

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2017-ENST-0005

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

Télécom ParisTech

Spécialité - Électronique et Communications

présentée et soutenue publiquement par

You WANG

le 13 Février 2017

Analyse de Fiabilité de Circuits Logiques et de Mémoire basés sur Dispositif Spintronique

Directeur de thèse: **Lirida Alves de Barros NAVINER**

Co-directeur de thèse: **Weisheng ZHAO**

Composition du Jury

Damien Deleruyelle, Professeur, Institut National des Sciences Appliquées de Lyon	Président
Ian O'CONNOR, Professeur, École Centrale de Lyon	Rapporteurs
Lionel TORRES, Professeur, Université Montpellier 2	
Christian Gamrat, Directeur de Recherche au CEA	Examinateur
Lirida Alves de Barros NAVINER, Professeur, Télécom ParisTech	Directeurs de Thèse
Weisheng ZHAO, Professeur, Université de Beihang	

THÈSE

TELECOM ParisTech

école de l'Institut Mines-Télécom - membre de ParisTech

46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

Reliability Analysis of Spintronic Device Based Logic and Memory Circuits

You WANG

©

Le Département Communications et Electronique (COMELEC)
LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay
46 Rue Barrault, Paris CEDEX 13, 75634, France

This thesis is set in Computer Modern 11pt,
with the LaTeX Documentation System

©You Wang 2017

February 2017

Remerciements

Je souhaite commencer ce manuscrit par adresser mes remerciements sincères aux personnes qui m'ont beaucoup aidé depuis trois ans et qui ont contribué à l'achèvement de ce mémoire. Cette thèse a été menée dans le cadre des travaux du groupe NANOARCHI (Circuits et Architectures pour Nanodispositifs émergents) du département Nanoélectronique au laboratoire Institut Fondamentale Electronique (IEF) et du groupe SEN (Systèmes Electroniques Numériques) du département COMELEC (communication et électronique) à Télécom ParisTech.

Je voudrais tout d'abord adresser mes remerciements à mes directeurs de thèse, Madame Lirida Naviner, Professeur à Télécom ParisTech et Monsieur Weisheng Zhao, Professeur à l'Universié de Beihang, qui m'ont accueilli dans leur équipe et qui m'ont soutenu tout au long des trois ans de travail. Leur confiance en mes capacités et leur grande patience lors de nos échanges régulières m'ont beaucoup encouragé et motivé. Je tiens à les remercier également de m'avoir aidé à exploiter des démarches de recherche scientifique et à résoudre des problèmes scientifiques ou administratifs.

Je voudrais ensuite remercier les membres de mon jury de leur temps consacré à ma thèse. Je remercie particulièrement, mes rapporteurs Professeurs Ian O'CONNOR et Lionel TORRES pour le regard critique et leurs remarques constructives sur mon travail. Je remercie également Professeur Damien Deleruyelle et M. Christian Gamrat, qui ont gentiment accepté d'examiner mes travaux de thèse.

J'aimerais aussi adresser mes remerciements à M. Hao Cai, pour les discussions efficaces et les suggestions originales de sa part durant la réalisation des idées innovantes tout au long de cette thèse, entre autre, sur les avantages de la techonologie FDSOI 28nm, M. Yue Zhang, pour ses aides me permettant approfondir la compréhension du sujet de thèse, de la modélisation du MTJ et des premiers pas de la rédaction scientifique, et Mlle. Erya Deng, pour le démarrage de cette thèse et la simulation des circuits de base à l'aide du logiciel Cadence.

Aussi, un grand merci aux services administratifs de l'IEF et de Télécom ParisTech, dont Mme. Chantal Cadiat, Mme. Yvonne Bansimba, Mme. Elisabeth Valensi, Mme. Sylviane

Thomas, Mme. Florence Besnard, Mme. Marianna Baziz... Grace à leur travail scrupuleux et efficace, j'ai pu profiter de ma vie à l'école sans me soucier des problèmes administratifs et concentrer sur mon sujet de recherche. Je tiens à remercier tous mes collègues de Télécom ParisTech et de l'IEF, de leur gentillesse et de leur accompagnement. Leurs conseils voire les causeries menés pendant les déjeuners ou les pauses ont beaucoup enrichi mes connaissances profesionnelles ainsi que ma vie quotidienne. C'est avec grand plaisir de travailler entre eux.

Naturellement, je voudrais aussi remercier tous mes amis français et chinois: la famille de Frédérique Bedouin, qui m'a montré la vie familiale française et les coins intéressants de Paris, Zhaohao Wang, Yu Zhang, Gefei Wang, Yao Wu, Mengying Ren, Pengwenlong Gu, Yimeng Zhao, ...pour leur accompagnement durant les trois ans.

Enfin, je souhaite exprimer profonde gratitude à ma famille, dont notamment, mes parents, M. Jianlin Wang et Mme. Yurui Yang, qui m'ont accompagné et soutenu san condition pour l'aboutissement de ce travail.

