to raise the bit error rate, uniqueness, stability, and reliability of SRAM PUF.

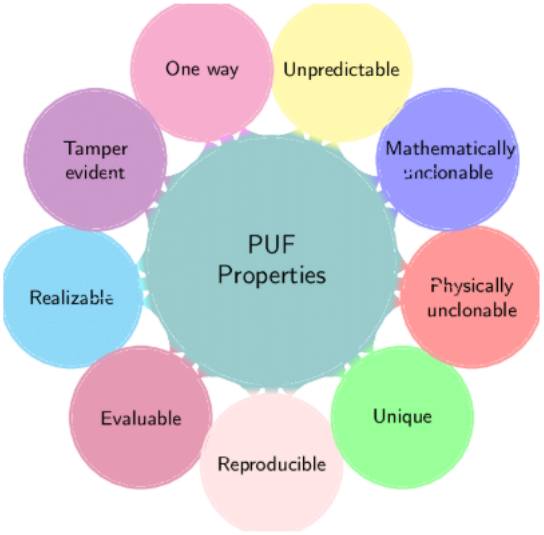

Figure 2.2: Basic function of PUF [20].

2.2.1 Conventional 6T SRAM PUF

The typical 6T SRAM cell serves as the basis for the most basic SRAM PUF. Transistor threshold voltage mismatches and process variances cause a cell to settle into either a 0 or 1 state at each power-up. Under nominal conditions, this behavior remains constant throughout several power cycles.

- **Benefits**: Makes use of current memory without the need for further equipment.

- **Drawbacks**: Bit instability and varying outputs are caused by environmental changes, such as temperature, voltage, and aging.

  In order to reduce instability, several works use methods like:

- **Majority voting**: Choosing the most frequent value after repeatedly turning on the SRAM.

- **Cell selection**: Only cells that reliably power up to the same state are chosen, discarding unstable cells.

2.2.2 8T SRAM PUF

Several researchers proposed 8T SRAM cells, which add two transistors to the conventional 6T cell, to increase robustness. In order to increase control over the stored value and noise immunity, these extra transistors create a feedback loop that can latch the cell value or help with read/write decoupling.
Benefits:

- **Enhanced dependability**: Under stressful environmental conditions, the latch aids in preserving the stored value.
- **Improved isolation**: More stable power-up states are made possible by the 8T configurations reduction of read disturb faults.
- **Lower BER**: Outputs that are more consistent and reliable over PVT changes.

2.2.3 SRAM PUF with assist techniques

Assist measures have been developed to further improve the SRAM PUF replies repeatability. In order to affect the metastability point and strengthen the desired power-up state, these strategies momentarily change the supply or access conditions during the power-up phase.
A few ways to help:

- **Negative bitline (NBL) assist**: To fix metastability, bitlines are pulled a little bit below ground during power-up.
- **Wordline overdrive**: This technique speeds up resolution by momentarily increasing wordline voltage.
- Back-biasing is the process of altering transistor body voltage to produce skew and increase the determinism of power-up values. By biasing the cell toward a more stable state, these aid techniques successfully lessen the likelihood of unstable bits.

2.2.4 Aging-Resistant SRAM PUFs

Over time, aging factors including Hot Carrier Injection (HCI) and Negative Bias Temperature Instability (NBTI) deteriorate transistor properties, causing drift in the PUF response. Several research have suggested strategies to combat instability brought on by aging:

- **Using aging-aware cell design**: SRAM cells are made with transistor sizes that reduce the influence of threshold drift.
- **Error correction over time**: ECC recalibration or periodic re-enrollment.
- **Adaptive cell selection**: Monitoring aging patterns and making necessary adjustments to the selection of dependable cells.

2.2.5 Reconfigurable and dynamic SRAM PUFs

Reconfigurable SRAM PUFs have been proposed to expand the challenge-response area or modify the PUF to suit various operational scenarios. By altering the cells activity or the measurement conditions, these architectures allow the same hardware to produce a variety of response patterns.
Among the methods are:

- Changing the timing or power-up sequence
- Changing the bitline precharge levels or supply voltage, adjusting the dynamic threshold

By facilitating dynamic reactions and lowering the possibility of modeling attacks, these techniques raise the security level of SRAM PUF.

2.2.6 Bitline discharge-based SRAM PUF

The bitline discharge behavior during read or power-up operations is the topic of a more modern class of SRAM PUF. These techniques track the transient voltage

behavior of bitlines to infer process variation fingerprints, rather than depending just on the cells ultimate state [17]. Summary of existing SRAM PUF methods is shown in table 2.1.

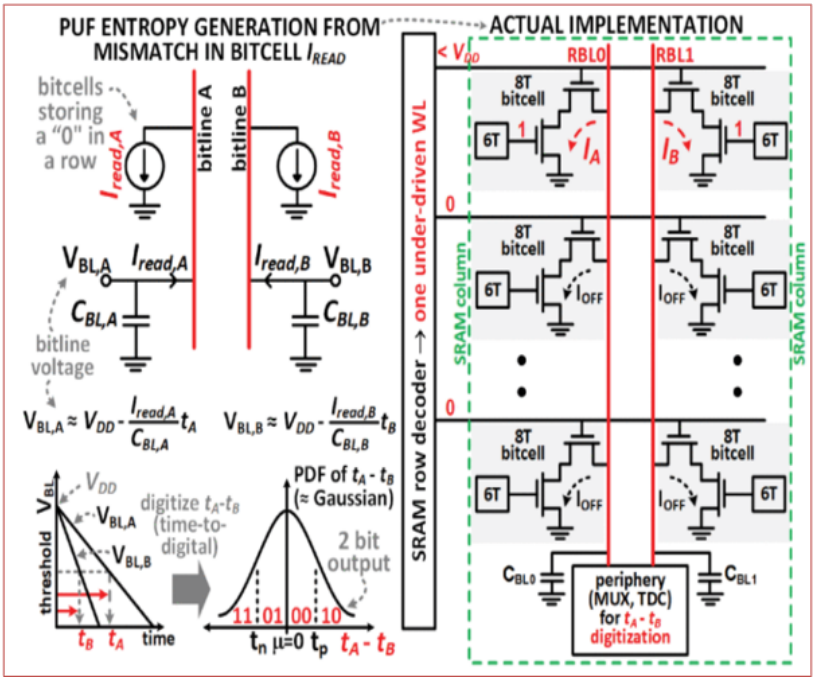

Figure 2.3: Working principle of in-memory static entropy generation (PUF). [1]

- As a PUF response, Bitline Discharge Timing (BDT) records the variation in discharge rate.
- Analog sensing is the process of employing comparators or sense amplifiers to sample intermediate voltages on bitlines.
- Higher entropy and distinct signatures can be obtained with this method, even from marginal or unstable cells that are often rejected in traditional SRAM PUFs.

2.3 TRNG Architectures

Since they produce unexpected and non-deterministic bitstreams that form the basis of safe key generation, authentication procedures and encryption techniques,

Method	Key feature	Advantage	Limitation
6T SRAM PUF	Standard cell	No hardware overhead	High BER, unstable
8T SRAM PUF	Added feedback or isolation	Improved stability, better noise immunity	Increased area
Assist techniques	External biasing	Reduced metastability	Requires control circuitry
Aging-Resistant SRAM PUF	Design or protocol-based aging countermeasures	Long-term reliability	May need re-enrollment
Reconfigurable SRAM PUF	Variable power	Dynamic responses, larger CRP space	Complexity, overhead
Bitline discharge SRAM PUF	Uses analog discharge behavior	High entropy, fine-grained signature	Requires precision sensing

Table 2.1: Summary of existing SRAM PUF methods

true random number generators are crucial hardware primitives in contemporary cryptographic systems. TRNG produce randomness from intrinsically unpredictable physical processes, including thermal noise, metastability, or jitter, as opposed to pseudo random number generators (PRNG) ultimate state, which rely on deterministic algorithms and a seed. In the literature, numerous TRNG architectures have been put out, each of which makes use of various circuit-level strategies and entropy

sources. Among the most well-known architectures are the following:

2.3.1 Ring oscillator-based TRNG

The jitter in the phase of free-running inverters coupled in a loop is exploited by ring oscillator TRNGs. The timing uncertainty caused by jitter results in a random output when one oscillator is sampled with another.
Benefits:

- Simple on-chip integration
- Adaptable to the chosen entropy rate or frequency

Limitations:

- Electromagnetic interference is a risk.
- To provide adequate unpredictability, a large number of oscillators are needed.

2.3.2 Amplifier or noise-based TRNG

These TRNGs work by amplifying shot noise or thermal noise from diode junctions, transistors, or resistive parts. Comparators or ADCs are used to digitize the analog noise signal.
Benefits:

- If noise is adequately isolated, a high entropy rate
- Fit for randomness of cryptographic quality

Restrictions:

- Analog front-ends require careful architecture to minimize interference, they also increase design complexity.

2.3.3 SRAM startup-based TRNG

This method takes use of SRAM cells erratic power-up condition. Each bits initial state (0 or 1) is random due to intrinsic manufacturing mismatches, and this can be harvested as entropy. This technique is occasionally utilized as a combined PUF-TRNG solution and is closely connected to PUF.

Benefits:

- Non-invasive entropy extraction in systems that already use SRAM, no additional hardware is needed.

Restrictions:

- Restricted to cold boot situations, susceptible to age or environmental factors that gradually decrease entropy.

