Emerging Memories for Dependable Computing

Elena Ioana Vatajelu

THÈSE

Pour obtenir l'Habilitation à Diriger des Recherches (HDR) de

L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : EEATS - Electronique, Electrotechnique, Automatique, Traitement du Signal (EEATS)
Spécialité : Nano électronique et Nano technologies
Unité de recherche : Laboratoire TIMA (Techniques de l'Informatique et de la Microélectronique pour l'Architecture des systèmes intégrés)

Mémoires émergentes pour les architectures de calcul fiable

Emerging Memories for Dependable Computing

Présentée par :

Elena-Ioana VATAJELU

Thèse soutenue publiquement le **24/01/2023**, devant le jury composé de :

Skandar BASROUR Professeur, Université Grenoble Alpes Laboratoire TIMA (Grenoble, France)	Président
Georges GIELEN KU Leuven (Belgium)	Rapporteur
Sylvain GUILLEY Professeur, Telecom ParisTech (France)	Rapporteur
Damien QUERLIOZ CNRS researcher - Université Paris-Saclay (France)	Rapporteur
Monica BURRIEL CNRS researcher - LMGP (France)	Examinateur
Luigi DILILLO CNRS researcher - LIRMM (France)	Examinateur
Ian O'CONNOR Professeur, Ecole Centrale de Lyon (France)	Examinateur

Emerging Memories for Dependable Computing

by

Elena-Ioana VATAJELU

October, 2022

Habilitation à Diriger les Recherches

Table of Contents

Table of Contents iii

List of Figures v

List of Tables viii

1 Summary 1
- 1.1 Curriculum Vitae . . . 2
  - 1.1.1 Personal Information . . . 2
  - 1.1.2 Education . . . 2
  - 1.1.3 Current Position . . . 2
  - 1.1.4 Past Positions . . . 3
  - 1.1.5 Executive Summary . . . 4
- 1.2 Summary of Scientific Activities . . . 6
  - 1.2.1 Memory – Robustness, Reliability & Test . . . 6
  - 1.2.2 Neuromorphic computing – design & test . . . 7
  - 1.2.3 In-Memory computing – design & security . . . 7
  - 1.2.4 Design of security primitives . . . 7
- 1.3 Supervision and Mentoring Activities . . . 8
  - 1.3.1 Ph.D. Candidates . . . 8
  - 1.3.2 Research Engineer . . . 9
  - 1.3.3 Interns (Master Students) . . . 9
- 1.4 Teaching Activities . . . 10
- 1.5 Scientific Projects and Collaborations . . . 10
  - 1.5.1 Project Coordination . . . 10
  - 1.5.2 Project Participation . . . 11
  - 1.5.3 Collaborations . . . 12
- 1.6 Dissemination of knowledge . . . 13
  - 1.6.1 Invited Presentations . . . 13
  - 1.6.2 Publications . . . 15
  - 1.6.3 Organisation of Scientific Events . . . 16
  - 1.6.4 Program Committee and Reviewer . . . 17
  - 1.6.5 Student Activities and Training Programs . . . 17
  - 1.6.6 Awards and Certificates . . . 18
- 1.7 Institutional Responsibility . . . 18

2 Research Activities 19
- 2.1 Preamble . . . 20
- 2.2 Introduction . . . 20

2.2.1 Memory - Robustness, Reliability Test . . . 21
2.2.2 Design and Evaluation of Hardware Security Primitives . . . 26
2.2.3 Design, Test and Security of Emerging Computing Paradigms . . . 28
2.3 Memory - Robustness, Reliability & Test . . . 32
2.3.1 SRAM Robustness Metrics . . . 32
2.3.2 Methodology for Statistical SRAM Robustness Analysis . . . 39
2.3.3 SRAM under SEU and Process Variability . . . 46
2.3.4 Variability Aware and Adaptive STAM Test . . . 47
2.3.5 STT-MRAM reliability evaluation and boosting techniques . . . 49
2.4 Design of Security Primitives . . . 56
2.4.1 Design and Evaluation of Phisical Unclonable Functions . . . 56
2.4.2 Design and Evaluation of Chip IDs . . . 64
2.4.3 Design and Evaluation of True Random Number Generators . . . 66
2.5 In-Memory Computing . . . 73
2.5.1 Comparative Study of Memristive Logic-in-Memory (LiM) Implementations . . . 73
2.5.2 Feasibility of complex Boolean functions using memristive-based LiM 75
2.6 Neuromorphic Computing - Design & Robustness Assessment . . . 78
2.6.1 Introduction . . . 78
2.6.2 Variability of fully-connected spiking neural network . . . 79
2.6.3 The design of a versatile analog spiking neuron in in 28nm UTBB FD-SOI technology . . . 84
2.6.4 The Design of a Probabilistic Spintronic Synapse . . . 85
2.6.5 Definition of Fault Models for Spiking Neural Networks . . . 88
2.6.6 Fault Injection Platform for Artificial Neural Networks . . . 89
2.6.7 Analysis of SNN Fault Tolerance . . . 90

3 Teaching and Supervision Activities **95**
3.1 Teaching Activities . . . 96
3.2 Supervision and Mentoring Activities . . . 98
3.2.1 Master Students . . . 98
3.2.2 PhD Candidates . . . 100
3.2.3 Supervision and Mentoring Strategy . . . 103

4 Funded Projects and Scientific Cooperation **105**
4.1 Funded Projects . . . 106
4.2 Research Collaborations . . . 113

5 Research Perspectives **117**
5.1 Perspectives on Reliability and Test . . . 118
5.1.1 In-Memory Computing (IMC) Solutions . . . 118
5.1.2 Neuromorphic Computing Solutions . . . 118
5.2 Perspectives on Security . . . 119
5.2.1 Physical Unclonable Functions (PUFs) . . . 119
5.2.2 Security of In-Memory Computing (IMC) Solutions . . . 120
5.3 Perspectives on Circuit Design for New Materials . . . 120

6 Publications **123**
6.1 Journal Publications . . . 124

6.2 Conference Proceedings 125
6.3 Workshops without Formal Proceedings 129

References **131**

List of Figures

1.1 Yearly publications record 15

2.1 IEEE reserch on SRAM Noise Margin and Statistic Yield Estimation Metrics . 23
2.2 IEEE research on SRAM Soft Error Rate measurement and estimation 24
2.3 IEEE research on STT-MRAM reliability evaluation and mitigation 25
2.4 IEEE research on PUFs (on the left) and TRNG (on the right) 28
2.5 IEEE research on In-Memory Computing 30
2.6 IEEE research on Neuromorphic Computing and Spiking Neural Networks . . . 31
2.7 Research topics developed during my research career 31
2.8 SRAM cell robustness analysis: a) SRAM core-cell in data retention mode with voltage noise disturbing its internal bodes; b) the "Butterfly Curve" of the SRAM core-cell; c) state space representation of an SRAM core-cell. 32
2.9 The SNM metrics compared: a) Maximum square, Piece-wise Linear Approximation and the Proposed Modified Negative Slope Criterion; b) Proposed Lower and Upper Bound Geometrical Estimation of SNM 34
2.10 a) Critical pulse width vs. noise amplitude for dynamic stability; b) Critical pulse energy vs. noise amplitude for dynamic stability 36
2.11 Comparison between the relative SNM, Qcrit and DNM metrics 37
2.12 SRAM under threshold voltage mismatch: a) The Meta-stable Equilibrium Point; b) Separatrix approximation 38
2.13 The time needed for the cells state to reach the separatrix a) in read operation mode, b) in write operation mode 38
2.14 a) Choice of the Significant Points on the Satisfiability Boundary: First Angular Bisection (FAB); b) the adjacent Significant Points on the Satisfiability Boundary: First Angular Bisection (FAB) 40
2.15 The polygonal estimation of the Satisfiability Boundary for different levels of angular bisection 41
2.16 Rectangular division of the parameter domain 43
2.17 2D Spec. Violation Metric (SVM) examples: a) best case scenario SVM = 0, b) SVM = 0.375, c) worst case scenario SVM = 1 (the total number of SPs in this illustration is 8) 43
2.18 Global and Operation Mode Contribution SVM in 2D analysis: a) Definition of the Overall Acceptance Region, b) Estimation of the Overall Satisfiability Boundary, c) Estimation of the Operation Mode Contribution Metrics and Global SVM . 45
2.19 Schematic representation of the proposed Statistical Failure Analysis Tool (SFAT) 45
2.20 SNM and SER variations due to random threshold voltage variation affecting the SRAM's transistors 46

2.21 SNM_{SB} (SNM_R) and SER variations due to random threshold voltage variation affecting only the PMOS transistors of the SRAM cell. 47
2.22 The STT-MRAM Memory cell: a) MTJ configurations; b) The $R_{MTJ} - V_{DC}$ hysteresis characteristic; c) Electric circuit of 1T1MTJ structure. 50
2.23 a) 2D illustration of failure mechanisms constraints during read and write operations of the 1T-1MTJ STT-MRAM. Here $W0F$ represents the write '0' failure region, $W1F$ represents the write '1' failure region, $R0F$ represents the read '0' failure region, $R1F$ represents the read '1' failure region, $TMR0$ represents the region where $TMR < 0$, and OK represents the NO failure region, i.e., the acceptance region; b) 3D representation of acceptance region in the (R_L, R_H, V_{TH}) parameter space. Three 'slices' are emphasized: the middle one corresponding to nominal value for V_{TH}, while the top and bottom ones correspond to V_{TH-MAX} and V_{TH-MIN}, respectively. 52
2.24 Graphical representation of the proposed Robustness Margin metrics (RM): a) RMs defined for read 0 and read 1 operation, respectively; b) RMs defined for write 0 and write 1 operation, respectively, c) RMs defined for read 0, read 1, write 0 and write1 operations no failures occur when the distributions of the MTJ electrical resistances are known, d) Failure probability estimation when some of the RM metrics take negative values - the distributions of the MTJ electrical resistances are known. 53
2.25 STT-MRAM cell reliability estimation under nominal values of the control voltages: a) reliability of the fresh cell affected by random variations in the MTJ resistance values; b) reliability degradation in time due to repetitive write stress, estimated assuming random variability of the MTJ resistance and NMOS threshold voltage (x-axis in logarithmic scale) 54
2.26 STT-MRAM cell reliability estimation under supply voltage (VDD) variation: a) reliability of the fresh cell affected by random variations in the MTJ resistance values; b) reliability degradation in time due to repetitive write stress, estimated assuming random variability of the MTJ resistance and NMOS threshold voltage. 55
2.27 The implementation strategy of the proposed PUF solution: 1) Write all cells to '1'; 2) Read each cell; 3) Use the read value 57
2.28 Schematic representation of the proposed PUF implementations: a) weak PUF, b) Strong PUF 58
2.29 Stability test detection: Classical test detects the most unstable cells, while the proposed test strategy detects the cells with high data stability 60
2.30 Phase space of the SRAM bit-cell and state evolution when the proposed method is implemented: a) for a perfectly symmetrical cell. b) for an asymmetrical bit-cell with inverter (M1&M2) stronger than inverter (M3&M4), c) for an asymmetrical bit-cell with inverter (M3&M4) stronger than inverter (MI&M2). 60
2.31 -Schematic representation of an SRAM memory array column with the dedicated circuit for stability testing marked in red 61
2.32 Left: correlation between minimum counter for reliable response and time of measurement. Right: Relationship between PUF reliability, counter difference and PUF entropy 63
2.33 a) Correlation matrix of sensor responses for one workload; b) Confusion Matrix for the FPGA recognition using the proposed method 65
2.34 TRNG implementation algorithm 68
2.35 Schematic circuit representation of TRNG implementation 68

2.36 The estimated percentage of '1' bits in a N-bit-string under different operation conditions (i) blue plots represent the results obtained without delay tuning, (ii) blue plots represent the results obtained with delay tuning in closed loop, (a) implementation at nominal current and temperature; (b) implementation under current variation at nominal temperature; (c) implementation at nominal current under temperature variation; (d) implementation under current ant temperature variation 70

2.37 Entropy of the MTJ state after performing a write operation 71

2.38 High-Entropy STT-MTJ-based TRNG implementation 71

2.39 NIST test results assuming various values of α_{XOR} 72

2.40 Primitive logic gates: Column 2 (CiM Solution) lists the considered LiM solutions and the corresponding primitive operations. The number of memory cells needed to implement a 2-input (1-bit) primitive operation is summarized in Column 3 (mem pts), while the schematic of the "primitive operation" gate for each solution is illustrated in Column 4 (Gate). Column 7 (Operations) lists the algorithm executed to obtain the result of the primitive operation. The executed operation can be input-destructive or not (see column 5, Remarks). An input-destructive operation is an operation that changes the value of the inputs after it is executed 74

2.41 State-of-the-art CIM solutions compared 76

2.42 Control voltage range for which the LiM Boolean operations are correctly executed within a memristor array 76

2.43 Map of operation conditions (amplitude and duration of the control pulse) for which the Boolean function is completed using LiM on memristive array. The upper figures (a) show the correctness of the operation and the retention of the input data (colored dots mark the conditions for which an operation is competed. The lower figures (b) illustrate the distribution of "perfect", "acceptable" and "worst correct" instances under the assumed control signals 77

2.44 Schematic representation of the neural network under study 80

2.45 Box-plot of STT-MTJ write probability estimation 82

2.46 The CMS (N=16) conductance evolution under 150 SLTP operations (in black) followed by 150 SLTD operations (in grey) 83

2.47 Standard design in 28 nm FD-SOI with a) schematic b) layout and area cost . 84

2.48 (a) Typical time simulations in 3 x 1μs period with nodes tracking, Vout (red), Vrefra (orange), Vm (blue) and Vctrl (grey); (b) Typical Monte Carlo simulations in 3 x 1μs period for global and local process variations. 85

2.49 (a) The MTJ resistance as a function of time during switching (the MTJ has only two stable states) together with a schematic of MTJ, the upper layer's magnetisation is pinned, whereas the lower one is controlled by the value of the current passing through it and its direction. Vpre and Vpost represent the control signals, which in the case of the SNN; (b) Schematic of a compound synapse with 4 MTJs in parallel, (c) schematic of the single layer SNN implement as a synapse crossbar array showing also the input and the output neurons 86

2.50 (a)The presynaptic neuron signal, The pulse Vpre has a triangular shape of 1µs. (b) The postsynaptic neuron signal, The pulse Vpost has positive and negative rectangular parts with a total duration of 100ns. (c)left: The voltage drop Vpre-Vpost across the synapse is positive and rises above the threshold when Vpost arrives short after Vpre resulting in a potentiation. (c)right: The voltage drop Vpre-Vpost across the synapse is negative and falls under the threshold when Vpost arrives at the end of Vpre pulse, resulting in a depression. (d) The average conductance of the synapse over 20 simulations for each position of the Vpost. The Vpre is fix, it always starts at 100ns, it is delimited by the two vertical lines. Whereas Vpost comes at diffident times (before, during and after Vpre) the conductance is probed only at the end of each simulation, (e) Potentiation and depression of a synapse initialized at an intermediate state of conductance, the presynaptic pulse is fixed at 100ns and is delimited by the two vertical line. Three simulations with different pulses are shown. A better result is obtained for slope of 0.326mV /ns. 87
2.51 (a) Generic IA chip; (b) schematic of the single layer SNN implement as a synapse crossbar array showing also the input and the output neurons; (c) Fault Injector 90
2.52 (a) STDP-based SNN recognition rate as a function of the size of the output layer; (b) STDP-based SNN recognition rate as a function of the number of output neurons affected by DNF; (c) STDP-based SNN recognition rate as a function of the number of input neurons affected by DNF; (d) STDP-based SNN recognition rate as a function of the number of synapses randomly affected by DNF; . . . 91
2.53 (a) STDP-based SNN recognition rate as a function of the size of the output layer; (b) STDP-based SNN recognition rate as a function of the number of output neurons affected by DNF; (c) STDP-based SNN recognition rate as a function of the number of input neurons affected by DNF; (d) STDP-based SNN recognition rate as a function of the number of synapses randomly affected by DNF; . . . 93

List of Tables

1.1 List of co-supervised Ph.D. candidates 9
1.2 Summary of the teaching activities 11
1.3 Summary of collaborations 13

2.1 Static Noise Margin estimated using the Piece-wise linear approach, Maximum Square Approach, the Proposed Modified Negative Slope Approach and the Proposed Geometrical Estimation 34

4.1 Summary of Research Grants 106

CHAPTER 1

Summary

1.1 Curriculum Vitae

1.1.1 Personal Information

- Name: Elena-Ioana VATAJELU
- Date and Place of Birth: 03 April 1981, Fagaras (Brasov, Romania)
- Nationality: Romanian
- Current position: "Chargé de Recherche classe normale", CNRS
- Work address: TIMA, UMR 5159, 46 Av Felix Viallet, 38000 Grenoble, France
- Work phone: +33 476.574.808
- Email: ioana.vatajelu@univ-grenoble-alpes.fr

1.1.2 Education

October 2007 - September 2011 PhD in Electronic Engineering at Universitat Politecnica de Catalunya, Barcelona, Spain. Supervisor: Prof. Joan Figueras

- Title of Doctoral Thesis: "Robustness Analysis of Nanometric SRAM Memories"
- Obtained distinction: cum laude

September 2004 - July 2005 MsC in Control Engineering at Technical University of Cluj-Napoca, Cluj-Napoca, Romania. Supervisor: Prof. Clement Festila.

- Title of Dissertation Thesis: "Robust Positioning – Robust Control Fundamentals applied on an elastic beam positioning system"

September 1999 - July 2004 BSc and MSc in Physics at Babes-Bolyai University of Cluj-Napoca, Physics Engineering Department, Cluj-Napoca, Romania. Supervisor: Prof. Sorin Dan Anghel

- Title of Master Thesis: "Timer Circuits – design, operation principles and applications"

1.1.3 Current Position

From October 2018 Full-time CNRS Researcher (FR. Chargée de Recherche Classe Normale) at TIMA Laboratory, Research group AMFORS, Grenoble, France

- Research activities focused on:
  - Design, test and reliability of emerging memories,
  - Design ad evaluation of hardware security primitives,
  - Design and dependability of emerging computing architectures.

1.1.4 Past Positions

September 2016 - September 2018 Research Assistant at TIMA Laboratory, Research group AMFORS, Grenoble, France

- Research activities focused on:
  - Design, test and reliability of emerging memories,
  - Design ad evaluation of hardware security primitives,
  - Design and dependability of emerging computing architectures.

January 2014 - August 2016 Research Assistant at Politecnico di Torino, Torino, Italy

- Research activities focused on:
  - Design, test and reliability of computer memories,
  - Design ad evaluation of hardware security primitives with emerging memories.

March 2012 - December 2013 Post-doctoral Research Fellow at LIRMM Laboratory, Montpellier, France

- Research activities focused on:
  - Process variability-aware test for nanometric SRAM memories,
  - Adaptive test for nanometric SRAM memories,
  - Reliability aspects of nanometric memory devices.

September 2011 - February 2012 Post-doctoral Research Fellow at Universitat Politecnica de Catalunya, Barcelona, Spain

- Research activities focused on:
  - Statistical analysis for reliability and yield estimation of VLSI circuits,
  - Design and reliability of ultra-scaled nanometric SRAM devices.

September 2010 - December 2010 Research Internship at University of Waterloo, Waterloo, Canada in the Research group of Prof. Manoj Sachdev

- Project: SRAM design for below threshold voltage operation and ultra-low power applications

September 2009 - December 2009 Research Internship at Purdue University, Purdue, USA, in the research group of Prof. Kaushik Roy.

- Project: SRAM dynamic robustness modeling and prediction

September 2004 - September 2007 Teaching and Research Assistant at Technical University of Cluj-Napoca, Cluj-Napoca, Romania

- Research activities focused on:
  - Modeling of distributed parameter systems,
  - Special structures for thermal process control,
  - Control strategies of electric oven for PCB fabrication.

1.1.5 Executive Summary

Main Research Topics :

- Memory design and test (CMOS and Beyond CMOS technologies),
- Techniques for circuit robustness and reliability estimation,
- Robust and variability-aware design,
- Neuromorphic Computing - design, reliability and test,
- In-memory Computing - design, reliability and test,
- Hardware security primitives (Phusically Unclonable Functions and True Random Number Generators) - design, reliability estimation and design for reliability.

Publications :

- 11 journal papers, 57 international conference papers with formal proceedings, 11 international workshops.

Invited Speaker :

- 2 keynotes, 3 tutorials, 13 invited presentations, 4 special sessions, 2 panels.

Research Projects :

- Project Leader of 4 projects:
  - 2020 - 2024 ANR-JCJC French (EMINENT) - 175.000€,
  - 2022 - 2025 80PRIME CNRS (SynConnect) - 135.000€,
  - 2021 - 2022 UGA IRS (CAD4RMD) - 14.000€,
  - 2018 CNRS INS2I-JCJC (RELeM) - 10.000€.
- Member of 14 projects: 1 European Horison Europe (Neuropulse), 1 French UGA IRS (SynConnect), 1 French ANR (POP), 1 French CNRS (PUF4IoT), 2 European FP7 (CLERECO, TRAMS), 1 European MC (NANOxCOMP), 1 French FUI (HiCool), 1 FEDER MICINN (TEC 2010-18384), 2 Spanish MICINN (TEC2013-41209-P, TEC2005-01027), 1 Italian CINI (Filiera Sicura), 1 Italian MIUR (STEPS), 1 Romanian ANCS (ANGIOTUMOR).

Monitoring and Supervision :

- 7 PhD candidates, 1 visiting PhD candidate, 1 engineer, 10 internships.

Teaching Activity :

- Overall +500 hours in: INP Grenoble (FR), Sorbonne University (FR), IMT Atlantique (FR), Politecnico di Torino (IT), University of Montpellier II (FR), Technical University Cluj Napoca (RO).

Organization of Scientific Events :

- Program: Program Chair (ETS), Program Chair (DTIS), Topic Chair (DATE), Track Chair (ETS), Topic Chair (ISVLSI) Track Chair (VLSID), Track Chair (DTIS), Track Chair (DDECS), Panel Chair (ETS), Special Session Chair (DTIS), Scientific Chair (TSS).

- Logistics: Publications Chair (DATE), University Booth Chair (DATE), Publication Chair (VTS), Publicity Chair (ETS), Organizing Chair (TSS).

Reviewing activities :

- Reviewer for International Journals: IEEE, ACM, Elsevier, Wiley.
- Conference Program Committee Member: DATE, VTS, ITC, ETS, ISCAS, ISVLSI, ICCD, VLSI-SoC, GLSVLSI, DTIS, DICS, AQTR, PRIME.

Editing activities :

- Associated Editor for Transactions on Computer Aided Design (IEEE),
- Associated Editor for Microelectronics Reliability (Elsevier),
- Guest Editor for Transactions on Emerging Technologies in Computing (ACM),
- Guest Editor for Microelectronics Reliability (Elsevier),
- Review Editor for Frontiers in Neuroscience/Neuromorphic Engineering.

Institutional Responsibilities :

- International: Steering Committee of the IEEE European Test Symposium, IEEE TTTC Technology Educational Program (TTEP), IEEE CEDA;
- FR National: CNRS GDR BioComp, CNRS GDR SoC2;
- Grenoble Local: FMNT Grenoble, TIMA Laboratory.

Societies :

- International: IEEE, ACM, IFIP;
- FR National: CNRS GDR BioComp, CNRS GDR SoC2.

1.2 Summary of Scientific Activities

My main research skills are related to dependability, test, fault tolerance and reliability of digital systems. At the beginning of my research carrier, these skills were applied to CMOS memories (SRAMs and DRAMs) and then extended to emerging memories (Magnetic and Resistive RAMs). During my postdoc at Politecnico di Torino, and even more during my postdoc at TIMA, I specialized in the use of such emerging memories for secure devices, as well as the design, test, and reliability of new computing paradigms (In-Memory and Neuromorphic Computing).

The memory devices are considered representatives for technology development. Therefore, advances in their fabrication (i.e., through the scaling for high densities and faster operations or through the introduction of new materials) are helpful for establishing the performance of other digital circuits. This is why **the CMOS and beyond-CMOS memories form the object of my research activities**.

Conventional Von-Neumann architectures and memories are not likely to fulfil all the needs of modern applications, due to inherent technological and conceptual limitations. Hence, in order to be at the forefront of the electronic industry in terms of design and manufacturing capabilities, it is essential to focus research and innovation efforts on the development of novel non-Von Neumann architectures enabled by emerging technology devices. This consideration motivates my interest **on memory-centric neuromorphic and in-memory computing architectures**.

Hardware dependability describes the ability of a system or component to function under stated conditions for a specified period of time. Dependability issues in hardware come from the manufacturing process, or operation and environmental conditions. Manufacturing induced variability, defects, stochastic effects, and aging degradation can cause important variations of the electrical characteristics of fabricated devices which can lead to device failure. Manufacturing test together with design-for-reliability solutions assure the quality and reliability of integrated circuits. This is why **my research activity is mainly focused on identifying the main reliability, robustness and stability issues faced by an electronic circuit (mostly memory-centered) and developing suitable test techniques for their detection and design solutions for their mitigation**.

Aside from reliability, an important aspect of device dependability is related to hardware security. Security systems use cryptographic protocols, frequently built on low-level cryptographic algorithms and primitives, such as Physically Unclonable Functions (PUFs) and True Random Number Generators (TRNGs). The Physically Unclonable Functions (PUFs) are emerging primitives exploited to implement low-cost authentication protocols and cryptographic primitives, such as secure key generators, key storing and one-way functions. State-of-the-art PUF solutions are memory-based and have piqued my interest, leading me to conduct research in the area of **design, evaluation and optimization of PUFs and TRNGs**.

1.2.1 Memory – Robustness, Reliability & Test

A considerable fraction of my research activity (PhD program and 3 years of post-doc) was focused on the modelling, the evaluation and the prediction of SRAM memory reliability and the effects of technology and supply voltage scaling, and process and temperature variations on the behaviour of a memory cell and on the parametric yield of a memory array. As a result I was able to propose new and easy-to-use metrics for robustness and functionality evaluation

of the SRAM cell for parametric yield estimation as well as adaptive and variability-aware test techniques. Later on I have adapted these metrics and tools to be suitable for the evaluation of beyond-CMOS memories (such as spintronic or memristive). In addition to working on the reliability and the test of CMOS and beyond-CMOS memories, I had the opportunity to bring my expertise to related fields, to explore the reliability aspects of alternate application domains, such as Service-Oriented Non-Volatile Memories (co-supervision of PhD student Marco Indaco) and Real-time Embedded Image Processing Systems (co-supervision of PhD student Pascal Trotta).

1.2.2 Neuromorphic computing – design & test

The Spiking Neural Networks (SNN) are widely studied nowadays due to the high level of realism they bring to neural simulation, their energy efficiency and their ability for on-line learning. I have studied, for the first time, the behavior of fully-connected, spintronic-based SNN under various sources of variability and unreliability and have demonstrated an important behavioral degradation of such neural networks under hardware variability, proving the need of further research towards building dependable hardware-based neuromorphic architectures. This work is now financed via an ANR JCJC project for which I am the project leader and the thesis director of PhD student Salah Daddinounou.

1.2.3 In-Memory computing – design & security

In-memory computing paradigm is an emerging concept based on the tight integration of traditionally separated memory elements and combinational circuitry. It allows the minimization of the time and the energy needed to move data across the processor. My research activities related to in-memory computing follow at the moment two directions: (i) design of SRAM-based in-memory computing hardware accelerators for low-power implementation of convolutional neural networks (co-supervision of PhD student Emilien Taly – CIFRE contract with ST) and (ii) analysis of reliability and security issues affecting memristive-based in-memory computing solutions (co-supervision of PhD student Pietro Inglese and project leader of a the CAD4RMD UGA-IRS grant).

1.2.4 Design of security primitives

In parallel to evaluating the detrimental effect of fabrication process and stochastic variability on the behavior of a memory cell, I am also investigating the means to use such variability to our advantage. Indeed, the fabrication process variability is an excellent source of entropy for the design of Physical Unclonable Functions, while stochastic variability at runtime is a great basis for Random Number Generators. During my research activities I have developed PUF and TRNG solutions based on the STT-MTJ devices and have proposed solutions to improve the reliability of memory-based PUFs (both CMOS and beyond-CMOS memories were targeted). Also, recently I have starting evaluating the feasibility of PUFs and TRNGs based on memristive devices (co-supervision of PhD student Sergio Vinagrero).

1.3 Supervision and Mentoring Activities

Up to now, I have co-supervised 7 Ph.D. candidates, 1 visiting Ph.D. candidate, 1 engineer and 10 master students.

1.3.1 Ph.D. Candidates

Graduated

- Graduated February 2014 : Marco Indaco *Service-Oriented Non-Volatile Memories* - ***Supervision quota: 33% during the second and third year of the thesis***, Doctoral school Politecnico di Torino, Torino, Italy,
- Graduated March 2016 : Pascal Trotta *Enhancing Real-time Embedded Image Processing Robustness on Reconfigurable Devices for Critical Applications* - ***Supervision quota: 33% during the second and third year of the thesis***, Doctoral school Politecnico di Torino, Torino, Italy

Ongoing

- 2021-2024: Sergio Vinagrero *Design and Evaluation of memristive-based Security Primitives* - ***Supervision quota: 80%*** , Doctoral school EEATS, UGA, Grenoble, France,
- 2021-2024: Emilien Taly *Design of a very low power AI System based on SRAM-based in-memory computing* - ***Supervision quota: 40%*** , Doctoral school EEATS, UGA, Grenoble, France,
- 2020-2023: Salah Daddinounou *Test and Reliability of Emerging Memory-based Spiking Neural Networks* - ***Supervision quota: 100% (Thesis Director)***, Doctoral school EEATS, UGA, Grenoble, France,
- 2019-2022: Pietro Inglese *Exploration of security threats in In-Memory Computing Paradigms* - ***Supervision quota: 34% first 6 months of the thesis, 67% afterwords***, Doctoral school EEATS, UGA, Grenoble, France,
- 2018-2021: Valerian Cincon *Neuromorphic systems in 28 nm FD-SOI technology* - ***Supervision quota: 33% for each of the three years of the thesis***, Doctoral school EEATS, UGA, Grenoble, France

Visiting

- January 2020-June 2020 : Vishal Gupta *Exploration of synaptic-like characteristics of memristor devices* - ***Supervision quota: 100%***, Doctoral School of Tor Vergata University, Rome, Italy

The complete list of the Ph.D. candidates I have co-supervised, as well as the supervision quota are given in Table 1.1.

Table 1.1: List of co-supervised Ph.D. candidates

Name	Supervision Period	Supervision quota [%]
Sergio Vinagrero	2021-	80
Emilien Taly	2021-	40
Salah Daddinounou	2020-	100
Pierto Inglese	2019-	67
Valerian Cincon	2018-	33
Marco Indaco	2013-2014	33
Pascal Trotta	2014-2016	33
Vishal Gupta	01.2020-06.2020	100

1.3.2 Research Engineer

- January 2021-April 2021 : Emilien Taly *Exploration of SRAM in-memory computing solutions for accelerating embedded Neural Networks*

1.3.3 Interns (Master Students)

Master II - 6 months internship

- 2022: Arnaud Degreze - *Analysis of the efficiency of LaNiO4 memristive devices for bio-inspired learning*, Supervision quota 50% - MSC. INSA Lyon Project financed by FMNT
- 2022: Sara Mannaa - *Design and Fault-Tolerance of a Spintronic-based Spiking Neuron*, Supervision quota 100% - MSC. WICS, Grenoble INP
- 2022: Marcelo Correa Cueto - *Design and Evaluation of Logic in Memory Solutions for FeFETs*, Supervision quota 50% - MSC. Phelma, Grenoble INP
- 2022: Luis Felipe Camponogara - *Design and Evaluation of Logic in Memory Solutions for ReRAMs*, Supervision quota 50% - MSC. Phelma, Grenoble INP
- 2021: Sergio Vinagrero - *Design and Evaluation of a OxRAM-based Security Primitives*, Supervision quota 100% - MSC. WICS, Grenoble INP
- 2020: Ihab Alshaer - *Physically Unclonable Function (PUF) and Random Number Generation (RNG) with Neural Networks (NN)*, Supervision quota 50% - MSC. CySEC ENSIMAG (2019-2020), Grenoble INPG/UGA, Grenoble, France - CyberAlps Institute
- 2019: Ali Bawab - *Compact Models for Synaptic-Compliant La2NiO4 Memristive Devices*, Supervision quota 80% - MSC. Advanced Electronic Systems Engineering (2018-2019), University of Burgundy, Dijon, France - Project financed by FMNT

Master I - 3 months internship

- 2022 : Matthieu Charles - *Experimental platform for ReRAM evaluation*, Supervision quota 100% - PHELMA INPG/UGA - Project financed by ANR/EMINENT

- 2022 : Jardel Kaique - *Evaluation of laser attacks on power-off CMOS devices*, Supervision quota 30% - PHELMA INPG/UGA - Project financed by ANR/POP

- 2020 : Solenn Guironnet - *Design of Emerging Memory-based Spiking Neural Networks*, Supervision quota 100% - PHELMA INPG/UGA - Project financed by ANR/EMINENT

1.4 Teaching Activities

After graduating the university I have performed various teaching activities at different levels: engineering students of the Technical University of Cluj-Napoca Romania, Montpellier University II France, master students of the Politecnico di Torino Italy, Polytechnic Institute of Grenoble (INPG) France, Sorbone University Paris France, PhD candidates at IMT Atlantique France. Table 1.2 summarizes my pedagogical activities. The column dedicated to the time spent for each subject is classified in 3 set: Lec.=lectures (i.e., "Cours magistraux"), Tut.=tutorials (i.e., "Travaux dirigés"), and Pra.=practical work (i.e., "Travaux pratique").

1.5 Scientific Projects and Collaborations

Starting from my Ph.D., I have actively cooperated to national- and european-funded scientific projects. The following list summarizes, for each project, the period, the project type and the type of involvement. Details concerning each project are given in Chapter 4.

1.5.1 Project Coordination

- 2022 - 2025: Project Leader of the CNRS 80PRIME Project SynConnect - Study and development of La2NiO4 memristive devices for bio-inspired computing. The project goal is to demonstrate, at small scale, the feasibility of a La2NiO4-based SNN and understand the main advantages and shortcomings (from technology and application) such that concomitant optimization of device and algorithm can be performed to guarantee an efficient bio-inspired electronic system.

- 2021 - 2022: Project Leader of the UGA IRS Project CAD4RMD - CAD Tool and Design Space Exploration for Resistive-based Memory Devices Integration in non-Von Neuman Architectures. The purpose of this project is to provide a tool (modular and expendable over different resistive memory devices and computing paradigms) which will guide the designer towards the optimal combination technology/computation for desired application.

- 2019 - 2023: Project Leader of the ANR JCJC Project EMINENT - Test and Reliability of Emerging Memory-based Spiking Neural Networks. The goal of the this project is to provide a dependable Emerging Memory-based Spiking Neural Network architecture.

- 2019: Project Leader of the CNRS INS2I JCJC Project RELeM - Reliable and Secure Emerging Memories for Dependable Computing Architectures. The aim of the project is to comprehensively model, at different abstraction levels, the behaviour of emerging memories with and without the presence of possible faults, in order to enable a reliability- and security-aware design space exploration.

Table 1.2: Summary of the teaching activities

Subject	Period	Hours			Location
		Lec.	Tut.	Pra.	
Emerging topics in computing	2019/20	4			IMT Atlantique
	2020/21	4			
	2021/22	4			
Emerging non-volatile memories	2019/20	4			IMT Atlantique
	2020/21	4			
	2021/22	4			
	2019/20	4			Sorbone Univ.
	2020/21	4			
	2020/21	4			
Hardware Trust – Physiscal Security	2017/18	4			Grenoble INP
	2018/19	4			
	2019/20	6			
	2020/21	6			
	2021/22	6			
Analyse et réalisation d'un système complexe	2017/18		28		Grenoble INP
	2018/19		28		
	2019/20		28		
	2020/21		28		
	2021/22		28		
Information Technology Tools	2012/13			80	Montpellier Univ. II
	2013/14			80	
Control Engineering	2005/06			80	Technical University of Cluj-Napoca
	2006/07			120	
System Theory	2005/06			40	Technical University of Cluj-Napoca
	2006/07			40	
Power Electronics in Automatic Control	2004/05			40	Technical University of Cluj-Napoca

1.5.2 Project Participation

- 2023 - 2027: Scientific Leader for TIMA Laboratory of the Horizon Europe Project Neuropuls - Neuromorphic energy-efficient accelerators based on Phase change materials augmented silicon photonics. The project goal is to develop secure hardware accelerators based on novel neuromorphic architectures and PUF-based security layers leveraging the benefits offered by the integration of photonics, PCMs and III-V materials.
- 2022 - 2023: Scientific Leader for TIMA Laboratory of the UGA IRS Project SynConnect - Fabrication of La2NiO4 memristive devices for bio-inspired computing. The project goal is to fabricate a small-scale La2NiO4 memristive array optimized for efficient bio-inspired electronic systems.
- 2022 - 2026: Member of the ANR PRC Project POP - Power-OFF laser attacks on security Primitives. The project goal is to evaluate the sensitivity of PUFs to power-off

laser attacks and develop countermeasures against these attacks.

- 2020: Scientific leader of the CNRS INS2I Project PUF4IoT - Physically-Unclonable Functions (PUF) for Secure IoT. The project goal was to design and evaluate highly reliable PUFs.

- 2017 - 2018: Member of the CISCO Granted Project FilieraSicura - Securing the Supply Chains of Domestic Critical Infrastructures from Cyber Attacks. The project aims to provide new methodologies, techniques and tools to protect the country's critical national infrastructure and companies from cyber attacks.

- 2016-2018: Member of the European Union's H2020, Marie Skłodowska-Curie Project NANOxCOMP - Synthesis and Performance Optimization of a Switching Nano-Crossbar Computer. This was a research mobility project focused on the use of beyond-CMOS technologies for the development and the hardware implementation of emerging computing paradigms.

- 2014-2015: Member of the European FP7 Project (STREP) CLERECO - Cross-Layer Early Reliability Evaluation for the Computing Continuum. This project is devoted to the early evaluation of the reliability of complex digital systems based on micro-processors, for both embedded systems and High Performance Computers.

- 2012-2014: Member of the French FUI Project HiCool - Solutions amont pour la conception orientée base consommation de circuits integrés complexes (Upstream solutions for consumer-oriented design of complex integrated circuits). The objective of the project is to improve the design process of system-on-chip devices that require a high level of control over their energy dissipation.

- 2012-2014: Member of the Spanish FEDER & MICINN Project LPMTR - Low power memory test and reliability. The project is focused on developing new methods and techniques for efficient reliability evaluation and memory test mechanisms.

- 2011-2012: Task Leader of the European Union's FP7 Project TRAMS - Terascale Reliable Adaptive Memory Systems. The TRAMS project is focused on the design of reliable, energy efficient and cost effective computing in the era of nanoscale challenges and teraflop opportunities.

- 2007-2011: Member of the Spanish MICINN Project LPCD TEC2005-01027 - Low power CMOS devices. This project was focused on the use of low-power CMOS devices across different applications and on the study of their reliability and development of suitable test techniques.

- 2006-2007: Task Leader of the Romanian ANCS Project ANGIOTUMOR - The prediction of the evolution and the assessment of the treatment's response for malignant tumors using morphological and hemodynamic modeling through imagistic, mathematical and artificial intelligence techniques.

1.5.3 Collaborations

Table 1.3 summarizes the main collaborations I had since I started my research career. I showed here only collaborations that led to either a publication or a common research project.

Table 1.3: Summary of collaborations

Institution	Name	Research Topic			
		Memory	IMC	Neuro.	HW Trust
Academia					
UPC	Antonio Rubio	√			
	Rosa Rodriguez	√			
LIRMM	Michel Renovell	√			
	Lionel Torres				√
	Giorgio Di Natale				√
IM2NP	Hassen Aziza	√			
	Jean-Michel Portal	√			
INL	Alberto Bosio	√	√	√	
	Ian O'Connor		√	√	
Spintec	Guillaume Prenat	√			
	Mihai Miron			√	
LEAT	Benoit Miramond			√	
CEA	Elisa Vianello			√	
TU Delft	Said Hamdioui	√	√		
	Motta Taouil	√			
U. Carlos III Madrid	Honorio Martin				√
U. of Southampton	Basel Halak				√
Aarhus University	Farshad Moradi	√			
U. degli Studi di Milano	Valentina Ciriani		√		
University of Ferrara	Cristian Zambelli	√			
U. Naples Federico II	Mario Barbareschi				√
Purdue University	Kaushik Roy	√		√	
Industry					
Intel	Javier Vera	√			
Infineon	Nabil Badereddine	√			
Blu5	Antonio Variale				√
ST Microelectronics	Philippe Galy			√	
	Pascal Urard		√	√	

1.6 Dissemination of knowledge

Starting from my Ph.D., I have been actively involved in the international scientific community. I have help organise and steer several scientific events. Besides my publication record, I have also contributed to the quality of the technical content by providing reviews, organising special sessions, participating in panels and delivering invited presentations.

1.6.1 Invited Presentations

Keynote Addresses

- 2020 Baltic Electronics Conference (BEC) - "Versatility of Emergent Memory Technolo-

gies: Friend or Foe?",

- 2019 Prague Embedded-Systems Workshop (PESW) - "Randomness in emerging technologies : Functional robustness vs. security".

Tutorials

- 2016 Training School on Trustworthy Manufacturing and Utilization of Secure Devices (TRUDEVICE) - "Emerging Memories and Randomness",
- 2015: IEEE Design and Technology of Integrated Systems (DTIS) Conference - Half Day Tutorial "The Memories of Tomorrow : Technology, Design, Test, and Dependability",
- 2014: IEEE European Test Symposium (ETS) - Embedded Tutorial "On the Impact of Process Variability and Aging on the Reliability of Emerging Memories"

Invited Talks at International Events

- 2022: 23rd IEEE Latin-American Test Symposium (LATS) - "Versatility of Emergent Memory Technologies",
- 2019: IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFTS) - "Rebooting Computing : the challenges for Test and Reliability",
- 2019: International Verification and Security Workshop (IVSW) - "STT-MRAM-based Security Primitives : PUF and TRNG",
- 2018: IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR) - "Emerging Memories for Security Primitives",
- 2017: International Mixed Signal Test Workshop (IMSTW) - "Test and Reliability in Approximate Computing",
- 2017: International Verification and Security Workshop (IVSW) - "Design Considerations for Spintronic-based Security Primitives",
- 2017: IEEE Design and Technology of Integrated Systems (DTIS) - "Memristive devices and their potential",
- 2016: International Mixed Signal Test Workshop (IMSTW) - "Randomness in emerging technologies : functional robustness vs. security",
- 2015: IEEE Design and Technology of Integrated Systems (DTIS) - "Physically Unclonable Function Implementation based on the variability of the Spin-Transfer-Torque magnetic memories",
- 2013: ESSCIRC, FP7 Variability and Reliability Show-Case - "Robustness Analysis of Nanometric SRAM Memories",
- 2011: Computing Architectures Software Tools and Nano-Technologies for Numerical Embedded and Scalable Systems (CASTNESS) - TRAMS: "Robustness of SRAM Memories".

Invited Talks at National Events (French)

- 2021: GDR SoC2 Scientific day on Security applications of emerging technologies - "PUF/TRNG avec des mémoires RRAM/MRAM",
- 2019: Inter GDR event (BioComp, SoC2, IQFA) Quantum and neuromorphic technologies meet - Faults in Spiking Neural Networks with Spike Timing Dependent Plasticity,
- 2019: GDR BioComp Conference - Fault Modeling of Spiking Neural Networks with STDP,
- 2018: GDR SoC2 BarCamp - Test and Reliability of Emerging Non-Volatile Memories.

Special Session Organiser

- 2019: IEEE VLSI Test Symposium (VTS) - "Testing Challenges for Neuromorphic Computing",
- 2018: IFIP/IEEE International Conference on Very Large Scale Integration (VLSI-SoC) - "Neuromorphic Computing - From Robust Hardware Architectures to Testing Strategies",
- 2018: International On-Line Test Symposium (IOLTS) - "Resistive and Spintronic RAMs : Device, Simulation, and Applications",
- 2016: IEEE VLSI Test Symposium (VTS) - "Spintronic Devices for Security Primitives"

Participation in Panels

- 2019: International Verification and Security Workshop (IVSW) - "Security Challenges for Neuromorphic Computing",
- 2016: International Mixed Signal Test Workshop (IMSTW) - "DFT vs. Security – Is it a Contradiction? How Can We Get the Best of Both Worlds?"

1.6.2 Publications

Figure 1.1 summarizes the number of my publications for each year including journal papers (with review process), papers published in official proceedings (coming from conferences or workshops with review process), and presentations given in national and international events without official proceedings. The complete list of publications is given in Chapter 6.

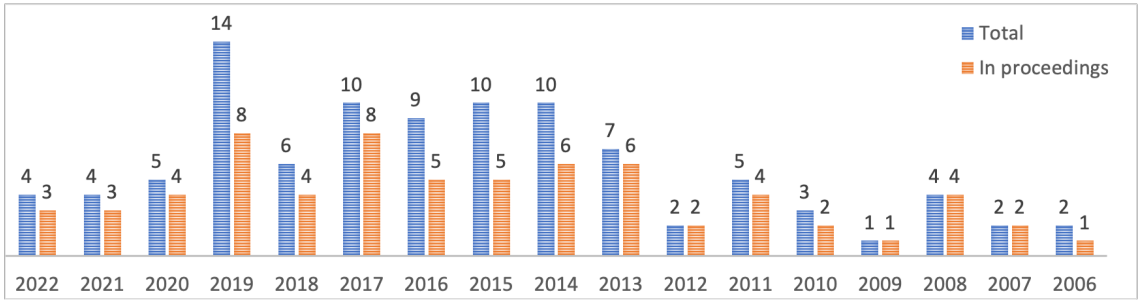


Figure 1.1: Yearly publications record

1.6.3 Organisation of Scientific Events

Program Chair

- 2021: IEEE European Test Symposium (ETS)
- 2019: IEEE International Conference on Design and Technology of Integrated Systems in Nanoscale Era (DTIS),
- 2019: Workshop on Emerging Memories: Technology, Design and Test (eMTDT)
- Program Vice-Chair 2020: IEEE European Test Symposium (ETS)

Topic/Track Chair

- 2023, 2022, 2020, 2019: IEEE European Test Symposium (ETS) - topic Memory and Emerging Technologies
- 2022 IEEE Computer Society Annual Symposium on VLSI (ISVLSI) - track Emerging and Post-CMOS Technologies,
- 2022 Conference on VLSI Design (VLSID) - track Test and Reliability,
- 2020, 2018: IEEE International Conference on Design and Technology of Integrated Systems in Nanoscale Era (DTIS) - track Design and Test of Emerging Technologies,
- 2019: IEEE Workshop on Design and Diagnostics of Electronic Circuits and Systems (DDECS) - track Design and Test of Emerging Technologies,
- 2017-2015: IEEE/ACM Design, Automation and Test in Europe Conference (DATE) - topic Design Verification and Validation

Panel Chair

- 2016 : IEEE European Test Symposium (ETS)

Special Session Chair

- 2015, 2014: IEEE International Conference on Design and Technology of Integrated Systems in Nanoscale Era (DTIS)

Publications/Proceedings Chair

- 2022-2020: IEEE/ACM Design, Automation and Test in Europe Conference (DATE)
- 2020-2015: IEEE VLSI Test Symposium (VTS)

Publicity Chair

- 2023: IEEE Latin American Test Symposium (LATS)
- 2018, 2017: IEEE European Test Symposium (ETS)

1.6.4 Program Committee and Reviewer

Editing Activities

- since 2022: Associate Editor for IEEE Transactions on Computer Aided Design
- since 2018: Associate Editor for Journal of Microelectronics Reliability (Elsevier)
- 2020: Guest Editor for Transactions on Emerging Technologies in Computing (ACM) - Special issue on Computing in Memory
- 2019: Guest Editor for Journal of Microelectronics Reliability (Elsevier) - Special Issue on Design and Technology of Integrated Systems
- since 2019: Review Editor for Frontiers in Neuroscience - Neuromorphic Engineering

Program Committee Member for the following Conferences

IEEE VLSI Test Symposium (VTS), IEEE European Test Symposium (ETS), IEEE International Conference on Design and Technology of Integrated Systems in Nanoscale Era (DTIS), IFIP/IEEE International Conference on Very Large Scale Integration (VLSI-SoC), IEEE Workshop on Design and Diagnostics of Electronic Circuits and Systems (DDECS), ACM Great Lake Symposim on VLSI (GLSVLSI), IEEE International Symposium on Circuits and Systems (ISCAS), IEEE International Conference on Computer Design (ICCD), IEEE/ACM Design, Automation and Test in Europe Conference (DATE), IEEE Computer Society Annual Symposium on VLSI (ISVLSI), IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH), Conference on Design of Circuits and Integrated Systems (DCIS), IEEE International Test Conference (ITC)

Reviewer for the following Journals

IEEE Design and Test of Computers, IEEE Transactions on Circuits and Systems, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, IEEE Transactions on Computer-Aided Design, IEEE Transactions on Computers, IEEE Transaction on Emerging Topics in Computing, IEEE Transactions on Nanotechnology, IEEE Embedded Systems Letters, IEEE Transactions on Embedded Computing Systems, ACM Journal on Emerging Technologies in Computing Systems, Elsevier Journal of Electronic Testing, Elsevier Microelectronics Journal, Elsevier VLSI Integration, Elsevier Journal of Parallel and Distributed Computing, MDPI Cryptography

1.6.5 Student Activities and Training Programs

University Booth for Design, Automation and Test in Europe Conference (DATE)

- 2019, 2017: Chair

Test Spring School at European Test Symposium (ETS)

- 2023, 2022: Chair of the Scientific Committee,
- 2021, 2020: Member of the Scientific Committee,

- 2017, 2016, 2015: Logistics Chair

Test Technology Educational Program (TTEP)

- 2022, 2021: General Chair,
- 2019, 2020: Program Chair

1.6.6 Awards and Certificates

Best Paper

- 2011: Workshop on Design and Diagnostics of Electronic Circuits and Systems (DDECS),
- 2010: IEEE Conference on Automation Quality Testing and Robotics (AQTR)

Organisation of Scientific Meetings

- 2020: IEEE TTTC Award "ITC 2020 Tutorials Program - Most Successful Technical Meeting"

1.7 Institutional Responsibility

Local

- TIMA Laboratory - Member of the Laboratory Council - since 2021,
- TIMA Laboratory - In charge of Gender Equality - since 2020,
- Micro and Nanotechnology Federation Grenoble (FMNT) - Scientific Coordinator (FR "Responsable d'axe scientifique - Microelectronique") - since 2021.