Abstract

Moore's law has successfully guided us in the research and development of integrated circuits (IC) for several decades. However, power dissipation issue has recently become the bottleneck for further scaling down of complementary metal oxide semiconductor (CMOS) technology node. This issue can not be overcome by the conventional semiconductor devices. Nanotechnologies and nano-devices are considered approaches to build up ultra low power IC in the next era of "More than Moore". For instance, spintronics devices such as magnetic tunnel junction (MTJ) feature non-volatility and 3D integration, which can turn off the standby power and reduce drastically the power dissipated in data traffic between memory and logic chips. Moreover, the spintronics devices are promising to operate normally in sub 0.1V, which is another bottleneck of semiconductor devices. Compared with its counterparts, MTJ nanopillar with interfacial perpendicular magnetic anisotropy (PMA-MTJ) becomes an outstanding candidate for spin transfer torque magnetic random access memory (STT-MRAM) because of its lower switching current, faster operation speed, high scalability and better thermal stability.

However, spintronics devices suffer from significant reliability issues, e.g., process variation, stochastic switching behavior, temperature fluctuation and dielectric breakdown. As a result, the performance of hybrid MTJ/CMOS circuits can be significantly degraded by these issues. The reliability becomes one of the most critical factors to limit the use in practical applications. Thus, it is essential to study, analyze and reduce/tolerate the impact of these issues on the yield of MTJ based circuits at the early design phase for economic reasons. For this purpose, a compact model including all of the reliability issues is required by the circuits designers. We proposed a model of PMA-MTJ switched by STT mechanism which comprises the main reliability issues. In order to achieve good agreement with the experimental measurements, several physical models which describe the issues and realistic parameters are integrated in the compact modeling. The model is programmed in VerilogA language for SPICE compatible simulation.

Based on the accurate model of PMA-STT-MTJ, the robustness of typical hybrid MTJ/CMOS circuits is entirely investigated, e.g., MRAM writing/reading circuit, magnetic flip-flop (MFF), magnetic full-adder (MFA). With detailed analysis of the simulation results, we proposed some design methodologies to improve the circuits robustness, such

as using high performance devices (e.g., fully depleted silicon on insulator (FDDOI)) and applying dynamic asymmetrical body bias in symmetrical circuits of FDSOI transistors.

Instead of weakening the impact of reliability issues, some of them can be beneficial to several special applications. For instance, the stochastic switching behavior can be used as a physical randomness source in the security area. We proposed a novel circuit design of true random number generator (TRNG) and compared the performance with conventional realization. Furthermore, the uncertainty in MTJ switching process also provides a new approach of low power inexact circuit design, e.g., approximate computing and stochastic computing.

Keywords: Magnetic tunnel junction, Reliability analysis, Compact model, Dynamic asymmetrical body bias, True random number generator, Approximate computing, Stochastic computing

Contents

Remerciements i

Abstract iii

List of Tables ix

List of Figures x

List of Acronyms xxi

1 Introduction 1

1.1 Motivations 1

1.2 Thesis contributions 4

1.3 Organization of the thesis 5

2 State of the art 7

2.1 Magnetic tunnel junction 7

2.1.1 MTJ working principles 7

2.1.2 MTJ switching approaches 10

2.1.2.1 Field-induced magnetic switching (FIMS) 11

2.1.2.2 Thermally assisted switching (TAS) 12

2.1.2.3 Spin transfer torque (STT) 13

2.1.2.4 Thermally assisted spin transfer torque (TAS+STT) 15

2.1.2.5 Spin Hall effect spin transfer torque (SHE+STT) 15

2.2 Magnetic tunnel junction based memories and logic circuits 17

2.2.1 Magnetic Random Access Memory 17

2.2.2 Logic in Memory 19

2.2.3 Other novel applications . . . 22
2.3 Reliability analysis of MTJ device and MTJ based applications . . . 23
2.4 Summary . . . 27

3 Compact modeling of reliability issues in STT-PMA-MTJ 29
3.1 Perpendicular magnetic anisotropy (PMA) MTJ . . . 29
3.2 Reliability issues of STT-PMA-MTJ . . . 31
3.2.1 Process variation . . . 31
3.2.2 Stochastic switching behavior of MTJ . . . 34
3.2.3 Temperature fluctuation behavior of MTJ . . . 34
3.2.4 Dielectric breakdown . . . 35
3.3 Physical models of PMA-MTJ . . . 36
3.3.1 Tunnel barrier resistance model . . . 37
3.3.2 Bias-voltage-dependent TMR model . . . 38
3.3.3 Model of static behavior . . . 39
3.3.4 STT switching dynamic model . . . 39
3.4 Physical models of reliability issues in MTJ . . . 40
3.4.1 Process variation . . . 40
3.4.2 Stochastic switching . . . 41
3.4.3 Temperature fluctuation behavior of MTJ . . . 41
3.4.3.1 Models of temperature sensitive parameters . . . 42
3.4.3.2 Temperature fluctuation due to Joule heating . . . 44
3.4.4 Dielectric breakdown . . . 46
3.4.4.1 Breakdown voltage . . . 46
3.4.4.2 Prediction of lifetime . . . 48
3.4.4.3 Breakdown probability . . . 49
3.4.4.4 TDDB phenomena submitted to voltage pulse stress . . . 49
3.5 Compact modeling in EDA tool Cadence . . . 51
3.5.1 Modeling language: Verilog-A . . . 51
3.5.2 Electrical Modeling of MTJ under Cadence . . . 53
3.5.2.1 Hierarchy of the physical models integrated in the compact model . . . 53
3.5.2.2 Parameters of the compact model and Component Description Format (CDF) . . . 53