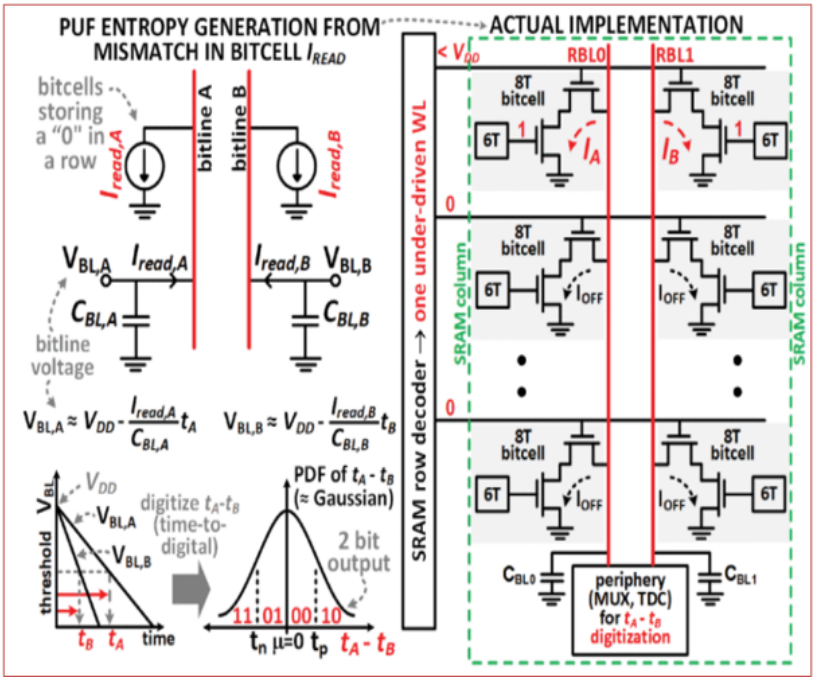

Figure 2.4: Working principle of in-memory dynamic entropy generation (TRNG) [1]

.

2.3.4 Bitline discharge-based TRNGs

This technique makes use of the intrinsic fluctuation in bitline discharge timing during read operations, which is especially pertinent to contemporary memory systems. The precise discharge trajectory may change due to device mismatch and noise in the precharge and discharge channels, which introduces entropy and working principle of TRNG is shown in figure 2.4.
Benefits:

- Ideal for incorporation into memory arrays such as 6T or 8T SRAM.
- For area efficiency, it can be integrated with the current PUF circuitry.

Restrictions:

- Randomness can be decreased via bitline capacitance and parasitic effects.
- Needs noise amplification and meticulous calibration.

2.4 Gaps identified (instability, high power, area)

Hardware-based true random number generators have been the subject of much research and development, but a number of issues still prevent their widespread and dependable use in resource-constrained and security-critical applications. These discrepancies frequently result from intrinsic trade-offs between scalability, environmental robustness, implementation overhead, and randomness quality. The following difficulties are revealed by a critical assessment of current TRNG architectures:

2.4.1 Instability and environmental sensitivity

Instability under changing environmental conditions is one of the most important design difficulties for TRNGs. Variations in temperature, variations in supply voltage, and the effects of aging can all drastically change the properties of the entropy source.

- Process, voltage, and temperature fluctuations can significantly affect the output of metastability-based and delay-based TRNG, thereby biasing them or lowering entropy.

- Why because cells may "lock-in" as a result of aging or systematic offsets, SRAM-based TRNGs experience decreased unpredictability over repeated power cycles.

- In addition to undermining security assurances, this instability calls for intricate post-processing techniques like hashing or von neumann correction, which raise design complexity and power consumption.

2.4.2 High power consumption

Some TRNG need a lot of dynamic or analog power, especially those that use ring oscillators or amplifier-based noise harvesting:

- Ring oscillator TRNG drain too much dynamic power because they need several high-frequency oscillators to operate continuously.

- TRNG based on analog noise rely on constant digitization and amplification of tiny signals, which is ineffective in low-power systems like internet of things edge devices.

For energy-constrained settings, such wearable or implanted devices, where every microwatt counts, such architectures are therefore less appealing.

2.4.3 Area overhead

Significant area overhead is frequently introduced when integrating TRNG blocks into pre-existing digital systems, particularly for designs that call for multiple entropy sources or analog front-ends.

- Avalanche noise-based and thermal noise-based analog TRNG, for example, need large parts like filters and precision amplifiers.

- Large RO arrays and inverter delay chains are examples of digital architectures that significantly increase gate count, routing complexity, and layout asymmetry.

In scaled CMOS technologies, where optimizing integration density is crucial, this becomes a limiting issue.

2.4.4 Limited integration with PUF/TRNG

Physical unclonable functions and TRNG both take advantage of circuit-level randomness, but few current solutions take advantage of a single architecture to generate PUF and TRNG at the same time.

- The majority of PUF/TRNG systems handle entropy sources independently, which results in hardware duplication and higher design overhead.
- Shared-infrastructure techniques have unrealized potential, particularly in memory-centric architectures like 8T SRAM, which can effectively serve both functions with very small design enhancements.

Chapter 3

Fundamentals and Theory of Hardware Security Primitives

3.1 SRAM cell structures: 6T vs 8T

Digital systems frequently use static random access memory cells because of their quick access times, low latency, and compatibility with common CMOS procedures. Different transistor layouts can be used to produce the memory cell, which is the basic building block of SRAM storage. The 6T and 8T SRAM cells are the most widely used of these. Although they are both volatile memory components, the structural distinctions between them have a big impact on stability, dependability, and compatibility with security primitives like PUFs and TRNGs [16].

3.1.1 6T SRAM cell

Six transistors make up the 6T SRAM cell: Two access transistors that link the storage nodes to two bitlines and two cross-coupled inverters that create a latch. Working of write and read operation of 6T SRAM :

- Write operation: This is accomplished by activating the wordline and forcing bitlines to follow the desired logic.

- Read operation: A single bitline is precharged, and the stored data uses an access transistor to conditionally discharge it.

Benefits:

- High-speed read/write operations and a compact layout with little space overhead are the limitations.
- Read disturb: Internal node voltage may change during read operation, particularly in weak noise or mismatch situations, read stability is limited because of the shared read/write path concurrent read and write isolation is not optimal and the system is susceptible to noise, variability, and aging effects, all of which are critical for PUF reliability.

3.1.2 8T SRAM cell

In order to create an independent read buffer, the 8T cell adds two more transistors to the 6T structure. This improves read stability by separating the read path from the storage node.
Benefits include:

- One read wordline (RWL) and one read bitline (RBL)
- Increased stability in reading: By separating the read channel from the storage node, read disturbance is avoided.
- Improved resistance to noise: Increased resilience to PVT shifts and process variations
- Ideal for PUF/TRNG: Repeated, non-invasive measurements are possible with isolated read access.
- Facilitates entropy extraction methods based on bitline discharge.

Limitations:

- A somewhat bigger area and more transistors.

- More intricate control signals (WL and RWL distinct).

3.1.3 Comparative summary

Feature	6T SRAM	8T SRAM
Transistor count	6	8
Read disturb	High	Very low
Area efficiency	High	Moderate
Read/Write path separation	No	Yes
Stability under PVT	Moderate	High
Suitability for TRNG/PUF	Limited (noisy and unstable)	Excellent (stable and isolated)
Bitline discharge use	Risk of state corruption	Enables safe and repeatable access

Table 3.1: Comparative summary of 6T SRAM vs 8T SRAM

Because of their small size and quick operation, 6T SRAM cells are still the best choice for high-density cache memory; nevertheless, they are less dependable for security applications that include noise exploitation or repetitive reads. On the other hand, 8T SRAM cells are more suitable for implementations of true random number generators and physical unclonable functions because to their higher stability and read integrity. The TRNG architecture suggested in this thesis is based on non-destructive entropy harvesting methods such bitline discharge analysis, which are made possible by their decoupled read path. Comparative summary of 6T SRAM vs 8T SRAM is shown in table 3.1

3.2 Role of process variation

The performance and dependability of nanoscale integrated circuits are greatly impacted by process variation, a fundamental feature in semiconductor production.

It describes how constraints in fabrication precision cause inherent oscillations in device parameters as threshold voltage, channel length, oxide thickness, doping concentration, and line edge roughness. These variances form the basis for constructing hardware security primitives such as physical unclonable functions, even though they are frequently seen as harmful to deterministic digital systems.

Process variation is used in PUF to provide distinct and surprising responses from identical circuit layouts that are manufactured on several chips. The subsequent power-up state of memory elements (such as SRAM cells) becomes intrinsically device-specific due to the little variations that each transistor and memory cell display due to random dopant fluctuations and lithographic irregularities. Because of its uniqueness, every chip can produce a different "fingerprint" that can be used for safe identification and verification.

Process variation also adds to entropy for TRNG by causing unpredictable behavior in metastable states, noise amplification circuits, or delay routes. When appropriately conditioned, these non-deterministic replies produce genuinely random bitstreams that are appropriate for use in cryptography. For example, in an 8T SRAM-based TRNG design, process variation amplifies the asymmetry in bitline discharge during the read/write operation and the mismatch in pull-up/pull-down routes, leading to spontaneous bit flips that increase unpredictability [19].

Excessive mismatches or systematic deviations may cause biased outputs or weaken the PUF/TRNG stability. Therefore, to minimize unwanted effects while still utilizing the unpredictability brought about by variation, careful design strategies are used, such as layout symmetry, transistor size, bitline equalization, and the inclusion of error correction codes.

3.3 Bitline discharge fundamentals

A key idea in SRAM operation, especially during read access, is bitline discharge. Data is transferred between memory cells and peripheral circuitry via bitlines in conventional SRAM layouts, like 6T or 8T cells. It is crucial to comprehend bitline

discharge dynamics in order to analyze memory performance and to take use of it in applications such as physical unclonable functions.

The bitlines are precharged to a preset voltage level VDD prior to access in a typical read operation. The internal node of the chosen SRAM cell links to the bitline via a pass transistor when the wordline is asserted. The voltage differential and current flowing through the access transistor cause one of the bitlines to discharge, depending on whether the value is stored as 0 or 1. After detecting the tiny differential voltage, the sense amplifier amplifies it to full logic levels.
A number of factors affect the discharge behavior:

- Process variation and sizing have an impact on access transistor strength.
- Bitline capacitance: A higher capacitance reduces discharge speed.
- Discharge route resistance varies among cells as a result of diversity brought about by manufacture.

These modest discharge properties can be used to introduce entropy in the context of TRNG and PUF. Even with the same control signals and beginning circumstances, various memory cells will show somewhat varied discharge rates due to process differences. Unpredictable or distinctive behavior can be retrieved by measuring the voltage level or the discharge time at a certain moment. For example, variations caused only by inherent mismatches in the physical characteristics of the transistors can be found by comparing the bitline voltages following a partial read time.

Bitline discharge is a more dependable entropy extraction process in 8T SRAM-based architectures because the read and write channels are separated, further isolating the read operation from disturbances. Better sensitivity to randomness brought on by process fluctuations is possible by optimizing the read buffer or sensing node to amplify little differences.