National

- GDR SoC2 - Vice-Scientific Coordinator (FR Responsable adjoint d'axe thematique Sécurité et intégrité des systèmes) - since 2021,
- GDR BioComp - Member of the board - since 2019.

International

- IEEE Council on Electronic Design Automation (CEDA) - Secretary of the Executive Committee - from 2022,
- IEEE European Test Symposium - Steering Committee Member - since 2019,
- International Federation for Information Processing (IFIP) - Member of Working Group 10.5 "Design and Engineering of Electronic Systems" - since 2020.

CHAPTER 2

Research Activities

2.1 Preamble

I have started my research career in the area of micro- and nano-electronics in October 2007 with my PhD at UPC (Universitat Politecnica de Catalunya), Barcelona, Spain, under the supervision of the Prof. Joan Figueras. His research activity mainly focused on the development of new test methodologies and the implementation of algorithms able to improve the development of highly dependable digital components. My work at the time was centred on SRAM memory design, reliability and test. At the same institution, I have worked in collaboration with Prof. Antonio Rubio and his research group in the framework of the FP7 TRAMS (Terascale Reliable Adaptable Memory Systems) project. My contribution was mainly on the investigation of the impact of statistical variability and reliability of near (and beyond) the end of the ITRS devices on terabit memory design. As main results, I have developed an ontology for robustness and reliability estimation to provide a comprehensive comparison between ultrascaled memory technologies. In the frame of this project, I have closely collaborated with Intel Barcelona, Imec Leuven-Belgium and Glasgow University. During my PhD I have been appointed a 3-month internship at Purdue University - USA with Prof. Kaushik Roy and another 3-month internship at Waterloo University - Canada with Prof. Manoj Sachdev. In March 2012 I started a contract for a post-doctoral position at LIRMM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier) in the research group of Dr. Patrick Girard (Directeur de Recherche at CNRS), in collaboration with Intel Mobile Communication (IMC) Lab in Sophia Antipolis, France. My research topic was centred around variability aware adaptive memory testing. As a side project, I have also worked on SRAM SEU modelling under fabrication induced process variability, in collaboration with L'Institut d'Electronique-UM Montpellier (IES). In January 2013, I started a second contract for a post-doctoral position at PoliTO (Politecnico di Torino), Italy, in the research group of Prof. Paolo Prinetto. Here, my research activity was focused on design, test and reliability of current and emerging memories, and implementation of hardware security primitives with emerging memories. During my time in Politecnico di Torino, I have maintained and further developed my research collaborations with Universitat Politecnica de Catalunya and LIRMM and I have managed to open the path to new industry and academia collaborations, such as Blu5, Delft University, Universita degli Studi di Napoli Federico II, Universita di Ferrara, Aix Marseille Université, Limerick University. In September 2016 I started a contract as an Expert Collaborator at TIMA Laboratory in Grenoble, France, in the research group AMFORS (Architectures and Methods for Resilient Systems), which in 2018 became a tenured research position with CNRS (CRCN). In TIMA I continue the research activity focused on test and reliability of emerging memories, and implementation of hardware security primitives with emerging memories. At the same time, I have started working on a new research direction, i.e., dependable computing architectures (von Neumann and non-von Neumann) build with emerging memory devices.

2.2 Introduction

Semiconductor memories are considered representatives for CMOS technology development; therefore, advances in their fabrication, through the scaling for high densities and faster operations are helpful for establishing the performance of other digital circuits. According to the International Technology Roadmap for Semiconductors stand-alone memory market increases drastically and embedded memory resources are continuously increasing and has approached 94% of the System-on-a-Chip (SoC) silicon area.

The semiconductor memories are divided into two large categories: volatile and non-volatile memories. The volatile memories are named like this because of their incapability of maintaining the stored data after the supply voltage was suppressed; meanwhile the non-volatile memories maintain the stored data under the same circumstances. These memories are typically used for the task of secondary storage, or long-term persistent storage. For the task of primary storage, the volatile memories are used, especially random access memories (static – SRAM or dynamic – DRAM) because they are faster and more compact than the non-volatile ones. SRAM is more expensive, but faster and significantly less power hungry (especially when in idle status) than DRAM. It is therefore used where either bandwidth or low power, or both, are principal considerations. SRAM is used for cache memories, both on and off the CPU chip. Through the years the SRAM memories have known a strong and rapid development due to their use in almost all large and complex digital circuitry.

2.2.1 Memory - Robustness, Reliability Test

There are two main concerns regarding the robustness of nanometric memories in general and to Static Random Access Memories (SRAM) in particular: technology scaling and supply voltage scaling. With the continuous technology scaling, the robustness of the SRAM memories becomes an important concern for the design and test engineers. Maintaining an acceptable level of robustness in SRAMs while scaling the minimum feature size and supply voltages of Systems-on-a-Chip (SoC) becomes increasingly challenging. Pushing the physical limits of scaling leads to device patterning challenges and non-uniform channel doping. In order to improve system performance, large arrays of fast SRAM are required, but the area impact of incorporating large SRAM into a chip translates into a higher chip cost. As a result, minimum-size SRAM cells are tightly packed, making SRAM arrays the densest circuitry on a chip. This is the reason why the SRAM arrays are the most susceptible and sensitive to manufacturing defects and process variations. Any asymmetry in the SRAM cell structure will make the cell less robust. The expected difficulties with scaling the SRAM are maintaining both acceptable noise margins (static and dynamic) and controlling its functionality.

In light of all these considerations, a large fraction of my research activity was focused on the modelling, evaluation and prediction of SRAM memory reliability and the effects of technology and supply voltage scaling and process and temperature variations on the behaviour of a SRAM cell and on the parametric yield of an SRAM array. The work was carried out by analyzing the SRAM cell robustness and functionality under process and environmental variations and providing new and easy to use metrics for robustness and functionality evaluation of the SRAM cell for parametric yield estimation (work carried out at UPC Barcelona, Spain). This work is briefly described in subsections 2.3.1 and 2.3.2 of this document. My main contributions were on (1) developing metrics for SRAM robustness evaluation and (2) developing methodology for fast SRAM yield estimation.

I have proposed metrics for both static and dynamic SRAM robustness evaluation. Regarding the static robustness evaluation, I have conducted my work following the trend in the research community, where the focuss was on the evaluation of the Static Noise Margin (SNM). At the time I conducted my reserch, expressions for above – threshold SNM were proposed by Seevinck et al. (1) and Bhavnagarwala et al. (2) and for sub – threshold SNM by Calhoun and Chandrakasan (3) and Welling and Zory (4). In (1) an explicit analytic expression for the SNM based on long channel models as a function of device parameters and supply voltage is derived. Bhavnagarwala et al. (2) study of SRAM cell SNM reduction due to intrinsic threshold voltage fluctuations in uniformly doped minimum geometry MOSFETs. The work

described in (4) investigates stability aspects of sub-threshold SRAM cells, deriving analytical expressions for the SNM as a function of circuit parameters, operating conditions and process variations. The noise margin of SRAM cells in low power conditions such as low supply voltage and source-body bias were experimentally determined, using a graphical approach in Cseveny et al. (5) and Hook et al. (6). Calhoun and Chandrakasan (3) have analysed the dependence of SNM on supply voltage, temperature, transistor sizes and global process variation in a commercial 65nm technology. My work builds on this previous work and extends it by examining the influence of lowering the supply voltage of a SRAM cell on its robustness (noise immunity), both in above– as in sub– threshold regions taking into account the short channel effects in MOSFET scaling, as velocity saturation and channel length modulation. Regarding the dynamic robustness evaluation, I have conducted my work both inside and outside of the trend imposed by the research community. The trend was on the evaluation of the dynamic stability when dynamic perturbations were considered (such as single event upsets) while part of my work was focus on deriving a dynamic stability estimation procedure and metric when static perturbations were considered (such as fabrication-induced variability). In this context, my work was mainly inspired by the work of Zhang et al. (7) and Ding et al. (8) where the non-linear circuit theory was first used to predict the transient noise immunity of an SRAM cell, as well as the work of Sharifkhani and Sachdev (9) who used the same principles to develop a dynamic noise margin model for SRAM working in sub-threshold regime.

Moreover, to be able to truly predict the reliability of an SRAM cell/array, I have researched methodologies for statistical yield estimation. The most common statistical failure analysis methodology for circuits under process variability is based on Monte Carlo (MC) simulation where the accuracy of the statistical failure analysis depends on the sample size. Extensive research has been devoted to the reduction of sample size for speed improvement while maintaining the accuracy of standard Monte Carlo simulations. One of the resulting methods is the Stratified Sampling technique (10), which consists in stratifying the sample space by choosing a partition of the input parameter space. The integrals in each stratum are than estimated and combined to obtain the overall integral. Another common method of SRAM analysis is Importance Sampling (11), (12) it is based on the fact that certain values of the input random variables have more impact on results than others. The work of R. Kanj, R. Joshi and S. Nassif (11) presents a methodology for statistical SRAM design and analysis based on Mixture Importance Sampling. Random variables using mixtures of distributions focusing on the failure region are generated. Two approaches can be taken for the statistical failure analysis of circuits under process variability (i) analysis in the performance domain (the failure probability is obtained by comparing the circuit operation metrics with desired performance constraints and determining the percentage of failing samples from the total number of samples) (ii) analysis in the parameter domain (a distinction is made between device parameters which lead to correct operation and device parameters which lead to failure and the failure probability is determined as the ratio between the hyper-volume occupied by the parameters in the failure region and the hyper-volume of the parameter domain). My work in this field is based on a parameter-domain analysis, while most approaches in the literature are Monte Carlo-based performance domain analysis. Nonetheless, some proposals existed for parameter domain analysis as well, such as the Most Probable Point Analysis. It consists in dividing the parameter domain into acceptance and failure regions and finding a particular point in the parameter domain that can be related to the probability of system failure, defined by a limit state (i.e. the Most Probable Point – MPP) (13). The failure probability estimation problem can be stated as finding the most likely point that causes the circuit to fail; that is, the point of maximum probability satisfying a certain failure criterion (MPP). The border between

the acceptance and failure regions is approximated by interpolation. An other proposal for statistical failure analysis in the parameter domain is the Yield Estimation Nonlinear Surface Sampling technique (YENSS). It was first presented by S. Srivastava and J. Roychowdhury in (14) and then improved by C. Gu and J. Roychowdhury in (15). The method locates the failure region boundary in the parameter domain and determines the failure probability as the ratio between the area (or volume) outside the bounded region and that of the parameter domain. The boundary points are determined by a local search algorithm. This technique assumes the partition of the parameters space into 2^N regions (N being the number of parameters, i.e. the dimension of the parameter domain), and uniform distribution of process parameters. My work on this topic is focused on the development of a methodology for SRAM failure probability estimation in the parameter domain, which, in contrast to other approaches, has tunable accuracy.

Figure 2.1illustrates the number of publications per year related to the SRAM cell robustness and yield metrics. It can be seen that the period where I worked on these topics was the period where they started gaining interest in the community. My research was in line with the community on the topic of Static Noise Margin, but disruptive on the topic of parametric yield analysis and dynamic noise margin study.

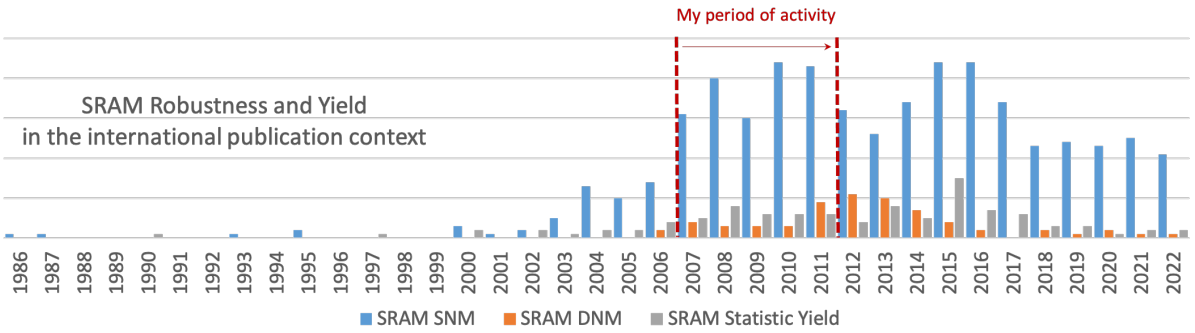


Figure 2.1: IEEE reserch on SRAM Noise Margin and Statistic Yield Estimation Metrics

Since fabrication induced process variability has a strong effect on the memory functionality, it follows that it has a strong influence on the memory behaviour under in the presence of fabrication-induced defects and/or cell upsets caused by radiation. *Building on previous work, I have analyzed and quantified the variability-induced deviations of the SRAM cell behaviour under the influence of resistive defects on one hand, and under the influence of particle strikes on the other hand. The former analysis has led to the development of an adaptive test technique, which accounts for the strength of variability affecting the device under test (work carried out at LIRMM Montpellier, France). This work is briefly described in subsections 2.3.3 and 2.3.4 of this document.*

Regarding IC radiation testing in general and SRAM in particular, the research community is divided in 3 main groups: radiation life testing; radiation accelerated testing - the main research efforts are focused here; simulation level radiation testing - my research focus. I have focused on simulation level testing since it allows the evaluation of novel technologies and the effect of process, voltage and temperature variations at very low cost and with acceptable precision. There are multiple studies dealing with the simulation level testing of radiation effects on electronic components. The response of a 6T SRAM cell to different current models induced by ions has been analysed in (16) by means of critical charge Q_{crit} characterization. Random variability effects on the critical charge of an SRAM cell have been investigated in (17) and (18). In (17) the random doppant fluctuation effects on the SRAM cell's Q_{crit} are

characterised besed on 3D models. In (18) the effect of random threshold voltage variation on the cell's Q_{crit} is evaluated by means of Monte Carlo SPICE simulations. The soft error rate (SER) of an SRAM cell to neutron induced transient currents modeled by MC-ORACLE tool is analysed in (19), published by my (then) reserch group. Building on this existing framework I have performed a comprehensive study SRAM reliability under under athmosferic neutron radiation by evaluating (by SPICE simulations) the rate at which soft errors occur in an SRAM cell when various reliability degradation factors are concomitently considered. In additio, continuing the effort of precise, realistic and fast estimation of the SRAM's stability we analise the possibility of using the classical SNM metric to evaluate the cell's resilience to soft errors in data retention mode under threshold voltage variations.

Figure 2.2illustrates the number of publications per year related to the measurement and evaluation of soft error rates affecting the SRAM memories. It can be seen that the period where I worked on this, the topic was already consolidated within the community. My research was in line with the community trends related to the modeling of SRAM SEU.

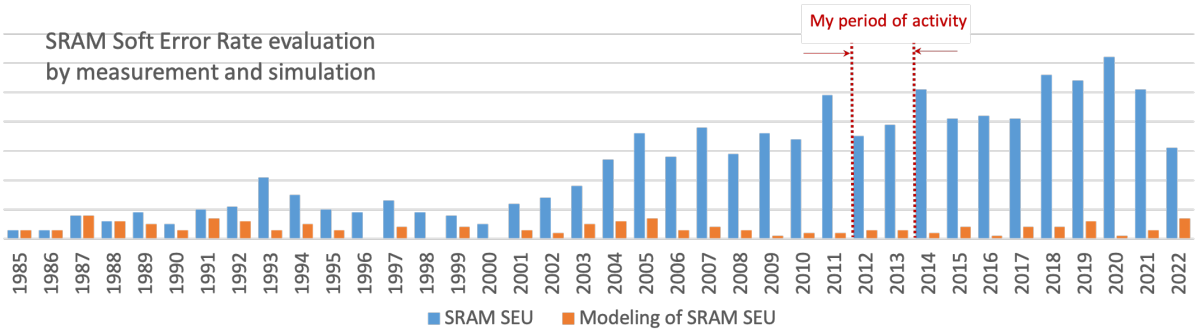


Figure 2.2: IEEE research on SRAM Soft Error Rate measurement and estimation

Regarding SRAM memory fault modeling and testing extensive research has been dedicated to the analysis of the SRAM cell affected by resistive-open defects in different technology nodes, stress conditions and process corners. Methods to test for these defects have been proposed, which show good fault coverage capabilities (20), (21), (22), (23), (24). However, with the continuous technology scaling, the effect of random variability has become more important than that of the systematic variability, especially in SRAMs. From the test point of view this variability leads to poor defect coverage capabilities. Another increasingly important issue when dealing with scaled technologies is the premature aging of the circuit. The aging affects both defect size and variability level. The presence of a resistive-open defect on an interconnect may induce a local increment of the heat (hot spot) due to the Joule heating effect. In such condition the phenomenon of electromigration takes place, and the resistive-open defect grows. Moreover, aging effects like Hot Carrier Injection (HCI), radiation induced damage and Bias Temperature Instability (BTI) can cause reliability degradation. In my work I focus on the analysis of permanent faults in SRAM arrays caused by resistive-open defects in the core-cell taking into account the random variability-induced threshold voltage mismatch between the transistors of the core-cell and IO circuitry. To the best of my knowledge, this was the first time a methodology for SRAM adaptive testing has been proposed, based on choosing the optimum bias conditions to maximize defect coverage with minimum over testing according to the process corner the memory is in after fabrication. This was a disruptive research, as no other publication up-to date was concerned with adapting the test strategy to the variability of the fabricated device.

According to the optimistic predictions on the technology scaling of the ITRS road-map the limits of consecrated memory technologies have become more and more evident. To overcome these issues, emerging memory technologies are being developed and implemented. Currently, the most promising emerging memory technologies are the magnetic type RAM, the resistive RAM and the Phase-change RAM. Nevertheless, being these technologies new, their characterization of reliability estimation is of utmost importance.

Building on my previous research experience, I have developed a methodology for memory reliability estimation focusing on the effects of process variability and aging phenomena at physical level. Based on a detailed and parameterized physical description of emerging memory technologies, I have performed an exhaustive study of the memory technology and have identified the resulting issues of the fabricated cell (work carried out at PoliTO, Torino, Italy). This work is briefly described in 2.3.5 of this document.

The Spin-Transfer Torque Magnetic Random Access Memory (STT-MRAM) offers read and write access times and compatibility with the CMOS process, however is facing a set of challenges that impact performance and reliability. These issues are mainly related to process variations of MOS and MTJ devices to the high write power consumption, and to the thermal fluctuations in the MTJ switching. Process variation effects on STT-MRAM cell have been extensively studied. For instance, (25) provides an overview of the interrelationship between design parameters and parametric failures of STT-MRAM cells in presence of process variations. The authors present a theoretical analysis for modeling the read and write failures of a cell due to the parametric variations of current pulse width and crosssection area. In (26), a magnetic- and electric- level STTMRAM cell model is proposed. The model is used to evaluate the effect of both thermal fluctuation and switching currents on the switching time of the device. The variations affecting the parameters of the NMOS transistor are not taken into consideration. Several circuit techniques have been proposed to improve the robustness of an STT-MRAM memory, like multi-terminal structures (27), new design paradigm decoupling conflicting design requirements between read stability and writability (25), or using complementary polarizers in the cell design for self-referencing and improved write current (28). Traditionally, the robustness of the STT-MRAM cell is expressed in terms of read stability and writability, concepts based on the operation current. In this context, I have proposed a methodology intended to support design and test engineers during predictive device simulation and memory characterization. Moreover, my methodology enables a significant understanding of design limits and can be integrated with the well-established statistical analyses to provide accurate information about the circuit limitations under process variability. We use the Robustness Margin metrics (RM) to determine whether a faulty operation can occur, to estimate maximum allowed process variability for correct operation, to extract current noise margins when the fabrication induced process variability is known, and to estimate the failure probability when the fabrication induced process variability causes faulty behavior.

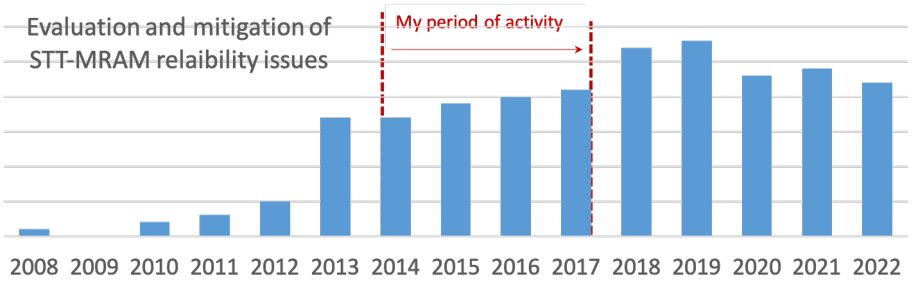


Figure 2.3: IEEE research on STT-MRAM reliability evaluation and mitigation

Figure 2.3illustrates the number of publications per year related to the reliability of STT-MRAM memories, both evaluation and mitigation. It can be seen that the period where I worked on this, the topic started to aquire recognition within the community. My research was in line with the community trends related to the modeling of STT-MRAM reliability.

2.2.2 Design and Evaluation of Hardware Security Primitives

Aside from reliability, an important aspect of device dependability is related to hardware security. Security systems use cryptographic protocols, frequently built on low-level cryptographic algorithms and primitives, such as Physically Unclonable Functions (PUFs) and True Random Number Generators (TRNGs). The Physically Unclonable Functions (PUFs) are emerging primitives exploited to implement low-cost authentication protocols and cryptographic primitives, such as secure key generators, key storing and one-way functions. PUFs exploit intrinsic manufacturing variability introduced in a device during the fabrication process to generate a signature, unique to each single device. One of the most investigated solutions uses SRAMs, since they provide high security (i.e., high inter-chip variation) and high stability (i.e., low intra-chip variation). Commercial devices and state-of-the-art studies exist for current SRAM CMOS technologies, however, they are not easily accessible for academia. *In this context, I have worked on the development of on open-access hardware platform based on STM32 micro-controllers, which allows the measurement and characterisation of embedded SRAMs to evaluate their usability for device identification. Also, I have worked on the reliability of classical PUFs and have proposed solutions for accurate reliability analysis and methods to mitigate well-known reliability issues, mostly focussing on SRAM and Ring Oscilator PUFs.* The rapid development of low power, high density, high performance SoCs has pushed the embedded memories to their limits and opened the field to the development of emerging memory technologies. The Spin-Transfer-Torque Magnetic Random Access Memory (STT-MRAM) has emerged as a promising choice for embedded memories due to its reduced read/write latency and high CMOS integration capability. *I have proposed an innovative PUF design based on STT-MRAM memory, which exploits the high variability affecting the electrical resistance of the Magnetic Tunnel Junction (MTJ) device in anti-parallel magnetization. The proposed solution is robust, unclonable and unpredictable. In parallel I have developed a solution for on-chip TRNG, based on STT-MRAM cells, which takes advantage of the stochastic-variations affecting the MTJ device.* I have started working on this topic while I was in PoliTO, Torino, Italy and have continued in TIMA. This work is briefly described in subsections 2.4.1 and 2.4.3 of this document.

Work has started in earnest to improve the reliability of generated SRAM-based PUF responses. The conventional method to improve PUF reliability is based on the use of error correction codes (ECC). Starting from the raw PUF response (i.e., the one obtained by reading every cells of the SRAM), an ECC is first calculated during the so-called enrollment phase. This code is then re-used as helper data to reconstruct the stable PUF response from the raw PUF response. These ECC blocks generally have significant area overheads, which scale quickly up as the number of correction bits increases (29). Further, they require the generation and handling of the helper data. In order to reduce and avoid the need of complex ECC mechanisms, various techniques have been proposed to intrinsically improve the reliability of the PUF. One proposal of increasing the PUF reliability is to use aging effects (NBTI - Negative Bias Temperature Instability) to change the circuit characteristics after manufacturing. By heating the device, the mismatch between the SRAM inverters is increased such that the difference in their electrical characteristics increases in magnitude, and hence strengthens the preference of the affected

cell for one of the stable states (30). However, this method requires a high temperature stress, which cannot be localized to the target cells, therefore aging the entire chip. To eliminate this problem, another method to improve PUF reliability has been proposed in (31) where Hot Carrier Injection (HCI) is used to reinforce the desired PUF response in short stress times without affecting the surrounding circuit. This method is based on a type of bi-stable element PUF that uses sense amplifiers (SA) as the core element. The offset of the SA is the indicator of PUF reliability. Post-manufacturing, HCI stress is applied to the SA to increase its offset, in this way assuring increased PUF reliability. Another approach to increase PUF response reliability is based on strategic selection of the bit-cells in the SRAM array to be used for PUF implementation (32), (33). In (32) the authors propose a strategy for selecting the words used to form the identifier in such a way that the bit flipping can be considerably reduced. In (32) they present a methodology to identify critical conditions for enrollment tests that can identify unreliable bits. Based on this, a bit-selection algorithm is developed that can assist with key selection from the reliable bits obtained at enrollment tests. These methods require multiple enrollment steps (in different environmental conditions) in order to identify the bits severely affected by noise. In contrast to these state-of-the art approaches we propose a novel and effective methodology to identify the unreliable bits in a SRAM-PUF based on a modified stability test strategy. When investigating more generaly the PUF reliability, Maes in (34) was among the first to demonstrate the the trade off between the PUF reliability and its entropy. Schaub et al. provide in (35) a generic probabilistic method for delay PUFs, where the trade off between reliability and entropy is modeled based on signal-to-noise ratio (SNR), and it is validated by real measurements.

One of the most investigated PUF solutions uses Static Random Access Memories (SRAMs), since they provide high security (i.e., high inter-chip variation) and high stability (i.e., low intra-chip variation). Commercial devices and state-of-the-art studies exist for current SRAM CMOS technologies. Nevertheless, very few studies exist for PUFs based on emerging memory technologies. There are some proposals exploiting resistive memories (36), (37), (38), and also magnetic memories (39), (40). The solution proposed in (39) uses a non-standard memory design, where two storage cells are used for one bit. The PUF implementation proposed in (40) takes advantage of the stochastic nature of Magnetic Tunnel Junction (MTJ) writing. This requires a very strict control of the enrollment phase and any variation in temperature or noise in the circuit, will affect the PUF reliability. On the other hand, in our proposal, we use nominal control signals, with large duration for write operation, to maximize the reproducibility of the operation by reducing, as much as possible, the effect of stochastic writing to assure a highly reliable PUF solution.

A true random number generator (TRNG) is built on an unpredictable stochastic process with high entropy. Common hardware TRNG implementations are based either on the exploitation of intrinsic noise or on the use of materials with special properties. The TRNGs based on thermal noise, jitter and meta-stability are today well-recognized and have real implementation solutions due to their CMOS compatibility. In contrast, the TRNGs based on materials with special properties are considered exotic and expensive solutions and are rarely used due to the lack of easy CMOS integration, even if they present better theoretical randomness. On special class of TRNGs combines the two main sources of randomness, i.e., intrinsic noise and materials with special characteristics. These TRNGs are designed using emerging memory technologies. TRNG solutions have been proposed based on the use of spin-transfer torque magnetic tunnel junctions (STT-MTJs) which take advantage of its intrinsic stochasticity. Indeed, the write operation has a stochastic behavior, i.e., the success of a write operation is a probabilistic phenomenon, due to the intrinsic thermal instability of all magnetic nanostructures.

This instability can be manipulated to set the write probability to 50%, thus generating random numbers. Solution proposed in (41) presents a simple STT-MTJ-based TRNG implementation, where a single MTJ is used. In this proposal, the authors have shown that if the MTJ is written with the 50% probability, a random number is generated, i.e., it behaves like a spintronic dice. This result validated by the work in (42), where the authors show measurement data of an MTJ device. However, this implementation solution is sensitive to environmental and fabrication variations. As a result, the generated numbers are random only under perfect conditions. To solve this issue, it is possible to tune the write operation in such a way that it offers 50% probability under all conditions. This can be done by monitoring the circuit operating conditions and sending feedback to the write driver. The solutions presented in (43), (44) feature the use of a closed-loop control of the MTJ write driver. In these solutions, basic randomness tests are performed on a limited set of responses and, according to the quality of the obtained randomness, a feedback is sent to the write driver. While these solutions prove useful, they still have issues such as high complexity of the write driver design, low efficiency of the randomness evaluation, and unavailability when extreme operating conditions are detected. Within my research I have tried to solve these issues.

Figure 2.4illustrates the number of publications per year related to the Phisical Unclonable Functions and True Random Number Generators, together with speciffic works focused on the STT-MRAM devices. I have started my reserch in this field whet the topic was getting embraced by the community and I am continuing working on it today, when the research is booming. My activities and results are in line with the community trends related to the design and reliebility of CMOS-based PUFs, but follow a disruptive direction when the the work is based on emerging technologies. Indeed, most of the hardware security primitive related research is focuses on CMOS circuits and not resistive devices, and, when looking from the resistive devices' perspective, most of the research is focused on their use as memories and little reserch is conducted towards alternate use such as security primitives.

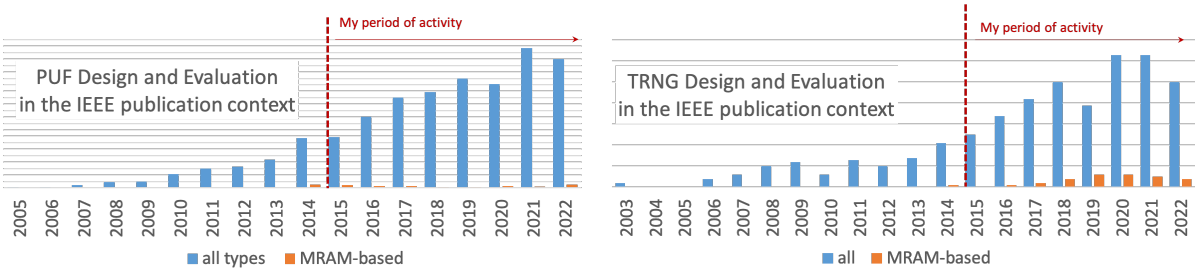


Figure 2.4: IEEE research on PUFs (on the left) and TRNG (on the right)

2.2.3 Design, Test and Security of Emerging Computing Paradigms

Since the appearance of modern computers, the widely adopted architecture has been based on the separation between the computing unit (or processor) and the memory storing the program to be executed and its data, i.e., the Von Neumann architecture. The separation between processor and memory has become an issue in modern computers due to the uneven evolution of processing speed and memory access times (also known as memory wall). With the technology advancements, the memory wall became increasingly important. Therefore, there is an urgent need to explore alternative architectures in the light of emerging non-volatile technologies (resistive memories), not only to further increase the computing efficiency at lower cost, but also to further reduce the overall energy. For example, moving the computation

to the memory (rather than doing it in the CPU) will significantly reduce the communication and therefore reduce the power consumption and increase the performance. This computing paradigm is referred to as Computation-in-Memory (CIM), an emerging concept based on the tight integration of traditionally separated memory elements and combinational circuits, that ensures the minimization of the time and the energy needed to move data across the processor. Computation-In-Memory (CIM) architecture, aims at eliminating the communication bottleneck while supporting massive parallelism. However, to achieve the ultimate objective of fully integrating the processing units and the memory in the same physical location, several technological challenges need to be overcome. There is a very wide variety of CIM solutions proposed today that exploit existing technologies. They enable logic and/or arithmetic operations directly inside the memory boundaries. The operations are performed without the need of transferring data to/from the CPU, thus saving time and energy. This can be achieved exploiting the physical characteristics of the memory and/or inserting computational elements in the peripheral logic (sense amplifiers). The research is mainly divided on levels of abstraction: (1) device level – to identify the optimum material combination to achieve the desired device behavior; (2) circuit level – to design logic gates and arithmetic circuits; (3) system level – to map applications on memory arrays, parallelize computations, design accelerators. *In this context, I have recently started working on the design and analysis of non-von Neumann architectures, such as neuromorphic computing and in-memory computing.*

Regarding the In-memory computing paradigm, research was directed towards identifying the main advantages and disadvantages of the existing CIM solutions exploiting traditional CMOS memories (such as SRAM) or beyond-CMOS devices (such as the memristive devices). This work is described in section 2.5. Regarding the neuromorphic computing paradigm, I have studied, for the first time, the behavior of fully-connected, spintronic-based spiking neural network under various sources of variability and unreliability. My study has shown important behavioral degradation of such neural network under hardware variability, proving the need of further research towards building dependable hardware-based neuromorphic architectures. In addition, I have also working on the hardware design of such neural network and have proposed design solutions for both functional entilies of a SNN: the spiking neuron and the synapse. This work is carried on at TIMA, Grenoble, France and it is described in section 2.6.

There is a very wide variety of CIM solutions proposed today that exploit existing technologies. They enable logic and/or arithmetic operations directly inside the memory boundaries. The operations are performed without the need of transferring data to/from the CPU, thus saving time and energy. This can be achieved by exploiting the physical characteristics of the memory and/or inserting computational elements in the peripheral logic (sense amplifiers). The research is mainly divided on levels of abstraction: (1) device level – to identify the optimum material combination to achieve the desired device behavior; (2) circuit level – to design logic gates (45), (46), (47), (48), (49) arithmetic circuits (50) neuromorphic computing synapse and/or neurons (51), (52), (53); (3) system level – to map applications on memory arrays, parallelize computations, design accelerators (54). In this context, my research activities are mainly on circuit level, where I am mainly interested in evaluating and demonstrating the advantages and shortcomings of the existing IMC solutions from the point of view of their performance, fault tolerance, ease of testing and security. While the circuit level investigation of IMC solutions is very present on the research community, the fault tolerance, testing and security issues are scarsely reaserched today as can be deduced from the IEEE publication record illustrated in Figure 2.6. In addition, even the few existing works are heavily focused on scouting logic or analog vector-matrix multiplication (where the result of the computation is obtained and manipulated outside of the memory array) (55), (56), (57), (58) and not on the solutions

that keep the result within the memory, as we do.

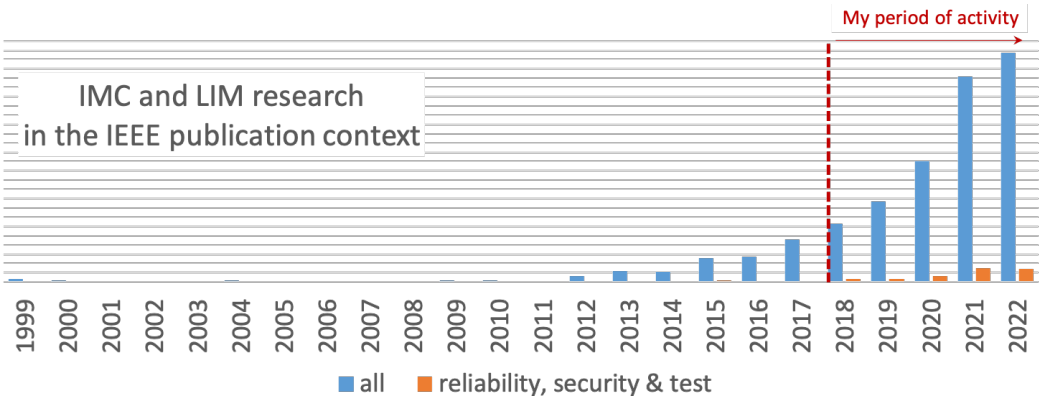


Figure 2.5: IEEE research on In-Memory Computing

The power and memory bottlenecks faced by today's computing architectures have motivated the research on Neuromorphic Computing systems. The ambition of Neuromorphic chips is to get closer to bio- inspired and even brain-inspired neuron and synapse computation models. Spiking Neural Networks (SNN) are an important class of bio-inspired computing paradigms offering promising solutions for on-chip cognitive applications. Leading projects in neuromorphic engineering that brought neuroscience and machine learning domains closer together have led to powerful brain-inspired chips able to simulate numerous spiking neurons to investigate new kinds of computer architectures (SyNAPSE (59), TrueNorth (60), Neurogrid (61), DYNAPs (62), Loihi (63), and Braindrop (64)), or to help neuroscientists through the Human Brain Project (SpiNNaker (65)). Products such as Loihi and SpiNNaker are using fully digital, core-based designs. Other proposals deal with mixed or even fully analog designs. More solutions are under study by academic research groups as a parallel effort. They focus on hybrid and heterogeneous architectures, targeting feedforward neural network but also more advanced models with less-well controlled learning rules (such as unsupervised or reinforcement learning, reservoir computing). In these cases, hardware architectures can rely either on formal coding to leverage the compatibility to well- known deep software frameworks or on spike frequency coding to reach better energy efficiency (66) and to allow local learning rules thanks to the bio-inspired Spike Time Dependent Plasticity (STDP). Preliminary solutions have shown promising results (67), (68), (69), (70). Hardware level solutions include explorations of state- of-the-art CMOS and emerging nanoelectronic technologies capable of mimicking the computational primitives of spiking neural networks where the most significant improvements come from the utilization of memristive arrays networks for synapse computation combined with CMOS implementations for neurons (71), (72), (73), (60). In this context, the neuromorphic computing paradigm has a huge potential when it makes use of emerging NV technologies (STT-MRAM, memristors), however, reliable and testable HW designs enabling the neuromorphic computing are still missing. In this context, my research is concerned with following research areas: (i) emerging memory technologies (memristors and spintronic devices) used in a non-Von Neumann context, (ii) hardware dependability (robustness, reliability and test) and design-for-dependability, (iii) hardware implementations of bio-inspired neural networks (Spiking Neural Networks). The first research topic is aligned with the international community, as it is today widely agreed that one of the best use cases for emerging technology is within neuromorphic computing. The third research area is also going with today's trend of looking for hardware alternatives to classical AI accelerators (mainly solving formal CNNs), mostly based on neuromorphic computing - which has gained a lot of interest in the last 10 years (see Figure 2.6). The

second reserch topic is somewhat more disruptive, since very few reserch groups are looking into the test and reliability of neuromorphic computing (see Figure 2.6-orange bars) and even fewer are concern with the systems capable of on-line learnig.

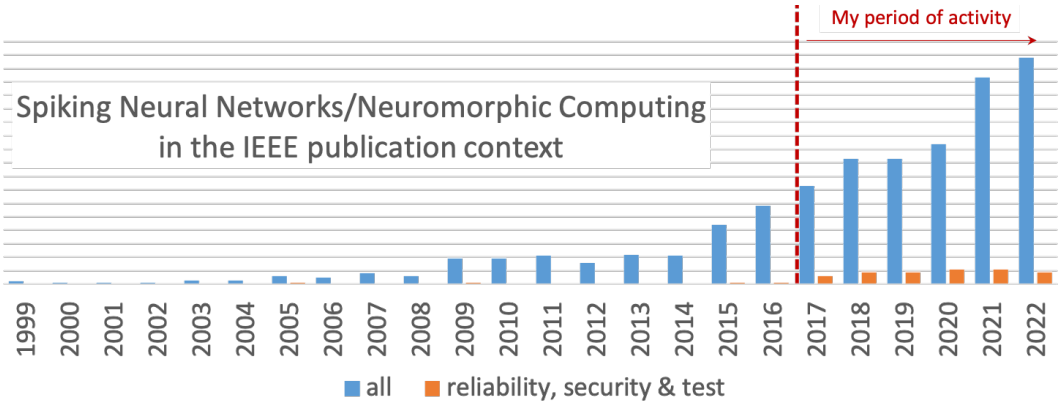


Figure 2.6: IEEE research on Neuromorphic Computing and Spiking Neural Networks

Figure 2.7 illustrates the diagram of the research topics I have been focusing on during my career up-to-date.

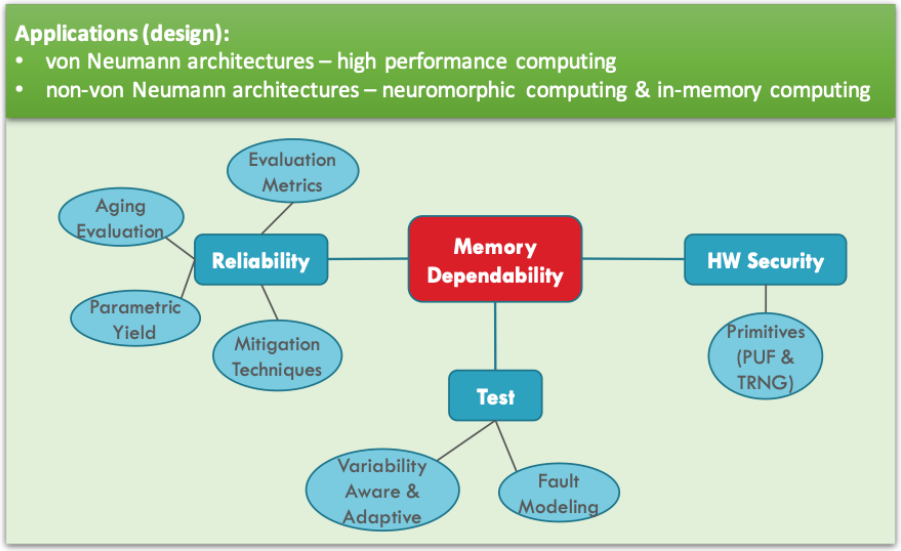


Figure 2.7: Research topics developed during my research career

The following sections of this document provide compact but comprehensive descriptions of my research activity over the course of the fifteen years I have worked in the field of micro- and nano- electronics, dealing with reliability, test and hardware security issues of current and emerging memory devices. It is not intended to present the totality of my published results, but a large subset which I consider as the most representative research results I have obtained during my career up-to-date.

2.3 Memory - Robustness, Reliability & Test

In this section I describe my activities related to the analysis of robustness and reliability issues in RAM memories as well as techniques for memory testing.

2.3.1 SRAM Robustness Metrics

This work was carried-out during my PhD, in the framework of the Spanish projects TEC2005-01027 and TEC 2010-18384, and it has been published in papers [J4], [J2], [J1], [C13], [C12], [C9], [C8], [C7], [C6], [C5], [C4], [W2] listed in Chapter 6. The work was focused on the definition of robustness metrics to characterize the behaviour of memory bit-cells under the effect of systematic and random process variability, voltage and temperature variations.

The main research objectives achieved are: (i) complete analysis of the main characteristics of the nanometric CMOS SRAM memories taking into consideration process, voltage and temperature variations; (ii) static and dynamic robustness analysis of the SRAM cell by means of static noise margin and dynamic noise margin; (iii) dynamic functionality analysis of the SRAM cell. My main original contributions in this area are: (1) new metrics for static robustness analysis, (2) new metrics for dynamic robustness analysis.

The SRAM robustness is defined as the maximum level of noise that can be tolerated by the cell when used in a system while still maintaining the correct operation. The robustness metric is called Noise Margin. Figure 2.8 illustrates (a) the basic structure of an SRAM bit-cell, who's internal nodes are perturbed by two voltage sources, (b) the voltage transfer characteristics of the 2 cross-coupled inverters of the core-cell in data retention mode, i.e., the butterfly curve, (c) the state space representation of the SRAM cell in data retention mode.

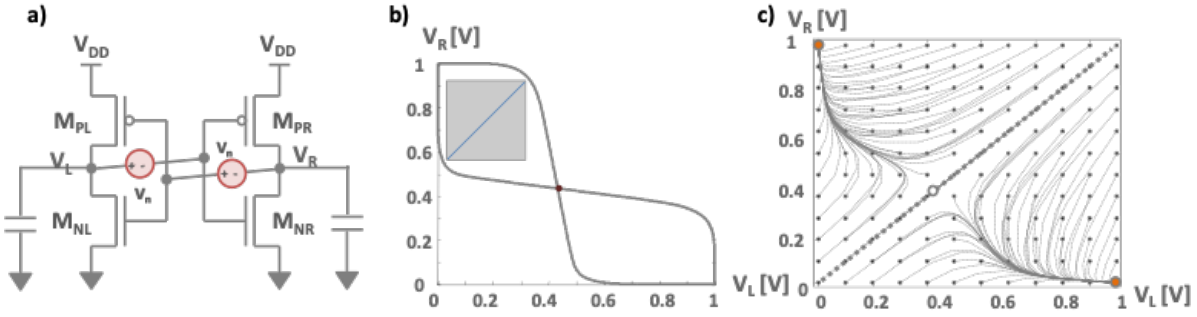


Figure 2.8: SRAM cell robustness analysis: a) SRAM core-cell in data retention mode with voltage noise disturbing its internal bodes; b) the "Butterfly Curve" of the SRAM core-cell; c) state space representation of an SRAM core-cell.

Metrics for SRAM Cell Static Robustness Analysis

Over the years, several techniques for statically analyzing the robustness of the SRAM cell have been developed. Among them, the Static Noise Margin, the Unity Gain Search, and the N-Curve are the most common. When referring to Static Noise Margin, a series-voltage noise is assumed. This, in the case of the 6T SRAM cell is translated as voltage noise sources at the inputs of the two crossed coupled inverters as shown in Fig. 2.8a). The worst case static noise is defined as the DC disturbance which is adversely present in all logic gates in an infinitely long chain of gates. Hence, the Static Noise Margin of the SRAM cell is defined as

the maximum amount of noise voltage that can be tolerated at the inputs of the cross-coupled inverters while the cell retains its data, assuming equal and opposite DC noise offsets.

As a result of my research activity, I have proposed two new strategies for static noise margin evaluation. The first strategy combines the classical negative slope approach with the maximum square criterion for an accurate estimation of the static robustness of the SRAM cell both in above- and sub- threshold operation. The second strategy proposes a geometrical estimation of upper and lower bound Static Noise Margin.

Modified Unity Slope Static Noise Margin Metric

Unlike other robustness evaluation methods, this methodology provides an analytical expression for Static Noise Margin estimation both in above- and in sub-threshold regimes. With this method it is relatively easy to integrate in the analysis the effect of systematic and random fabrication-induced process variations, voltage and temperature deviations. This method for robustness evaluation is fast (since it does not require electrical simulations) and it shows high accuracy when compared with results from HSPICE simulations. The models have as starting point the Kirchhoff's current low applied on the L and R nodes of the 6T SRAM memory cell, and the piece-wise linear approximation of the inverter voltage transfer characteristic (VTC), based on critical voltages.

Geometrical Estimation of Upper and Lower Bounds of Static Noise Margin

The voltage transfer characteristic (VTC) of an SRAM cell is a complex function which depends on the transistors characteristics. During my thesis, I have developed a geometrical upper/lower bound estimator of the static noise margin based on a piece-wise linear approximation of the inverter's VTC. These estimates based on the circuit characteristic voltages and VTC's slopes are used for SNM evaluation. This method combines electrical level simulation with analytical expression for fast and accurate robustness estimation. Approximating the VTC by segments, a piece-wise linear estimate is obtained. If the VTC is inside the contour obtained by the piece-wise approximation, an upper bound of the VTC is obtained. If on the other hand the contour obtained by piece-wise linear estimation is inside the VTC curve, a lower bound approximation is obtained. The SNM is obtained as a function of the characteristic voltages of the two cross coupled inverters (V_{IL}, V_{IH}, V_{DD}), their commutation threshold voltages (V_M) and their gains (g).

Proposed SNM metrics compared

Figure 2.9 shows the graphical representations of the SNM metrics described above. The static robustness of several SRAM cells has been determined using four approaches summarized in Figure 2.9 and the results are compared in Table **??**. The values of the SNM using the piece-wise linear approximation method and the proposed modified negative slope have been estimated by extracting the transistors' parameters using HSPICE and the current equations were solved in Matlab. The SNM estimation based on the maximum square criterion has been completed by means of HSPICE simulations. For the purpose of this comparison four SRAM cells were compared in 130nm, 90nm, 65nm and 45nm technology nodes. The predictive transistor models (PTM-bulk) were used.

The Static Noise Margin given by the maximum square approach is considered as reference, since no approximations or simplifications are made in its estimation. The compared metrics are:

$SNM_0 =$ *side of maximum square*

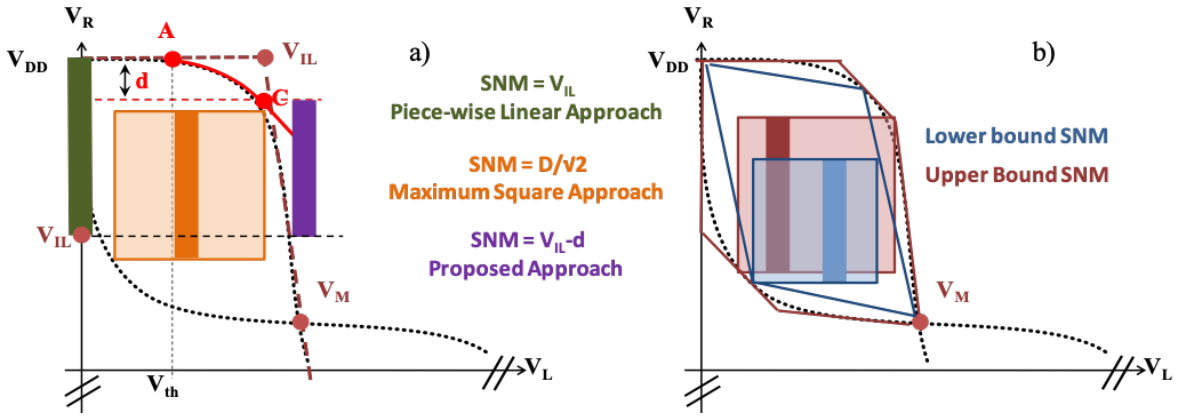


Figure 2.9: The SNM metrics compared: a) Maximum square, Piece-wise Linear Approximation and the Proposed Modified Negative Slope Criterion; b) Proposed Lower and Upper Bound Geometrical Estimation of SNM

$SNM_1 = V_{IL}$ obtained by piece-wise linear approximation

Proposed SNM_2 obtained with the modified negative unity slope approach

Proposed SNM_3 determined by means of geometrical estimation

The estimation errors for the proposed metrics when compared to the maximum square approach are determined using:

$$\epsilon_n = \left|\frac{SNM_n - SNM_0}{SNM_0}\right| 100; n = 1, 2, 3 \tag{2.1}$$

From Table 2.1 it can be seen that the proposed methods of SNM estimation are accurate when compared to SNM obtained by means of HSPICE simulations with higher accuracy than the simple piece-wise linear approximation. The accuracy difference between SNM_1 and SNM_2 is highly dependent on the cell's design. Actually, increasing the invertors' gain the accuracy increases in all three cases since for the ideal inverter $SNM_0 = SNM_1 = SNM_2 = SNM_3$.