3.5.2.3 Schematic view of model and relative circuit in the design environment . . . 55
3.5.3 Functionality validation of model . . . 59
3.6 Fast simulation model using worst-case corners . . . 64
3.6.1 Introduction of worst-case fixed corners model . . . 64
3.6.2 Worst-case fixed corners model of MTJ . . . 65
3.7 Conclusion . . . 67

4 Reliability analysis and variability-aware design of hybrid MTJ/CMOS circuits 69
4.1 Reliability analysis of MTJ based circuits . . . 69
4.1.1 Variability analysis of MTJ based circuits . . . 70
4.1.2 Influence of MTJ stochastic switching behavior on MTJ/CMOS circuits . . . 70
4.1.3 Temperature impact on MTJ based circuits . . . 72
4.1.4 Ageing of MTJ based circuits . . . 73
4.1.5 Application of non Monte-Carlo Methodology in hybrid MOS/MTJ Circuits . . . 77
4.1.5.1 Switching delay and time to failure estimation of 1T-1M memory array . . . 77
4.1.5.2 Variability-aware energy-delay analysis of PCSA based STT-MRAM cell . . . 78
4.1.5.3 Worst-case analysis of magnetic full-adder dynamic performance . . . 79
4.1.5.4 Results discussion . . . 80
4.2 Reliability-aware design of MTJ-based circuits . . . 81
4.2.1 Transistors with UTBB-FDSOI technology . . . 82
4.2.2 Circuit Design of non-volatile Flip-Flop using dynamic asymmetrical body bias of FDSOI . . . 83
4.2.3 Reliability analysis and performance evaluation . . . 84
4.3 Conclusion . . . 89

5 Novel applications of MTJ in conventional circuits 91
5.1 A novel circuit design of MTJ based true random number generator . . . 91
5.1.1 Traditional true random number generators . . . 91

5.1.2 Circuit design of true random number generator using MTJ 92
5.1.3 Simulation results 95
5.1.4 Performance evaluation and optimization 96
5.2 Realization of Stochastic computing using MTJ 98
5.2.1 Introduction of stochastic computing 98
5.2.2 Stochastic computation with combinational logic 99
5.2.3 Stochastic computing using STT-MTJ 99
5.2.4 Case Study: Polynomial function RTL synthesis 102
5.3 Approximate computing method using MTJ 103
5.3.1 Introduction of approximate computing 103
5.3.2 Design for Approximation 104
5.3.2.1 Reduced Logic Complexity 104
5.3.2.2 The Dual-mode MFA 105
5.3.2.3 Functional Simulation 106
5.3.3 Design Considerations 107
5.3.3.1 Supply Scaling Strategy 107
5.3.3.2 Performance Analysis 109
5.3.3.3 Reliability-aware Simulation 110
5.4 Conclusion 113

6 Conclusions and Perspectives **115**
6.1 Conclusions 115
6.2 Perspectives 117

Bibliography **119**

A Source code of STT-PMA-MTJ compact model **141**

B List of publications **153**

Résumé Français **156**

List of Tables

2.1 Comparison of performance for the different switching approaches [28] . . . 17

2.2 Performance comparison of the universal memory candidates widely used and appeared in the last decade: F represents feature size of the lithography, the energy estimation is on the cell-level (not on the array-level), the endurance is signified by the writing cycles [39, 42]. 19

2.3 Comparison of different compact models of MTJ 27

3.1 Parameters integrated in the compact model including constants, technology parameters, device parameters and reliability control flags. 58

3.2 Parameters settings of CMOS transistors 65

4.1 Design parameters settings 70

4.2 Worst-case corners setting of transistor and MTJ models for worst-case performance analysis of STT-MRAM cell 79

4.3 Model precision on function of n and transistors size 82

4.4 Comparison of performance between proposed methodology (DABB) and conventional method (NBB) 89

5.1 Comparison of performance in TRNG 98

5.2 Performance comparison of conventional MFA and proposed approximate adder. 112

B.1 Comparaison de performance de MFA conventionnel et MFA approximatif propsés. 185

List of Figures

1.1 Breakthroughs in spintronics research and development for memory. 3

2.1 Equivalent resistance model to describe GMR effect in the structure of non-magnetic (NM) layer sandwiched by two ferromagnetic (FM) layers: Anti-parallel (AP) state presents higher resistance value than parallel (P) state. 8

2.2 Spin-dependent tunneling of electrons in an MTJ while the magnetization directions in two FM layers are (a) parallel and (b) anti-parallel. 9

2.3 MTJ consists of three layers: two ferromagnetic layers separated by an oxide barrier. The nanopillar resistance (R_p, R_{ap}) depends on the corresponding state of the magnetization of the two ferromagnetic layers Parallel (P) or Anti-Parallel (AP). The MTJ state can be switched by modulating the magnetic field. 10

2.4 Field induced magnetic switching approach structure. 11

2.5 Thermally assisted switching approach structure. 12

2.6 Spin transfer torque switching approach structure. 13

2.7 Diagram of the LLG equation: $\Gamma_{damping}$ is the Gilbert damping torque, Γ_{STT} is the STT term and Γ_{field} is the effective field torque generated by effective magnetic field H_{eff}. 14

2.8 Thermally assisted spin transfer torque switching approach structure. . . . 15

2.9 Spin Hall effect spin transfer torque switching approach. 16

2.10 A schematic of the cross-point array. The selector is added in series with the MRAM cell at each cross-point. 18

2.11 1T/1MTJ memory cell architecture. 18

2.12 General architecture of logic in memory based on STT-MRAM: Mx represents the highest level of metal in CMOS technology. 20