As a result, the bitline discharge technique plays a crucial role in improving both uniqueness (for PUF) and randomness (for TRNG). It offers a high-speed, non-invasive method of taking advantage of the inherent diversity of devices without the need for intricate analog circuitry.

3.4 PUF metrics (Bit error rate, instability)

A set of precise statistical measures that measure a physical unclonable functions uniqueness, dependability, randomness, and resistance to temporal and environmental fluctuations are commonly used to assess the functions performance. Among them, bit error rate and instability are two crucial metrics that accurately represent a PUF resilience and usefulness in real-world situations, particularly for cryptographic applications like device authentication and key generation.

3.4.1 Bit error rate

One indicator of a PUF dependability is its bit error rate. It measures the degree to which a PUF reliably reproduces the same response under various operating conditions, such as variations in temperature, supply voltage, or time.
The Bit error rate is defined as:

$$\text{BER} = \frac{1}{N} \sum_{i=1}^{N} \left(R_i^{\text{ref}} \oplus R_i^{\text{test}} \right) \tag{3.1}$$

Where:

- N is the total number of bits in the PUF response.
- R_i^{ref} is the reference response bit obtained under nominal conditions.
- R_i^{test} is the test response bit under varying conditions.
- $\oplus$ denotes the bitwise XOR operation.

To guarantee that the answer can be reliably duplicated, a low BER (usually less than 5%) is preferred. However, some bits may flip over time as a result of power supply fluctuations, manufacturing variations, and thermal noise. In order to stabilize the output, BER analysis directs the development of error correction systems and aids in the identification of such unstable bits.

3.4.2 Instability

The intra-device variation in the PUF responsiveness over time or under stress from the environment is measured by instability. It shows the degree of variation in a single devices reaction to the same task across several measurements. High instability erodes the security guarantees provided by the PUF and is indicative of poor reproducibility. A percentage is a common way to express instability:

$$\text{Instability} = \frac{1}{K} \sum_{j=1}^{K} \left(\frac{H(R_j, R_{\text{ref}})}{N} \right) \times 100 \tag{3.2}$$

Where:

- K is the number of measurements.
- $H(R_j, R_{\text{ref}})$ is the Hamming distance between the j^{th} response and the reference response.
- N is the response length (i.e., number of bits).

Although a PUF of 0% would be ideal, in practice, values between 5 and 10% are acceptable and can be fixed with fuzzy extractors.

Chapter 4

Design and Methodology of PUF/TRNG

4.1 8T SRAM cell schematic and working

4.1.1 Introduction

Because of its small size and low power consumption, the traditional 6T SRAM cell is frequently utilized in memory architectures. It is less appropriate for security-related applications such as physical unclonable functions and true random number generators, where read-disturbance and resilience are crucial considerations, due to its limits in terms of read stability and noise margin. An improved 8T SRAM cell is used to get around these restrictions. By separating the read and write processes, the 8T design enhances stability.

4.1.2 Schematic description

By adding two more transistors for a distinct read path, the 8T SRAM cell is an expansion of the traditional 6T cell. The following elements make up the schematic:

- One bit of data is stored by the bistable latch made up of cross-coupled inverters (M1, M2, M5, M6). NMOS transistors are M1 and M2, while PMOS transistors

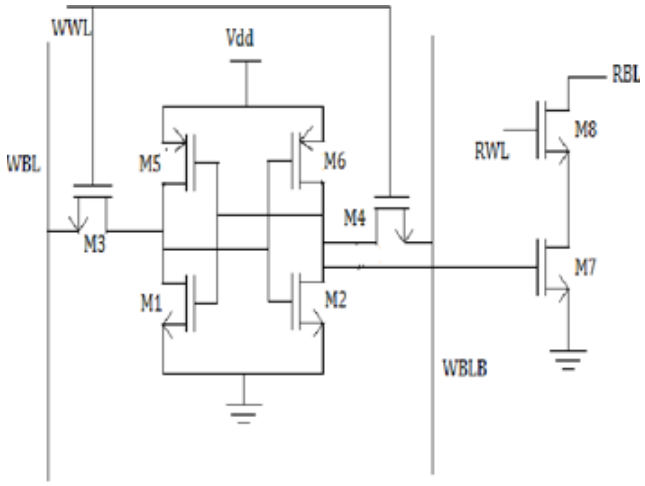

Figure 4.1: Schematic of standard 8T-SRAM [1]

are M5 and M6.

- Write access transistors (M3, M4): These are NMOS pass-gate transistors that are managed by the write word line and connected to the internal storage nodes and the bitlines (BL and BLB).

- The M7 and M8 read buffer transistors create an isolated read path. An NMOS transistor, M7, is linked between the storage node and the NMOS gate of M8. The read word line controls M8, which connects the read bitline to ground. Schematic of 8T-SRAM is shown in figure 4.1.

4.1.3 Working principle

4.1.3.1 Write operation

- The write word line is asserted high during the write phase.

- The bitlines BL and BLB (complementary values) receive the data.

- The cross-coupled inverters stored value can be overwritten by the fresh data thanks to the conductivity of the access transistors M3 and M4.

- WWL is deasserted to isolate the cell after the data has been latching.

- Figure 4.2 shows transient waveform of 8T SRAM cell during write operation.

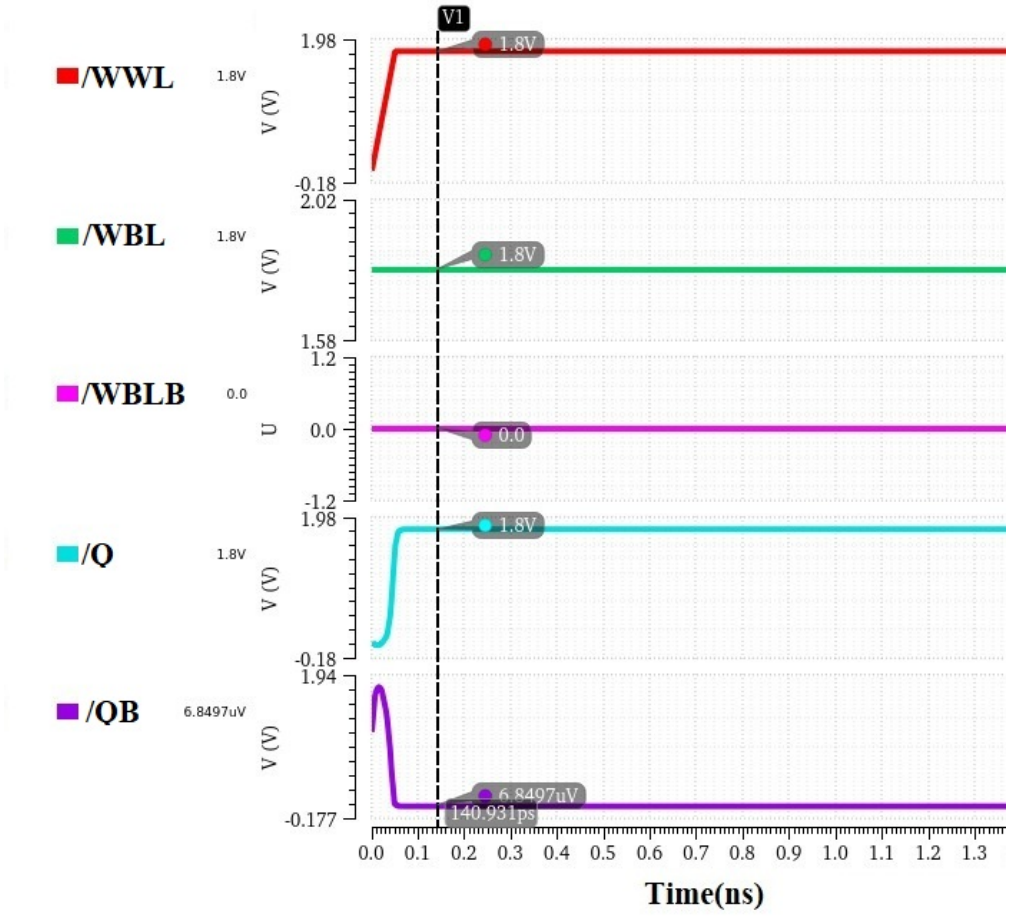

Figure 4.2: Transient waveform of 8T SRAM cell during write operation

4.1.3.2 Read operation

The read mechanism of the 8T SRAM cell is the main improvement:

- M8 is turned on when the read word line is triggered.
- The read bitline is precharged to VDD before to the read.
- M7 stays off and M8 doesn't conduct if storage node Q is low (logic 0), maintaining a high RBL.

- M7 activates when Q is high (logic 1), pulling M8 gate low to permit M8 to conduct and discharge the RBL to ground. Because internal nodes are not disrupted, this differential behavior permits non-destructive reading.

- Figure 4.4 shows transient waveform of 8T SRAM cell during read operation.

