Mixed-Signal Physically Unclonable Function

with CMOS Capacitive Cells

by
Kamal T. Y. Kamal

A Thesis
presented to
The University of Guelph

In partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in
Engineering

Guelph, Ontario, Canada
© Kamal Y. Kamal, October, 2019
ABSTRACT

MIXED-SIGNAL PHYSICALLY UNCLONABLE FUNCTION WITH CMOS CAPACITIVE CELLS

Kamal Kamal                      Advisor
University of Guelph, 2019       Radu Muresan, Ph.D., P.Eng

An electronic physically unclonable function usually includes an on-chip error-correcting code unit, which is vulnerable to security attacks and adds area, power, and data processing time overheads. This dissertation proposes two mixed-signal physically unclonable function circuits for authentication purposes. The first is called the capacitive physically unclonable function (C-PUF), which applies an expanding challenge-response pair approach; it processes a 21-bit challenge to generate a 128-bit response word. The second design is called the enhanced capacitive physically unclonable function (EC-PUF), which processes a 64-bit challenge word to generate also a 128-bit response word. The later design uses short-wide metal-oxide-semiconductor field-effect transistors within its capacitive cells, to allow for more intrinsic variations throughout the fabrication process, which improves the uniqueness of the response chunks at the cost of compatibility. Both designs divide the input challenge word over multiple computational groups to decrease processing time, increase security, and eliminate the need for error-correcting code units. Both the C-PUF and the EC-PUF designs were simulated using 45 nm complementary metal-oxide-semiconductor technology. The average power was 1.05 mW and 921.67 μW, the layout area was 3276 μm² and 22,470 μm², and the average data processing time was 30 μs and 118 μs, respectively.
I want to thank my advisor, Dr. Radu Muresan for the opportunity to pursue my Ph.D. degree. Without his assistance and patience, this work would not have been possible. I would also like to thank my advisory committee members, Dr. Stefano Gregori and Dr. Charlie Obimbo for their feedback and guidance. I cannot forget to thank the technical support staff, Joel Best, Matthew Kent, and Jeff Madge, for their prompt assistance during my degree course. Last but not least, I would like to thank my family for their invaluable support.
# Table of Contents

Abstract ........................................................................................................................................... ii

Acknowledgments .......................................................................................................................... iii

List of Contents .............................................................................................................................. iv

List of Tables .................................................................................................................................. x

List of Figures ................................................................................................................................ xi

List of Symbols and Abbreviations .............................................................................................. xiii

List of Appendices ..................................................................................................................... xviii

Chapter 1 ......................................................................................................................................... 1

Introduction ..................................................................................................................................... 1

1.1. Objective .............................................................................................................................. 3

1.2. Objectives and Contributions ............................................................................................... 3

Chapter 2 ......................................................................................................................................... 5

Background ..................................................................................................................................... 5

2.1. Notations and Definitions ..................................................................................................... 5

2.2. PUF Properties ..................................................................................................................... 8

2.3. Weak PUFs .......................................................................................................................... 10

2.4. Strong PUFs ....................................................................................................................... 11

2.5. Evaluation Metrics of PUFs ............................................................................................... 13
<table>
<thead>
<tr>
<th>2.5.1. Inter-Distance</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.2. Intra-Distance</td>
<td>15</td>
</tr>
<tr>
<td>2.5.3. Identifiability</td>
<td>16</td>
</tr>
<tr>
<td>2.6. Reproducibility and Authentication Tolerance</td>
<td>17</td>
</tr>
<tr>
<td>2.7. PUF Physical Unclonability</td>
<td>18</td>
</tr>
<tr>
<td>2.8. PUF Anti-Tamper Sensitivity (Tamper Evidence)</td>
<td>19</td>
</tr>
<tr>
<td>2.9. PUF Unpredictability (Mathematical Unclonability)</td>
<td>21</td>
</tr>
<tr>
<td>2.9.1. Randomness (Entropy)</td>
<td>22</td>
</tr>
<tr>
<td>2.9.2. Correlation Among Bits</td>
<td>23</td>
</tr>
<tr>
<td>2.9.3. Correlation Among Chips</td>
<td>23</td>
</tr>
<tr>
<td>2.9.4. One-Wayness</td>
<td>24</td>
</tr>
<tr>
<td>2.10. Smart Card Basics</td>
<td>24</td>
</tr>
<tr>
<td>2.11. PUF Usage Phases</td>
<td>28</td>
</tr>
<tr>
<td>2.11.1. Enrollment Phase</td>
<td>28</td>
</tr>
<tr>
<td>2.11.2. Authentication Phase</td>
<td>30</td>
</tr>
<tr>
<td>2.12. Untrusted Verifier</td>
<td>34</td>
</tr>
<tr>
<td>2.13. Sources of Variations</td>
<td>35</td>
</tr>
<tr>
<td>2.14. Sources of Permanent Variations</td>
<td>36</td>
</tr>
<tr>
<td>2.14.1. Global Mismatch Sources</td>
<td>36</td>
</tr>
<tr>
<td>2.14.2. Local Mismatch Sources</td>
<td>36</td>
</tr>
</tbody>
</table>
3.3.1. ICID PUF ..................................................................................................................... 60

3.3.2. Silicon Nanokey .......................................................................................................... 63

3.4. Problems in PUFs ............................................................................................................... 66

Chapter 4 ....................................................................................................................................... 68

Objective and Methodology of the C-PUF ................................................................................... 68

4.1. The C-PUF Design Objectives and Constrains ..................................................................... 68

4.1.1. Strong PUF Constrains ............................................................................................... 68

4.1.2. Design Constrains ........................................................................................................ 73

4.2. The CRP Expansion Concept ............................................................................................. 73

4.3. The C-PUF Authentication Protocol .................................................................................. 75

4.4. The C-PUF Diagram .......................................................................................................... 76

4.5. Combining Various Temporary Errors .............................................................................. 82

4.6. The Signed Error Rate ........................................................................................................ 82

4.7. The Verification Approach ................................................................................................. 83

4.8. Experimental Settings ........................................................................................................ 84

Chapter 5 ....................................................................................................................................... 86

Results and Discussions of the C-PUF ......................................................................................... 86

5.1. Identifiability (Inter- Distance) ........................................................................................ 86

5.2. Neutralizing Cross Talk and Other Noise Effects .............................................................. 88

5.3. The Temperature Variations .............................................................................................. 88
5.4. The Voltage Variations ...................................................................................................... 91

5.5. The C-PUF Compatibility .................................................................................................. 93

5.5.1. Area ............................................................................................................................. 93

5.5.2. Time ............................................................................................................................. 95

5.5.3. Power ........................................................................................................................... 96

Chapter 6....................................................................................................................................... 99

Objective and Methodology of the EC-PUF ................................................................................. 99

6.1. CRP Expansion .................................................................................................................. 99

6.2. CRP Enhancement .............................................................................................................. 99

6.3. EC-PUF Schematics ......................................................................................................... 100

6.4. Basic Authentication Protocol .......................................................................................... 104

6.5. The EC-PUF Model ......................................................................................................... 106

6.6. The EC-PUF Frequency Aspects ..................................................................................... 108

6.7. Error Handling .................................................................................................................. 111

6.8. Fabrication Variation Enhancement ................................................................................. 112

6.9. Other Applicable Fabrication Variation Enhancements ................................................... 113

6.10. Measurement Enhancement ........................................................................................... 113

6.11. The EC-PUF Stability Aspects ....................................................................................... 113

6.11.1. Thermal Aspects ...................................................................................................... 113

6.11.2. Transient Aspects .................................................................................................... 114
6.12. Security Aspects ............................................................................................................. 115
6.13. Experimental Setup ........................................................................................................ 116
6.14. Simulating the Environmental Variation ........................................................................ 117
6.15. Simulating the Fabrication Variation ............................................................................. 118

Chapter 7 ..................................................................................................................................... 119
Results and Discussions of the EC-PUF ..................................................................................... 119
7.1. Environmental Variation Simulation Results ................................................................... 120
7.1.1. Thermal Variation Simulation Results ...................................................................... 120
7.1.2. Voltage Variation Simulation Results ....................................................................... 129
7.2. Fabrication Variations Simulation Results ....................................................................... 135
7.2.1. Local Variation Simulation Results ........................................................................... 135
7.2.2. Global Variation Simulation Results ......................................................................... 137
7.2.3. Total Fabrications Variation Simulation Results ....................................................... 138
7.3. Discussions ....................................................................................................................... 139

Chapter 8 ..................................................................................................................................... 143
Conclusions and Future Work .................................................................................................... 143
8.1. Conclusions ...................................................................................................................... 143
8.2. Future Work ..................................................................................................................... 147
References ................................................................................................................................... 148
List of Tables

Table 3.1. Fabricated ICID ASIC properties .................................................................62
Table 3.2. ICID block measurements .........................................................................63
Table 3.3. Fabricated Nanokey ASIC properties .........................................................66
Table 3.4. Nanokey PUF measurements ...................................................................66
Table 4.1. Ranges of PUF experimental settings .......................................................84
Table 5.1. The C-PUF Power consumption details ......................................................98
Table 6.1. Acronyms and values of the parameters of each CC ..............................106
Table 6.2. Test Framework .......................................................................................117
Table 7.1. Area and Power Comparisons ................................................................141
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 2.1</td>
<td>Similar PUF individuals generate different responses</td>
<td>6</td>
</tr>
<tr>
<td>Fig. 2.2</td>
<td>PUF authenticity cause and effect block diagram</td>
<td>7</td>
</tr>
<tr>
<td>Fig. 2.3</td>
<td>PUF properties</td>
<td>8</td>
</tr>
<tr>
<td>Fig. 2.4</td>
<td>PUF authentication with trusted verifier</td>
<td>32</td>
</tr>
<tr>
<td>Fig. 2.5</td>
<td>Regions of planar n-MOSFET</td>
<td>48</td>
</tr>
<tr>
<td>Fig. 3.1</td>
<td>PUF classification</td>
<td>54</td>
</tr>
<tr>
<td>Fig. 3.2</td>
<td>Array of transistors producing a sequential random voltage</td>
<td>61</td>
</tr>
<tr>
<td>Fig. 3.3</td>
<td>ICID block diagram</td>
<td>62</td>
</tr>
<tr>
<td>Fig. 3.4</td>
<td>Row of the inverter array</td>
<td>64</td>
</tr>
<tr>
<td>Fig. 3.5</td>
<td>Silicon nanokey PUF block diagram</td>
<td>64</td>
</tr>
<tr>
<td>Fig. 4.1</td>
<td>Simplified diagram of the C-PUF</td>
<td>76</td>
</tr>
<tr>
<td>Fig. 4.2</td>
<td>Simplified capacitive cell</td>
<td>78</td>
</tr>
<tr>
<td>Fig. 5.1</td>
<td>Effect of 5% variation in the n-MOSFET capacitors</td>
<td>87</td>
</tr>
<tr>
<td>Fig. 5.2</td>
<td>Temperature effect on discharge time of the largest capacitive cells</td>
<td>89</td>
</tr>
<tr>
<td>Fig. 5.3</td>
<td>Voltage effect on discharge time of the largest capacitive cell</td>
<td>91</td>
</tr>
<tr>
<td>Fig. 5.4</td>
<td>Voltage effect on discharge time of the smallest capacitive cell</td>
<td>92</td>
</tr>
<tr>
<td>Fig. 6.1</td>
<td>Enhanced capacitive physically unclonable function block diagram</td>
<td>101</td>
</tr>
<tr>
<td>Fig. 6.2</td>
<td>Capacitive cell module</td>
<td>101</td>
</tr>
<tr>
<td>Fig. 6.3</td>
<td>MOSFET-based R-2R ladder</td>
<td>101</td>
</tr>
<tr>
<td>Fig. 6.4</td>
<td>Data dependency of a response chunk</td>
<td>106</td>
</tr>
<tr>
<td>Fig. 7.1</td>
<td>Central frequency vs temperature</td>
<td>120</td>
</tr>
</tbody>
</table>
Fig. 7.2. Discharge time and response of cell-A vs. temperature ...............................................121
Fig. 7.3. Discharge time and response of cell B vs. temperature ...............................................122
Fig. 7.4. Discharge time ratio of cell-B to cell-A vs. temperature ...............................................123
Fig. 7.5. Response ratio of cell-B to cell-A vs. temperature .......................................................123
Fig. 7.6. Discharge times trends of the smallest and the largest CCs vs. temperature ..............127
Fig. 7.7. Central frequency vs. $V_{DD}$ .....................................................................................129
Fig. 7.8. Discharge time and response chunk of cell-A vs. $V_{DD}$ ...............................................130
Fig. 7.9. Discharge time and response of cell-B vs. $V_{DD}$ ........................................................131
Fig. 7.10. Discharge time ratio of cell-B to cell-A vs. $V_{DD}$ ...................................................132
Fig. 7.11. Response ratio of cell-B to cell-A vs. $V_{DD}$ ..........................................................132
Fig. 7.12. Discharge times and trends of the smallest and the largest CCs vs. $V_{DD}$ ..........134
Fig. 7.13. Simulated effect of the local randomness on discharge time of the largest CC ....136
Fig. 7.14. Simulated effect of the local randomness on discharge time of the smallest CC ....136
Fig. 7.15. Simulated effect of the global randomness on discharge time of the largest CC ....137
Fig. 7.16. Simulated effect of the global randomness on discharge time of the smallest CC ....137
Fig. 7.17. Simulated effect of the total randomness on discharge time of the largest CC .......138
Fig. 7.18. Simulated effect of the total randomness on discharge time of the smallest CC .......138
List of Symbols and Abbreviations

AAC Application Authentication Cryptogram
ADC Analog-to-Digital Convertor
API Application Program Interface
ARQC Authorization Request Cryptogram
ASIC Application-Specific Integrated Circuit
ATM Automated Teller Machine
BCH Bose, Chaudhuri and Hocquenghem (code)
BER Bit Error Rate
BRR Bit Rejection Rate
CAP Chip Authentication Program
CC Capacitive Cell
\( C_{DB} \) Capacitance of a Drain-Bulk Junction
\( C_{SB} \) Capacitance of a Source-Bulk Junction
CMOS Complementary Metal-Oxide-Semiconductor
CPUF Controlled PUF
C-PUF Capacitive PUF
CRP Challenge-Response Pair
CV Coefficient of Variation
CVM Cardholder Verification Method
DAC Digital-to-Analog Convertor
DEC Decoder
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB</td>
<td>Database</td>
</tr>
<tr>
<td>DRAM</td>
<td>Dynamic Random-Access Memory</td>
</tr>
<tr>
<td>$\Delta L_D$</td>
<td>Channel length extension of depletion at the bulk-Source region</td>
</tr>
<tr>
<td>$\Delta L_S$</td>
<td>Channel length extension of depletion at the bulk-Source region</td>
</tr>
<tr>
<td>$\Delta V_{T0}$</td>
<td>Variation in the Initial Threshold Voltage</td>
</tr>
<tr>
<td>DEPA</td>
<td>Differential Electromagnetic Analysis</td>
</tr>
<tr>
<td>DPAuth</td>
<td>Dynamic Passcode Authentication</td>
</tr>
<tr>
<td>DPA</td>
<td>Differential Power Analysis</td>
</tr>
<tr>
<td>DRM</td>
<td>Digital Rights Management</td>
</tr>
<tr>
<td>ECC</td>
<td>Error Correcting Code</td>
</tr>
<tr>
<td>EC-PUF</td>
<td>Enhanced Capacitive Physically Unclonable Function</td>
</tr>
<tr>
<td>EEPROM</td>
<td>Electrically Erasable Programmable Read-Only Memory</td>
</tr>
<tr>
<td>EMA</td>
<td>Electro-Magnetic Analysis</td>
</tr>
<tr>
<td>EMV</td>
<td>Europay, MasterCard, and Visa, referring to their smart cards</td>
</tr>
<tr>
<td>EMVCo</td>
<td>Europay, MasterCard, and Visa Consortium</td>
</tr>
<tr>
<td>FAR</td>
<td>False Acceptance Rate</td>
</tr>
<tr>
<td>FF</td>
<td>Feed-Forward</td>
</tr>
<tr>
<td>FIB</td>
<td>Focused Ion Beam</td>
</tr>
<tr>
<td>FPGA</td>
<td>Field-Programmable Gate Array</td>
</tr>
<tr>
<td>FRR</td>
<td>False Recognition Rate</td>
</tr>
<tr>
<td>IACs</td>
<td>Issuer Action Codes</td>
</tr>
<tr>
<td>IC</td>
<td>Integrated Circuit</td>
</tr>
<tr>
<td>ICID</td>
<td>Integrated Circuit Identifier</td>
</tr>
</tbody>
</table>
ID Identity number
IP Intellectual Property
IPR Intellectual Property Rights
IoT Internet of Things
kbps kilo bit per second (data transfer rate)
L Channel Length of a MOSFET
LC Inductor(L)-Capacitor(C)
LFSR Linear Feedback Shift Register
Le Effective Channel Length of a MOSFET
LM Mask Length of a MOSFET Channel
L_{dep} Depletion Length of a Source/Drain junction
MAC Message Authentication Code
MGG Metal Gate Granularity
MOSFET Metal-Oxide-Semiconductor Field-Effect Transistor
MUX Multiplexer
NFC Near Field Communication
n-MOSFET n-channel MOSFET
NRC Non-Rouletted Response Chunk
NVM Non-Volatile Memory
OEPA Optical Emission Power Analysis
OPC Optical Proximity Correction
PIN Personal Identification Number
POK Physically Obfuscated Key
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-MOSFET</td>
<td>positive-channel MOSFET</td>
</tr>
<tr>
<td>PUF</td>
<td>Physically Unclonable Function</td>
</tr>
<tr>
<td>RC</td>
<td>Response Chunk</td>
</tr>
<tr>
<td>RDD</td>
<td>Random Descrete Dopant</td>
</tr>
<tr>
<td>RDF</td>
<td>Random Dopant Fluctuation</td>
</tr>
<tr>
<td>RFID</td>
<td>Radio-Frequency Identification</td>
</tr>
<tr>
<td>ROM</td>
<td>Read-Only Memory</td>
</tr>
<tr>
<td>RRC</td>
<td>Reference Response Chunk</td>
</tr>
<tr>
<td>RTL</td>
<td>Register Transfer Level</td>
</tr>
<tr>
<td>RSA</td>
<td>Rivest, Shamir and Adleman (algorithm)</td>
</tr>
<tr>
<td>SCADA</td>
<td>Supervisory Control And Data Acquisition</td>
</tr>
<tr>
<td>SEM</td>
<td>Scanning Electron Microscopes</td>
</tr>
<tr>
<td>SEPA</td>
<td>Simple Electromagnetic Power Analysis</td>
</tr>
<tr>
<td>SHIC</td>
<td>Super-High Information Content</td>
</tr>
<tr>
<td>SIM</td>
<td>Subscriber Identification Module</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal-to-Interference-plus-Noise Ratio</td>
</tr>
<tr>
<td>SOI</td>
<td>Silicon-on-insulator (technology)</td>
</tr>
<tr>
<td>SPA</td>
<td>Simple Power Analysis</td>
</tr>
<tr>
<td>SRAM</td>
<td>Static Random-Access Memory</td>
</tr>
<tr>
<td>TACs</td>
<td>Terminal Action Codes</td>
</tr>
<tr>
<td>TC</td>
<td>Transaction Certificate</td>
</tr>
<tr>
<td>TPM</td>
<td>Trusted Platform Module</td>
</tr>
<tr>
<td>TVR</td>
<td>Terminal Verification Results</td>
</tr>
</tbody>
</table>
\( \mu_n \)  
Mobility of Electrons

\( \mu_p \)  
Mobility of holes

ULSI  
Ultra Large Scale Integration

USA  
United State of America

USD  
United State of America’s Dollar

\( V_{DD} \)  
Drain Supplied Voltage

VLSI  
Very Large Scale Integration

VS  
Verifier Software

Vss  
Source Supplied Voltage

\( V_T \)  
Threshold Voltage

\( V_{T0} \)  
Initial Threshold Voltage

XOR  
Exclusive OR (logical)
List of Appendices

Appendix A: Average Hamming Distance ................................................................. 167
Appendix B: The Average Hamming Distance Distribution ........................................ 170
Appendix C: Clonability Metric of a PUF ............................................................... 171
Chapter 1

Introduction

Hardware authentication is getting rising importance to cope with the growing security concerns. Among the significant threats to supervisory control and data acquisition (SCADA) and the internet of things (IoT) is counterfeiting sensors, cameras, and radio-frequency identification (RFID) tags to access vital premises, exploit strategical projects, or to counterfeit machine-readable documents, and RFID-tagged goods [1]. This leads the tendency to adopt a verification policy for electronic systems to enable an authorized organization to remotely check a device’s authenticity [2]–[10]. Furthermore, there is a wide growth of contactless machine-readable passports, persons mobility monitoring tags, vehicle mobility monitoring tags [11], public transportation smart tickets [12], shipped goods mobility monitoring tags [13], physical access smart cards [14], banking smart cards, and near field communication (NFC) in smartphones [15]. Many of those objects have been venerable to security attacks [16]–[18].

On-chip hardware security applications for intellectual property (IP) protection copyright protection of integrated circuits and other physical objects for some applications may need hardware-based authenticity solutions. Storing a secret key on-chip is not a secure approach against invasive and side-channel attacks. Cloning an identification number (ID) stored on-chip on a non-volatile memory (NVM) of some smart cards is easy, instant, and costs only 25 USD for the multi-use reader/writer equipment and few cents for each blank smart card [19]. To secure a smart card or an RFID tag against counterfeiting, it is recommended to embed a physically unclonable function (PUF) within an electronic chip [20]–[25].
The mobility of such devices adds more security threats; for example, a smart card would sometimes operate in untrusted environments where an attacker may have time and tools to breach such systems [26], [27].

Digital PUFs suffer vulnerability to modeling (or mathematical/logic) attacks [28]. Furthermore, the analogous nature of the basic electronic devices, such as a metal-oxide-semiconductor field-effect transistor (MOSFET), capacitor, inductor, or a wire, influences the propagation delay of data signals within a delay-based arbitrator PUF circuit. This makes even a digital PUF vulnerable to temperature and voltage variations, and such temporary variations may alter the response; hence, most PUF chip designs include an error correction code (ECC) unit [26], [29]–[31]. Most PUFs also combine the authentication with identity and/or cryptographic-key generation [32], [33]. This implies no error tolerance, and thus ECC units were essential for such systems. ECC units cost area, power, and data processing time and pose security threats, such as invasive attacks [34] and side-channel attacks [34], [35].

A PUF would sometimes generate non-stable response bits, which might be non-correctable even by the embedded ECC unit. This reduces the number of reliable response bits among the total response bits. Flagging the non-stable bits requires extra on-chip complicated units, it also complicates the enrolment phase and raises its cost.

Motivated by that, this dissertation proposes the enhanced capacitive physically unclonable function (EC-PUF) for authentication purposes only, and not to generate an identity nor a cryptographic key. The basic protocol of the EC-PUF requires that the ID information be retrieved from on-chip NVM, like most of today’s ID-based smart cards and RFID tags but having a PUF on-chip makes ID secrecy less crucial. The EC-PUF generates its response by digitizing the discharge times of challenged capacitors. Dividing the input challenge word into chunks of bits
and processing them simultaneously through multi computational units, decreases processing time, increases security, and eliminates the need for an ECC unit.

The EC-PUF structures its response into chunks of bits, each response chunk (RC) is a binary string that represents the discharge time of a capacitive cell (CC). The EC-PUF design presumes effects for the environmental variations of temperature and supply voltage among all RCs are analogous, which result in correlated shifts in the values of all RCs. The design enables the verifier software (VS) at the remote authenticator side to analyze the environmental-driven response shifts without needing on-chip ECC unit. The dissertation introduces the basic equations for the remote VS. The EC-PUF design also presents a novel expanding concept which saves area, power, and data processing time, by generating a larger number of response bits against a smaller number of challenge bits.

1.1. Objective

PUFs usually require an on-chip embedded ECC unit, which adds more security threats. The main objective of this work is the development of a mixed-signal PUF on-chip module without an ECC unit. The design was developed using a general-purpose development kit library of 45 nm (gpdk045nm) for complementary metal-oxide-semiconductor (CMOS) technology.

1.2. Objectives and Contributions

In this dissertation, there are theoretical and experimental contributions. The theoretical contributions can be briefed as follows:

1) Refine the classification of strong vs. weak PUFs.

2) Present a refined mathematical representation for PUF challenge-response pair (CRP) reproducibility, PUF uniqueness, identifiability, and unclonability.
3) Propose a PUF design which helps the authenticating phase to cope with the effects of environmental variation on the PUF.

4) Propose a basic verifier algorithm to verify the PUF depending on comparisons of different RCs.

5) Simulate the fabrication variations of the proposed PUF chip to evaluate the feasibility of obtaining unique responses.

The experimental contributions include proposing a new PUF methodology that attains security, compactness, compatibility, and feasibility. The dissertation presents two designs of the proposed methodology and simulates their functionalities under various environmental conditions. These PUF designs are to support the authentication purposes of various embedded systems, such as secure smart cards, subscriber identity module (SIM) cards, trusted platform modules (TPMs), digital rights management (DRM) systems, intellectual property rights (IPR), and RFID tags.

List of publications:


This chapter presents the basic concepts of PUFs, properties, and terms.

### 2.1. Notations and Definitions

In this work, random variables are denoted in upper-case symbols; for example, $X$ represents the input challenge bits to the PUF system; while, $Y$ represents the noisy output response bits of the system. When a response is considered a reference, it is denoted as $R$. If $Y$ is a random vector, then its length (the number of its bits) is denoted as $\|Y\|$. Hamming distance, $D(Y_1; Y_2)$ is a distance metric between two vectors of equal length representing the number of bits which have different values at the same positions, i.e.,

$$D(Y_1; Y_2) \triangleq \{ j: Y_{1,j} \neq Y_{2,j} \},$$

where $(j)$ indexes the position of the compared bits within the two vectors, $Y_1$ and $Y_2$.

Defining a PUF requires first to define the random function.

**Definition 2.1:** A random function is a function for which knowing some inputs and outputs does not enable guessing the output against the other non-observed inputs [36]. Then, a PUF can be defined as:

**Definition 2.2:** A PUF is an unclonable physical entity that works as a random physical function whose output is hard to determine theoretically but can be measured experientially. In other words, a PUF is a unique physical function that is easy to evaluate, yet impossible to clone physically, and should have a degree of randomness to make it hard to model mathematically [37]. Furthermore, the PUF qualities are inheritable, thereby any device equipped with an
inseparable PUF becomes unique and unclonable [36], [38], [39].

The input word to a random physical function is called a *challenge*, and the output word is called a *response*, both are usually digital signals, forming strings of bits. Utilizing PUFs for authentication purposes does not require a unique challenge word, but a unique response word against each challenge word. A PUF individual (or instance) uniqueness is attributed to its CRP uniqueness. There have been several attempts to authenticate physical objects; however, most researchers consider the dawn of PUF was when Ravikanth [38] introduced the concept of a physical one-way function, which was proposed for low-cost physical tokens.

The fabrication process of similarly manufactured MOSFET devices would have various sources of interinsic variations [40]. This variation leads to a variation in response, as shown in Fig. 2.1.

![Diagram](image)

**Fig. 2.1.** Similar PUF individuals generate different responses [41]
A group of similarly manufactured PUF devices forms a *class* of instances [42] that is denoted here as \((P\text{class})\), which are conventionally considered identical, but in fact, they are merely similar and not identical, and hence each PUF instance can generate a different response. Fig. 2.1 illustrates the uniqueness of responses against a challenge applied to some similarly fabricated objects. In general, applying a set of bits as an input challenge \((C)\) to similarly fabricated PUF devices can generate distinct responses \((R1, R2, R3, \text{ or } R4)\) each output response is a set of \(N\)-bit binary bits [41].

![Fig. 2.2. PUF authenticity cause and effect block diagram](image)

For example, a MOSFET’s properties within a class can be distinctive due to fabrication variations in; 1) size, during etching and coating stages, 2) gate insulator thickness \((t_{ox})\), and 3) doping concentrations [43]. These random fabrication variations unintentionally form unique properties for each chip entity, which makes it possible to authenticate each individual chip. It is common to call security systems equipped with PUFs as physical disorder security-based systems [44]. It is important to emphasize that the distinct physical properties of a PUF system may not necessarily represent a unique ID per se, but each CRP is unique within a \(P\text{class}\); therefore, such a physical-based ID of a PUF system is considered a dynamic ID.
2.2. PUF Properties

There are various properties a PUF should have to be useful for authentication purposes. Fig. 2.3 categorizes those properties into the defining ones which a PUF must have and another nice to have properties [45].