2.13 Typical MOS/MTJ NV-LIM circuits based on pre-charge sense amplifier structure: logic gates, full adder and flip-flop. 21

2.14 Implementation of ASL Boolean gates. Only the net spin polarization is shown for spin current. (a) Inverter. (b) NAND. "F" denotes a magnet with fixed magnetization direction. 23

2.15 Low state resistance and high state resistance distributions of the 4kbit circuit with MTJ size of 120 x 170 nm as demonstrated in [41]. Bias voltage is kept at -0.1 V. 24

2.16 Experimental measurement of STT stochastic switching behaviors, the switching duration follows a certain distribution determined by the current and pulse duration. 25

2.17 TMR ratio at different temperature in experimental measurements in [7, 8, 11, 12]. 26

3.1 (a) Structure of PMA STT MTJ based on CoFeB/MgO stack. (b) Core of MTJ and switching mechanism. 30

3.2 Typical flow of magnetic tunnel junction (MTJ) device fabrication, which mainly consists of stack deposition, patterning, etching dielectric encapsulation, and connecting. 32

3.3 Magnetic curves (measured by NanoMOKE) of MTJ stacks annealed at different annealing times. The film stack deposited by magnetic sputtering processing are *exsitu* annealed at 300 °C for different annealing times (40, 60 and 90 min) with perpendicular H = 0.775 T in a high vacuum chamber. 33

3.4 The precession of magnetization under the influence of a spin current: Time dependence of (a) M_z and (b) M_x, (c) The reversal process of magnetic moment. θ and ϕ represent the initial state of free layer magnetic moment. For PMA-MTJ, the switching behavior is mainly dependent on initial value of θ. 35

3.5 Three main breakdown mechanisms for MTJ barriers: Pinhole and shunt are soft-breakdown mechanisms that are typically lower than the intrinsic dielectric breakdown voltage [71]. 37

3.6 The TMR ratio of MTJ is dependent on the bias voltage. 38

3.7 Switching probability P_{sw} as a function of pulse width: the lines are theoretical values plotted from (3.18) and the markers are statistical results from 1000 times of Monte Carlo simulation under Cadence. 42
3.8 Temperature dependence of effective anisotropy field H_k and saturation magnetization M_s. 44
3.9 Temperature dependence of thermal stability factor and of MTJ based chip failure rate with 8 bits per word, different reading duration and different reading current. 45
3.10 Switching voltage versus average switching time at different temperature conditions. 45
3.11 Temperature evaluation of MTJ during current pulses. 46
3.12 Breakdown voltage for different configuration (P or AP) and stress voltage (positive or negative) versus MgO thickness, the markers are experimental results from [72] (t_{ox}=1.8nm, 2.1nm) and [21] (t_{ox}=0.9nm). 47
3.13 The MTJ state is changed as the applied voltage is higher than the critical switching voltage (V_{cP}, V_{cAP}). The resistance is steeply degraded beyond breakdown voltage (V_{bP}, V_{bAP}). 48
3.14 Time-to-failure statistics of MTJ at different stress voltages (t_{ox}=1.25nm). The time value corresponding to 0 of Weibull function represents 63% failure time. 50
3.15 Number of voltage pulses before breakdown (η) versus interval between voltage pulses (Δt) with δ=30ns and τ=100ns. 50
3.16 Number of voltage pulses before breakdown (η) versus interval between voltage pulses (Δt) with δ=30ns and τ=100ns. 52
3.17 Architecture of PMA STT MTJ compact model integrating physical models of reliability issues. 54
3.18 Component Description Format (CDF) in Cadence. 56
3.19 (a)Symbol of the PMA-STT-MTJ compact model (b) Symbol at circuit level. 57
3.20 Schematic of pre-charge sense amplifier circuit. 57
3.21 MC simulations of (a)bias voltage dependent resistance and (b) 1000 complete writing process with process variations. 59
3.22 MC simulations of 1000 complete writing process with the stochastic behaviors. The switching duration is set following a normal distribution with variation of 0.02. 60

3.23 Switching probability as a function of applied switching voltage and switching time. . . . 61
3.24 TMR evolution with temperature increase and the experimental data (red points) in [99]. . . . 61
3.25 Resistance of MTJ versus bias voltage with different temperatures. Critical current is reduced by increasing temperature. . . . 62
3.26 Dependence of reading error rate on the thickness of oxide barrier tox and area of MTJ. . . . 63
3.27 Lifetime of MTJ without (dashed lines) and with (lines) consideration of self-heating. The dots are experimental data in [74]. . . . 63
3.28 Simulation results of statistical model (1000 MC simulations) and worst-cases model (TT, FF, and SS): (a) current and (b) switching delay of MTJ in different states (P or AP) with a voltage pulse(σ=0.01 and n=3). . . . 67

4.1 Architecture of pre-charge sense amplifier based STT-MRAM cell circuit proposed in [49]. It consists of two parts: writing control part and PCSA part. . . . 71
4.2 Transient simulations of 4T-2M writing circuit and PCSA circuit. . . . 71
4.3 Dependence of reading error rate on the thickness of oxide barrier tox and area of MTJ. . . . 72
4.4 Switching probability with different writing voltages: the dashed lines are theoretical values plotted from equation (3.18) and the markers are statistical results of MC simulation. . . . 73
4.5 (a) Switching probability as a function of applied switching voltage and switching time. (b) Switching voltage versus average switching time at different temperature conditions. . . . 74
4.6 Reading error rate of PCSA with different area of circuit (SA is the minimum size of PCSA circuit) under different temperature conditions. . . . 74
4.7 Cumulative breakdown probability distribution for theoretical case (dashed lines), the simulation results of MTJ under constant voltage (circles) and MTJ integrated in CMOS circuit (stars). . . . 75
4.8 The studied symmetrical MFF is composed of two parallel MTJs, the writing block, a clocked sense amplifier, NAND-based slave SR latch and feedback loop. . . . 76