Table 2.1: Static Noise Margin estimated using the Piece-wise linear approach, Maximum Square Approach, the Proposed Modified Negative Slope Approach and the Proposed Geometrical Estimation

Tech. [nm]	SNM_0 [V]	SNM_1 [V]	SNM_2 [V]	SNM_3 [V]	$\epsilon_1[\%]$	$\epsilon_2[\%]$	$\epsilon_3[\%]$
130	0.4178	0.6072	0.4322	0.4166	45.33	3.44	0.29
90	0.3843	0.5698	0.4032	0.3809	48.27	4.92	0.88
65	0.3525	0.5273	0.3681	0.3539	49.48	4.42	0.4
45	0.3261	0.4712	0.3375	0.328	44.49	3.49	0.58

The Modified Negative Unity Slope method which combines the negative unity slope approach with the maxim square criterion is used to evaluate the SNM of an SRAM cell in data retention mode both in above- and sub-threshold operations. The method has been proven to be accurate by comparison to HSPICE simulations (average relative error of 4%). This method

has been used to determine the critical supply voltages for various SRAM cells and the obtained results are also in good agreement with HSPICE simulations.

The second method is a geometrical method for upper and lower bound SNM estimation based on a piece-wise linear approximation of the inverter's voltage transfer characteristics. From the shape of the resulting piece-wise linear butterfly curve, different expressions for SNM estimation have been derived using the maximum square criterion. The lower and upper bound SNM estimators are obtained as functions of the critical voltages of the two cross coupled inverters. The average relative error in SNM estimation by means of the geometrical method is 0.54% when compared to HSPICE simulations improving state of the art estimators.

Metrics for SRAM Cell Dynamic Robustness Analysis

During my PhD program, I have developed two new strategies for dynamic robustness evaluation. The first strategy, proposes new *dynamic noise margin metric (minE-DNM)* for SRAM cell robustness analysis in data retention mode, assuming internal dynamic voltage noise sources disturbing cell. Simulation results show the use of the proposed metric as an indicator of cell robustness in the presence of transient voltage noise. The second strategy, evaluates the *functionality margins (Critical Access Time)* of the SRAM cell during data retention, read and write operation modes by means of phase-plane analysis. Dynamic Behavior of the SRAM Cell.

The noise margin is referred to as *Dynamic Noise Margin* if the cell is perturbed by a transient signal. Checking the robustness using dynamic noise margins requires time-dependent analysis. The spectral and time dependent properties of the specific noise patterns should be considered. When a transient noise of given amplitude affects a sensitive node of the SRAM cell, the bi-stable feedback-driven nature of the cell determines whether the noise will be filtered or will evolve to eventually flip the state. Stability analysis is concerned with identifying the minimum possible unintended violations that will destroy the data stored by the cell. In static stability analysis the noise margins are determined by identifying the maximum amplitude of a static voltage noise that can be tolerated by the cell without losing its data. This means that the noise may be present during an infinitely long time without flipping the cell's state. If the noise is a pulse, the noise amplitudes are allowed to be higher than the static margins without affecting the logic states.

Because of the nonlinear cross-coupling between the internal nodes of an SRAM cell, it is difficult to analytically characterize its dynamic behavior. Since there is usually no closed form analytical solution for a nonlinear differential equation, the phase space analysis is widely used in analyzing the circuit's behavior. A state space representation is a mathematical model of a physical system as a set of input, output and state variables related by differential equations. State space refers to the space whose axes are the state variables. The state of the system can be represented as a vector within that space. A phase space is a space in which all possible states of a system are represented, with each possible state of the system corresponding to one unique point in the phase space. Nonlinear systems often have multiple steady-state solutions. Phase space analysis of nonlinear systems provides an understanding of which steady-state solution that a particular system will converge to. A phase portrait is a geometric representation of the trajectories of a dynamical system in the phase plane.

Minimum Energy Dynamic Noise Margin Metric

In this approach, the noise perturbing the data stored by the cell is assumed to be a voltage noise which disturbs the internal nodes of the cell (as shown in Fig. 2.8a), similar to the static noise margin approach. The phase plane analysis (Fig. 2.8c) is used to characterize

the dynamic behaviour of the SRAM cell in data retention mode. The dynamic robustness metric (*minE-DNM*) of a SRAM cell in data retention mode is defined as the minimum energy of the voltage pulses able to flip the cell. The critical pulse width as a function of noise pulse amplitude is illustrated in Fig. 2.10a). The critical pulse width decreases exponentially with increasing the noise amplitude. For very large values of the noise the critical pulse asymptotically approaches zero. For small noises, close to SNM, the critical pulse width asymptotically approaches infinity. In signal processing, the energy E_S of a continuous-time signal $x(t)$ is defined as:

$$E_s = \langle x(t), x(t) \rangle = \int_{-\infty}^{\infty} |x(t)^2| \, dt \tag{2.2}$$

In this analysis, the continuous time signal whose energy is of interest is the voltage noise pulse with amplitude V_n and duration t_n. The energy of the noise signal is:

$$E = V_n^2 t_n \tag{2.3}$$

The energy which the critical noise pulse injects in the circuit can be expressed as:

$$E_{crit} = V_n^2 T_{crit} \tag{2.4}$$

Figure 2.10b) shows the dependence of the energy of the critical pulse as a function of its amplitude. For any cell, a similar energy curve is obtained to characterize its dynamic robustness. The dynamic robustness metric (*minE-DNM*) of a SRAM cell in data retention mode is defined as the minimum energy of the voltage pulses able to flip the cell.

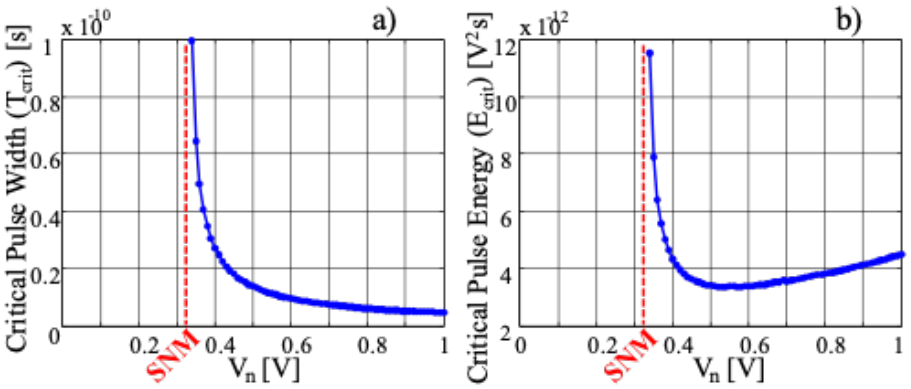


Figure 2.10: a) Critical pulse width vs. noise amplitude for dynamic stability; b) Critical pulse energy vs. noise amplitude for dynamic stability

The proposed Dynamic Noise Margin (*minE-DNM*) metric is compared with the Static Noise Margin (SNM) metric and with the Critical Charge metric in terms of SRAM relative robustness estimation. With this comparison the SRAM design space can be divided into radiation resilient cells (high critical charge), noise resilient cells and robust cells (which have high immunity to both current and voltage noises).

The three metrics used to evaluate the robustness of an SRAM cell (static noise margin, critical charge and the proposed dynamic noise margin) have been compared in order to assess their consistency over different cell designs (Fig. 2.11). In the white zones, all three metrics are in agreement with each other, i.e. the cells are less robust with respect to the reference

cell. The inconsistencies between metrics can be identified in the different zones resulting by the overlapping of the shaded areas (designs D_2 to D_6). For instance, it can be observed from Fig. 2.11 that design D_1 is less robust than the reference design (D_{ref}), under both static and dynamic disturbances, while design D_2 has a larger immunity to dynamic noise (both voltage at internal nodes and single event upset) than the reference cell, but is less robust under a DC voltage disturbance.

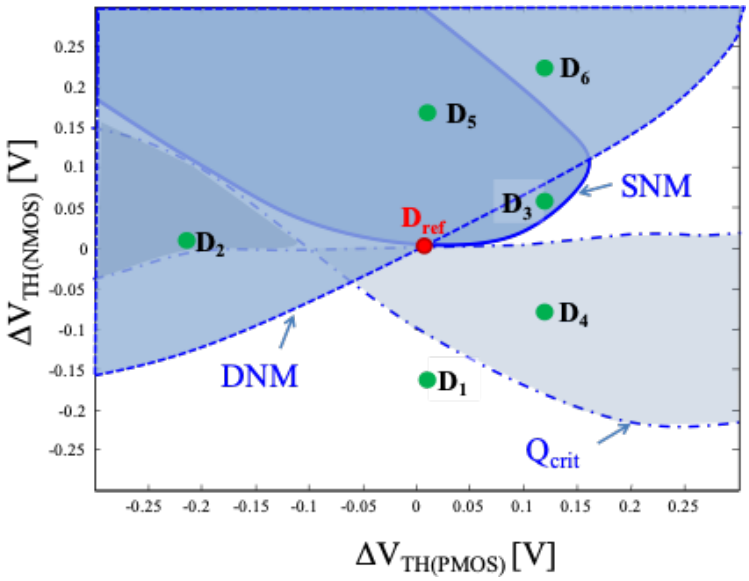


Figure 2.11: Comparison between the relative SNM, Qcrit and DNM metrics

Critical Access Time Functionality Metric

This approach is also based on phase plane analysis. However, no external noise source is assumed to affect the behaviour of the memory cell, the cell dynamic reliability is evaluated under the inherited disturbance of normal read and write operations. The proposed functionality metric is fully analytical, it is based on determining the cell equilibrium points and their regions of attraction. The proposed approach is based on the alpha power law MOSFET model (Sakurai – Newton) which takes into account the short channel modulation and the velocity saturation effects of carriers. It allows for dynamic reliability estimation for the nominal cell as well for integrating the effect of process, voltage and temperature variations. The method has been demonstrated to be fast and accurate when compared to HSPICE simulations.

The first step in analyzing the dynamic stability of the SRAM cell is to determine the equilibrium points and their regions of attraction. Determining the position of the two stable equilibrium points in the state space is straight forward as they are associated with the logic values of '0' and '1'. The meta-stable state (unstable saddle) is determined as the intersection point of the voltage transfer characteristics of the two crossed coupled inverters (Fig. 2.12a). Under nominal conditions, when the SRAM cell is symmetrical (identical inverters) the meta-stable point (V_M) coincides with the switching threshold (V_m) of the two inverters. However, under threshold voltage mismatch caused by process variations the SRAM is asymmetric and the two inverters have different switching thresholds. The switching threshold of an inverter is defined as the point on the VTC where the input voltage equals the output voltage. Under nominal conditions and symmetrical SRAM cell, the *separatrix* coincides with the 45° line – the first bisector, passing through the origin, the meta-stable point and the (V_{DD},V_{DD}) point. However, this is not the case under random threshold voltage variations. Due to the cell's asymmetry, the *separatrix* can deviate significantly from the first bisector. In this case the

separatrix will not pass through the origin but through point (V_{R0},V_{L0}) instead. It will neither pass through the (V_{DD},V_{DD}) but through (V_{R1},V_{L1}) –Fig. 2.12b).

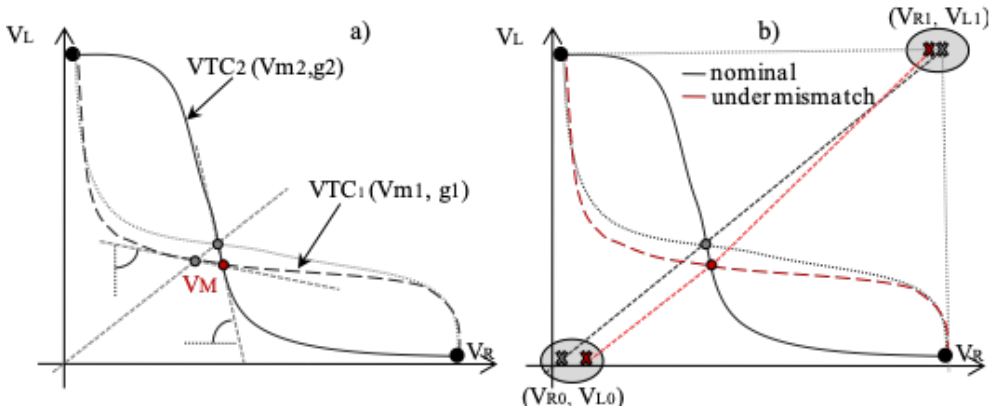


Figure 2.12: SRAM under threshold voltage mismatch: a) The Meta-stable Equilibrium Point; b) Separatrix approximation

Given a disturbing signal with a certain pulse width (T_w), the Dynamic Noise Margin is defined as the amplitude of the said signal for which $T_w = T_r$, with T_r the time needed by the system to reach the *separatrix* under the noise. If T_w is smaller than T_r the state trajectory will not reach the *separatrix*, remaining in the attraction region of its initial state and hence, no state-flip will occur – the SRAM cell is stable under read operation and unstable under write operation. If on the other hand, $T_w > T_r$, the state trajectory will cross the *separatrix* into the attraction region of the opposite stable equilibrium state, and state flip will occur – the SRAM cell is unstable under read operation and stable under write operation.

To validate the proposed mythology, the time needed for the cell's state to reach the separatrix is estimated by numerical integration under the disturbing influence of the access transistors during read and write operation assuming threshold voltage variability. The results obtained are illustrated in Fig. 2.13 for 10000 sample cells, assuming random threshold voltage variation.

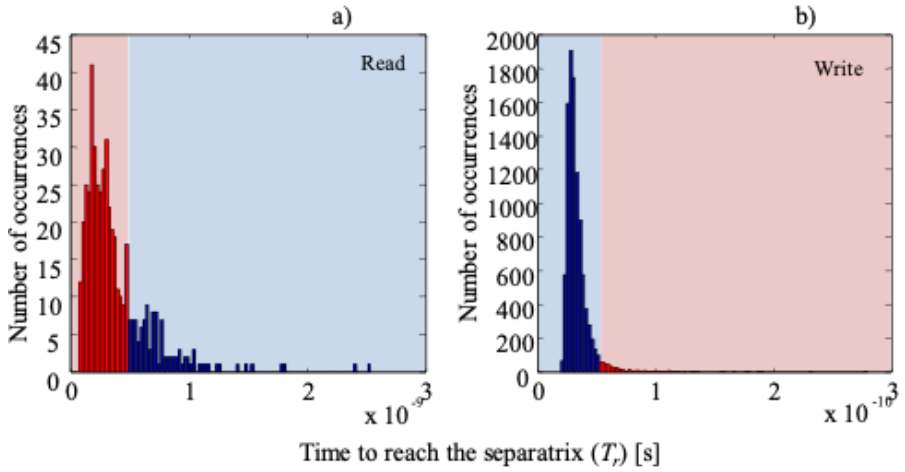


Figure 2.13: The time needed for the cells state to reach the separatrix a) in read operation mode, b) in write operation mode

This approach is based on a new analytical method for separatrix estimation used for functionality margin estimation. An SRAM cell under threshold voltage variability has been analyzed by means of static and dynamic functionality metrics. It has been observed that

for a sufficiently long disturbing signal (i.e., word line pulse width) the static and dynamic analyses give the same results. For shorter disturbing signal however (faster applications) the static analysis overestimates the probability of failure in read operation and underestimates the failure probability when the cell is written.

2.3.2 Methodology for Statistical SRAM Robustness Analysis

I have developed this work during my PhD program, in the framework of the Spanish projects TEC2005-01027 and TEC 2010-18384, and the European FP7 TRAMS project and it has been published in papers [C20], [C14], [C10] referenced in Chapter 6.

Satisfiability Boundary – Statistical Integration (SB-SI) Method

The essence of the proposed method consists in the decoupling between performance and statistics in the parameter domain. That is why it is hereafter referred to as SB-SI (*Satisfiability Boundary – Statistical Integration*). There is a range of values in the parameter domain for which the device satisfies the required performance. These values form the *Acceptance Region* while the remaining make up the *Rejection Region*. Robustness estimation is completed in two steps: first, the *Satisfiability Boundary* separating the Acceptance Region from the Rejection Region is found and second, a *Statistical Integration* over the two regions is performed to estimate probabilities.

Satisfiability Boundary (SB)

The Satisfiability Boundary is defined as the hyper-surface in the N dimensional space that separates the Acceptance Region (AR) from the Rejection Region (RR). The analysis is conduct in the parameter domain and the acceptance/rejection criterion is given by the circuit performance metric. Assuming a certain device with a parameter affected by process variability, its failure probability is given as the probability of not properly performing its specified function, while the correct operation of the device is reflected by performance satisfiability.

Assuming a circuit with N parameters ($p = [p_1 p_2 \dots p_N]$) affected by variability and whose performance metric must have larger values than a certain limit specified by $Perf(p) > P_{min}$. The two regions and the boundary between them are defined as follows:

$$AR = \{p | Perf(p) > P_{min}; RR = \{p | Perf(p) < P_{min}\} \tag{2.5}$$

$$SB = \{p | Perf(p) = P_{min} \tag{2.6}$$

where p is the N-dimensional vector of the process parameters. In the parameter domain, the AR and RR are N-dimensional hyper-volumes and the size SB is composed of N-1 dimensional hyper-surfaces. The Satisfiability Boundary search can be very expensive and time consuming if done by means of simulation. This is why I have proposed a method to estimate the SB by finding a set of boundary points (Significant Points – SP) and then interpolating to approximate the boundary surface with controllable error. To simplify the explanation of the method, we first consider a circuit with two (N = 2) parameters (p_1 and p_2) subject to variability (Fig. 2.14a). The figures in this subsection are obtained by analyzing a 6T SRAM cell in data retention mode and assuming that only two transistors parameters are affected by process variability.

In the parameter domain, the Satisfiability Boundary is approximated by a polygon whose vertices (hereafter referred to as *Significant Points (SP)*) are obtained by simulation. Evidently, the boundary is approximated more precisely with increasing the number of vertices. The first polygon obtained to approximate the Satisfiability Boundary in the 2D parameter domain is a quadrilateral. In this case, the SPs in this situation are determined by intersection between the SB and the parameter domain axes. If the obtained quadrilateral is not a good estimate for the boundary, an extra set of SPs will be added to the previous one. They are obtained at the intersection between the SB and the bisection lines (Fig. 2.14a) – *First Angular Bisection (FAB)*. For a more accurate approximation of the SB, more significant points can be found by further bisecting the resulting angles (Fig.. 2.15). For the two dimensional case the Significant Points (SPs) are determined by intersecting the satisfiability boundary with the segments given by the table in Fig. 2.14a. Assuming the parameter domain space defined by the parameter variation, the nominal parameter point ($p_{nom} = [p_{1nom}, p_{2nom}]$) defines the origin of the system axis, i.e. the [0 0] point. The maximum and minimum values on the two axes are given by $(+\Delta 1, -\Delta 1)$ and $(+\Delta 2, -\Delta 2)$ respectively.

Once the directions are established, the intersection points are found using a searching algorithm. For this particular case, the *bisection method* was chosen for its robustness and simplicity. This method is applied separately to find each Significant Point of the Satisfiability Boundary. The number of directions, and consequently the number of simulations grows with increasing the number of parameters, i.e. the dimension of the parameter domain.

In the 2D case, the Satisfiability Boundary can be approximated by the polygon obtained when applying an interpolation algorithm on the set of significant points (Fig. 2.14b). The first step to obtain the polygon is to find the adjacent points, which is intuitive for the two dimensional case but gets challenging for higher dimensions. The algorithm used in this work, which is extensible to the general N dimensional case, is described below for two dimensions.

In the *Quadrilateral Approximation (QA)*, the interpolation is straight forward, as the SPs are given by the intersection with the axes. For the first- and higher-order angular bisection approximations, the problem becomes more complex and an algorithm must be implemented. First, the extreme points in the parameter domain are mapped ($+\Delta$ values are mapped to 1 and $-\Delta$ values are mapped to -1), as shown in the table in Fig. 2.14b). Starting the search from the [1, 1] point, half of the adjacent points are found by decrementing 1 for each variable and the other half is found starting from the [-1,-1] point and incrementing 1 in each direction as shown in the diagram of Fig. 2.14b).

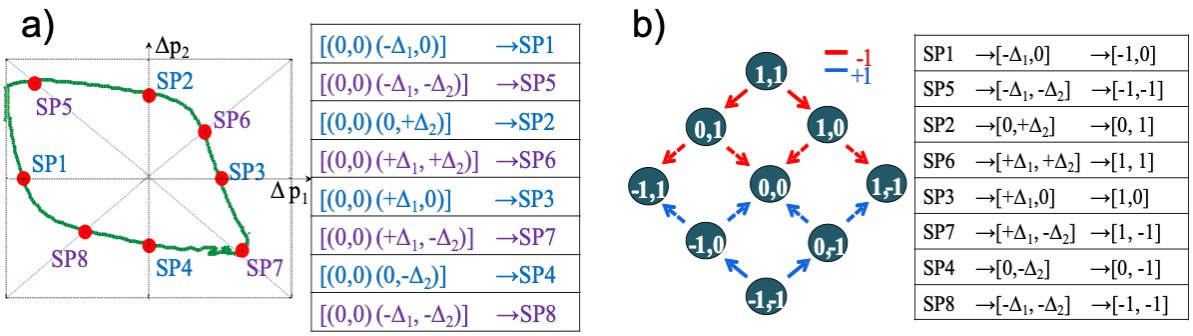


Figure 2.14: a) Choice of the Significant Points on the Satisfiability Boundary: First Angular Bisection (FAB); b) the adjacent Significant Points on the Satisfiability Boundary: First Angular Bisection (FAB)

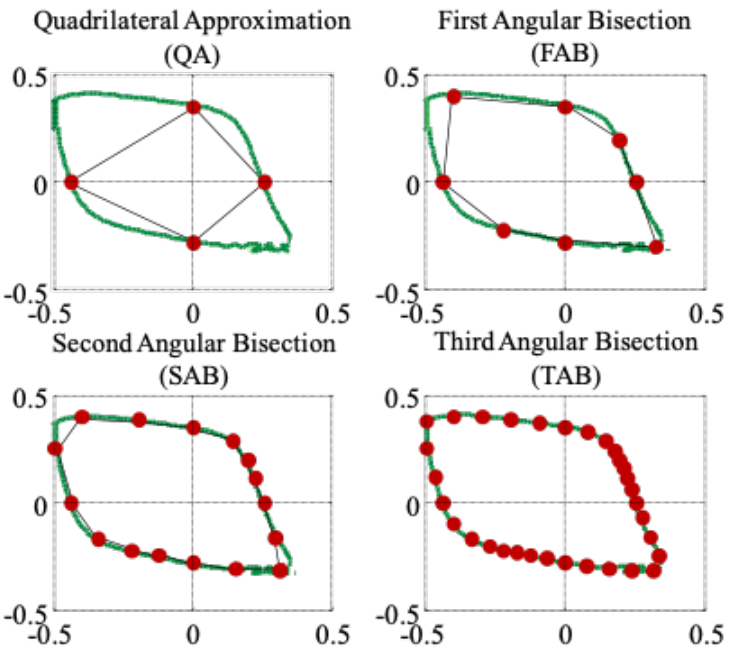


Figure 2.15: The polygonal estimation of the Satisfiability Boundary for different levels of angular bisection

Two points form a straight line. Thus, by connecting the adjacent points two by two the polygonal estimate of the Satisfiability Boundary is obtained (Fig. 2.14b). The general equation of a straight line is given by: $ax + by = 1$. The a and b coefficients for each of the straight lines forming the polygon are determined by solving the system:

$$\begin{bmatrix} x1 & y1 \\ x2 & y2 \end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{2.7}$$

where $(x1, y1)$ and $(x2, y2)$ are the coordinates of two of the adjacent points.

After obtaining the entire set of straight lines the boundary polygon (defined by the set of all a and b coefficients) is obtained (Fig. 2.15). The quadrilateral approximation (4 SPs), and the first (8 SPs), second (16 SPs) and third (32 SPs) angular bisections are illustrated in Fig. 2.15. An increase in the number of Significant Points results in improved accuracy of the boundary approximation but also in a larger number of simulations, that is, a longer estimation time.

Once the polygonal approximation of the SB is obtained, the acceptance and rejection regions are found. The condition that a random point in the parameter domain is in the acceptance or rejection region is given by

$$(x, y) \in AR \text{ if for all (a,b) } ax + by - 1 < 0 \tag{2.8}$$

$$(x, y) \in RR \text{ if for all (a,b) } ax + by - 1 > 0 \tag{2.9}$$

The method is extended to the general case of N parameters subject to variability. The significant points on the SB are determined analogously, using the bisection method. Once the significant points are found, the SB can be approximated by a hyper-surface. N by N adjacent points must be connected similarly to obtain the hyper-surface estimating the SB. Based on the above parameter domain partition, the robustness is determined by *Statistical Integration (SI)* as described below.

Statistical Integration (SI)

Assuming a multivariate distribution of N random variables, the joint (cumulative) distribution function is given by:

$$F(x_1, x_2, ..., x_N) = P(X_1 \leq x_1, X_2 \leq x_2, ..., X_N \leq x_N) \tag{2.10}$$

where $X_1, \ldots, X_N$ are the random variables under analysis and the probability density function is $f(X_1 \ldots X_N)$. The probability that the variables are between certain limits is given by:

$$P(x_1 \leq X_1 \leq y_1, x_2 \leq X_2 \leq y_2, ..., x_N \leq X_N \leq y_N) = \int_{y_1}^{x_1} ... \int_{y_N}^{x_N} f(X_1, ...X_N)\, dX_1 ... dX_N \tag{2.11}$$

The parameter domain is an N-dimensional hyper-rectangle. Given the complex shapes of the two regions (AR and RR), a computation strategy is required to determine the value of P in the equation given above.

As integration over a regular, well defined space is straight-forward and it is relatively easy to consider correlation, the parameter domain must be divided into hyper-rectangles. Three regions are obtained, i.e. Hyper-rectangle Acceptance Region (HAR), Hyper-rectangle Rejection Region (HRR) and Hyper-rectangle Satisfiability Boundary (HSB) as shown in Fig. 2.16 for the two dimensional case. The partition is performed using a bisection-like method. The left side of Fig. 2.16 illustrates the first step of space division: each edge of the initial rectangle is divided in half and by connecting the resulting points, four rectangles are obtained. In order to determine to which of the three regions these rectangles belong, the positions of the vertices are checked using (2.8) and (2.9). The rectangles in HAR and HRR are left untouched, whereas those in HSB are further divided (right side of Fig. 2.16) until the desired accuracy is reached.

- If all vertices are in the acceptance region, the rectangle (R) is in HAR.
- If all vertices are in the rejection region, the rectangle (R) is in HRR.
- If there are vertices both in the acceptance and rejection regions, the rectangle (R) is in HSB.

After completing the space division the statistical integration is performed by overlapping the parameter distribution of the parameter domain. By integrating the probability density function in each of the obtained rectangles using (2.11), the acceptance, rejection and boundary probabilities are obtained.

Spec Violation Metrics: Global and Operation Mode Contribution

The Spec. Violation Metric is an extension of the Satisfiability Boundary – Statistical Integration Method. It gives a fast algorithm to be implemented when comparing designs or the efficiency of variability mitigation circuit techniques. This method uses the boundary points of the Acceptance Region to decide on the cell robustness and reliability, without the need for statistical integration. The results are intended to offer a preliminary qualitative estimation of the cell robustness and not a quantitative one. This metric evaluates the localization of the acceptance region with respect to the maximum variability range in parameter variation

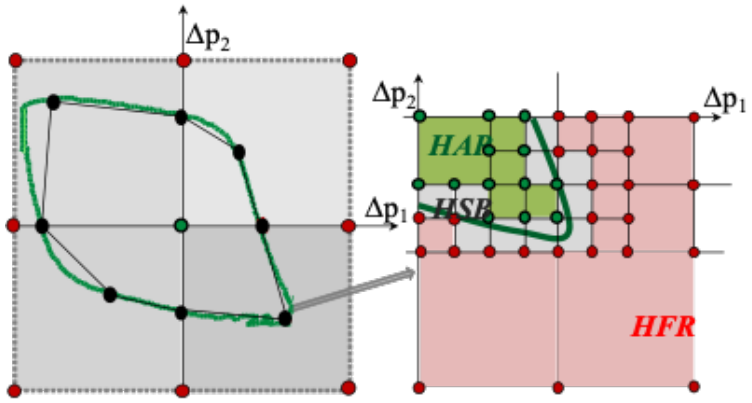


Figure 2.16: Rectangular division of the parameter domain

space, more precisely, how much of the AR is inside the maximum variability range. The Spec Violation Metric is maximum (=1) in the worst case scenario (AR inscribed in the maximum variability range) and zero in the best case scenario (AR circumscribes the maximum variability range).

Since the acceptance region is determined as the area of the parameter variation space delimited by the satisfiability boundary, the analysis can be performed by studying the position of this boundary with respect to the maximum variability range in the parameter variation space. Seeing as the satisfiability boundary is approximated by a polygon whose vertices are the significant points, the position of the AR with respect to the maximum variability range, is given by the localization of the significant points. The Spec Violation Metric is defined as the ration between the number of significant points inside the maximum variability range and the total number of significant points. In the example in Fig. 2.17b) 3 of the 8 significant points are inside the maximum variability range, hence SVM=3/8. Assuming the maximum variability to be 3σ, the Spec Violation Metric is defined as:

$$SVM = \frac{\#SP^{3\sigma}}{\#SP} \tag{2.12}$$

where SVM is the Spec Violation Metric, $\#SP^{3\sigma}$ is the number of significant points in the 3σ variability range, and $\#SP$ is the total number of significant points.

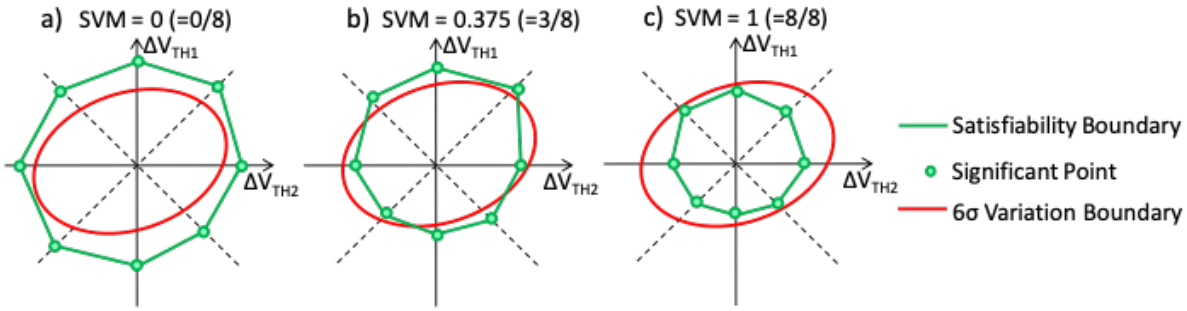


Figure 2.17: 2D Spec. Violation Metric (SVM) examples: a) best case scenario SVM = 0, b) SVM = 0.375, c) worst case scenario SVM = 1 (the total number of SPs in this illustration is 8)

In order for the device under analysis to be considered reliable under process variation, it has to be reliable in all operation modes. By evaluating in the parameter domain the functionality metrics which characterize the device the acceptance regions in all operation modes are determined. The region in the parameter space where all specifications are complied with is given by the intersection of these acceptance regions (as illustrated in Fig. 2.18a in a 2D representation). The resulting acceptance region is defined as the Overall Acceptance Region. This acceptance region is delimited by the overall satisfiability boundary. The significant points on the overall satisfiability boundary ($SP_{overall}$) are determined by identifying among the significant points (SP) the closest point to the origin of parameter space for each searching direction. The searching direction is indicated by $j = 1\ddot{o}J$, with J denoting the total number of directions (in the 2D representation in Fig. 2.17 and Fig. 2.18, J is 8). The module of the vector from the parameter space origin to the significant point $SPk(j)$ (where k denotes the operation mode satisfiability boundary to which the significant point belongs), obtained for the j-th searching direction is given by $|SPk(j)|$. The vector from the origin of the parameter space to the significant point $SP_{overall}$ is determined by:

$$|SP_{overall}| = min(|SPk(j)|) \tag{2.13}$$

Equation (2.13) is illustrated in Fig. 2.18b in a two dimensional representation. Moreover, for the proposed analysis methodology the points inside the 3σ variability range are of interest, so the overall significant point inside the maximum variability range for each searching direction is given by the set of significant points defined as:

$$\{SP^{3\sigma}_{overall}(j)\} = \{SP_{overall}(j) \text{ with the condition that } |SP_{overall}(j)| < 3\sigma\} \tag{2.14}$$

Each of these points belongs to one of the individual satisfiability boundaries as it can be observed from Fig. 2.18b. In this way, the operation mode that is more sensitive to random threshold voltage variation can be identified, by finding out which of the individual satisfiability boundaries has more significant points inside the maximum variability range (spec. violation points). The spec violation metric is defined for each operation mode: Operation Mode Contribution Metrics. The operation mode contribution Spec Violation Metric is defined as the number of points among the overall satisfiability points given by each of the operation modes divided by the total number of significant points.

$$SVM_k = \frac{\#\{SP^{3\sigma}_{overall} \cap SP_k\}}{\#SP_{overall}} \tag{2.15}$$

with k denoting the operation mode ($k = \{Mode1, Mode2, Mode3, Mode4\}$)

Statistical Failure Analysis Tool (SFAT)

A MATLAB tool has been developed, for straightforward and fast estimation of SRAM cell robustness and parametric yield under process variability. A schematic representation of SFAT and its applications is given in Fig. 2.19. The tool has been be used for different cell designs: 6T SRAM, 8T SRAM and 10T SRAM and 3T1D DRAM. Once the type of cell is chosen, the design parameters are defined and a static and dynamic robustness analysis is performed. If the results are not satisfactory the design parameters can be readjusted and another robustness evaluation performed. The tool allows for comparison between different cells in terms of

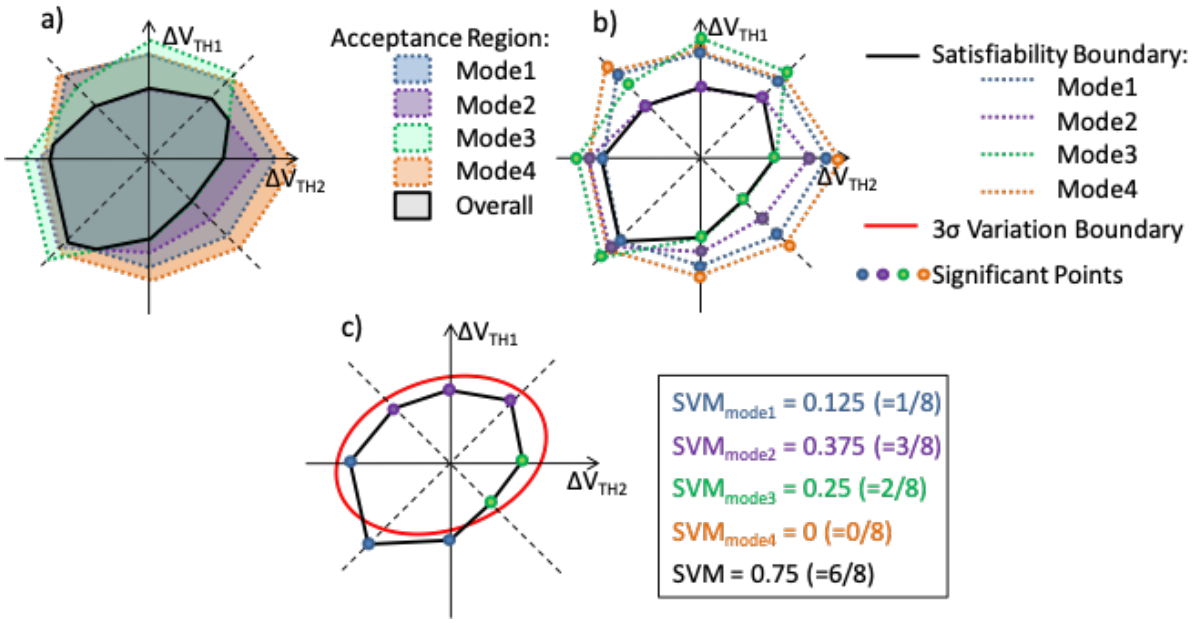


Figure 2.18: Global and Operation Mode Contribution SVM in 2D analysis: a) Definition of the Overall Acceptance Region, b) Estimation of the Overall Satisfiability Boundary, c) Estimation of the Operation Mode Contribution Metrics and Global SVM

robustness. Once the design and cell performance are decided, the process variability can be introduced in the analysis and by means of SB-SI method, the parametric yield determined for different variability scenarios in each operation mode.

Another feature of this tool is that it allows the incorporation of aging effects, assist techniques for performance improvement or techniques for variability mitigation and redundancy. There is no need of new simulations for evaluating the aging effects, since they imply variation in the nominal supply voltage.

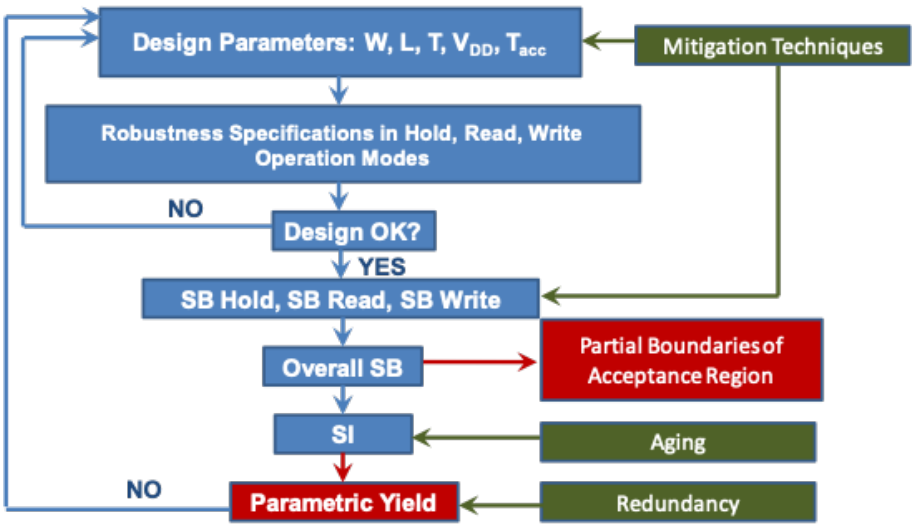


Figure 2.19: Schematic representation of the proposed Statistical Failure Analysis Tool (SFAT)

The effect of aging on the parametric robustness on an SRAM cell can be estimated using the same satisfiability boundary obtained for the 'fresh' cell and integrating using the new parametric distribution. The column/row redundancy can also be evaluated without need

of extra simulation if the assumption is made that a failing cell means a failing column/row and this failing column/row is replaced by a redundant one. In the case of assist techniques for performance improvement or techniques for variability mitigation and redundancy the Satisfiability Boundary needs to be estimated in different scenarios.

2.3.3 SRAM under SEU and Process Variability

I have developed this work during my post-doc at LIRMM Lab in Montpellier, France, in collaboration with L'Institute d'Electronique-UM Montpellier and it has been published in papers [C18], [C17], referenced in Chapter 6.

Device degradation due to technology scaling, has brought forth major issues related to process variation such as stability and reliability degradation that are especially problematic for the Static Random Access Memories (SRAM). An accurate and fast estimation of memory reliability is required to ensure its correct operation under extreme conditions. For this reason, several metrics have been proposed, such as the Static Noise Margin (SNM) to evaluate the stability of the SRAM cell under static noise; and the Soft Error Rate (SER), to evaluate the reliability of the memory under radiation. While accurate in predicting the memory reliability, the cell's SER estimation requires lengthy simulations for each cell configuration. On the other hand, a single simulation is necessary to estimate its SNM. For this reason, I have analysed the possibility of using the classical SNM as a first estimator of SRAM cell's reliability under neutron radiation while its transistors are affected by random threshold voltage (V_{th}) variation. I have carried out a study based on stability sensitivity analysis to V_{th} variations, which has led to a new way of evaluating the SNM metric. In this way, I have achieved high correlation between SNM and SER.

The stability of an industrial 40nm SRAM cell under random threshold voltage variability has been estimated, by means of SNM and SER. It has been shown that the two metrics are not correlated when considering V_{th} variations in all the transistors of the cell's loop - see Fig. 2.20.

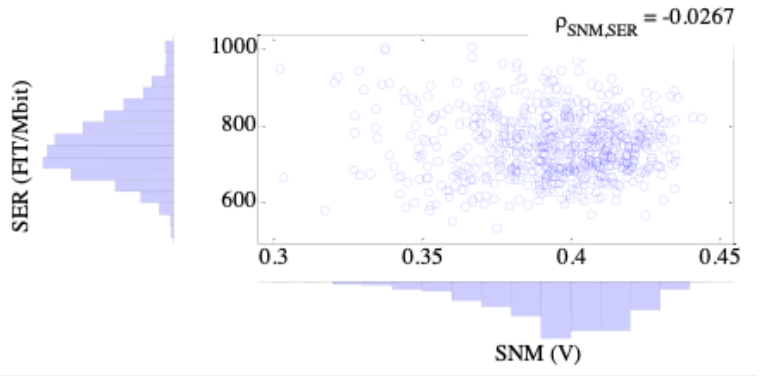


Figure 2.20: SNM and SER variations due to random threshold voltage variation affecting the SRAM's transistors

The sensitivity analysis shows that the Static Noise Margin is strongly affected by variations in the threshold voltages of the four transistors. On the other hand, the Soft Error Rate is mainly affected by changes in the threshold voltages of the two PMOS transistors when the upset affects the NMOS. This is a reason for the low correlation observed when the two metrics are compared. The way to obtain a better correlation between the two metrics is to differentiate between the static stability of the two states of the cell. As the particle strike occurs when the cell is in $< V_L = 0, V_R = V_{DD} >$ state, the SNM metric to be considered for correlation

analysis is SNMR. The histograms and the scatter plot obtained in this case are shown in Fig. 2.21. The Pearson coefficient is -0.8881.

As a result, using the described method, the SNM can be used as a first estimator of the SRAM cell's reliability to particle strikes under random threshold voltage variability, while the SER can be evaluated afterwards as way to refine the analysis.

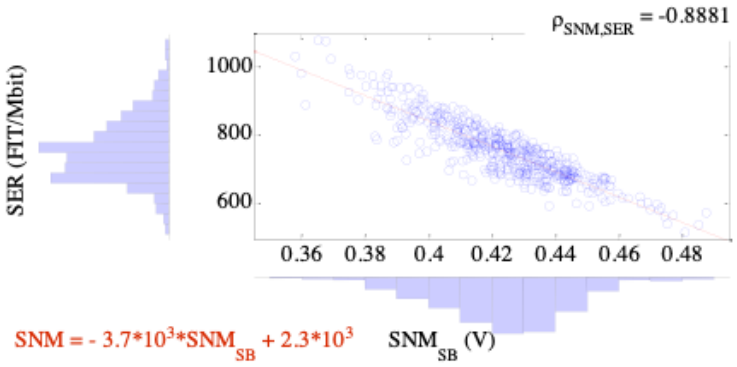


Figure 2.21: SNM_{SB} (SNM_R) and SER variations due to random threshold voltage variation affecting only the PMOS transistors of the SRAM cell.

2.3.4 Variability Aware and Adaptive STAM Test

I have developed this work during my post-doc at LIRMM Lab in Montpellier, France, in collaboration with Intel Mobile Communications (IMC) Lab in Sophia Antipolis, France, and it has been published in papers [C19], [C16], [C15] referenced in Chapter 6.

Functional operations of a Static Random Access Memory (SRAM) are strongly affected by random variability in core-cell transistors and by the variability-induced threshold voltage mismatch between the transistors of the Input-Output (IO) circuitry (especially Sense Amplifiers). This variability also affects the faulty behaviour of the SRAM array. In this context, my research was focussed on the analysis of static and dynamic faults due to resistive-open defects in the SRAM core-cell, taking into account the effects of random process variability in core-cells and IO circuitry. I have performed statistical analyses to evaluate the SRAM failure probabilities accounting for defects at each possible location. The results show that random process variability in the SRAM core-cell and IO circuitry have an important effect on the behaviour of an SRAM array and also on the defect coverage of various commonly-used test sequences. It is shown that under variability, the minimum defect size detected with maximum probability is more than 2X larger than the minimum size detected in nominal conditions, thus leaving a large range of defects undetected. Several stress conditions during test have been evaluated to assess their capability to increase the defect coverage under random process variability.

I have demonstrated that the failure probability when joint variability scenario (SRAM core-cell and sense amplifier) is considered is slightly smaller than the failure probability obtained when variability is assumed to affect only the core cell. This is due to the fact that the increasing the word line voltage improves the sense amplifier detection capability due to the resulting increase of the differential bit line voltage. The results show a higher increase of the cell's failure probability saturation value when the supply voltage is used as stress parameter during test. However, this too has its drawbacks much like the frequency tuning.

In order to improve the SRAM test reliability, the circuit has to be stress in such a way as to reduce the range of intermediate failure probabilities, reduce the maximum defect value for fault-free behavior. While design for yield techniques strive to reduce the effect of process variability on the functionality of the SRAM array, design for test techniques should aim to boost the variability effect while the SRAM is in test mode. In this way a narrower range of defects cause intermediate failure probably during test, providing more reliable results. The test environment has to be harsh enough to increase the defect coverage capabilities of the test sequence but not too much so as to reduce the fabrication yield.

Traditionally, bias conditions have been used to improve the behavior of the SRAM under process variations by applying body bias to compensate for the effect of variability. Based on the same principle, bias conditions also affect the cell's behavior when resistive-opens are present, hence affecting test's defect coverage capability. I have analyzed both body- and source-bias conditions to find the way to improve defect detectability in the SRAM cell. First, I have evaluated the minimum defect values detected in all operation modes for each possible location. Sorting the defects in ascending order leads to a better understanding of which of the operation conditions (fabrication – process corner and stress – temperature, supply voltage and operation frequency) has a stronger effect on the cell's sensitivity to resistive-opens. Since the defects are sensitized by either read or write operations, the stress and fabrication conditions can be classified in function of the operation mode they affect. By studying the obtained results several conclusions can be drawn. Testing under low temperatures leads to higher detectability of transition faults while testing under high temperatures leads to higher detectability of read faults when compared to room temperature testing. Testing under high supply voltage leads to higher detectability of read faults and transition faults when compare to testing under nominal supply voltage. The effect of operation frequency on the defect detectability cannot be clearly classified.

When the fabrication conditions are analyzed, a clearer classification can be done. The defect detectability in the Fast NMOS-Fast PMOS ('FF') process corner is larger than in the typical corner for faults sensitized by read and/or write operations. Reduced defect detectability is observed in the Slow NMOS-Slow PMOS ('SS') when compared to the typical corner for both read and transition faults. The process corner, in which the fabricated cell is in, cannot be controlled but it can be identified after fabrication, based on the circuit's characteristics. The biasing conditions can be adjusted in such a way that a cell fabricated in a certain process corner behaves similarly to a cell fabricated in a different process corner.

The results show that the body-bias does not bring an important improvement in resistive-open defect detectability. On the other hand, the data shows that applying source-bias to the cell's transistors during test has a strong effect on the minimum value of the detected defect. When using source-bias, the detectability of defects causing read failures is boosted when the NMOS transistors are forward source-biased and the PMOS transistors are reverse source-biased (i.e. fast NMOS and slow PMOS transistors). The detectability of defects causing write failures is boosted when the NMOS transistors and the PMOS transistors are forward source-biased (i.e. fast NMOS and fast PMOS transistors). These results are consistent with the ones observed when corner analysis is performed.

Based on the above considerations, we propose a new test strategy for detecting restive-open defects in the SRAM cell based on adjustable source-bias conditions. The test algorithm proposed is:

$$\Updownarrow (w0) \Uparrow ((r0)_{SBr}, (w1)_{SBw}, (r1)_{SBr}, (W0)_{SBw}) \Updownarrow (r0)$$

where SBr indices indicate that special source-bias conditions are used during read operation, while the SBw indices indicate that special source-bias conditions are used during write operation.

Using transistor source-bias during test leads to improved defect coverage. However, the same bias conditions adversely affect the functionality of the SRAM cell. That is why the functionality metrics of the cell have been analyzed together with defect detectability under different source-bias conditions. The functionality metrics considered in this analysis are the write time (Tw) and the zero level degradation (ZLD). We define the write time, as the time between the activation of the word line for write operation and the moment the cell flips its state. We define the zero level degradation as the maximum increase of the voltage at node storing zero due to the discharge of the corresponding bit line during read operation.

Due to electromigration effects, a resistive-open defect increases in time. Therefore, even if a fresh cell passes the test, the same cell, when aged, can fail due to an undetected defect, whose resistive value increased in time. This is when increased detectability is needed. However, tuning the test conditions to detect small value defects can lead to yield loss due to over testing.

In conclusion, two limitations have to be imposed to the source-bias conditions during test to avoid over testing: maximum 10% deterioration of the functionality metrics (so to avoid unwanted parametric fails) and $Em\%$ defect detectability improvement. The Em factor represents the electromigration effect on defect value and it can be obtained either using the existing analytical models, or statistically extrapolated from data obtained from previous tested chips.

The proposed algorithm is adaptable in two ways:

- From chip to chip: the bias conditions are chosen according to the process corner the circuit is in after fabrication;
- From batch to batch: the Em coefficient is adjusted according to collected data from the previous batch of tested chips and new source-bias conditions are obtained for each process corner.

By using the proposed test algorithm, and data collected from the chip under test and previously tested chips, an adaptive test algorithm can be devised. This adaptive testing is fast and guarantees that a chip which passed the test after fabrication, will continue to function correctly even under aging effects.