Fig. 2.3. PUF properties [45]

- **Defining Properties for PUFs**

  The following points highlight the most basic properties for PUFs when used for authenticity applications:

  - A PUF should be physically unclonable.
  - A PUF should be identifiable among other instances.¹
  - Robustness (or reproducibility), i.e., a PUF repeatedly gives the same output response against the same input challenge.
  - The defining properties of a PUF should be unique among other instances.²
- A PUF response should be evaluable; otherwise, it cannot be used as a PUF. Some PUFs are easier to evaluate than others. Though there has been no standardized measure to the evaluability of a PUF, the average time to obtain the PUF response has been adopted in most PUF articles.

- A PUF should be feasible to construct.

- **Nice-to-Have Properties**

A PUF may or may not have some additional qualities, such as:

- Truly Unclonable, this implies the following concepts:
  1) Physical unclonability
  2) Mathematical unclonability, i.e., not modellable by model attacks.
  3) Unpredictability; this refers to the randomness of the generated responses and the input challenges as well.

- Temper Evident

  A PUF is considered having such quality if its response changes after an invasive attack.

- One-way Function

  A PUF is considered a one-way function if observing its response does help to know the challenge, or else this indicates the function is mathematically modellable and therefore not secure.

  Regarding identification, authentication, one-wayness, and unpredictability, there are two notes to emphasize:
1) As marked by ¹ and ², many researchers try to make PUFs do both identification and authentication tasks, using a PUF’s response as an ID [45], [46]. Furthermore, some utilize the response as an encryption key [33].

2) Although Fig. 2.3 has followed a classification approach to present the PUF characteristics, the one-way and the mathematical unclonability qualities can be classified as parts of the unpredictability. A PUF’s properties and formulations can take many various forms [24], [45], [47]–[50]. Using PUFs for identification and/or key generation eliminates the tolerance for errors, which are commonly induced by environmental variations.

### 2.3. Weak PUFs

A weak PUF can have no input challenge or a few challenges, it is usually used to generate an ID. Weak PUFs are vulnerable to modeling attacks whether they have a small number of CRPs [51] or a large number of CRPs [52]. Beside extending the number of CRP words, longer challenge words also can slow but cannot stop modeling attacks [34]. It is clear that digital-based PUFs whether they are delay-based or memory-based, they are vulnerable to modeling attacks [28], [51]–[57]. Examples of weak PUFs are; the Flip-flop PUF [58], Buskeeper PUF [59], static random-access memory (SRAM) PUF [60], dynamic random-access memory (DRAM) latency PUF [61], butterfly PUF [62], latch PUF [63], artificial fingerprint device (AFD) [64], [65], coating PUF [66], etc. An extreme case of a weak PUF-like construction is a construction which only has a single challenge word. Such a construction is called physically obfuscated key (POK) [45], [24], [67]. In [68], the authors considered that a strong PUF can serve in challenge-response authentications; while, a weak PUF is used to generate a physically obfuscated key. This key is used for encryption; if the verifier side can decrypt the message with the archived key (public or private), then the PUF is authenticated [37]. It is not safe to send a response out of a weak PUF
device unless the response is encrypted or hashed, and the communication is done through an application program interface (API). That is to protect the PUF against man-in-the-middle attack (MITM), therefore such PUFs are known as controlled PUFs (CPUFs) [36], [69].

2.4. Strong PUFs

A Strong PUF is an ideal concept, and yet no known PUF has been considered as one. All the proposed PUFs are just candidates. The distinction between a strong and weak PUF was first introduced by Guajardo et al. [60] and discussed further by Rührmair et al. [70], then by Konigsmark et al. [71]. Strong PUFs are sometimes called physical random functions [72], [46] or physical one-way functions [38], [73]. A PUF is considered strong if even when tested by an attacker for a long time, it sustains its unpredictability for the non-observed CRP sets [44], [45]. The strength of PUFs only relate to the security qualities; that is, a strong PUF concept implies the following security aspects:

- **Physical Unclonability**

  Statistical and random disorders of manufacturing make it infeasible to produce identical PUFs even by the same manufacturing process. That is why such physical entities are considered physically unclonable.

- **Mathematical Unclonability (Unpredictability)**

  A strong PUF needs a strong unpredictability, so even mathematical modeling is supposed to be difficult; otherwise, it is useless to have a physically unclonable function that is easy to model mathematically. A mathematical unclonability of a PUF depends on:
1) Anti-Modeling Complexity

The interactions of challenge bits within a strong PUF should be complicated enough to make it infeasible for a computer program to predict the response with high probability (Pr), and with a reasonable time, even if many previous CRPs were observed [74]. The term high probability refers here to \(1 - \text{negligible}\).

Many modeling attacks have succeeded to model various digital-based PUFs; both memory and delay-based PUFs [28], [75], [51]–[57], while that is not the case for mixed-signal PUFs, even though they were proposed earlier by Lofstrom et al. [76]. This can be attributed to the difficulty of modeling mixed-signal PUFs.

2) Number of CRPs

It is important to highlight that having a large number of CRPs does not always help, as, it did not help protect the robust ring oscillator PUF from modeling attack [52]. However, a strong PUF should have a significant number of CRPs to make it infeasible to measure all CRPs within a reasonable time frame [77]. Some references recommended that a strong PUF should have 30-bit CRP sets (one billion possibilities) or above to be immune against modeling attacks [78], [79]. More recent literature suggested a higher limit of 64-bit CRP sets [80].

• Accessibility

Another condition which a strong PUF must fulfill is that even when an attacker can observe the input challenges and read out their corresponding response without restriction, it remains immune to modeling attacks (to predict the unobserved responses). In other words, a strong PUF does not need to restrict the access to its input nor its output with a secure
communication protocol. The non-restricted accessibility can enable an attacker to test several CRPs; however, having a large number of probable CRP sets makes it infeasible for an attacker to test them all within a reasonable time. Encrypting or hashing the responses is usually done on-chip before sending it to the verifier side; however, for strong PUFs, encrypting or hashing is recommended rather than essential [81].

- **Anti-Tamper Sensitivity**

  A strong PUF should have a self-distracting quality when an invasive attack is applied to its physical components, in other words, the PUF responses after an invasive attack should not be the same as before the attack. The authors in [54], [82] considered all strong PUF candidates are vulnerable to modeling attacks through machine learning algorithms; however, the authors did not revise the strong PUF nomination of those PUFs, nor established new criteria of strong PUF classification. Furthermore, [70] considered the weakness of a weak PUF does not refer to its unclonability or randomness properties but to the number of challenges it accepts. This dissertation adopts strict conditions for strong PUF properties, as if a PUF does not meet the four basic conditions discussed above, then it is considered a weak PUF. However, up to date still, no known PUF is error-free against the environmental variations.

**2.5. Evaluation Metrics of PUFs**

There are many metrics to evaluate PUFs [25], but in this dissertation, a PUF construction is useful for authentication purpose if it has a challenge-response behavior that meets at least the following three basic measures [83], [84]:
2.5.1. Inter-Distance

Hamming distance between responses \((Y_{1,i,\alpha}, \text{ and } Y_{2,i,\alpha})\) of two individual instances of a PUF construction against a challenge \(X_i\) under the same environmental conditions \((\alpha)\) is called the *inter-distance* between the two PUF individuals \((D_{1-2,i}^{\text{inter}} (X_i))\).

Noting that whenever the inter-distance is discussed, \(\alpha\) should be identical for both of the compared PUF instances; therefore, the notation \(\alpha\) although it usually appears within the measured response as \(Y_{1,i,\alpha}\), but it does not usually appear within the inter-distance notation as \((D_{1-2,i}^{\text{inter}})\), as both measurements of the two responses, \(Y_1\) and \(Y_2\) are supposed to be performed at the two PUF chips under \(\alpha\) set of environmental conditions. The inter-distance parameter characterizes the *uniqueness* of a PUF instance among the other tested PUFs within a Pclass.

The average inter-distance \((D_{\text{ave},i}^{\text{inter}})\) is usually referred to as the average Hamming distance, it is based on comparing the digital differences among responses of all PUF pairing combinations of all tested PUFs within a Pclass. Thereby, \(D_{\text{ave},i}^{\text{inter}}\) indicates the average uniqueness of the Pclass, which refers to the relative uniqueness of any PUF individual among the other individuals within that class. The sources of uniqueness are not attributed to the uniqueness of the response, as the response by itself is not necessarily unique, but the CRP is supposed to be unique within the tested Pclass [49]. It is common to test the inter-distance among a number of non-repeated pairs \((N_{\text{pair}})\) of individuals within the same Pclass. It is important to emphasize that whenever determining the average uniqueness of a PUF class, it is simpler to assume that any change in \(\alpha\) would affect the PUF individuals equally, thereby \((D_{\text{ave},i}^{\text{inter}})\) remains the same; however, this assumption might not be precise and implies a margin of tolerance. The average inter-distance \((D_{\text{ave},i}^{\text{inter}})\) formulation is detailed in Appendix A.
**Definition 2.3:** An evaluable (P_{class}) shows a high Pr of uniqueness if:

\[ Pr(D_{ave,i} \text{ is large}) \text{ is high [49], [45].} \]

This refers to a high Pr for that CRP within that P_{class} is unique for each PUF individual. The uniqueness of a P_{class} is usually tested at normal conditions \( \alpha \).

Based on \( D_{ave,i}^{\text{inter}} \), it is also common to determine the average inter-distance distribution \( \mu_{\text{inter}} \) of the P_{class}. Detailed formulations are in Appendix B.

Through simulation, it is also possible to determine \( \mu_{\text{inter}} \) using only one PUF chip by randomly varying some properties of its devices. This is commonly applied through Monte-Carlo randomness simulation. The optimum value for \( \mu_{\text{inter}} \) to achieve the most randomness of the responses within a P_{class} is 50\%. This means that among the tested PUF individuals within the class against a challenge; 1) on average, half of the responses’ bits are not correlated; 2) the \( \mu_{\text{inter}}=50\% \) represents an unbiased uniqueness and identifiability of the individuals within the class [85]–[87]; and 3) for a single individual instance of a PUF construction, the distance between two separately measured responses against the same challenge is zero. A PUF designer usually aims to approach such an optimal value.

### 2.5.2. Intra-Distance

For an individual PUF chip, the reproducibility of a response against a challenge can be measured by re-applying the same challenge, under the same environmental conditions (\( \alpha \)). Hamming distance \( D(Y_1; Y_2) \), in this case, represents the *intra-distance*.

The distance notation between the two noisy responses may refer to different environmental conditions (\( \alpha_1 \)) and (\( \alpha_2 \)), then the distance notation becomes \( D(Y_{1,\alpha_1}; Y_{2,\alpha_2}) \). A small rate of faulty bits can be ignored if the authentication is designed in a way that permits a degree of tolerance, or
an ECC unit is to be used to correct such bits. In general, regardless of the differences in the environmental conditions, the rate of the correct bits to the total response bits is referred to as *reproducibility*/ *steadiness*/ *stability*/ or *robustness* of a PUF. The average intra-distance distribution ($\mu_{\text{intra}}$) is determined by re-challenging the same PUF individual many times to measure multiple $D$ values at various environmental conditions regarding $Y_{1,\alpha_1}$ which is measured at standard conditions. The ideal value of $D$ is zero, so as for its average distribution, $\mu_{\text{intra}}$.

2.5.3. Identifiability

Identifying a PUF chip instance among its siblings is considered as uniqueness. It is important to state that neither the challenge sets, nor the response sets are required to be unique, but the CRP combination of each PUF unit should be unique to achieve identifiability (or uniqueness).

**Definition 2.4:** An evaluable $P_{\text{class}}$ shows *identifiability* if it is reproducible and unique and the probability:

$$\Pr (D_{i,\text{intra}} < D_{\text{ave,inter}})$$

is high [49], [45].

In other words, when applying the same challenge $X_i$, under the same environmental conditions $\alpha$ to various PUF individuals within a $P_{\text{class}}$, and the measured noisy response $Y_{i,\alpha}$ of that PUF individual (PUF$_1$) is distinct among the responses of other individuals of that $P_{\text{class}}$, then that PUF$_1$ is considered *identifiable* among its $P_{\text{class}}$ members. Such a response can be an identity and would usually have a high $\Pr$ of being unique within that $P_{\text{class}}$. This means, with a high $\Pr$, the CRP within a $P_{\text{class}}$ is identifiable for each PUF individual. Noting that identifiability of a $P_{\text{class}}$ is usually tested at normal working conditions, then $\alpha$ here refers to normal working conditions (voltage, temperature, etc.). The term ‘normal’ working conditions may carry different indications depending on the technology, here it refers to a supply voltage of 1 V and a temperature of 25° C.
to 27° C, and that is usually specified for each test performed. The assumption here is that the environmental effects on a PUF unit when it generates its response against a randomly chosen challenge \( X_i \) is the same against any other challenge, then \( D_{\text{ave},i,\alpha}^{\text{intra}} \) remains the same for any \( X_i \). The other assumption is considering any change in \( \alpha \) would equally affect all the individuals within a \( \text{Pclass} \), and therefore \( D_{\text{ave},i}^{\text{inter}} \) remains the same.

2.6. Reproducibility and Authentication Tolerance

For a single PUF construction such as a PUF chip, the uncertainty about the expected response to a particular challenge should be significant when an attacker does not have access to that particular PUF instance.

For a single individual PUF instance, the distance between two separately measured responses (repeatedly) against the same \( X_i \), under the same \( \alpha \), should ideally be zero or should be small. This can be denoted as: \( D_i^{\text{intra}} = D(Y_{1,\alpha}; Y'_{1,\alpha}) \)

**Definition 2.5:** An evaluable PUF class shows *reproducibility* if:

\[ \Pr(D_i^{\text{intra}} \text{ is small}) \text{ is high} \quad [49], \quad [45]. \]

A PUF usage relies on measuring the response, and there is always a chance of having added noise, and random errors. Depending on the PUF application and its required error tolerance; error-correcting code can even be omitted in some cases. In this case, only a predefined number of bit errors is accepted; i.e. if the allowable Hamming distance as a predetermined error threshold (\( \Delta \)) is not exceeded, then the PUF authentication is declared as ‘pass’. If \( ||Y_i - R_i|| \leq \Delta \), then the authentication is considered successful. This implies ignoring the errors if their value is less than the predetermined threshold (\( \Delta \)), without needing to use any helping data (\( W_i \)) to correct those
errors; where (i) is the CRP index.

The permitted tolerance threshold (Δ) is related to the following four parameters [50], [88], [89]:

1) **Bit Rejection Rate (BRR):** The ratio of the non-reliable bits to the total number of response bits. Such non-reliable bits are usually ignored.

2) **Bit Error Rate (BER):** The ratio of the false bits to the total number of response bits.

   **False Acceptance Rate (FAR):** The probability for a biometric security system to incorrectly accept an access attempt by an unauthorized user.

   **False Recognition Rate (FRR):** The probability that a biometric security system will incorrectly reject an access attempt by an authorized user. It is determined by dividing the number of false acceptances by the number of authentication attempts.

### 2.7. PUF Physical Unclonability

**Definition 2.6:** An evaluable PUF class P shows **physical unclonability** if it is hard to apply or to control the physical creation procedure of the class.

This definition is also applicable to the mathematical/logical unclonability if it is hard to find a precise mathematical model that can emulate the physical functionality of a PUF structure.

Hypothetically, if the first (original) process created an individual (PUF1) and the clone individual is (PUFclone); then the clone can be checked by an inter-distance rule. Considering the physical clone is (PUF2clone), and as discussed within the mathematical unclonability in Section 2.6, the physical clonability can be tested in the form of inter-Hamming distance $D_{\text{inter}}^{i-\text{clone},i}$ between the responses of the $Y_{1,i,\alpha}$, $Y_{\text{clone},i,\alpha}$ of both the original and the cloned PUFs. In this test, responses are measured against the same challenge (X) indexed by (i), under the same environmental conditions $\alpha$. Detailed formulations are in Appendix C.
2.8. PUF Anti-Tamper Sensitivity (Tamper Evidence)

It is important to emphasize that for a physical cloning approach; the attacker is considered having the PUF class blueprint or access to the internal devices to get reverse engineer the PUF structure to create a cloned PUF (PUFclone). However, practically; there will always be many implicit random variations along with any manufacturing process, which nobody can totally really control. However, detecting the cloned PUF depends on the precision of response measuring and the permissible tolerance (Δ), where (Δ) influences the values of FAR and FRR of the authentication process. Tamper-resistant physical security is usually easier when a smart card is not in action; otherwise, complexity rises [88]. However, there is always a chance that an attacker may find a way to disable such prevention approaches; therefore, a strong PUF should have a second defense front in the form of ‘Tamper Evidence’ quality feature, i.e. changing (PUF1) hardware parameters yields a change to its output response, in fact, in this case, the PUF even be considered as another PUF, it can be called (PUF2) and it can no more generate the expected response (Y1), then it can never be authenticated. That is a very important property to ‘self-destroy’ a PUF as a defensive approach against invasive attacks. A tamper-sensitive PUF is usually called ‘fragile PUF’ [90]. On the other hand, using a less tamper sensitivity PUF or adopting a high FAR by the verifier, which means more tolerance. Such PUF might still be able to generate some response (Y2) which still has an acceptable degree of similarity to (Y1) and with the help of (W1), the authentication would still be possible, then the verifier can be deceived.

A convenient approach to study the tamper effect on a PUF is to compare the output response before and after being tampered. This is usually done by determining $D_{intra}^{1-2}(x,)$ between the original noisy response (Y1) of the untampered (PUF1), and the response after tampering (PUF2) is (Y2), and the reason behind applying the intra-distance checkup rather than the inter-distance...
one is simply because it is assumed that the verifier side does not know about the tamper attempt and would still consider the PUF as if it the same one. By extending the tamper concept discussed in [49], then it is possible to test the anti-tamper sensitivity, as:

\[ D_{1-2,i,\alpha}^{\text{intra}} = || Y_{1,i,\alpha} - Y_{2,i,\alpha} || \]

Then for tamper sensitivity, there will be three possibilities:

1) A very good tamper evidence indication is when \((Y_2)\) is very much different from \((Y_1)\), in other words, both are less correlated to each other, which leads to a very high intra-distance \((D_{1-2,i,\alpha}^{\text{intra}})\) among their responses. In other words;

If: \( D_{\text{ave}}^{\text{inter}} < D_{1-2,i,\alpha}^{\text{intra}} \)

That means (PUF\(_2\)) is distinct, not only from (PUF\(_1\)) but also from the entire \((\text{Pclass})\). As larger \((D_{1-2,i,\alpha}^{\text{intra}})\) is, as more anti-tamper the PUF class \((\text{Pclass})\) would be.

2) A good tamper-wise security indication is when \((Y_2)\) does not differ much from \((Y_1)\), in other words, still there is some correlation between \((Y_2)\) and \((Y_1)\); which leads to a medium \((D_{1-2,i,\alpha}^{\text{intra}})\), in other words;

If: \( D_{\text{ave}}^{\text{inter}} \leq D_{1-2,i,\alpha}^{\text{intra}} \leq D_{\text{inter},\alpha}, \exists 1 \leq i \leq N_{\text{challenge}} \), then: \( \text{PUF}_2 \neq \text{PUF}_1 \)

However, still \( \text{PUF}_2 \in \text{Pclass} \)

Then; the probability \( \Pr(D_{1-2,i,\alpha}^{\text{intra}}) = 0 \) for at least one challenge \((X_i)\) is high.

3) A bad tamper-wise security indication is when the response \((Y_2)\) is very near to \((Y_1)\) in term of intra-distance \((D_{1-2,i,\alpha}^{\text{intra}})\) such that;

If: \( D_{1-2,i,\alpha}^{\text{intra}} = 0, \forall 1 \leq i \leq N_{\text{cha}}, \) then \((\text{PUF}_2(X_i)) \equiv (\text{PUF}_1(X_i))\).
Alternatively, in less restricted mode:

If: \( (D_{1-2,i,\alpha}^{\text{intra}}(X_i) = 0, \exists 1 \leq i \leq N_{\text{challenge}} \), then: \( (PUF_2(X_i)) \cong (PUF_1(X_i)) \).

Where \( (N_{\text{challenge}}) \) represents the number of the tested challenges.

In both last cases (2 and 3), it is concluded that tested PUF design (the whole \( P_{\text{class}} \)) is not sensitive enough against tampering, in other words, it does not have a tamper evidence quality [45].

2.9. PUF Unpredictability (Mathematical Unclonability)

Beside physical unclonability, a secure PUF must have an unpredictability quality, i.e., the unobserved responses are supposed to be sufficiently random [49]. Mathematical cloning tends to be an easier approach of PUF cloning, where an attacker does not create a physical clone of the PUF, but an algorithmic model. For this purpose, an attacker would study the PUF behavior depending on CRPs analyses applying black-box testing approaches to design a mathematical clone model. Obviously, as many observations the model is based on, the higher probability of a successful imitation for a PUF behavior can be, and thereby an attacker’s successful model can expect the upcoming response for the new challenges. In general, for both mathematical and physical clone, it is possible to assess the predictability of a clone being able to generate the correct responses, by using the inter-distance in Hamming form \( D_{1-\text{clone}}^{\text{inter}}(X_i) \), considering the attacker has two PUF individuals, the genuine (\( PUF_1 \)), and the clone \( PUF_{\text{clone}} \); and their respective responses \( (Y_{1,i,\alpha}, Y_{\text{clone},i,\alpha}) \) are measured against the same challenge (X) indexed by (i), under the same environmental conditions \( \alpha \). The cloning success can be checked by an inter-distance rule as follows:

\[
D_{1-\text{clone}}^{\text{inter}}(X_i) \triangleq \text{dist} [PUF_{1,\alpha}(X_i); PUF_{\text{clone},\alpha}(X_i)]
\]

or \( D_{1-\text{clone}}^{\text{inter}}(X_i) \triangleq \text{dist} [PUF_{1,i,\alpha}; PUF_{\text{clone},i,\alpha}] \)
or \( D_{\text{inter}, \alpha}^{\text{1-clone}, i} \triangleq || \text{PUF}_{1, \alpha}(X_i) - \text{PUF}_{1, \alpha}(X_i) || \)

, then \( D_{\text{inter}, \alpha}^{\text{1-clone}, i} \triangleq || \text{PUF}_{1, \alpha} - \text{PUF}_{\text{clone}, i, \alpha} || \)

or \( D_{\text{inter}, \alpha}^{\text{1-clone}, i} \triangleq || Y_{1,i, \alpha} - Y_{\text{clone}, i, \alpha} || \)

Then the cloning is considered successful if:

\( D_{\text{inter}, \alpha}^{\text{1-clone}, i} = 0 \), for \( \forall 1 \leq i \leq N_{\text{challenge}} \), then (PUFclone) is an identical clone of (PUF1).

Alternatively, for less restriction case, if:

\( \Pr (D_{\text{inter}, \alpha}^{\text{1-clone}, i} = 0, \text{ for more than one instance (i)}, \text{ then it is highly probable that (PUFclone) is an identical clone of (PUF1).} \)

2.9.1. Randomness (Entropy)

To maintain unpredictability, by achieving high randomness, a CRP should fulfill the following conditions:

1) **Randomness of Challenges**: The verifier is supposed to adopt an adaptive random challenge selection approach, to achieve unpredictability of the next challenge.

2) **Randomness of Responses**: Responses to different challenges are independent. That not only implies the random correlation between independent responses, but it also implies there is no historicity in the PUF, i.e. no previous challenge or response can affect the next response.

3) **Large Number of Unique CRPs**: The PUF is supposed to have very large CRP sets. Otherwise, the attacker can try all possible challenges, then records their responses, then no secret response key is left undisclosed. The number of CRPs is usually, but not always, exponential to the number of the components used to build a PUF [44].
The randomness of data is usually assessed by the entropy concepts. Entropy is a measure of the uncertainty one has about the outcome of a random variable distributed according to a given distribution; it also represents the unpredictability of a secret random variable [45]. There is Shannon’s entropy [91], which was formulated further by Cover et al. [92], and a less minimum entropy margin proposed by Rényi [93], which Dois discussed further [94], [95].

2.9.2. Correlation Among Bits

Neighboring cells must not influence each other to avoid security threats. Correlation between cells would reduce the number of possible combinations and thus would increase the risk of successful brute-force attacks. To check whether a correlation exists between neighboring response bits \((R_n, R_{n-j})\), with a lag of \((j)\), and an autocorrelation function \((AC_{XX})\), then: \(AC_{XX}(j) = \sum_{n=1}^{N} R_n R_{n-j}\), where \((N)\) is the total number of response bits.

If \(AC_{XX}(j) = 1\), or \((-1)\), then the data at lag \((j)\) are completely correlated; while, if \(AC_{XX}(j) = 0\), then the data are completely uncorrelated [50].

2.9.3. Correlation Among Chips

When a layout is not done carefully, it may happen that the output of the cells depends on the position in the layout. Then, in this case, some specific cells tend to a certain value. This can lead to security threats since knowing the output of one chip makes it easier to guess the output of other chips within a PUF class. For all the previously known PUFs, response bits are generated individually, the correlation among chips in the form of inter-Hamming Distance \((D_{1}^{\text{inter}})\) can be determined similarly. If a response bits-string of a PUF chip individual is considered as a reference, then the mean value of all Hamming distances to the responses of the other tested PUF chips against the same challenge \((X_i)\) and under the same environmental conditions \((\alpha)\) can be
represented as \((D_{ave,i}^{inter})\). Its optimum target is 50\%, which means that there is no correlation among the tested PUF chips within the class [50]. However, some tolerance is usually accepted, especially when a PUF generates a large number of response bits, such as 64 or 128, then having \((D_{ave,i}^{inter})\) of \((50 \pm 2)\%\) for example, would not help the attacker much to model the PUF; after all, adopting a designated tolerance rate depends on the application nature.

2.9.4. One-Wayness

In computer science, a one-way function is defined as a function that is easy to compute on every input, but hard to invert to determine each input [38], [49]. The terms "easy" and "hard" here are supposed to be understood in the sense of computational complexity, especially as polynomial-time problems. A function is usually considered a one-way function if for a given input challenge \((X_{1,i,\alpha})\), indexed by \((i)\), under environmental conditions \((\alpha)\), it is easy to measure the noisy output response \((Y_{1,1,\alpha})\), but for a given response \((Y_{1,2,\alpha})\), it is hard to predict the input challenge \((X_{predict,2,\alpha})\) as an inversion of the response:

\[
X_{predict,2,\alpha} = \text{inv}(Y_{1,2,\alpha})
\]

In other words, the inversion has a negligible probability if:

\[
\Pr(\|X_{1,2,\alpha} - X_{predict,2,\alpha}\| = 0), \text{ for more than one } (i)\text{ is low, then the PUF function can be considered as a one-way function until proving the opposite by figuring out an inversion model.}
\]

2.10. Smart Card Basics

It is important to review the vulnerabilities of the financial smart cards, to know the role that PUFs can play. When a financial card is inserted into a point of sale’s terminal, the terminal (represented by the card reader) reads the cardholder’s information from the card. Information includes card details (e.g. primary account number, start, and expiry date), backward compatibility
data to support older technologies (e.g. magnetic strip), and control parameters for the verification protocol.

In Europe, Canada, and lately in the United States of America (USA), the majority of the contemporary financial smart cards were first originated by Europay, MasterCard, and Visa incorporated which jointly developed the ‘EMV’ standard for financial smart cards, payment terminals, and automated teller machines (ATM) [96]. This EVM standard is managed by an EMV consortium (EVMCo) association [97] and its recent version is EMV 4.3 [98]–[102]. The EVM card is known as the chip card, chip and personal identification number (PIN) card, or chip and signature card.

This card usually has contact pins (for card insertion mode), RFID (for contactless mode), and a magnetic stripe (for swiping mode). The EVM standard complies with ISO/IEC 7816 for contact mode, and ISO/IEC 14443 for the contactless mode.

MasterCard has chip authentication program (CAP) for secure e-commerce, which is known as EMV-CAP and supports some verification modes, while Visa has dynamic passcode authentication (DPAuth) scheme, which is their equivalent version of the CAP.

A terminal usually runs a risk management algorithm to decide whether a transaction should be authorized online or offline. The algorithm uses a combination of data known as terminal action codes (TACs) held in the terminal and issuer action codes (IACs) which are read from the card. According to the terminal verification results (TVR), the terminal decides whether to accept, decline or to request an online verification for a payment transaction. An online-only device such as an ATM always attempts to go online with the authorization request, unless an IAC declines that. After the terminal runs its risk management algorithm and decide which type of verification should be adopted, it requests one of the following cryptograms from the card:
- transaction certificate (TC)- offline approval
- authorization request cryptogram (ARQC)- online authorization
- application authentication cryptogram (AAC)- offline decline.