4.9 HBD failure gate current density: the breakdown sensitivity of transistor in MFF circuit. 76

4.10 Switching delay and time to failure of memory arrays: the cross are from 1000 elements of memory arrays with statistical model; the blue dot line is the Weibull function of the cross in which F signifies the failure probability of the memory elements; the triangles are from 1 element with the worst-case model (σ=0.01 and n=3). 78

4.11 Writing and sensing performance of STT-MRAM cell: the stars and dots are from statistical model (1000 MC simulations); the frames are from worst-case model of MTJs (σ=0.01 and n=4.5) and CMOS transistors. 80

4.12 Performance of delay time and dynamic energy in MFA circuit: the stars and dots are from statistical model (1000 MC simulations); the frames are from worst-case model of MTJs and CMOS (σ=0.01 and n=3). Two different discharge transistor sizes are considered: W/L=200nm/30nm and W/L=500nm/30nm. 81

4.13 The thin film devices in FDSOI technology with a cross-section view of planar/2D structure FDSOI CMOS. Body bias voltage can impact transistor performance. Poly bias is achieved by additional gate length. 83

4.14 Variability FDSOI: A single NMOS transistor works in saturation region. The coefficient of V_{th} variation is analyzed among FBB, nominal design (no body bias) and FBB. 84

4.15 (a) Pre-charge sense amplifier with dynamic asymmetrical bias bias (b) RC circuits generate the body bias voltages for transistors in PCSA. 85

4.16 Waveform of proposed Non-volatile flip-flop circuit using dynamic asymmetrical body bias. 86

4.17 Reading error rate of the proposed NVFF versus different TMR value: FBB means dynamic asymmetrical body bias, and NBB means nominal body bias. 86

4.18 Reading error rate of the NVFF versus process variations: MTJ parameters t_{sl} , t_{ox}, TMR follow normal distribution around the mean value μ with the deviation σ. 87

4.19 Reading error rate of the proposed NVFF versus different supply voltage: FBB means asymmetrical forward body bias, and NBB means nominal body bias. 88

4.20 Reading error rate of the proposed NVFF versus different thermal conditions: FBB means asymmetrical forward body bias, and NBB means nominal body bias. 88

5.1 Switching probability of MTJs on function of switching current with 10ns pulse. The current with 50% switching success is indicated above. This figure is obtained by 1000 runs of Monte-Carlo simulation with the same MTJ under voltage pulses. 93

5.2 Architecture of proposed MTJ-based true random number generator: The random writing circuit generates a switching current to write the MTJs with 50% success and it is controlled by the correction block; The MTJ writing part enables MTJ switching (generating random number or resetting to initial state); The correction block composed of counter and comparator is used to execute real-time output probability tracking and send feedback to writing block. 94

5.3 MTJ writing circuit and PCSA: I_{sw} is the switching current flowing through MTJ during random writing phase and I_r is the switching current during reset phase. N_{c0}, N_{c1} and N_{c2} modulate the switching current according to the random number probability obtained in the precedent cycle. 95

5.4 The phase transition diagram of proposed circuit design: The three states in blue frame are with different output random number probability after reset phase; The state in green signifies the unknown switching probability after random writing phase; The three states in yellow are with different known random number probability after sensing phase. For MTJs, '0' represents P state and the resistance is relatively low. '?' represents unknown information. 96

5.5 Time-domain diagram of proposed true random number generator. During each cycle, the MTJs are firstly reset to the initial state (with I_r=178μA for P state and I_r=142μA for AP state), then randomly switched, and finally sensed at the output. The initial current is set for 50% of switching success. 97

5.6 Output '1' probability versus number of clock cycles: The output random number probability becomes stable after 30 cycles (The probability of '1' occurrence stays around 50%) for all the five corner models. 97

5.7 Examples: SC based on two input combinational logic (AND, OR, XOR and scaled addition). 100

5.8 Proposed stochastic bit generator with 4T1M structure. 1000 runs monte-carlo simulation illustrates the stochastic behavior of MTJ. 100

5.9 Simulation result: switching probability versus MTJ operation current. . . . 101

5.10 Layout of 4T1M SNM with 28nm FDSOI process. 101

5.11 Polynomial function synthesis with traditional binary signal. 102

5.12 An example of polynomial function synthesis. 103

5.13 Conventional CMOS approximate adders: AXA1, AXA2 and AXA3 [183]. . 105

5.14 Circuit implementation of two approximate MFAs: AX-MFA1 (without dashed line box) and AX-MFA2. The first approximate AX-MFA1 is implemented with conventional simplified logic: input C_i in dashed rectangle is eliminated to get an approximate $Sum = A \otimes B$. The second dual-mode approximate AX-MFA2 is implemented with the whole schematic. 106

5.15 The transition simulation waveforms of approximate adder with reduced logic complexity (AX-MFA1). Output Sum is with errors. 107