2.3.5 STT-MRAM reliability evaluation and boosting techniques

I have developed this work during my post-doc at PoliTO in Torino, Italy, partly in collaboration with UPC Barcelona, Spain, Universita di Ferrara, Italy, LIRMM, Montpellier, France, and Aix Marseille Université, France, and it has been published in papers [J7], [J6], [C36], [C28], [C27], [C26], [C24], [C23], [C22], [C21], [C20], [W7], [W4], [W3], [W2] referenced in Chapter 6.

In Spin-Transfer Torque Magnetic Random Access Memories (STT-MRAMs), information is stored into devices called Magnetic Tunneling Junction (MTJ). An MTJ is usually composed of two ferromagnetic layers separated by one oxide barrier. One ferromagnetic layer has a pinned magnetization direction (fixed ferromagnetic layer), set at fabrication time. The magnetization

direction of the second ferromagnetic layer is unpinned (free ferromagnetic layer), i.e., it has a freely rotating magnetic orientation that can be dynamically changed by forcing a sufficiently large spin polarized current through the device. The relative Magnetization Directions (MDs) of the two ferromagnetic layers defines the conductance of such a tunneling junction. This effect is called Tunneling Magnetoresistance (TMR) and it is characterized by the TMR ratio, defined as the relative resistance change between the two magnetized states. When the magnetization directions of the two layers are parallel, the MTJ device exhibits low electrical resistance (R_L), while when they are anti-parallel, the MTJ device exhibits high electrical resistance (R_H). The TMR ratio is, therefore, defined as: $TMR = (R_H - R_L)/R_L$. As shown in Fig. 2.22a) (for an MTJ with perpendicular magnetic anisotropy), the parallel and the anti-parallel relative magnetization directions are conventionally associated with the logic states '0' and '1', respectively.

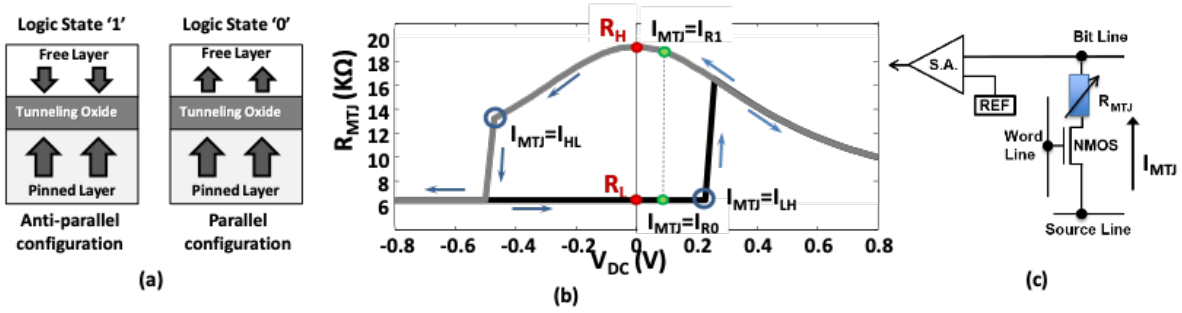


Figure 2.22: The STT-MRAM Memory cell: a) MTJ configurations; b) The $R_{MTJ} - V_{DC}$ hysteresis characteristic; c) Electric circuit of 1T1MTJ structure.

The electrical resistance of the MTJ device (R_{MTJ}) changes with the voltage drop across the device; the voltage-resistance behavior exhibits a hysteresis characteristic. Figure 2.22b illustrates the hysteresis characteristic of an MTJ device with perpendicular magnetic anisotropy. From this characteristic the electrical properties of the MTJ pillar can be extracted. The device resistance in parallel (R_L) and anti-parallel (R_H) magnetizations are given by the RMTJ values at zero-volt bias voltage ($V_{DC} = 0V$). From this, the TMR(0) at zero volt bias voltage can also be evaluated. From the same hysteresis characteristic, the switching conditions of the relative magnetization can be extracted. When there is a positive voltage drop ($V_{DC} > 0V$) on the MTJ device in parallel magnetization (i.e., the free ferromagnetic layer is at high voltage, while the pinned ferromagnetic layer is at low voltage), a current flow is generated (I_{MTJ}) through the device. When this voltage drop is large enough for the current to reach the switching threshold value, the magnetization direction of the free ferromagnetic layer is switched. The relative magnetizations are now anti-parallel. A similar situation is encountered when the MTJ initially in anti-parallel magnetization is negatively biased. The switching conditions for the magnetic states are indicated with blue circles in Fig. 2.22b). Here, I_{HL} represents the switching threshold current from anti-parallel to parallel state, while I_{HL} represents the switching threshold current from parallel to anti-parallel state.

Several STT-MRAM cell implementations have been proposed. In this work we target the popular 1T1MTJ structure. In this topology, the memory cell consists of one MTJ device connected to one NMOS transistor in series. The cell is accessed by the corresponding control lines, i.e., Bit Line (BL), Source Line (SL) and Word Line (WL). The equivalent electric circuit is provided in Fig. 2.22c.

An STT-MRAM cell can fail due to an unsuccessful write operation (write failure – WF), a destructive read operation (read disturb – RD) or a wrong decision during the read operation

(read failure – RF), or due to spontaneous magnetic direction flip during data retention (data retention failure – DRF). In this work, I have assumed the probability of data retention failure to be insignificant, since the analysis and the proposed methodology are focused on a STT-MRAM cell designed with high thermal stability factor.

Fault modeling and robustness metrics for STT-MRAM evaluation

Since the MTJ is in fact the data storing device, and the NMOS transistor acts as the control device, we first focus on the analysis of the constraints to be imposed on the R_{MTJ} for guaranteeing correct cell operation. Once this analysis is concluded, we evaluate the constraints on the NMOS threshold voltage variations as well.

In the case of a writing operation, the current flowing through MTJ has to be large enough, and of sufficiently long duration to allow the switching of the magnetization direction of the free ferromagnetic layer. To allow for a correct write operation (sufficient current) the high and low values of the MTJ electrical resistance must be below R_{HMAX-W} and R_{LMAX-W}, respectively. The region in the MTJ resistance space in which the R_H values are higher than R_{HMAX-W} indicates incorrect low state writing operation (write '0' fault: $W0F$), while the one in which R_L values are higher than R_{LMAX-W} indicates incorrect high state writing action (write '1' fault: $W1F$). Furthermore, (R_L, R_H) pairs must respect that $R_H > R_L$ ($TMR > 0\%$). All these boundaries are illustrated in Fig. 2.23a in the two dimensional space of the MTJ resistance. During read operation the current flowing through the cell (I_R) is compared with a reference value (I_{REF}). The reference current is assumed ideal and equal to the average current flowing through two ideal cells in complementary states, biased for read operation. If $I_R < I_{REF}$, the state is read as '1', i.e., the MTJ is in its anti-parallel state, $R_{MTJ} = R_H$. In this case, R_H must be high enough ($> R_{MIN-R}$) for the current condition to be satisfied (otherwise a read '1' fault occurs: $R1F$ in Fig. 2.23a). If $I_R > I_{REF}$, the state is read as '0', i.e., the MTJ is in its parallel state ($R_{MTJ} = R_L$). In this case, RL must be small enough ($< R_{MAX-R}$) for the current condition to be satisfied (otherwise a read '0' fault occurs: $R0F$ in Fig. 2.23a).

In order to assure sufficient separation between the IR currents for read '0' and read '1' operations, the MTJ should have a TMR ratio of 100% or higher (TMR=1 in Fig. 2.23a). TMR values lower than 100% do not imply read failure, just devices sensitive to perturbations, sense amplifier variability, and so on.

The union of faulty write and faulty read operation regions (red regions in Fig. 2.23a) represents the overall failure region for the cell under analysis, while the remainder of the parameter space represents the acceptance region (green region in Fig. 2.23a), the region in the MTJ resistance space, where the cell operates correctly.

The NMOS transistor also suffers from process variations mainly affecting its threshold voltage. For a more comprehensive characterization of the cell failure mechanisms in the parameter space, a 3rd dimension is added (for the V_{TH}) as depicted in Fig. 2.23b. The discussion on the RMTJ does not change; the acceptance region is bounded by the same constraints. However, the values of these constraints (i.e., R_{HMAX-W}, R_{LMAX_W}, R_{HMAX-R}, R_{LMAX_R}) are dependent on the driving capability of the NMOS transistor. A low value V_{TH} means higher driving capability of the NMOS, which translates into relaxation of write operation constraints (R_{HMAX-W} and R_{LMAX-W}) and also read operation constrains (R_{HMAX_R} and R_{LMAX_R}). This leads to a larger acceptance region (as shown in Fig. 2.23b bottom cross-

section). The situation is reversed when the NMOS threshold voltage is large (see Fig. 2.23b upper cross-section).

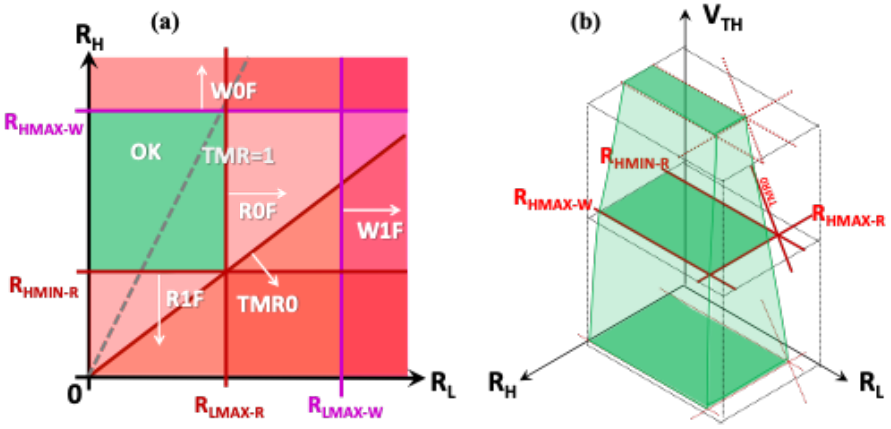


Figure 2.23: a) 2D illustration of failure mechanisms constraints during read and write operations of the 1T-1MTJ STT-MRAM. Here $W0F$ represents the write '0' failure region, $W1F$ represents the write '1' failure region, $R0F$ represents the read '0' failure region, $R1F$ represents the read '1' failure region, $TMR0$ represents the region where $TMR < 0$, and OK represents the NO failure region, i.e., the acceptance region; b) 3D representation of acceptance region in the (R_L, R_H, V_{TH}) parameter space. Three 'slices' are emphasized: the middle one corresponding to nominal value for V_{TH}, while the top and bottom ones correspond to V_{TH-MAX} and V_{TH-MIN}, respectively.

The robustness of an STT-MRAM cell is defined as the degree to which the correct operation conditions are satisfied. In this section we introduce and describe our proposed metrics for robustness evaluation (Robustness Margins - RM). The Robustness Margin metric is evaluated in the MTJ electrical resistance space. We define the Robustness Margin of one STT-MRAM cell as the minimum distance between its resistance value and the corresponding resistance margin for read and write operation, respectively (Fig. 2.24a and Fig. 2.24b, respectively). If the Robustness Margins take positive values, the cell can be correctly written and read from, while a negative RM signals a failing operation.

The RM defines the variability margins of a memory cell. Evaluated at designed time, the cell Robustness Margin provides an estimate of the maximum allowed process variability (in terms of MTJ electrical resistance) for assured robustness after fabrication, i.e., the maximum allowed parameter deviation from its nominal value.

If the fabrication induced variability of the two resistances is known, the Robustness Margins are evaluated as the minimum distance between the maximum resistance variation point and the first resistance value causing a failure for each operation (Fig. 2.24c). The RM can be used to define the noise margins of a memory cell, i.e., the maximum current noise tolerated by a cell affected by process variability, before a failure is observed.

If one or all RM take negative values, we estimate the failure probability under fabrication induced process variability for each operation mode with negative RM and the total cell failure probability. We say that a memory operation if faulty if its corresponding Robustness Margin takes negative values. In this case, the failure probability (fP) is estimated by integrating the probability distribution function defining the variability of the MTJ resistance over the area belonging to the failure region (Fig. 2.24d).

$$fP_{RF\&WFP} = 1 - \int_0^{min(R_{LMAX-R}, R_{LMAX-W})} \int_{R_{HMAX-R}}^{R_{HMAX-W}} \int_{V_{THmin}}^{V_{THmax}} f(R_L, R_H, V_{TH})\, dR_L dR_H dV_{TH} \tag{2.16}$$

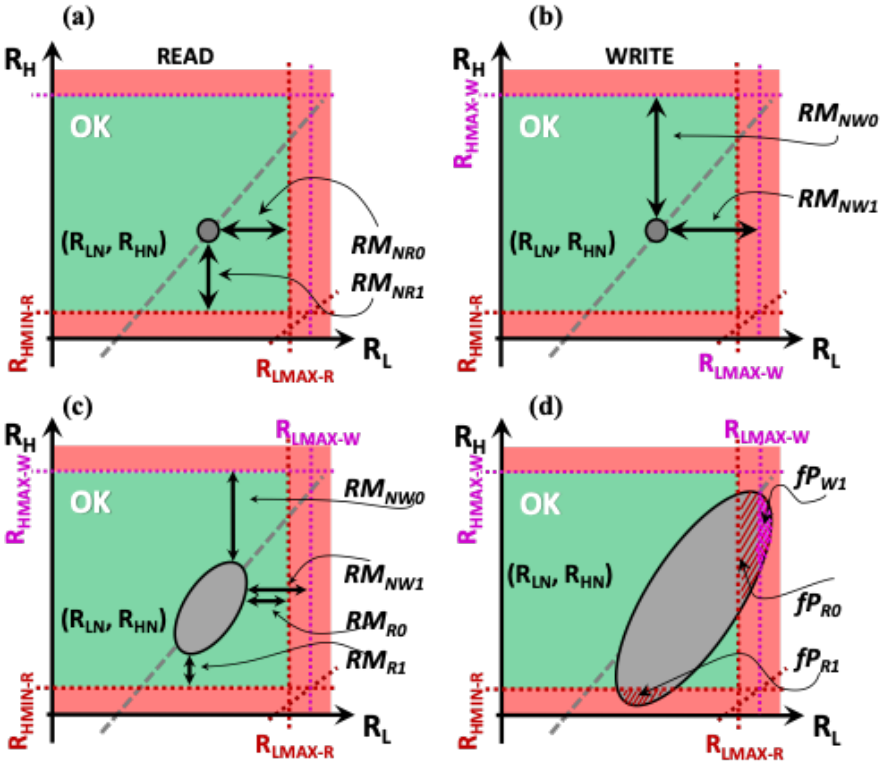


Figure 2.24: Graphical representation of the proposed Robustness Margin metrics (RM): a) RMs defined for read 0 and read 1 operation, respectively; b) RMs defined for write 0 and write 1 operation, respectively, c) RMs defined for read 0, read 1, write 0 and write1 operations no failures occur when the distributions of the MTJ electrical resistances are known, d) Failure probability estimation when some of the RM metrics take negative values - the distributions of the MTJ electrical resistances are known.

STT-MRAM cell reliability to aging

The main mechanism of the MJT degradation is the breakdown phenomenon. At each write operation the tunnel barrier is exposed to an electrical stress which might cause an electrical breakdown. During the anti-parallel to parallel write operation, the MTJ is subjected to a larger voltage stress then in the opposite write operation. For this reason, the analysis is focused on the degradation of R_H. For this analysis, I have used the percolation model to characterize the time-dependent dielectric breakdown. In this approach, the dielectric material is modelled as a large number of parallel conducting paths, whose formation follow a Weibull distribution.

Building on these hypotheses, we have evaluated the failure probability of the STT-MRAM cell under study at fabrication time, i.e. the fresh cell ($P_{RF\&WF}(0)$) and under repeated write stress, i.e. the aged cell ($P_{RF\&WF}(t)$).

To demonstrate the effect of the access transistor threshold voltage variation on the reliability of the STT-MRAM cell, the failure probability of the fresh cell due to incorrect read and write

operations has been estimated assuming discrete values for V_{TH}. The results are shown in Fig. 2.25a. As expected, the failure probability ($P_{RF\&WF}(0)$) increases as the V_{TH}, due to the resulting decrease in driving current. If the threshold voltage is by only 50mV larger than its nominal value, the failure probability decreases to about 10^{-3}, which is unacceptable for a cell in a memory array. The reliability further worsens for larger positive deviations of V_{TH}.

The reliability of the cell is obtained by statistically integrating the joint probability density function of the electrical parameters of the cell after different cumulative stress periods. By evaluating (2.16) the reliability curve is obtained and shown in Fig. 2.25b. We observed that the cell reliability degradation is almost insignificant during a large number of operation cycles (approximately 10^{16} under our assumptions) and then it is fast falling to zero. The cell reliability degradation in time is governed by the MTJ time-depended degradation.

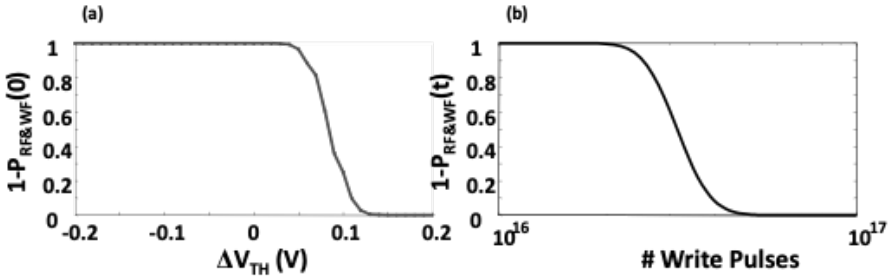


Figure 2.25: STT-MRAM cell reliability estimation under nominal values of the control voltages: a) reliability of the fresh cell affected by random variations in the MTJ resistance values; b) reliability degradation in time due to repetitive write stress, estimated assuming random variability of the MTJ resistance and NMOS threshold voltage (x-axis in logarithmic scale)

Control Voltage Influence on the STT-MRAM Reliability

In order to change the magnetic state of the STT-MRAM cell, there must be sufficient current (I_{MTJ}) flowing through the MTJ element to be able to switch the magnetic orientation of the free-layer. During a write operation, the power supply voltage (V_{DD}) is applied to the Word Line (WL) and sets the voltage drop between Bit Line (BL) and Source Line (SL). During the read operation, a small voltage drop is applied between Bit Line (BL) and Source Line (SL).

$$V_{WL} = V_{DD}$$
$$\mathrm{V}_{BL-SL} = V_{DD} \text{ for W0 operation}$$
$$\mathrm{V}_{BL-SL} = -V_{DD} \text{ for W1 operation}$$
$$\mathrm{V}_{BL-SL} = 0.3V_{DD} \text{ for R0 \& R1 operation}$$

I have estimated the reliability of the STT-MRAM cell under test assuming different values for the control voltages. A first analysis is performed by estimating the cell reliability under supply voltage (V_{DD}) variation. This translates into a variation of the MTJ current caused by two joint effects: i) different voltage drop (V_{BL-SL}) on the same resistance; ii) different gate voltage on the same NMOS transistor. The failure probability of the fresh cell due to incorrect read and write operations has been estimated assuming discreet values for V_{TH} (results shown in Fig. 2.26a). It has been noted that the cell failure probability decreases

with increasing the supply voltage and it is widely spread across different threshold voltages. The effect of threshold voltage variation is more pronounced at lower supply voltages. The same decrease in cell reliability with supply voltage scaling has been observed when the joint effects of resistance and V_{TH} variations are considered. In Fig. 2.26b, the cell reliability curves obtained after different cumulative stress periods are shown. For the fresh cell we observe a monotonic decrease of the failure probability when the supply voltage increases. This was to be expected, since larger V_{DD} means larger MTJ current, hence larger read and write capabilities. However, close to the breakdown point (at and beyond the endurance limit, $t \geq 10^{16}$ write cycles) and beyond the failure probability is not monotonic with supply voltage variations. We observe an increase in reliability up to a certain point, after which the reliability decreases. This reliability decrease is mainly due to the added stress exerted on the tunnel junction, which has a detrimental effect on the cell endurance to write operations.

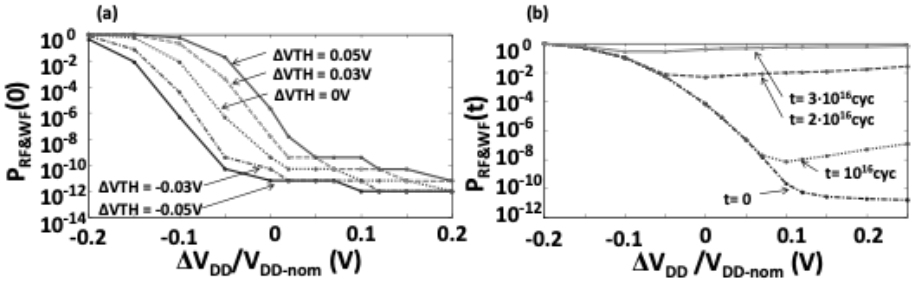


Figure 2.26: STT-MRAM cell reliability estimation under supply voltage (VDD) variation: a) reliability of the fresh cell affected by random variations in the MTJ resistance values; b) reliability degradation in time due to repetitive write stress, estimated assuming random variability of the MTJ resistance and NMOS threshold voltage.

The same analyses have been performed by varying each of the control voltages. We observe that the effect of voltage drop between Bit Line (BL) and Source Line (SL) and of body bias on the cell reliability are less relevant than the effect of Word Line and Supply Voltage. From these data we conclude that for reliability mitigation, the most efficient (among the ones we have analyzed) is supply voltage boosting. A close second is word line boosting, which shows almost the same efficiency as V_{DD} boost.

2.4 Design of Security Primitives

I have started working on this topic during my post-doc at PoliTO in Torino, Italy and I continue to work on it today. The most significant results are published in papers [J9], [J5], [C57], [C52], [C51], [C49], [C47], [C38], [C33], [C32], [C31], [C30], [C29], [C25], [W11], [W10], [W9], [W8], [W6], [W5] referenced in Chapter 6.

Security primitives are low-level cryptographic algorithms used to build cryptographic protocols for computer security systems. As an alternative to classical mathematical cryptography primitives, novel hardware security primitives are used, such as Physically Unclonable Functions (PUFs) and True Random Number Generators (TRNGs).

- Physically Unclonable Functions (PUFs) are emerging cryptographic primitives used to implement low-cost authentication protocols and cryptographic primitives, such as secure key generators, key storing and one-way functions. PUFs exploit intrinsic manufacturing variability introduced in a device during the fabrication process to generate a signature, unique to each single device.
- True Random Number Generators (TRNGs) are cryptographic primitives used to generate random numbers from a physical process, rather than a computer program. They provide random keys, device identification and seeds for Pseudo Random Number Generators (PRNGs).

2.4.1 Design and Evaluation of Phisical Unclonable Functions

STT-MRAM-based PUF Implementation

The electrical characteristics of an STT-MRAM bit-cell are strongly affected by the fabrication process. Variations in the doping profile, line edge roughness, and poly grains affect the threshold voltage of the access transistor. Variations in the thickness of the tunneling oxide layer, in the thickness and the cross-sectional area of the free ferromagnetic layer, and in the lattice dislocation during thin film growth have a direct effect on the value of the MTJ resistance (both in parallel and anti-parallel magnetization). The variation of the MTJ resistance in anti-parallel magnetization is larger than in parallel magnetization due to the intrinsic quantum tunneling process and to the extrinsic scattering process.

In our PUF solution, we take advantage of the high variability and uncontrollability of the MTJ resistance in anti-parallel magnetization and that of the transistor threshold voltage. For this, we resort to a dedicated memory array consisting of N active cells (to generate and respectively store the PUF value), and M reference cells (to generate the reference current for read operations). Since the active and the reference cells have the same design characteristics and are fabricated in the same way, they are similarly affected by variability. Also, the references cells are interleaved with the active cells to further assure similar variability under stated environmental conditions.

In our PUF solution we take advantage of the differential sensing during read operation, based on read current comparison against a reference value. The proposed solution is to be implemented as follows (see also Figure 2.27):

1. Write all cells to '1': all cells (active and reference) are set to anti-parallel magnetization. As a result, due to fabrication variability, both the active and the reference cells are

characterized by high electrical resistance, with random values distributed around the nominal, resulting in random values for the read current (IR). The reference current (Iref) takes a single value given by the average current flowing through the reference cells during the read operation.

2. Read each cell: the second step consists in reading the active cells of the memory array. This operation is performed the same way a standard differential read operation is performed, i.e., by comparing the current passing through the active cell (active current: IR) with the reference current (Iref). If IR>Iref, the value stored in the cell is interpreted by the sense amplifier as '0'. If, on the other hand, IR<Iref, the value stored in the cell is interpreted as '1'.

3. Store the security key: immediately after the read operation is performed, the outputs of the sense amplifiers (a string of '1' and '0' values) are used as security key. The PUF is now programmed and the unique signature is obtained by reading the data stored in the memory array.

A PUF solution is considered robust if after each run of the algorithm on a same device, the results obtained match with high probability. Since the MTJ and access transistor parameter values are set by fabrication, the read current distribution (IR) and the reference current (Iref) are the same for one device operated under the same conditions at each run. However, due to the unavoidable and unpredictable noise in the circuit, there is an uncertain sensing zone for the sense amplifier. The bits falling in this zone, i.e. the bits for which |IR-Iref| is smaller than the sensing margin of the SA, can induce a meta-stable state, which will be randomly stabilized to '0' or '1' depending on the noise in the circuit. These cells can be read randomly as '1' or '0' at different runs of the algorithm, and are denoted as nondeterministic active cells in Figure 2.27. To reduce the probability of such occurrences, we use a 3-stage sense amplifier (first stage for current sensing and the other two for voltage sensing), which is designed to counteract the effect of fabrication variability.

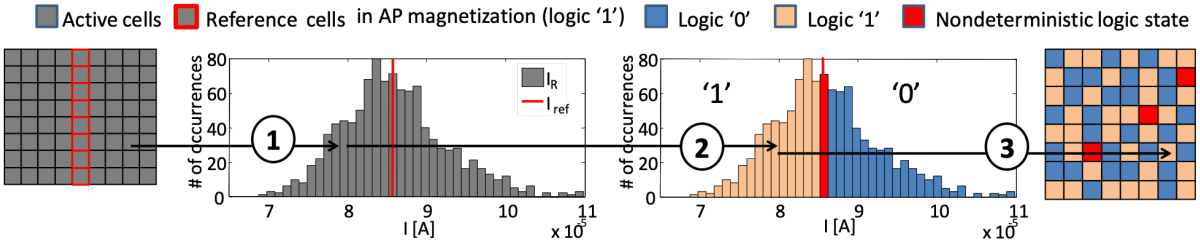


Figure 2.27: The implementation strategy of the proposed PUF solution: 1) Write all cells to '1'; 2) Read each cell; 3) Use the read value

The proposed PUF solution is based on the emerging STT-MRAM memory technology. It takes advantage of the high variability affecting the electrical resistance of the MTJ device in anti-parallel magnetization and the peculiarity of the reading operation. The PUF solution is implemented on a classical STT-MRAM memory array design, which makes it possible to be used as general-purpose memory as well. We show that the STT-MARM-based PUF is a promising solution for secret key generation since statistical simulations demonstrate high unpredictability, robustness, and limited sensitiveness to environmental variations. Since the reproducibility of our PUF is somewhat degraded by environmental operation conditions, we have also proposed a solution that allows dealing with uncertainty of the PUF response.

Besides the reproducibility, the STT-MARM-based PUF architecture can be considered an effective secure storage since tampering, model-based and side-channel attacks cannot be easily succeeded. Micro-probing tampering technique lags compared to the shrinking of silicon technology nodes and each MRAM cell measures as a single transistor. Model-based attacks exploit a subset of the CRPs set to extract knowledge about the PUF characteristics. The huge cardinality of such a subset makes this attack not viable for the proposed architecture since the CRP set is not disclosed and, moreover, no challenge mechanism is implemented. As for the side-channel, since accessory operations are performed in a differential way and since the sense amplifier is symmetric, there are not observable external phenomena.

With this implementation, one signature is generated per device (Figure 2.28a). Based on this, I have devised an alternate PUF architecture, which allows for generating multiple challenge/response pairs per device, i.e., a strong PUF implementation (Figure 2.28b).

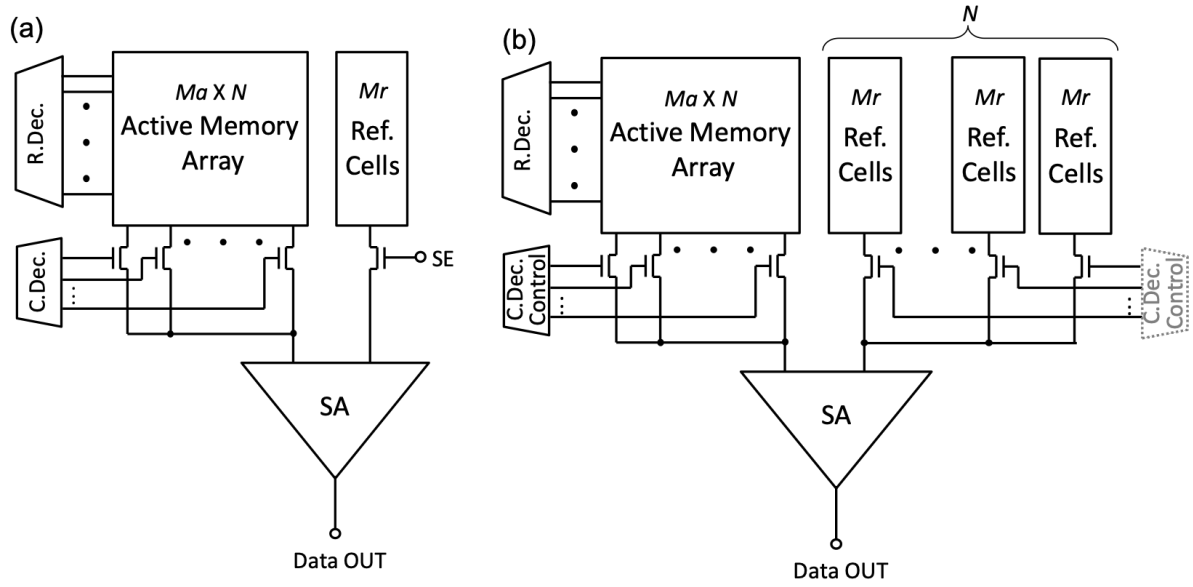


Figure 2.28: Schematic representation of the proposed PUF implementations: a) weak PUF, b) Strong PUF

The proposed architecture consists of an active memory array designed with Ma rows and N columns and a reference array designed with N groups of reference cells. All reference cell groups are identical and consist of Mr MTJ devices. These Mr MTJ devices are connected in the serial/parallel configuration described in the previous section, and are used to generate the reference current.

The novelty of the proposed approach consists in using multiple reference cell groups (the number of reference cell groups is equal to the number of columns in the active memory array). In this situation, different signatures can be generated, following the idea previously described, of comparing two currents, nominally identical but in actuality different due to process variability. The signatures can be generated by (for each active row Ma):

- reading one column at a time using one reference group (this is in fact how the weak PUF is generated),
- reading one column at a time using each of the reference groups (in this case, for each combination column/reference group we obtain a set of signatures),

- reading two columns at a time using a pair of reference groups. In this case, the current fed to the sense amplifier is double compared to the previous cases, and the condition of having nominally identical currents compared by the SA is met (in this case, we obtain new signature from any combination of 2 columns/2 reference array),
- this last situation extends to larger number of columns to be read at once. The maximum number of columns we can combine for signature generation is N (the total number of columns in the memory array), in which case we will also use N reference groups.

The challenges are generated as combinations of columns and reference cell groups. The Column Decoder Controller has the role of generating the combinations and selecting the corresponding active columns and reference cell groups used for the read operation, and, therefore, for generating the response. This number of combination grows exponentially with the number of columns, as required for a strong PUF.

SRAM-based PUF Reliability

We have proposed a novel and effective methodology to identify the unreliable bits in a SRAM-PUF based on a modified stability test strategy. In SRAM-based PUF implementations we are interested in the behavior of the bit-cell at power-up. In this scenario we are not concerned with the access transistors, since the control signals are disabled. The same situation is found when the memory cell is in data retention mode.

Classical stability tests are designed to detect the unstable SRAM cells causing memory faulty behavior. One of the most common cell-stability tests is the data retention test. This test is performed by writing all cells to '1'. Next, the supply voltage of the memory core is decreased to its critical value and the memory array is held in data retention for a predetermined number of clock cycles. Next, the supply voltage is increased to its nominal value and the data is read from the memory array. The unstable cells will flip during data retention at low supply voltage and detected during the read operation. The procedure is repeated with the memory array being initialized with all cells to '0'.

This method, however, is impractical for the purpose of this work, i.e., to detect the most stable SRAM cells, the ones that will cause unreliability of the PUF response. In order for this method to be relevant, the data retention voltage has to be decreased beyond the critical value used for traditional data retention test and data retention time has to be considerably increased to guarantee that all targeted cells have flipped. This method, as well as any other method dedicated to stability test for memory operation, is not efficient when the highly symmetrical cells are targeted. We propose a test strategy based on dynamic evaluation of bit-cell stability. Figure 2.29 illustrates the underlying difference between classical stability test and our proposed method. It shows that the classical test detects the most unstable cells, while the proposed test strategy detects the cells with high data stability. As a measure of data stability we use the static noise margin metric described in the previous section, more precisely, the difference between the metrics determined for each of the wings of the butterfly: SNM1-SNM2. For perfectly symmetrical cells this difference is zero, and its absolute value increases with the mismatch between the two cross-coupled inverters.

Our proposed solution for cell stability estimation is based on the dynamic evolution of the cell's state at power-up. The purpose of this analysis is to identify the bit-cells most likely to cause unreliability of the PUF response, i.e. the cells with high symmetry. The underlying principle of our proposed test strategy is illustrated in Figure 2.30. If the internal state of the

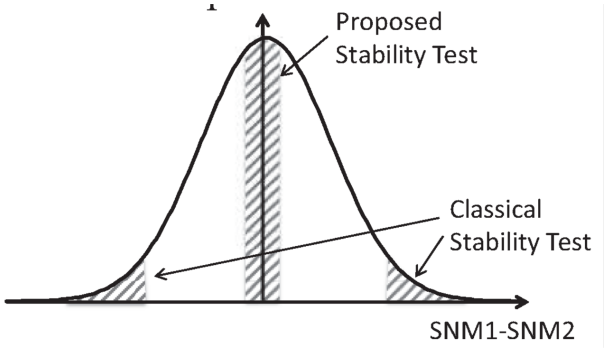


Figure 2.29: Stability test detection: Classical test detects the most unstable cells, while the proposed test strategy detects the cells with high data stability

memory cell before power-up is skewed towards the stable state SI, i.e., $V_{LO} = Vskew$ and $V_{RO} = 0$, after power-up a symmetrical cell will evolve towards the SI state and at equilibrium its internal node voltages will be $V_L = V_{DD}$ and $V_R = O$. In the same way, if its internal state is slightly skewed towards the stable state S2, i.e., $V_{LO} = 0$ and $V_{RO} = V_{skew}$, after power-up it will evolve towards the S2 state and at equilibrium its internal node voltages will be $V_L = 0$ and $V_R = V_{DD}$ (as shown in Figure 2.30a).

Nevertheless, if the cell is not symmetrical, for instance the inverter (M1&M2) is stronger than inverter (M3&M4), after power-up, the internal state of the cell will evolve towards stable state S1, regardless of the initial state skew (Figure 2.30b) direction. Figure 2.30c shows the opposite case where the inverter (M1&M2) is weaker than inverter (M3M4).

The proposed method uses a 2-step procedure where the memory is first powered-up with initial state of the cells skewed towards the state S1, and then the memory is powered-up again with initial state of the cells skewed towards the state S2. If the state of the cell stabilizes in S1 in the first step and in S2 in the second step, the cell is highly symmetrical and thus not suitable for PUF purposes (unreliable bit in the PUF response). On the contrary, if the state of the cell stabilizes in the same state in both cases, the cell is not asymmetrical, therefore very stable in the PUF context.

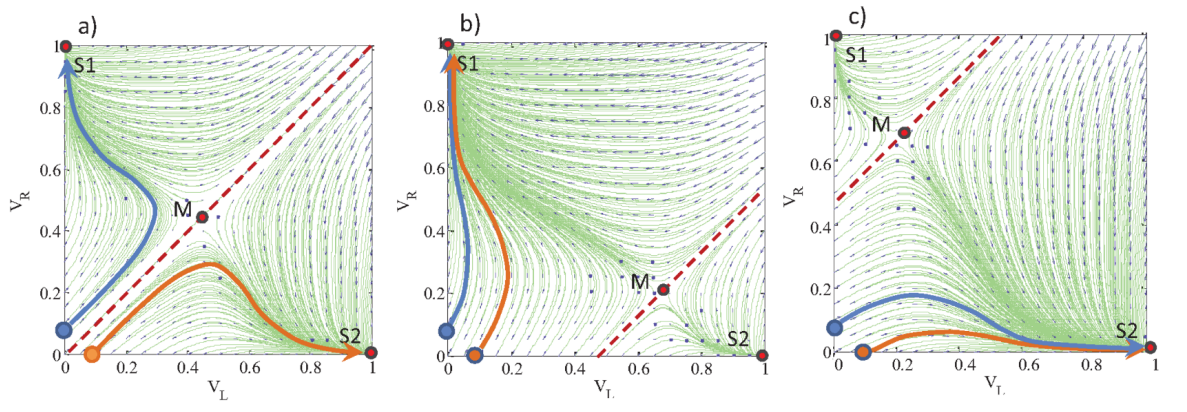


Figure 2.30: Phase space of the SRAM bit-cell and state evolution when the proposed method is implemented: a) for a perfectly symmetrical cell. b) for an asymmetrical bit-cell with inverter (M1&M2) stronger than inverter (M3&M4), c) for an asymmetrical bit-cell with inverter (M3&M4) stronger than inverter (MI&M2).

In order to implement the proposed test strategy, the classic SRAM memory design is modified

as shown in Figure 2.31. The added hardware (marked in red) is used to set the initial state of the cell internal nodes before power-up. In today SRAM memory design the supply voltage of the peripheral circuit is decoupled from the supply voltage of the memory array to save power when the memory is in data retention and in some cases to increase the reliability. We take advantage of the double supply voltage to perform the power-up in two steps. First, the peripheral supply voltage is enabled and the word lines are selected, i.e., $V_{WL} = V_{DD}$. In this way the access transistors are ON and the cross-coupled inverters are OFF. The bit-lines are decoupled from the pre-charge circuit (PRE) and also from the read/write driver (SA/WD) since no bit-line is selected. Next, the test sequence is initiated by enabling the signal TE ($V_{TE} = V_{DD}$ and $V_{TEB} = 0$). In this way, the bit-line BLB is connected to ground ($V_{BLB} = 0$) and the bit-line BL is connected to the proposed V_{skew} generator ($V_{BL} = V_{skew}$). These voltages are transferred to the cell's internal nodes through the access transistors. Next, the TE signals (TE and TEB) and WL signals are disabled separating the core cells from the bit-lines and the supply voltage of the memory array is enabled. During ramp-up of the memory array supply voltage, the memory cells will stabilize in one of the two stable states. After the power-up process is complete, the memory content is normally read. Once the read operation is completed, the supply voltage of the memory array is disabled and the data is erased from the array. The same procedure is repeated, and this time the signal TEB will be enabled before power-up, connecting the bit-line BLB to the V_{skew} generator and the bit-line BL to ground. After the power-up process is complete, the memory content is normally read.

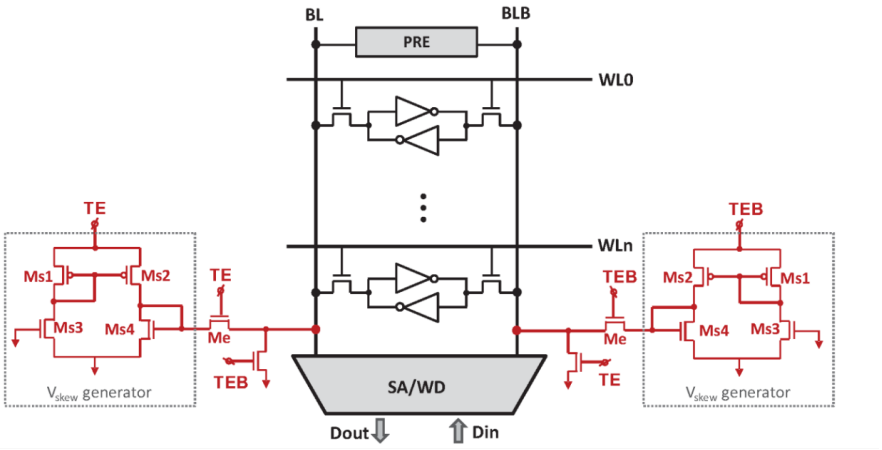


Figure 2.31: -Schematic representation of an SRAM memory array column with the dedicated circuit for stability testing marked in red

The addresses for which the read operations output complementary values are either stored in a nonvolatile memory or on a server, and are used as helper data when the memory is used for PUF implementation. They are the addresses not to be taking into account when a PUF response is generated. By eliminating these bits from the PUF response, its reliability is greatly improved. The value of the V_{skew} voltage directly affects the efficiency of the method. Using a low V_{skew} value while performing the proposed test allows detecting only the highly symmetrical cells. Increasing this value leads to the detection of slightly asymmetrical cells as well. If the value chosen for V_{skew} is too low, there is the risk of under-testing, i.e., some of the cells which cause unreliability of the PUF response remain undetected during test. If the value chosen for V_{skew} is too high, there is the risk of over-testing, i.e., more bits are eliminated from the PUF response then necessary leading to over sizing the PUF implementation.

To demonstrate the viability of our proposal, we have performed statistical HSPICE simulations accounting for local and global fabrication induced variability. We generate 500 instances of

the SRAM-based PUF implementation, designed with 1024-bit memory. Each PUF instance was once enrolled and then challenged 100 times assuming different noise seeds in the circuit. The results show the number of unreliable cells is between 71 and 108. The 500 instances of our PUF implementation have been tested using the proposed methodology. The test was performed assuming different values for the V_{skew} voltage. When a low value was chosen for the V_{skew} voltage (i.e., 15% of the nominal value of the static noise margin) the number of detected unreliable bits for PUF implementation is between 72 and 107. All cells deemed unreliable by the stability test have shown unreliability when the PUF response was challenged. However, the test coverage is lower than 100%, i.e., not all unreliable bits have been detected. Increasing the value of the V_{skew} voltage (to 20% of the nominal value of the static noise margin) the number of detected unreliable bits for PUF implementation is between 76 and 112. All cells deemed unreliable by the stability test have shown unreliability when the PUF response was challenged and the test coverage is 100% in all cases. Choosing the right V_{skew} voltage is paramount in the efficiency of the proposed stability test in the detection of PUF unreliable bits. However, since at design time the expected variability range of the fabricated circuit is known, so is the nominal static noise margin, an educated estimation of the optimum V_{skew} is possible.

RO-based PUF Reliability

The goal of this work is to define a methodology to evaluate the reliability of the RO-PUF responses, based on the measured differences of the oscillation frequencies. This method will provide at run-time, besides the response to the challenge, the information whether the response is reliable or not. We have proposed a method that will improve the state-of-the-art as it provides a methodology to estimate reliable responses on-the-fly, based on an off-line study under different environmental conditions.

Reliability is defined as the ability of the PUF to produce the same response for a given challenge under different operation conditions and aging. In the case of a RO-PUF, the frequencies of two ROs are compared to generate a response. By convention, if the frequency of the first RO is larger than the frequency of the second, the PUF response is 1, otherwise is 0. If the two frequencies are very similar, the response is prone to be unreliable since a small shift in the frequency in one of the ROs due to noise or environmental conditions can alter the response. Therefore, analyzing the frequency differences of all ROs in a PUF can give us a good measure of PUF reliability. Based on the general agreement, the oscillation frequencies (fdiff) of all ROs in the PUF can be fitted to a normal distribution. Frequency differences close to 0Hz are possibly unreliable. For this case, we define a threshold T such that pairs for which T $<$ fdiff $<$ T are considered unreliable. Thus, reliability is calculated as Reliability $= 1$ [P (T) P(T)]. Furthermore, we can use the distribution of frequency differences to estimate the time needed to obtain a response for a certain challenge. The measurements of each RO frequency are performed resorting to a counter, which count at each rising edge of the RO, and the PUF response is obtained by comparing two counters. It has been observed that RO pairs whose frequency difference is very large can assure a meaningful counter difference early on, while pairs whose frequency difference is very small take more time to provide a meaningful counter difference since the frequency difference might be masked by the sampling effect of the counters, therefore, the two counters can register the same value, until the frequency lag becomes significant enough to counteract this effect. Our methodology is based on the observations that two RO start oscillating at the same time and any two sine waves with different frequencies will experience simultaneous zero-crossing periodically, at intervals Tsync

= 1/fdiff. As a result, any expected change in the counter difference must happen in a Tsync interval. If we observe the counter difference at certain intervals tsample, we can define the expected number of samples until the counter difference changes as E = Tsync/tsample = 1/(fdiff · tsample). By introducing the notion of frequency difference threshold for reliability (i.e., T) we can correlate the lag of meaningful counter difference with the reliability of the corresponding response.

We based our study on 200 identically designed ROs with three CMOS inverters in 65nm technology provided by ST Microelectronics. We assume all ROs are affected by process variability. The output of a RO is connected to a counter which increments its value at every rising edge of the oscillation. In this study, there is a total of 10.000 possible CPRs. The state of the counter has been sampled at each 1ns. To evaluate the PUF reliability, all ROs have been simulated under nominal conditions and environmental perturbations. The results show that the effect of temperature variation is negligible as compared to supply voltage variation. Additionally, we can consider that there is no gradient of temperature inside the region of the circuit for the ROs, so we can consider that the working temperature is practically uniform. The distribution of frequency differences widens with increasing the supply voltages. This means that at lower supply voltage the expected times to obtain a response are longer than for nominal supply voltage, which, in turn, can translate to lower reliability.

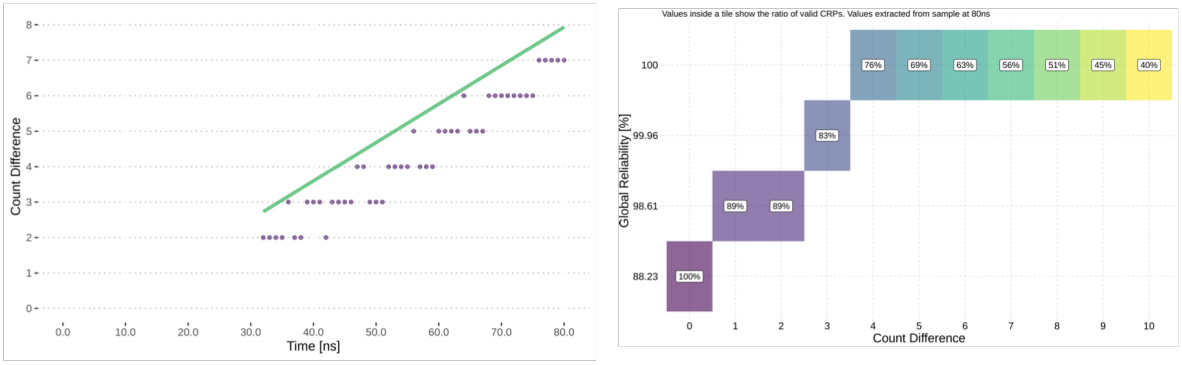


Figure 2.32: Left: correlation between minimum counter for reliable response and time of measurement. Right: Relationship between PUF reliability, counter difference and PUF entropy

Based on our observations from simulations and the general agreement on the variability distribution, the oscillation frequencies of all ROs in the PUF can be fitted to a normal distribution. RO pairs for which the frequency difference is very small are possibly unreliable and take more time to provide a meaningful counter difference. For this reason, when simulating the RO, we do not only retain the counter value at the end of the simulation time, but we also record intermediary counter values at a fixed time-step. In this way, when applying a challenge to the RO-PUF, we are able to calculate its response and also determine how fast this response can be obtained. Using this method, the reliability of the RO PUF can be computed at every counter sample. Moreover, we can calculate the minimum counter difference required to achieve 100% reliability for every sample. This correlation is show in Figure 2.32-left. A challenge whose counter difference is under the green line for any given time is considered unreliable and since the frequency difference of a pair will not change through time. This allows us to know early on if a challenge is going to be reliable. The results after applying the proposed methodology to our simulations of the RO PUF are displayed in Figure 2.32-right. It is shown the relationship between the reliability of the RO PUF, the minimum count different and the valid number of CRPs after filtering the responses that do not fall higher than the

green line from Figure 2.32-left. We can see that in our case, if we filter responses with a count difference lower than 4 at 80ns, we obtain while maintaining 79% of the CRPs.

This study can be used to implement an online test methodology for RO-PUF reliability estimation. We propose a design-for-test solution which modifies the classical RO-PUF by adding two blocks: an absolute-value subtraction block which computes the difference between the two counters in absolute value; and a comparator, which compares this difference against a constant value, which is the counter difference obtained by simulation as described previously. This new design will provide, besides the response to a challenge, the information whether the response is reliable or not.

2.4.2 Design and Evaluation of Chip IDs

Device fingerprinting, consisting on obtaining a set of attributes from a device, can be a powerful mechanism for device identification. We have proposed a device identification method based on measuring physical and electrical properties of the device, while controlling its switching activity. The method is enabled by the fact that both the sensors and the effects of the switching activity on the circuit are uniquely affected by manufacturing-induced process variability.

On-chip sensors are embedded elements widely deployed in many computing systems (e.g., MCUs, GPUs, FPGAs, etc.) which allow self-regulation of operation conditions. Among the most commonly used sensors are the temperature sensors and the voltage sensors (core voltage, power supply noise, etc.). The measurements recorded by these sensors are strongly dependent on the electric activity of the hosting device. As a rule of thumb, increased activity in the vicinity of the sensors can be observed as a temperature increase, as disturbance in the core voltage, etc. The exact values of these measurements are influenced by the fabrication-induced process variability affecting the sensors, which can induce effects like simple linear bias (like offset) or more complex effects such as cross-dimensional effects, clock-skew, tolerance or timing bias. Moreover, the electric activity of the hosting device depends on the switching activity and it is highly affected by the fabrication-induced parametric variations of the underling circuit. Therefore, the measurements from the embedded sensors are doubly affected by the fabrication-induced process variability.