The card can accept the terminal's request or decline it or force the transaction to be online. In the case of online authorization, the card generates ARQC, which is a digital signature of the transaction details. The verifier side (i.e. the card issuer) verifies this signature to check whether the card is genuine. Then, the verifier sends an authorization response cryptogram (ARPC) to accept or decline the transaction.

According to the risk management algorithm’s request, a PIN is required in some cases, and it can be one or both of the following types:

1) Offline PIN verification, which is based on comparing the inserted PIN with the pre-stored one on an NVM within the card, and the comparison is usually done within the card; then the result is sent to the card reader (i.e. the terminal).

2) Online PIN verification, that does not require pre-stored PIN on the card and is verified by the issuer of the card. Then the issuer sends the verification result to the card reader to authorize or deny the transaction.

A card may support either an offline, online or both authentications. If it supports both, the user PIN is usually (but not necessary) the same. Although chip cards have the ability to store a PIN, it is up to the bank or the national regulations whether to allow an offline PIN or not. The offline transaction facilitates the sale for points of sale having an unreliable internet connection to its card reading terminal. Later, the terminal can communicate with the card issuer to accomplish the financial transactions. At that time, smart card fraud in the USA was not as rife as in Europe.
Furthermore, the communication infrastructures were well established in most areas of the USA, so there were no many problems in keeping the magnetic stripe cards which accomplish the transaction online only. Therefore, there were no enough justifications to spend billions of dollars to upgrade the banking systems and the sale points’ terminals. That may explain why the American banks, unlike the European, did not encourage migration to EMV cards at the beginning of their release.

The details of any online transaction are usually encrypted with a cryptographic message authentication code (MAC) by the ATM or any terminal of sale using a symmetric key which is shared between the card and the verifier side (i.e. the bank which issued the payment card). EMV cards use Rivest, Shamir and Adleman algorithm (RSA) encryption.

In a type of EMV, known as the static data authentication (SDA), the card itself is not capable of performing RSA operations, so it only provides the terminal with a static certificate. Since 2009, some banks have started issuing cards featuring dynamic data authentication (DDA) which contains RSA private keys to sign a nonce sent by the terminal. DDA cards can sign a terminal-provided nonce (any message) with an asymmetric key when a transaction is done offline [103], [104].

From all the above, it is concluded that all verification approaches depend on the ability of the smart card to use an encryption key that is stored in its NVM. The invasive attacks can use sophisticated equipment such as scanning electron microscopes (SEM) and focused ion beams (FIB) techniques to find the secret keys to either break the hardware security or to reverse-engineer the devices to gain access to the data or the funds they protect. Also, what if the attacker observed the PIN code and cloned the smart card? Then unless the card user checks his list of purchases before paying the monthly debt, no one can know about the cloned card that the attacker keeps
utilizing. In some PUF-based smart cards, a PUF generates a response that helps to authenticate the smart card. Furthermore, in some PUF systems, either the response itself or depending on it, a new encryption key is generated and utilized for each session, then the key is removed from all internal registers of the smart card. Most of these PUFs generate their keys depending on the random challenge the PUF receives from the authenticator side.

2.11. PUF Usage Phases

It is essential to understand the dependency as well as the difference between the authentication (or verification) and the identification of an object.

In the identification case, the system must recognize the object. For example, in a smart card, by reading a serial number code stored on a non-volatile memory (NVM), the system then matches against every known ID code template.

While in authentication case, an authentication process is based on comparing the received object’s biometric pattern with the object’s template. That template was archived during the training phase of the system (i.e. the enrollment phase) as a reference to be used later (i.e. at the verification phases) to check the similarity degree among the two phases' responses.

For each of the identification and authentication cases, the system usually adopts a certain FAR and FRR to decide whether the similarity is sufficient or not. The usage of a PUF consists of two phases [105]:

2.11.1. Enrollment Phase

The enrollment phase, also known as learning or training phase, should be performed before releasing the PUF to the end-user, and it can be briefed as:

1) Assigning a known ID to identify a chip, this ID code or number is stored on a server. Linked lists against each PUF ID are archived within the database (DB). archived on the
Each chip ID will be used later within the verification phase so each chip can locate its related linked lists within the DB.

2) The PUF entity is challenged with a challenge $X_i$, to generate output response data $R_i$, and an error-correcting or helping data ($W_i$). Usually, $W_i$ is stored on-chip in an NVM, to be used later during the authentication phase to correct the noisy response $Y_i$ to generate the corrected response ($S_i$). Storing $W_i$ on-chip poses a security threat, a safer approach is to store $W_i$ on a secure server, either of the verifier side or of an independent third party, especially in the case of untrusted verifier side. [50]. The verifier can correct the received noisy response $Y_i$ using $W_i$ to determine the related corrected response ($S_i$).

3) $R_i$ is usually archived off-chip.

4) $W_i$ list can either be stored internally on-chip or externally within a safe DB.

5) The verifier stores an initial, small, or large set of CRPs securely in his DB. Along with $W_i$, getting $(X_i, R_i, W_i)$ template against each CRP index (i) for each PUF’s ID.

The archived number of CRPs would depend on the application. For example, for a PUF that is embedded into a credit/debit card, the verifier should test and save many CRPs. Otherwise, if that PUF is embedded in RFID tag for an on-shelf item in a supermarket, then such verifier DB would only need one or few CRPs if considering the possibility of returning the purchased item, then resell it for multiple times.

Due to noise during the enrolment phase, the reference response ($R_i$) is measured against a challenge ($X_i$) for several times. Later, those independent challenges ($X_1, \ldots, X_{N_{cha}}$), and their corresponding reference responses ($R_1, \ldots, R_{N_{cha}}$) are saved in the verifier DB, where ($N_{cha}$) is the maximum number of possible CRP sets. If the number of the challenge bits (n=21), then the number of possible CRP sets is:
For each \( X_i \), it is expected that the measured response \( Y_i \) during the verification phase would be slightly different from \( R_i \), which was measured during the enrolment phase. That relation can be represented as:

\[ Y = R + E \]

Where \((E)\) represents the error, and it is assumed to be zero during the enrolment phase. In this dissertation, \( X, R, Y, \) and \( E \) are digital numbers of \( n \) bits.

To handle the difference problem between \( Y \) and \( R \), the authentication system should include a signal processing function \( S_i = G(W_i, Y_i) \) to enhance the reliability and reproductively of the response detection during the verification phase. Where \((W_i)\) is any form of helping data, which assists error-correction. When \( S_i = R_i \), then the authentication is successful.

To sum up, the verifier DB must have data indexed with product ID (i.e., serial number) which is linked to the three following lists:

- Challenge \( X: X_1, \ldots, X_{N_{cha}} \), here, \((X_i)\) denotes the challenge instead of \((C_i)\).
- Response \( R: R_1, \ldots, R_{N_{cha}} \)
- Helper data \( W: W_1, \ldots, W_{N_{cha}} \)

### 2.11.2. Authentication Phase

During the authenticating phase, \( S_i \) can be compared to \( R_i \), or if no on-chip helper data is used, then \( Y_i \) can be compared to \( R_i \) instead.

There are many suggested authentication protocols; however, this dissertation presents the PUF core, and it can be attached to any extra circuits if needed. The simplest authentication protocol, is when the verifier side is trusted (i.e. immune with security against external attacks,
and with honesty against internal attacks); however, for the communication security, communication protocols still have to use hash, or encryption functions to communicate the information between the PUF (at user-side) and the verifier-side [106]. Furthermore, at the user side, to operate the PUF on RFID tag, it implicitly considers the existence of other complementary hardware parts; such as antenna, modulation/demodulation circuits, similar to any non-authenticatable RFID tag. The serial number from a magnetic strip is considered as equivalent to the (ID user name) in internet terminologies, while the PIN code is equivalent to the password.

In most communication protocols when the client does not tap the smart card (does not use the RFID feature); instead, the clients enter the PIN code, that adds more security as that proves the client knows the correct PIN which is stored on-chip in an NVM within the smart card. When the entered PIN matches the stored one, the card reader sets a local secure communication with the smart card. After that, the card reader uses a public or symmetric key, which is usually stored in on-chip NVM, to encrypt/decrypt the communication with the verifier side. The basic authentication protocol in Fig. 2.4 can be briefed as follows:

**User:** The authentication phase starts when a package that contains the PUF device is scanned by an RFID scanner at a point of sale, for example. Similarly, for financial cards, the reader can be an RFID scanner or a smart card reader. The ID is usually embedded within an RFID tag/smart card in various forms [50]:

- If the ID is stored outside the chip, then it is to be read from a magnetic strip or barcode.
- The ID is stored on-chip in an NVM, which is the case of most RFID tags and smart card chips.

The ID can be generated using another on-chip PUF. The claimed ID is sent to the verifier side; it would mostly be encrypted.
Verifier: Using the received ID leads the VS to the related CRP list, which was stored during the enrollment phase on the safe server, to randomly choose a challenge (X_i) and send it to the PUF individual. Where (i) indexes (X_i, R_i, W_i) within the archived CRP list.

In most PUF authentication protocols, some helper data (i.e., error-correcting) data (W_i) is sent to the user side, along with the challenge, while for others, the helper data is kept within the verifier DB exclusively. The second approach is more secure, and it also reduces the chip cost regarding area, power, and complexity.

User: The user reader challenges the PUF embedded within the submitted smart card or the RFID tag, with challenges X_i, measures (possibly noisy) response Y_i, then sends it to the verifier.

Verifier: Due to the environmental driven errors during the verification phase, the noisy measured response Y_i usually differs from the referenced response R_i which has been measured.

Fig. 2.4. PUF authentication with trusted verifier
during the enrolment phase. Therefore, even the simplest authentication protocol, needs to include the following basic steps:

- To save the computational time of non-correctable responses, usually the verifier first determines whether $|| Y_i - R_i || \leq \Delta$, where $\Delta$ is a predetermined threshold, which sets the FAR and FRR limits. Where $||Y_i - R_i||$ is the difference among bits of the response during verification (i.e., authentication) phase $Y_i$ to the response measured during enrolment phase $R_i$. This difference is called Hamming Distance (HD, or D for short), and since the test in the above case involves all response bits of both phases, so it is just like comparing the entire response of two different PUF instances, then such a distance is considered as inter-Hamming Distance ($D_{i,a}^{\text{inter}}$) [50].

- When the difference $|| Y_i - R_i ||$ exceeds $\Delta$, then $Y_i$ is far from being correlated with $R_i$. If this difference (or error) is uncorrectable nor ignorable, then the verifier (usually software) denies the authentication. While, if the difference is less than $\Delta$, then the verifier approves the authentication claim.

- When $|| Y_i - R_i || \leq \Delta$, the verifier calculates the corrected response $S_i$ using the helper data $W_i$ of the measured $Y_i$, as $S_i = G(W_i, Y_i)$, then the verifier checks whether $|| S_i - R_i || = 0$; i.e. $(S_i = R_i)$; if not, then the authentication is denied, else it is successful.

To maintain a high-security level, the used $(X_i, R_i, W_i)$ templet can be deleted from the DB, so it may not be used again.

1) After a successful authentication phase, an encrypted access approval code number is sent to the user, sometimes by using $Y_i$ as a session encryption key. On the user side, the smart card, for example, uses its PUF session response $Y_i$ to decrypt that permission code. The PUF design should make sure that the generated response $Y_i$, which is sent to the verifier side, is
stable during the authentication phase.

2) If the transaction requires exchanging secret messages between the smart card and the verifier (e.g., for a financial transaction), a secure authenticated channel can be set between the verifier and the card using the response $Y_i$ as an encryption key for the transaction session.

### 2.12. Untrusted Verifier

There are usually two sides of the transaction; an end-user, which is to be verified, and the verifier side, a bank for example. There are various verification protocols to protect the verifier side from outsiders’ frauds and MIT attacks [107], while for the user side, there is only chip-and-PIN protection, and is proven to be broken [103]. Also, still there is less sufficient protection for the user-side (i.e., the client) against insider frauds within the bank, for example. Therefore, in some systems, to avoid compromising the DB at the verifier side, the verifier might not have all the required DB on its server; instead, some parts of the authentication DB are kept at another independent server. This 3rd party server has only the complementary piece of the verification process and can provide it only upon a request from one/ or both verification sides.

For example, it can have $(i, W_i, R_i)$ lists; while, the verifier server only has the $(ID, i, X_i)$ lists, so the verifier for each verification time, when it receives the claimed ID from the user-side (can be a smart card or RFID tag), then the verifier chooses an arbitrary one-time verification index $(i)$, sends its related challenge $(X_i)$ to the user-side, at the same time, the verifier sends the chosen $(i)$ to the 3rd party server asking for the related helper data $W_i$. When the verifier gets back the non-corrected noisy response $Y_i$ from the user-side, at the same time, it gets from the 3rd party server the related $W_i$, and the verifier can correct $Y_i$ to guess $R_i$ and gets the approval from the 3rd party. The verifier then can authorize the transaction with the end-user.
In this case, the user and the verifier both are protected and responsible at the same time; exactly as having a safe box in a bank, the box cannot be open without inserting the client’s and the bank’s keys, so the access cannot be disclaimed by any party. However, discussing various authentication protocols is outside the scope of this dissertation.

2.13. Sources of Variations

All known PUFs are based on components’ mismatch, the use of the term ‘mismatch’ here is interchangeable with term ‘variation’. A PUF identity relies on the variations during the fabrication process, and since the research here is about silicon electronic PUFs; therefore, variation and error sources of electronic components will be considered in this study. The terms ‘mismatch’ or ‘variation’ here refers to the inequity of responses among multiple PUF individuals within the PUF class, which is the essential quality for chip identification purposes, and these types of mismatch sources are permanent. During the PUF authentication process, if the environmental conditions differ from those at the enrollment phase, the produced response may differ as well. The difference between these responses is usually considered an error, and the collective effect of all the environmental variations on a PUF is considered a temporary source of error. Intrinsic variation in MOSFET properties wafer-to-wafer and across the wafer (die-to-die) influence connections, like the variations in metal thickness due to damascene process (e.g., copper polishing), is around (10-20)% and the variation in the threshold voltage ($V_T$) is around 10% [108], while the variation in dielectric thickness at wafer level is around 5% [108], [109]. There is also an example of an arbiter PUF which is based on the variations in propagation delay has reported an inter-chip (die-to-die) response variation of (17-23)% for the regular arbiter, and (28-38)% for the feedforward arbiter PUF [110], such variation depends on the die and the wafer locations. There are two main sources of mismatches; permanent and temporary [50], [111], [112].
2.14. Sources of Permanent Variations

The permanent mismatches are attributed to variations during fabrication, and from manufacturing view side, there are two main types of manufacturing variations: global/systematic/extrinsic variation sources, and local/stochastic/ intrinsic variation sources.

2.14.1. Global Mismatch Sources

Global mismatches come from inaccuracies of the production process. These include lot-to-lot, wafer-to-wafer, and chip-to-chip variability depending on the source of variation. In [40] it is stated that about 20% of the variabilities come from production inaccuracy. The global/extrinsic process variability is caused for example by temperature gradients across the wafer during annealing, by photoresist development, etching process, or photolithographic process variations. These variations lead to global effects on the produced device characteristics. For example, gradients of threshold voltage values can occur over the whole wafer toward a certain direction or device properties may change depending on their placing/direction.

2.14.2. Local Mismatch Sources

Local mismatches are attributed to random (or stochastic) atomic level variations during the fabrication process of MOSFET devices by the inevitable variations of charge, material, non-uniformity of interfaces, and granularity of fabrication materials [113]. This kind of variation to build up the fingerprint (uniqueness) of a device among the other devices within the class. Local variations in nano MOSFET processes are classified as Systematic local process variations, and Statistical local process variations [112]. Local mismatch sources can be categorized into:
A. Local Systematic Process Variations

The main source of systematic local process variations is attributed to the low resolution of the lithography process. The imprecision issues of lithography are treated with Resolution Enhancement Techniques (RETs), such as phase shift masking (PSM) by adding phase information to the mask, off-axis illumination (OAI) to optimize the angles of light illuminating the mask, source polarization (to control of the polarization of the illumination), source mask optimization (SMO), and optical proximity correction (OPC) to optimize the mask pattern shape [114]–[117]. To assess the importance of (RETs), it would be surprising to know that by using immersion lithography and (OPC) techniques, the fabrication foundries fabricated the 45-nm nodes using 193-nm steppers and that mainly through employing the effects of pattern direction and mask shadow [112]. Local systematic variances are caused by lithography issues such as light phase shift, layout-mediated strain, and well proximity. Although new lithographic developments have narrowed the local systematic variances of lithography for some extent, RETs still preserve their importance. For instance, even when using Extreme Ultraviolet Lithography (EUL) with a laser beam wavelength of (13.4 nm), still (OPC) is needed to enhance pattern pitch and direction.

The introduction of strain silicon to move silicon atoms farther apart reduces the atomic forces that interfere with the movement of electrons through the transistors and thus improving the mobility, which enhances the chip performance and lowers the power consumption. Therefore, electrons can move faster allowing strained silicon transistors to switch faster [118]. Straining silicon technology has been started with 90-nm technology, which led toward more variation; as besides the lithography-related geometry variations, additional variations were added due to spacing, distances to the shallow trench isolation, and positions of contacts.
Using regularized design and dummy features can greatly reduce the impact of systematic variabilities [112].

**B. Local Statistical (Random) Process Variations**

Local mismatches come from stochastic atomic level differences. Up to date, it is not possible to fully control the process on an atomic level. Moreover, the smaller the minimal feature size, the higher the influence of each atom. Fluctuations in physical gate oxide thickness is an important source of local mismatch, as when the transistor is scaled-down and has a gate oxide thickness is of several atomic layers with a typical semiconductor-to-insulator interface roughness of about 1–2 atomic layers only; this thickness can vary up to 50% within the transistor [119]–[121]. This entails variation of carriers’ mobility, which varies the gate current. Sub-100 nm-transistors have different mismatch sources which vary with the type of the transistor. Some examples are:

- In polysilicon gated transistors, the mismatch is usually attributed to the variation in doping concentration in the gate region; unlike metal gate transistors with a high-k dielectric material.
- In fully depleted silicon on isolator (SOI) transistors, the threshold voltage varies with the body thickness [122]; therefore, when comparing with the MOSFET, a MOSFET with a metal-gate and high-k insulator would have less threshold voltage fluctuation [123].
- In planar bulk MOSFET, there are three major local mismatch sources [50]:
  
  1) **Random Discrete Doping (RDD) (or Random Dopant Fluctuation (RDF))**  
     
     It is originated during ion implanting and redistributing during high annealing temperature. (RDD) is the main source of mismatch (variability), mainly at channel
region; however, the 22-nm node technology and beyond, have thinner bodies and lightly doped or intrinsic channels, which reduces RDD effect dramatically [119], [124].

2) **Line Edge Roughness (LER), and Line Width Roughness (LWR)**

The LER is referred to the fluctuation of edges along patterned lines, such as gate lines, fin, or nanowire channel lines. In other words, LER describes the gate length variation along the width. The LWR is referred to as the fluctuations in both line edges causes variation in the width of lines. Both LER and LWR are caused by tolerances in materials and tools used in lithography processes, mainly due to photoresist (jaggedness, striations, and rippling), and then during etching step [119], [122], [125], [126]. The granularity of the photoresist, with other factors, introduce unavoidable LER in the gate [112].

3) **Poly-Gate Granularity (PGG)**

It refers to the effect of poly-silicon grain boundary distribution on the threshold voltage. When a transistor has a polysilicon gate, then PGG is an important source of variation (or mismatch). This variability is attributed to the doping non-uniformity (fluctuation) when having a rapid diffusion. The granularity becomes a significant source of variability when the grain size becomes comparable to the transistor feature size (L/W) [122], [127], [128].

In bulk MOSFETs, the doping concentration in the polysilicon gate region play an important role, where the surface Fermi-level pinning at the Poly-Si grain boundaries dominates the poly-Si.; while that is not the case in a metal gate with high-k dielectric insulation material. It is concluded that for a PUF device to have a larger PGG effect as a source of variance, using a poly
gate is recommended over metal gate.

In general, for a planner transistor, RDD usually dominates the mismatching behavior. However, RDD and PGG induced competing variabilities at 35 and 25-nm channel lengths. For shorter channel lengths, if the LER could be scaled according to the International Technology Roadmap for Semiconductors (ITRS) as 1.2, 1, 0.75, and 0.5 nm for the 35, 25, 18, and 13 nm channel length transistors requirements, then RDD becomes the dominant source of variability of the threshold voltage \(V_{TH}\). However, if LER remains at \(LER\approx 4\) nm, then LER becomes the dominant source of induced potential variabilities, which pins the channel due to the very high concentration of mobile carriers (especially the electrons in n-MOSFETs) below the 25-nm channel length [129], [113]. In addition to the variability in the transistor characteristics, the discreteness of dopants in combination interface roughness introduces significant statistical variability in the gate leakage current [112]. With the further reduction of the gate oxide thickness and doping concentration at the interface as low as possible, the impact of the surface potential pinning at the Poly-Si gate grain boundaries will increase. The use of Poly-Si micro-grains can reduce the fluctuations, as variability will be reduced if the characteristic size of the grain is much smaller than the device dimensions; thus, the associated potential fluctuations will be self-averaged in the device. However, the best approach would be using amorphous Poly-Si or a uniformly structured metal gate [129]. Using a metal gate eliminates the PGG variability; but, it introduces both high-K granularity (morphology) and work-function variations instead [130], [131].

Materials, technologies, and factors contribute to variation; therefore, still, no absolute answer is found to the question: Which transistor type has the most statistical local variations?

In general, to obtain a PUF class where the PUF individuals have high identifiability (uniqueness), it is recommended that the designers and the manufacturers act exactly in the
opposite direction against any variation limiting approach, and with scaling down the transistor. Furthermore, the manufacturers can apply the PUF-aware OPC to improve the uniqueness of the PUF individuals among the class [132], [133]. The 45-nm based PUFs would have a better uniqueness than 65 nm based PUFs, and the 28-nm PUFs would be far better than 45 nm scale PUFs [134]. Furthermore, as the MOSFET device scale goes down, as more the intrinsic variation of a device’ parameters would be [134].

2.15. Sources of Temporary Variations

Temporary mismatches are also classified as reversible mismatch sources and irreversible mismatch sources.

2.15.1. Reversible Mismatch Sources

These variations happen during the authentication phase, when the environmental conditions vary, such as the supplied voltage ($V_{DD}$) and the ambient temperature are not as standard as in the laboratory [24].

2.15.2. Irreversible Mismatch Sources

The properties of an electronic device vary during the device lifetime, and such variations are usually referred to as transient variations. Such variations are not temporary, nor exist at the manufacturing time, and are not even constant throughout the lifetime of the object. For instance, an electronic device ages during all of its lifetimes, and there will always be continuous gradual changes in the device properties.

Experiments done on ring oscillator PUF based on 90 nm Field-Programmable Gate Array (FPGA) have proven the large effect of aging on responses, which reduces a PUF reliability. However, aging did not affect the randomness, which maintains a PUF uniqueness [111].
2.16. Temperature Effect on MOSFET Capacitances

The electronic charge which occupies a MOSFET’s gate surface can cause a shift in flat-band voltage ($V_{fb}$) [117]. However, as the applied voltage is varied, the Fermi energy at the oxide-semiconductor interface changes also and affects the occupancy of the surface states. Considering that ($\varphi_{st}$) is the surface potential at the threshold voltage level ($V_T$), then:

$$\varphi_{st} = \pm 2 \frac{kT}{q} \ln \frac{N_A}{n_i} \quad (2.16.1)$$

$$V_T = V_{fb} + \varphi_{st} \pm \frac{\sqrt{2qN_A\varepsilon_s|\varphi_{st}|}}{C_{ox}} \quad (2.16.2)$$

Where, ($C_{ox}$) is the oxide electric capacitance, ($\varepsilon_s$) is the silicon permittivity, ($k$) is Boltzmann’s constant, ($T$) is the temperature, ($q$) is the electron charge, ($n_i$) the intrinsic carrier concentration, ($N_A$) is the acceptor doping concentration in p-substrate, which is replaced by ($N_D$) as donor doping concentration in case of n-substrate.

For (2.16.1) and (2.16.2), the positive signs are for p-substrate and the negative signs for n-substrate. For p-substrate, $V_T$ in Eq. 2.16.2 becomes lower at low temperatures [117]. At the inversion region of a MOSFET, the total capacitance (when the drain and the source are shorted externally) equals the series combination of gate poly-silicon capacitance ($C_{poly}$), the oxide electric capacitance ($C_{ox}$) and the inverted channel capacitance ($C_{inv}$). If the applied voltage at the gate is ($V_G$), and the threshold voltage is ($V_T$), then:

$$Q_{inv} = C_{ox}(V_G - V_T) \quad (2.16.3)$$

$$Q_{inv} = \frac{\varepsilon_{ox}}{t_{ox}}(V_G - V_T) \quad (2.16.4)$$
The channel charge \( Q_{inv} \) increases at low temperatures, and this produces a rise in the total capacitance of the series combination \( C_{poly} \), \( C_{ox} \), and \( C_{inv} \) [117]. It is clear that the total charge is negatively proportional to the MOSFET temperature, which means that any increment to the temperature causes an increment in the discharge time.

2.17. Security Prospects

Data accessing to PUFs can be controlled by an application program interface (API) and hash or encrypting units to restrict any unauthorized applying of the challenges to the PUF.

The API may also prevent repeating a response to add unreliability to the side-channel modeling attacks [135]; therefore, such PUFs are called controlled PUFs [69]. Communications to the remote verifier are then protected along the communication path from end to end [136]. Security attacks are generally classified into invasive, semi-invasive, and non-invasive attacks.

2.17.1. Invasive Attacks

Invasive attacks require removing the chip’s packaging by etching, drilling, or laser cutting, then using a micro-probing workstation to probe the chip [137]–[142]. An active post-process spray coating containing inhomogeneous particles can be included as introduced in [143], [66] to; 1) embed a unique signature for smart cards by measuring some electrical properties at certain spots within the inhomogeneous coating material, 2) protect the chip from the invasive attacks, and 3) harden the basic optical non-invasive attacks. The active coating included particles of various permeabilities, shapes, and sizes. In a later work, metal sensors were layered beneath the passivation layer to form an active-coating capacitive PUF [144]. In [145], the metal sensors were reshaped into pairs of metal comb capacitors to increase the surface area exposed to the coating. This model was elucidated further in [146]. A newer active-coating capacitive PUF stacked more than one layer of metal comb capacitors [147].
2.17.2. Semi-Invasive Attacks

Semi-invasive attacks attempt to inject a temporary error into the PUF, either by a faulty instruction or by operating the chip in certain conditions. They would include testing the targeted chip with various input data, supply voltages, temperatures, and frequencies. However, such attacks do not permanently alter the properties of the chip [148]–[152].

2.17.3. Non-Invasive Attacks

Non-invasive attacks include optical, brute-force, man-in-the-middle (MITM), and side-channel attacks.

- **Optical Attacks:** These attacks aim to view the internal structure of the chip. Ray-resistant encapsulant shields can be used to resist such attacks.

- **Brute-Force Attacks:** An attacker tries multiple challenges to generate a specific response. This attack is more relevant to weak PUFs with a single CRP or a few CRPs. By contrast, it is not an imminent threat to strong PUFs, where each CRP is only used once or a few times.

- **MITM Attacks:** These attacks from eavesdropping, manipulating, to relaying are more relevant to uncontrolled PUFs, where communications to the verifier side are not encrypted. [153], [154]. Encrypted communications greatly decrease the viability of such threats.

- **Side-Channel Attacks:** These include optical emission power analysis (OEPA) [155], electromagnetic (EM) attacks [156], and differential power analysis (DPA) [157], [158]. In these attacks, an attacker analyzes a large number of power traces of the chip to discern its inward function. Decoupling capacitors can oppose such attacks [159]–[163].

In general, applying a post-processing coating can protect internal data communication between the PUF and the microcontroller from invasive attacks. Encrypting the communication between the smart card and the verifier side resists MITM attacks.
2.17.4. Randomness Tests

A PUF’s response bits should have a degree of randomness to protect the PUF against modeling attacks, thus responses of many PUFs were tested using the statistical test suite for random and pseudorandom number generators.

These 16 tests were suggested by the national institute of standards and technology (NIST) in its special publication 800-22 in 2001 [164] and were revised and republished in 2010 [165]. However, responses usually do not pass all those randomness tests, as not all of these tests are relevant. Furthermore, these tests are debatable and still under continuous development [166], [167].