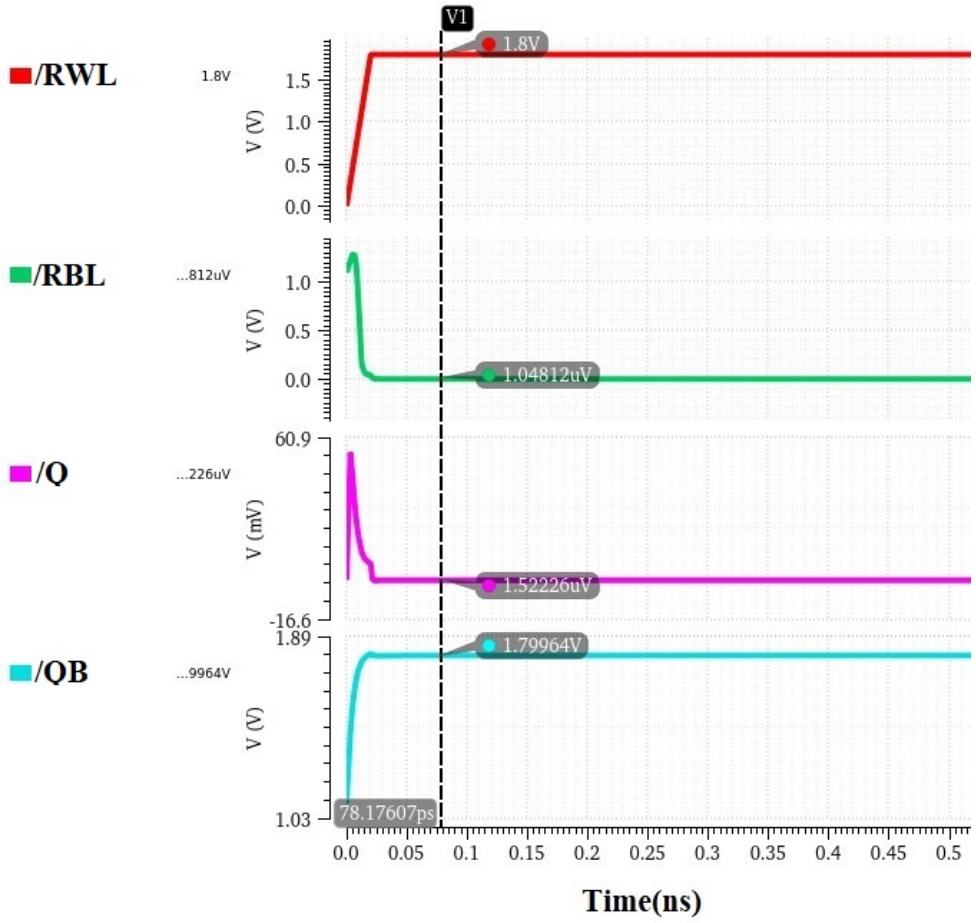

Figure 4.3: Transient waveform of 8T SRAM cell during read operation

4.1.4 Advantages of 8T cell in PUF/TRNG applications

- **Read stability**: Bit flipping during readout is reduced by isolating the read channel, which preserves the internal node voltages. Butterfly curve of 8T SRAM cell during read operation is shown in figure 4.4.

- **Greater entropy source**: The 8T cell has a somewhat higher sensitivity to process changes, which helps it produce distinct reactions in PUFs.

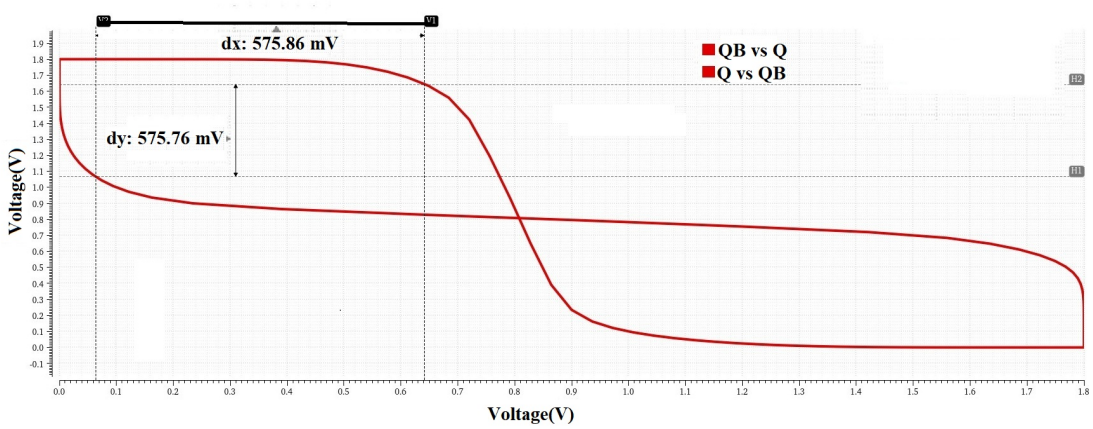

Figure 4.4: Butterfly curve of 8T SRAM cell during read operation

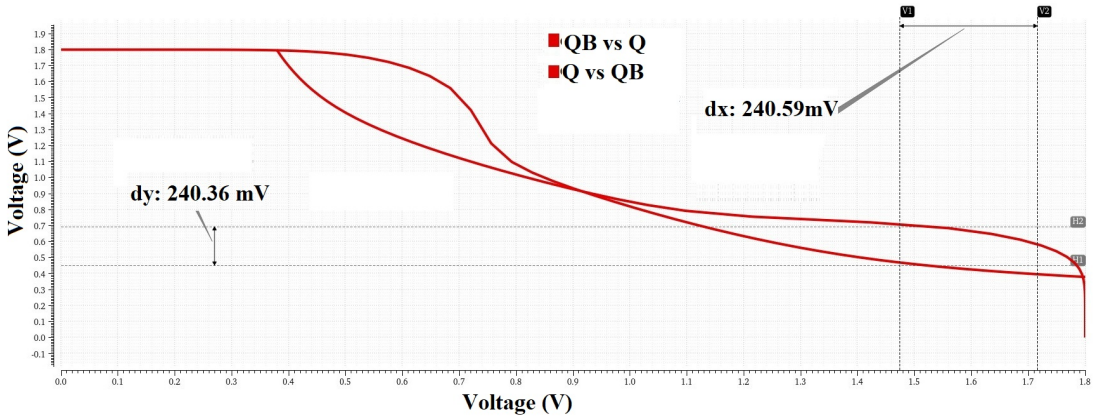

Figure 4.5: Butterfly curve of 8T SRAM cell during write operation

- **Lower bit error rate**: The bit error rate in repeated readouts is substantially lower than with 6T cells because of better noise margins and less read-disturbance.

- **Robustness to environmental fluctuations**: In cryptographic and security applications, the decoupled read path increases tolerance to changes in supply voltage and temperature.

- Figure 4.6 shows butterfly curve of 8T SRAM cell during write operation.

4.2 Explanation of bitline discharge method

4.2.1 Introduction

A reliable solution for memory-based security primitives such as true random number generators and physical unclonable functions is the bitline discharge method. The bitline discharge method makes use of analog properties, namely the discharge behavior of bitlines during a controlled read operation, in contrast to traditional SRAM based techniques that only use the power-up status of memory cells. This technique is ideal for entropy extraction and device fingerprinting since it captures the effects of noise, process fluctuations, and transistor mismatches.

4.2.2 Principle of operation

Precharge and controlled discharge are the two stages of the bitline discharge method. The method mainly tracks the voltage drop on a precharged bitline when an SRAM cell read path is activated.

4.2.2.1 Precharge phase

- The bitline, or more precisely the read bit line, or RBL, in an 8T SRAM, is precharged to the supply voltage VDD prior to the start of the read operation.
- This ensures constant circumstances for comparison by producing a uniform initial voltage level across all cells.

4.2.2.2 Discharge phase

- The read path is enabled when the read word line is turned on.
- Whether the read transistor stack (such as M7 and M8 in 8T SRAM) conducts depends on the data stored in the SRAM cell.

- The precharged RBL begins to discharge through the transistor stack in the direction of ground if the path is conductive.
- The strength (mobility, threshold voltage, and channel length variation) of the NMOS transistors determines the discharge rate.
- The cells stored state ($Q = 1$ or 0).
- Noise and random changes in the process.

4.2.3 Sensing and output generation

After a predetermined amount of time, the bitline voltage is sampled. Usually a voltage comparator or a sense amplifier are used for this:

- The output is regarded as a logical 1 if the bitline voltage VBL has fallen below a predetermined threshold V_{th}.
- VBL is regarded as a logical 0 if it stays above V_{th}.

The random bit (for TRNG) or the unique bit (for PUF) is formed by this binary output, which is based on the discharge dynamics of each SRAM cell.

4.2.4 Application in TRNG

Because of the effects of thermal noise, power supply variations, and transient device instability, the discharge behavior of TRNGs is intrinsically random:

- Different evaluations can yield different findings, even when the circuit topology is the same.
- It is possible to create a high-quality random bitstream by sampling this behavior across time and across several cells.

The sensing circuit may incorporate oscillator-based sampling techniques or metastability components to increase unpredictability.

4.2.5 Application in PUF

The bitline discharge approach for PUFs uses process-induced changes and device mismatch as a source of uniqueness:

- When the discharge is measured under controlled, repeatable conditions, the output stays stable across multiple evaluations enabling dependable key generation or device identification.
- Static physical variations, such as channel doping, oxide thickness, and threshold voltage differences, determine the discharge rate of each SRAM cell.

To lessen any instability, error correction methods (such as fuzzy extractors) or environmental compensation may be applied.

4.2.6 Advantages

Beyond digital beginning states, fine-grained analog fluctuations are captured by higher entropy extraction.

- **Increased robustness**: Read-disturbance is decreased by separating sensing from core storage.
- **Scalable technology**: Entropy is improved by greater variances at smaller nodes.
- **Adaptable configuration**: PUF or TRNG modes can be tuned via movable thresholds and sensing times.

4.2.7 Challenges

- **Circuit complexity**: Needs extra analog sensing and timing circuitry.
- **Environmental sensitivity**: Temperature and voltage have an impact on discharge rate, this needs to be adjusted or compensated.

- **Power consumption**: If precharging and sensing logic are not optimized, they may use more power.

4.2.8 Summary

An approach that shows promise for improving the entropy and dependability of SRAM-based PUF and TRNG architectures is the bitline discharge method. It successfully captures physical randomness and chip-specific variation by looking at the analog discharge behavior of bitlines, which makes it extremely pertinent for secure hardware applications. Table 4.1 shows the performance metrics of the designed 8T SRAM cell.

Table 4.1: Performance metrics of the designed 8T SRAM cell

Metric	**Value**	**Description**
Read delay	0.5 ps	Time taken to read a bit
Write delay	17.39 ps	Time taken to write a bit
Hold power	831 pW	Power consumed during hold operation
Hold static power	828 pW	Leakage during hold state
Read power	238.45 W	Power during read operation
Read static power	185 nW	Leakage during read
Write power	3.01 mW	Power during write operation
Write static power	163.72 pW	Leakage during write
Hold SNM	428.75 mV	Noise margin while holding data
Read SNM	422.18 mV	Noise margin during read
Write SNM	240.59 mV	Noise margin during write

4.3 Peripherals of memory array

Peripheral circuits are essential for enabling dependable read and write operations in memory systems, particularly SRAM-based arrays used for PUF applications. The write driver, precharge circuit, and sense amplifier are examples of these peripherals. During memory access activities, these peripheral circuits guarantee proper functionality, speed, and power efficiency, while the core memory cell (such as 6T or 8T SRAM) is in charge of storing the data. Furthermore, the precision and timing of these peripherals have a direct impact on the dependability and quality of output bits in entropy-based applications such as PUF.