5.16 The transient simulation waveforms of approximate adder by insufficient writing current (AX-MFA2). 108

5.17 Supply voltage strategy in bi-mode MFA. 108

5.18 Cross-sectional view of dynamic well FDSOI MOS devices. Different well configurations impact circuits performance. 109

5.19 4.41μm*1.98 μm Layout with planar 28nm FDSOI technology. A 16nm poly bias is used to reduce leakage power and enhance yield. LVT-RVT strategy is performed in layout, single P-well covers nMOS RVT transistor (in sense amplifier) and pMOS LVT transistor (in logic network). 110

5.20 Latency simulation of dual-mode MFA. A 152.7 ps latency is realized in approximate Sum operation when V_{dd}=0.5V. Continuous supply scaling down to sub-V_t region leads to large latency (1.27 ns when V_{dd}=0.36V). . . 110

5.21 Probability with respect to V_{dd} scaling. MOS/MTJ process variations and MTJ stochastic effect influence MFA probability. 111

5.22 Sensing probability with respect to V_{dd} considering process variation. Different well configurations impact sensing error rate. Single N-well doping method achieves the extra V_{dd} margin. 112

B.1 Évènements importants de la recherche et du développement en spintronique. 158

B.2 Effet tunnel dépendant du spin des électrons dans un MTJ, tandis que les directions d'aimantation dans les deux couches FM sont (a) parallèles et (b) en antiparallèle. 160

B.3 Structure standard de MTJ. 161

B.4 Schématique de (a) cross-point array et (b) 1T/1MTJ mémoire cellule architecture 163

B.5 Circuits typiques de MOS/MTJ NV-LIM basés sur une structure d'amplificateur de détection de pré-charge : portes logiques, additionneur complet et bascule. 164

B.6 Architecture du modèle compact de PMA STT MTJ intégrant des modèles physiques de problèmes de fiabilité. 168

B.7 Les simulations MC de (a) la résistance dépendante de la tension de polarisation et (b) 1000 processus d'écriture complet avec des variations de processus. 169

B.8 (a) Simulations de MC de 1000 processus d'écriture complète avec les comportements stochastiques. (b) Probabilité de commutation en fonction de la tension de commutation et du temps de commutation. 170

B.9 (a) Evolution du TMR avec augmentation de la température et données expérimentales (points rouges) dans [99]. (b) Résistance du dispositif MTJ par rapport à la tension de polarisation à différentes températures. Le courant critique est réduit en augmentant la température. 170

B.10 Durée de vie de MTJ sans (lignes pointillées) et avec (lignes) prise en considération de l'auto-échauffement. Les points sont des données expérimentales dans [74]. 171

B.11 (a) Taux d'erreur de lecture versus l'épaisseur de MgO et la surface de MTJ. (b) Probabilité de commutation avec différentes tensions d'écriture. 172

B.12 (a) Probabilité de commutation en fonction de la tension de commutation et du temps de commutation appliqués. (b) Tension de commutation en fonction du temps moyen de commutation à des températures différentes. . 173

B.13 (a) Taux d'erreur de lecture du PCSA avec la surface différente du circuit (SA est la taille minimum du circuit de PCSA) dans différentes conditions thermiques. b) Distribution cumulative de probabilité de claquage. 173

B.14 Performance d'écriture et de lecture de la cellule STT-MRAM: les étoiles et les points sont issus du modèle statistique (1000 simulations MC); Les trames viennent du modèle le plus défavorable de MTJs ($sigma = 0.01$ et $n = 4.5$) et de transistors CMOS. 175
B.15 Variabilité FDSOI: Un seul transistor NMOS fonctionne dans la région de saturation. Le coefficient de V_{th} variation est analysé entre FBB et la conception nominale (sans biais corporel) et FBB. 176
B.16 (a) Amplificateur de détection de pré-charge avec polarisation du substrat asymétrique dynamique (b) Les circuits RC génèrent les tensions de polarisation du substrat pour les transistors dans PCSA. 177
B.17 Taux d'erreur de lecture du NVFF en fonction de différentes (a) tensions d'alimentation et (b) conditions thermiques. 178
B.18 Architecture du circuit de TRNG proposée. 179
B.19 Diagramme temporel du circuit proposé. 180
B.20 Résultat de simulation: probabilité de commutation par rapport au courant de fonctionnement MTJ. 181
B.21 Un example de synthèse de fonction polynomiale. 182
B.22 Les simulations de transition de l'additionneur approximatif avec une complexité logique réduite (AX-MFA1). La sortie Sum est avec des erreurs. . . 183
B.23 Les simulations de transition de l'additionneur approximatif avec double mode (AX-MFA2). La sortie Sum et C_o est avec des erreurs. 183
B.24 Simulation de latence de MFA à double mode. 184