2.18. Fabrication Variations of a MOSFET

This section discusses some major variations which occur throughout the fabrication process of a MOSFET-based chip. These random variations create distinct parameters for each MOSFET. A parametric variation is usually a source of disturbance for most applications, while for PUFs, it can be an excellent source of entropy to generate device-specific response bits [22]. This section discusses the intrinsic variations’ effects on the threshold voltage ($V_T$). The initial threshold voltage value of a MOSFET at zero-bias ($V_{T0}$) depends on the flatband voltage, the bulk potential, the oxide electrical capacitance per unit area ($C_{ox}$), and the depletion layer charges. The variation in the initial threshold voltage ($\Delta V_{T0}$) plays a vital role in forming distinctive properties for a PUF chip. The polarities of $V_T$ and $V_{T0}$ are positive for an n-MOSFET and negative for a positive-channel metal–oxide–semiconductor field-effect transistor (p-MOSFET). In both MOSFET types, there is a roll-off voltage that contributes to $|\Delta V_{T0}|$, which reduces $|V_{T0}|$.

For generic (non-PUF) applications, chip designers try to reduce the intrinsic variations by
taking into consideration some factors, which are briefed here at the technology, layout and mask, and the schematic levels. The introduction of strained silicon since 90 nm complementary metal-oxide-semiconductor (CMOS) technology and below enabled the electrons to move faster, which ultimately made MOSFETs switch faster. Straining the silicon results in better MOSFET performance and lowers energy consumption; however, it causes more deterministic, or systematic variability in a MOSFET’s parameters, such as $C_{ox}$, MOSFET channel width ($W$), and the charge carrier mobility ($\mu$), which can be specified as mobility of electrons/holes for n/p channel as $(\mu_n/\mu_p)$, respectively [168].

Statistical variability is another type of variation, it attributed to random variations in parameters like $V_{T0}$, $\mu$, $C_{ox}$, $L$, or $W$. However, the variation in $V_{T0}$ is the most prominent parameter and is primarily ascribed to random dopant fluctuation (RDF) within the channel and the gate; beside other factors, such as the random fluctuation of surface roughness, oxide charge, and the lithography driven line edge roughness (LER) [115]. The fluctuation in $V_{T0}$ mainly depends on RDF, as the dopant concentration can vary up to $\pm10\%$ [169].

2.18.1. Variations at Technology Level

The variation of $t_{ox}$ gained more importance as the technology shrinks it down. Using high-k insulators with larger physical $t_{ox}$ abates its influence on variation. Unlike $t_{ox}$, the effect of the oxide charge on $V_T$ value does not represent a significant factor in modern fabrication technologies, especially when a MOSFET works at strong inversion mode [170]. Let $K' = \mu C_{ox}(W/L)$, then an RDF of 10% with 5% variation in $t_{ox}$ can lead to variation of 100 mV in $V_{T0}$, 15% in $K'$, and 5% in the source-bulk (C$_{SB}$) and the drain-bulk (C$_{DB}$) junction capacitances [171]. Silicon-on-insulator (SOI) technology mitigates RDF, but an ultra-thin body SOI MOSFET raises more uniformity
concerns. The high-k/metal gate technology reduces RDF but creates metal gate granularity (MGG), which becomes among the major sources of $\Delta V_{T0}$ in scaled bulk-MOSFETs [172]. In general, a MOSFET fabrication technology tends to decrease $| \Delta V_{T0} |$ by: 1) strict control over doping fluctuation of the source and the drain; 2) strict control over the critical geometrical dimensions, such as the gate length and width of the MOSFET; 3) scaling up the gate insulator thickness; and 4) scaling up the junction depth [173]. Choosing a MOSFET manufacturing technology with less control of characteristics can better serve chip authenticity applications.

2.18.2. Variations at Layout and Masking Levels

Low lithography resolution is a major source of systematic local process variations. Such imprecision is treated with resolution enhancement techniques, such as phase shift masking by adding phase information to the mask, off-axis illumination to optimize the angles of light illuminating the mask, source polarization to control of the polarization of the illumination, source mask optimization, and optical proximity correction (OPC) to optimize the mask pattern shape [96], [114]–[117], [174]. Besides the lithography-related geometry variations, additional variations are added due to spacing, distances to the shallow trench isolation, and position variances of contacts. A regularized design and dummy features can greatly reduce the impact of systematic variability [112]. Among the basic layout approaches that are commonly applied to reduce such variations are; 1) matching the orientation of the transistors’ layouts to decrease mobility variations, 2) laying out transistors as close as possible and using common centroid layout to minimize gradients throughout the fabrication process, 3) using dummy gates at the sides of transistor layouts, 4) not laying out contacts on top of active gates, and 5) not laying out metal lines across active gates [50].
Within an electronic PUF circuit, wherever the intrinsic variations are desired, the layout and masking processes can follow special approaches to oppose or at least avoid any variation-limiting effects, whereas the typical anti-variation approaches can still be implemented elsewhere.

2.18.3. Variations at Schematic Design Level

An intrinsic-based variation of a MOSFET’s parameter among similarly designed MOSFETs within a chip is usually referred to as *intra-die variation* or *mismatch*.

These variations can be due to random local effects (such as the non-flatness of the polysilicon gate due to granularity) and to $t_{ox}$ gradients over the chip. The general trend of MOSFET technology evolution is toward smaller horizontal dimensions, smaller physical $t_{ox}$, smaller junction depth, heavier substrate doping, and (although not always) lower power supply voltage.

![Fig.2.5. Regions of planar n-MOSFET](image)

In Fig. 2.5, a planar bulk MOSFET has as a mask length ($L_M$). This length is reduced due to depletion of the source and drain junctions by a length ($L_{dep}$) on each side, leaving a channel length ($L$), which
is determined as [175]:

\[ L = L_M - 2 L_{dep} \]  \hspace{1cm} (2.18.1)

It is important to emphasize that \( L, L_M, \) and \( L_{dep} \) exist even without applying any external voltage at the MOSFET. After applying a voltage on the drain/source, lateral extents of the p-n depletion regions of the bulk-source and the bulk-drain are produced, they are denoted here as \( (\Delta L_S) \) and \( (\Delta L_D) \), respectively. When a voltage is applied at the gate, there will be a shared gate-source/gate-drain influence on the depletion charges of the channel at the source/drain sides, respectively. The remaining distance between the source and drain depletion is considered the effective channel length \( (L_e) \), where the diffusion charges are influenced by the gate voltage only [134]. Then the \( L_e \) is calculated as:

\[ L_e = L - (\Delta L_S + \Delta L_D) \]  \hspace{1cm} (2.18.2)

The effective length exists only when the channel is formed by applying external voltages at the MOSFET, while the bulk-source and the bulk-drain depletions exist even without applying any external voltage to any of the MOSFET’s terminals.

A typical channel region contains only about \( L_e^{1.5} \) dopant atoms; therefore, as a MOSFET fabrication process is scaled down, mismatch due to dopant fluctuation increases [176]. Furthermore, since \( VT \) is inversely proportional to the square root of the device area [177], therefore, the current will have a similar dependency on the area. For example, in analog designs, to improve the current matching by a factor of 2, the device area is quadrupled. One of the approaches to achieve that is doubling both \( W \) and \( L \).
In small-geometry devices, there are various factors which contribute to $\Delta V_{T0}$, such as non-uniform vertical and lateral doping concentrations, short channel effect (SCE), narrow width effect (NWE), and drain-induced barrier lowering (DIBL) [171].

At schematic design level, a designer can choose L and W of the MOSFETs. In short-channel MOSFETs, the SCE varies $V_T$ by a value denoted as $(\Delta V_{T0}^{SCE})$, then:

$$V_T = V_{T0} \pm \Delta V_{T0}^{SCE}$$  \hspace{1cm} (2.18.3)

Where $\pm \Delta V_{T0}^{SCE}$ is considered positive, despite the channel type. The variation’s sign in (2.18.3) is negative for n-channel and positive for p-channel. In other words, the term $\pm \Delta V_{T0}^{SCE}$ always carries an opposite sign to $V_{T0}$, as the depletion regions around the junctions (the wells) reduce $V_T$, where $\Delta V_{T0}^{SCE} \propto (X_j / L)$ [178].

On the other hand, in narrow-channel MOSFETs, when W is on the same magnitude order as the maximum depletion region thickness $X_{dm}$, another source of variation that contributes to $V_T$ must be considered. In addition to the oxide thickness above the channel ($t_{ox}$), there is a thick field oxide (FOX), which covers the region around the channel to prevent the surface leakage currents between adjacent MOSFETs. The overlapped area between the gate electrode and FOX develops a low depletion region, which raises $V_T$. This phenomenon is known as the narrow width effect (NWE) [179]. The narrow-channel-based variation of $V_{T0}$ is denoted as $(\Delta V_{T0}^{NWE})$, which always carries a positive sign. Then it affects $V_T$ as [178]:

$$V_T = V_{T0} \pm \Delta V_{T0}^{NWE}$$  \hspace{1cm} (2.18.4)
Where the sign in (2.18.4) is positive for n-channel and negative for p-channel. In other words, the term $\pm \Delta V_{T0}^{\text{NWE}}$ always has a similar sign to $V_{T0}$, and if the absolute values are considered, the narrow channel causes extra depletion charge that ultimately increases $|V_T|$. Combining SCE and NWE, then $V_T$ is determined as:

$$V_T = V_{T0} \pm \Delta V_{T0}^{\text{SCE}} \pm \Delta V_{T0}^{\text{NWE}}$$ (2.18.5)

In the case of short-narrow n-channel, then the variations ($-\Delta V_{T0}^{\text{SCE}}$) and ($+\Delta V_{T0}^{\text{NWE}}$) tend to cancel each other out, and a similar argument (but with opposite signs) is valid for a short-narrow p-channel [178].

At the design level, a PUF designer should oppose the general anti-variation approaches which most analog designers follow, such as:

- To avoid the intra-chip mismatch among MOSFETs, which is attributed to statistical randomness, if the current should match, a high gate-to-source voltage ($V_{GS}$) is usually recommended to decrease the influence of $V_T$ [180]. Another approach to decrease the current mismatch by a factor of 2 is by squaring the MOSFET’s area [171]
- On the other hand, if the voltage should match, it is recommended to keep low $V_{GS}$. That can be done by increasing the channel width-to-length (W/L) ratio [180], [115].
- In some cases, a designer can choose between using n-MOSFET or p-MOSFET. Usually, n-MOSFETs have more intrinsic variations than p-MOSFETs. The random discrete dopants, the LER, the polysilicon granularity of the gate electrode, and surface potential pinning at the poly-Si grain boundaries play an important role in the statistical variation within n-MOSFETs while playing a negligible role in p-MOSFETs [127].
For resistors, an N-well diffusion resistor is less susceptible to the intrinsic variations due to its lower doping and larger volume. Furthermore, it is made of monocrystalline materials, which reduces the impact of defects and grain borders [181]. Alternatively, a MOSFET-base resistor is more susceptible to the intrinsic variations.

From all above, to improve the uniqueness of MOSFET-based PUF design, it is significant to include short-channel MOSFETs and/or narrow-channel MOSFETs. In either way, $V_T$ would be more distinct among the equally scaled MOSFETs, which can eventually lead to more distinctive PUF chips. Since such differences are attributed to random intrinsic variations, it is quite probable to have unique properties for each PUF chip. Reversing or at least avoiding any variation-limiting approach allow for more intrinsic variations, which can create unique PUF chips. Furthermore, scaling down a MOSFET increases the sensitivity to threshold voltage variations [182].
Chapter 3

Literature Review

This chapter reviews the most basic types of PUFs in general and mixed-signal PUFs in particular. It also presents various PUF classifications.

3.1. PUF Overview

Most electronic devices nowadays are fabricated using sub-micron technologies, which are highly susceptible to the impacts of process variations. These variation cause variations in the physical parameters of the manufactured devices. These parametric variations are usually annoying for most circuit applications, but here they turn out to be an excellent entropy source to generate device-specific authentication bits. These variations are not controlled by the manufacturer, therefore they are considered random (or stochastic) variations. The physical parameters of devices after the manufacturing process may refer to different kinds of physical effects depending on the media used for implementing the PUF [106]. Physical parameters of such fabricated devices are often stochastically distributed [38], frequencies of oscillating structures [37], the difference in transmission-speed of racing signals in some propagation paths [46], property disparities of coating layers [143], the randomness of the initial state of memory or memory-like circuitry [183]. This chapter reviews some PUF implementations at the reconfigurable, and the very large scale integration (VLSI) implementation levels. Although several anti-counterfeit approaches have been presented earlier, the term ‘physically uncleanable function’ was not coined yet, and those approaches were not presented along with comprehensive mathematical representation forms. Since Ravikanth proposed the optical PUF in 2001 as a
physical one-way function and discussed its unclonability [38], many PUFs were proposed [184]–[187].

Fig. 3.1 shows a basic classification approach where a PUF can be classified either according to their security levels, applications, physical contents, or according to their sources of variation. The main interest here is about the electronic PUFs, with more focus on the mixed-signal ones. PUFs can also be broadly classified depending on their physical contents into electronic and non-electronic PUFs.

Fig. 3.1. PUF classification
• **Electronic or Silicon PUFs**

In this type of PUFs, the challenge-response determining methodology is determined based on the electronic properties of an object such as delay of a gate, the threshold voltage of a transistor, etc. Among the well-known PUFs, the *integrated circuit identifier* (ICID) [188], [76], [189], static random access memory (SRAM) PUF [60], [190], arbiter PUF (APUF) [72], [46], [110], [110], ring oscillator PUF (ROPUF) [191], butterfly PUF [62], data (D) flip-flop PUF [58], power distribution PUF (or resistive PUF) [192], coating PUF [66], LC PUF [193], MEmory cell-based chip authentication (MECCA) PUF [194], glitch PUF [195], and the bistable ring PUF [196].

Electronic PUFs are categorized further according to their implementation approaches, into:

- **Digital PUFs:** It is first important to emphasize that all electronic PUFs receive, and generate digital inputs and outputs, but internally, all challenging processes are analog in nature due to the inherent physical characteristics of the electronic devices. That may explain the limitations which all current PUFs face, such as their susceptibility to environmental conditions. The terminology ‘digital PUFs’ can carry various meanings; however, in this work, it refers to all non-mixed-signal PUFs. In other words, it refers to any electronic PUF which generates its response in a digital form directly without needing an analog-to-digital converting stage. Recently, the term ‘digital PUFs’ in most research works would refer to reconfigurable PUFs, which usually implemented as FPGA designs [197], [198]. Digital PUFs are discussed further in Section 3.2.
Mixed-Signal PUFs: This type of PUFs is based on an analog operation and uses embedded analog measurement techniques; then ultimately, it uses an analog-to-digital conversion stage to end up having a digital response. Examples of such PUFs are the ICID (or the Threshold Voltage PUF) [188], [76], [189] and the Inverter Gain PUF [199]. Mixed-signal PUFs are discussed further in Section 3.4.

Non-electronic PUFs

When a PUF is not based on variations of electronic devices, then such a PUF is a non-electronic PUF or non-silicon PUF. In this type of PUFs, the challenge-response mechanism is determined based on some non-electronic properties of an object. The optical PUF family is an example of non-electronic PUFs. This family includes the optical PUF of proposed by Pappu [38], then by Pappu et, al.[73] then further developed by Rührmair [186], CD PUF [200], and the paper PUF [201], [202]. The optical PUF proposed by Pappu [38] has been made up of transparent optical medium containing bubbles. Shining a laser beam through the medium produces a speckle pattern (the response) behind the medium that depends on the exact position and direction of the incoming beam (the challenge).

PUFs are also categorized according to their source of variations into intrinsic and non-intrinsic PUFs [45].

The Intrinsic PUFs

For the intrinsic PUFs, the variations are obtained unintentionally due to production process uncontrollable sources of randomness.

For example, in electronic devices, the variation of doping, the thickness of the gate oxide, and geometrical variations among the fabricated MOSFET devices can cause variation in the device’s parameters, such as the threshold voltage ($V_T$) parameter. If measuring such parameter
is made possible, that presents readability of a unique identity for such devices, and that uniqueness is inherited to each object which includes such a device. In general, a PUF is considered intrinsic when it fulfills both of the following conditions:

- Randomness is inherited during production process implicitly, without any added effort or cost.
- Response evaluation is performed internally (even if externally corrected). That means less chance for measurement errors and less external influences during response evaluation. The other advantage is security; as for example, in case of weak PUFs; it is essential to have the response evaluation internally, then encrypted before disclosing it to outside the chip.

The intrinsic PUFs include: the Arbiter PUF [72], [46], Exclusive OR (XOR) Arbiter PUF [203], Forward Feed (FF) Arbiter PUF [110], Ring Oscillator (RO) PUF [39], [191], [204], Improved ROPUF [205], Enhanced ROPUF [206], Glitch PUF [207], [195], [208], SRAM PUF [60], [183], [190], Latch PUF [63], Flip-flop PUF [58], [209]. Butterfly PUF [62], Bistable Ring PUF [196], ICID [188], [76], [189], SRAM Failure PUF [210].

- **The Non-intrinsic PUFs**

The non-intrinsic PUFs are called so because they either are not completely embedded in the object, or they are not produced in the standard manufacturing process of their embedding object, or both [45]. The non-intrinsic PUFs can be in various forms, such as:

- The optical-based PUFs, which include the CD PUF [200], paper PUF [201], [202], and the phosphor PUF [211], [212].
- The electronic-based PUFs, which include the magnetic PUF [213], acoustical PUF [214], coating PUF [66], [144], power distribution network PUF [192].
- The RF-based PUFs, which include the RF-DNA [215] and the LC PUF [193].

PUFs can also be classified according to their application type, into stand-alone PUFs, and integrated PUFs [105]. Finally, PUFs are sometimes classified according to their security levels, into strong and weak PUFs [26]. Weak PUFs usually protect themselves by an algorithm which controls the access to the PUF and thus called controlled PUFs [153], [69]. The focus of this research is on Electronic PUFs, which are intrinsic as the variations are attributed to manufacturing processes.

3.2. Digital PUFs

Digital PUFs are part of the electronic PUFs. They are subdivided further into memory-based PUFs (whether at the register level, or at the array level) and delay-based PUFs (whether the delay is at the transistor level, or at gate level).

3.2.1. Memory PUFs

Memory-based PUFs are conceptually the simplest kind of PUFs. A digital secret key K is embedded in a tamper-proof package along with some logic circuit which computes Response = RF (K, Challenge), where RF is a random function. Relying on an embedded key makes such a system, not a compelling PUF. Most of the known digital PUFs are memory-based. Memory PUFs can be further classified into two categories according to the stability of the embedded key:

- **Initial state value memory-based PUFs**

They are based on the random initial digital value at the power-on time of some memory elements such as latches and flip-flops; therefore, such PUFs are called metastable- (or bi-stable) memory-based PUFs. Examples of such PUFs, the SRAM PUF [190], [60], Flip-flop PUF [58], [209], Butterfly PUF [62], [216], and the Buskeeper PUF [59]. In this category,
there are some weaker PUFs which are known as ‘Physically Obfuscated Keys’ [37] which are based on a type of NVM which is harder to attack than the regular Electrically Erasable Programmable Read-Only Memory (EEPROM).

• Steady-state memory-based PUFs

They are based on reading a steady digital value stored in an address that is randomly chosen by the challenge word. The memory is commonly a large read-only memory (ROM), such as super high information content (SHIC) PUF [217], [218], SRAM Failure PUF [210], and static memory-based PUFs [63], [219], which also known as mono-stable memory-based PUFs.

3.2.2. Delay PUFs

Some silicon PUFs are based on the propagation delay variations of digital signals across separated paths across identically designed devices [39], [46], [110], [191]. The racing signals are usually manipulated by a chain of switches and interconnects to influence manipulate their delays, which ultimately leads to generating a non-predicted output response [77]. A ring of an even number of inverters has one of two possible states when powered up; based on that, the bistable-ring PUF (BR-PUF) was proposed [196].

3.3. Mixed-Signal PUFs

A PUF in this category is considered as a mixed-signal PUF if it adopts an analog measurement technique, then an analog-to-digital converter to convert its response into digital form. Most known mixed-signal PUFs are based on fabrication variations such as (geometric, and doping) which their influence is obvious mainly on the threshold voltage parameter of the MOSFET. In this research work, the mixed-signal PUFs can be subdivided into:
• **Memory-based Mixed-Signal PUFs**

They include:

- super-high information content (SHIC) PUF [217], [218]
- SRAM Failure PUF [210]
- MECCA PUF [194]

• **Non-Memory-based Mixed-Signal PUFs**

They include:

- threshold voltage PUF (or. ICID PUF) [188], [76], [189]
- inverter gain PUF (or nanokey PUF) [199], [220]

3.3.1. **ICID PUF**

The ICID PUF was introduced even before coining the ‘PUF’ terminology [221]. In 1999, Keith Lofstrom filed the ICID patent which has been issued in 2000 [188]. It was implemented first as 128 bits using 350 nm MOSFET technology as presented in 2000 [76]. In 2007, it was extended to 256 and implemented using 90 nm technology [189]. Lofstrom also announced in [189] that SiidTech company [222] with Hitachi Solutions Technology Ltd. (former Hitachi ULSI Systems Co.,Ltd.) [223] have deploying ICID blocks in millions of integrated circuit chips. The ICID is based on variations in threshold voltage among addressable MOSFETs in an array with common gate and source and sequentially selected drains, driving a resistive load. These MOSFETs are switched (according to the submitted input challenge bits) to derive a resistive load. Device mismatches cause randomly different drain currents which produce a sequence of random voltages across the load.
The response sequences are different for every integrated circuit die because every transistor has a different distribution of random dopant atoms. The random analog voltage sequence across the driven load is converted to a binary identification sequence by an auto-zeroing comparator, which is controlled by a successive approximation register (SAR). Fig. 3.2 shows an ICID array of devices producing a difference voltage sequence. The MOSFETs which form the sequence are addressed like a memory to generate the response bit-by-bit using a sequencer unit, as shown in Fig 3.3.

![Fig.3.2. Array of transistors producing a sequential random voltage](76)
Experimental chips were fabricated using a 0.35µm single-poly N-Well process. The layout consisted of two metal layers for the ICID blocks. Each ICID block has 112 identification cells, made of minimum-sized transistors, and a static logic has been used. The properties of the fabricated PUF are briefed in Table 3.1.

### Table 3.1. Fabricated ICID ASIC properties

<table>
<thead>
<tr>
<th>Implementation Technology</th>
<th>0.35 µm single poly CMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chip area</td>
<td>$252 \times 93 , \mu m^2 = 23436 , \mu m^2$</td>
</tr>
<tr>
<td>Chip area with pins</td>
<td>$2060 \times 1820 , \mu m^2 = 3749200 , \mu m^2$</td>
</tr>
<tr>
<td>ID bits</td>
<td>128</td>
</tr>
<tr>
<td>Number Samples</td>
<td>55 chips- 7260 ID blocks</td>
</tr>
</tbody>
</table>

The chip area in Table 3.1 includes 112 identification cells plus the test logic. The reported average distance distribution ($\mu_{\text{inter}}$) is very close to 50%. The worst-case drift of the response has been estimated by variations in frequency, temperature, and bias. The ICID block tolerates a
wide range of power supplies, biases, clock frequencies, and temperatures. Typical operating ranges are shown in Table 3.2.

Table 3.2. ICID block measurements

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Min.</th>
<th>Nom.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{DD}$</td>
<td>1.1V</td>
<td>2.5V</td>
<td>5.0V</td>
</tr>
<tr>
<td>ID</td>
<td>50 µA</td>
<td>100 µA</td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td>-</td>
<td>25º C</td>
<td>125º C</td>
</tr>
<tr>
<td>Frequency</td>
<td>30 Hz</td>
<td>500 kHz</td>
<td>25MHz</td>
</tr>
<tr>
<td>Bit Rate</td>
<td>2 bps</td>
<td>30 kbps</td>
<td>1.5 Mbps</td>
</tr>
<tr>
<td>Bit Drift Error</td>
<td>0 %</td>
<td>1.3 %</td>
<td>5 %</td>
</tr>
</tbody>
</table>

The 112-bit experiment, with less than 4% drift, can reliably distinguish more than 1 million IDs with less than a $10^{-7}$ error rate. ICID is based on mixed-signal techniques, but the inputs and outputs are digital. To check the nominal drift, five Application-Specific Integrated Circuits (ASICs) each of 660 ID block were heated to 250º C for 100 hours to test the drift. ID shifts of 6 or fewer bits were observed, with an average of 1.5 bits, which means a drift (or an average intra-distance, or a sample mean), $(\mu_{\text{intra}})$ of 1.3%. 55 chips were packaged and tested, for a total of 7260 ID blocks. No failures were discussed, with each ICID block producing an ID differs from all others by an inter-distance of at least 27 bits.

3.3.2. Silicon Nanokey

In 2008, the silicon nanokey PUF (or Inverter Gain PUF) has been first proposed by Puntin [199], then extended in [220] and its immunity against brute force attacks was tested in [224]. This PUF is based on the variation in gain of similar inverters. The difference in gain between a pair of inverters is determined in an analog electronic measurement ($V_{\text{out}}$) from an array of inverters, as shown in Fig. 3.4.
Fig. 3.4. Row of the inverter array\cite{199}, \cite{224}.

This voltage is then converted into a single bit response. One p-MOSFET is used for each row, to save the area; furthermore, each p-MOSFET should be large enough to neglect the mismatch, unlike what is required for the n-MOSFET in the array of invertors as they should be at the minimum feature size to obtain the maximum possible intrinsic variation during the fabrication process\cite{199}.

Fig 3.5. Silicon nanokey PUF block diagram \cite{199}
The inverters are set in (N) rows and (M) columns, where the challenge through X decoder can address the rows by word-lines (WL_i), and through the decoder (Y) it can address the columns by bit-lines (BL_n), as shown in Fig. 3.5. The entire system including voltage regulator (process regulator), sample and hold analog to digital converter (bit extractor), and the error detector to flag any faulty response-bit; however, the system did not include an error correcting unit. Every response bit is generated by the sign of the difference between consecutive output voltages from the sample and hold unit (or analog processing unit). When any two consecutive voltages are almost equal, in other words, when the difference between them is less than the decision margin voltage difference (VDM), then noise, temporary (i.e. operation temperature, or voltage), and transit variations (i.e. aging) can easily flip the response bit value. Here comes the role of the detector unit (the output comparator unit) during the enrolment phase, as it generates a response flag (RF) bit whenever it detects such a non-stable response.

The string of RF bits acts as a mask to exclude the unreliable response bits, to exclude them from the reference list of the reliable CRPs’ list at the database. In [199], the VDM was set to 44 mV, and then it was reduced to 35 mV in [224]. At room temperature and VDD of 600 mV, the bit error rate (BER) in [199] was lower than 0.01% and less than 0.009% in [220]. In worst conditions of 125°C or with 10% variation in VDD, the BER in both [199] and [220] was lower than 0.01%. The consumed power in [199] and [220] was 30 µW and 38 µW, respectively. The chip area was 18,000 µm² and 35,000 µm². The authors did not present the inter-distance, but they mentioned that the correlation between response strings from different PUF instances is less than 1%. In [220], experimental devices were fabricated using a 90 nm, N=256 inverter rows, and M= 16. The chip area in Table 3.3 includes sequencer units which interact with each M/2 challenge bits to address a new response bit among the whole 256 challenge bits.
The input bit rate was 768 kilobit per second (kbps), and the output bitrate was 6.25 kbps. The effect of voltage and temperature variations on the BER is briefed in Table 3.4 [220].

### 3.4. Problems in PUFs

In general, PUFs have various problems. First, they usually adopt a parallel individual bit approaches to generate the response, sacrificing chip area, power, and simplicity of design, just to narrow the environmental-variation errors caused by temperature and voltage variations. Nevertheless, not only all known PUFs still suffer from environmental-variation-caused errors, but also the individual bit approach opens a wide door for individual bits’ random errors. There is no previous attempt to reduce the number of random errors by following a multi-chunk approach to generate processable chunks of bits, where each chunk has a significance and can be treated as a physically meaningful value instead of merely a bunch of random bits. Second, most previously known PUFs use on-chip ECC units, scarifying a chip’s area, power, design simplicity, process

### Table 3.3: Fabricated nanokey PUF ASIC properties

<table>
<thead>
<tr>
<th>Implementation Technology</th>
<th>90nm CMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chip area with pins</td>
<td>155 µm × 226 µm = 35,030 µm²</td>
</tr>
<tr>
<td>ID (response) bits</td>
<td>256</td>
</tr>
<tr>
<td>Number Samples</td>
<td>24 chips</td>
</tr>
</tbody>
</table>

### Table 3.4. Nanokey PUF measurements

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Min.</th>
<th>Nom.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>V&lt;sub&gt;DD&lt;/sub&gt;</td>
<td>0.54V</td>
<td>0.6V</td>
<td>0.66V</td>
</tr>
<tr>
<td>Power</td>
<td>38 µW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td>/</td>
<td>25° C</td>
<td>125° C</td>
</tr>
<tr>
<td>BER in standard conditions</td>
<td>&lt; 0.009%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BER in worst conditions</td>
<td>&lt; 0.01%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
time, and security at most. However, their responses are still not error-free. Third, most weak PUFs are digital, and they are simpler to attack, but they require API protection, and they rely on encryption or hashing units, which themselves are vulnerable to attacks, besides the overheads of area, power, and data processing time. On the other hand, strong PUFs are mostly mixed-signal and more complex to attacks and they would be safe to access without even an API control. However, this comes with complexity to correct errors and to check the non-reliable bits during the enrolment phase, which complicates such phase and raises its cost. Besides, such PUFs would ultimately have less reliable bits than the total generated response bits, which raises another issue of reliability Fourth, in propagation delay-based PUFs, having the challenging bits propagate simultaneously within a small chip area can cause crosstalk errors, where an induced voltage can flip a bit in an adjacent path. This is a saddling issue in many PUFs, for example, an arbiter PUF is generally more vulnerable to crosstalk-induced errors compared to memory-based PUFs, even compared to the star-up ones. This is because of the temporary electromagnetic coupling among the adjacent logical paths, which can increase the propagation delay of some paths more than others, which would ultimately bias the arbitration of a delay-based PUFs [25].
Chapter 4

Objective and Methodology of the C-PUF

This chapter presents the objectives, constraints, methodologies, schematics, and the modeling diagrams of the proposed C-PUF.