The proposed identification method is based on this double effect (variability on the circuit, plus variability on the sensors measuring properties of the circuit), which results in unique features that can be used to construct a reliable hardware fingerprint of devices. In a nutshell, we generate a specific workload using different elements of the system and we record the measurements of the on-chip sensors of the system. This, in turn, allows us to build a completely self-contained fingerprinting scheme. Nevertheless, modeling the impact of the switching activity on the on-chip sensor measurements is a very challenging task. Such a model depends on many factors, as the initial state of the system, the workload, the physical location of the switching elements and on-chip sensors, crosstalk effects and the imperfections of the device. Because of the aforementioned model complexity, we propose the use of an artificial neural network (ANN) for device classification. The input of the ANN will be the raw on-chip sensor measurements generated during the application of a specific workload, while the output is the device ID. We have acquired a total of 660 FPGA fingerprints corresponding to all the measurements from the on-chip sensor, under the 4 different workloads and 5 different delays, in three different days. The signatures have been generated at room temperature using always the same configuration for the power supply. We have computed the correlation between

the different sensors and depicted in Figure 2.33a). A high positive or negative correlation means that some sensors could be removed from the configuration. In this case, there is no correlation on the sensors obtaining values close to 0. 440 FPGA signatures corresponding to two different days have been used for the training of the neural network. The rest of the signatures (220) have been used for validation purposes. Figure 2.33b) depicts the confusion matrix for the 220 signatures (20 per FPGA) where a 99.1% of accuracy is reached in the FPGA identification.

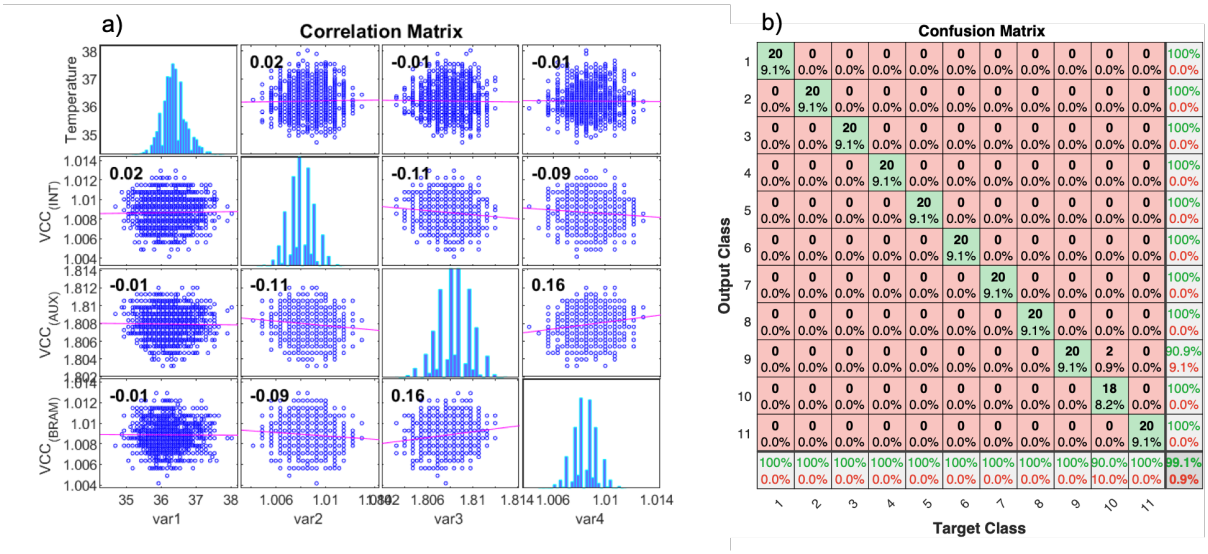


Figure 2.33: a) Correlation matrix of sensor responses for one workload; b) Confusion Matrix for the FPGA recognition using the proposed method

These preliminary experimental results are very promising, nevertheless, some issues are still open, in particular:

- **Scalability of the system:** if the population of devices is increased to a thousand of them, with the previous results, we cannot conclude that there is a unique fingerprint for each device. Experiments shall be performed on thousands of devices in order to study the suitability of the identification method for a large device population. These new experiments will include not only more devices but also will explore a wider range of switching patterns and new architectures for the NN that could improve the identification accuracy.

- **Versatility:** the experimental setup uses ad-hoc elements (i.e., ROs) to generate different workloads, and to emulate real circuit behavior. Future research will also include the adaptation of the proposed identification to other platforms, such as microprocessors or GPUs, by activating elements existing in the system (e.g., use of an AES module to generate different switching activities).

- **Operating conditions and aging:** the results have shown the effectiveness of the proposed identification method at room temperature and normal operating conditions. It is necessary to conduct further experiments using different workloads and operating conditions. As we have access to the information of the sensors, we will normalize the measurements taking into account the initial operating conditions (voltage and temperature) in order to reduce the complexity of the training and facilitate the identification. Aging of different elements could be modeled in order to take into account their effects

on the identification. Other solutions could include the retraining of the NN during the lifetime of the different devices.

- **Security:** from this perspective, the main question to be answered is related to the clonability of the system. In other words, is it possible to impersonate a legitimate device and replicate their outputs? At a first glance, the proposed system seems to be secure. Indeed, we have stated the difficulty of modeling the behavior of the entire system due to its complexity. Moreover, the unlimited number of possible workloads that can be applied makes very difficult for an adversary to predict all sensor measurements. In addition, the preliminary results presented on the subsection *Workload removal* are very promising because according to these results, the workload used to the identification must be used during the training. Thus, it will be quite difficult for an adversary to produce the correct response in an identification system where the server asks for the response of a specific workload. Other future lines could include the use of masks that select a subset of measurement points, unknowns for the adversary, for the identification. This idea is based on the fact that it is easier to obtain a higher average accuracy when replicating a large number of points than a subset of them.

2.4.3 Design and Evaluation of True Random Number Generators

Taking advantage of the MTJ stochastic switching, and using the optimal pulse width/amplitude combination to control the MTJ programming, we can obtain a 50% probability of switching. In this case, the final state of the MTJ device will be solely dependent on the internal thermal agitation, producing an unbiased output bit. In practice, these random bits are generated in three steps. First, since the state of the MTJ device is unknown, large enough pulse is applied to the MTJ to assure that the device is in the parallel state (SET operation). Then, a voltage of opposite polarity is applied to the MTJ, intended to switch its state to anti-parallel (PROGRAM operation). However, in this case, the amplitude and duration the pulse are set (by solving the equation below) in such a way assure 50% probability of switching. This means that the MTJ switching with this controlled pulse has a 50% chance to succeed (the state of the MTJ is now anti-parallel) and 50% chance to fail (the state of the MTJ remains parallel). The third step is to read (digitalize) the state of the MTJ (READ operation). By convention, if the MTJ switches to ant-parallel state, the read operation yields a logic '1', while if the MTJ keeps its parallel state, the read operation yields a logic '0'.

$$P_{WR} = exp\left(\frac{-\pi^2 \cdot \Delta \cdot (\frac{I_W}{I_{C0}} - 1)/4}{\frac{I_W}{I_{C0}} \cdot exp(2\alpha \cdot \gamma \cdot H_k \cdot t_w \cdot \frac{\frac{I_W}{I_{C0}} - 1}{1+\alpha^2}) - 1}\right) \tag{2.17}$$

Here, P_{WR} is the probability of completing the write equation, Δ is the thermal stability factor (dependent on the device internal temperature and on its geometry), I_W is the writing current, t_w is the writing duration, I_{CO} is the critical current required to achieve state switching, α is the Gilbert dumping coefficient, γ is the gyro-magnetic factor and H_k is the effective anisotropy field.

By repeating this set of operations (SET-PROGRAM-READ), on identically designed MTJ devices, assuming ideal fabrication process and unperturbed environmental conditions, we obtain strings of truly random bits, since their logic values are entirely dependent on a truly random process, i.e., the thermal fluctuations inside the MTJ device. This is due to the fact that, by assuming an ideal fabrication process, and unperturbed environmental conditions,

choosing a value of the programming current, the programming speed can be easily (and precisely) estimated.

However, in practice, the fabrication process is far from being ideal and at this moment the STT-MTJ devices are predicted to be fabricated with high process variability. This means that all parameters which are technology and fabrication dependent, will vary from device to device. Moreover, since the environmental temperature and operation voltage are subjected to perturbations (either due to internal issues like gate switching, or external issues, like hardware attack) during device operation, the parameters temperature (T), write current (I_W) and write time (t_w) required to achieve 50% switching probability, will not only be different from device to device, but also from cycle to cycle for the same device.

Based on the above considerations, we propose two solutions for STT-MTJ based TRNG design: (i) solution based on on-line tunable programming duration, achieved with a digitally controlled delay element, (ii) solution which exploits the stochastic nature of the MTJ switching, and the behavior of an XOR gate dealing with probabilistic signals. We target an MTJ switching probability of 50% with a resulting bit stream with 50% logic '1's. For statistical relevance, the generated bit stream should be rather large, since the probability of equal proportion of '1's and '0's increases with the number of trails.

TRNG based on on-line tunable programming duration of STT-MTJ

The implementation of such a TRNG is achieved by performing the following steps (as illustrated in Figure 2.34):

1. According to the nominal MTJ design parameters, set the nominal value for I_{W_n} and t_{w_n}, at room temperature, in such a way that the writing probability is $P_{WR} = 0.5$;
2. Run N cycles of SET-PROGRAM-READ operations and count the number of '1's in the resulted N-bit stream;
3. If the resulting N-bit stream has 50% of the bits at logic '1' continue by running the next N cycles of SET-PROGRAM-READ operations without changing the operation conditions. If the resulting N-bit stream has less than 50% of the bits at logic '1', increase write pulse (tw) and continue by running the next N cycles of SET-PROGRAM-READ operations with the modified operation conditions. If the resulting N-bit stream has more than 50% of the bits at logic '1', decrease the write pulse (tw) and continue by running the next N cycles of SET-PROGRAM-READ operations with the modified operation conditions;
4. Repeat step 3 until the desired bit-string length has been achieved. The situation in which the resulting bit-stream has bias towards one of the logic states can arise due to a plethora of factors:
    - the parameters of the fabricated cell differ from the nominal MTJ design parameters, for which the nominal values of amplitude and duration of the PROGRAM pulse have been estimated;
    - the operation temperature is different for room temperature;
    - the supply voltage suffers fluctuations;
    - the peripheral circuit itself (specially the write driver) is affected by fabrication induced variability.

Regardless of the reason for which the resulting N-bit stream is bias towards a logic value, our proposed scheme compensates by modifying the duration of PROGRAM operation, in such a way that the overall bit-stream (obtained by running multiple times N X SET-PROGRAM-READ operation) is unbiased, with high entropy and uncorrelated bits. This is achieved by performing a preliminary worst-case analysis of the targeted MTJ device (accounting for environmental- and fabrication- induced process variation) thus estimating the range of delay our digitally-controlled element has to cover, as well as the required minimum precision.

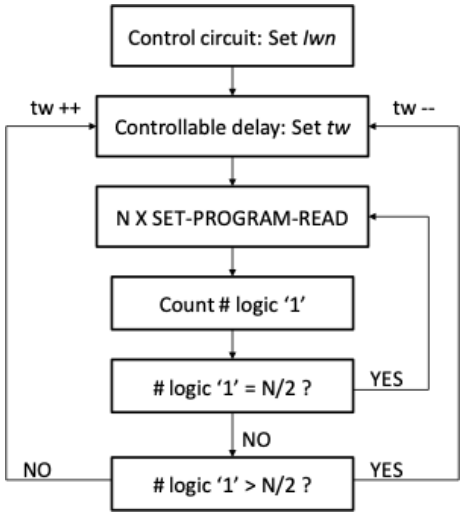


Figure 2.34: TRNG implementation algorithm

The schematic circuit representation of the TRNG implementation (described by the algorithm in Figure 2.34) is depicted in Figure 2.35. Here we use one MTJ element, with its ferromagnetic layers set in relative parallel orientation, i.e., initialized at logic state '0'. The stochastic nature of the switching process is the entropy source for the proposed implementation. The output noisy signal is digitalized by performing a traditional read operation, and, under ideal conditions a random bit-stream is generated. The read operation is performed by means of a Sense Amplifier (S.A. in Figure 2.35) which compares the current passing through the MTJ element against a reference current and decides on the state of the MTJ. By convention, a read current larger than the reference value is output as logic '0', while a read current smaller than the reference value is output as logic '1'. This is the conventional read operation, and we need not make any adjustments for the purpose of our implementation.

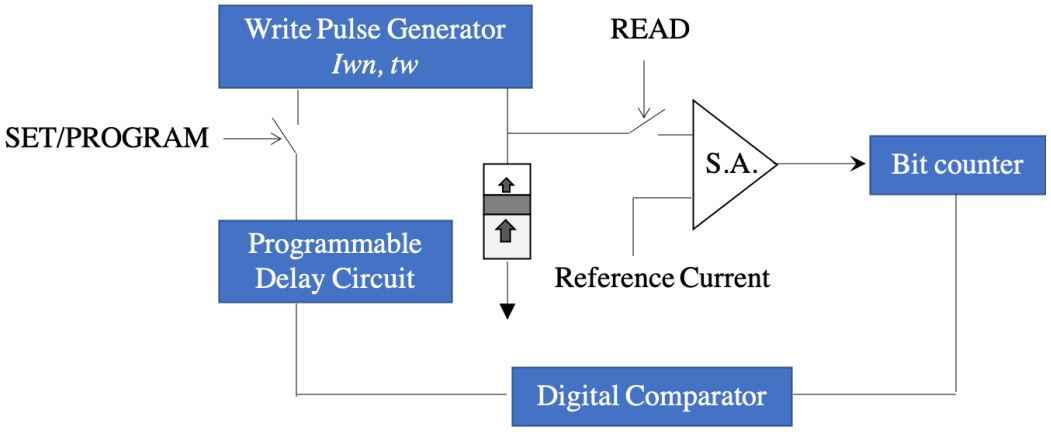


Figure 2.35: Schematic circuit representation of TRNG implementation

However, due to unavoidable fabrication-induced process variation, and inadvertent variability of the operating conditions, i.e., temperature and current, the output bit-stream will be biased towards one logic state or the other, according to the magnitude of the variability, resulting in low quality randomness of the bit-stream. To compensate for this issue, we are proposing a closed-loop circuit implementation, which allows to tune, on-the-fly, the PROGRAM time in such a way to compensate for the biased output. A Bit Counter is introduced at the output of the Sense Amplifier which counts the number of bits resolved by the Sense Amplifier as logic '1'. The result is fed to a digital comparator, which evaluates whether the number of bits in state '1' is larger or smaller than half of the bit-stream length. The result is fed to a digitally controlled delay element, which is used to tune the PROGRAM time accordingly, in order to reduce, or, ideally, eliminate the output bit-stream bias. If the bit-stream is biased towards logic '1' (number of '1' bits larger than '0' bits), the delay element is controlled towards decreasing the delay, i.e., decreasing the write pulse width. If, on the other hand, the bit-stream is biased towards logic '0' (number of '1' bits smaller than '0' bits), the delay element is controlled towards increasing the delay, i.e., increasing the write pulse width. The bit counter and digital controller circuits are implemented using conventional design, and we need not make any adjustments for the purpose of our implementation we will, however, require a custom-designed delay element.

The accuracy of the delay element will directly affect the bias of each generated N-bit stream. A poor accuracy will lead to oscillations of the programming pulse width, which will cause the N-bit stream to be strongly biased towards '1' or '0', directly affecting the uniformity of the generated number, in turn affecting its randomness. Our delay element is designed in such a way that this problem can be avoided.

As a proof of concept, using the algorithm in Figure 2.34 and the simulation set-up in Figure 2.35, we have performed M=100 sets of N=512 SET/PROGRAM/READ sequences. For illustration purpose only, and due to limited space, I present here the results obtained in 4 case studies: (a) nominal MTJ device operating at nominal conditions; (b) nominal MTJ device operating at room temperature under write current drop; (c) nominal MTJ device operating at nominal current under temperature drop; (d) nominal MTJ device operating under the effect of current and temperature drop (as depicted in Figure 2.36). The simulations were performed, under the same conditions, with and without the feedback loop, with the purpose of demonstrating the efficiency of our proposed implementation. It can be seen that when under nominal environmental conditions the output bit-stream is overall insignificantly bias, the number of '1' and '0' bits is practically the same. This statement is confirmed in both situations: with and without the feedback loop. This is to be expected since no operation condition changes, and the output is solely dependent on the inherent stochasticity of the write operation. In this scenario, the feedback loop plays a role only for the initial set of t_w value, which is dependent on the fabricated MTJ characteristics. In all the other case studies, the importance of using the feedback loop becomes apparent. The results show that performing the SET/PROGRAM/READ sequences without the use of the feed-back loop, results in an output bit-stream strongly biased towards '0' logic. While, using the feed-back loop we manage to compensate for environmental variations and eventually manage to re-equilibrate the bit-stream. It is worth noting that the compensation is fast, and its speed can be further improved by possibly improving the design of the delay element.

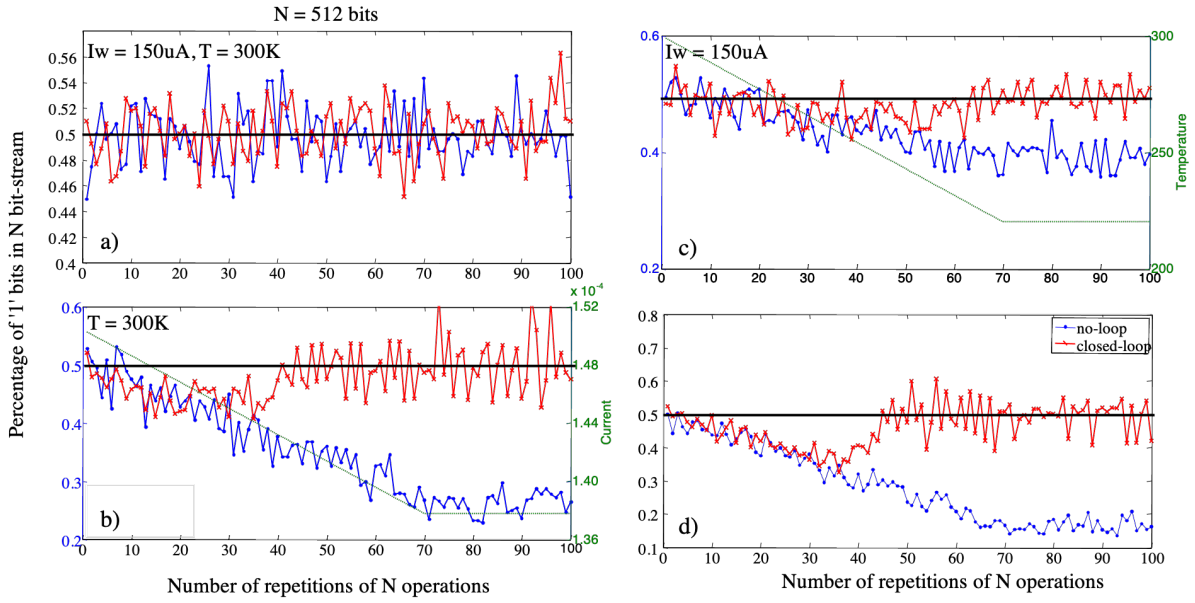


Figure 2.36: The estimated percentage of '1' bits in a N-bit-string under different operation conditions (i) blue plots represent the results obtained without delay tuning, (ii) blue plots represent the results obtained with delay tuning in closed loop, (a) implementation at nominal current and temperature; (b) implementation under current variation at nominal temperature; (c) implementation at nominal current under temperature variation; (d) implementation under current ant temperature variation

High-Entropy STT-MTJ-based TRNG

We have proposed an innovative TRNG design by combining multiple spintronic dices whose outputs are XORed in order to mitigate both manufacturing process variability and variations in the write logic.

When assuming the MTJ write operation as the source of randomness, it is possible to calculate its entropy in different operation conditions. As an example, Figure 2.37 shows the entropy of the MTJ state after performing a write operation with constant write current (I_W=150μA) and variable write duration (t_w), under the assumption of process variability (I_{C0}, which is strongly dependent on fabrication-induced material and geometry variations). The entropy has been calculated using the following equation:

$$H(X) = -P_{X=0} \cdot log_2(P_{X=0}) - P_{X=1} \cdot log_2(P_{X=1}) \tag{2.18}$$

where where X is the datum with possible values 0 and 1, $P_{X=0}$ is the probability of the datum to be 0 and $P_{X=1}$ is the probability of the datum to be 1. In our specific case, the state probability is given by eq. (2.17).

Under these conditions, we have identified the maximum entropy conditions, represented by the black line in Figure 2.37. Ideally, the entropy of the data generated by a TRNG should respect the maximum entropy condition. In reality, the data generated by the TRNG is situated in the vicinity of the black line due to uncontrollability of the manufacturing process and noise.

The solution proposed in this paper mitigates these issues, by XORing multiple MTJ devices, as illustrated in Figure 2.38.

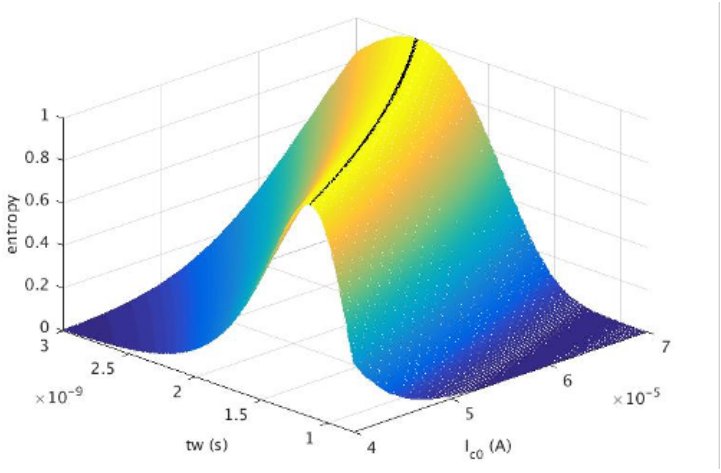


Figure 2.37: Entropy of the MTJ state after performing a write operation

Let us consider a probabilistic digital signal, with probability p_i being in logic state '1'. For this signal to be perfectly random, the probability pi should be equal to 50% ($=\frac{1}{2}$). In the case of non-ideality, we define the probability deviation from the ideal value as $\alpha_i=|\frac{1}{2}\text{-}p_i|$, with $\alpha_i[0;\frac{1}{2}]$. Combining two such signals as such in an XOR gate, a probabilistic signal will result, for which the probability of being in the logic state '1' is given by equation the following equation, directly obtained from the XOR truth table

$$P_{XOR} = p_1 \cdot (1 - p_2) + p_2 \cdot (1 - p_1) = \frac{1}{2} \pm 2 \cdot \alpha_1 \alpha_2 \tag{2.19}$$

This equation can be extended to the case of N probabilistic sources, as follows:

$$P_{XOR} = \frac{1}{2} \pm (-2)^{N-1} \alpha_1 \alpha_2 ... \alpha_N \tag{2.20}$$

The deviation of the output signal from the ideal randomness can be defined as:

$$\alpha_{XOR} = 2^{N-1} \alpha_1 \alpha_2 ... \alpha_N \tag{2.21}$$

with $\alpha_{XOR} \in [0;\frac{1}{2}]$. In other words, the magnitude of α_{XOR} represents the quality of the signal randomness, i.e., the smaller its value the better the randomness.

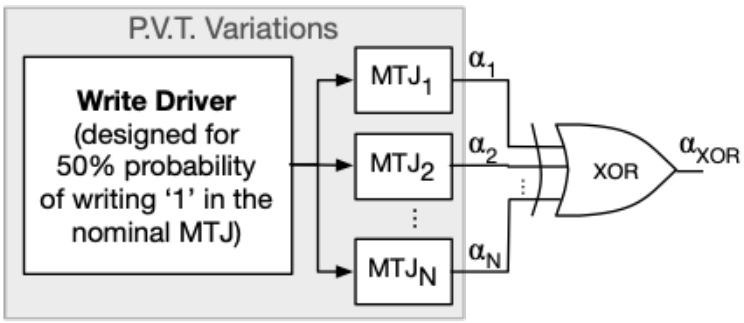


Figure 2.38: High-Entropy STT-MTJ-based TRNG implementation

The challenge now is to understand how many MTJ devices are needed to guarantee that the resulting α_{XOR} is small enough to assure the quality of the resulting random sequence.

We have applied the NIST test to identify the minimum value of α_{XOR} for which the quality of the random sequence is acceptable. When running the NIST tests with multiple test sequences, data is considered random if at least a certain number of sequences have passed the test (threshold row in Figure 2.39.). Results show that the optimum value of α_{XOR} is $\alpha_{(XOR-opt)}$=10^{-5}.

α	Frequency	BlockFrequency	CumulativeSums	Runs	LongestRun	Rank	FFT	NonOverlapTemplate	Universal	ApproxEntropy	RandomExcursions	RandomExcVariant	Serial	LinearComplexity
10^{-2}	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10^{-3}	0	188	0	0	889	1009	1014	0	992	0	0	0	770	1014
10^{-4}	724	1011	726	991	1011	1011	1023	1024	1011	1006	387	388	1013	1015
10^{-5}	1010	1014	1012	1015	1010	1010	1011	1024	1011	1008	641	636	1015	1012
10^{-6}	1018	1019	1015	1013	1012	1013	1012	1024	1016	1011	655	655	1017	1016
10^{-7}	1014	1015	1017	1012	1011	1014	1012	1024	1015	1009	633	631	1011	1009
Threshold	**1004**	**1004**	**1004**	**1004**	**1004**	**1004**	**1004**	**1004**	**1004**	**1004**	**610**	**610**	**1004**	**1004**
10^{-2}	0	0	0	0	0	0	0	0	0	0	0	0	0	1
10^{-3}	0	0	0	0	0	1	1	0	0	0	0	0	0	1
10^{-4}	0	1	0	0	1	1	1	1	1	1	0	0	1	1
10^{-5}	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10^{-6}	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10^{-7}	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Figure 2.39: NIST test results assuming various values of α_{XOR}

We have performed a large number of simulations, taking into account process variability, voltage and temperature variability and we have found that with 24 MTJ devices we can assure the value of $\alpha_{(XOR-opt)}$=10^{-5}.

This work proposes a TRNG implementation based on STT-MTJ devices. The solution exploits the post-processing power of XORs and does not require complex analog design to compensate the effects of variability and noise. This makes the design straightforward and guarantees the quality of the random numbers from the first generated bit. In contrast, previous solutions which require a control setup time of many cycles before the high entropy can be guaranteed. In addition, this solution filters temperature attacks without any loss in throughput and randomness quality, while the other solutions suffer massive loss in throughput, not being able to filter strong attacks.

2.5 In-Memory Computing

In this section I describe my activities related to the analysis of state-of-the art computing in memory solutions. I have carried out this research at TIMA, and it has been published in papers [C53] and [C54] listed in Chapter 6.

Since the appearance of modern computers, the widely adopted architecture has been based on the separation between the computing unit (or processor) and the memory storing the program to be executed and its data, i.e., the Von Neumann architecture. The separation between processor and memory has become an issue in modern computers due to the uneven evolution of processing speed and memory access times (also known as memory wall). With the technology advancements, the memory wall became increasingly important. Therefore, there is an urgent need to explore alternative architectures in the light of emerging non-volatile technologies (resistive memories), not only to further increase the computing efficiency at lower cost, but also to further reduce the overall energy. For example, moving the computation to the memory (rather than doing it in the CPU) will significantly reduce the communication and therefore reduce the power consumption and increase the performance. This computing paradigm is referred to as Computation-in-Memory (CIM), an emerging concept based on the tight integration of traditionally separated memory elements and combinational circuits, that ensures the minimization of the time and the energy needed to move data across the processor. Computation-In-Memory (CIM) architecture, aims at eliminating the communication bottleneck while supporting massive parallelism. However, to achieve the ultimate objective of fully integrating the processing units and the memory in the same physical location, several technological challenges need to be overcome. There is a very wide variety of CIM solutions proposed today that exploit existing technologies. They enable logic and/or arithmetic operations directly inside the memory boundaries. The operations are performed without the need of transferring data to/from the CPU, thus saving time and energy. This can be achieved exploiting the physical characteristics of the memory and/or inserting computational elements in the peripheral logic (sense amplifiers). The research is mainly divided on levels of abstraction: (1) device level – to identify the optimum material combination to achieve the desired device behavior; (2) circuit level – to design logic gates and arithmetic circuits; (3) system level – to map applications on memory arrays, parallel computation and design computing accelerators.

In this context, my recent research was directed towards identifying the main advantages and disadvantages of the existing CIM solutions exploiting traditional CMOS memories (such as SRAM) or beyond-CMOS devices (such as the memristive devices).

2.5.1 Comparative Study of Memristive Logic-in-Memory (LiM) Implementations

In the last five years, the number of research papers dealing with this topic on different levels of abstraction has increased exponentially. In this context, research is focused on the design of LiM architectures, the development of LiM-compatible instructions set, the methods for system integration and development of the programming model for LiM integration in computing systems. Nevertheless, the actual status of the research is fragmented and the reproduction of the reported results, along with the choice of an implementation to be adopted, are not trivial. We have studied the existing LiM implementations in order to perform a fair comparison in terms of resources and efficiency. More in particular, we study the simple Boolean functions

implemented in-memory resulting in a comprehensive comparison of LiM solutions in terms of required number of memristors and number of operations.

Several LiM proposals exist in literature, some are general, and can be used with any memory technology, others take advantage of the device physics and are only suitable to be implemented in a specific technology. In addition, some of the existing LiM solutions are designed to implement specific logic functions henceforth called "primitive operations", while others propose solutions for the implementation of any Boolean function. The existing LiM solutions can be classified depending on the way the inputs are stored (the memory content, i.e., stateful logic, or an electrical signal, i.e., non-stateful logic) and depending on where the operations are performed (in the memory array, or in the periphery). In this context, three main LiM classes can be distinguished. They are described in the following and their characteristics are summarized in Figure 2.40.

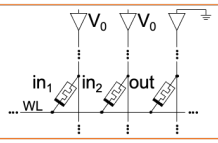
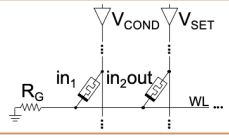
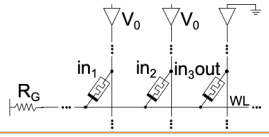
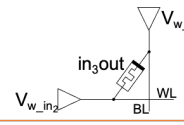
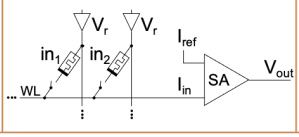
	CiM Solution	*#mem pts*	*Gate*	*Remarks*	*Technology*	*Operations*
LIM Array Stateful Logic	**MAGIC (NOR2)**	3		input non-destructive, parallelizable	Memristive crossbar (1R-RRAM)	1. Initialize R_{out} 2. Apply proper voltages to R_{in}, R_{out} 3. Obtain output in R_{in_out}
	IMPLY A → B	2		input destructive, parallelizable		
	Stateful Three-Input Logic (ORNOR3)	3		input destructive		
LIM Array + Periphery – Non Stateful Logic (Hybrid inputs)	**PLiM (RMAJ3)**	3 (1 in the array, 2 external)		input destructive, requires input pre-processing	Memristive crossbar (1R-RRAM)	1. Read inputs R_{in} and convert to voltages 2. Apply the voltages to R_{in_out} 3. Obtain output in R_{in_out}
LIM Array + Periphery Stateful Logic	**Logic in Periphery**	2		input non-destructive, additional step to store output in memory	SRAM 1T1R-RRAM STT-MRAM	1. Apply voltage on multiple rows 2. Obtain output (via SA) 3. [Store output in memory]

Figure 2.40: Primitive logic gates: Column 2 (CiM Solution) lists the considered LiM solutions and the corresponding primitive operations. The number of memory cells needed to implement a 2-input (1-bit) primitive operation is summarized in Column 3 (mem pts), while the schematic of the "primitive operation" gate for each solution is illustrated in Column 4 (Gate). Column 7 (Operations) lists the algorithm executed to obtain the result of the primitive operation. The executed operation can be input-destructive or not (see column 5, Remarks). An input-destructive operation is an operation that changes the value of the inputs after it is executed

Stateful Logic in LiM Array – The operations are performed within the memory array and the data are stored as memory content. This solution is proposed only for memristive crossbar (1R-RRAM) arrays. The input data are stored within the memory array, and the output (computation result) is obtained as memory content within the memory array. Inputs and outputs are coded as the resistive states of the storing memristors. In order to enable LiM operations within the memory array, several conditions need to be respected: (1) the memory cells containing the input data and the memory cells to store the result of the computation must share the same row (column); (2) access to multiple memory cells should be enabled; (3) specific control voltages (different than the memory read/write voltages) to be applied for

the completion of logic operations. As a consequence, the write driver, the voltage regulator and the address decoders of standard memory array have to be modified to enable LiM. LiM solutions pertaining to this class include: Memristor-Aided Logic - MAGIC, FELIX, IMPLY, and Stateful Three-Input Logic (ORNOR3).

Stateful Logic in LiM Array and its Periphery – The operations are performed within the memory array periphery or by means of additional logic and the input data are stored as memory content. This solution can be used with any memory technology. The input data are stored within the memory array, while the output (computation result) is obtained as a voltage (or current) outside of the memory array. In order to enable LiM operations within the periphery of the memory array, several conditions need to be respected: (1) the memory cells containing the input data must share the same column; (2) access to multiple memory cells should be enabled by modifying the address decoders; (3) the sense amplifiers should be modified such that different references are allowed.

Non-Stateful Logic in LiM Array and its Periphery – The operations are performed within the memory array and by using additional logic, and the data are coded partially as memory content and partially as voltage levels. This solution can be used with resistive technology only. It uses two types of input data: (1) memory content, (2) voltage level, while the output (computation result) is obtained as memory content within the memory array. In order to enable this type of LiM operation several conditions need to be respected: (1) specific control voltages (different from the memory read/write voltages) to be applied for the completion of logic operations, (2) specifi registers to store the inputs to be given as voltage levels. As a consequence, the write driver, the voltage regulator and the address decoders of standard memory array have to be modified to enable LiM.

The goal of this work is to provide a comprehensive comparison of existing LiM solutions and understand their implementation complexity. The analysis has been performed on basic Boolean functions, in order to be as generic as possible and to provide the designer an indication of implementation complexity and cost of each LiM solution. In addition, this study can give an indication of which LiM solution is more suitable for a target application, depending on the most used Boolean functions. In order to achieve a fair comparison among all solutions, we mapped all the 0-input logic functions (TRUE and FALSE), 1-input logic functions (COPY, NOT), 2-input logic functions (NOR, OR, NAND, AND, XNOR, XOR, NIMPLY, IMPLY) and the Full Adder as 3-input logic function, by using the primitive operations of MAGIC (and its extensions), IMPLY (and ORNOR3) and PLiM solutions. The Figure 2.41 summarizes the results of this study.

The obtained results show big discrepancies among LiM solutions in the number of steps to execute the operations. For instance, the XOR requires many more steps if implemented with IMPLY logic compared to FELIX. These results reflect the complexity of each operation but do not directly translate into an estimation of the actual execution time. This is due to the fact that, due to physical and electrical characteristics of the memristive devices, the timing of each operation can vary significantly.

2.5.2 Feasibility of complex Boolean functions using memristive-based LiM

In addition to comparing LiM solutions, we have also performed an in-depth analysis of memristive-based LiM implementations and identified the advantages and challenges with which a designer would confront when adopting this computing paradigm. We demonstrate

	Input			#memristors						#steps					
2 bit	0 0 1 1 0 1 0 1			MAGIC	FELIX	IMPLY	IMPLY destr.	ORNOR3	RMAJ3	MAGIC	FELIX	IMPLY	IMPLY destr.	ORNOR3	RMAJ3
Gate	Output	#in	#out												
TRUE (write 1)	1 1 1 1	0	1	1 + 0	1 + 0	1 + 0	1 + 0	1 + 0	1 + 0	1	1	1	1	1	1
FALSE (write 0)	0 0 0 0	0	1	1 + 0	1 + 0	1 + 0	1 + 0	1 + 0	1 + 0	1	1	1	1	1	1
in1 (COPY)	0 0 1 1	1	1	2 + 1	2 + 1	2 + 1	2 + 1	2 + 1	2 + 0	4	3	4	4	4	2 + 1
NOT *in1*	1 1 0 0	1	1	2 + 0	2 + 0	2 + 0	2 + 0	2 + 0	2 + 0	2	2	2	2	2	2 + 1
in1 NOR *in2*	1 0 0 0	2	1	3 + 0	3 + 0	3 + 1	3 + 0	3 + 0	3 + 1	2	2	9	5	2	6 + 4
in1 OR *in2*	0 1 1 1	2	1	3 + 1	3 + 0	3 + 1	2 + 1	3 + 1	3 + 1	4	2	7	3	4	4 + 3
in1 NAND *in2*	1 1 1 0	2	1	3 + 2	3 + 0	3 + 0	3 + 0	3 + 0	3 + 1	10	2	3	3	3	6 + 5
in1 AND *in2*	0 0 0 1	2	1	3 + 2	3 + 1	3 + 1	3 + 1	3 + 1	3 + 1	6	4	5	5	5	4 + 3
in2 IMP *in1*	1 0 1 1	2	1	3 + 1	3 + 1	3 + 1	2 + 0	3 + 1	3 + 0	6	4	5	1	1	2 + 2
in2 NIMP *in1*	0 1 0 0	2	1	3 + 1	3 + 1	3 + 1	3 + 0	3 + 1	3 + 1	4	4	7	3	3	4 + 3
in1 XOR *in2*	0 1 1 0	2	1	3 + 2	3 + 0	3 + 2	3 + 2	3 + 2	3 + 1	10	3	13	13	10	7 + 4
in1 EQUAL *in2* (XNOR)	1 0 0 1	2	1	3 + 2	3 + 1	3 + 2	3 + 2	3 + 2	3 + 1	12	5	15	15	12	9 + 5
	Input														
3 bit	0 0 0 0 1 1 1 1 0 0 1 1 0 0 1 1 0 1 0 1 0 1 0 1														
	Output														
FA (sum)	0 1 1 0 1 0 0 1	3	2	5 + 4	5 + 1	5 + 2	5 + 2	5+5	5 + 1	36	12	49+37	49+37	24	9 + 10
FA (c_out)	0 0 0 1 0 1 1 1														

Figure 2.41: State-of-the-art CIM solutions compared

that: (i) in some cases a very precise control is needed to guarantee the correct completion of logic functions; (ii) assuring the perennity of input data (i.e., performing non-input destructive operations) is not always guaranteed; (iii) concatenating logic operations is not trivial; (iv) when performing complex logic operations, there is an important tradeoff between the robustness of data (both input and output), the complexity of the controller and the number of clock cycles required for the completion of the operation.

As mentioned before, specific conditions need to be complied with to guarantee the correct operation of the existing LiM solutions. The dynamic behavior of the memristive devices needs to be accounted for when Boolean operations are desired within the memory array. By performing a simple circuit analysis, we can identify the operation voltage ranges for which LiM implementations become viable. An example for a specific memristor is illustrated in Figure 2.42.

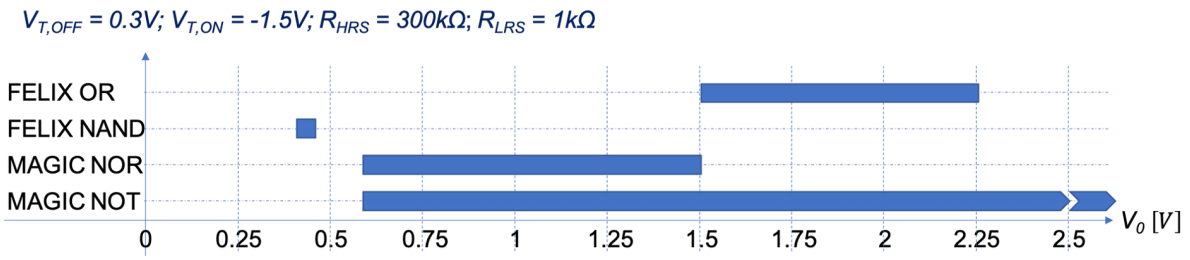


Figure 2.42: Control voltage range for which the LiM Boolean operations are correctly executed within a memristor array

It should be noted here that while MAGIC NOT operation allows for a very wide range of voltages, the FELIX NAND requires very precise control of the voltage. In addition, when the dynamics of the circuit are accounted for and we analyze the circuit behavior under real conditions, when the control signals are transient (traditionally clocked rectangular pulses) and not DC we discover that the completion of the Bolen logic is in some cases extremely sensitive to the parameters of the control signal (both amplitude and duration) as illustrated in Figure 2.43.

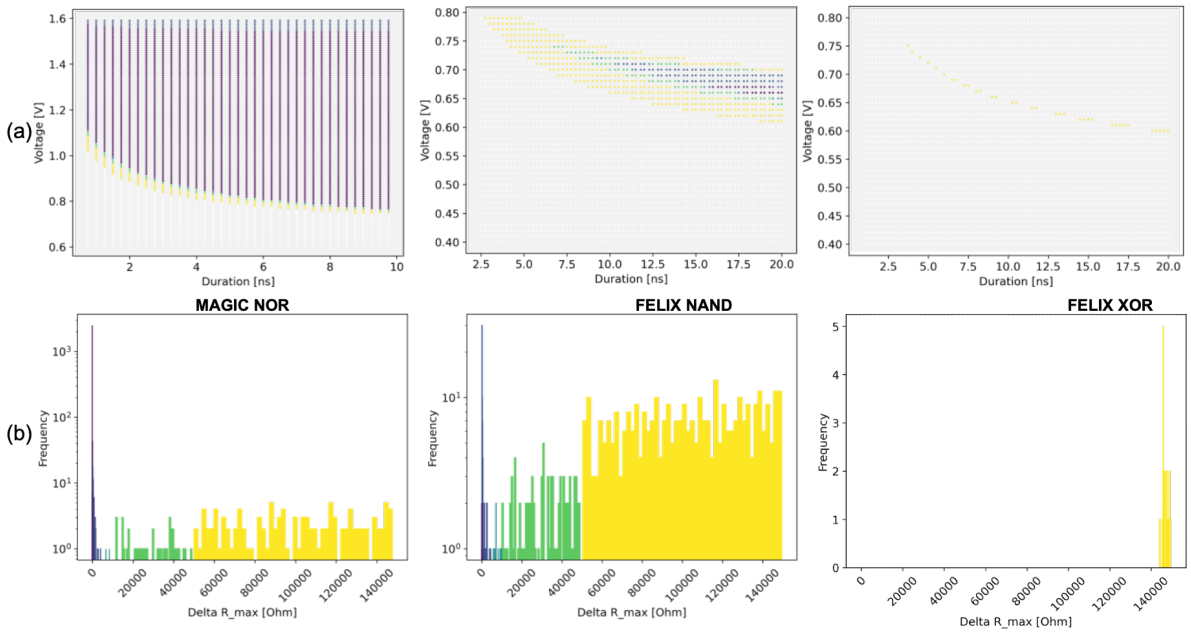


Figure 2.43: Map of operation conditions (amplitude and duration of the control pulse) for which the Boolean function is completed using LiM on memristive array. The upper figures (a) show the correctness of the operation and the retention of the input data (colored dots mark the conditions for which an operation is competed. The lower figures (b) illustrate the distribution of "perfect", "acceptable" and "worst correct" instances under the assumed control signals

For each control voltage setup (i.e., couple of amplitude and duration) we have simulated (using Cadence Spectre with the VTEAM memristor model) the execution of a set of LIM operations (MAGIC NOT, MAGIC NOR, FELIX OR, FELIX NAND, FELIX XOR, IMPLY) applying all input combinations (i.e., 00, 01, 10, 11) and analyzing the final state of the three memristors (the two inputs and the output). The results for the MAGIC NOR, FELIX NAND and FELIX XOR are illustrated in Figure 2.43. Here, in blue we illustrate the "perfect" results, i.e., input and output memristor are in nominal LRS/HRS when the operation is completed; in yellow we illustrate the "worst correct operation", i.e., the input data is lost but the output memristor stores the correct bit value, however with a very degraded LRS/HRS, in green we illustrate the "acceptable" operation, i.e., input and output memristor store the correct Boolean value but their LRS/HRS are degraded compared to nominal case. The worst situation is observed in the case of the FELIX XOR implementation. This is due to the fact that this operation requires the completion of 2 consecutive Boolean functions, both being very sensitive to the control signal. Our results showed that while operations are performed correctly from a logic point of view, the resistance of the output memristor does not always reach its ideal value. This can be problematic in multiple situations, such as when we concatenate logic functions, or when the design suffers from marge fabrication-induced variability. In both cases this resistance off-set can cause the wrong digital result.

2.6 Neuromorphic Computing - Design & Robustness Assessment

In this section I describe my activities related to the design and the analysis of robustness issues in hardware-based Neural Networks. My research is mainly focused on the Spiking Neural Networks (SNN) based on memristive devices for synaptic storage and modulation. I have carried out this research at TIMA, and it has been published in papers [C56], [C55], [C50], [C48], [C46], [C42], [C40], [C39], [C37] listed in Chapter 6.

2.6.1 Introduction

The computing performance needed by emerging electronic applications (such as Internet-of-Things and Big Data analytics) is posing a serious challenge to current computer architectures and technologies, which are required to provide increasing computing power while withstanding severe constraints on size, energy consumption and reliability. Conventional Von-Neumann architectures and memories are not likely to fulfil all the needs of modern applications, due to inherent technological and conceptual limitations. Hence, in order to be at the forefront of the electronic industry in terms of design and manufacturing capabilities, it is essential to focus research and innovation efforts on the development of novel non-Von Neumann architectures enabled by emerging technology devices. In this context, the neuromorphic computing paradigm has a huge potential when it makes use of emerging NV technologies (STT-MRAM, memristors), however, reliable and testable HW designs enabling the neuromorphic computing are still missing.

In this context, my research activity is spans three main domains: (i) emerging memory technologies (memristors and spintronic devices) used in a non-Von Neumann context, (ii) hardware dependability (robustness, reliability and test) and design-for-dependability, (iii) hardware implementations of bio-inspired neural networks (Spiking Neural Networks).

The emerging memory technologies favor increasing system complexity and performance, opening the scientific community to great improvements in state-of-the-art computing (such as high-performance computing or approximate computing) but also to new applications and computation paradigms (such as in-memory computing or neuromorphic computing) which had been unfeasible a few years back due to technological limitations. This project mainly focuses on the use of memristors or spintronic devices as artificial synapses for bio-inspired computing architectures. In this context, they have double functionality: memory (to save the values of the synaptic weights) and computation (to facilitate the on-line learning process, i.e., update synaptic weights).

Due to the aggressive technology scaling and the introduction of new steps at back-end-of-line for the fabrication of the storage element, both components of the memory cell (access device and storage element) suffer from fabrication-induced variability. Moreover, due to intrinsic properties of these technologies, they are more susceptible to variations and defects, so there is a need for high quality test and fault tolerance. In addition, because of the different operation modes (analog storage and computing) compared to traditional digital memories, they require fundamentally new testing schemes.

Hardware dependability describes the ability of a system or component to function under stated conditions for a specified period of time. Dependability issues in hardware come from the manufacturing process, or operation and environmental conditions. Manufacturing induced variability, defects, stochastic effects, and aging degradation can cause important

variations of the electrical characteristics of fabricated devices which can lead to device failure. Manufacturing test together with design-for-reliability solutions assure the quality and reliability of integrated circuits.

The Spiking Neural Networks (SNN) show good potential due to the high level of realism they bring to neural simulation, their energy efficiency and their ability for on-line learning. The related bio-inspired learning rule is known as STDP (Spike Based Dependent Plasticity) and is applied on each synapse independently of the global state of the network. In return, the synapse must be doted of computation capabilities. A hardware implementation of an SNN requites architectural co-localization of the processing and memory (non-Von Neumann architecture). The circuits solutions used to implement silicon neurons are application depended, but the vast majority are built with a temporal integration block, a spike generation block, a refractory period mechanism, and a spike adaptation block. Synapses are required to exhibit plasticity (i.e., modulation in their efficacy) and to support online learning algorithms, that manifest in changes in their strengths. Emerging memory devices can be used as synaptic elements thanks to their tuneable conductivity, compatibility with advanced CMOS fabrication process, low power consumption, non-volatility and scalability. The synaptic conductance modulation can be emulated using: (i) the analog approach (cumulative decrease and increase of resistance), where multiple resistance states emulate long-term potentiation and depression; or (ii) the binary approach, uses two distinct resistance states per device associated and probabilistic programming.

The main research results I have obtained related to this topic are: (i) the analysis of a fully connected SNN under process variability, (ii) the design of a versatile analog spiking neuron in in 28nm UTBB FD-SOI technology, (iii) the design of a probabilistic spintronic synapse and the development of a specific control sequence which allows for online modification of the synaptic weight, (iv) the development of a fault injection platform, which allows for the analysis of artificial neural networks (both spiking and formal) under the effect of various types and densities of faults.

2.6.2 Variability of fully-connected spiking neural network

I have performed a comprehensive reliability analysis of spintronic-based functional modules for spiking neural networks by providing an in-depth study of meaningful threats. In addition, I have analysed the behaviour of a fully-connected network under the observed effects of variability. I have performed the reliability analysis on a functional module made of a compound spintronic synapse and spiking neuron. I have evaluated the effect of cycle-to-cycle and device-to-device variability, as well as the effect of non-nominal environmental conditions on the learning and read operation modes of the spintronic functional module.