List of Acronyms

AFM Anti-Ferromagnetic

AP Anti-Parallel

ASL All Spin Logic

ASIC Application-Specific Integrated Circuit

BER Bit Error Rate

BL Bit Line

BSIM Berkeley Short-channel IGFET Model

CAD Computer-aided Design

CAM Content Addressable Memory

CDF Component Description Format

CIMS Current-induced magnetization switching

CIP Current In Plane

CMOS Complementary Metal-Oxide-Semiconductor

CPP Current Perpendicular to Plane

CPU Central Processing Unit

DC Direct Current

DRAM Dynamic Random Access Memory

DW Domain Wall

ECC Error Correction Circuit

EDP Energy-Delay Product

FIMS Field Induced Magnetic Switching

FM Ferromagnetic

FPGA Field Programmable Gate Array

GMR Giant MagnetoResistance

HBD Hard Breakdown

HDD Hard Disk Drive

HKMG High-K Metal-gate

IC Integrated Circuit

IEEE Institute of Electrical and Electronics Engineers

ITRS International Technology Roadmap for Semiconductors

LLG Landau-Lifshitz-Gilbert

LUT Look Up Table

MC Monte-Carlo

MFA Magnetic Full Adder

MOSFET Metal Oxide Semiconductor Field Effect Transistor

MRAM Magnetoreisistance Random Access Memory

MTJ Magnetic Tunnel Junction

MTTF Mean-Time-to-Failure

NM Non-Magnetic

NML Nanomagnetic Logic

NMOS N-Channel Metal Oxide Semiconductor

OTV Oxide Thickness Variation

OxRAM Oxide Random Access Memory

P Parallel

PCRAM Phase-Change Random Access Memory

PCSA Pre-Charge Sense Amplifier

PDF Probability Density Function

PMA Perpendicular Magnetic Anisotropy

PVT Process-Voltage-Temperature

RA Resistance-Area Product

RAM Random Access Memory

RDF Random Dopant Fluctuations

RM Racetrack Memory

RRAM Resistive Random Access Memory

RV Resistance variation

SBD Soft Breakdown

SC Stochastic Computing

SHE Spin Hall Effect

SL Source Line

SoCs Systems-on-Chip

SOT Spin Orbit Torque

SPICE Simulation Program with Integrated Circuit Emphasis

SRAM Static Random Access Memory

STT Spin Transfer Torque

TAS Thermally Assisted Switching

TBD Time-to-Breakdown

TCAD Technology Computer-aided Design

TDDB Time-Dependent Dielectric Breakdown

TMR Tunnel MagnetoResistance

TRNG True Random Number Generator

TTF Time-to-Failure

WL Word Line

Chapter 1

Introduction

1.1 Motivations

Charge and spin are the two intrinsic attributes of an electron, which determine its macroscopic behaviors. Before the discovery of giant magnetoresistance (GMR), the investigations on the charges and spins of electrons were usually considered to be independent of each other and little attention was paid to the correlation between these two attributes [1]. The charge-based devices have changed the way we create, produce and even think since their birth in 1947. With the quick development dominated by Moore's law, the number of transistors in a dense integrated circuit (IC) have successfully doubled approximately every two years (Or 18 months from aspect of chip performance) for decades. Among all the transistor devices, complementary metal-oxide-semiconductor (CMOS) technology is the most widely used in the ICs nowadays. The development of IC in the digital age is determined by the scaling down of CMOS technology node. Moore's prediction has been used in the semiconductor industry to guide long-term planning and to set targets for research and development for several decades.

However, the scarcity of resources such as power consumption and interconnect bandwidth has become the bottleneck to continue Moore's scaling [2]. It was predicted by the International Technology Roadmap for Semiconductors (ITRS) that the memory static power in 2026 will be triple that in 2016 [3]. This trend is due to the increasing contribution of the leakage current to the total power consumption as CMOS technology node shrinks blow 90 nm [4]. Thus, the off-state leakage is considered as the critical obstacle for further scaling down of CMOS technology node. Meanwhile, with the emergence of cloud and internet of things (IoT), seamless interaction of big-data and instant data have become necessary. For the essential elements of IoT (e.g., sensors), emerging devices with

features of ultra low power and high performance are required to generate the data instantly with few consumption. From the other part, abundant computing and memory resources are required in the Big data to generate the service and the information that clients need. The conventional CMOS circuits can not meet these urgent requirements. In this background, the emerging spintronic devices which combine the two attributes of electron (charge and spin) are considered as a promising solution because of non-volatility and fast speed operation. Compared with the conventional CMOS based memories, spintronics based memories can retain the stored information without power supply. Moreover, with easy 3D integration, spintronic devices are deposited on the top of arithmetic units, which avoids the large data traffic of the conventional Von-Neumann architecture and thus reduces the operation latency and improves energy efficiency.

The development of spintronics devices originates from the discovery of Giant Magnetoresistance (GMR) effect in 1988 by Albert Fert and Peter Grünberg [5, 6]. From then on, many academic and industial researchers have concentrated on the emerging materials to explore better energy efficiency of spintronics devices. As one of the most important spintronics devices, magnetic tunnel junction (MTJ) is a promising candidate for the next generation of non-volatile memories. MTJ consists of one nonmagnetic layer sandwiched by two ferromagnetic layers in which the Tunnel MagnetoResistance (TMR) effect was discovered for the first time by [7] in 1975. The resistance of MTJ depends on the relative magnetization orientation of the two ferromagnetic layers (R_p at parallel state and R_{ap} at antiparallel state). As the MTJ resistance can be configured comparable with CMOS transistors, it can be integrated in the memories and logic circuits to represent logic '0' or '1'. Its characteristic is quantified by TMR ratio $((R_{ap} - R_p)/R_p)$. The research and development of MTJ has become intensive since the first experimental demonstration of TMR effect based on the amorphous AlxOy barrier at room temperature (TMR ratio was 18% and 11.8%) in 1995 [8, 9]. Even though the TMR ratio has been improved up to 70% at room temperature (RT) [10] with materials and technology optimization, this low value limited the application of MTJ into CMOS circuits. The single-crystalline MgO was introduced into MTJ by Shinji Yuasa in 2004, which increased the TMR ratio up to 180% at RT [11]. A TMR ratio as high as 604% at 300 K in Ta/Co20Fe60B20/MgO/Co20Fe60B20/Ta pseudo-spin-valve magnetic tunnel junction was observed by Shoji Ikeda in 2008 [12], which is obtained by optimizing annealing temperature and suppressing the Ta diffusion into CoFeB electrodes and in particular to the CoFeB/MgO interface.