4.3.1 Write driver circuit

Function: During a write operation, the write driver regulates the voltages on the bitlines (BL and BLB) in order to write data into the SRAM cell.

- The write word line, which uses access transistors to link the internal nodes of the SRAM cell to the bitlines, is activated when a write operation is started.
- In accordance with the input data, the write driver applies complimentary logic levels to BL and BLB.
- For instance, the write driver sets BL = VDD and BLB = 0V to write a 1 (and vice versa for a 0).
- If necessary, the driver must be powerful enough to reliably flip the state and overcome the cell inherent latch.

Design considerations:

- Power consumption and driving strength must be balanced.
- It should be made to prevent overdrive, which shortens the cells lifespan.
- Precise control of bitline voltages can also be used to reduce or produce variability in PUF applications.

4.3.2 Precharge circuit

Circuit structure: Usually made up of two PMOS transistors that are controlled by a precharge enable signal and coupled to the bitlines. Equalization transistors are frequently added to the circuit to guarantee that both bitlines begin at the same voltage.

- By setting the bitlines to a predetermined voltage, usually VDD or a common-mode voltage, before access, the precharge circuit gets them ready for the subsequent read or write cycle. Figure 4.6 shows schematic of precharge circuit, layout of precharge circuit and waveform of precharge circuit.
- Both BL and BLB (or RBL in 8T cells) are precharged to a predetermined voltage level at the start of each memory cycle, guaranteeing that following read operations begin from a constant state.
- This aids in lowering read time and power in conventional SRAM. It establishes the initial condition for controlled entropy extraction in TRNG designs that employ the bitline discharge approach.

Important aspects of PUF include:

- Precharging establishes the starting voltage level at which differential sensing for PUF starts.
- Bias or decreased entropy might result from any mismatch or timing fault in precharge.

4.3.3 Sense amplifier

- During a read operation, the sense amplifier is in charge of picking up on minute voltage variations between the bitlines and boosting them into complete logic levels, or digital 1 or 0. Figure 4.7 shows schematic of precharge circuit, layout of precharge circuit and figure 4.8 shows waveform of precharge circuit.

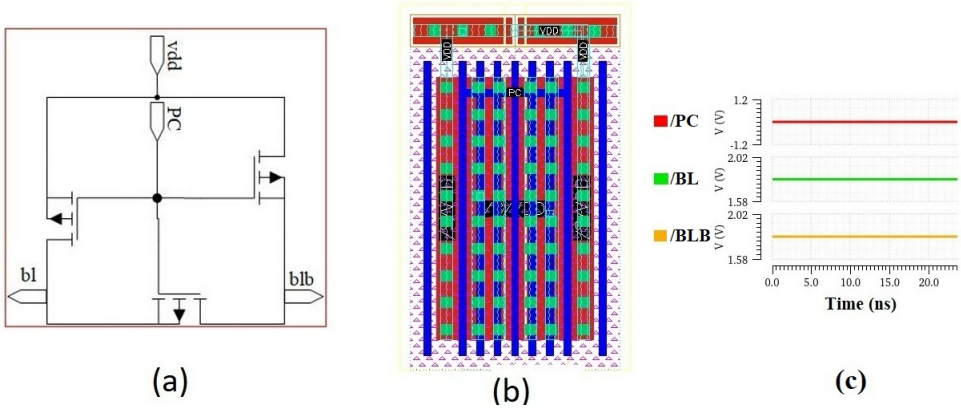

Figure 4.6: (a) Schematic of precharge circuit (b) Layout of precharge circuit (c) Transient response of precharge circuit

- One of the bitlines (BL or BLB) begins to discharge dependent on the stored data once the word line is enabled for a read operation.

- The sensing amplifier swiftly resolves this little difference to a legitimate output after detecting it, which is usually in the range of 100 mV or less.

- Both BL and BLB are inputs for differential sensing; one side is compared to a reference for single-ended sensing (such as RBL in an 8T cell).

- Among the types are latch-type sense amplifiers, which are popular in high-speed memory because they are quick and energy-efficient.

- Current-mode sense amplifiers: Used in sophisticated systems, these have improved noise immunity.

- Dynamic sensing amplifiers: For energy savings, rely on timed operation.

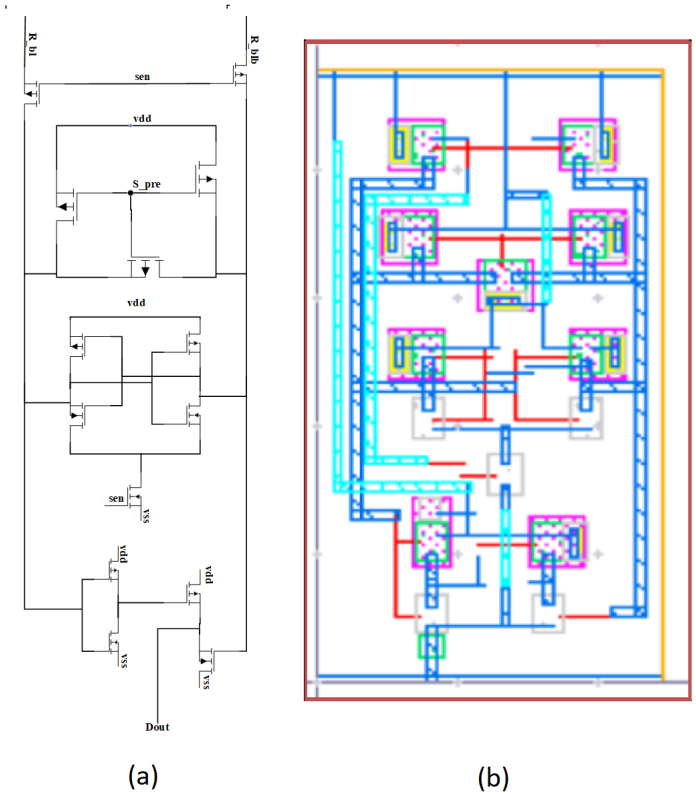

Figure 4.7: (a) Schematic of sense amplifier (b) Layout of sense amplifier

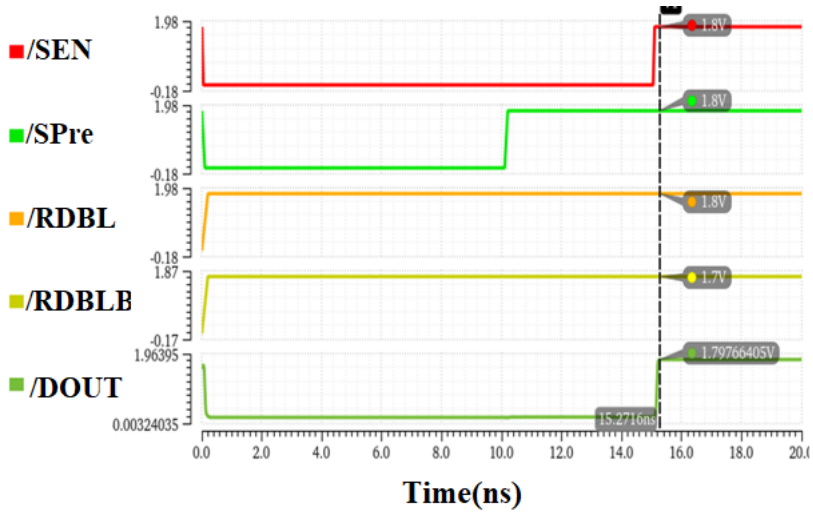

Figure 4.8: Waveform of sense amplifier

4.3.4 Summary

Table 4.2: Peripheral circuits and their role in PUF systems

Peripheral	**Role in memory**	**Key importance in PUF systems**
Write driver	Drives bitlines during write operations	Must ensure reliable state flipping for entropy storage
Precharge circuit	Prepares bitlines for next operation	Provides consistent starting conditions for discharge/sensing
Sense amplifier	Amplifies bitline voltage difference	Enables accurate bit resolution for randomness or identity

Table 4.2 shows that peripheral circuits have a direct influence on the entropy quality, bit stability, and dependability of security primitives like PUFs in addition to being necessary for the fundamental operation of SRAM arrays.

4.4 Tools used

The suggested SRAM-based PUF system was designed, simulated, analyzed, and post-processed utilizing a combination of software platforms and industry-standard Electronic Design Automation (EDA) tools. Every tool was essential to various stages of the project, ranging from data analysis and simulation to schematic design and layout. The following equipment was utilized:

4.4.1 Cadence virtuoso

Goals: Schematic design, circuit simulation, layout design, and post-layout simulation.

Application:

- Developed the 8T SRAM cell and its auxiliary circuits (sensing amplifier, write driver, and precharge).

- Analyzed discharge characteristics and process variation using Monte Carlo, parametric, and transient models.
- To verify the physical design, a layout and DRC/LVS checks were created.
- Spectre serves as the simulator backend for analog and mixed-signal simulations.

Chapter 5

Implementation and Results of PUF/TRNG

5.1 Introduction

This chapter describes the full implementation and outcomes of the suggested SRAM-based physical unclonable function and true random number generator using 8T SRAM cells and the bitline discharge approach. The cadence virtuoso design environment and UMC 40nm technology were used for the simulations. This chapter aims to analyse the outcomes under various voltage and layout settings and confirm the suggested architectures functionality, unpredictability, and stability.

5.2 PUF implementation

An 8T SRAM cell is used to build the PUF, which uses the controlled bitline discharge and power-up state to provide distinct responses. By creating unpredictability in bitline discharge duration as a result of process differences, the discharge approach increases randomization. Precharge, discharge control, and read logic are all included in the schematic. A two bit PUF output in analog form by using bitline discharge method is shown in Figure 5.1 as a waveform that shows random bit creation.