4.1. The C-PUF Design Objectives and Constrains

The main objectives of the proposed PUF are to be strong, without on-chip ECC, and compact in terms of area, power, and process time to suit for the RFID applications. The objectives are discussed with details through the following subsections.

4.1.1. Strong PUF Constrains

All the previously proposed PUFs are candidates for being strong PUFs, and none was recognized as one. This can be attributed to the fact that no standard test-bench has been agreed upon yet, but since the response of most PUFs are also used as identification ID/cryptographic key; therefore, most researchers test such PUFs’ securities in terms of randomness, exactly as a random number generator.

In particular, most digital PUFs have been demonstrated to be mathematically modellable; therefore, the PUF in this research was designed as a mixed-signal PUF. This PUF should fulfill the constraints of a strong PUF such as repeatability, accuracy, robustness, reliability, and randomness. That can be framed into four major aspects:
• **Physical Unclonability**
  
  Statistical and random variations during a MOSFET manufacturing process make it imposable to reproduce an identical physical copy of a chip even by the same manufacturer, at least with the currently known technologies.

• **Accessibility**

  The proposed PUF includes ADC and DAC stages; therefore, it is considered a mixed-signal PUF. This category of PUFs is known for its strength; therefore, a mixed-signal PUF does not usually need any access protection protocol. In such a case, the design can allow any external entity to apply multiple challenges and read out their corresponding response without representing a security threat to the strong PUF. However, the proposed PUF is generic and can comply with any suggested communication protocol for the authentication phase and having an encryption unit to encrypt the response before sending it to the verifier side.

• **Anti-Tamper Sensitivity**

  It is important for any key generating/storing hardware to have a self-destructive ability to serve as an anti-tamper feature for the system [225]. The proposed PUF is capacitive-based, which can increase the anti-tamper sensitivity to counter invasive attacks by implementing one or both of the following approaches:

  - To have a tamper-resistant shielded PUF, in the new PUF, it is recommended to have the capacitors as a multi-finger MOSFETs to reduce the physical area compared to the with the metal-comb capacitors that were used in [146]. It is also possible to connect some metal combs to the electrodes of the MOSFET.
- To have a tamper-resistant coated PUF, the C-PUF is to support the active coating of the chip with an opaque coating layer that contains a random mixture of conductor and insulator particles of random sizes, shapes, locations, and dielectric constants.

Both of metal-comb capacitors and active coating approaches can help to:

- counter the non-invasive optical attacks using optical microscopic inspection toward the underneath layers of the PUF chip;

- counter electric measuring probes, as such an action can change the capacitance causing temporary error responses, which means the PUF can generate the correct response as soon as the probe is removed;

- counter invasive attacks such as displacing any metal bar or coating layer can permanently change the capacitance of the affected portion and therefore influences the PUF response;

- beside the enhancement to the tamper-resistivity feature mentioned above, the metal combs with the active coating can improve the inter-distance among the capacitors within the same chip, and that increases the inter-distance of the total response among the other tested chips within the class. In other words, the PUF class average uniqueness, which depends on the inter-distance \( D_{\text{ave},i}^{\text{inter}} \) will increase. This increases the identifiability of the PUF individuals within that PUF class.

- **Mathematical Unclonability (Unpredictability)**

  To harden the mathematical clonability approach (i.e., modeling attacks) the design takes into consideration three aspects:

  1) **Anti-Modeling Complexity**: All the known digital PUFs have been proven vulnerable to model attacks; therefore, they are considered weak.
2) To avoid that, the proposed PUF will utilize analog circuits to add more ‘confusion’ as it is one of the basic principles of cryptography [226]. Furthermore, the new capacitive PUF (C-PUF) adds another source of confusion as it is not necessarily that the largest capacitor generates the largest count; that is done by choosing some counters of one certain size (in this work, fourteen counters each is of 8 bits), such that most of these similar counters have to recount from zero for many times to represent some long discharge times. That depends on the challenge sets, environmental situations (for example, in low temperatures, and the range of supplied voltages). This adds more confusion against an attacker and saves in the design area.

Successful modeling for such a mixed-signal PUF may need an invasive attack to discover the internal structure of the PUF. Such as the MOSFET capacitors, their fingers, and areas. Hypothetically, even if an attacker does that without touching the coating layers by using an x-ray screening, then an attacker should estimate the confusion effects caused by variations in temperature, voltage, and aging. Furthermore, an attacker needs to precisely estimate the effect of both of discharge load control bits, and the comparator reference control bits, on the generated chunks of response bits.

3) **Number of CRPs:** The presented PUF here has only a 21-bit input challenge (i.e. 2,097,152) challenge-response sets), which is supposed to be sufficient for most everyday applications. However, in case of needing higher security levels, 64-bit CRPs, for example, then an extended version of the proposed circuit can be used. The uncontrollable variations during fabrication, besides the environmental disorders during verification, both are expected to make this PUF immune against physical and mathematical (or modeling)
cloning attacks. It is necessary to emphasize that the variations affect all devices within the proposed circuit

(not only the capacitors); therefore, all the PUF chip devices contribute to making a unique response. Furthermore, using a system clock pulse of 100 MHz or above to the counters (which is used to count the discharging times) would make any minute variation generate a noticeable variation in the counter output response.

Having 21-bit challenges (i.e. 2,097,152 CRPs) does not necessarily mean that it must be used all; for some RFID tag applications, either the product is sold once, or in the worst-case scenario the buyer return it to the shop, then the product is sold again; therefore, for such PUF applications even few CRPs can be sufficient.

The other important concept to clarify here is that for the RFID for such applications, the same few challenges can be applied to other PUF siblings within the class, and the response will be different for each. The question here, within the DB, how the verifier knows which lists to look into? The answer is by having an ID number to claim. This code or serial number is stored on a memory chip within the RFID chip. The verifier then picks a challenge from the DB titled by that claimed ID. After challenging the PUF and getting the response as a reply from PUF chip to the verifier side, then the verifier compares the received response with the proper response against that used challenge.

This is why having a limited number of input challenges, does not restrict the possible population number of PUF siblings within a particular product family (i.e., class).

4) **Correlation Among Response Bits:** All the previous PUFs have a standard approach, which is to generate (N) individual bits from (N) different units. Applying the autocorrelation can help to assess the relationships among adjacent response bits.
In this PUF, the case is different, as the bits represent real numbers which refer to the discharge time rather than being merely random bits. However, it is possible here to test the correlation among bits having similar positions among the chunks, in particular among chunks, which are generated from similarly sized units, i.e. the challenged CCs. Furthermore, it is possible to study the correlation among the RCs generated from similarly positioned counters among other similar PUF chips within the class.

4.1.2. Design Constrains

The C-PUF design is meant to meet some physical and computational constraints, i.e., area, power, and data processing time, to serve the authentication purposes of various embedded systems. Besides the tight area and process time constraints, the C-PUF chip should comply with the tight power constraints of the limited induced power by the embedded coil within smart cards and fobs. The proposed C-PUF is meant to be embedded into RFID tags, smart cards, subscriber identity module (SIM) cards, TPMs, and DRM systems.

4.2. The CRP Expansion Concept

If a high-resolution authentication gadget can verify a person by scanning one finger only, would it need to scan the other fingers?

Verifying an electronic PUF chip would be even more secure because; 1) it is harder to counterfeit the internal structure of a chip physically or to model it mathematically, as it is not as exposed as a human’s fingerprint; 2) a challenge word of a PUF is only used once or for few times then it is altered, unlike a human’s fingerprint. This CRP expansion concept makes it possible to challenge a chip using a smaller number of bits for the challenge word. This approach can be recommended under these conditions:
1) High-resolution measuring approach for the physical properties of the chip, generating a large number of bits.

2) A PUF structure that is hard to model, as the case of mixed-signal PUFs.

Applying the CRP expansion concept is not recommended for weak PUFs, which interact with a small number of challenge words and are easy to model.

All previously known PUFs (weak and strong) have used an equal number of challenge and response bits. In this PUF, an expansion idea is proposed by handling input of a 21-bit challenge to generate a 128-bit response. It is important to emphasize that using a smaller number of bits for a challenge does not lemmatize the product population. For example, having only 21-bit of challenges within the verifier DB does not limit the PUF product population to (2,097,152) only, as an overlapping of input challenges among the product population is allowable since each response is unique, which eventually creates a unique CRP for each PUF chip. Then ideally, the product (class) population is limited by the number of the response bits. The other unaccepted repetition is within the claimed ID code of the smart card, which is usually read from non-volatile memory to initiate the verification phase by referring to the related list of CRPs within the DB at the verifier side. The assumption here is that the anti-modeling complexity is a more vital factor to maintain security than the number of CRP words, and the number of bits per challenge and response words. It is assumed that for most applications, 2,097,152 CRPs are enough to immune the PUF against modeling attacks, as long as the proposed mixed-signal PUF is complex enough. In this dissertation, a 21-bit challenge word is used to generate 128-bit response word; this expansion approach reduces the design area.
4.3. The C-PUF Authentication Protocol

The main features of this PUF are explained by going through a basic verification protocol step-by-step:

1) After reading the ID of a smart card, the ID is sent to the verifier side. Then, according to that received ID, the verifier’s software randomly picks a challenge among those which were archived for this claimant PUF during the enrollment phase.

2) The PUF receives an input challenge of 21-bit, to generate an output response of 128-bit.

3) The PUF sends the 128-bit raw response to the verifier side without correcting it on-chip. From the PUF side, the verification phase is done by generating the response; then, the verifier correlates the received response with its related archived one.

There are two major concepts to emphasize here:

A. To enable the verifier’s software to handle a non-corrected response, among the 128-bit response, there is a 16-bit chunk, which conveys the resultant of all effects (i.e., temporary and any other temporary variations) to the verifier side. This enables the verifier’s software to analyze the PUF response flexibly and reliably. That also eliminates the need for an on-chip ECC unit. The 128-bit response is generated in 15 chunks, one is of 16 bits, and the rest 14 chunks are of 8 bits each. Each chunk of bits is generated by an individual counter that presents the discharge time of its related CC.

B. To maintain security, the excess of discharge time of any of the 14 CCs makes its related counter behave like an 8-bit roulette; that is to maintain the unpredictability of the response. Only the verifier’s software has the accessibility to the related archived response at the DB, so it will be able to correlate the received response to the archived one and decide whether to approve or deny the claimant PUF.
4.4. The C-PUF Diagram

The proposed PUF is based on charging a set of n-MOSFET capacitors, then counting the system clocks during the discharging time. Each capacitor-counter pair individually yields chunks of response bits that is an RC, which represent the discharge time of that individual capacitor.

![Diagram of the C-PUF](image)

**Fig. 4.1. Simplified diagram of the C-PUF**

In Fig. 4.1, the proposed PUF diagram is shown. The design should maintain the same K and M challenge bits for all (capacitor-challenging) bits to simplify the error-correcting algorithm at the verifier side. A capacitor challenging bit is a bit that decides whether its controlled capacitor will be charged or not for an authentication phase (or verification session). To add some unpredictability, each challenge-controlled capacitor has a built-in (i.e. parallel) always enabled MOSFET capacitor. Charging capacitors at the initiation time of the authentication phase would
theoretically represent a weakness toward power analysis attacks, i.e., electromagnetic analysis (EMA), simple electromagnetic power analysis (SEPA), differential electromagnetic analysis (DEPA), simple power analysis (SPA), and differential power analysis (DPA)).

However, some concept must be kept in mind:

- All capacitors are charged simultaneously.
- Charging is only for once per an authentication phase.
- Charging period is 10 ns only; an adversary can not collect enough power traces to analyze the circuit behavior.
- Challenging bits’ combination is usually changed for every authentication phase and would never be used again.

All the above points countermeasure the power analysis attacks.

In the proposed PUF, 15 CCs were used; each CC includes a pair of two n-MOSFET capacitors. Each unit includes; one constant (always enabled) capacitor and another challengeable capacitor. That forms 15 MOSFET CCs, which are challengeable with 15 challenge bits. A CC includes MOSFETs in various sizes and can be networked in various configurations to randomize the output responses, to attain less predictable responses. This helps the C-PUF to be more immune to modeling attacks.

There are two challenging subsets L and M, each of 3-bit subset of challenge bits. The L-subset challenges PUF through controlling the non-linear discharge load, which affects the discharge time. The M-subset challenges the PUF by controlling the reference voltage level of an analog comparator. L and M subsets have to be decoded first through 3x8 decoder, then drives an 8-bit DAC.
The proposed PUF circuit was simulated using Cadence with the BSIM4 model and general-purpose development kit (gpdk045) library. Fig. 4.1 shows that the C-PUF diagram consists of 5 main modules; 1) challenge-controlled capacitor; 2) challenge-controlled discharge load; 3) challenge-controlled reference voltage; 4) comparator; and 5) N-bit counter. The PUF employs 15 CCs within the challenge-controlled capacitor module. The structure of each CC is shown in Fig. 4.2.

![Simplified capacitive cell](image)

**Fig. 4.2. Simplified capacitive cell**

Each CC is composed of $2 \times n$-MOSFET transistors. The primary capacitor (NM0) is charged each time an authentication phase is initiated, and this will make the related counter generate a non-zero value every time. The charging operation of the secondary capacitor (NM4) is controlled by the challenge signal (ch) which is a bit among the field (ch20-ch6); this bit drives a gate of p-MOSFET transistor (PM0) to control (NM4). A single charge trigger pulse and discharge trigger pulse are supposed to be generated by an implicit control module within the smart card or the RFID tag.
The discharge-trigger signal initiates the discharge cycle through an array of 15 transmission gates, one per CC. The challenged-controlled comparator reference voltage in Fig. 4.1 is controlled by the output of an R-2R ladder DAC. Each discharging terminal (Vcap) of the 15 CCs is related to an individual discharge load, comparator, and counter. An individual comparator’s output is 1V when the discharged voltage (V_{dis}) is more than or equals to the reference voltage (V_{ref}). Otherwise, it would be 0 V.

Each comparator output is connected to a low-pass filter (LPF) of a 3fF shunt n-MOSFET capacitor, which is connected to the ground potential of the circuit; that is to filter the 100 MHz switching noise of the driven clock-gating NAND. Counter modules are implemented as synchronous counters to cope with the 100MHz of the clock pulse. Each counter measures the discharge time of the related CC. For the simulated circuit, the challenge bit field values are K=15, L=3, and M=3 bits. Regarding (N), the counter module in Fig. 4.1 contains (14x 8-bit) counters and (1x 16-bit) counter. The 16-bit counter measures the discharge time of the smallest CC that includes a pair of two n-MOSFET transistors; each is sized as (length=45nm, width=120nm). This 16-bit response chunk also serves as an environmental (mainly temperature) indication which conveys the PUF status to the remote authenticating algorithm.

Each 8-bit counter is used to measure the discharge time for a larger CC and functions as modulo (overflow) counter, or in other words, just like an electronic roulette wheel. Thus, to guess the winning number, i.e. the response to the challenge word, it is required to guess the 128-bit correct combination of the 14 (8-bit) roulette wheels plus the 16-bit of the reference counter as well. All responses of all counters are manipulated by a randomly selected combination of 21-bit input (i.e. a challenge word), which is used once per lifetime of the product. Within a challenge word, the 16 least significant challenge bits (ch5-ch0) control the 15 subdivisions of the entire
PUF. The bits (ch5-ch3) are decoded through the 3×8 decoder (DEC1) to obtain 8 bits, which are then converted to an analog voltage by the first digital-to-analog convertor (DAC1) which controls the discharging load of all the 15 CCs. In other words, the bits (ch5-ch3) control eight levels of discharging speeds. In other words, the discharge load is challenged by the output of the R-2R ladder (DAC1) of L bits. In the designed PUF, L= 3 bits and analog output voltage of DAC1 controls the discharging speed by controlling the gates of 15x n-MOSFET within the (challenge-controlled discharge load) block, as shown in Fig. 4.1. The n-MOSFET transistors in this discharge path array are of minimum feature size (45nm × 120nm) to seek the maximum possible intrinsic variation during the manufacturing process. However, each discharge load n-MOSFET transistor can be of a different size to promote the unpredictability of the discharge speeds for the 15 subdivisions of the PUF. In our simulation, the focus was to study the effect of slight variation in a single n-MOSFET capacitor, so our transistors were of a minimum size and equal.

The M-field in Fig. 4.1 is composed of the challenge bits (ch2-ch0). These bits are applied to the second decoder (DEC2). The obtained 8 bits out of DEC2 are converted to an analog voltage by DAC2 to generate an equal reference voltage for the 15 comparators.

It is to be noted that among other factors that influence all those 15 chunks of response bits are the working environment of the PUF, such as the temperature variation. The temperature factor would affect all counters in the same direction, either up or down. The 16-bit reference counter and its related CC (n-MOSFET pair each of 45nm × 120nm) have been chosen in such a limited size in order not to have a roulette behavior. The non-rouletted 16 bits will convey any environmental effects to the verifier side. Based on their position within the response word, the verifier’s software can refer each chunk of 8 bits to their sub-division within an RC and correlate them to the corresponding CRP from the DB.
Using p-MOSFET capacitors, contributes to cryptography confusion as well, as a challenging bit of (0 V) enables charging the challenged capacitor. In other words, negative logic is used in this challenge bits’ subset, while K and M subsets operate their ADCs according to positive logic. The K-set of the challenging bits is converted from digital to analog to control the discharging load. The M-set of the challenging bits is converted from digital to analog to have an analog reference voltage for the comparator to compare the discharged voltage at the K-controlled load.

Having two ADC stages is an attempt to immune the proposed PUF against modeling attacks, which usually apply various artificial intelligence techniques to model PUFs. A comparator will generate an output of (V\textsubscript{DD}) as long as the challenged capacitor discharging voltage at the K-controlled load is more than the K-controlled reference voltage. As long as the comparator output is high (i.e. V\textsubscript{DD}) then the next counter counts the system clock pulses. V\textsubscript{DD}, GND, and the system clock pulse would be gotten from the smart card reader. Each counter would generate different chunk of bits according to the K-controlled discharge time, on condition that the discharge voltage value is over, or equal the M-controlled reference voltage. Having the comparator-counter pair which acts as DAC adds another hurdle against modeling attacks; besides, it generates a subset of the PUF response. Since each capacitor-challenging bit turned out to be a set of bits out of its counter, that adds a ‘diffusion’ concept of cryptography [226]. The total PUF response is the total counted bits of all counters, i.e. 128-bit. Any variation in temperature is expected to affect each response subset linearly. That would be easily noticed by the verifier error-correcting algorithm. A similar discussion can be made about (V\textsubscript{DD}). The environmental variations during the authentication phase cause temporary errors, which are considered systematic and not random. During the authentication phase, environmental variations can make a PUF chip generates RCs with different strings (shifted values) than the RCs archived during the enrollment phase.
The most dominant sources of environmental variation are due to the variations in the supplied voltage \(V_{DD}\), and the temperature of the PUF chip between the enrolment and authentication phases.

### 4.5. Combining Various Temporary Errors

In previous PUFs, where bits were generated individually, the common approach to compare between any two responses of the same PUF measured under different environmental conditions is the intra-Hamming between the nominal response which is measured at nominal conditions \((\alpha_1)\) and the verification phase chunks under any other different conditions \((\alpha_2)\). For the new PUF, the response bits are grouped in chunks to construct a meaningful binary value and any variation in temperature or/and voltage or any other systematic influences which influences the whole PUF chip during the verification phase it can shift up/down the response chunks depending on the magnitude and the sign of each factor, and the size of the CC. For example, as discussed earlier, the counted values which represent the discharge times are expected to decrease with the increment of the temperature, while increasing the voltage will have an opposite effect on the counted values.

### 4.6. The Signed Error Rate

In the C-PUF, it is possible to measure any shift rate in the chunk value in decimal, and such value is called here the signed error rate (SER). During the enrolment phase, the DB should keep the reference responses as chunks representing the actual number of clock pulses as if there is no roulette effect. The proposed verification approach does not deal with the CCs from the electrical point of view, but as a discharge time rate that varies with the size of the capacitor, and for simplicity, the shifting error ratio (SER) would be considered the same for all other non-rouletted response chunks (NRCs) of all counters.
4.7. The Verification Approach

The proposed remote verification approach to verify the C-PUF can be briefed as follows:

1) The SER can be determined from the reference response chunk RRC, which is resulted from discharging CC-A through counter-A. Noting that, only the noisy reference chunk \( (Y_{i,a2,A}) \) is a non-rouletted response chunk \( (NRC_{i,a2,A}) \), which is also considered as the reference response chunk at \( \alpha_2 \) environment \( (RRC_{i,a2,A}) \).

\[
SER = \frac{(Y_{i,a2,A} - R_{i,a1,A})}{R_{i,a1,A}} = \frac{(Y_{i,a2,A} - RRC_{i,a1})}{RRC_{i,a1,A}}
\]

Where \( R_{i,a1,A} \) is archived at the DB. The VS would be set to tolerate an error \((\pm \Delta R)\), to compare the creditability of the calculated SER, it should be within an acceptable margin of tolerance, as:

\[
|SER| \leq |\pm \Delta R|
\]

If the above condition has not been fulfilled, then the verifier aborts the verification phase. Assuming a similarity in the environmental influence on all RCs, the SER is then can approximately be considered the same for all NRCs.

2) An SER is calculated by comparing the received response \( Y_{i,a2,A} \) to the archived \( R_{i,a1,A} \), which also represents the non-rouletted response chunk \( (NRC-A) \) or \( (RRC_{i,a1}) \).

3) The SER enables the VS to calculate the expected non-rouletted count \( (ENRC_{i,a2,B}) \) at the environment \( \alpha_2 \) based on the archived non-rouletted response count \( (NRC_{i,a1,B}) \) at the environment \( \alpha_1 \):

\[
SER = \frac{(EN_{i,a2,B} - NRC_{i,a1,B})}{NRC_{i,a1,B}}
\]
\[ ENY_{i,\alpha 2, B} = SER \times NRC_{i,\alpha 1, B} + NRC_{i,\alpha 1, B} \]

The term \((SER \times NRC_{i,\alpha 1, B})\) whether it was positive or negative it would be smaller than \(NRC_{i,\alpha 1, B}\), therefore \(ENY_{i,\alpha 2, B}\) would always be positive.

4) The expected noisy response chunk of counter-B would be \(EY_{i,\alpha 2, B}\), and it can be calculated as:

\[ EY_{i,\alpha 2, B} = \left\lfloor \text{MOD}((ENY_{i,\alpha 2, B}), 255) \right\rfloor \]

Where the 255 is the maximum number that can be obtained from an 8-bit counter.

5) The VS can compare the determined \(EY_{i,\alpha 2, B}\) to the received \(Y_{i,\alpha 2, B}\) and should regards the preset margins for

\[ |SER \times EY_{i,\alpha 2, B}| \leq |\Delta R \times Y_{i,\alpha 2, B}| \]

It is also possible that each RC would have different \(\Delta R\) value archived at the DB.

4.8. Experimental Settings

The purpose of the experiments performed on the proposed C-PUF chip is to determine and compare the important properties as realistically and as accurately as possible. Despite the minor differences among researchers when setting the PUF experimental parameter settings, most ASIC-based research works consider the room temperature 27°C. The functionality of the proposed PUF is to be tested within the industrial temperature range \((-25^\circ C - 125^\circ C)\). The supplied voltage \((V_{DD})\) to be varied within \(\mp 20\%\) of its standard value.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Minimum</th>
<th>Nominal</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>(-25^\circ C)</td>
<td>27°C</td>
<td>125°C</td>
</tr>
<tr>
<td>(V_{DD})</td>
<td>(-20%)</td>
<td>1 V</td>
<td>(\mp 20%)</td>
</tr>
</tbody>
</table>
Table 4.1 summarizes the experimental parameters ranges of the proposed PUF. For this work, since the (Cadence /Virtuoso/ gpdk library 045nm/ low power silicon devices) is used to design the PUF, then the $V_{DD}$ ideal voltage is set to (1 V). The designed was made using 10 capacitor challenging bits, with three challenge-bit for each of K, and M sets. The total response bits are 128-bit. Various simulation tests were performed on the C-PUF design. First, the effect of temperature variation was studied within the industrial range of temperatures (−25) to (125) °C. The other simulation test was to study the identifiability of the PUF when having two different capacitive cells. The two n-MOSFETs of the first cell are sized as (600nm × 600nm) each, and the other cell at 5% extra area that is (630 nm × 630 nm) for each n-MOSFET. Finally, the average power consumption of the PUF design was measured.
Chapter 5

Results and Discussions of the C-PUF

This chapter presents the resulted response chunk versus the size of a capacitive cell; this is to test the influence of the geometrical variations during the fabrication on the response. The chapter also presents the results of varying temperatures and supply voltage on discharge time of both the largest and the smallest capacitive cells of the C-PUF. The chapter also presents an approach that a VS can use to handle the shifts caused by environmental variations.

5.1. Identifiability (Inter-Distance)

The first performed test was to measure the identifiability of the C-PUF. It is important to remind that the variation among the fabricated chips is to influence all the fabricated devices within the PUF; however, to verify the PUF identifiability, the worst-case scenario was considered, that only one n-MOSFET capacitive cell was varied. Again, various intrinsic variations were modeled as a geometrical one, and it is assumed as a variation rate of 5%. Fig. 5.1 presents the results for this test. The controlling challenging bits (ch5-ch0) have 64 possible values; however, to have a clear graph, only the first 33 possibilities were presented. The Y-axis in Fig.5.1 represents the generated response at the counters in decimal, where each 8-bit counter acts as an 8-slot roulette. In other words, the shown amplitudes equal (MOD255 of the real discharge time representation). Fig. 5.1, shows that against each decimal number that represents the challenge bits (ch5-ch0) at the X-axis, there is a distinct difference between the generated response values (of the
two individual 8-bit roulette counters), which appears as an amplitude at the Y-axis. The blue line in Fig. 5.1 represent the decimal value of the 8-bit response of a capacitive cell that has a pair of n-MOSFET transistors each of (600nm × 600nm) area, and the orange line represents the response of the 5% larger cell of two transistors each of (630 nm × 630 nm) area. The entire design consumes an average power of 1.05 mW, which is mostly consumed through the 15 comparators. The circuit was run under an ambient temperature of 27°C to simulate the discharge time for a capacitive cell of two n-MOSFETs each of (630nm × 630nm) when its secondary n-MOSFET (NM4) is challenged with (0) through its p-MOSFET (PM0), then the average discharge time was around 30 µs. Assessing the in-class (inter-distance) identifiability of a PUF is usually counted in term of Hamming distance among the total response bits, i.e. 128 bits of samples of the fabricated chips; however, for more strict assessment, it was determined between only two RCs of 8 bits each.
The C-PUF has achieved a 50% Hamming distance at the longest discharge times. Then when (ch5-ch0) gets higher values, the Hamming distance becomes above 50%, which is attributed to the absence of the roulette effect in shorter discharge times. The overall average Hamming distance for the results of Fig. 5.1 was 57%, which indicates that under certain conditions, the design has potential identifiability for intrinsic variation rates, which are even below 5%. It is important here to emphasize that since the presented PUF is based on counters, then the correlation between adjacent bits can have different meanings and significances, as each represents part of a binary number of real value, rather than a random bit in a string. This approach can facilitate handling the environmental-driven drifts of responses.

5.2. Neutralizing Cross Talk and Other Noise Effects

In the proposed PUF, there is no bit propagating along the presented PUF circuit at all, as this PUF generates the bits from slow discharging capacitors, instead of propagating the bit signals. The only rapidly moving charge would be within the counter, and such noise has never been a source of error in a counter output; furthermore, the capacitive cells of the proposed PUF have slow discharge rates, which would cause low electrical noise. From this, it is concluded that the response of the proposed PUF is not susceptible to cross-talk errors.