The neural network under study is a two-layer (one input, one output), fully-connected (all input neurons connected to all output neurons) neural network constructed with spiking neurons and analog-like synapses (as illustrated in Fig. 2.44a). It is a spiking neural network in which the input signals are independent trains of spikes and the output neurons function with lateral inhibition on a winner-takes-all strategy. When an input neuron spikes, a voltage pulse is applied on the pre-synaptic terminal (lack of spike translates in the absence of the voltage pulse). This voltage pulse is modulated through the corresponding synaptic weight. All resulting weighted currents are applied to the post-neurons until one of them fires. This prevents all other post-neurons from firing via the lateral inhibition.

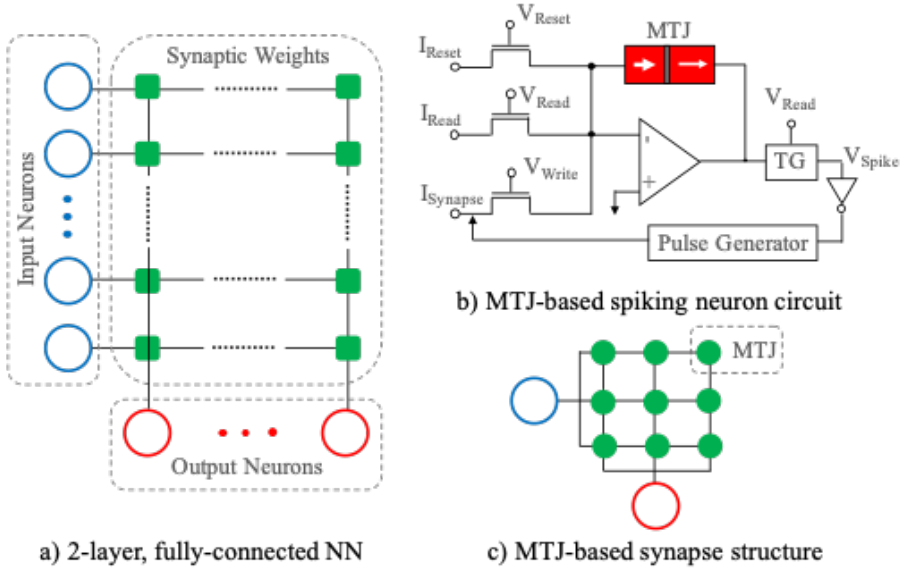


Figure 2.44: Schematic representation of the neural network under study

Artificial Spintronic Neurons

The SNN under study is built using spiking neurons. The spiking neuron uses an MTJ device which can be written to, read from, or reset (see Fig. 2.44b). The MTJ device is initialized at low resistance state (parallel configuration). During the write operation, the signal $Vwrite$ is enabled while the $Vread$ and $Vreset$ signals are disabled. As a result, the weighted potential is applied to the MTJ device via the inverting amplifier. If this signal is large enough, will switch the state of the device. The magnitude of the weighted potential determines the probability that the MTJ device switches states. After the write operation is completed, the neuron state is read by enabling $Vread$ signal and disabling $Vwrite$ signal. As a result, a voltage signal $Vout$ is generated which depends on the MTJ resistance state and on the $Iread$ current. If the MTJ has switched to high resistive state (anti-parallel magnetization), the $Vout$ signal will trigger the pulse generator circuit. Once the information is transferred, the MTJ device is reset (re-written to parallel magnetization state) by enabling the $Vreset$ and disabling the other two signals.

Artificial Spintronic Synapses

The link between the input neuron (also called pre-synaptic neuron) and the output neuron (also called post-synaptic neuron) is implemented with a compound magnetoresistive synapse (CMS) as shown in Fig. 2.44c. This synaptic device employs multiple (N) binary MTJ elements connected in parallel. They operate in stochastic regime and act as one single synapse. This CMS is expected to exhibit $CL = N + 1$ discrete conductance levels obtained by summing up the parallel conductance, ranging from the minimum compound synaptic weight – achieved when all MTJs are in high resistive state (anti-parallel magnetization) – to maximum compound synaptic weight – achieved when all MTJs are in low resistive state (parallel magnetization).

Unsupervised learning process is used to achieve the desired synaptic weight, i.e., the simplified stochastic spike timing-dependent plasticity (STDP) learning rule. The synapse plasticity is controlled by the joint effect of pre- and post-synaptic pulses, i.e., the signals coming from the input/output neurons. The voltage drop at the proposed CMS reads the difference between the pre- and post- synaptic voltages. A synapse connecting the fired post-neuron and the spiked

pre-neurons increases its weight with a certain probability, resulting in a stochastic long-term potentiation (SLTP). A synapse connecting the fired post-neuron with an idle pre-neuron decreases its weight with a certain probability, resulting in a stochastic long-term depression (SLTD). The probabilities for synaptic potentiation and depression depend on the programming conditions of the MTJ (the duration and magnitude of the programming voltage pulse), the physical characteristics of the device and the environmental conditions (such as temperature).

MTJ - Sources of Variability

- *Physical device parameters*: the Gilbert damping coefficient (α), the electron polarization percentage (P) and the uniaxial anisotropy (Kv);
- *Geometrical device parameters*: the thickness and area of the free ferromagnetic layer (tfr, Afr), and the thickness of the tunneling oxide barrier (tox). The volume of the free layer directly impacts the height of the energy barrier, hence the efficiency of the write operation, the area of the free layer affects the MTJ resistance, while the thickness of the tunneling oxide barrier affects the TMR of the MTJ.
- *Operation conditions*: write time (tw), write current (Iw) and temperature (T).

The major sources of process variations in the MTJ electrical response include variations in the tunneling oxide thickness (tox), the thickness and the cross-section area of the free ferromagnetic layer (tfr, Afr). Variations in these parameters result in a spread of RH and RL values.

Synapse $\&$ Neuron Reliability Analysis

I have implemented the MTJ-based spiking neuron circuit and the MTJ-based synaptic structure in $Cadence6.1.6$ $Spectre$ using a compact model for the MTJ device. An MTJ device with Perpendicular Magnetic Anisotropy (PMA) was used, composed of $CoFeB/MgO/CoFeB$ thin films which, involve low threshold switching currents and relatively high TMR ratio.

In this context, I have performed comprehensive analysis to evaluate the effect of fabrication-induced and environmental parameter variation (i.e., operation temperature and supply voltage) on the probability of successful MTJ write operation. Figure 2.45 illustrates the dependence of write probability on the applied control voltage under different scenarios. In Fig. 2.45a the effect of write time on the write probability is illustrated. As expected, it can be seen that the write probability increases with control voltage amplitude and its duration. In Fig. 2.45b the effect of temperature variations on the write probability is illustrated. A constant pulse duration of 60ns is chosen for illustration purpose only. It should be noted that the write probability increases when the temperature increases. This temperature effect is stronger at intermediary switching probabilities. In Fig. 2.45c the effect of fabrication-induced process variability (physical and geometrical parameters) on the write probability is shown. The simulations were performed on 1000 devices by performing Monte Carlo simulations at a temperature of 300K with a control pulse duration of 60ns. The results show that process variability has a dominant impact on the MTJ switching probability, resulting in very wide distribution of switching probabilities, practically covering the entire value range between 0% and 100%.

In continuation, I explain and illustrate how the variable MTJ switching probability affects the behavior of the neuromorphic functional modules under study.

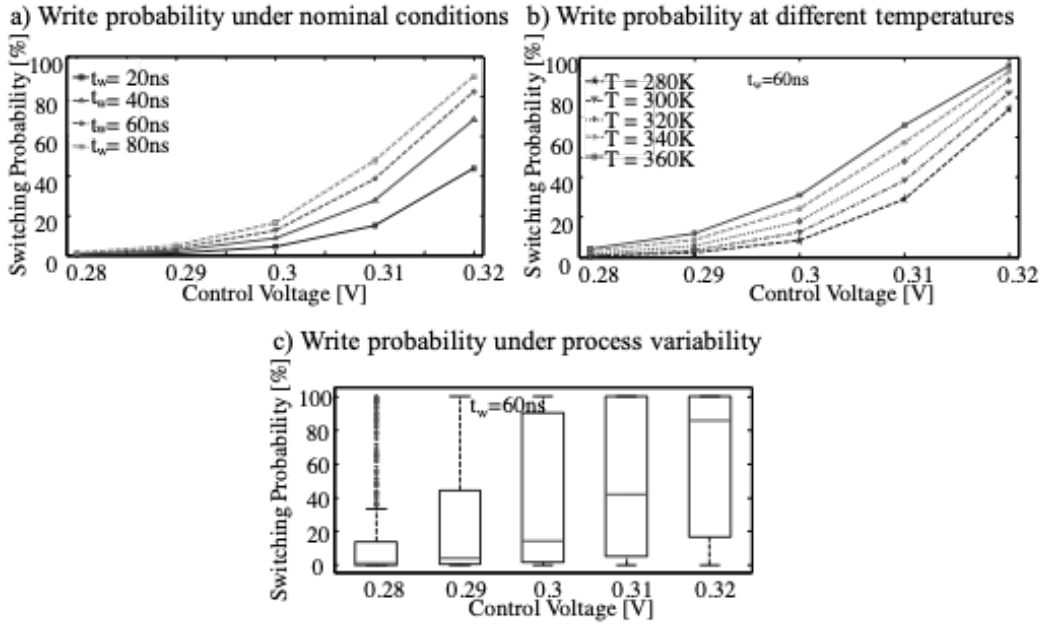


Figure 2.45: Box-plot of STT-MTJ write probability estimation

Reliability of Artificial Neuron

The spiking neuron is built with an MTJ device and an inverting amplifier, as illustrated in see Fig. 2.44b. This neuron has three operation modes: write, read, and reset. During the write operation, the equivalent synaptic voltage is inverted and applied to the MTJ, initially in low resistive state. This inverted voltage can switch the state of the MTJ with a certain probability. A write operation is followed by a read operation. If the MTJ has switched to high resistive state, the voltage at the output of the amplifier is large enough to activate the pulse generator (i.e., the neuron spikes). This operation is followed by a reset, in which the state of the MTJ is returned to low resistive. The reset operation is performed with high control voltage to maximize the switching probability. If, on the other hand the MTJ's state was not flipped by the write operation, the subsequent read operation will not generate enough voltage at the output of the amplifier to activate the pulse generator (i.e., neuron does not fire). In this case, the read operation is followed by another write operation.

While the read and reset operations are controlled in such a way to compensate for reliability issues, the stochastic write operation is strongly affected by process parameter variations and variations in the environmental conditions. Extrapolating from the curves illustrated in Fig. 2.45, the switching probability of the neuron's MTJ (i.e., the probability that the neuron will spike) is dependent on the amplitude of the synaptic current and on the operation temperature and it is highly dependent on process variability. The sum of these effects, can cause a decrease in the neuron firing rate, which in turn, slows the learning process and decreases its accuracy, and, in extremis, causes an evaluation error of the neural network.

Reliability of Artificial Synapse

The compound synaptic device under study is designed with multiple (N) binary MTJ elements connected in parallel. These MTJs all operate in the stochastic regime and are expected to exhibit $N + 1$ discrete conductance levels obtained by summing up the parallel conductance. The synaptic weight is given by this equivalent conductance:

$$W_i = \frac{N-i}{R_H} + \frac{i}{R_L} \Rightarrow \begin{cases} min(W) = \frac{N}{R_H} \\ min(W) = \frac{N}{R_L} \end{cases} \tag{2.22}$$

with i the synaptic state, given by the number of MTJ devices in low resistive state ($R_{MTJ} = R_L$). The minimum compound synaptic weight is achieved when all MTJs are in high resistive state ($i = 0$), while the maximum compound synaptic weight is achieved when all MTJs are in low resistive state ($i = N$). Therefore, larger number of MTJ devices translates in larger number of possible synaptic weight levels, which in turn translates into a larger precision of the learning process, therefore better neural network success rate.

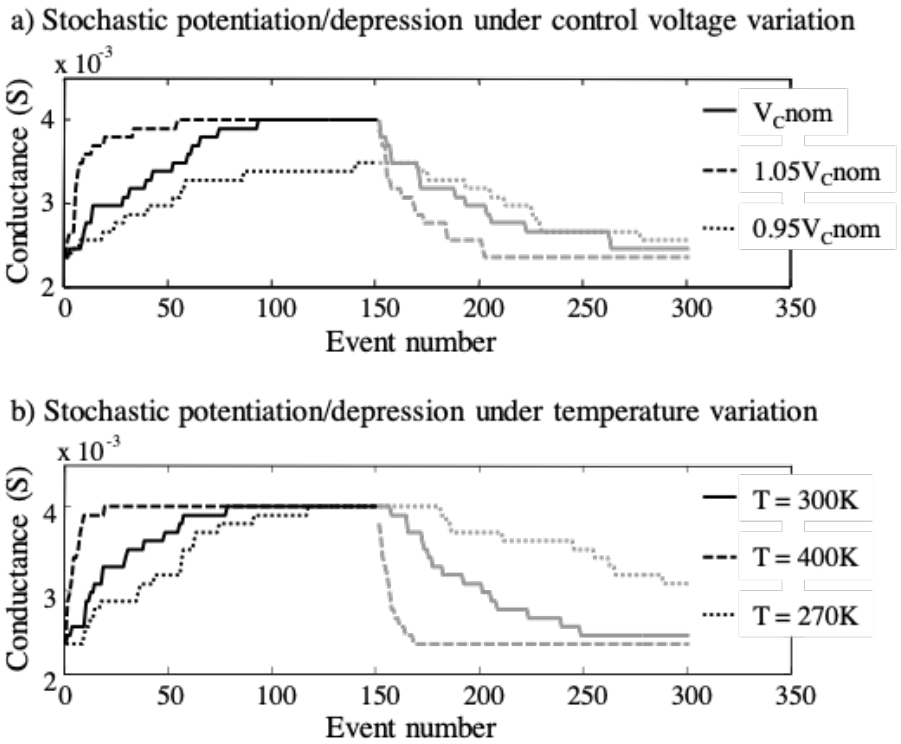


Figure 2.46: The CMS (N=16) conductance evolution under 150 SLTP operations (in black) followed by 150 SLTD operations (in grey)

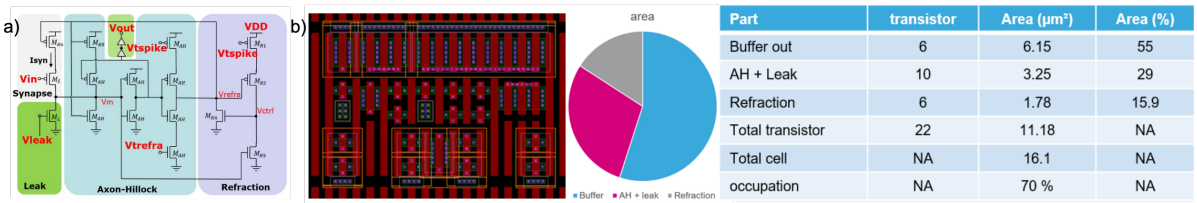
In addition, the number of conductance (synaptic weight) levels is dependent on the amplitude and duration of the control signal (V_C). A large control signal or a large pulse duration (or both) result in a large MTJ switching probability (Fig. 2.45a). This can lead to a reduction in the number of conductance levels, since the probability of two or more MTJ devices to flip during a single SLTP/SLTD operation increases. This effect is illustrated in Fig. 2.46, where the results obtained by performing 150 stochastic potentiation (SLTP) operations followed by 150 stochastic depression (SLTP) operations on a CMS with N=16 MTJ devices assuming different amplitudes of the control voltage (same pulse duration). When the operations were performed at nominal control voltage, the CMS achieved 17 conductance levels during potentiation, and 17 conductance levels during depression (continuous lines in Fig. 2.46a). On the other hand, when the operations were performed at 5% larger than nominal control voltage, the CMS achieved 12 conductance levels during potentiation, and 11 conductance levels during depression. On the other hand, a small control signal results in a low MTJ switching probability. This can lead to an increase in the number of operations (events) needed for weight increase/decrease, making the learning process slower. In other words,

boosting the control signal decreases the learning precision, while lowering the control signal increases the learning time. A similar behavior is observed under temperature variation and illustrated in Fig. 2.46b. We observe that at high temperature the number of conductance levels decreases, whereas at low temperature larger number of events are required to reach the same conductance level.

2.6.3 The design of a versatile analog spiking neuron in in 28nm UTBB FD-SOI technology

Analog based neuromorphic circuits, such as analog SNNs are bio-inspired and show great potential to provide energy-efficient data-centric computing solutions. This has led to massive research efforts being directed on one side towards brain-inspired learning algorithms that can be performed by these structures and on the other side towards the actual hardware design and the optimal implementation of SNNs. In this context, our research was focused on a low power, analog spiking neuron circuit design which it is implemented and validated in in 28 nm ultra-thin body and buried oxide (UTBB) FD-SOI with High-K metal gate. We have performed a comprehensive analysis of an analog spiking neuron designed in 28 nm UTBB FD-SOI CMOS technology from STMicroelectronics. Thick oxide MOS transistors with regular VT are selected to perform a low power design enhanced by using sub-threshold bias. The neuron operation is analysed through SPICE simulations performed at 0.7 V nominal voltage bias. Several parameters control the shape of the neuron response, such as: the synaptic excitation, the spike duration, and the refractory period. The neuron under study was fabricated and characterized for bias voltages ranging from 1.8 V standard bias to the measured limit of 0.19 V. We have also demonstrated the full functionality of the analog spiking neuron in FD-SOI.

The proposed optimized analog neuron design and its custom layout based on CMOS transistors with thick oxide and regular threshold voltage (VT) are shown in Figure 2.47 as follows: the circuit schematics, the circuit layout, and the evaluation of area consumption per neuron sub-functionality. In the circuit schematic, five regions are highlighted: the synapse region, generating the synaptic current, the leak region that generates and controls the membrane leakage current, the out region containing the output buffer chain, the refraction region ensuring the neuron compliance to the refraction period, and finally the Axon-Hillock region that performs the core functions of the neuron, i.e., integration and comparison. The refraction region is an add-on which produce a clear delimitation at the end of the refraction period. The circuit uses VDD, Gnd to bias the design in the $1.8 \div 0.19$ V voltage range; Vin is the input signal (synaptic excitation); Vout is the output signal; Vtrefra controls the duration of the refractory period; Vtspike controls the spike duration; Vleak controls the neuron membrane leakage; Vm is the neuron membrane potential; Vrefra is an integrating node used for spike and refraction duration; Vctrl is a control signal to end the refraction period.



Part	transistor	Area (µm²)	Area (%)
Buffer out	6	6.15	55
AH + Leak	10	3.25	29
Refraction	6	1.78	15.9
Total transistor	22	11.18	NA
Total cell	NA	16.1	NA
occupation	NA	70 %	NA

Figure 2.47: Standard design in 28 nm FD-SOI with a) schematic b) layout and area cost

The circuit design is initially performed at 0.7 V supply voltage and the integration duration of 1 µs, followed by 1μs spike width and the duration of the refractory period of 1 μs. To

validate the functionality of the proposed neuron design, SPICE simulations were performed until deep subthreshold regime, expressed by bias voltage lower than 0.7 V down to the 0.19 V limit at room temperature and on typical process corner. Figure 2.48a) reports the temporal response for relevant nodes of the proposed neuron design. It should be noted that the circuit behaves in agreement with its specifications. The spike energy has been computed and is equal to 8 fJ for 141 fW static power at nominal bias 0.7 V on VDD for the neuron without the output buffers. For this structure, Monte Carlo simulations are performed to give an idea of the robustness of the design. Figure 2.48b) depicts the responses dispersion in agreement with design specifications. It appears that the relative standard deviation of cycle duration is 55%.

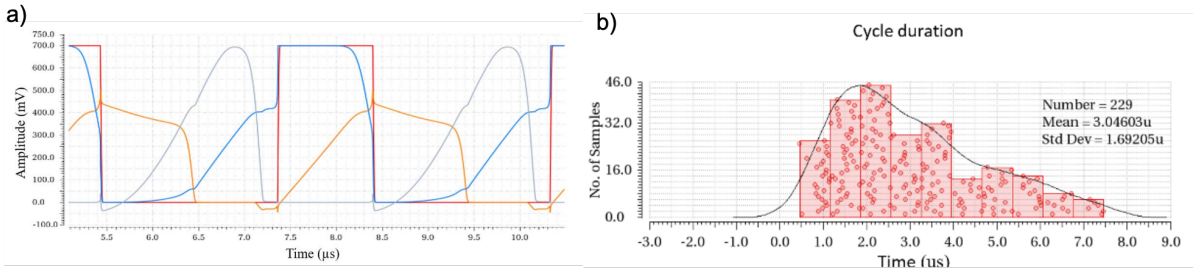


Figure 2.48: (a) Typical time simulations in 3 x 1μs period with nodes tracking, Vout (red), Vrefra (orange), Vm (blue) and Vctrl (grey); (b) Typical Monte Carlo simulations in 3 x 1μs period for global and local process variations.

The exploration of neuron functionality has been performed in at different voltage regimes: at VDD=0.7V (nominal value), at VDD = 0.19V (sub-threshold regime) and at VDD = 1.8V. We have observed that the neuron functions correctly at all supply voltages and that the energy per spike ranges between 2.8fJ and 28pJ depending on the supply voltage and synaptic excitation.

2.6.4 The Design of a Probabilistic Spintronic Synapse

Recent developments in neuromorphic computing aim to implement these SNNs in hardware to fully exploit their potential in terms of low energy consumption. We have demonstrated the plasticity of a multi-conductance state synapse in SNN is shown. The synapse is a compound of multiple Magnetic Tunnel Junction (MTJ) devices connected in parallel. The network performs learning by potentiation and depression of the synapses. In this work we show how these two mechanisms can be obtained in hardware-implemented SNNs. We developed a methodology to achieve the Spike Timing Dependent Plasticity (STDP) learning rule in hardware by carefully engineering the post- and pre-synaptic signals. We demonstrate synaptic plasticity as a function of the relative spiking time of input and output neurons only.

In the SNN under study, MTJ-based synapses are used. The synaptic weights are expressed in conductance levels. Various conductance levels for a single synapse are obtained by connecting multiple MTJs. Indeed, a single MTJ device is capable of only 2 conductance levels (Figure 2.49a)), but by connecting several MTJ devices in series or parallel, multiple conductance levels can be achieved. We have proposed a compound synapse designed with multiple MTJs connected in parallel. For demonstration purpose only, I show here the study of a synapse built with 4 MTJs connected in parallel as shown in Figure 2.49b), the two terminals Vpre and Vpost are the spiking voltages of presynaptic and postsynaptic neurons respectively. Four MTJs per synapse is largely sufficient to learn simple datasets like MNIST, this latter

has even already been trained with only two states synapses using binarized neural networks. By using (n) number of MTJs in parallel, we get (n+1) levels of conductance in the synapse. This synapse is capable of 5 conductance levels, the smaller conductance is achieved when all MTJs are in anti-parallel orientation (high restive state) and the larger conductance is achieved when all MTJs are in parallel orientation (low restive state). Intermediary conductance values are obtained when some MTJs are in antiparallel state while the others are in parallel state. In order to achieve the different conductance levels, the MTJs should be able to switch their states independently.This condition can be met by controlling the MTJs in the probabilistic region, by taking advantage of the thermal switching.

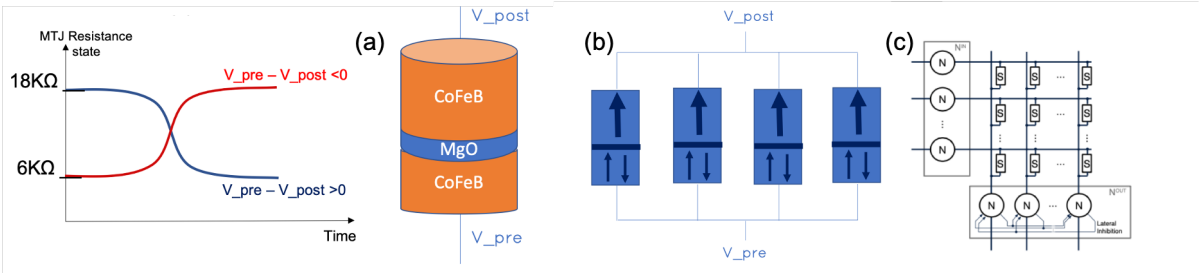


Figure 2.49: (a) The MTJ resistance as a function of time during switching (the MTJ has only two stable states) together with a schematic of MTJ, the upper layer's magnetisation is pinned, whereas the lower one is controlled by the value of the current passing through it and its direction. Vpre and Vpost represent the control signals, which in the case of the SNN; (b) Schematic of a compound synapse with 4 MTJs in parallel, (c) schematic of the single layer SNN implement as a synapse crossbar array showing also the input and the output neurons

We focus in this study only on the synapse and we analyse its behavior during the network training process. Our goal is to design the signal profiles for the pre- and post-synaptic spikes respectively, which allow for the implementation of the STDP learning. We carried out electrical simulations of the synapse presented in Figure 2.49b), where the signals Vpre and Vpost are chosen so that the voltage-drop (Vpre – Vpost) can change the conductance of the synapse to emulate time-dependent potentiation and depression. In order to tune the synapse conductance, the MTJs should be able to switch their states independently. This condition can be met by controlling the MTJs in the probabilistic region. The probability of MTJ switching depends directly on the applied voltage: amplitude and pulse width. In other words, we can use the MTJ stochasticity to our advantage and control the compound synapse in such a way, that when applying a voltage pulse of a certain amplitude, each MTJ will have a different switching delay. In this case, if the synapse is at minimum conduction level (all MTJs in high resistive state) there exist a positive voltage of amplitude Vprob which, when applied across the synapse, will cause the MTJs to switch one by one to the low-resitive state. In this way, the conductance of the synapse will gradually increase to its highest value (all MTJs in low resistive state), thus effectively emulating the synaptic potentiation. In a similar manner, the depression can also be emulated. The purpose of this work is to identify the value of Vprob for synaptic potentiation and depression and to optimise it for STDP learning. Network training consists of adjusting the synaptic weights (i.e., the conductance). In order to assure time-dependent learning the voltage drop (VpreVpost) across the MTJs should be delay-dependent as shown in Figure 2.50. As the pre- to post-synaptic spiking delay increases, we distinguish five behaviours in this order: 1) High potentiation: If the postsynaptic neuron spikes immediately after the presynaptic neuron, the voltage drop (Vpre Vpost) $>$ 0 is maximum so the conductance of the synapse is raised significantly. 2) Low potentiation: The voltage drop still positive but with smaller amplitude as the delay increases,

the conductance of the synapse is then raised by a small amount accordingly. 3) Unchanged conductance: This is a transition region between potentiation and depression, the voltage drop decreases, it can no longer perform potentiation. It decreases more as the time delay increases, it becomes negative, but dose not perform depression either because it doesn't exceed the negative threshold voltage yet. 4) Low depression: The large delay allows the voltage drop (Vpre – Vpost) to have a negative amplitude which is enough to lower the conductance level of the synapse. 5) High depression: If the postsynaptic pulse arrives very late compared to the presynaptic pulse, the voltage drop will have a very large negative amplitude, the synaptic connection is then penalized by lowering its conductance drastically. The Figure 2.50shows how the plasticity in the synapse can be obtained by tailoring the shapes of presynaptic and postsynaptic pulses. The Vpre pulse has a triangular shape with a positive and a negative part, this decreasing voltage shape converts the increasing delay into a gradual decrease in voltage drop.

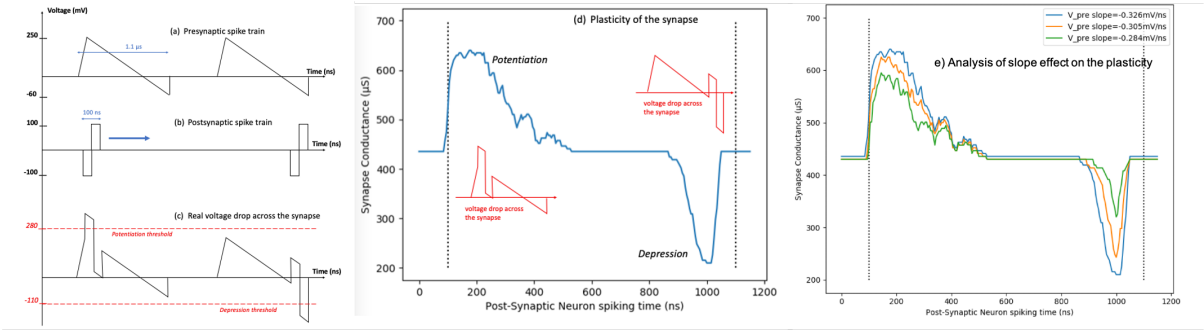


Figure 2.50: (a)The presynaptic neuron signal, The pulse Vpre has a triangular shape of 1µs. (b) The postsynaptic neuron signal, The pulse Vpost has positive and negative rectangular parts with a total duration of 100ns. (c)left: The voltage drop Vpre-Vpost across the synapse is positive and rises above the threshold when Vpost arrives short after Vpre resulting in a potentiation. (c)right: The voltage drop Vpre-Vpost across the synapse is negative and falls under the threshold when Vpost arrives at the end of Vpre pulse, resulting in a depression. (d) The average conductance of the synapse over 20 simulations for each position of the Vpost. The Vpre is fix, it always starts at 100ns, it is delimited by the two vertical lines. Whereas Vpost comes at diffident times (before, during and after Vpre) the conductance is probed only at the end of each simulation, (e) Potentiation and depression of a synapse initialized at an intermediate state of conductance, the presynaptic pulse is fixed at 100ns and is delimited by the two vertical line. Three simulations with different pulses are shown. A better result is obtained for slope of 0.326mV /ns.

This work demonstrates a simple way to implement the synaptic connections using the magnetic tunnel junctions. We successfully reproduced the most important parameter for the STDP training rule which is time-dependence. The training which consists of adjusting the synapse conductance is only controlled by the time delay between the spikes of the connected neurons to the synapse. This way the synapse is designed to be used in a bigger network to allow unsupervised learning. The voltage pulse shapes of input and output neurons are chosen in such a way to allow the occurrence of the desired operation (potentiation or depression) and its magnitude at the right time. Moreover, potentiation and depression occur only when the two incoming pulses arrive, making the design suitable also for inference, where the synapse conductance stays unchanged when only one spike arrives in the synapse.

2.6.5 Definition of Fault Models for Spiking Neural Networks

In respect to network reliability, it's commonly assumed that neural networks have a built-in fault-tolerance property due to their parallel structures, or fault-tolerant training algorithms. However, recent works have shown that with the possibility to integrate neural networks more thorough control of fault tolerance is required. In this context, we have developed fault models that will enable fault injection campaigns and that will allow identifying scenarios of faulty operations, happening before and after the on-line learning. In this work, we have only considered permanent faults caused by manufacturing defects and aging-related phenomena. Due to the fact that there are a large number of SNN circuit implementations, and the number keeps growing, we chose to define realistic fault models, without the need of the full knowledge of the hardware implementation. Thus, we define a functional set of possible faults that can affect the elements belonging to the SNN, neurons and synapses. In particular, we define how the inputs and outputs of the functional interface of the neurons and synapses can be affected by the faults, while considering the hardware root causes that can lead to those faults. These faults are similar to, for instance, the stuck-at, where the fault is defined at the interface of a logic gate, without the knowledge of the actual transistor-level implementation of the gate, but still being representative of the majority of physical defects that may appear at the transistor level.

Modeling of synaptic faults

We define the following fault models:

- *Dead Synapse Fault (DSF)*, defined as a synaptic connection that does not allow information transfer from input to output neuron. This is a permanent fault, representative of defects strongly affecting the non-volatile resistive or magnetic memory element (such as breakdown) or faults affecting either the switch or the OR gate in such a way that the switch is always opened.
- *Degraded Plasticity Fault (DPF)*, defined as a synaptic weight that is not able to store the whole range of possible values. This fault is representative of defects affecting the non-volatile resistive or magnetic element (such as agingrelated effects). This fault model is a permanent, parametric fault, and has three parameters, i.e., the minimum and maximum weight that can be reached by the synapse, and the number of conductive levels.
- *Synapse-Stuck-At-0 and Synapse-Stuck-At-1 (SSA0, SSA1)*, defined as a synaptic weight that stores permanently either the minimum or the maximum weight (these two faults are two extreme cases of the above-defined DPF).

Modeling of neuronal faults

We define the following fault models:

- *Dead Neuron Fault (DNF)*, defined as a neuron that does not fire under any conditions. This permanent fault is representative of defects affecting any element within the neuron that would prevent the pulse generator to generate spikes.
- *Input/Output Stuck Lateral Inhibition Fault (ISLIF, OSLIF)*, defined as a neuron that is not able to receive the lateral inhibition information from other neurons (ISLIF) or not

able to transmit such an information (OSLIF). These are also considered as permanent faults.

- *Input/Output Delayed Spike Fault (IDSF and ODSF)*, defined as a spike (as PPost for the input, and as PPre for the output) that happens with a delay or an anticipation, compared to the correct behavior. This fault is representative of defects affecting the time constant of the integrator, or the comparator within the neuron, leading to an anticipated or delayed triggering of the spike. This particular fault is a parametric, delay fault type.
- *Input/Output Delayed Synapse Activation Fault (IDSAF and ODSAF)*, defined as a synaptic activation that happens with a delay or an anticipation. This fault is similar to the IDSF/ODSF, but it affects the signal enabling the synapse activation. IDSAF is pertaining to output Neurons, while ODSAF is pertaining to input Neurons. These faults can be random transient or intermittent faults.
- *Input/Output Delayed Lateral Inhibition Fault (IDLIF and ODLIF)*, defined as a lateral inhibition signal that happens with a delay or an anticipation. These delay faults can also be random or intermittent type.

2.6.6 Fault Injection Platform for Artificial Neural Networks

Neural Networks (NNs) are becoming ubiquitous today. They are (or will be in the near future) present in many applications, including safety-critical ones, such as autonomous driving, aircraft collision avoidance, and malware detection. In this context, the presence of faults might lead to dreadful consequences. Even if NN algorithms are intrinsically fault tolerant due to their structure redundancy, it has been shown that hardware implemented NNs don't fully inherit this property. This is the case for both formal and spiking neural networks.

Figure 2.51a) shows a generic architecture for ANN computation. In this architecture, the synaptic weights and biases are stored in the non-volatile memory (NVM) and they are loaded into the main memory (DRAM) at power on. The compute engines load the weights from the DRAM to their caches and perform the neural computation. Intermediate computations are stored in the local memory. Figure 2.51b) shows the generic schematic representation for a fully connected layer of a Spiking Neural Network (SNN). The correctness of the network operation depends on the precision of neural calculation and on the correctness of the weight and bias storage. In the case of an ANN, since the synaptic weights and bias values are stored and transferred through four types of memories, any one of them can contribute to an erroneous calculation if affected by a fault. Moreover, a fault affecting the neural calculation can be modeled as an equivalent fault in the local memory. The same assumption stands for the SNNs as well. These considerations bring us to the conclusion that the study of the network accuracy under the assumption of faulty memory has the potential of giving relevant insights on the operation of a neural network (NN) on faulty hardware. Here, the considered fault model is the bit-flip. The goal of this work is to evaluate the effect of memory faults on the accuracy of neural computation.

We have developed a fault injector to evaluate the effect of faults affecting the memory storing the synaptic weights on the overall accuracy of an NN. The proposed fault injector receives as input the characteristics of the trained network (i.e., its topology and the values of the synaptic weights and biases) and the injection parameters (i.e., the number, the distribution and the location of faults). In addition, in the case of ANNs or digital SNNs, the fault

injector is in charge of the digitalization of the synaptic weights, to allow the analysis of fault injections on different coding resolutions. The tool is able to inject bit flips, with a user-defined probability. The random fault locations are uniformly distributed over the targeted memory arrays. Figure 2.51c) shows the implementation details of the injector.

The first step is the digitalization of the synaptic weights and biases (line 1). Then the tool calculates the total number of bits on which the fault injection is performed (line 2). Here we can distinguish between a fault injection on the entire memory storing the weights, or limiting the injection to a particular layer. The total number of faults to be injected is calculated based on the targeted fault probability and the total number of bits on which the fault injection is performed (line 3). Due to the randomness of the fault injection, several campaigns need to be conducted, in order to have statistically relevant results (line 4). Each fault campaign needs to be initialized with the correct values of the weights (line 5). For each fault to be injected, the tool randomly selects its position (line 6) and inverts the corresponding bit of the weight (line 8), emulating a bitflip. Finally, the network (with faulty weights) is fed-forward for inference and its accuracy is evaluated.

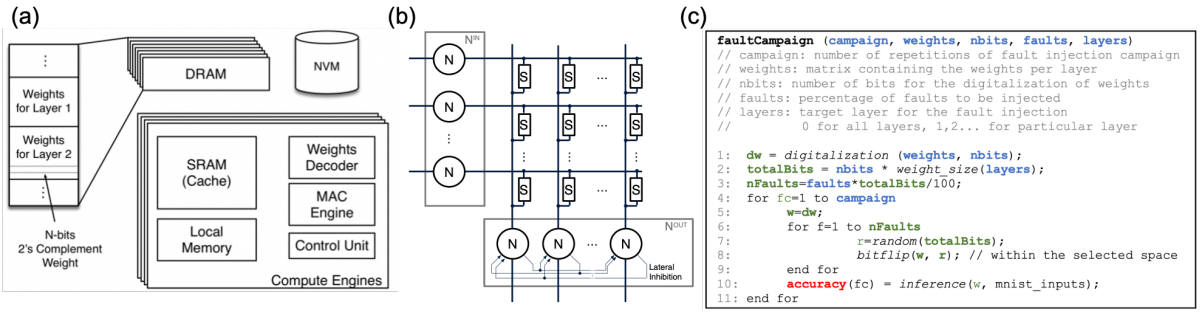


Figure 2.51: (a) Generic IA chip; (b) schematic of the single layer SNN implement as a synapse crossbar array showing also the input and the output neurons; (c) Fault Injector

2.6.7 Analysis of SNN Fault Tolerance

We have implemented a spiking neural network with learning strategy based on spike-timing dependent plasticity. The network is designed to solve the MNIST database, i.e., to be trained to recognize hand written digits. This data base has 60000 examples for the network training and 10000 examples for testing the network. Each example consists in the image of a hand-written digit. The hand-written digit is a 28x28 pixels image in grey-scale (256 tones of grey from white to black). The information carried by each image is transmitted to network in the form of spikes. The spike encoding is performed by frequency encoding of each pixel's tone of grey. With this encoding, the black pixels carry no information, while the white pixels carry the maximum amount of information, i.e., maximum frequency (255 spikes per time unit). Each image is presented to the network for 10 time units. In order to respond to the requirements of this data base, the network is designed with 784 input neurons, one for every image pixel. The input neurons are connected in a one-to-all fashion to the output neurons. We have performed an initial study to assess the efficiency of the implemented training strategy, by using different sizes of the output layer. The results are shown in Figure 2.52a).

As expected, the network precision is higher with the size of the output layer is large, as more patterns are learned. However, this increase is not linear and it becomes less significant as the number of neurons increases. For example, when the output layer size increases from

400 neurons to 600 neurons, the recognition rate is increasing by 1% only. In addition, the simulation time becomes more significant, as the number of computations performed during learning increases. Thus, for the future analysis presented in this paper we are using the network with 400 output neurons (with 313600 synaptic connections) since the simulation time is manageable and the network accuracy is high enough.

Further to that, fault injection campaign is performed, for different scenarios of fault occurrence assuming clustered faults or unclustered (random or deterministic distributed) faults. The functionality of the network has been evaluated and the severity of the fault occurrence related to its frequency and location has been assessed. It should be noted that in this work we only consider the case where either neurons (input or output) or synapses are faulty. The scenario where a combination of such faults occurs is not considered, being one direction of future research. Another assumption used in this work is that all faulty components are suffering from the same fault (unique fault model is used in this analysis). In addition, all faults have the same magnitude. The scenario where a combination of such faults may occur (with different magnitudes) is not considered and is part of our future research. These assumptions limitations come from the complexity of the problem that may lead to extreme exploration space for complete analysis. These setups are applied of the network HW implementation where the synapses are placed in a grid, controlled column-wise by output neurons and row-wise by input neurons and where the neurons are placed in rows/columns around the synaptic array. This present setup is sufficient to demonstrate the relevance of the proposed fault models in the operation of the targeted SNN.

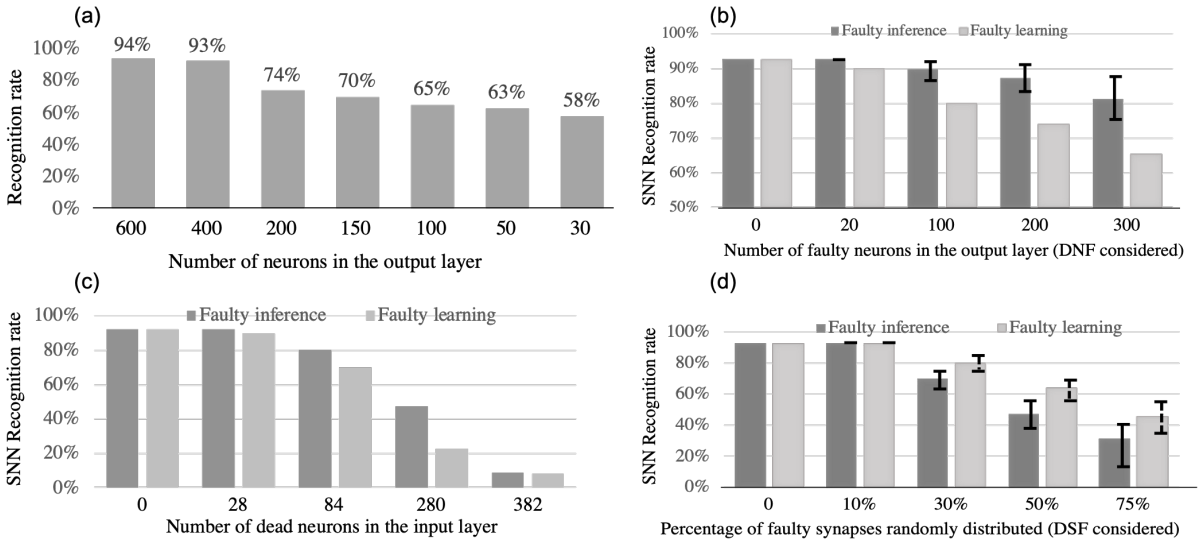


Figure 2.52: (a) STDP-based SNN recognition rate as a function of the size of the output layer; (b) STDP-based SNN recognition rate as a function of the number of output neurons affected by DNF; (c) STDP-based SNN recognition rate as a function of the number of input neurons affected by DNF; (d) STDP-based SNN recognition rate as a function of the number of synapses randomly affected by DNF;

First, we show how DNFs affect the recognition accuracy of the SNN under study. If the faults are injected during the learning phase, their location is irrelevant, since dead neurons equals smaller size output layer. Indeed, the results show the same accuracy whether the network is simulated with dead neurons or by eliminating them from the network all together, as well as their corresponding synaptic connections. However, when DNFs are injected during the inference phase, their location is somehow relevant and the accuracy of the pattern recognition

is dependent on the patterns learned by each dead neuron. In Figure 2.52b) we represent the average values, together with the minimum and maximum values of the recognition rates obtained when DNFs are injected at inference. We have observed that the range of these values increases at the number of injected DNFs increases. This is due to the fact that depending the location of the injected fault the recognition accuracy of one or multiple digits in the data base can be affected. In addition, it is very important to note that the injection of DNFs in the output layer during inference is less critical (has less effect on the SNN recognition rate) than the injection of DNFs in the output layer during learning. It is worth noting also that one DNF in the output layer has the equivalent effect as full-column DSF. Indeed, if all synapses in a column of the synaptic array are affected by DSFs, the corresponding output neuron is not connected to any of the input neurons, hence it is like it is affected by a DNF itself.

A second analysis shows how DSFs affect the recognition accuracy of the SNN under study. We consider here 2 scenarios, random distribution of DSFs and DSFs affecting full synaptic row. Results show that random distribution of the DSFs has lesser impact on the recognition rate than row-wise clustered faults. Injecting DSFs on a full row of the SNN is equivalent to injecting one DNF on the corresponding input neuron. This translates in losing the information carried by 1-pixel of the input images. Therefore, one outcome of our experiments is to understand to which extent the location of these faults can be important. For simplicity, in Figure 2.52d) we illustrate the accuracy of the SNN under random DSFs injection during learning and inference, for different fault densities. The simulations were repeated 50 times to assure that different locations of the injected faults are considered. The figure includes the average values of the achieved recognition rate together with the corresponding maximum and minimum values. We have observed that if less that 10% of the synapses are affected by DSF, the SNN accuracy is not significantly affected. For larger fault densities, we observed that the accuracy decreases rapidly. In addition, results show that the network manages to learn around these faults. The study show higher network accuracy if the DSFs are injected during learning than in the case where DSFs are injected during inference.

It is important to note that different faults have different effects if they happen during the learning or during the inference stages of a network operation. Indeed the synaptic faults (DSF, DPF and SSAx) have a stronger influence during the inference stage of the SNN than during the learning stage. This is due to the fact that the network manages to learn around the faulty synapses due to the on-line learning algorithm (STDP). If the fault occurs during the inference stage, we observe a fast degradation of the recognition rate, due to the fact that the network is found in the situation of recognizing degraded patterns. The location of occurrence of synaptic faults is also very important as stated in the most-right column of the table in Figure 2.53. Indeed, if a fault (DSF, DPF or SSA0) occurs on a minimum weight - depressed synapse no effect will be observed on the network behavior. However, if a fault such as DSF, DPF or SSA0 occurs on a maximum weight - excited synapse a strong effect will be observed on the network behavior. The situation is the exact opposite for the synapses affected by SSA1.

During neuron-related fault injection campaigns (DNFin, DNFout, IDSF/ODSF, IDLIF/ODLIF and ISLIF/OSLIF) we observe the opposite effect, i.e., a stronger influence during the learning stage of the SNN than during the inference stage. This is due to the fact that at this stage, the computation element is affected, which means that during learning mode the injected fault leads to wrong behavior learning, while a fault injected during the inference leads to less recognition accuracy. Faults affecting the input neurons are the most critical, since these neurons encode the information. The effect of DNFin is strongly dependent on the location of

the faulty neuron. Faults affecting the output neurons are less catastrophic due to the intrinsic redundancy of the SNN networks with STDP, where a pattern is learned by multiple output neurons. Stuck-at fault occurring at lateral inhibition stage is the most critical, since even a single fault can cause full system failure. Indeed, if a OSLIF fault occur on one neuron, it will prevent all other neurons from firing, hence only a single pattern will be learned by the network containing features from multiple patterns, making the network unusable.

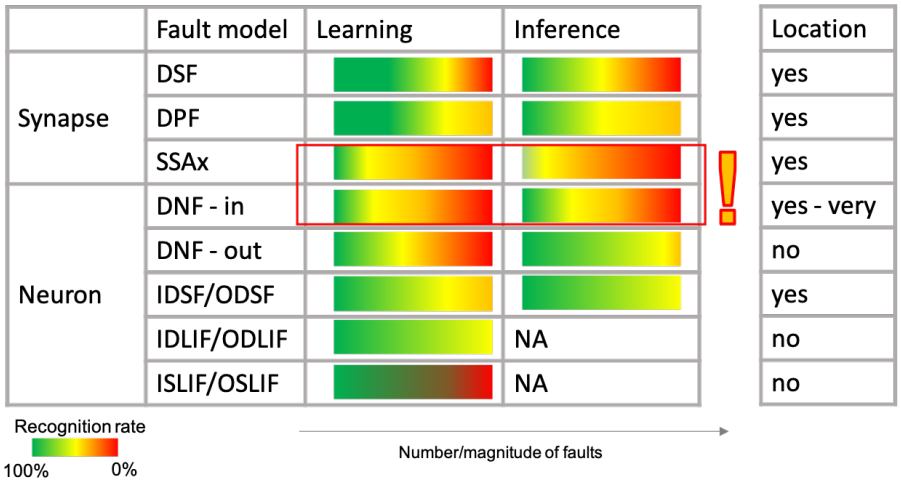


Figure 2.53: (a) STDP-based SNN recognition rate as a function of the size of the output layer; (b) STDP-based SNN recognition rate as a function of the number of output neurons affected by DNF; (c) STDP-based SNN recognition rate as a function of the number of input neurons affected by DNF; (d) STDP-based SNN recognition rate as a function of the number of synapses randomly affected by DNF;

This analysis represents a preliminary study of the fault tolerance of SNNs. Further evaluations are necessary to be able to evaluate, with high confidence the reliability of a SNN. Multiple fault injection scenarios need to be further performed to have a full picture of the network accuracy: different locations, different fault magnitudes should be studied as well as plausible clustering scenarios and combinations between synaptic and neural faults. In addition, the network should be evaluated under different application scenarios (or databases with same dimensionalities) to evaluate the fault effects also independently of the application.

CHAPTER 3

Teaching and Supervision Activities

3.1 Teaching Activities

Starting from my masters, I have performed various teaching activities at different levels: engineering students of the Technical University of Cluj-Napoca Romania, Montpellier University II France, master students of the Politecnico di Torino Italy, Polytechnic Institute of Grenoble (INPG) France, Sorbone University Paris France, PhD candidates at IMT Atlantique France, as well as embedded tutorials on specific scientific topics in international conferences. The topics of my teaching activities (focusing on academic courses only), as well as the description of the course and the targeted students are described in the following paragraphs.

Emerging topics in computing is an advanced class targeted to graduate students which have an interest in research. It introduces the student to the issues faced by traditional computer architectures and mitigation solutions currently under study by the research community.

- The covered topics are: (i) a summary description of the main concerns related to traditional computing mainly the CMOS technology limitations, the memory and the power wall; (ii) basic notions and an introduction to unconventional computing such as: approximate computing, normally-off computing, near memory computing, in-memory computing, neuromorphic computing.
- Course prepared for and presented to PhD and Master students of IMT Atlantique, FR; 2019/20, 2020/21, 2021/22

Emerging non-volatile memories (ENVM) is an advanced class targeted to graduate students which have an interest in research. It introduces the student to the issues faced by by the CMOS technology in general and the CMOS memories in particular and presents beyond-CMOS devices and their ability to eventually mitigate many of these well known issues.