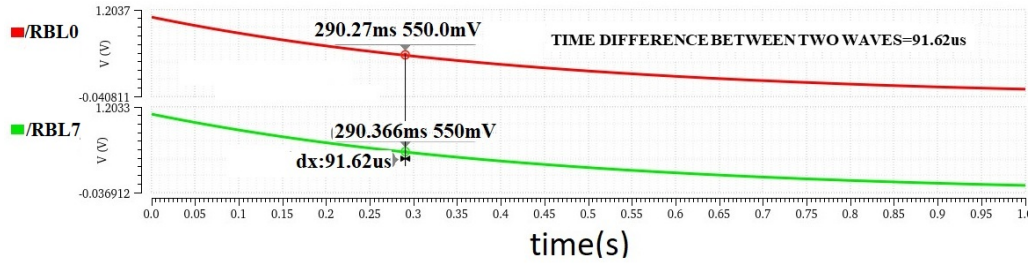

Figure 5.1: Voltage vs time for analog PUF output

5.3 TRNG implementation

The intrinsic randomness found in bitline discharge and metastability events is used in the construction of the TRNG. The delay and discharge behavior under changing process, voltage, and temperature conditions is the cause of entropy. The waveform recording the random bitstream produced by the TRNG circuit is displayed in Figure 5.2.

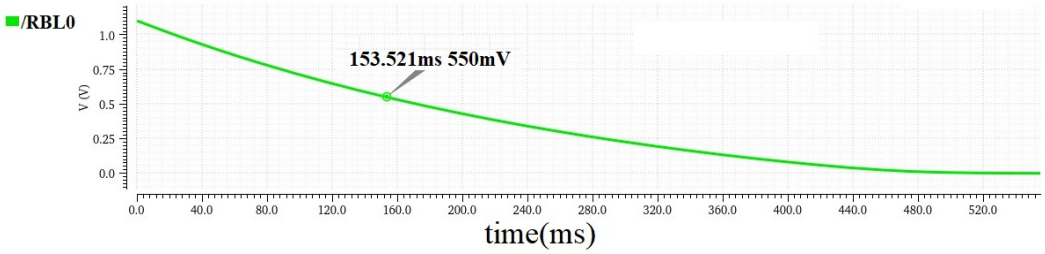

Figure 5.2: Voltage vs time for analog TRNG output

5.4 Time-to-Digital converter

The PUF circuits use a TDC to digitize the delay difference between matched pathways. It measures the time difference in digital form using a counter mechanism and a delay chain. Figure 5.3 shows the schematic of time to digital converter.

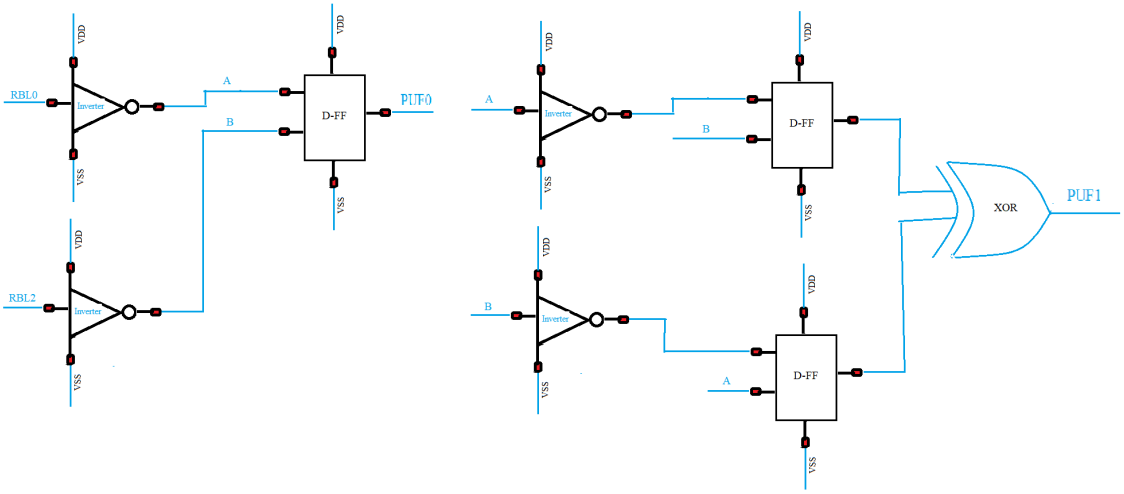

Figure 5.3: Schematic of time to digital converter

5.5 Digitization of PUF output

A time to digital converter is used to digitize the discharge-based PUF analog behavior. This guarantees a consistent 0 or 1 result for every bit. Sharp logic levels are ensured by comparing the digital output with the analog discharge voltage. The output waveform of the TDC is displayed in Figure 5.4, illustrating how sensitive it is to changes in discharge.

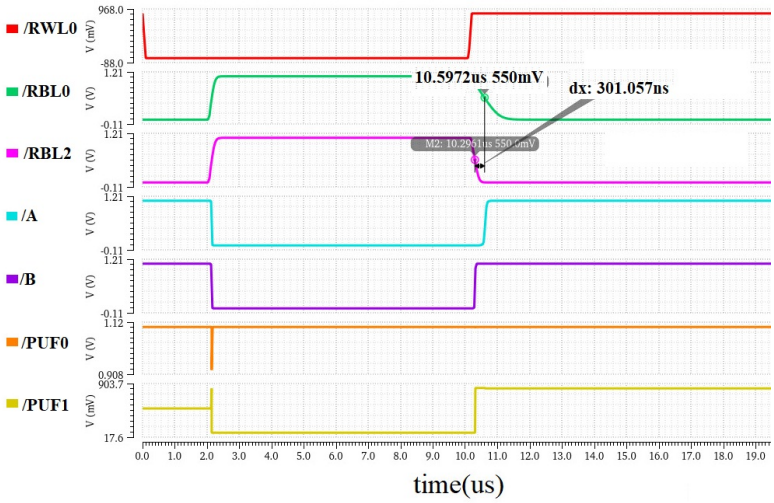

Figure 5.4: Waveform of time to digital converter

5.6 8-bit PUF output waveform

Figure 5.5 displays the collected 8-bit PUF output. Each SRAM instances uniqueness is confirmed by the waveform, which verifies steady and distinct bits produced across several runs.

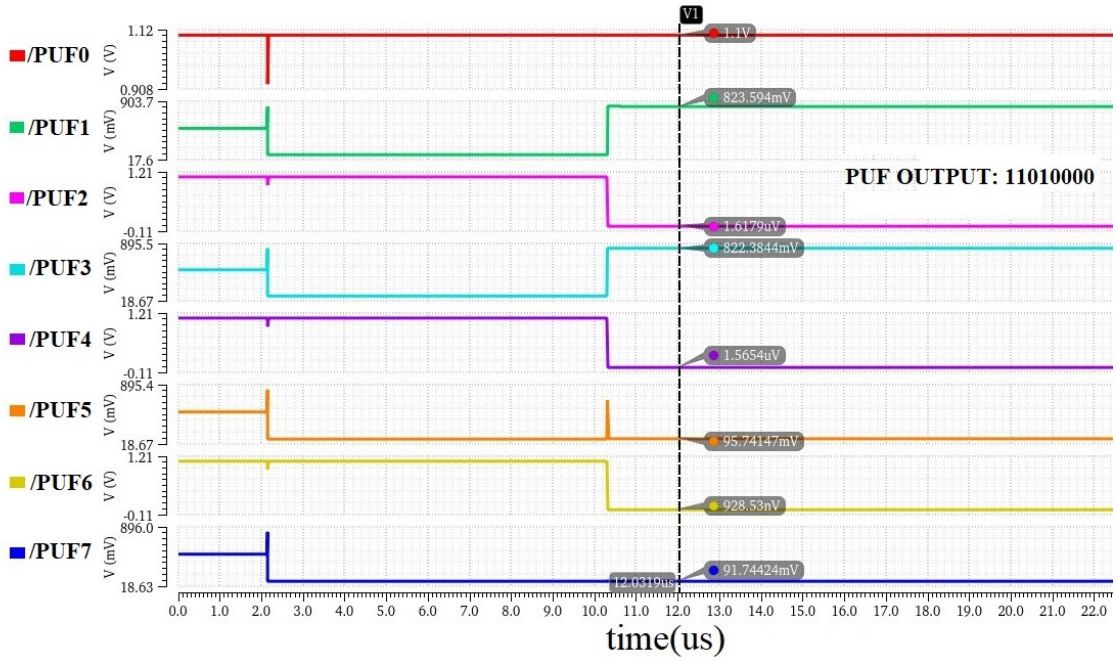

Figure 5.5: Waveform of 8-bit PUF output

5.7 Bit error rate vs supply voltage

The bit error rate is tested at several supply voltages between 0.8V and 1.3V in order to assess resilience. Figure 5.6 illustrates how the BER rises at lower voltages because of smaller noise margins while staying within reasonable bounds under normal circumstances.

5.8 Instability vs supply voltage

PUF bit instability is examined across voltage ranges. The number of unstable bits when voltage varies is shown in Figure 5.7. At nominal voltage (1.1V), the

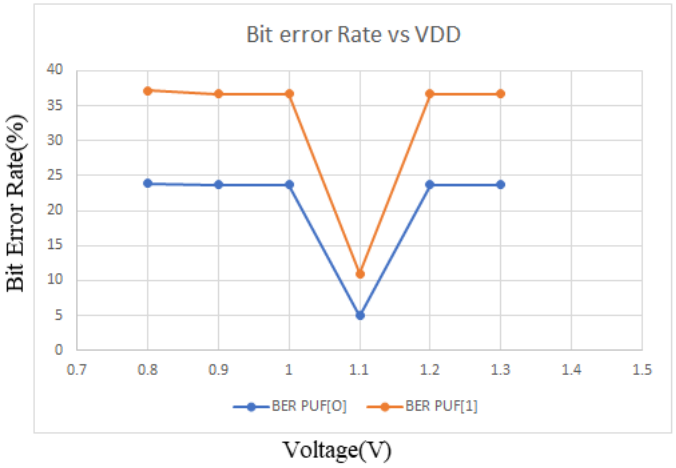

Figure 5.6: Bit error rate vs supply voltage

instability is reduced, confirming the designs dependability.