5.3. The Temperature Variations

The C-PUF was tested for a minimum temperature of (−25° C), and a maximum of (125° C), and the nominal temperature value of (27° C). Fig 5.2 presents a MOSFET capacitor and the effect on its discharge time through a discharge load. The simulation was run to produce the reference 16-bit response for the longest discharge time scenario. That is, the related capacitive cell was challenged with logic (0) to enable the auxiliary n-MOSFETs capacitor as well.
For that case, there were two charged n-MOSFET transistors, each of the minimum feature size (Length=45nm, Width=120nm). The simulation result is shown in Fig. 5.2. The dotted black line represents the mentioned shorter discharge time of a capacitive cell. Fig. 3 illustrates the response of the regular capacitive cells at designed sized and 5% oversized. This variation in the area is assumed as a modeling approach to represent the various intrinsic variations, such as geometrical, doping, and gate insulator thickness (t_{ox}) or any other parameter that influences the semiconductor device.

Fig. 5.2. Temperature effect on discharge time of various capacitive cells
The blue-dashed line represents the response of a capacitive cell that contains two n-MOSFETs each of (600nm x 600nm) dimensions, and the orange solid line accounts for a capacitive cell that contains two n-MOSFETs each of (630nm x 630nm) dimensions. The result shows that rising the temperature decreases the stored charge within the capacitor; this result complies with Eq.2.18.1 to 2.18.4. The results in Fig. 5.2 were recorded for a maximum discharge time; this required:

- ch5-ch0 were (000000), which controls all discharged loads, and hence the discharge times.
- The secondary (challengeable) capacitor of the 45x120 nm² was charged (i.e. challenged with 0).

The discharge time of the parallel pair of the similar primary and secondary capacitors of that capacitive cell is shown in the dashed black line.

- The secondary (challengeable) capacitor of the (600x600) nm² is charged (i.e., challenged with 0). The discharge time of the parallel pair of the similar primary and secondary capacitors of that capacitive cell is shown in the dashed blue line.
- The secondary (challengeable) capacitor of the (630x630) nm² is charged (i.e., challenged with 0). The discharge time of the parallel pair of the similar primary and secondary capacitors of that capacitive cell is shown in the solid orange line.

The obvious linearity of discharge times against the temperature variation can help the verifier side to adapt its expectations of the drift in response value based on the non-overflowed value of the least significant response bits (R15-R0) which are generated from the capacitive cell which contains two capacitors each of (45x120) nm², and its discharge time is represented within Fig. 5.2 as the dashed black line. Noting that the equity of capacitor sizes of any pair of capacitors within any capacitive cells is not necessary, but it was chosen arbitrarily. Furthermore, it is possible to
increase the temperature range over 125°C by increasing the sizes of the capacitors within the smallest capacitive cell and increasing the related counter number of bits. It is clear from Fig. 5.2 that there is a correlation between the temperature and the discharge time of a capacitive cell.

### 5.4. The Voltage Variations

The C-PUF was tested under V\textsubscript{DD}, 80\% V\textsubscript{DD}, and 120\% V\textsubscript{DD}. Fig. 5.3 illustrates the discharge times of two n-MOSFETs, each of (600 × 600) nm\textsuperscript{2} versus various input challenges at the DEC1 and DEC2. The dashed black line represents the discharge time under V\textsubscript{DD} of 1V. The solid blue line represents discharge time under V\textsubscript{DD} of 0.8 V, and the solid red line represents discharge time under V\textsubscript{DD} of 1.2 V.

![Fig. 5.3 Voltage effect on discharge time of the largest capacitive cell](image-url)
Similarly, Fig. 5.4 shows the discharge times of the smallest capacitive cell that has two n-MOSFETs, each of \((45 \times 120) \text{ nm}^2\) under the same conditions mentioned above.

The upper solid-blue line represents the discharge time of the mentioned capacitive cell when the \(V_{DD}\) is 20% more than its regular value of 1V. The low solid-red line represents the discharge time of the mentioned capacitive cell when the \(V_{DD}\) is 20% less than its regular value of 1V. The least steep regions represent the effect of varying the comparator reference voltage; while, the steeper regions represent the effect of varying the discharge load. These steeper regions can be flattened.
by adjusting the values of the related digital-to-analog ladder’s resistors, and since that would increase the discharge time, it would require larger non-rouletting counter of let say 32-bit instead of the current 16-bit counter. The other effect of increasing the discharge time is the increase in the average time of response generation over the 30 μs average. The average of the error’s root mean square between the discharge times at 100% \( V_{DD} \) and 80% \( V_{DD} \) of the least steep regions in Fig 5.3 equals 7.185 %, while the average of the error root mean square between the discharge time at 100% \( V_{DD} \) and 120% \( V_{DD} \) is 4.764 %.

5.5. The C-PUF Compatibility

The compatibility aspects of the new PUF is discussed here in terms of area, data processing time, and the consumed power of the chip.

5.5.1. Area

It refers here to the PUF chip area, and there are generally two kinds of a design area; the first is the actual area which depends on the semiconductor technology and number of bits, while the other is the absolute area that is suitable to compare each design among others, fairly, without being drift by the number of bits nor the semiconductor scaling technology. A PUF chip size varies according to:

- The number of the processed bits, which refers to (the maximum number of bits between the challenge string number of bits and response string number of bits, if they are not equal). In the proposed PUF, the number of challenge string bits is (21-bit), while the number of response string bits is (128-bit), so the PUF is considered to have a size of (128-bit).

- The semiconductor technology, as the minimum feature size \( F \) is considered equal to the minimum channel length \( L_{min} \) of the used semiconductor technology. \( F^2 \) represents the minimum feature area [50]. For the 45 nm process, it is then:
\[ F^2 = (45 \text{ nm})^2 = 2.025 \times 10^{-15} \text{ m}^2 = 2.025 \text{ fm}^2 \]

To approach a fair comparison with any other PUF size, regardless of its number of bits, and its semiconductor scaling technology; after measuring the PUF chip area, it is divided over \( F^2 \).

The proposed PUF has an area of \((3,276 \mu\text{m}^2)\), then the PUF area per bit is:

\[
\text{PUF area in } F^2 \text{ units} = \frac{\text{PUF area}}{\text{Feature area}} = \frac{3276 \text{ pm}^2}{2.025 \text{ fm}^2} = 1.6177 \times 10^6 \text{ } F^2 = 1.61777 \text{ MF}^2
\]

Then the PUF area in \((F^2)\) units per maximum number of processed bits (here it is the response bits):

\[
\text{PUF area in } \left(\frac{F^2}{\text{bit}}\right) \text{ units} = \frac{\text{PUF area in } F^2 \text{ units}}{N_{\text{res}}} = \frac{1.61777 \text{ M } F^2}{128 \text{ bit}} = 12638.82812 \text{ } \left(\frac{F^2}{\text{bit}}\right)
\]

So, when comparing with other PUF designs, it is fair comparison when comparing using \(( \frac{F^2}{\text{bit}} )\) scale of each PUF.

It is still possible to compare the area of a chip among different PUFs regardless of the \( F^2 \) standard, only when both designs use the same feature size. For example, In [227], the authors simulated a digital PUF design, which was also based on 45 nm MOSFET technology. That design works with an 8-bit challenge word to generate an 8-bits output. The design in [227] has been laid in an area of 36 \( \mu\text{m} \) by 50 \( \mu\text{m} \). When counting the area over the number of (challenge or response) bits, the result is \((225 \text{ } \mu\text{m}^2/ \text{bit})\). For our proposed design, the input is a 21-bit challenge to generate a 128-bit response. The layout dimensions of our design are 42 \( \mu\text{m} \times 78 \mu\text{m} \), the simulated chip area gives \((156 \text{ } \mu\text{m}^2/ \text{ challenge bit})\), and when considering the 128-bit response bits the area per bit will be \((25.6 \text{ } \mu\text{m}^2/ \text{bit})\) only.
5.5.2. Time

There are two aspects of time here:

1) **Response Chunk Generation Time**

Whether the response bits are generated simultaneously (in parallel) or sequentially (in series), in both cases, the speed is measured in term of the time interval between starting to challenge the PUF until the last bit of response string reached to the PUF output stage. In most previously known PUF designs, that time is considered constant, and if there is any variation due to ambient or aging variations, it would not be so noticeable. However, in the proposed PUF design, the generation time is more variable, it depends on the ambient, aging, beside the submitted challenge at most. For the proposed PUF the average running time is 30 us. Noting that the difference in the used semiconductor technology (especially lowering the threshold voltage) may contribute to this speed; however, to make the comparison possible between PUFs, such technology-wise speed differences can be ignored. Furthermore, many PUF chips carry also a hash, encrypt/ decrypt, and sometimes error-correcting stages; while, others don’t. Therefore, when comparing speed with other PUFs, if the experimental results give such delay details of these stages, then such non PUF times can be excluded; otherwise, they are included in the overall response generation time.

2) **Challenge/ Response Data Read/Transmit Time**

Whether the challenge string bits are submitted to the PUF in serial or parallel; and whether the responses are generated in serial or parallel, in both cases the reading and transmission of challenge string bits and response string bits are done serially. For example, if a smart card’s microcontroller read-out frequency is 1 MHz, and there are 128 bits to
read, thus the readout time is \( t = \frac{128}{1 \text{ MHz}} = 128 \, \mu\text{sec} \). Also, this time is supposed to be excluded from the comparison between PUFs, as it depends on the number of bits and the read-out frequency of the microcontroller and does not depend on the PUF element. Similarly, the transmission time is supposed to be dealt with, and since most of the previous experimental results don’t include read/ transmit times; therefore, such times are already excluded from the cited literature and do not interfere with time comparisons of the PUFs’ performances.

5.5.3. Power

The power consumption is an important factor, especially for passive-RFID applications, where an electromagnetic scanner is supposed to operate the PUF wirelessly. The induced voltage through the RFID antenna is supposed to run the PUF circuit. While if the PUF is designed for smart cards’ applications, the power factor becomes less critical. For the proposed PUF, the average power of \((1.05 \text{ mW})\), and to have a fair comparison, it is essential to evaluate the power per bits[50].

\[
\text{PUF Average Power per bit} = \frac{\text{PUF Average Power}}{N_{\text{res}}} = \frac{1.05 \text{ mW}}{128} = 8.203 \, \mu\text{W/ bit}
\]

Since some PUFs need more operation time to generate their response string than others, then evaluating the power does not reflect the exhausted energy over time, and the comparison between different PUFs cannot be fair, unless considering the running time of PUFs, in other words, evaluating their exhausted energies. For example, the average running time for the proposed PUF to generate its response string is \((30 \, \mu\text{s})\), then:

\[
\text{PUF Energy per bit} = \text{PUF Power per bit} \times \text{PUF Running Time} = 8.203 \, \mu\text{W/ bit} \times (30 \, \mu\text{s}) = 246.09 \, \text{pJ/ bit}
\]
In the C-PUF design, the input is a 21-bit challenge word, and the output is a 128-bit response word. The layout dimensions of this design are 42 µm by 78 µm which give (156 µm²/ bit) when considering the 21 challenge bits, while when considering 128-bit response bits, the area per bit is then (25.6 µm²/ bit) only. The entire circuit consumes an average power of 1.05 mW. The details of the power consumption are shown in Table 5.1.

The C-PUF design has an innovative error indication technique to eliminate the need for an on-chip ECC unit. The design also proposed an expanding concept in the relation between the number of challenge and response bits to decrease the design area. The proposed C-PUF design was published in [228]. The proposed C-PUF proved compatibility in terms of area; however, a 64-bit response may not be high enough for some applications. The C-PUF has shown high sensitivity to thermal variations, especially below \(-25^\circ\) C. For compatibility, the C-PUF design requires multi voltages (1, 0.5, 0.35, and 0.25) V. Practically, each voltage may vary independently, which can complicate the authentication and any further experiments on the design.

Since the RCs of this design can also be treated as meaningful decimal numbers, then the distances among chunks can also be determined in the form of signed decimal numbers rather than as Hamming distance among random bits.
The SER for the CCs indicates the correlations among the RCs, and these correlations can be analyzed to determine the effects of the environmental variations among the RCs. The SER authentication approach may require a high rate of tolerance, which can come along with a high FAR. This tolerance and 21-bit challenge-word limitations can suit some applications, but other applications may require less FAR and more than $2^{21}$ CRPs. The C-PUF design can be adapted to the application requirements and the operational environments.

### Table 5.1: The C-PUF Power consumption details

<table>
<thead>
<tr>
<th>Voltage Signal</th>
<th>Supplied Circuits</th>
<th>Vol. (V)</th>
<th>Power (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital $V_{DD}$</td>
<td>Counters+ NAND $\times 15$+DEC1</td>
<td>0.5</td>
<td>9.180E-07</td>
</tr>
<tr>
<td>Clk</td>
<td>Counters (16-bit $\times 1$+ 8-bit $\times 14$)</td>
<td>0.5</td>
<td>1.476E-08</td>
</tr>
<tr>
<td>Reset</td>
<td>Counters (16-bit $\times 1$+ 8-bit $\times 14$)</td>
<td>0.5</td>
<td>2.847E-08</td>
</tr>
<tr>
<td>Enable</td>
<td>Counters (16-bit $\times 1$+ 8-bit $\times 14$)</td>
<td>0.5</td>
<td>6.805E-08</td>
</tr>
<tr>
<td>$V_{DD_Ref}$</td>
<td>DAC2</td>
<td></td>
<td>1.786E-06</td>
</tr>
<tr>
<td>$V_{DD_L}$</td>
<td>DAC1</td>
<td>0.25</td>
<td>2.435E-12</td>
</tr>
<tr>
<td>Ch2-Ch0</td>
<td>Comp. Reference Challenge</td>
<td>1</td>
<td>3.480E-11</td>
</tr>
<tr>
<td>Ch5-Ch3</td>
<td>Load Ctrl Average Challenge</td>
<td>0.5</td>
<td>3.256E-12</td>
</tr>
<tr>
<td>Dis_Trig</td>
<td>Discharge Trigger</td>
<td>1</td>
<td>5.954E-11</td>
</tr>
<tr>
<td>$V_{DD}$</td>
<td>Comparator$\times 15$+ DEC2</td>
<td>1</td>
<td>1.045E-03</td>
</tr>
<tr>
<td>Ch20-Ch6</td>
<td>Logic (1) → Enables Capacitor Cell $\times 7$</td>
<td>0.5</td>
<td>7.935E-12</td>
</tr>
<tr>
<td>Ch20-Ch6</td>
<td>Logic (0) → Enables Capacitor Cell $\times 8$</td>
<td>0</td>
<td>2.954E-11</td>
</tr>
<tr>
<td>Charge</td>
<td>Disabled Cells$\times 7$ + Enabled Cells$\times 8$</td>
<td>1</td>
<td>741E-09</td>
</tr>
</tbody>
</table>
This chapter introduces the enhanced capacitive physically unclonable function (EC-PUF) system and schematics. It discusses the EC-PUF design, protocol, and security. The chapter explains the EC-PUF’s error handling and the CRP expansion approaches. The chapter also describes the applied variation, measuring, and stability enhancements.

6.1. CRP Expansion

A class of PUF chips can be challenged by one 64-bit challenge since each chip generates a unique response, which eventually creates a unique CRP for each chip. Only the number of response bits can limit the population of a class. The output response word of the proposed EC-PUF consists of 128 bits, and hence a difference of one bit permits a population of about 3.4E+38 chips. The CRP expansion aims to balance constraints of area, power, population, and security. Increasing the total challenge bits to 64 in the EC-PUF, instead of 21-bit in the C-PUF, gives the EC-PUF larger set of challenge bits. This leads to; 1) longer string for the challenge word, which enhances security; 2) this gives a VS more CRP alternatives, which also promotes security; 3) since it is more secure throughout a PUF’s lifetime to use a CRP once only, then a larger set of CRPs can support an extended period for a PUF as well as for its attached product, such as a smart card or an RFID tag.

Increasing the total challenge bits to 64 in the EC-PUF, instead of 21-bit in the C-PUF [228],

6.2. CRP Enhancement

enhances the security by two aspects; longer challenge word string and more possible CRP choices.
Furthermore, since it is safer to use each CRP only once throughout a PUF’s lifetime, then extending the number of the possible CRPs allows more authentications. This can support an extended lifetime for a PUF chip as well as for its attached product, such as a smart card or an RFID tag.

6.3. EC-PUF Schematics

The EC-PUF is a mixed PUF signal based on challenging groups of networked capacitors. Each group is referred to as a capacitive cell (CC). In the simulated design, the EC-PUF has eight CCs, as shown in Fig. 6.1. A time-to-digital converter (TDC) digitizes the discharge time of each CC into a 16-bit RC. Concatenating all RCs yields a total response of 128 bit. The uniqueness of the response is attributed to the uniqueness of the intrinsic variations throughout the fabrication process of the EC-PUF chip, especially of the analog units within the chip.

The proposed EC-PUF comprises eight challenge-controlled CCs, eight comparators, eight 16-bit counters, one oscillator, one Schmitt trigger, and eight frequency dividers. It can accommodate an optional substitution box (S-box), as shown in Fig. 6.1. The EC-PUF also includes two DACs, which are DAC1 and DAC2, each of eight challenge bits referred to as M and N, respectively. The challenge-controlled reference voltage is controlled by DAC1, while the challenge-controlled discharge load is controlled by DAC2.

Each CC consists of six sections of several parallelly networked n-MOSFET capacitors. Each section is to be challenged by a challenge-bit in the field (ch63-ch16). Fig. 6.2 illustrates a simplified CC model with only one section that has only one n-MOSFET capacitor. The CCs were constructed with n-MOSFETs to reduce the silicon area since an n-MOSFET has a greater charge per unit area than a p-MOSFET. The primary n-MOSFET capacitor (NM0) is charged each time an authentication phase is initiated; this enables a counter to generate a non-zero RC, even when every challenge (ch) bit within the capacitive challenge bits group (L’) is set to a high voltage (typically 1 V), which
Fig. 6.1. Enhanced capacitive physically unclonable function block diagram

Fig. 6.2. Capacitive cell module

Fig. 6.3. MOSFET-based R-2R ladder
represents a logic 1.

The charging operation of the auxiliary n-MOSFET capacitor (NM1) is controlled by the ch bit. When ch is of a high voltage, and the charging control signal is applied at the gate of the main switch p-MOSFET (PM0), this makes the switch p-MOSFET (PM2) pass the charging current toward NM1.

To store larger charges in each capacitive section, NM0 and NM1 were extended parallelly with other n-MOSFETs. However, sizes and numbers of the n-MOSFETs within each section are random, to maintain unpredictability of the discharge times of each capacitive section enabled by a ch bit within an L’ group of bits. The drains of all n-MOSFETs capacitors of all capacitive sections within a CC are joined into one terminal whose voltage is the capacitive cell’s voltage (V\text{cap}), as shown in Fig. 6.2. The electric path of each V\text{cap} terminal toward the discharge load is controlled by a pass (or transmission) gate as a discharging control switch. Each of the eight pass gates consists of two parallelly connected MOSFETs; one p-MOSFET to better pass the high voltage around the drain voltage (V\text{DD}) value, and one n-MOSFET to better pass the low voltage around the source voltage (V\text{SS}) value.

The authentication phase starts when a low control signal of 0 V is applied at the gate of PM0. This control signal is generated from a timer unit within a microcontroller embedded within the smart card’s chip. The width of the low-level pulse controls the charging time of the eight CCs and resets the eight counters. For our design, the minimum time required to fully charge the CCs in the proposed design is about 1 ns. It is not critical if the microcontroller is not capable of generating such a narrow pulse, as a wider charging time will not change the value of the stored charge (Q) of a CC.

The discharge phase starts when a high-level control signal of 1 V is applied by the microcontroller unit. This high-level signal shuts down PM0 of the eight CCs to stop the charging
phase, enables the eight 16-bit counters and simultaneously turns on the eight pass gates to connect each CC to its discharge load. Each discharge load is merely an n-MOSFET with a grounded source terminal.

When the control signal turns the eight pass gates on, this applies $V_{\text{cap}}$ of each of the eight CCs at its related discharge load n-MOSFETs. In this case, the discharging voltage across the drain-to-source ($V_{\text{dis}}$) equals $V_{\text{cap}}$ at the drain of the discharge load n-MOSFET.

The gates of all eight n-MOSFETs are driven by one discharge controlling voltage ($V_{\text{Dis-Ctrl}}$), as shown in Fig. 6.1, $V_{\text{Dis-Ctrl}}$ is the analog output voltage (Analog) of DAC2, as shown in Fig. 6.3. The eight challenge bits (ch15-ch8) control 256 levels of $V_{\text{Dis-Ctrl}}$, which drives the gates of the eight discharge-load n-MOSFETs. Alternatively, the analog voltage of DAC1 represents the reference voltage $V_{\text{ref}}$ for all comparators, as shown in Fig. 6.1. In other words, the challenge bits (ch15-ch0) control 256 levels of $V_{\text{ref}}$ at the eight comparators.

The voltage across each discharge load n-MOSFET that is $V_{\text{dis}}$ drives a non-inverting input of a comparator. Each comparator gives an output voltage equals to $V_{\text{DD}}$ when the applied $V_{\text{dis}}$ at its non-inverting input is greater or equal to $V_{\text{ref}}$. Each comparator’s output represents an enabling signal (E) that controls a 16-bit counter. The schematic in Fig. 6.3 illustrates DAC1/DAC2, where each is an 8-bit DAC and constructed as a MOSFET-based R-2R ladder. Each DAC was designed to output low analog voltages, especially DAC2, which drives all eight discharge loads. Each discharge load is implemented here as an n-MOSFET that functions within the subthreshold margin to maintain a discharging time that is large enough to be measured by digitizing it through a 16-bit counter that is clocked by a high frequency. The final digital value that appears at a counter is an RC, which does not necessarily represent the discharge time, as a large discharge time and a high clock frequency collaborate toward overflowing the related counter. The actual digital representation for a discharge
time of a CC is referred to here as the non-rouletted response chunk (NRC). For a 16-bit counter, the
decimal representation of an RC is the remainder after dividing NRC by \(2^{16}\). Such modular (MOD)
operation is represented here as \(RC = \text{MOD}_{2^{16}}(NRC)\). Among the eight CCs, there is one CC with
small charge storage capacity, whose discharge time is digitized by a relatively lower frequency to
prevent its related 16-bit counter from overflowing. This CC is called the reference capacitive cell,
and its non-rouletted response is called the reference response chunk (RRC).

The discharge load n-MOSFETs were sized similarly in this design, although they can be scaled
differently. Nevertheless, since these n-MOSFETs are scaled at minimum feature size, the intrinsic
mismatch variations of the fabrication process would still contribute to the individuality of each RC
locally within the same chip and globally among the similar EC-PUF chips. The eight comparators
were also scaled identically, but they were designed, laid out, and to be masked according to
conventional anti-variation approaches, yet some variation is also expected to contribute to the
individuality of the response.

A supply voltage of analog and mixed-signal chips is usually controlled by a bandgap reference
(BGR) regulating circuit, which can sustain the voltage within a variation window of a few
millivolts. The proposed design did not include a voltage regulator, as it is assumed to be pre-existing
among the other system-on-chip units; however, a BGR circuit was only used for the comparators.
This enables the simulation to emphasize the effect of the voltage variation on the CCs and the DAC
units, which ultimately shifts the response value up or down.

6.4. Basic Authentication Protocol

Beside authentication purposes, most PUFs also use the response as an ID and/or cryptographic
key [144], [32], [33]. Combining the authentication with the identification/cryptography eliminates
error-tolerability, and hence an on-chip ECC unit becomes essential. Instead, the basic protocol of
the EC-PUF requires to retrieve the ID information from an on-chip NVM, as most of today’s smart cards and RFID tags, but embedding an on-chip PUF makes the ID data security less crucial. The basic verification phase protocol of the EC-PUF can be summarized in three main steps:

1) The smart card/tag reader sends the ID to the verifier side. According to that ID, the VS randomly picks a 64-bit challenge and sends it to the EC-PUF. These challenges were measured and archived during the enrollment phase.

2) The PUF receives a 64-bit input challenge and generates a 128-bit output response. The PUF sends the response to the verifier side without correcting it on-chip.

3) The VS correlates the received RCs to the archived ones. To analyze the environmental-driven shifts within the NRCs’ values, the VS would analyze the RRC’s shift to evaluate the shifting trend of all NRCs. It can also apply an extrapolation approach to better achieve its verification task. It may also include an artificial intelligence structure to achieve its verification analysis, which can also help to handle the aging-driven shifts within the NRCs. Ultimately, if the authenticator approves the authenticity of the EC-PUF chip, it allows the smart card to proceed to the next requested step.
6.5. The EC-PUF Model

The major influential parameters on each RC within the EC-PUF design are listed in Table 6.1.

Table 6.1. Acronyms and values of the parameters of each CC

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Details</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Number of bits of each counter</td>
<td>16 bits</td>
</tr>
<tr>
<td>f</td>
<td>Clock pulse frequency for a counter</td>
<td>varies</td>
</tr>
<tr>
<td>E</td>
<td>Comparator output</td>
<td>0/1 V</td>
</tr>
<tr>
<td>K</td>
<td>Total challenge bits</td>
<td>64 bits</td>
</tr>
<tr>
<td>L</td>
<td>Total challenge bits for C</td>
<td>48 bits</td>
</tr>
<tr>
<td>L'</td>
<td>Challenge bits for each capacitive cell (CC)</td>
<td>6 bits</td>
</tr>
<tr>
<td>M</td>
<td>Challenge bits for discharge control</td>
<td>8 bits</td>
</tr>
<tr>
<td>N</td>
<td>Challenge bits for reference voltage</td>
<td>8 bits</td>
</tr>
<tr>
<td>O</td>
<td>Organization of within CC</td>
<td>varies</td>
</tr>
<tr>
<td>P</td>
<td># of CCs, discharge loads, comparators, and counters</td>
<td>8 units</td>
</tr>
<tr>
<td>Q</td>
<td>The stored charge of a CC</td>
<td>varies</td>
</tr>
<tr>
<td>R</td>
<td>Total response bits of 8 counters</td>
<td>128 bits</td>
</tr>
<tr>
<td>R'</td>
<td># of bits of each counter</td>
<td>16 bits</td>
</tr>
<tr>
<td>S</td>
<td>Size of the charged capacitors in a CC</td>
<td>varies</td>
</tr>
<tr>
<td>T</td>
<td>Temperature</td>
<td>-55°C to 125°C</td>
</tr>
<tr>
<td>V</td>
<td>Supplied voltage (VDD)</td>
<td>1 V to 1.2 V</td>
</tr>
<tr>
<td>v</td>
<td>Intrinsic variations of a CC</td>
<td>varies</td>
</tr>
<tr>
<td>v'</td>
<td>Intrinsic variations of DAC1&amp;2</td>
<td>varies</td>
</tr>
<tr>
<td>v''</td>
<td>Intrinsic variations of discharge load</td>
<td>varies</td>
</tr>
</tbody>
</table>

Fig. 6.4. Data dependency of a response chunk.
Fig. 6.4 shows the basic parameters which control an RC. The analog units of the EC-PUF are drawn in red. The upper red block represents the factors that $Q$ of a CC depends on, such as; 1) the CC related challenge bits ($L'$), where each bit decides which capacitive section to enable to be charged and which to leave disconnected; 2) the total size of the charged capacitors of all enabled sections ($S$) within each CC; 3) the structure organization ($O$) of each charged capacitive cell, although our basic capacitive cell consists only of capacitive devices networked parallely, a cell can include other components such as inductors and resistors, and can be networked in various ways; 4) the intrinsic variation of each component within a CC, denoted here as ($v$); 5) the supplied voltage ($V$); and 6) the temperature ($T$). The total charge can then be depicted as $Q (L', S, O, v, V, T)$.

The lower red block in Fig. 6.4 refers to the parameters which affect the period of the enable signal ($E$), where a high voltage level of 1 V represents an active $E$ signal. A related counter is active by $E$ and counts the applied clock pulses for the period of active $E$. Since the model in Fig. 6.4 refers to a single CC, the lower red block includes a single discharge load n-MOSFET, the output analog voltage of DAC1($v'$), the output analog voltage of DAC1($v''$), the supply voltage $V$, and the temperature $T$. Ultimately, $E$ can be represented as a function $E (Q, M, N, v', v'', V, T)$. The intrinsic variation of the comparator is another parameter that can affect $E$. However, the generic comparators such as the eight used in our EC-PUF are less affected by the intrinsic variations. Furthermore, the BGR that regulates the voltage for all comparators minimizes the effect of the voltage variation on the comparator.