- The covered topics are: (i) a summay of the main concerns related to the CMOS technology scaling and the limited performance improvement with miniaturisation (ii) the main issues faced by the computer memory and storage and the definition of the memory wall, (iii) device description and operation principles of the 3 main types of emerging non-volatile memories the PCM, the ReRAM and the MRAM, (iv) overview of the application domains which can benefit by the introduction of the ENVM.
- Course prepared for and presented to Master students of Sorbonne University, FR; 2019/20, 2020/21, 2021/22

Hardware Trust – Physiscal Security is an advanced class targeted to graduate students which have an interest in hardware security. It introduces the student to the issues related to hardware trust and provides a brief overview of the hardware security primitives such as Physical Unclonable Functions (PUFs) and True Random Number Generators (TRNGs) as roots of trust.

- The covered topics are: (i) the definition and analysis of the main types of HW counterfeiting (ii) the main detection and prevention strategies targeting these counterfeiting methods, (iii) Physical Unclonable Functions (PUFs) - types, designs, quality metrics and applications (iv) True Random Number Generators (TRNGs) - types, designs and quality metrics.

- Course prepared for and presented to Master students of Grenoble INP, FR; 2017/18, 2018/19, 2019/20, 2020/21, 2021/22.

Analyse et réalisation d'un système complexe it is a practical work (project) intended to help students master the complete design of an integrated system (digital, analog or mixed digital/analog), from the writing of the detailed specifications to the prototype, including the architecture specification, the definition of the validation plan, the implementation and the verification. The work is in the form of a group project.

- Project supervision of Master students of Grenoble INP, FR; 2017/18, 2018/19, 2019/20, 2020/21, 2021/22.

Information Technology Tools it is a practical work (laboratory) dedicated to first year students and it is targeted at bringing the student at a minimum level in the use of digital tools (web tools, data processing software, writing documents, etc.), to have a basic knowledge in computer science.

- The covered topics are: (i) Information and data - Conduct research and information monitoring (search engine, social networks), manage data (file manager, databases), process data (spreadsheet) (ii) Communication and collaboration - interact (email, video-conferencing), share and publish (sharing platforms, forum and comment space), collaborate in a group (collaborative work platform and document sharing); (iii) Content creation - develop text documents (word processing, presentation), develop multimedia documents (image/sound/video/animation capture and editing), adapting documents to their purpose (format conversion tools), programming (simple computer development, solving a logical problem)
- Laboratory work/supervision of first year students of University of Montpellier, FR; 2012/13, 2013/14.

Control Engineering it is introductory class dedicated to second year students in Computer and Control Engineering. It gives basic notions on control theory for embedded systems, properties of linear time-invariant continuous systems, feedback, feedforward control, open loop and closed loop systems.

- The covered topics are: (i) Process specifications and characteristics; (ii) P-controller and static behaviour of a control loop; (iii) Dynamic behaviour of a control loop: open loop and closed loop transfer function; (iv) Phase margin, gain margin, Bode stability criterion, loop shaping, effect of dead time in a control loop; (v) Control structures.
- Laboratory work/supervision of second year students of the Technical University of Cluj-Napoca, RO; 2005/06, 2006/07 - mainly working in Matlab and Simulink.

System Theory it is introductory class dedicated to second year students in Computer and Control Engineering. It gives basic notions on system theory emphasizing the unified treatment of continuous and discrete time systems from a state-variable viewpoint.

- The covered topics are: (i) Introduction to state-space systems and the differences between state-space and input-output models of systems; (ii) Linear state-space models; (iii) Linear time-invariant systems; (iv) Stability, controllability and observability of linear systems
- Laboratory work/supervision of second year students of the Technical University of Cluj-Napoca, RO; 2005/06, 2006/07 - mainly working in Matlab and Simulink.

Power Electronics in Automatic Control it is an advanced class dedicated to third year students in Control Engineering. It introduces the principles of torque/speed control in DC and AC motors alternatively, torque/speed control principles for special electric motors, the problem of supplying electrical circuits.

- The covered topics are: (i) Introduction to power electronics in control systems; (ii) Switching of the by-polar transistor, the MOSFET, the tyristor and Gate turn-off thyristor, the IGBT transistor; (iii) AC and DC voltage inverters as execution elements in automatic systems; (iv) controlled rectifiers as execution elements in automatic systems.
- Laboratory work/supervision of second year students of the Technical University of Cluj-Napoca, RO; 2004/05.

3.2 Supervision and Mentoring Activities

Up to now, I have co-supervised 10 master students, 7 Ph.D. candidates, 1 visiting Ph.D. candidate and 1 engineer.

3.2.1 Master Students

Arnaud Degreze: *Analysis of the efficiency of LaNiO4 memristive devices for bio-inspired learning*. Arnaud has held an internship in TIMA and LMGP from April to September 2022. His internship was focused on electrical characterization and compact modeling of LaNiO4 memristive devices and the analysis of their efficiency in the context of spiking neural networks with local learning. On the TIMA side, under my supervision, he has worked on the development and optimization of memristive compact models to be used in electrical simulations. Currently he is performing humanitary work in Peru.

Sara Mannaa: *Design and Fault-Tolerance of a Spintronic-based Spiking Neuron*. Sara has held an internship in TIMA from March to July 2022. Her internship was focused on the design and evaluation of Spintronic-based Spiking Neuron with stochastic behavior. She has worked on the evaluation of magnetic tunnel junction stochasticity when multiple devices are connected in series/parallel and how such structures can be used to emulate the accumulation/integration behaviour of a spiking neuron. Currently she is a PhD Candidate at Ecole Centrale de Lyon.

Marcelo Correa Cueto: *Design and Evaluation of Logic in Memory Solutions for FeFETs*. Marcelo has held an internship in TIMA from April to July 2022. His internship was focused on the evaluation of FeFET operation modes and their variability, designing a FeFET memory array including write drivers and sense amplifiers and performing in-memory multiply accumulate operations. During his internship he was part of a "Student Exchange" program France-Brasil and currently he is master student at Grenoble Institute of Technology.

Luis Felipe Camponogara: *Design and Evaluation of Logic in Memory Solutions for ReRAMs*. Luis has held an internship in TIMA from April to July 2022. His internship was focused on the development of in-memory computing solutions for neural network accelerators. He has developed a ReRAM-based architecture able of performing multiply-accumulate operations. He has evaluated the accuracy of the proposed architecture under several levels of precision. During his internship he was part of a "Student Exchange" program France-Brasil and currently he has returned to Brasil to complete his engineering diploma.

Matthieu Charles: *Experimental platform for ReRAM evaluation*. Matthieu has held an internship in TIMA from April to July 2022. The main objective on his internship was to set-up a measurement platform for Knowm Memristive devices. He has then used this platform (i) to characterise the electrical behavior of independent knowm devices during read and write operations and (ii) to characterise the static and dynamic device variability. Currently he is finishing his master studies and performs in alternation with working at STMicroelectronics.

Jardel Kaique: *Evaluation of laser attacks on power-off CMOS devices*. Jardel has held an internship in TIMA from May to August 2022. The main objective on his internship was to perform a thorough literature review to summarise the effect of laser attacks o the behavior of CMOS devices in order to perform a taxonomy of possible fault attacks. Currently he is finishing his master studies at LCIS.

Sergio Vinagrero: *Design and Evaluation of a OxRAM-based Security Primitives*. Sergio has held an internship in TIMA from March to July 2021. His internship was focused on the evaluation of OxRAM efficiency when used as security primitives. He has evaluated the extend at which the fabrication induced variability of the OxRAM devices is stable enough to be used for PUF design. In addition he has worked on the estimation of OxRAM cycle-to-cycle variability and evaluated the amount of entropy which can be extracted for the design of TRNGs. Sergio currently holds a PhD position in our group, working on the same topic.

Ihab Alshaer: *Physically Unclonable Function (PUF) and Random Number Generation (RNG) with Neural Networks (NN)*. Ihab has held an internship in TIMA from March to July 2020. His internship was focused on investigating the possibility of using SRAM-PUFs in conjunction with Multi-level Perceptrons (MLP) to improve the quality of PUF and increase the space of challenge-response pairs (CRP), at the same time increasing the nonlinearity of CRPs. Ihab currently holds a PhD position in LCIS, in collaboration with TIMA lab working on the topic of fault attacks on Risk V.

Solenn Guironnet: *Design of Emerging Memory-based Spiking Neural Networks*. Solenn has held an internship in TIMA from April to June 2020. Her internship was focused on the design and evaluation of Spintronic-based synapses for Spiking Neural Networks with on-line learning. She has worked on the evaluation of magnetic tunnel junction stochasticity when multiple devices are connected in parallel and how such structures can be used to emulate the plasticity of a probabilistic synapse. Currently she is finishing her master at Grenoble Institute of Technology.

Ali Bawab: *Compact Models for Synaptic-Compliant La2NiO4 Memristive Devices*. Ali has held an internship in TIMA in collaboration with the LMGP and IMEP-LaC from April to September 2019. His internship was focused the creation of an efficient compact model for newly developed LaNiO4 memristive devices. On the TIMA side, under my supervision, he has worked on the development and optimization of memristive compact models to be used in electrical simulations. Currently he is graduating his PhD at the Femto ST institute.

3.2.2 PhD Candidates

Sergio Vinagrero (supervision quota 80%) State Scholarship

Design and Evaluation of Resistive-based Security Primitives

The Physically Unclonable Functions (PUFs) exploit intrinsic manufacturing variability introduced in a device during the fabrication process to generate a signature, unique to each single device. In order to guarantee its security, the generated secret key must be unique from device to device (unclonable), and, for a same device, it must be robust with respect to aging and environmental variations (reproducible). True Random Number Generators (TRNGs) are used to generate random numbers from a physical process, rather than a computer program. They are implemented by taking advantage of thermal noise or other quantum phenomena and are expected to generate random bits with very high entropy and zero correlation. An on-chip TRNG design should occupy small area, give high bit rate, and have low power consumption, while assuring un-biased bit streams with high entropy per bit and low (no) correlation among them. The rapid development of low power, high density, high performance SoCs has pushed the embedded memories to their limits and opened the field to the development of emerging memory technologies. The Resistive Random Access Memory (ReRAM) has emerged as a promising choice for embedded memories due to its reduced read/write latency and high CMOS integration capability. Inner properties of ReRAMs make them suitable for the implementation of basic security primitives such Physically Unclonable Functions (PUFs) and True Random Number Generators (TRNGs). This thesis will explore solutions which exploit (i) the high variability affecting the electrical resistance of the resistive device to build a robust, unclonable and unpredictable PUF, and (ii) the stochastic nature of the write operation in the resistive device to generate randomly distributed numbers.

Emilien Taly (supervision quota 40%) CIFRE with STMicroelectronics

Design of a very low power Artificial Intelligence system (Tensor Processing Unit - TPU) based on in-memory computing

The data-centric computing paradigm requires the design and implementation of dedicated hardware solutions for AI gas pedals to cope with the large amount of data to be processed with minimal latency. Nevertheless, strict requirements on the non-functional characteristics of such implementations are necessary to have a usable and competitive product. Specifically, an AI gas pedal core must meet the following requirements: reusability and versatility, very low power consumption, high precision, low silicon area, while allowing parallel operations. In order to reduce energy consumption (at least 2x) and latency (at least 3x), a completely different approach must be investigated. In this context, the in-memory computing paradigm is a promising technique that minimizes data transport, the main performance bottleneck and energy cost of most data-intensive applications. This thesis will take advantage of the proven benefits of CIM to design a very low power AI accelerator.

Salah Daddinounou (supervision quota 100%) Project ANR EMINENT

Test and Reliability of Emerging Memory-based Spiking Neural Networks

This PhD thesis focuses on the robustness of hardware implementations of bio-inspired neural networks (Spiking Neural Networks) by using emerging technologies (memristors and/or spintronic devices). The innovative aspects of this research topic are related to the test and reliability aspects for HW implemented Spiking Neural Networks with on-line, unsupervised learning. The main objectives of this thesis are: (i) to provide an in-depth study of meaningful

robustness threats in SNNs: complete analysis of the faults that can occur in a Spiking Neural Network (SNN) and the assessment of fault-tolerance quality of such a network; (ii) to conduct a reliability estimation campaign: At gate level, evaluate the effect of aging phenomena, noise injection, and severe operation and environmental conditions. This analysis will culminate in a fault injection campaign. Faults will be injected in the network architecture and the system behavioral and its robustness, failure rate and mean time to failure (MTTF) will be evaluated. (iii) to provide a manufacturing test strategy and design-for-test solutions: Both functional and structural test strategies will be targeted; (iv) to provide a strategy for architecture robustness improvement: demonstrate the characteristics of a testable, reliable architecture of SNN with on-line learning and optimize the architecture for fault tolerance, accuracy, size and power consumption.

Pietro Inglese (supervision quota 65%) State Scholarship

Exploration of security threats in In-Memory Computing Paradigms Security is critical today for information and communication technologies

It is the basis for obtaining confidentiality, authentication, and integrity of data. Improving the attack resilience of secure devices is a major challenge today due, in part, to the accelerated race among the developers and the attackers, and also to the heterogeneity of new systems and their ever-increasing numbers. The vulnerabilities of electronic devices that implement cryptography functions have been well studied in the last decade for von Neumann computing architectures, designed with CMOS technology. However, there is little evidence that these studies hold true for novel computing paradigms with new technologies. In-memory computing paradigm is an emerging concept based on the tight integration of traditionally separated memory elements and combinational circuitry. It allows the minimization of the time and the energy needed to move data across the processor. The most promising solutions for in-memory computing architectures are based on the use of emerging technologies (Spin Transfer Torque RAM and Redox RAM) that are able to act as both storage and information processing unit. Despite the promising nature of the in-memory computing- based architectures many issues related to the devices themselves and to their double usage (storage and computing unit) need still to be solved. In particular, a correct evaluation of security threats targeting systems based on in-memory computing is still missing. The objective of the thesis is to map a crypto-processor algorithm in the corresponding in-memory architecture by using different mapping approaches, thus obtaining several in-memory crypto-processor versions with different characteristics. We will then compare the resilience of these circuits to the proposed attacks.

Valerian Cincon (supevision quota 34%) CIFRE with STMicroelectronics

Design of a neuromorphic circuit in 28nm FDSOI

Based on the advanced FD-SOI CMOS technology from STMicroelectronics and the first demonstration of an original 28nm pulse neuron transistor, this thesis proposes: (i) To consolidate an analog spike-time dependent plasticity (STDP) demonstrator by TCAD 3D physical simulation: The objective is to propose and make the proof of concept via numerical simulation tools on an STDP with pre and post synapses. (ii) to design efficient spiking neurons: The objective is to study an impulse response following the same reserch methods. It is also a question of choosing the number of synapses for integration into a network. We are looking for a new method using the local gradient approach. (iii) To propose integration solutions with pre and post synaptic function: The objective is to grasp the notion of synaptic weighting and to extract a relevance of integration solution. (iv) To study a multi-layer neural network: The objective is to approach the neural architectures with the functional elements

previously elaborated. The problem of connections will be an important point in this step. The local gradient would be a way of optimization. (v) To perform simulations and propose a silicon demonstrator: The objective is to consolidate and realize prototypes in order to concretize the theoretical and numerical approaches.

Marco Indaco (supervision quota 33%) Graduated in 2014, currently he is the Global Delivery Manager at Moovit, Rome, Italy

Service-Oriented Non-Volatile Memories

The widespread usage of NAND flash memory technology has faced a surprising increment, far beyond what it was originally expected, mainly thanks to the advances in the manufacturing processes. Introducing for the first time in the literature the concept of the Service-Oriented Non Volatile Memories (SONVMs), the goal of this PhD thesis is to enhance the degree of runtime reconfigurability of an MLC NAND Flash controller, through the provision of userselectable differentiated memory access modes. In particular, the proposed solutions envision adaptivity at two different layers: architectural level and device level. The architecture layer adaptivity is based on adaptive ECC decoding structure. The physical layer adaptivity is based on an adaptive high-voltage sub-system of the Flash memory device. After having considered the flexibility and the trade-offs in the physical layer and in the ECC sub-system in isolation, joint parameter tuning has been used for high memory adaptivity to application requirements in the reliability/ performance/power optimization space.

Pascal Trotta (supervision quota 33%) Graduated in 2016, currently he is Sr. Staff Digital Design Engineer at TDK InvenSense, Milano, Italy

Enhancing Real-time Embedded Image Processing Robustness on Reconfigurable Devices for Critical Applications

Image processing is increasingly used in several application fields. As example, in automotive, cameras are becoming key sensors in increasing car safety, driving assistance and driving comfort. They have been employed for infotainment (non-critical), as well as for some driver assistance tasks (critical). The complexity of these algorithms brings a challenge in real-time image processing systems, requiring high computing capacity, usually not available in processors for embedded systems. Hardware acceleration is therefore crucial, and devices such as Field Programmable Gate Arrays (FPGAs) best fit the growing demand of computational capabilities. In addition, the reconfigurable nature of FPGA devices can be exploited to increase the system reliability and robustness by leveraging Dynamic Partial Reconfiguration features which is a great asset for safety critical applications. This thesis focuses on the development of techniques for implementing efficient and robust real-time embedded image processing hardware accelerators and systems for mission-critical applications. In order to ensure real-time performances, efficient FPGA-based hardware accelerators implementing selected image processing algorithms have been developed. Functionalities offered by the target technology, and algorithm's characteristics have been constantly taken into account while designing such accelerators, in order to efficiently tailor algorithm's operations to available hardware resources. Dynamic reconfiguration features of modern reconfigurable FPGA have been extensively exploited in order to integrate run-time adaptation into the designed hardware accelerators.

Vishal Gupta visiting young researcher (January-June 2020) PhD Candidate at the Doctoral School of Tor Vergata University, Rome, Italy

Exploration of synaptic-like characteristics of memristor devices

At the time of his visit with my group, Vishal was completing his PhD studies on the topic of memristive devices and their use in tomorrow's computing systems. His reserch work was directed towards developing compact models of memristive devices based on experimentally validated data. This memristor model is based on the device structure and also on the mechanism responsible for its switching. Additionally, he has implemented an architecture that allows the exploration of synaptic-like characteristics of memristors.

3.2.3 Supervision and Mentoring Strategy

To be able to accurately supervise and guide all my students I am organising several individual and group meetings as follows:

- Individual meetings: (i) I meet with each of my students (PhD and Master) once per week in ad-hoc meetings which can last from 10 minutes to 2 hours, depending on the progress of their work and on their state of mind. In these meetings we mostly discuss specific operational issues. (ii) Another type of individual meeting (a bi-yearly meeting) is reserved to the PhD candidates under my supervision. This is a half-day meeting intended to follow the overall progress of the thesis work and adjusting the schedule as needed.

- Group meetings: (i) Every week I host a "Journal Club" in which every student will present shortly a paper of their choosing strictly related to their research project. This presentation is generally followed by a discussion on the work presented. (ii) Twice a month I organise a "Research progress meeting" in which every student will present the progress of their work and difficulties encountered, such that the group can give their opinions, ask questions or suggest ways of improvement or solutions for the encountered issues. (iii) At the end of the year I organise a "Journal Club - Make it funny keep it short" where every student will present a scientific paper on any topic not related to their research.

In addition to meting the students under my supervision, I help organise, at the laboratory level:

- **TIMA PhD Conference (2021 and 2022)**: This event is intended to familiarise the young researchers with the full review process in an international event. During this event, a PhD candidate will have both the role of author as well as the role of reviewer. This activity is organised in 3 steps (i) the students write a paper describing the subject of their thesis, the motivation, a very short state-of-the-art, the scientific method they are using, and results. This paper is then submitted through a review managing platform by a certain deadline. (ii) each student will be assigned for review several papers submitted by its peers. (iii) all involved students will attend the Program Committee (PC) Meeting, where papers are discussed among the reviewers in order to decide whether the paper is fit for publication. This activity will help students better understand the miscommunication that can occur when the referees (their fellow PhD students) ask questions about a paper. In addition they can put themselves in the shoes of a reviewer. The ultimate goal of this activity is to help young researchers improve their written communication skills.

- **TIMA PhD Day (2021 and 2022)**: This event is organised in 4 sections: (i) scientific communications by every PhD candidate - "My research year in 180seconds"; (ii) Welcome the new-comers - first year PhD candidates introduce themselves to the lab; (iii) Team building games (iv) round table "How to improve the quality of life for young researchers"

To improve my supervision and mentoring skills, I have followed a training program "Thesis ethical supervision" which had the following objectives: (i) to acquire the skills to select and supervise the thesis student, (ii) to acquire the tools and methods that facilitate the supervision relationship, (iii) to conduct the thesis as a project and use adapted supervision techniques, (iv) to accompany the PhD student in the development of his/her professional project.

CHAPTER 4

Funded Projects and Scientific Cooperation

4.1 Funded Projects

Starting from my PhD., I have carried-out my research in the framework of national- and European-funded scientific projects. In some cases I was hired on the project (as PhD student, Post-doc or Researcher), in others I have contributed as project member or even task leader, while for the most recent ones I have actively contributed to the conception and the writing of the project proposal and I am work-package leader or project leader. An exhaustive list of the project is presented in Table 4.1, as well as my role within these projects (M=member, TL=Task Leader, WPL=Work Package Leader, Co-PI=co-Project Leader, TIMA-PI=Partner Project Leader, PI=Project Leader of the entire project), the organism in charge of financing the research and the the allocated budget of which I am in-charge. Some details concerning each project are presented in the following.

Table 4.1: Summary of Research Grants

Publicly Funded Reserch Grants					
Period	Acronym	Topic	Role	Organism	Grant (€)
2023/27	Neuropuls	SNN & PUF Photonics	TIMA-PI	HEurope	150.000
2022/24	SynConnect	Neuromorphic HW	PI	CNRS-MITI	135.000
2022/23	SConnect	Neuromorphic HW	TIMA-PI	UGA-MIAI	50.000
2022/26	POP	Security of PUFs	WPL	ANR-PRC	97.000
2021/22	CAD4RMD	In-memory Computing	PI	UGA-IRS	14.000
2020	PUF4IoT	Reliability of PUFs	Co-PI	CNRS-INS2I	5.000
2019/23	EMINENT	Neuromorphic HW	PI	ANR-JCJC	175.000
2019	RELeM	Reliability NVM	PI	CNRS-INS2I	10.000
2017/18	FilieraSicura	HW Security	M	CISCO	-
2016/18	NANOxCOMP	In-memory Computing	M	H2020-MC	-
2014/15	CLERECO	Reliability	M	FP7	-
2012/14	HiCool	Low-power Design	M	FR FUI	-
2012/14	LPMTR	Reliability	M	MICINN ES	-
2011/12	TRAMS	Reliability	TL	FP7	-
2007/11	LPCD	Reliability	M	MICINN ES	-
2006/07	ANGIOTUMOR	System Control	WPL	ANCS RO	-
Industrial Collaboration					
2021/24	STM Taly	ANN Accelerators	PI	CIFRE	50.000
2018/21	STM Cincon	Neuromorphic HW	M	CIFRE	30.000

NEUROPULS "Neuromorphic energy-efficient accelerators based on Phase change materials augmented siLicon photonics" – Horizon Europe RIA – Partners: CNRS, CEA-LETI, Ghent University, Politecnico di Torino, Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC-ID), Barcelona Supercomputing Center, UNIVR, Hewlett Packard Enterprise, Albora Technologies, Ethniko kai Kapodistriako Panepistimio Athinon, Ludwig-Maximilians-Universität München, Argotech, Università di Verona - total budget 8.000.000€

NEUROPULS will develop -for the first time- secure hardware accelerators based on novel neuromorphic architectures and PUF-based security layers leveraging the benefits offered by

the integration of photonics, PCMs and III-V materials. This integration will provide superior security, energy-efficiency and speeds for spiking and formal recurrent NNs when compared to current available technology for the selected use-cases.

The project goal will be acheived by fulfilling the following objectives: Objective 1: Development of a CMOS-compatible platform addressing the integration of Silicon Photonics with PCMs and III-V materials. Objective 2: Development of a low-power and secure RISC-V interfaced neuromorphic accelerator based on the integration of Silicon Photonics, novel PCMs, and Q-switched III-V lasers. Objective 3: Development of a system-level simulation platform for PCM-based photonic low-power accelerators using photonic security layers.

My contribution: I will contribute to this project on 2 main topics: (i) the development of novel photonic NN architectures suitable for SNNs and RNNs. More in detail, starting from the specifics of the architecture (such as connectivity and versatility of processing units) I will analyse the efficiency of various flavours of SNN-compatible leaning algorithms which are lightweight and hardware friendly. (ii) Development of security layers based on attack resilient photonic PUFs. More specifically I will work on the reliability aspects of CMOS and photonic based PUFs.

SynConnect: "Study and development of La2NiO4 memristive devices for bio-inspired computing" – CNRS 80PRIME Grant – Partners: TIMA and LMGP lab Grenoble

This project focuses on the use of La2NiO4 devices as artificial synapses for bio-inspired computing architectures, i.e., Spiking Neural Networks (SNNs) with a bio-inspired learning rule known as STDP (Spike Time Dependent Plasticity). The learning is applied on each synapse independently of the global state of the network; therefore, the synapse must be doted of computation capabilities. The researchers at LMGP have demonstrated that La2NiO4 can be used to build artificial synapses with analog, highly multilevel set and reset transitions with memristive characteristics which can be tuned by varying the material's oxygen content. The goal of the project is to demonstrate, at small scale, the feasibility of a La2NiO4-based SNN and understand the main advantages and shortcomings (from technology and application) such that concomitant optimization of device and algorithm can be performed to guarantee the achievement of a truly efficient bio-inspired electronic system.

The completion of this project will offer technological insights on how build an array of L2NO4 devices for neuromorphic hardware, opening the path towards the large-scale fabrication of such devices and possibly their integration with CMOS process. During the SynConnect project we will demonstrate, at small scale, the feasibility of L2NO4-based bio-inspired system and understand the main advantages and shortcomings (from technology and application) such that concomitant optimization of device and algorithm can be performed to guarantee the achievement of a truly efficient bio-inspired electronic system.

My contribution: Within the SynConnect project I will work on the circuit design and neuromorphic algorithm aspects. I will work on the development of the SNN based algorithm with STDP learning to identify the critical parameters affecting the learning efficiency and their interconnections. I will also contribute to the design of a small scale hardware-based SNN (using the mersitor models), then I will develop an environment to test the L2NO4 arrays in network environment and propose solutions to optimize the SNN and STDP algorithm.

POP: Power-Off laser attacks on security Primitives" – ANR PRC Grant – Partners: Mines Saint- Étienne, Laboratoire Hubert Curien, TIMA, LCIS - total budget 396.000 €

The POP project (Power-Off laser attacks on security Primitives) aims at: (i) experimentally assessing the feasibility of power-off laser attacks, (ii) modelling at electrical level the associated physical mechanisms for simulation and prediction purposes, (iii) designing countermeasures to address the new threats posed by these attacks.

To accomplish our research objectives, we plan to: (1) design a test chip embedding the basic blocks found in various security primitives for test and modelling, (2) perform laser experiments on circuits already available (and on the manufactured test chip later on) to build models and methods, (3) assess how power-off laser attacks may be used to alter PUFs and forge clones, or to deactivate attack sensors, (4) propose and validate countermeasures against these new attacks, and (5) underline threats and solutions considering our two use cases.

The POP project aims to open a new way to use power-OFF laser illumination on electronic integrated circuits, it could become an excellent choice of attack medium to target quickly and efficiently mobile electronic devices. As a consequence, the countermeasures which will be developed during the POP project could help the hardware designer to protect the future electronic devices against this new threat and the results.

My contribution: Within this project project I am actively involved in the design of the basic blocks found in various security primitives to be included in the test chip. In addition I am responsible for creating models of laser-induced faults and I will contribute to the development and design of countermeasures against the identify attacks. I will co-supervise the activity of trainee-students (internships) and of one PhD student.

CAD4RMD CAD Tool and Design Space Exploration for Resistive- based Memory Devices Integration in non-Von Neuman Architectures" – UGA IRS Grant – TIMA single partner

The purpose of this project is to provide a tool (modular and expendable over different resistive memory devices and computing paradigms) which will guide the designer towards the optimal combination technology/computation for desired application. This goal has been be achieved by developing a simulation platform to be integrated with standard CAD tools which supports introducing resistive memory devices into the standard integrated circuit design flow. More precisely, the following objectives have been be fulfilled: Objective 1. Create a data-base of primitive gates for computing-in memory for different types of resistive-memory devices and arrays. Objective 2. Develop an automated tool for design space exploration and circuit analysis – includes voltage and frequency tuning, device-to-device and cycle-to-cycle variability, noise and aging effects. Objective 3. Develop an automated tool to map critical computation kernels (specific to application) on appropriate compute primitive to achieve desired performance metrics.

The main benefits introduced during the project are the following: (i) Technology-aware design of CIM Hardware compute primitives: it gives the characteristics of the different HW CIM primitives for different technologies such as power consumption, reliability, robustness, etc., (ii) Technology/application benchmarking for CIM-based compute kernels: it gives which technology is the best to use for which application. This will enable appropriate trade-off and technology selection when targeting specific applications.

My contribution: as project leader I am responsible of all administrative and research tasks. I am actively involved in all the project tasks, I coordinate and supervising the activity of trainee-students (internships).

PUF4IoT "Physically-Unclonable Functions for Secure IoT" – CNRS INS2I Soutien pour le développement de relations internationales – Partners: TIMA single partner

The PUF4IoT project's goal is to enable consistent trustworthy authentication and authorization services across all IoT devices by developing secure PUF-based authentication and mutual- authentication protocols between any pair of components of an IoT (things, devices, communication objects and servers).

This project was targeted at consolidating the collaboration with with Dr. Honorio Martin (UC3M, Universidad Carlos III de Madrid, Spain). As a result, we have published several papers and submitted a H2020 project proposal on the same topic, in response to the call H2020-SU-ICT- 2018-2020. This project was be beneficial for TIMA since it allowed us to join the skills and know- how to improve the state-of-the-art on PUF-based authentication protocols.

My contribution: I have worked on developing novel integrated circuit identification techniques as well as on techniques to improve the robustness of existing PUF solutions.

EMINENT "Test and Reliability of Emerging Memory-based Spiking Neural Networks" – ANR JCJC Grant – TIMA single partner

The research hypothesis of the EMINENT project: the strong restrictions on the size of embedded Spiking Neural Network architectures (limited silicon area and interconnectivity ability) require minimization of the network redundancy which in turn reduces its the intrinsic fault tolerance. I postulate that there is an acute need to evaluate the reliability and perform manufacturing test of the neuromorphic hardware architectures to guarantee their correct operation and robustness.

The overall goal of the EMINENT project is to provide a dependable emerging memory-based Spiking Neural Network architecture. This goal will be achieved by fulfilling the following objectives: Objective 1: Provide an in-depth study of meaningful dependability threats in SNNs, Objective 2: Conduct a reliability estimation campaign, Objective 3: Provide a manufacturing test strategy and design-for-test solutions, Objective 4: Provide a strategy for architecture dependability improvement.

The results of the EMINENT project will have a direct impact in the field of device physics, hardware design and dependability, and system integration, and will indirectly affect all domains requiring manipulation of large amounts of data such as natural resources management, vehicle control, human-robot interactions, traffic optimization, radar systems, financial applications and so on. The successful completion of the project is expected to bring contributions to its corresponding research fields by: (i) providing a comprehensive study of behavioural faults and defect and fault models. This will offer a correct understanding of the limitations of SNNs with respect to their intrinsic fault tolerance and offer an insight in the ability to redesign and dimension the network to limit the use of resources or use unreliable resources. (ii) providing a test methodology for hybrid architectures build with devices with both deterministic and stochastic behavior. Such a test methodology could be potentially used in other contexts such as hardware implemented Bayesian networks, stochastic computing, or in-memory computing. (iii) providing the means to improve circuit dependability such that the minimum neural algorithm (for a given application) can be implemented in hardware without performance penalty. This will minimize the area footprint of neural computation modules, lessen the burden of connectivity, and as a consequence, will be a big step towards achieving hardware implemented artificial brain.

My contribution: as project leader I am responsible of all administrative and research tasks. I am actively involved in all the project tasks, I coordinate and supervising the activity of

trainee-students (internships) and of the PhD student, as well as the activity of the 2 senior researchers involved in this project.

RELeM "Reliable and Secure Emerging Memories for Dependable Computing Architectures" – CNRS INS2I Grant – TIMA single partner

The aim of the project is to comprehensively model, at different abstraction levels, the behaviour of emerging memories with and without the presence of possible faults, in order to enable a reliability- and security-aware design space exploration. Subsequently, I will develop Design-for-Reliability (DfR) solutions and hardware security solutions for systems using non-volatile emerging memories to meet the dependability requirements of the application. The main tasks to be accomplished within this project are: (i) to provide means for increasing the reliability of the memory devices by developing Design-for-Reliability solutions, and (ii) to provide means for increasing the security of the computing systems by identifying circuit vulnerabilities, especially to fault and to side-channel attacks.

My contribution: as project leader I am responsible of all administrative and research tasks.

FilieraSicura "Securing the Supply Chains of Domestic Critical Infrastructures from Cyber Attacks" – CISCO Italy

The project aims to provide new methodologies, techniques and tools to protect the country's critical national infrastructure and companies from cyber attacks. FilieraSicura aims to speed up the technology transfer of innovations developed by research towards the market, to address the evolution of cyber threats to companies and critical infrastructures with the same speed with which they evolve.

My contribution: Within this project I held a researcher position. My work was focused on investigating how scaled technologies and beyond-CMOS devices will modify the computing performance and behaviour and, most importantly, what is the impact on its security.

NANOxCOMP "Synthesis and Performance Optimization of a Switching Nano-Crossbar Computer" – European Union's H2020, Marie Skłodowska-Curie Project

The main goal of this project is developing a complete synthesis and optimization methodology for switching nano-crossbar arrays that leads to the design and construction of an emerging nanocomputer. New computing models for diode, FET, and four-terminal switch based nanoarrays are developed. The proposed methodology implements both arithmetic and memory elements, necessitated by achieving a computer, by considering performance parameters such as area, delay, power dissipation, and reliability. With combination of arithmetic and memory elements a synchronous state machine (SSM), representation of a computer, is realized. The proposed methodology targets variety of emerging technologies including nanowire/nanotube crossbar arrays, magnetic switch-based structures, and crossbar memories. The results of this project will be a foundation of nano-crossbar based circuit design techniques and greatly contribute to the construction of emerging computers beyond CMOS.

My contribution: This was a research mobility project and I was involved in cultivating the relationship with Purdue University. In this context I held 2 visiting-researcher grants visit Purdue Univerity for a total of 3 months to develop the topic of neuromorphic computing. In addition to this networking activity, I have worked on developing fault tolerance techniques for nanocrossbar computing arrays.

CLERECO "Cross-Layer Early Reliability Evaluation for the Computing Continuum" – European FP7 Project (STREP)

The CLERECO project aims at addressing the problem of having an early, fast, and accurate evaluation of computing systems reliability to support design decisions for hardware and software reliability enhancing mechanisms in the system. Such solutions can be only developed if the expected reliability of the system can be quickly and accurately assessed: (a) at different stages of the design flow (from design concept and early design stages through first silicon validation and eventually during operation in the field), (b) considering the impact of all hardware and software components, and the different modes of operation (use cases) of the system. Finally, the development of a reliable and dependable product that employs mechanisms to detect and handle possible faults can reduce the overall life cycle costs.

The proposed CLERECO framework for efficient reliability evaluation and therefore efficient exploitation of reliability oriented design approaches starting from early phases of the design process will enable circuit integration to continue at exponential rates and enable the design and manufacture of future systems for the computing continuum at a minimum cost (e.g., up to 50% less area, up to 50% less energy, etc.) contrary to existing worst-case-design solutions for reliability. The applications of such chips will play a major role in our society and can be seen through the prism of future computing systems ranging from avionics, automobile, smartphones, mobile systems, Personal Computers (PCs) and future servers utilized in the settings of Data Centers, Grid Computing, Cloud Computing and other types of HPC systems.

My contribution: My work was focused on on memory and storage reliability evaluation as well as preliminary work on techniques for implementing efficient and robust embedded image processing hardware accelerators.

HiCool "Upstream solutions for consumer-oriented design of complex integrated circuits" – FUI FR

The objective of the project is to improve the design process of systems-on-chip requiring a high degree of control over their power dissipation. Mobile devices, high performance computing and telecommunications infrastructure, will benefit from the solutions provided by this project. The expected benefits are improved productivity for the design teams and increased quality of the circuits developed with regard to their consumption. The project proposes automated techniques dealing with the aspects of power estimation, verification of circuit representations and insertion of low power structures. These techniques are unified by a formalised, bottom-up and top-down flow, centred on the architecture level and covering the RTL and logic gate levels in addition.

My contribution: Within this project I held a Post-Doc position. My work was focused on low-power memory and storage circuits and on reducing their testing costs.

LPMTR "Low power memory test and reliability" – Spanish FEDER MICINN

The project is focused on developing new methods and techniques for efficient reliability evaluation and memory test mechanisms.

My contribution: Within this project I held a Post-Doc position. My work was focused on developing methodology to evaluate the reliability of SRAM memories and develop test and design for test solutions which will maximise fault coverage.

TRAMS "Terascale Reliable Adaptive Memory Systems" – European Union's FP7

The project is concerned with the Terascale Computing era, where a single chip will be able to perform trillions of operations per second and provide trillions of bytes per second in off-chip bandwidth. However, these implies smaller devices which are much more susceptible

to faults and their performance exhibits a significant degree of variability. As a consequence, to unleash these impressive computing capabilities, a major hurdle in terms of reliability has to be overcome. The TRAMS project is the bridge for reliable, energy efficient and cost effective computing in the era of nanoscale challenges and teraflop opportunities. Both the Late CMOS and the Beyond CMOS technologies hold the promise of a significant increase in device integration density complemented by an increase in system performance and functionality. However, a dramatic reduction in single device quality is also expected, complemented by increase in statistical variability, severe reduction of the signal to noise ratio, and severe reliability problems. Therefore, alternative device solutions and computation paradigms need to be investigated to keep the technology evolution pace in such a challenging scenario. The TRAMS project addresses a specific variability and reliability-aware analysis and design flow as well as a hierarchical tolerance design. In such a tera-device multicore system the main idea will be to define countermeasure techniques at circuit and architecture design levels. The objective of this project is to investigate in depth potential new design alternatives and paradigms, which will be able to provide reliable memory systems out of highly unreliable nanodevices at a reasonable cost and design effort.

My contribution: Within this project I held a Post-Doc position. My work was focused on developing methodology to evaluate the reliability of SRAM memories in ultra-scaled technology and develop methods for fast yield estimation.

LPCD "Low power CMOS devices" – Spanish MICINN

This project was focused on the use of low-power CMOS devices across different applications and on the study of their reliability and development of suitable test techniques.

My contribution: Within this project I held a PhD Candidate position. My work was focused on analyzing the SRAM cell's robustness and functionality under process and environmental variations and providing new and easy to use metrics for robustness and functionality evaluation of the SRAM cell for parametric yield estimation. More in particular, i have worked on (i) analysing the principal characteristics of the nanometric CMOS SRAM memories taking into consideration process, voltage and temperature variations. (ii) evaluating the Static and dynamic robustness of the SRAM cell by means of static noise margin and dynamic noise margin, (iii) the analysis of the dynamic functionality of the SRAM cell, and on (iv) developing a method for parametric and statistical evaluation of SRAM cell failures under process variability.

ANGIOTUMOR "The prediction of the evolution and the assessment of the treatment's response for malignant tumors using morphological and hemodynamic modeling through imagistic, mathematical and artificial intelligence techniques" – Romanian ANCS

ANGIOTUMOR is a fundamental research project applied clinical character who pursue complex mathematical models and imaging of tumor neovascularization using information extracted from clinical practice by ultrasonography for high performance. Research for this purpose include the latest developments in medical imaging, information and communication technology, image processing and data analysis and processing using artificial intelligence systems.

The ultimate goal of the ANGIOTUMOR project is to generate abstract simulation systems spatial and temporal development of tumor angiogenesis process for: Accurate noninvasive diagnosis of disease; assessing the potential evolutionary and cancer prognosis and estimation of treatment response. To fulfill their intended purpose ANGIOTUMOR project has set the following objectives: (i) Assessment of malignant tumor neovascularization using high performance ultrasound techniques, immunohistochemistry, cell culture, mathematical modeling, imaging and evolving. (ii) Analysis of angiogenesis as a method for predicting the development

of malignant tumors. (iii) Development of quantitative models for describing and predicting therapeutic response dynamics of malignant tumors.

My contribution: I was involved in the development of mathematical models of malignant tumor neovascularization as well as in the creation of predictive models for its development.

The two projects under Industrial Collaborations Table 4.1 i.e., **STM Taly** "Design of a very low power Artificial Intelligence system (Tensor Processing Unit - TPU) based on in-memory computing" and **STM Cincon** "Design of a neuromorphic circuit in 28nm FDSOI" are ANRT CIFRE grants in collaboration with STMicroelectronics which finance PhD thesis. The scope of these projects is briefly described in Section 3.2 where the respective PhD candidates are presented.

My contribution: as project leader I am responsible of all administrative and research tasks. I coordinate and supervising the activity of the PhD candidates.

4.2 Research Collaborations

In this section I shortly describe my active and past collaborations within the research community. I list here only the collaborations which lead to common publications but were not funded by any research grant. The list of collaborators is presented in chronological odored referring to the start of the collaboration.

Joan Figueras (Professor) and **Rosa Rodriguez** (Professor) at Universitat Politecnica de Catalunya (UPC), Spain

Joan was my PhD supervisor and Rosa is a professor within the same department at UPC. The cooperation with Joan lasted until his retirement in 2017, while the cooperation with Rosa is still ongoing. Our collaboration is centered on developing techniques for the evaluation and the improvement of the robustness and reliability of emerging memories. The results of these collaborations are published in [C28], [C27], [C26], [C22] and [C21] listed in Chapter 6.

Michel Renovell (DR CNRS) at Laboratoire d'informatique, de robotique et de microélectronique de Montpellier (LIRMM), France

I have started my collaboration with Michel while I was still a PhD student at UPC Barcelona and this cooperation continued while I held my Post-Doc possitions at LIRMM and PoliTO. Our collaboration was on the evaluation of the SRAM robustness under the effect of process variability and severe environmental conditions. The results of this collaboration are published in [J14], [C36], [C28], [C26], [C21], [C13], [C12], [C09], [W7] and [W2] listed in Chapter 6.

Kaushik Roy (Professor) at Purdue University, US

My collaboration with Kaushik started in 2009 when I visited him at Purdue. Our cooperation was on the analysis and modeling of the SRAM dynamic behaviour and methods to evaluate the SRAM dynamic stability. The results of this collaboration are published in [C8] listed in Chapter 6.

Farshad Moradi (Professor) at Aarhus University, Denmark

I have collaborated with Farshad during my PhD. In 2009 we were both visiting scientists at Purdue University in the US and there we have started a collaboration of the design of low-power circuits. The results of this collaboration are published in [J03] listed in Chapter 6.

Lionel Torres (Professor) and **Giorgio Di Natale** (DR CNRS) at Laboratoire d'informatique, de robotique et de microélectronique de Montpellier (LIRMM), France

I have started my collaboration with Lionel and Giorgio while I was Post Doc at PoliTO. Our collaboration was on the design and analysis of spintronic-based Physical Unclonable Functions (PUFs). The collaboration with Giorgio continues today as we are i the same team at TIMA and are supervising several students together. The results of the common collaboration with Lionel and Giorgio are published in [J5], [C29], [W8] listed in Chapter 6.

Mario Barbareschi (Resercher) at University of Naples Federico II, Italy

I have started my collaboration with Mario while I was Post Doc at PoliTO. Our collaboration was on the design and analysis of spintronic-based Physical Unclonable Functions (PUFs) as well as on the design of computing primitives for memristive-based in-memory computing. The results of our collaboration are published in [J5], [C35] listed in Chapter 6.

Cristian Zambelli (Resercher) at University of Ferrara, Italy

I have collaborated with Cristian while I was a Post Doc at PoliTO on the topic of relibility of non-volatile memories. The results of these collaborations are published in [C24] listed in Chapter 6.

Jean-Michel Portal (Professor) and **Hassen Aziza** (Associate Professor) at Institut Matériaux Microélectronique Nanosciences de Provence (IM2NP), France

I have collaborated with Hassen while I was a Post Doc at PoliTO on the topic of relibility of non-volatile memories, while my collaboration with Jean-Michel started when I joined TIMA. This collaboration was centered on the evaluation of the efficiency and robustness of memristive devices used for the In-Memory-Computing paradigm. The results of these collaborations are published in [C40], [C24] listed in Chapter 6.

Alberto Bosio (Professor) at Institut des Nanotechnologies de Lyon (INL), France

Alberto was one of my supervisors when I was a Post-Doc resercher at LIRMM. We have continued to collaborate after my contract was over until today. Since I have integrated TIMA laboratory, we have collaborated on the development of emerging computing paradigms such as approximate computing, in-memory computing and neuromorphic computing (also in collaboration with **Ian O'Connor** Professor at INL). The results of these collaborations are published in [J8], [C56], [C48], [C35], [C34] listed in Chapter 6.

Said Hamdioui (Professor) and **Mottaqiallah Taouil** (Asistant Professor) at Delft University of Technology (TUDelft), The Netherlands

My collaboration with Said and Motta started during my Post Doc at PoliTO. This collaboration was focused on the robustness, the reliability and the test of emerging memory technologies. This collaboration has lead to the submission of an H2020 collaborative project which has not received financing so far. The results of this collaboration are published in [J07], [J06], [C35] listed in Chapter 6. In addition during my Post Doc at PoliTO I have held a 3-month Visiting Research grant at Delft.

Guillaume Prenat (Research scientist) at SPINTEC, France

I have collaborated with Guillaume since my arrival at TIMA. This collaboration was centered on the evaluation of the circuit design and test strategies for spintronic-based memories. This collaboration has lead to the submission of an ANR collaborative project which has not received financing so far. The results of this collaboration are published in [C40] listed in Chapter 6.

Benoit Miramond (Professor) at Laboratoire d'Electronique, Antennes et Télécommunications (LEAT) and **Elisa Vianello** (Research scientist) at CEA LETI, France

My collaboration with Benoit and Elisa started shortly after my arrival at TIMA. This collaboration was focused on the study of existing algorithms for neuromorphic computing and on finding solutions for hardware implementation of neuromorphic computing using memristive synapses. This collaboration has lead to the submission of an ANR collaborative project which has not received financing so far. The results of this collaboration are published in [C42] listed in Chapter 6. In addition, both Benoit and Elisa are part of the advisory board of the EMINENT project which I coordinate.

Honorio Martin (Associate Professor) at University Carlos III of Madrid, Spain

My collaboration with Honorio started in 2017. Our collaboration is centered on the design and reliability study of physical unclonable functions for circuit authentication. More in particular, we worked on the development of reliability mitigation techniques and circuit designs for PUF. The results of this collaboration are published in [C52], [C47] listed in Chapter 6.

Basel Halak (Associate Professor) at University of Southampton, UK.

My collaboration with Basel started in 2018 and it is centered on the design and reliability study of physical unclonable functions for circuit authentication. The results of this collaboration are published in [C47] listed in Chapter 6.

Aside of the aforementioned collaborations which yielded scientific publications, I have started cultivating new relations. In the following I will list the most promising of my new collaborations, collaborations which, I hope, will help move froward with my research plan (described in Chapter 5).

Monica Burriel (CR CNRS) and **Celine Ternon** (Assistant Professor) at Laboratoire des Matériaux et du Génie Physique (LMGP), France

The collaboration with Monica and Celine started in 2021 with the co-tutorship of an intern. Our collaboration is focused on the design and optimization of neuromorphic circuits based on memristive devices. My role in this collaboration is to identify the best application and circuit design for the memristive devices fabricated at LMGP, but also to give feedback on their efficiency in application. This collaboration allowed us to be awarded 2 research grants on the topic. More details are presented in section 4.1.

Mihai Miron (CR CNRS) at SPINTEC, France

I started collaborating with Mihai in 2021 on the topic of design and optimization of neuromorphic circuits based on spintronic structures. My role in this collaboration is to give feedback on the device efficiency when used in spiking neural networks and to study and identify the application which will boost this technology. This collaboration allowed us the write and submit a patent and a research paper.

Damien Deleruyelle (Professeur) at Institut National des Sciences Appliquées (INSA), France

I started collaborating with Damien in 2022 with the co-tutorship of an intern. Our collaboration is centered on the design and optimization of neuromorphic circuits based on Ferroelectric devices. My role in this collaboration is to perform electrical simulations and contribute to the circuit design and analysis.

CHAPTER 5

Research Perspectives

My main research skills are related to dependability, test, fault tolerance and reliability of digital systems. During my career I have applied these skills to CMOS memories (SRAMs and DRAMs) and then extended to emerging memories (Magnetic and Resistive RAMs) and later on I specialized in the use of such emerging memories for secure devices, as well as the design, test, and reliabi- lity of new computing paradigms (In-Memory and Neuromorphic Computing). My research perspectives are organized following 3 main directions : (i) Reliability and Test, (ii) Hardware Security and Trust, and (iii) Circuit Design for New Materials.

5.1 Perspectives on Reliability and Test

5.1.1 In-Memory Computing (IMC) Solutions

Many non-volatile memory device-based circuits have been proposed in the last decade to enable the implementation of some primitive computation functions in-memory. Most of this work addressed logic bitwise operations such as NAND, AND, OR, NOR and XOR. Recent work is focused more on some (limited) arithmetic operations such as vector-matrix multiplication, addition and multiplication. In my research I will focus on the development of fault models, reliability improvement, characterization and test methods targeting at enhancing the dependability of existing and future IMC primitives and IMC-based accelerators for various applications.