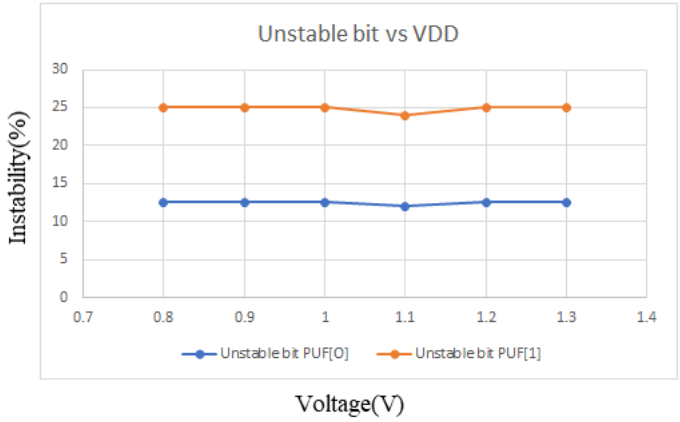

Figure 5.7: Instability vs supply voltage

5.9 Layout of 8T SRAM cell

A single 8T SRAM cell fully customized architecture is shown in Figure 5.8. Every transistor is sized appropriately to strike a balance between stability and speed. Layout accuracy was ensured by the successful completion of the DRC and LVS inspections.

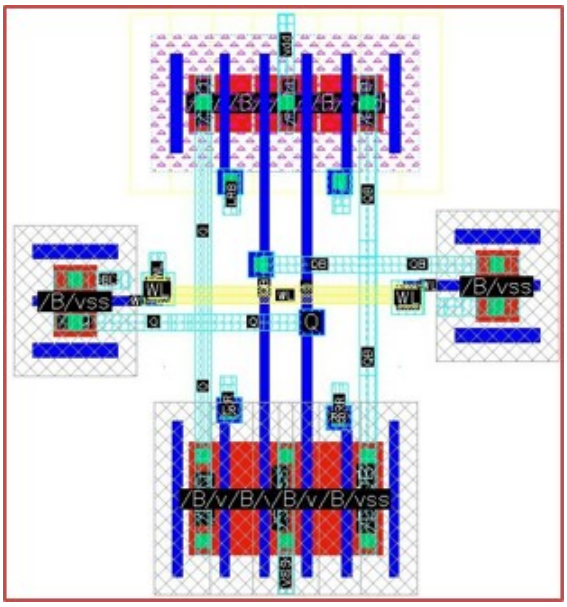

Figure 5.8: Layout of 8T SRAM

5.10 Layout of 8×4 SRAM array for PUF implementation

To ensure scalability, an array consisting of eight rows and four columns was created for PUF implementation. Area of 8×4 8T SRAM Array is 105um^2. The arrangement is depicted in Figure 5.9, emphasizing the minimal routing complexity and regular organization.

5.11 Layout of 64×32 SRAM array for PUF implementation

The implementation used a large-scale array with 64 rows and 32 columns. The hierarchical layout method, which enables effective placement and routing, is depicted in Figure 5.10. Metrics like area and anticipated power are also covered. Area of 64×32 8T SRAM Array is 1905 um^2.

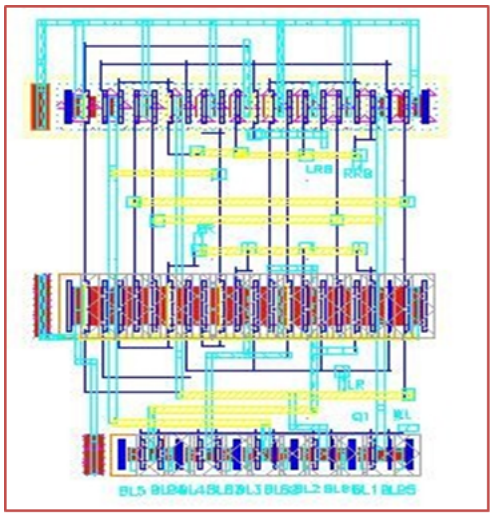

Figure 5.9: Layout design of 8×4 8T SRAM array for PUF implementation

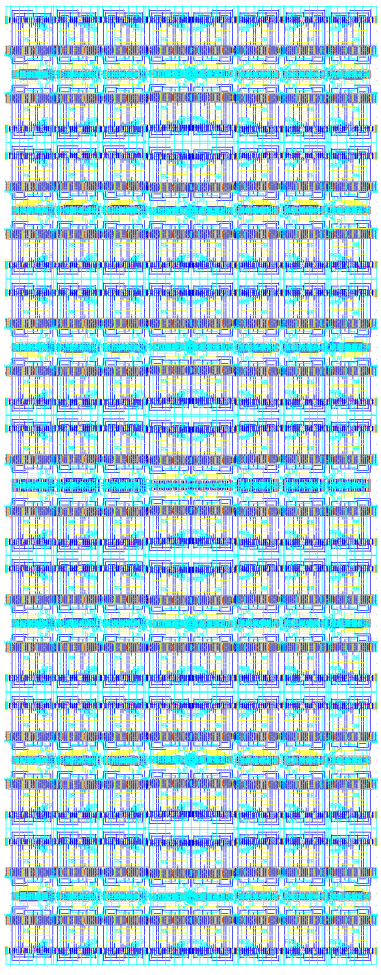

Figure 5.10: Layout design of 64×32 8T SRAM array for PUF implementation

5.12 Post-layout simulation

Extracted parasitics were used in post-layout simulations. The effect of layout-induced delays on the PUF output is seen in Figure 5.11. Although there were slight timing deteriorations, functionality was unaffected. Table 5.1 represent comparison of pre-layout and post-layout simulation results for PUF implementation.

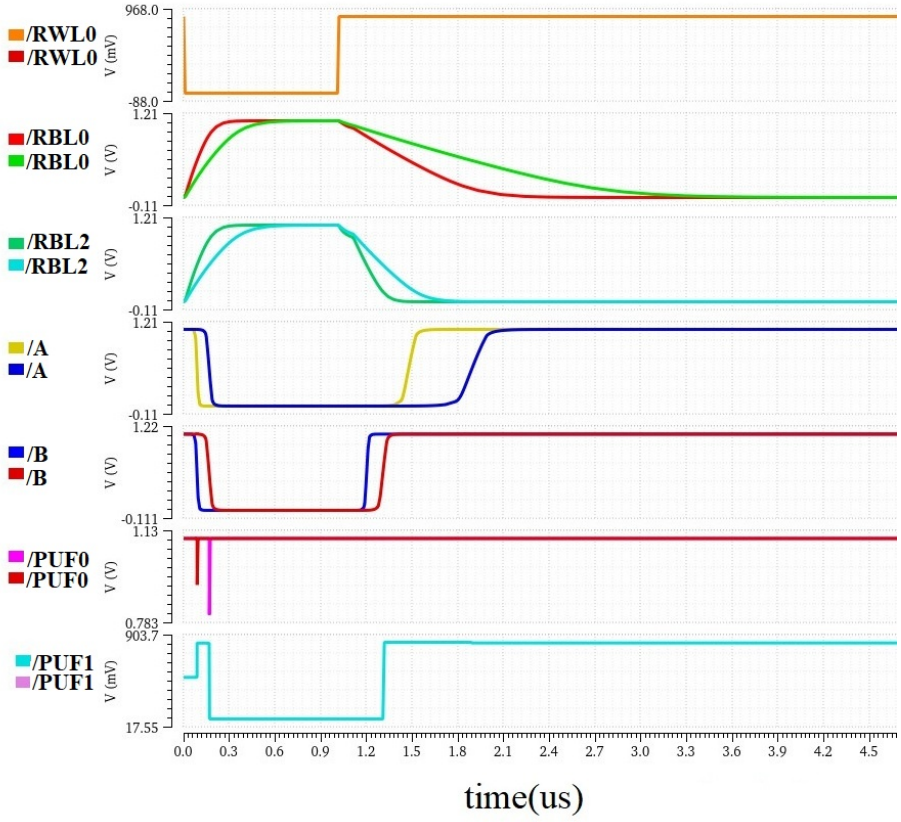

Figure 5.11: Transient reaponse of post layout simulation for PUF implementation

5.13 Improved BER and instability

Findings post-layout adjustments resulted in increased stability and BER. The BER and instability before and after arrangement are contrasted in Figures 5.12 and 5.13.

Table 5.1: Comparison of pre-layout and post-layout simulation results for PUF implementation

Parameter	Pre-layout simulation	Post-layout simulation
Delay between RWL0 and RBL0 (ns)	444.45	842.36
Delay between RWL0 and RBL1 (ns)	180.922	282.8
Static power (mW)	1.22	1.18
Peak power (mW)	129	126.26

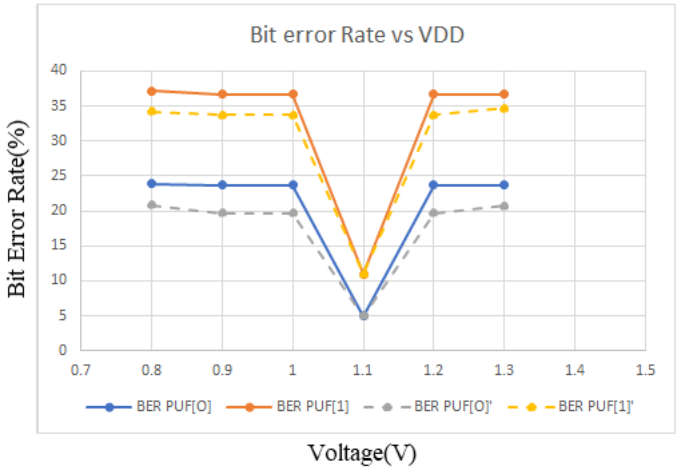

Figure 5.12: Comparision of BER vs supply voltage before and after error reduction technique

5.14 Summary

This chapter used statistical analysis, physical layouts, and precise waveforms to validate the PUF architecture implementation. The outcomes demonstrate how well the 8T SRAM and bitline discharge approach work together to produce reliable, steady, and random outputs that are appropriate for hardware security applications. Table 5.2 shows the key parameters of the designed 8T SRAM-based PUF.