The blue blocks in Fig. 6.4 refer to the digital portions of the EC-PUF model. Unlike the analog portions, the intrinsic variations usually do not alter the digital data processed with these units. The right blue block represents the parameters that impact an RC. It includes the period of counting activation $E$, the number of the bits of the counter ($C$), the clock pulse frequency applied at the
counter after dividing the central frequency $f$ by a divisor $D$. The RC of a CC is then represented as $R(E, C, f/D)$.

The left blue block in Fig. 6.4 illustrates the possibility of adding an S-box to add more confusion against model attacks. Furthermore, it is also possible to insert an authentication key at the S-box to add another level of security. The key in Fig. 6.4 can be provided from the authenticator side either as a permanent security code within an on-chip NVM (which represents a security threat) or through a secure communication channel as a short-term or even as a one-time password [229], [230]. Embedding an S-box is optional, and its structure can go from a few basic gates to plenty of various combinatorial logic blocks. It can be constructed in such a way that a chunk of the challenge bits ($L'$) at one CC can influence ($L'$) of other CCs.

In standard advanced encryption standard (AES), the S-box only applies the confusion concept of cryptography [226]. In our proposed EC-PUF design, an optional S-box can be added to serve both confusion and diffusion concepts. The diffusion within the EC-PUF can be applied when the design wires some of the challenge bits ($L'$) of some or all CC as input to the S-box. For the EC-PUF, since the complexity of the S-box can vary widely depending on the application requirements, and since adding an S-box here is optional, the S-box is not included within the area and the power measurements.

6.6. The EC-PUF Frequency Aspects

Many microcontrollers operate at low internal frequencies. For example, the LPC8N04 microcontroller has a maximum internal frequency of 8 MHz [231]; then, its timer can control the EC-PUF with a minimum period of 125 ns. In other words, although the EC-PUF’s capacitors require only 1 ns to be fully charged, the charging time will be 125 ns. This is not a problem for the EC-PUF; instead, this makes it compatible with most of today’s microcontrollers despite their internal
frequencies. After the charging control signal rises, the microcontroller can count a number of its internal clock pulses as waiting time and then read the 128-bit response, without the need for handshaking of control signals nor synchronization. The waiting time depends on the environmental conditions and corners of the chip. For the proposed chip, a waiting time of 750 μs is sufficient.

The clock pulse frequency here represents the sampling rate of the time-to-digital conversion. The higher a counter’s frequency, the more precise the time-to-digital conversion would be. This ultimately enhances the ability to detect the distinct properties of each EC-PUF chip. The proposed EC-PUF includes an oscillator to run the counters at higher frequencies. High frequency is also preferred to roulette the seven (non-reference) counters as part of the confusing aspect.

On the other hand, there are limitations on frequency, as high frequency costs extra dynamic power. Besides, there is limitation of the propagation delay throughout each synchronous counter. This delay primarily depends on MOSFET technology, design, manufacturing, and environmental variations. Furthermore, the reference counter, which is affiliated with the smallest CC to generate the (non-rouletted) reference response, will also be overflowed at a greater frequency.

The maximum frequency that can run the 16-bit reference counter without rouletting was measured under these conditions; 1) the six input challenge bits which control the reference CC were all set to 0 V to achieve the maximum charge; 2) the eight input challenge bits which control the discharge loads were all set to 1 V to achieve the largest resistive load; 3) the eight input challenge bits which control the reference voltage at all comparators were all set to 1 V to achieve the largest time for the enabling signal E; 4) the simulation temperature was set at a minimum of −55º C; and 5) the supply voltage (VDD) was set at a minimum value of 1 V. Under those conditions, the discharge time of the reference CC was 29.38703 μs. Then, for the reference counter to count 65,535 without rouletting, its clock should not exceed 2.23 GHz. The initial frequency (f) of the oscillator at those
conditions was 3.95 GHz. This frequency is central to all counters.

Our EC-PUF design includes a frequency division unit that contains eight digital frequency dividers. The central frequency is divided by 1, 2, 3, 4, 5, 6, 7, or 8 divisors. Some divisors can be selected by the challenge, either directly (as implemented here) or through the S-box. For example, one of the challenge bits in this design opts one of the divisors to be 3 or 2. The intrinsic and environmental variations within the oscillator and the Schmitt trigger add randomness and uniqueness for the EC-PUF chip; even though the Schmitt trigger per se does not alter the frequency, its loading effect on the oscillator does.

A digital frequency divider preserves a precise frequency division, which both the intrinsic and the environmental variations cannot alter. This helps the authorized VS to handle the environmental-driven effects that are supposed to shift the eight RCs.

In our EC-PUF, the clock frequency that runs the 16-bit counter of the smallest CC (CC-A) is divided by 3 to avoid overflowing the related counter (counter-A). The low clock frequency \( f_A = f/3 \) and the short discharge time \( t_A \) make the related 16-bit counter (counter-A) counts less and avoids overflowing. Alternatively, in our EC-PUF, the central frequency \( f \) directly runs counter-B, which is related to the largest CC (CC-B), which has a large discharge time \( t_B \). The high clock frequency \( f_B = f/1 \) and the large discharge time \( t_B \) make counter-B counts for an extended time to digitize the discharge time \( t_B \) caused by discharging CC-B and the digital representation for \( t_B \) equals NRC-B.

The microcontroller reads the final count and sent it to the VS, that is RC-B, where \( RC-B = \text{MOD}_{2^{16}}(NRC-B) \).

If the enrollment phase of the EC-PUF is done with a central frequency \( f \), then the response ratio (RR) of \( (NRC-B/RRC) \), that is denoted here as \( (RR-B) \), and determined as:
Response Ratio \( \left( \frac{B}{A} \right) = \frac{t_B}{t_A} \times \frac{f_B}{f_A} = 3 \frac{t_B}{t_A} \) 

From equation (6.1), the variation of the frequency between the enrollment and the verification phases does not alter the response rates of a larger CC to the reference CC. This is a key concept toward using the response rates of the seven CCs, with respect to the reference CC as helper data stored within the remote database along with the 16-bit RC of each CC. This concept enables the design to eliminate the on-chip ECC unit.

6.7. Error Handling

During the authentication phase, PUFs generate responses with some inevitable error rate. Some errors are attributed to the temporary variation of the environment, such as temperature, voltage, or frequency. Others can be attributed to the slow, irreversible, and lasting change in the chip's physical properties, usually known as aging. To handle this, an on-chip ECC is usually embedded, but since it is insecure, eliminating it is one of the objectives of the proposed EC-PUF.

The proposed EC-PUF exploits the correlations among the individual RCs to handle the environmental-driven shifts of the RCs’ values and ultimately to authenticate the EC-PUF chip. Two main features in the proposed EC-PUF design support the VS to handle the environmental-driven shifts in the generated RCs:

- Within the analog portion, there is a central M-bit set that controls all discharge loads, and a central N-bit set that controls the reference voltage \( V_{\text{ref}} \) of all comparators.
- Within the digital portion, there is a central \( f \) for all counters, although it is divided by various divisors since they are known to the VS.
Among the 128-bit response, the total effect of the environmental effects is conveyed to the remote VS by the 16-bit RRC. The only specialty about the RRC is that its CC is relatively small, and its counter’s frequency should be limited to avoid rouletting its 16-bit counter. Alternatively, there are no limitations to the CC size nor the counter’s frequency regarding the other seven 16-bit RCs. A VS is supposed to analyze each RC with a degree of tolerance, which eliminates the need for an on-chip ECC unit.

The VS can reconfigure its tolerance margins to tolerate the environmental effects on the EC-PUF chip. The tolerance margins control FAR and FRR of the authentication phase. The proposed EC-PUF not only eliminates on-chip ECC but also gives the authorized VS flexibility to handle errors.

In the EC-PUF schematic, variation-aware aspects are applied, based on the type and the scale of some components:

- The C-PUF in [228] has square-shaped n-MOSFET capacitors to save the area, whereas the EC-PUF has short-wide n-MOSFET capacitors with a minimum channel length to support attaining more intrinsic variation throughout the fabrication process.

- The two DAC ladders introduced in [228] were based on diffusion N-well resistors. Instead, the EC-PUF ladders were designed using MOSFETs in the subthreshold region as resistors. Such ladders can cause more random variations and save area. Each of these MOSFET-based resistors has a long, narrow channel. The non-square MOSFETs in the two DAC ladders achieve a high electrical resistivity with less physical size and more intrinsic variation.

6.8. Fabrication Variation Enhancement
6.9. Other Applicable Fabrication Variation Enhancements

At layout and masking levels, causing fabrication variations requires the layout and masking engineer to act exactly in the opposite direction against any variation-limiting approaches. For example, PUF masks are better with PUF-aware OPC applied to improve their uniqueness [132], [232]. This approach can be applied wherever a variation is required in a portion of the circuit, whereas a regular OPC can normally be applied elsewhere.

Choosing a fabrication technology with a high rate of intrinsic variations can help acquire distinct PUF chips. Besides, it is sometimes possible to influence certain fabrication stages to achieve more variations as a kind of PUF-aware fabrication.

6.10. Measurement Enhancement

Most generic microcontrollers adopt low-power approaches, including lowering the clock pulse of the internal oscillator. Our first C-PUF [228] relied on the internal frequency of the microcontroller, assuming a frequency of 100 MHz. This frequency represented the sampling rate of the TDC units. Alternatively, the EC-PUF includes a high-frequency oscillator for higher sampling rates, which enhances the precision of the time-to-digital conversion to better indicate tinier intrinsic variations in the form of RCs. Increasing the total response bits to 128 in the EC-PUF instead of 64-bit in the C-PUF raises the overall resolution of the PUF system.

6.11. The EC-PUF Stability Aspects

The stability-enhancing is discussed here in terms of thermal aspects and transient aspects.

6.11.1. Thermal Aspects

The magnitudes of the drains’ currents vary with temperature due to variations of the threshold voltage and the mobilities of the charged carriers. It is well known that both the threshold voltage and the mobility are inversely proportional to the temperature. In linear and saturation modes,
increasing the threshold voltage increases the drain current; decreasing the mobility decreases the current. At lower $V_{GS}$, the variation in $V_T$ dominates; therefore, increasing the temperature increases the drain current. At higher $V_{GS}$, mobility dominates; therefore, increasing the temperature decreases the drain current. At some $V_{GS}$, both effects cancel each other out, and the drain current does not change with temperature. As a rule of thumb, when $V_{GS}$ is much less than $V_T$, the temperature rises the drain current, whereas when $(V_{GS} = V_{DD})$, the temperature reduces the drain current.

6.11.2. Transient Aspects

Furthermore, when working in a subthreshold mode, a subthreshold drain current of p-MOSFET is very much less affected by temperature variations than n-MOSFET. On the other hand, at saturation mode, the drain current of a p-MOSFET is slightly more affected than in n-MOSFET [233]. In the proposed EC-PUF, most MOSFETs within both R-2R ladders (DAC1 and DAC2) operate in subthreshold mode; therefore, the ladders’ implementation was mostly based on p-MOSFETs, for more thermal stability compared to n-MOSFETs; however, to implement better pull-down circuit, the foot of each ladder is an n-MOSFET, as shown in Fig. 6.3.

The DAC ladders are to be initiated before the charge/discharge controlling pulse, to assure stability when the discharge begins. The DAC ladders do not have any switching activity through the entire challenging time of the EC-PUF circuit. Therefore, using slow resistive MOSFETs does not set back the DAC performance. The pulse width time of the charge/discharge control signal depends on the microcontroller’s internal clock frequency. On the other hand, for the MOSFET capacitors, even the time of 2 ns was enough to have the capacitors fully charged; however, most microcontrollers already run at much slower pulse rates. This makes the EC-PUF maintain its stability with most generic microcontrollers.
6.12. Security Aspects

Similar chips are referred to as a class. They can be challenged by one 64-bit challenge, as the unique response of each chip is what matters most. However, for security measures, each CRP is safer to be used once only, to thwart MITM attacks. At each authentication attempt, the authorized authenticator randomly chooses a challenge among the archived CRP list, so an attacker cannot predict the upcoming challenge. In addition, the EC-PUF can have an S-box, as shown in Fig. 6.1 and Fig. 6.4. This optional S-box serves both confusion and diffusion purposes at the same time. Furthermore, a key can either be inserted by the smart card reader’s keypad, stored in on-chip NVM, or received from the verifier side. It is also possible to relate that key to the encrypting key of the communication with the verifier side. A brute-force modeling attack should overcome the environmental variations of voltage, temperature, and frequency, besides the large response possibilities of $2^{128}$. The communications between the PUF and the microcontroller are internal, and the communications with the verifier side over the internet are secured as well. Exploring the internal structure of the targeted chip, whether by basic optical attacks or by x-rays, is a vital step in planning a suitable approach to attack the targeted chip. However, knowing the internal structure still does not ensure accurate modeling due to the different doping rates and other complex factors of each MOSFET within the EC-PUF design. Besides, in general, it is difficult to model the environmental effects on mixed-signal PUFs. It is worth mentioning that to the best of the authors’ knowledge, no successful attack on any mixed-signal PUF was reported until the writing time.

For the EC-PUF, the difficulties a DPA attack faces start first with the short charging phase, about 1 ns; second, all CCs withdraw their charging currents simultaneously. Although the total current is correlated to the challenge-activated capacitors, but their distribution among the CCs is not known to an outsider unless the attacker can probe the individual current of each CC. This
hardens the DPA attacks unless achieving direct contacts for the microprobes to each CC. The physical access to the internal details of the EC-PUF chip implies an invasive attack, which is not feasible against the suggested active coating packaging for the EC-PUF chip.

The best bet for an attacker would be applying EM probing, measuring the eight different frequencies and the counters’ stop times to reveal the NRC of all counters. However, among the 128 flip-flops, only 16 work simultaneously as a 16-bit counter, and some counters may meet at the same rising edge at different times. This involves determining the physical wiring of each response bit, which can be more difficult if a designer arbitrarily wires the 128 response bits to the microcontroller. The VS is to be programmed to sort out the proper position of the received response bit among the eight RCs. The PUF-to-microcontroller data bus is internal and to be protected by an active coating; therefore, if an attacker tries to indicate the stop time of each counter, this requires removing the active coating and possibly some higher metal layers, which will destroy the EC-PUF system.

6.13. Experimental Setup

This section presents the tests which were performed on the proposed EC-PUF design. It also discusses the basic verification approaches for the EC-PUF. The simulations mostly considered the discharge time and RC of the smallest and largest CCs. Table 6.2 lists the related objective and metrics of each test. The table also refers to the results section of each test. Two categories of variation tests were carried out; the environmental and the fabrication variations.
Table 6.2. Test Framework

<table>
<thead>
<tr>
<th>Test</th>
<th>Objective</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature variation</td>
<td>Measuring the thermal effect on the response</td>
<td>Discharge time and digital response</td>
</tr>
<tr>
<td>Voltage variation</td>
<td>Measuring the voltage effect on the response</td>
<td>Discharge time and digital response</td>
</tr>
<tr>
<td>Monte Carlo simulation–local</td>
<td>Measuring the effect of local intrinsic variations on the discharge time of the capacitive cells (CCs)</td>
<td>Discharge time’s coefficient of variation</td>
</tr>
<tr>
<td>Monte Carlo simulation–global</td>
<td>Measuring the effect of global intrinsic variations on the discharge time of the CCs</td>
<td>Discharge time’s coefficient of variation</td>
</tr>
<tr>
<td>Monte Carlo simulation–total</td>
<td>Measuring the collective effect of the local and the global intrinsic variations on the discharge time</td>
<td>Discharge time’s coefficient of variation</td>
</tr>
</tbody>
</table>

6.14. Simulating the Environmental Variation

The EC-PUF chip is intended to be embedded in smart cards and other portable devices that would work under various environmental conditions; therefore, the effects of the variations in temperature and voltage were tested. The RCs of the smallest and the largest CCs were compared as part of the ECC elimination approach. The carried out environmental variation tests are categorized as follows:

1) **Thermal Variation Tests:** The thermal variation effect on the discharge times and the RCs of the smallest and largest CCs were simulated in the range (−55° to 125°) C.

2) **Voltage Variation Tests:** The effect of the supply voltage variation on the discharge times and the RCs of the smallest and the largest CCs was simulated in the range (1-1.2) V.
3) Thermal and Voltage Variation Tests: The collective effect of both thermal and voltage variations was measured to study the worst error case. This was to attain the maximum tolerance rate, which the VS should adopt to maintain a suitable FAR and FRR.

6.15. Simulating the Fabrication Variation

Monte Carlo simulations were conducted to test the effect of the intrinsic randomness of the manufacturing process on the uniqueness of EC-PUF chips. That is by measuring the variances of their discharge times, which ultimately reflect on the EC-PUF chips’ responses. This research adopted the coefficient of variation (CV) of the discharge time as a metric for the uniqueness of the EC-PUF chip. The first Monte Carlo simulation was to measure the effect of the global, inter-die, or process variation among the EC-PUF chips within a silicon wafer on the discharge times of the CCs. The second simulation of Monte Carlo also added the local, intra-die, or mismatch variation among devices within each EC-PUF chip. This was to assess the effect of the total variations on the discharge times.

In both simulations, the discharge times of the smallest and largest CCs were measured to evaluate the influence of a CC’s size on the uniqueness of the discharge time.
Chapter 7

Results and Discussions of the EC-PUF

This chapter discusses the area of the layout and the average consumed power. The section also
discusses the effect of simulations of temperature, voltage, and intrinsic variation on the time of
discharge and the responses generated.

It would be erroneously assumed that a lower supply voltage would store less charge in the
capacitors; therefore, a shorter discharge time would be expected. The influence of the supply
voltage on the two DAC R-2R converters is more significant therefore supplying a lower supply
voltage leads to longer discharge times for all eight CCs.

The longest discharge time was attained at a minimum temperature of $-55^\circ \text{C}$. The discharge
load controlling bits ch15–ch8 were set to a high voltage of 1V as logic 1s. Similarly, the
comparator’s controlling bits ch7-ch0 were set to logic 1s; all the capacitors within that CC were
enabled and charged by a minimum $V_{\text{DD}}$ of 1V. Under typical corners’ conditions for the MOSFETs,
the average data processing time was 118 $\mu$S. To determine the maximum time which a
microcontroller should wait before reading the RCs, the longest discharge time was determined using
Monte Carlo simulation for the total intrinsic variations of 100 chip samples. The simulation showed
that the largest CC required 715 $\mu$s to discharge, and this is the required time to generate its RC. It
would, therefore, be safe to program a microcontroller timer to wait for about 750 $\mu$s before reading
the 128-bit response.
7.1. Environmental Variation Simulation Results

The proposed design processes a 64-bit input challenge to generate a 128-bit output response. The design area is 22,470 µm², and it consumes an average power of 921.67 µW. The analog portion consumed an average power of 131.5 µW; this includes the oscillator, Schmitt trigger, two DACs, and eight comparators. Charging all CCs consumed an average power of 67 nW, with a peak instantaneous current of 385.5 µA. The digital units, which include the frequency dividers, clock gating, and counters, consumed 792.4 µW. The results and discussion are divided into environmental and intrinsic variations.

7.1.1. Thermal Variation Simulation Results

It is essential to study the effects of the environmental variations on the proposed EC-PUF chip. Under the environmental variations, the simulation results are divided into thermal and voltage variations. To study the EC-PUF at its worst-case scenario of thermal variations, the basic EC-PUF design did not include any frequency stabilization. Fig. 7.1 shows that the central frequency (f) varies inversely to the temperature. However, it is assumed here that the supplied voltage is regulated to 1V at both the enrollment and the verification phases.

![Central frequency vs temperature](image-url)

**Fig.7.1. Central frequency vs temperature**
Fig. 7.2. shows the temperature variation effect on the discharge time and the decimal representation of the RRC. It is clear from Fig. 7.2 that the time-to-digital conversion produces an RC that is linear to the discharge time. It also shows that the discharge time varies reversely to the temperature; however, the response value does not overflow counter-A.

![Graph showing discharge time and response vs. temperature](image)

**Fig. 7.2. Discharge time and response of cell-A vs. temperature**
On the other hand, Fig. 7.3 shows that the excess of $t_B$ overflows counter-B; therefore, it acts as a digital roulette, and the value of $R\_B$ varies even against the same applied $L'$ challenge chunk at $CC\_B$.

![Fig. 7.3. Discharge time and response of cell B vs. temperature](image-url)
Fig. 7.4 shows the ratio of the discharge times’ ratio (t_B/t_A).

![Graph showing the discharge time ratio of cell-B to cell-A vs. temperature](image1)

**Fig. 7.4. Discharge time ratio of cell-B to cell-A vs. temperature**

Fig. 7.5 shows the ratio of the non-rouletted value of RC-B to RC-A, that is (NRC-B/RRC), which is denoted as (RR-B), both vary with the temperature.

![Graph showing the response ratio of cell-B to cell-A vs. temperature](image2)

**Fig. 7.5. Response ratio of cell-B to cell-A vs. temperature**

The equation of the response ratio is given by:

\[
y = 3E-07x^3 - 7E-05x^2 + 0.0087x + 12.782
\]
One of the reasons behind such inconsistency is that the generic comparator used in this design has an overdrive of 50 mV with a response time of 38 ns. This delay enables the related counter to count for an extended time, and with such a high frequency, this makes a significant difference in the counted value. Each different CC creates a different input impedance, which also varies with temperature, and impacts the transition slope from 1 V to 0 V of each comparator’s output. Each output represents an enable signal E for its related counter. With such high frequencies, even a slight change in the slope can impact the NRC of the counter. The worst-case scenario was studied with the largest CC, which is CC-B, which at certain conditions, generates the largest NRC. Dividing this NRC by the RRC yields the largest RR, which is related to the smallest CC, which is labeled here as CC-A. All the bits of N challenge field were set to 1s to have the least V_{ref}, this is to achieve the maximum counted values related to t_A and t_B. Through these tests, it has been decided to store the maximum possible charge in both CC-A and CC-B, which was done by setting each challenge bit in L’ field (for both CCs) to 0. Each challenge bit in M field was set to ones to make all the discharge loads (implemented by n-MOSFETs) at their maximum impedance. When the discharge phase is initiated by the charging/discharging control signal, the impedance of each CC becomes in parallel to its related n-MOSFET discharge load. The total impedance of each CC with its parallel discharge n-MOSFET affects the input impedance of its related comparator. Having a larger n-MOSFET impedance gives the impedance of the CC more influence. Furthermore, since each CC consists of tens of parallel n-MOSFETs, this makes a CC have less impedance and more susceptibility to the thermal effects. This explains the variations in the ratio of discharge times (t_B/t_A) in Fig. 7.4 and RR-B in Fig. 7.5 when the temperature varies. However, the RR variation shrinks against other challenges that cause shorter discharge times and when the size difference between the CCs is less, as such conditions reduce the discharge time ratios, hence the RR.
When the C-PUF design was proposed, it came with the SER verification approach. Here, two more precise verification approaches are proposed. The VS can apply any of those three approaches to verify the authenticity of a CC, like CC-B, despite the thermal-driven shifts in its RC-B and NRC-B values.

**A. Ratio-Based Verification**

The enrollment phase was simulated at 25° C and VDD at 1 V for the challenge that causes the maximum discharge time. Counter-B generated a binary NRC-B, which in decimal equals 197375. The authorized enroller knows that counter-B counted 197375 and shows a final count of 767. Alternatively, counter-A generated a 16-bit binary RRC, which equals 15232 in decimal, and since it is smaller than 65536, it did not roulette counter-A. In other words, counter-A did not apply a MOD function to the base 65536. Then RR-B is 12.9579.

During the enrollment phase, the seven RCs, NRCs, and their RRs beside the RRC are to be archived in a safe server where the enrollment data are only accessible by the authorized authenticator or a trusted third party, depending on the authentication protocol.

At the verification phase, under unknown temperature, the authenticator can assume that all the received RCs were shifted similarly, trend-wise, but not quantity-wise, and all RCs are generated under similar but not necessarily identical, environmental factors.

As an example of the verification phase at a different temperature, the temperature was set to −10° C. The worst-case scenario was considered, where the microcontroller does not convey the temperature to the verifier side. Counter-A gave 22090. Then the authenticator expects NRC-B to be around \( \lceil 12.9579 \times 22090 \rceil \), that is 286240. After applying the MOD function to base 65536, this yields 24096. In the simulation, RC-B at counter-B was 18161.
The difference between the expected and the generated values can be considered for that CRP as a tolerance margin for NRC-B. The frequency was not stabilized; at 25° C, it was 3.09684 GHz, while at −10° C was 3.59,985 GHz; however, this should not affect the response rate, as the frequency is eliminated as in equation (6.1). A more extreme case was simulated under −55° C, counter-A had 43775, then the authenticator expects NRC-B of approximately $\lceil 12.9579 \times 43775 \rceil$, that is 567232, whereas counter-B had an RC-B of only 526675.

**B. Ratio-Based Verification Guided by Individual Trends**

If the VS finds the shifting trends among the eight RCs consistent, such shifts can be attributed to natural causes rather than security threats. This comparison helps the VS to decide whether to accept or reject the entire 128-bit response. The consistency among the eight RCs also played a key role in eliminating the on-chip ECC unit.

Using the same RR of an NRC at different temperatures can save both enrollment and verification phases’ time, cost, and computational resources. However, it requires from the VS to adopt wider margins of error tolerance. This raises the FAR and lowers the FRR, which may still be acceptable depending on the authentication application, especially since analyzing the general shift trend among all NRCs plays a major role in the authentication process.

As RRs, such as RR-B, vary with temperature, as shown in Fig. 7.5, a more precise approach requires archiving the individual trend of the RR and the individual trend of each RC at different temperatures. This implies:

- At the enrollment phase, each challenge is applied at several temperatures; this creates a set of different RRs for each CC. These data sets of all CCs are archived at the server. This enrollment phase is slow and requires controlling the thermal condition. It is possible to reduce the required number of measurements by choosing the most dominant temperatures
for the targeted application. For example, for banking smart cards in Europe and North America, the trend of the RCs through temperatures $-25, 0, 25^\circ C$ can make a good fit. The trend equation ($y$) of the RRC and NRC of each of the other seven chunks against the temperature ($x$) are driven and archived at the server.

Fig. 7.6 shows the trends of chunk-A (RRC) and the non-rouletted value of chunk-B (NRC-B) against the temperature variation.

![Graph showing discharge times trends of the smallest and the largest CCs vs. temperature.](image)

**Fig. 7.6.** Discharge times trends of the smallest and the largest CCs vs. temperature.

The number of the measured temperature points was ($N = 37$). When relating the NRC-B data set to the RRC data set, this makes the degree of freedom (DF) equal to ($N - 2 = 35$). Pearson’s correlation factor ($r$) was determined from the related data and is usually represented as $r(35) = 0.9998$. This correlation value represents a largely positive relationship between the two sets of responses. The statistical correlation coefficient, or the probability value ($P$), had a value of $5.8998E-63$. This extremely small $P$ indicates a significant statistical correlation between the two response sets.
• At the verification phase, the VS can utilize the archived trends. Starting with the received value of chunk-A, the VS substitutes the RRC value into $y$ to find the temperature $x$. If RRC is out of that archived range of responses, that means the temperature at the EC-PUF is out of the archived range. In such cases, the VS can apply the extrapolation approach based on the closest archived RRC value to estimate the temperature $x$.

• Since all CCs are assumed to have the same temperature, the VS can substitute the calculated value of $x$ into the trends of the other chunks, like chunk-B, to calculate $y$ which represents NRC-B. The VS applies the MOD function to the base 65536 to find RC-B, to compare this estimated value to the received one. Similarly, this process applies to all seven rouletting RCs.

• The VS can analyze the shift trends of all eight RCs to check their consistency, then ultimately decide whether to accept or reject the entire 128-bit response.

• The VS can also substitute the determined value of $x$ into the trend equation, as shown in Fig. 10, to determine the RR at this temperature. If the value of $x$ is out of the archived range, the VS can apply the extrapolation approach based on the closest archived $x$ value to estimate the RR. This way, RR is determined at that specific temperature, which makes it more accurate. The VS can then apply the ratio-based verification approach explained in (a), but this time, the difference diminishes between the determined RC and the received one. This enables the VS to set finer margins of error tolerance and become more definite about the authenticity of the EC-PUF chip.
7.1.2. Voltage Variation Simulation Results

Voltage variation is another vital environmental variation factor that influences all the RCs. An increment in the supplied $V_{DD}$ increases the oscillator’s frequency, as shown in Fig. 7.7.

![Graph: Central frequency vs. $V_{DD}$](image)

**Fig. 7.7. Central frequency vs. $V_{DD}$**

The frequency in Fig. 7.7. represents the central frequency, which drives all frequency dividers. The enroller does not need to measure the frequency, as it is eliminated when determining the RR between the RR (chunk-A) and each of the other seven RCs.
Fig. 7.8 shows that the discharge time varies inversely to the voltage; however, the response value does not overflow counter-A.