On the **short-term**: I will work on a complete analysis of possible reliability challenges, defects, defect mapping and fault modelling, related to memory devices used in an IMC context. This will culminate in an analysis of circuit failure modes and a comprehensive evaluation of the spatial and temporal dependencies of the bit-cell failures for IMC primitives. On the **medium-term**: I will focus on the development of test strategies and methodology development aiming at detecting all faulty behaviours of the memory array and the corresponding computation module to simultaneously assure robust data retention and correct result of the on-site computation. On the **long-term**: I want to develop test-oriented structures to be inserted early in the IMC design cycle, i.e., Design-for-Test (DfT), and will optimize such structures with the purpose of improving the testability, diagnostics and test time for a reduced number of test pins. In addition, I will work on the development and implementation of design-for-reliability solutions aiming at reducing in-field failure rate and ensuring robust operation under variability and stochastic effects as well as device aging and temporary failures.

On-going and envisioned collaborations: Delft University (NL) ; ST Microelectronics (FR) ; Oxford Brooks University (UK) ; University of Rome "Tor Vergata" (IT), IM2NP, Aix-Marseille Université (FR) ; Karlsruhe Institute of Technology (DE), Aachen University (DE).

5.1.2 Neuromorphic Computing Solutions

I will work on the reliability and test of bio-inspired hardware systems (spiking neural networks as well as formal neural networks). Regarding the formal neural networks, I will focus on their hardware implementation based on IMC accelerators and therefore the main activities that I intent to carryout in this domain are described above (subsection 5.1.1) with the important observation that the DfT and DfR techniques will be optimized for application. Regarding the spiking neural networks, I will focus on implementations which allow for on-chip, on-line learning, since they are fundamentally different from the formal neural networks and therefore present interesting challenges for hardware implementation and its dependability.

On the **short-term**: I will work on providing a full characterization of SNN architectures under manufacturing imperfections, aging phenomena and severe environmental conditions. I will develop pertinent fault models and methodologies for conducting a fault injection campaign and will identify scenarios of faulty operation happening before and after the on-line learning. On the **medium-term**: I will conduct a comprehensive manufacturing test campaign. I will target both functional and structural test strategies due to the special nature of these networks. SNN architectures resort to the combination of devices with both deterministic and stochastic behaviors and they include both digital and analog elements. A test strategy suitable for SNNs should be able to test the correct operation of both neurons and synapses, it should be economical (in area, power and performance overhead), it should not depend on the training data nor on the context the network would be used. On the **long-term**: I would like to focus on designing test access mechanisms which ideally will allow for (i) independent testing of the synaptic array and neuron layer for structural integrity, and (ii) jointly testing the SNN for functional correctness. Moreover, I will devise design-for-test (DfT) and built-in-self-test (BIST) techniques.

On-going and envisioned collaborations: EEA/INL Ecole Centrale de Lyon (FR) ; LIP6 Sorbonne University (FR) ; ST Microelectronics (FR) ; Smartnvy (FR); Delft University (NL); University of Rome "Tor Vergata" (IT), C2N Université Paris-Sud (FR); LIRMM Laboratory (FR) ; University of Ferrara (IT) ; University of California, Irvine (US)

5.2 Perspectives on Security

5.2.1 Physical Unclonable Functions (PUFs)

I plan to continue my research in the field of design and evaluation of PUFs with special focus on their use in the IoT to enable consistent trustworthy authentication of all IoT objects in the ecosystem. This goal can be reached by implementing a Physical Unclonable Function (PUF) in every single device in the IoT system and use these PUFs to authenticate all devices, and all com- munications. The PUF will act as the trust anchor as it returns a fingerprint self-generated due to manufacturing process, and is intrinsically robust compared with the non-volatile memory storing a serial number. In this context, I will focus my research efforts towards enabling consistent trustworthy authentication and authorization services across all IoT devices by guaranteeing their required properties in order to enable secure PUF-based authentication and mutual-authentication protocols between any pair of components of an IoT (things, devices, communication objects and servers).

On the **short-term**: Define a taxonomy of resource constraints and security requirements in IoT devices for the profiling of the targeted applications and communication protocols. The objective is to identify typical stringent constraints of IoT devices, e.g., cost, accessibility, computing resources, power consumption, energy availability. These constrains will be then translated for new PUF solutions (CMOS and beyond-CMOS) suitable for IoT components and the development of authentication protocols able to exploit the new PUF designs for security-resource optimization. On the **long-term**: I will work on the development of reliability and security boosting techniques for PUFs and PUF-based authentication protocols. A complete analysis of PUF reliability and its effects on the overall system reliability will be performed in order to create realistic PUF reliability models regarding temperature variations, voltage variations and aging. I will propose mitigation techniques in order to overcome reliability issues at different levels: technological, architectural, protocol at enrolment phase. This approach

should allow the PUF designers and users to select the most appropriate method to make the PUF steady and still compliant with the complexity and security requirements.

On-going and envisioned collaborations: Mines Saint- Étienne (MSE) (FR), Laboratoire Hubert Curien (FR), EEA/INL Ecole Centrale de Lyon (FR) ; Universidad Carlos III de Madrid (ES), University of Southampton (UK), KU Leuven (BE), LMU Munich (DE).

5.2.2 Security of In-Memory Computing (IMC) Solutions

I will continue working on the security aspects to be considered for IMC implementations based on memristive and spintronic devices. Indeed, the vulnerabilities of the tightly coupled memory-computing elements have to be analyzed in hybrid NV-CMOS technologies. The main objectives of this work are twofold : (i) establish a taxonomy of attacks – identification and classification of possible attacks that can lead to exploitable errors (i.e., that can jeopardize the security of the application) or leak of information in an in-memory crypto-processor architecture; (ii) provide justified comparison between in-memory mapping techniques.

On the **short**- and **medium-term**: I will focus on the definition of fault models of at bit-cell level and at circuit level, and on the information leakage analysis in order to address both fault and side-channel attacks. Currently I am working on the mapping of logic functions on In-Memory matrix and architectures. Once this activity I completed I will implement various versions of crypto-cores by synthesizing the building blocks of encryption algorithms on memristive-based In-Memory architectures. On these cores (i) I will study the correlation between in-memory computation power and switching profiles to understand if side-channel analysis will allow the leakage of information, and (ii) will analyze internal perturbations (timing, power supply, temperature) to unders- tand the circuit's intrinsic resistance in front of perturbation attacks. On the **long-term**: I will develop countermeasures against the side-channel and perturbation attacks.

Envisioned collaborations: Aix-Marseille Université (FR), KU Leuven (BE)

A long-term research ambition *(spanning over these 2 main research directions, i.e. (i) Reliability and Test, and (ii) Security) is to develop a methodology (possibly to be integrated in commercial CAD tools) for design space exploration of memory and in-memory-based computing accelerators (crypto-, neuromorphic-), which will allow for design optimizations in terms of area, power consumption, performance, endurance, testability and security. This methodology should allow architects to select the most appropriate technology and memory-based computation paradigm to fulfil their application requirements.*

5.3 Perspectives on Circuit Design for New Materials

In the future, I expect the nanoelectronics industry to move towards new materials, circuits and computing paradigms which will be able to increase the efficiency of computers and, in the same time, guarantee the sustainability of circuits fabrication.

On the **short**-, **medium**- and **long-term**: I will monitor closely the nanomaterials domain, to identify suitable candidates to alleviate the issues of today's electronics. These materials should be CMOS compatible, with an easy fabrication process, offer high endu- rance when used for computation, function at very low energy (or even have the ability to harvest energy), use bio-degradable materials, etc. On the **medium**- and **long-term** : I will work closely with

materials scientists to design and optimize devices and circuits for achieving ultra-low power computing systems with self-healing abilities (on-line test, diagnosis and repair).

On-going and envisioned collaborations: Spintec (FR), LMGP (FR), Comenius University (SK), imec (BE), Aachen University (DE).

CHAPTER 6

Publications

6.1 Journal Publications

J11 L. Anghel, A. Bernasconi, V. Ciriani, L. Frontini, G. Trucco, E.-I. Vatajelu, "Stuck-At Fault Mitigation of Emerging Technologies Based Switching Lattice," ***Journal of Electronic Testing*** 36, 313–326 (2020)

J10 M. C. Morgül, L. Frontini, O. Tunali, L. Anghel , V. Ciriani, E.-I. Vatajelu, C.A. Moritz, M. Stan, D. Alexandrescu, M. Altun, Circuit Design Steps for Nano-Crossbar Arrays: Area-Delay-Power Optimization with Fault Tolerance, ***IEEE Transactions on Nanotechnology***, Ed. IEEE, Vol. , DOI: 10.1109/TNANO.2020.3044017, décembre 2020

J9 E.-I. Vatajelu, G. Di Natale "High-Entropy STT-MTJ-based TRNG," ***IEEE Transactions on Very Large Scale Integration (VLSI) Systems***, Ed. IEEE, 2019

J8 L. Anghel, M. Benabdenbi, A. Bosio, M. Traiola, E.-I. Vatajelu (alphabetical order), "Test and Reliability in Approximate Computing", ***Journal of Electronic Testing: Theory and Applications***, Ed. Springer , Vol. 34, No. 4, pp. 375-387, 2018

J7 E.-I. Vatajelu, P. Prinetto, M. Taouil M., S. Hamdioui, "Challenges and Solutions in Emerging Memory Testing", ***IEEE Transactions on Emerging Topics in Computing***, Ed. IEEE, Vol. PP, No. 99, 2017

J6 E.-I. Vatajelu, P. Pouyan, S. Hamdioui, "State of the art and challenges for test and reliability of emerging nonvolatile resistive memories", ***International Journal of Circuit Theory and Applications***, Ed. Wiley, Chichester, UK, Vol. 46, No. 1, pp. 4-28, 2017

J5 E.-I. Vatajelu, G. Di Natale, M. Barbareschi, L. Torres, M. Indaco, P. Prinetto, "STT-MRAM-Based PUF Architecture exploiting Magnetic Tunnel Junction Fabrication-Induced Variability", ***ACM Journal on Emerging Technologies in Computing Systems (ACM JETC)*** - Special Issue on Secure and Trustworthy Computing – ACM J. Emerg. Technol. Comput. Syst. 13, 1, Article 5, 2016

J4 E.-I. Vatajelu, Á. Gómez-Pau, M. Renovell and J. Figueras, "SRAM Cell Stability Metric under Transient Voltage Noise," ***Microelectronics Journal***, Volume 45, Issue 10, pp 1348–1353, 2014

J3 F. Moradi, T. Vu Cao, E.-I. Vatajelu, A. Peiravi, H. Mahmoodi, D. T. Wisland, "Domino logic designs for high-performance and leakage-tolerant applications," vol 46, issue 3, pp. 247-254 ***Integration, The VLSI Journal***, 2013

J2 E.-I. Vatajelu, J. Figueras, "The Impact of Supply Voltage Reduction and Process Variability on the SRAM Static Noise Margins," ***Journal of Automation Computers and Applied Mathematics***, vol 17, no.4, pp. 579-587, ISSN 1221-437X, 2008

J1 E.-I. Vatajelu, J. Figueras, "The Impact of Supply Voltage Reduction on the Static Noise Margins of a 6T-Sram Cell," ***Journal of Control Engineering and Applied Informatics***, vol. 10, no.4, pp. 49-55, ISSN 1454-8658, 2008

6.2 Conference Proceedings

C57 S. Vinagrero Gutierrez, G. Di Natale, E.-I. Vatajelu "On-line reliability estimation of ring oscillator PUF" at IEEE European Test Symposium (***ETS***), 2022.

C56 Cristiana Bolchini, Alberto Bosio, Luca Cassano, Bastien Deveautour, Giorgio Di Natale, Antonio Miele, Ian O'Connor, E.-I. Vatajelu "Dependability of Alternative Computing Paradigms for Machine Learning: hype or hope?," at International Symposium on Design and Diagnostics of Electronic Circuits and Systems (***DDECS***), 2022.

C55 S. Daddinounou, E.-I. Vatajelu "Synaptic Control for Hardware Implementation of Spike Timing Dependent Plasticity," at International Symposium on Design and Diagnostics of Electronic Circuits and Systems (***DDECS***), 2022.

C54 P. Inglese, E.-I. Vatajelu, G. Di Natale "On the Limitations of Concatenating Boolean Operations in Memristive-Based Logic-In-Memory Solutions," at International Conference on Design and Technology of Integrated System in Nanoscale Era (***DTIS***), 28-30 June, 2021.

C53 P. Inglese, E.-I. Vatajelu, G. Di Natale"Memristive Logic-Iin-Memory Implementations: A Comparison," at Conference on PhD Research in Microelectronics and Electronics (***PRIME***), 19-22 July, 2021.

C52 H.Martin, E.-I. Vatajelu, G. Di Natale "Identification of Hardware Devices based on Sensors and Switching Activity: a Preliminary Study," at Design, Automation and Test in Europe (***DATE***), 1-5 February, 2021.

C51 F. Regazzoni, E.-I. Vatajelu et al. "Machine Learning and Hardware security: Challenges and Opportunities" to be published at IEEE International Conference On Computer Aided Design (***ICCAD***), 2020

C50 V. Cincon, E.-I. Vatajelu, L. Anghel, Ph. Galy, "From 1.8V to 0.19V voltage bias on analog spiking neuron in 28nm UTBB FD-SOI technology" to be published at IEEE ***EUROSoI-ULSI*** Conference 2020

C49 E.-I. Vatajelu, G. Di Natale, M. S. Mispan, B. Halak, "On the Encryption of the Challenge in Physically Unclonable Functions", IEEE International On Line Testing (***IOLTS***), pp. 115-120, 2019

C48 A. Bosio, E.-I. Vatajelu, et al., "Rebooting Computing: The Challenges for Test and Reliability", IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (***DFT***), pp. 8138-8143, Noordwijk, NETHERLANDS, 2019

C47 E.-I. Vatajelu, G. Di Natale, O. Keren, H. Martin, "On the Reliability of the Ring Oscillator Physically Unclonable Functions", IEEE 4th International Verification and Security Workshop (***IVSW***), pp. 25-30, Rhodes Island, GREECE, 2019

C46 E.-I. Vatajelu, G. Di Natale, L. Anghel, "Reliability of Hardware-Implemented Spiking Neural Networks (SNN)", IEEE VLSI Test Symposium (***VTS***), Monterey, UNITED STATES, 2019

C45 G. Di Natale G., E.-I. Vatajelu, K. Senthamarai Kannan, L. Anghel, "Hidden-Delay-Fault Sensor for Test, Reliability and Security", IEEE Design Automation and Test Conference in Europe (***DATE***), Florence, ITALY, 2019

C44 Eggersglüß S., E.-I. Vatajelu, et al., "IEEE European Test Symposium (ETS)", IEEE International Test Conference (***ITC***), Washington DC, UNITED STATES, 2019

C43 L. Anghel, A. Bersnasconi, V. Ciriani, L. Frontini, G. Trucco, E.-I. Vatajelu, Fault Mitigation of Switching Lattices under the Stuck-at-Fault Model, IEEE Latin American Test Symposium (***LATS***), 2019

C42 L. Anghel, G. Di Natale, B. Miramond, E.-I. Vatajelu, E. Vianello (alphabetical order), "Neuromorphic Computing - From Robust Hardware Architectures to Testing Strategies", 26th IFIP IEEE International Conference on Very Large Scale Integration (***VLSI SOC***), Verona, ITALY, 2018

C41 C.Morgül Muhammed, L. Frontini, E.-I. Vatajelu, L. Anghel, "Integrated Synthesis Methodology for Crossbar Arrays", ACM/IEEE International Symposium on Nanoscale Architectures (***NANOARCH***), Athens, GREECE, 2018

C40 E.-I. Vatajelu, L. Anghel, J.-M. Portal, M. Bocquet, G. Prenat, "Resistive and Spintronic RAMs: Device, Simulation, and Applications", IEEE International On Line Testing (***IOLTS***), Platja d'Aro, SPAIN, 2018

C39 E.-I. Vatajelu, L. Anghel, "Fully-Connected Single-Layer STT-MTJ-based Spiking Neural Network under Process Variability", ACM/IEEE International Symposium on Nanoscale Architectures (***NANOARCH***), Newport, RI, UNITED STATES, 2017

C38 E.-I. Vatajelu, G. Di Natale, P. Prinetto, "Zero bit-error-rate weak PUF based on Spin-Transfer-Torque MRAM memories", IEEE 2nd International Verification and Security Workshop (***IVSW***), pp. 128-133, Thessaloniki, GREECE, 2017

C37 E.-I. Vatajelu, L. Anghel, "Reliability Analysis of MTJ-based Functional Module for Neuromorphic Computing", IEEE International Symposium on On-Line Testing and Robust System Design (***IOLTS***), Thessaloniki, GREECE, 2017

C36 E.-I. Vatajelu, R. Rodriguez-Montanes, M. Renovell, J. Figueras, "Mitigating Read and Write Errors in STT-MRAM Memories under DVS", IEEE European Test Symposium (***ETS***), Limassol, CYPRUS, 2017

C35 M. Barbareschi, A. Bosio, S. Hamdioui, A. Nguyen Hoang, M. Traiola, E.-I. Vatajelu (alphabetical order) "Memristive devices: Technology, Design Automation and Computing Frontiers", International Conference on Design and Technology of Integrated Systems in Nanoscale Era (***DTIS***), Palma de Mallorca, SPAIN, 2017

C34 L. Anghel, M. Benabdenbi, A. Bosio, E.-I. Vatajelu (alphabetical order), "Test and reliability in approximate computing", Invited paper, Mixed Signals Testing Workshop (***IMSTW***), Thessaloniki, GREECE, 2017

C33 E.-I. Vatajelu, G. Di Natale, P. Prinetto, "STT-MTJ-based TRNG with On-the-Fly Temperature/Current Variation Compensation," IEEE International On-Line Test Symposium (***IOLTS***), Spain, 2016

C32 E.-I. Vatajelu, G. Di Natale, P. Prinetto, "Security Primitives (PUF and TRNG) with STT-MRAM," IEEE VLSI Test Symposium (***VTS***), Las Vegas, USA, 2016

C31 E.-I. Vatajelu, G. Di Natale, P. Prinetto, "Towards Highly Reliable SRAM-based PUFs," IEEE/ACM Design, Automation and Test in Europe Conference and Exhibition (***DATE***), Dresden, Germany, 2016

C30 A. Variale, E.-I. Vatajelu, G. Di Natale, P. Prinetto, P. Trotta, T. Margaria, "SeCube: An Open-Source Security Platform in a single SoC", Design and technology of Integrated Systems (***DTIS***), Istanbul, Turkey, 2016

C29 E.-I. Vatajelu, G. Di Natale, L. Torres, P. Prinetto, "STT-MRAM-Based Strong PUF Architecture," International Symposium on VLSI (***ISVLSI***), Montpellier, France, 2015

C28 E.-I. Vatajelu, R. Rodriguez-Montañés, S. Di Carlo, M. Indaco, M. Renovell, P. Prinetto, J. Figueras, "Power-Aware Voltage Tuning for STT-MRAM Reliability," European Test Symposium (***ETS***), Cluj-Napoca, Romania, 2015

C27 E.-I. Vatajelu, R. Rodriguez-Montañés, M. Indaco, P. Prinetto, J. Figueras, "STT-MRAM Cell Reliability Evaluation under Process, Voltage and Temperature (PVT) Variations," Design and technology of Integrated Systems (***DTIS***), Napoli, Italy, 2015

C26 E.-I. Vatajelu, R. Rodriguez-Montañés, M. Indaco, M. Renovell, P. Prinetto, J. Figueras, "Read/write robustness estimation metrics for spin transfer torque (STT) MRAM cell," Design, Automation and Test in Europe Conference and Exhibition (***DATE***), Grenoble, France, 2015

C25 E.-I. Vatajelu, G. Di Natale, M. Indaco, P. Prinetto, "STT MRAM-Based PUFs," Design, Automation and Test in Europe Conference and Exhibition (***DATE***), Grenoble, France, 2015

C24 E.-I. Vatajelu, H. Aziza, C. Zambelli, "Nonvolatile Memories: Present and Future Challenges," International Design and Test Symposium (***IDT***), Algiers, Algeria, 2014

C23 M. Indaco, S. Di Carlo, E.-I. Vatajelu, P. Prinetto, S. Arcaro and D. Pala, "Integration of STT-MRAM model into CACTI simulator," International Design and Test Symposium (***IDT***), Algiers, Algeria, 2014

C22 S. Di Carlo, M. Indaco, P. Prinetto, E.-I. Vatajelu, R. Rodriguez-Montañés, J. Figueras, "Reliability Estimation at Block-Level Granularity of Spin-Transfer-Torque MRAMs," Conference on Design and Fault Tolerance (***DFT***), Amsterdam, The Nederlands, 2014

C21 E.-I. Vatajelu, R. Rodriguez-Montañés, M. Indaco, M. Renovell, P. Prinetto, J. Figueras, "On the impact of supply voltage variation on the statistical reliability of a Spin-transfer-torque MRAM (STT-MRAM)," Conference on Design of Circuits and Integrated Systems (***DCIS***), Madrid, Spain, 2014

C20 E.-I. Vatajelu, M. Indaco, P. Prinetto, "On the Impact of Process Variability and Aging on the Reliability of Emerging Memories," Embedded Tutorial, European Test Symposium (***ETS***), Paderborn, Germany, 2014

C19 E.-I. Vatajelu, L. Dilillo, A. Bosio, P. Girard, A. Todri, A. Virazel, N. Badereddine, "Adaptive Source Bias for Improved Resistive-Open Defect Coverage during SRAM Testing," Asian Test Symposium (***ATS***), Yilan, Taiwan, 2013

C18 E.-I. Vatajelu, G. Tsiligiannis, L. Dilillo, A. Bosio, P. Girard, S. Pravossoudovitch, A. Todri, A. Virazel, F. Wrobel, F. Saigné, "On the Correlation between Static Noise Margin and Soft Error Rate evaluated for a 40nm SRAM Cell," Conference on Design and Fault Tolerance (***DFT***), New York, USA, 2013

C17 E.-I. Vatajelu, G. Tsiligiannis, L. Dilillo, A. Bosio, P. Girard, S. Pravossoudovitch, A. Todri, A. Virazel, F. Wrobel, F. Saigné, "SRAM Soft Error Rate Evaluation Under Atmospheric Neutron Radiation and PVT variations," International On-Line Test Symposium (***IOLTS***), Chania, Greece, 2013

C16 E.-I. Vatajelu, A. Bosio, L. Dilillo, P. Girard, A. Todri, A. Virazel, N. Badereddine, "Analyzing the Effect of Concurrent Variability in the Core Cells and Sense Amplifiers on SRAM Read Access Failures," Conference on Design and Technology of Integrated Systems (***DTIS***), Abu Dhabi, UAE, 2013

C15 E.-I. Vatajelu, A. Bosio, L. Dilillo, P. Girard, A. Todri, A. Virazel, N. Badereddine, "Analyzing SRAM Resistive-Open Defects Under the Effect of Process Variability," European Test Symposium (***ETS***), Avignon, France, 2013

C14 E.-I. Vatajelu, J. Figueras, "Efficiency Evaluation of Parametric Failure Mitigation Techniques for Reliable SRAM Operation," Design, Automation and Test in Europe Conference and Exhibition (***DATE***), Dresden, Germany, 2012

C13 E.-I. Vatajelu, Á. Gómez-Pau, M. Renovell and J. Figueras, "SRAM Stability Metric under Transient Noise," Conference on Design of Circuits and Integrated Systems (***DCIS***), Avignon, France, 2012

C12 E.-I. Vatajelu, Á. Gómez-Pau, M. Renovell and J. Figueras, "Transient Noise Failures in SRAM Cells: Dynamic Noise Margin Metric," Asian Test Symposium (***ATS***), New Delhi, India, 2011

C11 E.-I. Vatajelu, J. Figueras, "Robustness Analysis of 6T SRAMs in Memory Retention Mode under PVT Variations," Design, Automation and Test in Europe Conference and Exhibition (***DATE***), Grenoble, France, 2011

C10 E.-I. Vatajelu, J. Figueras, "Statistical Analysis of 6T SRAM Data Retention Voltage under Process Variation," IEEE Design and Diagnostics of Electronic Circuits and Systems (***DDECS***), Cottbus, Germany (Best Paper Award), 2011

C9 E.-I. Vatajelu, M. Renovell and J. Figueras, "Robustness of SRAM to Power Supply Noise under Dynamic Voltage Scaling," Conference on Design of Circuits and Integrated Systems (***DCIS***), Algarve, Portugal, 2011

C8 E.-I. Vatajelu, G. Panagopoulos, K. Roy, J. Figueras, "Parametric Failure Analysis of Embedded SRAMs using Fast and Accurate Dynamic Analysis," IEEE European Test Symposium (***ETS***), Prague, Czech Republic, 2010

C7 E.-I. Vatajelu, J. Figueras, "Statistical Analysis of SRAM Parametric Failure under Supply Voltage Scaling," IEEE International Conference on Automation, Quality and Testing, Robotics (***AQTR***), Cluj-Napoca, Romania, 2010

C6 E.-I. Vatajelu, J. Figueras, "Data Retention Failures in SRAMs Caused by Lowering the Power Supply Voltage," Conference on Design of Circuits and Integrated Systems (***DCIS***), 2009

C5 E.-I. Vatajelu, J. Figueras, "Supply Voltage Reduction in SRAMs: Impact on Static Noise Margins," IEEE International Conference on Automation, Quality and Testing, Robotics (***AQTR***), Cluj-Napoca, Romania (Best Paper Award), 2008

C4 E.-I. Vatajelu, J. Figueras, "Impact of Process Variability on the SRAM Static Noise Margin," Conference on Design of Circuits and Integrated Systems (***DCIS***), 2008

C3 E.-I. Vatajelu, M. Manisor, P. Raica, L. Miclea, T. Pop, O. Mosteanu, R. Badea, "Time Dependent Mathematical Model and Simulation of Tumor-Induced Angiogenesis based on Enzyme Kinetics," International Conference on Advancements of Medicine and Health Care through Technology (***ICAMHCT***), pp. 303-308, Cluj-Napoca, Romania, 2007

C2 M. Manisor, E.-I. Vatajelu, P. Raica, L. Miclea, O. Mosteanu, T. Pop, R. Badea, "Analysis of the Dynamic Behavior of Tumor Induced Angiogenesis based on Continuous Models," pp. 365-370, International Conference on Advancements of Medicine and Health Care through Technology (***ICAMHCT***), Cluj-Napoca, Romania, 2007

C1 E.-I. Vatajelu, T. Buzdugan, C. Festila, I. Nascu, "Theoretical Preliminaries Regarding the Modeling and Identification of an Electric Oven with Silicon Carbide Heating Elements," IEEE International Conference on Automation, Quality and Testing, Robotics (***AQTR***), Cluj-Napoca, Romania, 2006

6.3 Workshops without Formal Proceedings

W11 E.-I. Vatajelu, G. Di Natale, "High-Entropy STT-MTJ-based TRNG", 8th Workshop on Trustworthy Manufacturing and Utilization of Secure Devices (TRUDEVICE'2019), Baden Baden, GERMANY, 2019

W10 E.-I. Vatajelu, G. Di Natale, "High Entropy STT-MTJ-based TRNG," TRUDEVICE workshop (FCTRU), Barcelona, Spain, 2016

W9 E.-I. Vatajelu, G. Di Natale, P. Prinetto, "STT-MTJ-Based True Random Number Generator," TRUDEVICE workshop, Dresden, Germany, 2016

W8 E.-I. Vatajelu, G. Di Natale, L. Torres, P. Prinetto, "Exploiting the variability of the magnetic tunnel junction for security purposes," Workshop on Leading Edge Embedded Non Volatile Memories (eNMV), Gardanne, France, 2015

W7 E.-I. Vatajelu, R. Rodriguez-Montañés, M. Indaco, M. Renovell, P. Prinetto, J. Figueras, "Spin Transfer Torque MRAM Cell: Metrics for Robustness Estimation," Workshop on Leading Edge Embedded Non Volatile Memories (eNMV), Gardanne, France, 2015

W6 E.-I. Vatajelu, G. Di Natale, P. Prinetto, "Zero Bit-Error-Rate Weak PUF based on Spin-Transfer-Torque MRAM Memories," TRUDEVICE workshop, Saint-Malo, France, 2015

W5 E.-I. Vatajelu, M. Indaco, G. Di Natale, P. Prinetto, "MRAM based PUF," Joint MEDIAN-TRUDEVICE Open Forum, Amsterdam, The Netherlands, 2014

W4 E.-I. Vatajelu, Marco Indaco, Paolo Prinetto, "On the Impact of Control Voltage Variation on the Reliability of Spin Transfer Torque Magnetic RAM," Joint MEDIAN-TRUDEVICE Open Forum, Amsterdam, The Netherlands, 2014

W3 E.-I. Vatajelu, M. Indaco, R. Rodriguez-Montañés, S. Di Carlo, P. Prinetto, and J. Figueras, "Spin-Torque Transfer MRAM Statistical Reliability Prediction," STEM Workshop, Paderborn, Germany, 2014

W2 E.-I. Vatajelu, M. Renovell, J. Figueras, "Robustness of SRAM to Power Supply Noise under Dynamic Voltage Scaling, Low Power on Test and Reliability," LPonTR Workshop (with ETS), Prague, Czech Republic, 2010

W1 E.-I. Vatajelu, "Identifying an electric oven with two heating zones," International PhD Students' Workshop, (IWCIT), Gliwice Poland, 2006

References

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

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

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

[4] A. Wellig and J. Zory, "Static noise margin analysis of sub-threshold sram cells in deep sub-micron technology," in *Integrated Circuit and System Design. Power and Timing Modeling, Optimization and Simulation* (V. Paliouras, J. Vounckx, and D. Verkest, eds.), (Berlin, Heidelberg), pp. 488–497, Springer Berlin Heidelberg, 2005.

[5] S. Cserveny, J. M. Masgonty, and C. Piguet, "Noise margin in low power sram cells," in *Integrated Circuit and System Design. Power and Timing Modeling, Optimization and Simulation* (E. Macii, V. Paliouras, and O. Koufopavlou, eds.), (Berlin, Heidelberg), pp. 889–898, Springer Berlin Heidelberg, 2004.

[6] T. Hook, M. Breitwisch, J. Brown, P. Cottrell, D. Hoyniak, C. Lam, and R. Mann, "Noise margin and leakage in ultra-low leakage sram cell design," *IEEE Transactions on Electron Devices*, vol. 49, no. 8, pp. 1499–1501, 2002.

[7] B. Zhang, A. Arapostathis, S. Nassif, and M. Orshansky, "Analytical modeling of sram dynamic stability," in *2006 IEEE/ACM International Conference on Computer Aided Design*, pp. 315–322, 2006.

[8] L. Ding and P. Mazumder, "Dynamic noise margin: definitions and model," in *17th International Conference on VLSI Design. Proceedings.*, pp. 1001–1006, 2004.

[9] M. Sharifkhani and M. Sachdev, "Sram cell stability: A dynamic perspective," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 2, pp. 609–619, 2009.

[10] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*. USA: Society for Industrial and Applied Mathematics, 1992.

[11] R. Kanj, R. Joshi, and S. Nassif, "Mixture importance sampling and its application to the analysis of sram designs in the presence of rare failure events," in *2006 43rd ACM/IEEE Design Automation Conference*, pp. 69–72, 2006.

[12] T. Doorn, E. ter Maten, J. Croon, A. Di Bucchianico, and O. Wittich, "Importance sampling monte carlo simulations for accurate estimation of sram yield," in *ESSCIRC 2008 - 34th European Solid-State Circuits Conference*, pp. 230–233, 2008.

[13] *A Most Probable Point Based Method for Uncertainty Analysis*, vol. Volume 2: 26th Design Automation Conference of *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 09 2000.

[14] S. Srivastava and J. Roychowdhury, "Rapid estimation of the probability of sram failure due to mos threshold variations," in *2007 IEEE Custom Integrated Circuits Conference*, pp. 229–232, 2007.

[15] C. Gu and J. Roychowdhury, "An efficient, fully nonlinear, variability-aware non-monte-carlo yield estimation procedure with applications to sram cells and ring oscillators," in *2008 Asia and South Pacific Design Automation Conference*, pp. 754–761, 2008.

[16] R. Naseer, Y. Boulghassoul, J. Draper, S. DasGupta, and A. Witulski, "Critical charge characterization for soft error rate modeling in 90nm sram," in *2007 IEEE International Symposium on Circuits and Systems*, pp. 1879–1882, 2007.

[17] A. Balasubramanian, P. R. Fleming, B. L. Bhuva, O. A. Amusan, and L. W. Massengill, "Effects of random dopant fluctuations (rdf) on the single event vulnerability of 90 and 65 nm cmos technologies," *IEEE Transactions on Nuclear Science*, vol. 54, no. 6, pp. 2400–2406, 2007.

[18] A. Griffoni, P. Zuber, P. Dobrovolny, P. J. Roussel, D. Linten, M. L. Alles, R. D. Schrimpf, R. A. Reed, L. W. Massengill, D. Kobayashi, E. Simoen, and G. Groeseneken, "Impact of process variability on the radiation-induced soft error of nanometer-scale srams in hold and read conditions," in *2011 12th European Conference on Radiation and Its Effects on Components and Systems*, pp. 195–201, 2011.

[19] G. Tsiligiannis, L. Dilillo, A. Bosio, P. Girard, A. Todri-Sanial, A. Virazel, F. Wrobel, and F. Saigné, "A Novel Framework for Evaluating the SRAM Core-Cell Sensitivity to Neutrons," in *RADECS: European Conference on Radiation and Its Effects on Components and Systems*, (Biarritz, France), pp. 1–4, Sept. 2012.

[20] A. van de Goor and J. Simonse, "Defining sram resistive defects and their simulation stimuli," in *Proceedings Eighth Asian Test Symposium (ATS'99)*, pp. 33–40, 1999.

[21] L. Dilillo, P. Girard, S. Pravossoudovitch, A. Virazel, S. Borri, and M. Bastian Hage-Hassan, "Efficient March Test Procedure for Dynamic Read Destructive Fault Detection in SRAM Memories," *Journal of Electronic Testing: : Theory and Applications*, vol. 21, no. 5, pp. 551–561, 2005.

[22] Q. Chen, H. Mahmoodi, S. Bhunia, and K. Roy, "Modeling and testing of sram for new failure mechanisms due to process variations in nanoscale cmos," in *23rd IEEE VLSI Test Symposium (VTS'05)*, pp. 292–297, 2005.

[23] Q. Chen, H. Mahmoodi, S. Bhunia, and K. Roy, "Efficient testing of sram with optimized march sequences and a novel dft technique for emerging failures due to process variations," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 13, no. 11, pp. 1286–1295, 2005.

[24] S. Hamdioui, R. Wadsworth, J. Delos Reyes, and A. van de Goor, "Importance of dynamic faults for new sram technologies," in *The Eighth IEEE European Test Workshop, 2003. Proceedings.*, pp. 29–34, 2003.

[25] J. Li, P. Ndai, A. Goel, S. Salahuddin, and K. Roy, "Design paradigm for robust spin-torque transfer magnetic ram (stt mram) from circuit/architecture perspective," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 12, pp. 1710–1723, 2010.

[26] Y. Chen, X. Wang, H. Li, H. Xi, Y. Yan, and W. Zhu, "Design margin exploration of spin-transfer torque ram (stt-ram) in scaled technologies," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 12, pp. 1724–1734, 2010.

[27] X. Fong and K. Roy, "Robust low-power multi-terminal stt-mram," in *2013 13th Non-Volatile Memory Technology Symposium (NVMTS)*, pp. 1–4, 2013.

[28] X. Fong and K. Roy, "Low-power robust complementary polarizer stt-mram (cpstt) for on-chip caches," in *2013 5th IEEE International Memory Workshop*, pp. 88–91, 2013.

[29] M.-D. Yu and S. Devadas, "Secure and robust error correction for physical unclonable functions," *IEEE Design Test of Computers*, vol. 27, no. 1, pp. 48–65, 2010.

[30] A. Garg and T. T. Kim, "Design of sram puf with improved uniformity and reliability utilizing device aging effect," in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1941–1944, 2014.

[31] M. Bhargava and K. Mai, "A high reliability puf using hot carrier injection based response reinforcement," in *Cryptographic Hardware and Embedded Systems - CHES 2013* (G. Bertoni and J.-S. Coron, eds.), (Berlin, Heidelberg), pp. 90–106, Springer Berlin Heidelberg, 2013.

[32] S. Eiroa, J. Castro, M. C. Martínez-Rodríguez, E. Tena, P. Brox, and I. Baturone, "Reducing bit flipping problems in sram physical unclonable functions for chip identification," in *2012 19th IEEE International Conference on Electronics, Circuits, and Systems (ICECS 2012)*, pp. 392–395, 2012.

[33] K. Xiao, M. T. Rahman, D. Forte, Y. Huang, M. Su, and M. Tehranipoor, "Bit selection algorithm suitable for high-volume production of sram-puf," in *2014 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*, pp. 101–106, 2014.

[34] R. Maes, "An accurate probabilistic reliability model for silicon pufs," in *Cryptographic Hardware and Embedded Systems - CHES 2013* (G. Bertoni and J.-S. Coron, eds.), (Berlin, Heidelberg), pp. 73–89, Springer Berlin Heidelberg, 2013.

[35] A. Schaub, J.-L. Danger, S. Guilley, and O. Rioul, "An improved analysis of reliability and entropy for delay pufs," in *2018 21st Euromicro Conference on Digital System Design (DSD)*, pp. 553–560, 2018.

[36] P. Koeberl, Kocabaş, and A.-R. Sadeghi, "Memristor pufs: A new generation of memory-based physically unclonable functions," in *2013 Design, Automation Test in Europe Conference Exhibition (DATE)*, pp. 428–431, 2013.

[37] G. S. Rose, N. McDonald, L.-K. Yan, B. Wysocki, and K. Xu, "Foundations of memristor based puf architectures," in *2013 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, pp. 52–57, 2013.

[38] W. Che, J. Plusquellic, and S. Bhunia, "A non-volatile memory based physically unclonable function without helper data," in *2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 148–153, 2014.

[39] L. Zhang, X. Fong, C.-H. Chang, Z. H. Kong, and K. Roy, "Highly reliable memory-based physical unclonable function using spin-transfer torque mram," in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2169–2172, 2014.

[40] J. Das, K. Scott, S. Rajaram, D. Burgett, and S. Bhanja, "Mram puf: A novel geometry based magnetic puf with integrated cmos," *IEEE Transactions on Nanotechnology*, vol. 14, no. 3, pp. 436–443, 2015.

[41] A. Fukushima, T. Seki, K. Yakushiji, H. Kubota, H. Imamura, S. Yuasa, and K. Ando, "Spin dice: A scalable truly random number generator based on spintronics," *Applied Physics Express*, vol. 7, p. 083001, jul 2014.

[42] W. H. Choi, Y. Lv, J. Kim, A. Deshpande, G. Kang, J.-P. Wang, and C. H. Kim, "A magnetic tunnel junction based true random number generator with conditional perturb and real-time output probability tracking," in *2014 IEEE International Electron Devices Meeting*, pp. 12.5.1–12.5.4, 2014.

[43] S. Oosawa, T. Konishi, N. Onizawa, and T. Hanyu, "Design of an stt-mtj based true random number generator using digitally controlled probability-locked loop," in *2015 IEEE 13th International New Circuits and Systems Conference (NEWCAS)*, pp. 1–4, 2015.

[44] Y. Wang, H. Cai, L. A. B. Naviner, J.-O. Klein, J. Yang, and W. Zhao, "A novel circuit design of true random number generator using magnetic tunnel junction," in *2016 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, pp. 123–128, 2016.

[45] S. Kvatinsky, D. Belousov, S. Liman, G. Satat, N. Wald, E. G. Friedman, A. Kolodny, and U. C. Weiser, "Magic—memristor-aided logic," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 61, no. 11, pp. 895–899, 2014.

[46] S. Kvatinsky, G. Satat, N. Wald, E. G. Friedman, A. Kolodny, and U. C. Weiser, "Memristor-based material implication (imply) logic: Design principles and methodologies," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 10, pp. 2054–2066, 2014.

[47] K. M. Kim and R. S. Williams, "A family of stateful memristor gates for complete cascading logic," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 11, pp. 4348–4355, 2019.

[48] S. Angizi, Z. He, A. Awad, and D. Fan, "Mrima: An mram-based in-memory accelerator," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 39, no. 5, pp. 1123–1136, 2020.

[49] P.-E. Gaillardon, L. Amarú, A. Siemon, E. Linn, R. Waser, A. Chattopadhyay, and G. De Micheli, "The programmable logic-in-memory (plim) computer," in *2016 Design, Automation Test in Europe Conference Exhibition (DATE)*, pp. 427–432, 2016.

[50] H. Assaf, Y. Savaria, and M. Sawan, "Vector matrix multiplication using crossbar arrays: A comparative analysis," in *2018 25th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pp. 609–612, 2018.

[51] M.-H. Wu, M.-C. Hong, C.-C. Chang, P. Sahu, J.-H. Wei, H.-Y. Lee, S.-S. Shcu, and T.-H. Hou, "Extremely compact integrate-and-fire stt-mram neuron: A pathway toward all-spin artificial deep neural network," in *2019 Symposium on VLSI Technology*, pp. T34–T35, 2019.

[52] A. F. Vincent, J. Larroque, N. Locatelli, N. Ben Romdhane, O. Bichler, C. Gamrat, W. S. Zhao, J.-O. Klein, S. Galdin-Retailleau, and D. Querlioz, "Spin-transfer torque magnetic memory as a stochastic memristive synapse for neuromorphic systems," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 9, no. 2, pp. 166–174, 2015.

[53] D. Fan, Y. Shim, A. Raghunathan, and K. Roy, "Stt-snn: A spin-transfer-torque based soft-limiting non-linear neuron for low-power artificial neural networks," *IEEE Transactions on Nanotechnology*, vol. 14, no. 6, pp. 1013–1023, 2015.

[54] B. Yan, M. Liu, Y. Chen, K. Chakrabarty, and H. Li, "On designing efficient and reliable nonvolatile memory-based computing-in-memory accelerators," in *2019 IEEE International Electron Devices Meeting (IEDM)*, pp. 14.5.1–14.5.4, 2019.

[55] T.-L. Tsai, J.-F. Li, C.-L. Hsu, and C.-T. Sun, "Testing of in-memory-computing 8t srams," in *2019 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*, pp. 1–4, 2019.

[56] S. M. Nair, C. Münch, and M. B. Tahoori, "Defect characterization and test generation for spintronic-based compute-in-memory," in *2020 IEEE European Test Symposium (ETS)*, pp. 1–6, 2020.

[57] K. Cheng, J. Song, X. Zhang, Y. He, R. Wang, and Y. Wang, "A reliability-concerned compute-in-memory behavior model for convolutional neural network," in *2021 IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, pp. 1–4, 2021.

[58] J. Yu, H. A. Du Nguyen, M. Abu Lebdeh, M. Taouil, and S. Hamdioui, "Enhanced scouting logic: A robust memristive logic design scheme," in *2019 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, pp. 1–6, 2019.

[59] A. S. Cassidy, P. Merolla, J. V. Arthur, S. K. Esser, B. Jackson, R. Alvarez-Icaza, P. Datta, J. Sawada, T. M. Wong, V. Feldman, A. Amir, D. B.-D. Rubin, F. Akopyan, E. McQuinn, W. P. Risk, and D. S. Modha, "Cognitive computing building block: A versatile and efficient digital neuron model for neurosynaptic cores," in *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–10, 2013.

[60] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 668–673, 2014.

[61] T. T. W. D. O. D. G. G. M. G. B. D. Khodagholy, J. N Gelinas, "Neurogrid: recording action potentials from the surface of the brain," in *Nature Neuroscience*, vol. 18, p. 310–315, 2015.

[62] S. Moradi, N. Qiao, F. Stefanini, and G. Indiveri, "A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors (dynaps)," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 1, pp. 106–122, 2018.

[63] M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C.-K. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y.-H. Weng, A. Wild, Y. Yang, and H. Wang, "Loihi: A neuromorphic manycore processor with on-chip learning," *IEEE Micro*, vol. 38, no. 1, pp. 82–99, 2018.

[64] A. Neckar, S. Fok, B. V. Benjamin, T. C. Stewart, N. N. Oza, A. R. Voelker, C. Eliasmith, R. Manohar, and K. Boahen, "Braindrop: A mixed-signal neuromorphic architecture with a dynamical systems-based programming model," *Proceedings of the IEEE*, vol. 107, no. 1, pp. 144–164, 2019.

[65] X. Jin, M. Lujan, L. A. Plana, S. Davies, S. Temple, and S. B. Furber, "Modeling spiking neural networks on spinnaker," *Computing in Science Engineering*, vol. 12, no. 5, pp. 91–97, 2010.

[66] L. Khacef, N. Abderrahmane, and B. Miramond, "Confronting machine-learning with neuroscience for neuromorphic architectures design," in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2018.

[67] G. W. Burr, R. M. Shelby, A. Sebastian, S. Kim, S. Kim, S. Sidler, K. Virwani, M. Ishii, P. Narayanan, A. Fumarola, L. L. Sanches, I. Boybat, M. L. Gallo, K. Moon, J. Woo, H. Hwang, and Y. Leblebici, "Neuromorphic computing using non-volatile memory," *Advances in Physics: X*, vol. 2, no. 1, pp. 89–124, 2017.

[68] P. Diehl and M. Cook, "Unsupervised learning of digit recognition using spike-timing-dependent plasticity," *Frontiers in Computational Neuroscience*, vol. 9, 2015.

[69] T. Masquelier and S. J. Thorpe, "Unsupervised Learning of Visual Features through Spike Timing Dependent Plasticity.," *PLoS Computational Biology*, vol. 3, p. e31, Feb. 2007.

[70] G. Srinivasan, S. Roy, V. Raghunathan, and K. Roy, "Spike timing dependent plasticity based enhanced self-learning for efficient pattern recognition in spiking neural networks," in *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 1847–1854, 2017.

[71] E. B. B. M. L. Sung Hyun Jo 1, Ting Chang, "Nanoscale memristor device as synapse in neuromorphic systems," in *Nano Letters*, vol. 10, p. 1297–1301, 2010.

[72] W. K. R. Abhronil Sengupta, Priyadarshini Panda, "Magnetic tunnel junction mimics stochastic cortical spiking neurons," in *Nature Scientific Reports*, vol. 6, 2016.

[73] A. Sengupta, A. Banerjee, and K. Roy, "Hybrid spintronic-cmos spiking neural network with on-chip learning: Devices, circuits, and systems," *Phys. Rev. Applied*, vol. 6, p. 064003, Dec 2016.

Résumé

Mes principales compétences de recherche sont liées à la sûreté de fonctionnement, au test, à la tolérance aux fautes et à la fiabilité des systèmes numériques. Au début de ma carrière de chercheur, ces compétences ont été appliquées aux mémoires CMOS (SRAMs et DRAMs), puis étendues aux mémoires émergentes (RAMs magnétiques et résistives). Ces recherches ont été menées dans un contexte où les architectures et les mémoires Von-Neumann conventionnelles ne sont plus susceptibles de répondre à tous les besoins des applications modernes, en raison de limitations technologiques et conceptuelles inhérentes. Par conséquent, afin d'être à la pointe de l'industrie électronique en termes de capacités de conception et de fabrication, j'ai concentré mes efforts de recherche et d'innovation sur l'étude de nouvelles architectures non Von Neumann rendues possibles par les dispositifs émergents. De plus, la variabilité induite par la fabrication, les défauts, les effets stochastiques et la dégradation due au vieillissement peuvent entraîner d'importantes variations des caractéristiques électriques des dispositifs fabriqués, ce qui peut conduire à leur défaillance. Il est donc naturel que mon activité de recherche se concentre sur l'identification des principaux problèmes de fiabilité rencontrés par les nouveaux circuits intégrés, sur le développement de techniques de test appropriées pour leur détection et sur la conception de solutions pour les atténuer. Un autre aspect important de la fiabilité est lié à la sécurité du matériel. Les systèmes de sécurité utilisent des protocoles cryptographiques, souvent construits sur des algorithmes et des primitives cryptographiques de bas niveau, tels que les fonctions physiquement inclinables (PUF) et les générateurs de nombres aléatoires (TRNG). Les solutions PUF de pointe sont basées sur la mémoire et ont suscité mon intérêt, ce qui m'a conduit à mener des recherches dans le domaine de la conception, de l'évaluation et de l'optimisation des PUF et des TRNG. Dans cette présentation HDR, je donnerai un aperçu de mes activités de recherche jusqu'à présent, je présenterai brièvement les principaux résultats que j'ai obtenus dans chacun de ces quatre sujets principaux, et je présenterai à l'auditoire mon projet de recherche.

Mots-clés : Mémoires CMOS, Mémoires émergentes, Calcul non-Von Neumann, Fiabilité, Test, Sécurité du matériel

Abstract

My main research skills are related to dependability, test, fault tolerance and reliability of digital systems. At the beginning of my research carrier, these skills were applied to CMOS memories (SRAMs and DRAMs) and then extended to emerging memories (Magnetic and Resistive RAMs). This research was conducted in a context where conventional Von-Neumann architectures and memories are no longer likely to fulfil all the needs of modern applications, due to inherent technological and conceptual limitations. Hence, in order to be at the forefront of the electronic industry in terms of design and manufacturing capabilities, I focussed my research and innovation efforts on study of novel non-Von Neumann architectures enabled by emerging technology devices. Moreover, manufacturing induced variability, defects, stochastic effects, and aging degradation can cause important variations of the electrical characteristics of fabricated devices which can lead to device failure. So naturally, my research activity is focused on identifying the main dependability issues faced by memory-centered ICs and developing suitable test techniques for their detection and design solutions for their mitigation. Aside from reliability, an important aspect of device dependability is related to hardware security. Security systems use cryptographic protocols, frequently built on low-level cryptographic algorithms and primitives, such as Physically Unclonable Functions (PUFs) and True Random Number Generators (TRNGs). The Physically Unclonable Functions (PUFs) are emerging primitives exploited to implement low-cost authentication protocols and cryptographic primitives, such as secure key generators, key storing and one-way functions. State-of-the-art PUF solutions are memory-based and have piqued my interest, leading me to conduct research in the area of design, evaluation and optimization of PUFs and TRNGs. In this HDR presentation, I will give an overview of my research activities up-to-date, I will briefly present the main results I have obtained in each of these four main topics, and I will introduce the audience to my research project.

Keywords: CMOS Memories, Emerging Memories, non-Von Neumann Computing, Reliability, IC Test, Security Primitives