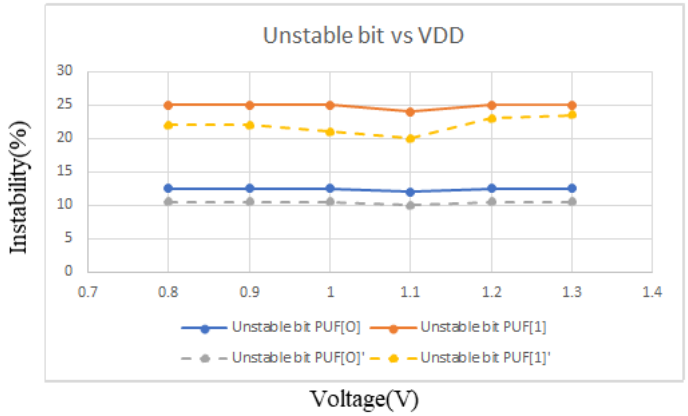

Figure 5.13: Comparision of instability vs voltage before and after error reduction technique

Metric	This Work	JSSC 2022 [1]	JSSC 2021 [16]	ISSCC 2021 [5]
Technology	40nm	28nm	130nm	40nm
Entropy source	Bitline discharge method	SRAM bitcell read current	Hybrid SRAM	Hybrid ring oscillator
Layout area of 64×32 8T SRAM array (μm^2)	1905	1125	2307	21,675
V_{DD} (V)	0.8–1.3	0.75–1.05	0.5–0.7	0.7–1.4
Temperature (°C)	-25 to 105	-25 to 100	-40 to 120	-40 to 125
Unstable bits (%) @ Nominal V	10.2 (LSB), 20.4 (MSB)	11.4 (LSB), 29.5 (MSB)	2.71	0.39
Bit Error Rate (BER %) @ Nominal V	5 (LSB), 12 (MSB)	1.8 (LSB), 3.78 (MSB)	0.29	0.03
PUF energy (fJ/bit)	85	72	16,76	39

Table 5.2: Comparison of this work with state-of-the-art PUF designs

Chapter 6

Conclusion and Future Work

6.1 Summary of achievements

In this thesis, a physical unclonable function architecture based on 8T SRAM cells using bitline discharge techniques was designed, simulated, and analyzed. The study effectively shown that high entropy responses appropriate for secure key generation and device authentication can be achieved by carefully managing the bitline discharge path and taking use of process variation in 8T SRAM-based memory cells.
The following are some of the main achievements:

- Design and simulation of 8T SRAM-based PUF: This design uses a small and effective 8T SRAM cell to combine hardware fingerprinting and randomness extraction. Pre-layout and post-layout simulations were used to evaluate the PUF module, guaranteeing its dependability in authentic parasitic scenarios.

- Methods for reducing bit error rates: A number of methods were investigated to increase the PUF responses stability and repeatability. Among these were improvements at the circuit level and the use of lightweight error correction to guarantee resilience to changes in the environment and voltage.

- Post-layout verification: The entire layout design was put into practice in UMC 40nm technology, and then post-layout simulations were run to assess timing,

power, and bit error rate. This confirmed the design functional correctness and process robustness.

6.2 Contributions to the field

This thesis makes the following notable contributions:

- **Bitline discharge technique for entropy enhancement**: Bitline discharge was shown to be a useful technique for extracting and amplifying entropy, which greatly enhances the uniqueness needed for PUF applications.
- **Increased BER and stability trade-off**: Introduced and assessed circuit-level techniques that help achieve a workable equilibrium by addressing the trade-off of repeatability for PUF.
- **Validation via layout and post-layout simulations**: By doing full-layout design and parasitic-aware simulation, this method helped close the gap between theoretical modeling and real-world application.

6.3 Possible improvements

Although the current architecture provides a solid basis, there are a number of opportunities for improvement:

6.3.1 Area optimization

Compared to a traditional 6T cell, the 8T SRAM-based device takes up more silicon area yet providing better stability and control.

- Reducing the overhead of peripheral circuitry could be one area of future effort.
- To reduce the area footprint, transistor-level improvements are being investigated.
- To cut down on redundancy, several SRAM cells can use shared control logic.

6.3.2 Low-power version

Future work may concentrate on the following areas to facilitate deployment in ultra-low-power contexts, such as wearable sensors or IoT edge devices:

- Reducing leakage and dynamic power usage using strategies like clock gating or power gating.
- Using adaptive biasing techniques to operate at sub-threshold or near-threshold voltages.

6.3.3 Integration into SoC

For practical security applications, the PUF-TRNG module can be directly integrated into a larger system-on-chip (SoC) to further improve the architecture. Standard interface development, such as advanced microcontroller bus architecture or APB-based access control, would be necessary for this.

- Evaluating interoperability with cryptographic engines and secure boot modules.
- Dealing with possible noise interference or cross-talk when digital and analog blocks are placed together.

References

[1] Taneja, Sachin, Viveka Konandur Rajanna, and Massimo Alioto. "In-memory unified TRNG and multi-bit PUF for ubiquitous hardware security." *IEEE Journal of Solid-State Circuits* 57, no. 153-166, Dec 2021.

[2] Taneja, Sachin, and Massimo Alioto. "PUF architecture with run-time adaptation for resilient and energy-efficient key generation via sensor fusion." *IEEE Journal of Solid-State Circuits* 56, no. 2182-2192, Mar 2021.

[3] Holcomb, Daniel E., Wayne P. Burleson, and Kevin Fu. "Power-up SRAM state as an identifying fingerprint and source of true random numbers." *IEEE Transactions on Computers* 58, no. 9, pp. 1198-1210, Nov 2008.

[4] Gao, Bin, Bohan Lin, Xueqi Li, Jianshi Tang, He Qian, and Huaqiang Wu. "A unified PUF and TRNG design based on 40-nm RRAM with high entropy and robustness for IoT security." *IEEE Transactions on Electron Devices* 69, no. 2, pp. 536-542, Jan 2022.

[5] Park, Jaehan, and Jae-Yoon Sim. "36.4 A physically unclonable function combining a process mismatch amplifier in an oscillator collapse topology." *In 2021 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 64, pp. 504-506. IEEE, Feb 2021.

[6] Jeon, Duhyun, Jong Hak Baek, Yong-Duck Kim, Jaeseong Lee, Dong Kyue Kim, and Byong-Deok Choi. "A Physical Unclonable Function With Bit Error Rate $< 2.3 \times 10^{-8}$ Based on Contact Formation Probability Without Error Correction

Code." *IEEE Journal of Solid-State Circuits*, vol. 55, no. 3, pp. 805–816, Nov 2019.

[7] Riya, S. S., and V. Lalu. "Stable cryptographic key generation using SRAM based Physical Unclonable Function." *In 2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 653-657. IEEE, Sep 2020.

[8] Cortez, Mafalda, Apurva Dargar, Said Hamdioui, and Geert-Jan Schrijen. "Modeling SRAM start-up behavior for physical unclonable functions." *In 2012 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*, pp. 1-6. IEEE, Oct 2012.

[9] Satpathy, Sudhir K., Sanu K. Mathew, Raghavan Kumar, Vikram Suresh, Mark A. Anders, Himanshu Kaul, Amit Agarwal, Steven Hsu, Ram K. Krishnamurthy, and Vivek De. "An all-digital unified physically unclonable function and true random number generator featuring self-calibrating hierarchical Von Neumann extraction in 14-nm tri-gate CMOS." *IEEE Journal of Solid-State Circuits* 54, no. 4, pp. 1074-1085, Jan 2019.

[10] Chen, Zhuojun, Ming Wu, Yifeng Zhou, Renlong Li, Jinzhe Tan, and Ding Ding. "Puf-cim: Sram-based compute-in-memory with zero bit-error-rate physical unclonable function for lightweight secure edge computing." *IEEE Transactions on Very Large Scale Integration (VLSI)* Systems 31, no. 8, pp. 1234-1247, Jun 2023.

[11] Zhang, Ruilin, Xingyu Wang, Kunyang Liu, and Hirofumi Shinohara. "A 0.186-pJ per bit latch-based true random number generator featuring mismatch compensation and random noise enhancement." *IEEE Journal of Solid-State Circuits* 57, no. 8, pp. 2498-2508, Mar 2022.

[12] Taneja, Sachin, and Massimo Alioto. "Fully synthesizable unified true random number generator and cryptographic core." *IEEE Journal of Solid-State Circuits* 56, no. 10, pp. 3049-3061, Sep 2021.

[13] Wu, Meng-Yi, Tsao-Hsin Yang, Lun-Chun Chen, Chi-Chang Lin, Hao-Chun Hu, Fang-Ying Su, Chih-Min Wang et al. "A PUF scheme using competing oxide rupture with bit error rate approaching zero." *In 2018 IEEE International Solid-State Circuits Conference-(ISSCC)*, pp. 130-132. IEEE, Feb 2018.

[14] Mahmud, Md Ishtyaq, Pintu Kumar Sadhu, Venkata P. Yanambaka, and Ahmed Abdelgawad. "Vxorpuf: a vedic principles-based hybrid XOR arbiter PUF for robust security in IoMT." *In IFIP International Internet of Things Conference*, pp. 246-261. Cham: Springer Nature Switzerland, Oct 2023.

[15] Choi, Yunhyeok, Bohdan Karpinskyy, Kyoung-Moon Ahn, Yongsoo Kim, Soonkwan Kwon, Jieun Park, Yongki Lee, and Mijung Noh. "Physically unclonable function in 28nm fdsoi technology achieving high reliability for aec-q 100 grade 1 and iso 26262 asil-b." *In 2020 IEEE International Solid-State Circuits Conference-(ISSCC)*, pp. 426-428. IEEE, Feb 2020.

[16] Liu, Kunyang, Xinpeng Chen, Hongliang Pu, and Hirofumi Shinohara. "A 0.5-V hybrid SRAM physically unclonable function using hot carrier injection burn-in for stability reinforcement." *IEEE Journal of Solid-State Circuits* 56, no. 7, pp. 2193-2204, Nov 2020.

[17] Taneja, Sachin, and Massimo Alioto. "PUF architecture with run-time adaptation for resilient and energy-efficient key generation via sensor fusion." *IEEE Journal of Solid-State Circuits* 56, no. 7, pp. 2182-2192, Mar 2021.

[18] McGrath, Thomas, Ibrahim E. Bagci, Zhiming M. Wang, Utz Roedig, and Robert J. Young. "A puf taxonomy." *Applied physics reviews* 6, no. 1, pp. 011303, Mar 2019.

[19] Zhang, Ji-Liang, Gang Qu, Yong-Qiang Lv, and Qiang Zhou. "A survey on silicon PUFs and recent advances in ring oscillator PUFs." *Journal of computer science and technology* 29, no. 4, pp. 664-678, Jul 2014.

[20] Zerrouki, Fahem, Samir Ouchani, and Hafida Bouarfa. “A survey on silicon PUFs.” *Journal of Systems Architecture* 127, pp. 102514, Jun 2022.