---

**Fig. 7.8. Discharge time and response chunk of cell-A vs. V_{DD}**
On the other hand, Fig. 7.9 shows that excess of $t_B$ overflows counter-B; therefore, it acts as a digital roulette.

![Fig. 7.9. Discharge time and response of cell-B vs. $V_{DD}$](image-url)
Ideally, both the rate of discharge times \((t_B / t_A)\) in Fig. 7.10 and the rate of responses \((NRC-B / BRC)\) in Fig. 11 should be constant; however, both vary inversely to the voltage.

\[
y = 35.625x^3 - 154x^2 + 229.24x - 97.978
\]
The output slope for each comparator varies with the voltage, the CC size, and the structure of the networked capacitors within each CC. Increasing the supplied voltage significantly affects the eight discharge loads and the reference voltage at the comparator inputs, which are centrally controlled by DAC1 for all eight comparators. In comparison to the other analog components of the EC-PUF, the comparators are less affected by voltage variation, as their supplied voltage is regulated by a central BGR circuit.

The VS can handle the voltage variation effects similar to the temperature variation. However, a large voltage variation is not such an inevitable factor, as precise voltage regulation is possible with circuits such as the BGR, which can regulate the supplied voltage with an error of just a few microvolts. Voltage regulation is mainly essential for the analog portion of the EC-PUF chip. On the other hand, the digital units such as the S-box, the eight AND gates, and the eight counters do not require such a tight voltage regulation.

The worst-case was studied considering the voltage supplied to the oscillator, the Schmitt trigger, the CCs, and the DAC ladders is not regulated. As a worst-case scenario for the voltage variation of 100 mV, the verification phase was simulated at a voltage of 1.1 V; cell-A generated 4525 counts. In this case, the VS presumes counter-B should have an NRC-B of \[12.9579 \times 4525\], that is 58635 counts. Since it is less than 65536, the VS does not need to apply MOD function, or even if it does, it also gets 58635 counts. Through simulation, counter-B gave an NRC-B of 69035 counts. This difference of 10400 counts can be tolerated by the VS. A more extreme voltage variation of 200 mV was simulated as well; at 1.2 V counter-A had 1415 counts. Then the VS presumes that counter-B should have NRC-B of \[12.9579 \times 1426\], that is 18478 counts. The simulation showed that counter-B gave an NRC-B of 24001 counts. This 5623 difference can be considered within the tolerance margin for NRC-B. As shown in Fig. 7.11 that increasing the voltage makes the discharge
times of all CCs become shorter. This was the worst-case scenario for the voltage variation, where the voltage varies by 200 mV, the largest NRC was studied. In that case, CC-B was challenged with L’ which causes the largest discharge time, as well as M and N fields.

Although the simulated EC-PUF design can function at the voltage range (1–1.2) V, it was optimized for 1 V and assumed the enrollment and the verification phases to be done in the range (1–1.1) V. This means the VS can set narrower tolerance margins for the NRCs than in the worst-case scenario studied above. The correlation between the response rates under various voltages is shown in Fig. 7.12.

The number of the measured voltage points was (11). When relating NRC-B data set to the RRC data set, this makes DF equal to (9), and r(9)= 0.9989. This correlation value represents a largely positive relationship between the two sets of responses. It was also found that P has a value of 2.47E-13. This extremely small P indicates a significant statistical correlation between the two response sets.

7.12. Discharge times and trends of the smallest and the largest CCs vs. $V_{DD}$
The total shift in the value of any NRC depends on both temperature and voltage shifts. It is clear from Fig. 7.5 and Fig. 7.10 that temperature influences the discharge time oppositely to the voltage. The response, however, is also influenced by the frequency if it is not stabilized. The influence of the frequency is eliminated anyway, as in equation (6.1).

7.2. Fabrication Variations Simulation Results

Even when considering the worst-case scenario which is when neither the temperature nor the voltage is conveyed to the verifier side, and neither the voltage nor frequency is firmly stabilized, the proposed EC-PUF is still able to handle these variations, and the VS can authenticate the chip.

The intrinsic fabrication variations were simulated to estimate the randomness contribution to the individuality of each EC-PUF chip if fabricated by a typical MOSFET technology. 100 hundred chips were generated and tested using Monte Carlo simulation guided by the generic process design kit gpdk045 for oxide-polysilicon gated planar MOSFET with BSIM4-model. Throughout the simulation, a supply voltage of 1 V and a temperature of 25°C were considered. The intrinsic variation is usually divided into global and local variations. For the EC-PUF chip, Monte Carlo simulation was run for the global and total variations.

7.2.1. Local Variation Simulation Results

The effects of the local variations of the largest CC (cell-B) and the smallest CC (cell-A) were simulated. An input challenge was applied to both CCs. The resulted discharge time distributions of both CCs are shown in Fig. 7.13 and Fig. 7.14, respectively. The discharge time $t_B$ in Fig. 7.13 is to be digitized by a counter runs at a frequency $f/1$; similarly, the discharge time $t_A$ in Fig. 7.14 is to be sampled by a frequency $f/3$. For cell-B, Fig. 7.13 shows that the discharge time’s mean of 100 chip samples against an arbitrary challenge was 68.1642 µs, and the standard deviation was 17.8477 µs.
Fig. 7.13. Simulated effect of the local randomness on discharge time of the largest CC.

Fig. 7.14. Simulated effect of the local randomness on discharge time of the smallest CC.

Fig. 7.14 illustrates the discharge time distribution of cell-A; its mean was 17.255 µs, and its standard deviation was 4.95769 µs. When the standard deviation was compared to the mean of each CC, the largest CC had a CV of 26.18%, compared with 28.73% for the smallest CC.
7.2.2. Global Variation Simulation Results

The effects of the global variations of the largest CC (cell-B) and the smallest CC (cell-A) were simulated.

![Graph showing simulated effect of global randomness on discharge time of the largest CC](image1)

![Graph showing simulated effect of global randomness on discharge time of the smallest CC](image2)

Fig. 7.15. Simulated effect of the global randomness on discharge time of the largest CC.

Fig. 7.16. Simulated effect of the global randomness on discharge time of the smallest CC.

Fig. 7.15 shows that the discharge time’s mean of 100 chip samples against an arbitrary challenge was 73.5474 µs and the standard deviation was 43.9664 µs. Fig. 7.16 illustrates the discharge time distribution of CC-A; its mean was 18.1928 µs, and its standard deviation was 12.5404 µs. When the standard deviation was compared to the mean of each CC, the largest CC had a CV of 59.78%, compared with 68.93% for the smallest CC.
7.2.3. Total Fabrications Variation Simulation Results

The effect of the local variation was added to the global one to assess the effect of the total variation on the discharge time of 100 simulated EC-PUF chips against the same challenge.

Fig. 7.17 and Fig. 7.18 show the discharge time variation for the largest and the smallest CC, respectively.

![Fig. 7.17. Simulated effect of the total randomness on discharge time of the largest CC.](image1)

![Fig. 7.18. Simulated effect of the total randomness on discharge time of the smallest CC.](image2)

The randomness simulation exhibited the mean discharge time of the largest CC as 75.0486 µs, and its standard deviation was 54.7207 µs. The smaller CC yielded 20.2737 µs and 15.3738 µs, respectively. The collective anticipated CV of both the mismatch and the total variations is then 72.91% for the largest CC and 75.83% for the smallest CC. The high variation in discharge times
shows the validity of the proposed EC-PUF design, as applying a higher clock pulse frequency at a counter means a higher sample rate for the time-to-digital conversion, which can more precisely indicate the physical uniqueness of the analog units related to that RC. The variations of the discharge time comply with the theoretical concepts of the size effect of MOSFET-based objects, as the smallest CCs showed more CV. The CV values also exhibit the far greater impact of the global intrinsic variations compared to the local variations. This can help draw a road map to guide the variation-aware layout and manufacturing processes to further enhance the uniqueness of the EC-PUF chips.

The CV of a discharge time is a generic parameter of an RC uniqueness. The frequency of the clock pulse to each counter is not of concern here, as it is mostly high enough to indicate even the minute differences in discharge times, and in the end, the concatenation of the eight distinct 16-bit RCs results in a unique 128-bit response for each EC-PUF chip. This is how the continuous number counting and the roulette-effect raise the number of possible 128-bit response combinations, so as the difficulty bar against attackers. It is important to emphasize that there is no specific number for the combinations, as it depends on the central frequency and divisions, the applied challenge and voltage, and the PUF’s temperature.

In this dissertation, various design, layout, and manufacturing aspects have been discussed toward acquiring more intrinsic variations throughout a MOSFET-based manufacturing process, to form distinctive properties for a MOSFET-based PUF chip.

7.3. Discussions

The proposed EC-PUF is based on the discharge time delay of networked MOSFET-based capacitors. An independent high-frequency oscillator drives eight frequency dividers, and then each different frequency runs a counter. A higher frequency means a higher sampling rate, which makes
a more precise time-to-digital conversion. This helps the EC-PUF chips to generate distinct responses even when meager intrinsic differences are obtained throughout their fabrication process.

To maintain differences among the eight 16-bit RCs, the frequency is divided for eight different rates; furthermore, the CCs are configured differently at each submitted challenge.

The capacitive nature of the EC-PUF has pros and cons. The CCs are interactive to the metal comb arrays and the random post-process sprayed coating. This can inherently increase the intrinsic randomness and chip uniqueness, optically oppose the basic optical inspecting attacks, and electrically shield the chip against the invasive attacks. The CCs can be used for other purposes than at the authentication time, including security-based applications such as decoupling and current flattening to counter DPA attacks. On the other hand, the capacitive nature of the EC-PUF includes constraints, which require adopting a broad margin of tolerance at the verification phase.

A large CC and high clock pulse rate can cause an overflow in the related 16-bit counter, which then acts as a roulette wheel; this helps to maintain the unpredictability of the RCs against modeling attacks. The confusion and diffusion concepts of cryptography can be enhanced further by including a custom-made S-box to raise the security level against modeling attacks. The on-chip microcontroller encrypts the data using conventional security schemes. In this mixed-signal PUF, errors are not in the form of flipped bits, but in the form of shifts in discharge times of the CCs. Decrypting the chip-to-verifier communications does not directly lead the adversary to model the PUF, as the 128 bits of the counters are hard-wired in a secret sequence to the microcontroller; therefore, an invasive attack would still be needed to disclose the physical meaning of each bit. This approach enables the remote VS to control the acceptance/rejection tolerance margins of the responses. This can also help the VS to handle the environmental shifts. Furthermore, it helps to detect the non-uniform shifts among the RCs, which would be attributed to invasive attacks on
the PUF chip. The EC-PUF eliminates the need for an on-chip ECC unit.

The environmental parameters such as voltage and temperature can be read from on-chip sensors and sent to the remote verifier. However, even without measuring those parameters, based on a 16-bit RR, the proposed design enables the VS to analyze the collective shift caused by the environmental factors.

The design was simulated under temperatures in the range (−55° to 125°) C and supplied voltage in the range (1-1.2) V. The design was simulated using 45 nm CMOS technology with a chip area of 22,470 µm². The average response time was 118 µs, measured at 1 V, −55° C, and typical corners for both n and p MOSFETs. At a temperature of 25° C and a voltage of 1 V, the average power was 921.67 µW.

Lin et al. [227] presented a digital PUF using 45 nm CMOS technology. It has an input challenge word length of 8 bits, so as the response word. The dimensions of the design were 36 µm by 50 µm. When counting the area over the number of (challenge or response) bits, the result is (225 µm²/ bit).

The C-PUF [228] has a 21-bit challenge and it generates a 128-bit response. The layout dimensions

<table>
<thead>
<tr>
<th>Factor</th>
<th>Lin et al. [227]</th>
<th>C-PUF [228]</th>
<th>EC-PUF [234]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area (µm²)</td>
<td>1,800</td>
<td>3,276</td>
<td>22,470</td>
</tr>
<tr>
<td>CRP (bit)</td>
<td>8-8</td>
<td>21-128</td>
<td>64-128</td>
</tr>
<tr>
<td>Area/ bit (µm²/ bit)</td>
<td>225</td>
<td>156-25.6</td>
<td>351-176</td>
</tr>
<tr>
<td>Power (µW)</td>
<td>136.4 @1 GHz</td>
<td>1050 @ 100 MHz</td>
<td>921.67 @ 4 GHz</td>
</tr>
<tr>
<td>Power (µW/ bit)</td>
<td>17.05</td>
<td>50-8.2</td>
<td>14.4-7.2</td>
</tr>
</tbody>
</table>

Lin et al. [227] presented a digital PUF using 45 nm CMOS technology. It has an input challenge word length of 8 bits, so as the response word. The dimensions of the design were 36 µm by 50 µm. When counting the area over the number of (challenge or response) bits, the result is (225 µm²/ bit). The C-PUF [228] has a 21-bit challenge and it generates a 128-bit response. The layout dimensions
of our design were 42 µm by 78 µm, which gives (156 µm²/ challenge bit), and when considering
the 128-bit response bits, the area will be (25.6 µm²/ response bit) only, as in Table 7.1.

The EC-PUF [234] has a layout dimension of 149.899 µm by 149.899 µm, which makes (351
µm²/ challenge bit), and when considering the 128-bit response bits the area will be (176 µm²/
response bit). When comparing the C-PUF versus the EC-PUF, the later design has a larger area,
but it allows for 2⁶⁴ CRPs, instead of 2²¹ CRP for the C-PUF. The EC-PUF also achieves a more
precise time-to-digital conversion by having much higher pulse rates. The EC-PUF also has a more
optimized design, which reduces power consumption. It also reduced the sensitivity against the
environmental variations of temperature, as it and was able to function at low temperature down to
(−55°C) with compared to a minimum temperature of (−25°C) for the C-PUF.

Security-wise, the larger CC and a high-frequency clock pulse of the EC-PUF can cause more
overflow in the counters; therefore, it enhances the RCs unpredictability against simulation attacks.
In addition, cryptography principles of confusion and diffusion can be further improved by including
a custom-made S-box to increase protection against modeling attacks. From all the above, the EC-
PUF has presented better features as a capacitive PUF compared to the C-PUF.
Chapter 8

Conclusions and Future Work

This chapter presents the conclusion statements regarding the presented research work in this dissertation. Summaries of the two presented designs, differences, and contributions are briefed first, followed by suggestions for possible future developments.

8.1. Conclusions

In this dissertation, two versions of a mixed-signal capacitive physically unclonable function were proposed. Both designs depend on challenging MOSFET-based capacitors through digitizing their discharge times into RCs. Both designs are proposed for authentication purposes and not for identification, as they require a margin of tolerance to handle the environmental-based shifts in the generated RCs. Both designs adopted an innovative error indication technique to eliminate the need for on-chip error-correcting code (ECC) units. They also proposed an expanding concept regarding the number of challenge and response bits, which saves area and power, and it also adds difficulty against modeling attacks. The first design is called the capacitive physically unclonable function (C-PUF); it processes a 21-bit input challenge word to generate a 128-bit output response word. It has proved compatibility in terms of area; however, a 21-bit challenge limits the challenge-response pairs (CRPs) to $2^{21}$ despite having a 128-bit response. The second proposed version is called the enhanced capacitive physically unclonable function (EC-PUF) [234]; it processes 64-bit challenge word, which allows for $2^{64}$ possible CRPs.
The dissertation has discussed layout, masking, and fabrication sources of intrinsic variations of MOSFETs in general, and the planar type in particular. Various permanent, transitional, and temporary MOSFET variations were discussed. The C-PUF included few but large n-MOSFET capacitors, this approach reduces the design area, but it also reduces the margin of fabrication variations, which are the source of a PUF response uniqueness, where a CC is composed of hundreds of short-wide n-MOSFET capacitors. This approach allows for more intrinsic variations throughout the fabrication process. The C-PUF design depends on an external clock pulse, which can be provided by a microcontroller. Most microcontrollers run at a frequency that is much below 1 GHz. Such a low clock pulse rate for the counters in the C-PUF represents a low sampling rate for the time-to-digital conversion. Alternatively, the EC-PUF has an embedded oscillator that generates about 4 GHz square wave. A higher sampling rate makes a more precise time-to-digital conversion for the discharge time. This helps make an EC-PUF chip generate distinct RCs among other EC-PUF chips, even when only meager intrinsic differences are obtained throughout the fabrication process. In the EC-PUF design the central frequency generated by the oscillator is digitally divided over eight divisors, then the resulted frequencies individually run the eight counters; this adds confusion against timing attacks.

Security-wise, a large CC and a high-frequency clock pulse can cause an overflow in a counter, which then acts as a roulette wheel. That helps the unpredictability of an RC against the modeling attacks. Furthermore, the confusion and diffusion concepts of cryptography can be enhanced further by including a custom-made S-box to raise the security level against the modeling attacks. The on-chip microcontroller would still encrypt the communications to the remote verifier side using conventional security schemes. Decrypting the chip-to-verifier communications does not directly lead the adversary to model the PUF, as the 128 bits of the counters are hard-wired in a secret
sequence to the microcontroller; therefore, an invasive attack would still be needed to disclose the physical meaning of each bit.

The capacitive nature of both designs has pros and cons. The CCs are interactive to the metal comb arrays and the random post-process sprayed coating. That can inherently increase the intrinsic randomness and the chip uniqueness, optically oppose the basic optical inspecting attacks, and electrically shield the chip against the invasive attacks. Other than the verification phase time, the CCs can be utilized for other purposes, such as decoupling and current flattening applications to counter the DPA attacks. On the other hand, the capacitive nature of both designs causes environmental-based constrains, which imply that the VS should adopt a broad margin of tolerance at the verification phase. In this dissertation, various design, layout, and manufacturing aspects were discussed to acquire more intrinsic variations throughout a MOSFET-based manufacturing process to form distinctive properties for a MOSFET-based PUF chip.

The dissertation proposed three authentication approaches.

- The first approach is based on a signed-shift error rate (SER) rate concept, as the environmental effect would have similar (but not identical) effects on the RRC and the other RCs. The SER indicates the correlations among the RCs; these correlations can be analyzed to determine the effects of the environmental variations among the RCs. The SER authentication approach requires a rate of tolerance, which may mean a high false acceptance rate (FAR) as well.

- The second authentication approach is based on applying an extended enrollment phase by challenging the PUF under various temperatures and voltages and archiving the trends of each RC on the server. At the authentication phase, referring the received reference response chunk (RRC) to its archived trend can help the VS to compare the other received RCs
to their archived trends. Although the later authentication approach is more precise, it raises the cost of the enrollment phase.

- The third authentication approach aims to balance the precision with the enrollment cost, that is by measuring the RCs at the most dominant range for the application, such as archiving the trendlines in the thermal range -25°, 0°, and 25° C, as this range could be the most common thermal points for commercial smart cards and RFID-tags, then the VS can apply an extrapolation to estimate the RC at different temperatures. All the above approaches considered the worst-case scenario, where the temperature and the voltage are not conveyed to the VS, while most microcontrollers would already have such features, which can be extremely helpful for the VS to handle the effects of the environmental-based temporary variations that appear in the form of shifts in the generated RCs.

Monte Carlo randomness simulated the fabrication variations of 100 EC-PUF chip samples. The global variations have shown much more impact on the discharge time of a CC than the local variations. Using a high frequency to digitize the discharge time improves the time-to-digital converting accuracy. The fabrication variations within the embedded ring oscillator will add more uniqueness to the generated response.

The EC-PUF design was simulated under temperatures in the range (−55° to 125°) C and supplied voltages in the range (1-1.2) V. The design was simulated using 45 nm CMOS technology with a chip area of 22,470 µm². The average response time was 118 µs, measured at 1 V, 25° C, and typical corners for both n and p-MOSFETs. The average power was 921.67 µW. Based on the layout area, average power, and compatibility with a range of voltages, temperatures, and microcontrollers, the proposed designs are feasible for authentication and anti-counterfeiting purposes.
8.2. Future Work

As future research work, it is possible to fabricate prototypes for the C-PUF and the EC-PUF to physically test the effect of the environmental factors on their functionalities. Also, it would be possible to practically analyze the effect of the aging on the generated responses. Furthermore, power analysis attacks can be performed to study the immunity of the proposed systems.
References


J. S. Kim, M. Patel, H. Hassan, and O. Mutlu, “The DRAM latency PUF: quickly evaluating physical unclonable functions by exploiting the latency-reliability tradeoff in modern


Appendices

Appendix A

Average Hamming Distance

The Hamming distance \( (D_{1-2,i}^{\text{inter}}) \) between each possible pairing group of the tested PUFs within that \( (P_{\text{class}}) \) is first found as:

\[
D_{1-2,i}^{\text{inter}} \triangleq \text{dist} \{ PUF_1(X_i) ; PUF_2(X_i) \}
\]

or

\[
D_{1-2,i}^{\text{inter}} \triangleq \text{dist} \{ PUF_{1,i}^{\alpha} ; PUF_{2,i} \}
\]

or

\[
D_{1-2,i}^{\text{inter}} \triangleq || PUF_1(X_i) - PUF_2(X_i) ||
\]

, then \( D_{1-2} (X_i) \triangleq \text{dist} \{ Y_{1,i} ; Y_{2,i} \}
\]

or

\[
D_{1-2} (X_i) = || Y_{1,i} - Y_{2,i} ||
\]

The relation above should be repeated for each possible pair combination of tested PUF within \( (P_{\text{class}}) \) [49].

If the experiments are done under the same condition \( (\alpha) \), and \( \forall 1 \leq i_1 \neq i_2 \leq N_{\text{puf}} \); for different PUF individuals \( (i_1, i_2) \) within the PUF class, the minimum value for \( (N_{\text{puf}}) \) is 2, and if the experiments are repeated using different pair combination within the PUF class, then the value of \( (N_{\text{puf}}) \) increases by one each time a new \( (D_{1-2,i}^{\text{inter}}) \) is tested between two different PUF entities.

\( \forall 1 \leq i \leq N_{\text{cha}} \), and both measurements are for the same challenge input.
Every measurement here is done once only, then the number of the measured PUFs \((N_{\text{meas}})\) doesn’t affect the inter-distance \((D_{1-2,i}^{\text{inter}})\). The value of \((D_{1-2,i}^{\text{inter}})\) is determined by comparing the two (e.g. 128-bit) responses digit-by-digit then counting the flipped digits; however, in this dissertation \((Y_{1,i,\alpha}, \text{ and } Y_{2,i,\alpha})\) each can be represented as a decimal integer.

If there are \(N_{\text{puf}}\) individuals, and the response \(Y_{i,\alpha}\) was measured once for each PUF individual, and each experienced CRP is indicated by a subscript (i).

Among those \(N_{\text{puf}}\) objects, and it is required to determine Hamming distance between each two different PUF individuals \((D_{1-2,i}^{\text{inter}})\) against the same challenge, this means that the Hamming distance is determined for each pair of two \((L= 2\) elements\) within that class. The total number of all possible pairs \(N_{\text{pair}}\) within a class can be calculated as:

\[
N_{\text{pair}} = \frac{n!}{L! \cdot (n-L)!} = \frac{n \cdot (n-1) \cdot (n-2)!}{2! \cdot (n-2)!} = \frac{n \cdot (n-1)}{2} = \frac{(N_{\text{puf}})(N_{\text{puf}}-1)}{2}
\]

For example, if the noisy response \(Y_{i,\alpha}\) is measured once for each of four PUF individuals \((1, 2, 3, 4)\), and for the same challenge indexed by \((i =150)\) as \((X_{150})\), and the tests are done under some certain set of environmental conditions \((\alpha3)\), then the measured responses can be denoted as \(Y_{1,150,3}, Y_{2,150,3}, Y_{3,150,3}, Y_{4,150,3}\), and the number of non-repeated PUF combinations can be determined as:

\[
N_{\text{pair}} = \frac{(N_{\text{puf}})(N_{\text{puf}}-1)}{2} = \frac{(4)(4-1)}{2} = 6 \text{ pairs}
\]

, and the inter-distances are:

\[
D_{1-2,150}^{\text{inter}}, \; D_{1-3,150}^{\text{inter}}, \; D_{1-4,150}^{\text{inter}}, \; D_{2-3,150}^{\text{inter}}, \; D_{2-4,150}^{\text{inter}}, \; D_{3-4,150}^{\text{inter}}
\]
If the pairs are numbered as (p), then the inter-distance can be denoted as \(D_{p,i}^{\text{inter}}\):

\[
D_{1,150,3}^{\text{inter}}, D_{2,150,3}^{\text{inter}}, D_{3,150,3}^{\text{inter}}, D_{4,150,3}^{\text{inter}}, D_{5,150,3}^{\text{inter}}, D_{6,150,3}^{\text{inter}}
\]

, and the average inter-distance \(D_{\text{ave},i}^{\text{inter}}\) against a challenge \(X_i\), within an ambient \((\alpha)\), is:

\[
D_{\text{ave},i}^{\text{inter}} = \frac{\sum_{p=1}^{\text{Pairs}} D_{p,i}^{\text{inter}}}{\text{No. of Pairs}}
\]

For example, if the above parameters are applied to find \(D_{\text{ave},i}^{\text{inter}}\), then:

\[
D_{\text{ave},i}^{\text{inter}} = \frac{\sum_{p=1}^{\text{Pairs}} D_{p,i}^{\text{inter}}}{\text{No. of Pairs}} = \frac{D_{1,150}^{\text{inter}} + D_{2,150}^{\text{inter}} + D_{3,150}^{\text{inter}} + D_{4,150}^{\text{inter}} + D_{5,150}^{\text{inter}} + D_{6,150}^{\text{inter}}}{6}
\]

To avoid biasing the responses, an ideal inter-distance of a chip should be around 50%.
Appendix B

The Average Hamming Distance Distribution

The average distance distribution ($\mu_{i}^{\text{inter}}$) is determined as:

$$\mu_{i}^{\text{inter}} = \frac{D_{\text{ave},i}^{\text{inter}}}{N_{\text{res}}}$$

Alternatively, if substituting the $D_{\text{ave},i,\alpha}^{\text{inter}}$ as a formula (not as a calculated value), then:

$$\mu_{i}^{\text{inter}} = \frac{D_{\text{ave},i}^{\text{inter}} \cdot \sum_{p=1}^{\text{No. of Pairs}} D_{p,i}^{\text{inter}}}{N_{\text{res}} \cdot \text{No. of Pairs}} = \frac{1}{N_{\text{res}}} \cdot \sum_{p=1}^{\text{No. of Pairs}} D_{p,i}^{\text{inter}}$$

$$= \frac{1}{N_{\text{res}}} \cdot \frac{\sum_{p=1}^{\text{No. of Pairs}} D_{p,i}^{\text{inter}}}{(N_{\text{puf}}) \cdot (N_{\text{puf}} - 1)} = \frac{2}{N_{\text{res}} \cdot N_{\text{puf}} \cdot (N_{\text{puf}} - 1)} \cdot \sum_{\text{pair}=1}^{\text{No. of Pairs}} D_{p,i}^{\text{inter}}$$
Hamming distance between the response of the original PUF instance (PUF₁) and the response of the cloned PUF instance (PUF Clone) can be formulated as [49]:

\[
D_{\text{1-clone},i}^{\text{inter}} \triangleq \text{dist} \left[ \text{PUF}_{1,\alpha}(X_i); \text{PUF}_{\text{Clone},\alpha}(X_i) \right]
\]

or
\[
D_{\text{1-clone},i}^{\text{inter}} \triangleq \text{dist} \left[ \text{PUF}_{1,i,\alpha}; \text{PUF}_{\text{Clone},i,\alpha} \right]
\]

or
\[
D_{\text{1-clone},i}^{\text{inter}} \triangleq \| \text{PUF}_{1,\alpha}(X_i) - \text{PUF}_{1,\alpha}(X_i) \|
\]

, then
\[
D_{\text{1-clone},i}^{\text{inter}} \triangleq \| \text{PUF}_{1,i,\alpha} - \text{PUF}_{\text{Clone},i,\alpha} \|
\]

or
\[
D_{\text{1-clone},i}^{\text{inter}} \triangleq \| Y_{1,i,\alpha} - Y_{\text{Clone},i,\alpha} \|
\]

If: \(D_{\text{1-clone},i}^{\text{inter}} = 0\), for \(\forall 1 \leq i \leq N_{\text{cha}}\), then (PUF Clone) is an identical clone of (PUF₁), where \((N_{\text{cha}})\) is the maximum number of possible challenges (i.e., CRP) sets. Alternatively, for less restriction, if \(\Pr (D_{\text{1-clone},i}^{\text{inter}} = 0)\), for more than one \(X_i\) is high, then there is a high probability that (PUF Clone) is an identical clone of (PUF₁) for the other untested challenges and under other environmental conditions.