As a promising memory candidate, the switching approaches of MTJ are always with intensive research. Field Induced Magnetic Switching (FIMS) was firstly employed in the early realizations of MTJ based magnetoresistive random access memory (MRAM) [13, 14]. Too high currents (∼10mA) are required by this switching method to generate magnetic fields, which becomes a critical constraint for FIMS to realize high density and low power memory due to high power consumption, large die area and high disturbance. Thermally Assisted Switching (TAS) was proposed and by Bernard Dieny and Jean-Pierre Nozières in 2003, which largely decreases the threshold of switching current [15]. In this method, a current flows into MTJ to heat the MTJ and facilitates the switching by another. TAS has effectively decreased the power consumption of writing operation (switching current∼1mA), but the scalability issue still remains unsolved and the switching speed is lower due to the necessary cooling down after the heating. To address the power and scalability issue, a novel switching approach of Spin Transfer Torque (STT) was firstly predicted theoretically by John Slonczewski and Luc Berger in 1996 [16, 17] and observed experimentally by many research groups in 2000 [18, 19]. This method uses a relatively low current (∼100uA) flowing through the MTJ to switch its state. Without the need of magnetic field, STT makes it possible to achieve high density and low power MRAM. MTJ with interfacial perpendicular magnetic anisotropy (PMA-MTJ) was discovered by Shoji Ikeda in 2010 [20] which features low switching current (49uA), and high thermal stability. Figure 1.1 demonstrates the evolution of the most significant breakthroughs of spintronics research and development.

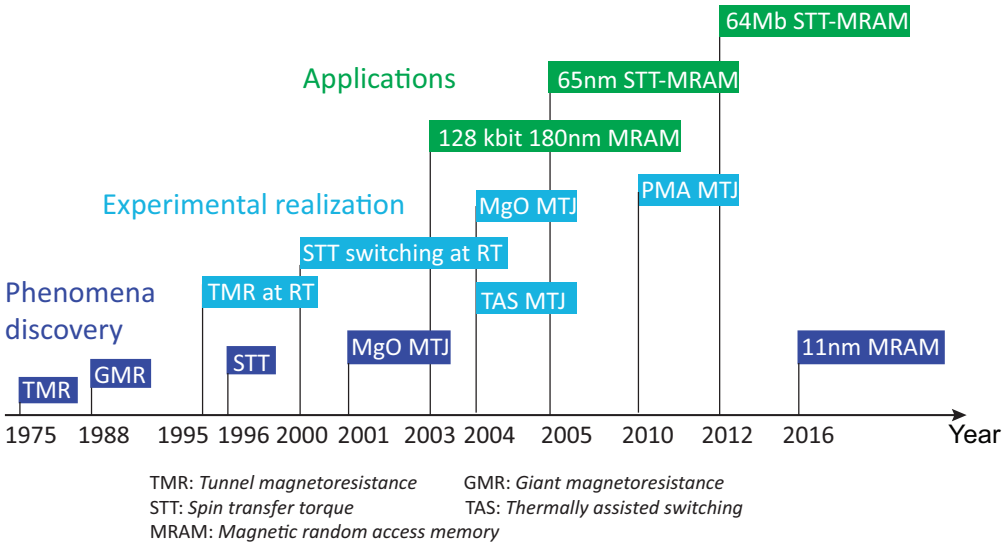


Figure 1.1: Breakthroughs in spintronics research and development for memory.

Despite the outstanding potentials in STT-MRAM, its wide commercialization still

remains very challenging due to poor reliability. As STT switching method has been demonstrated intrinsically stochastic [21], a relatively high current density is required for successfully switching in writing process. With ultra-thin layers(∼1nm) and small die area in MTJ, MTJ suffers from extreme work conditions such as intense electric field across oxide barrier and high current density flowing through it. As a result, the performance is severely degraded in terms of self-heating effect, process variations and aging mechanisms. The reliability risks can be involved from initial design to tape-out, till the final wear-out. All will have significant impact on quality and yield of MTJ based circuits.

Research work on reliability mainly concentrates on defects modeling, reliability analysis, reliability-aware methodology, and failure prediction [22]. Defects modeling characterizes physical defects and maps the degradation to parameters at device level (e.g., BSIM4 model), which is the basic work of the latter three. With the fast evolution of STT-MRAM, the reliability has attracted the attention of researchers([21, 23, 24]). The reliability issues of MTJ have always been well characterized theoretically and experimentally. However, there exists not yet a compact model comprising all the possible reliability issues for circuit designers. With the expensive cost of MTJ fabrication, it is very profitable to identify and address the possible reliability issues and thus provide reliability-aware circuits at the early design phase.

This thesis is dedicated to provide a thorough understanding of the sources of the possible reliability issues in MTJ and propose an accurate compact model for circuit designers. This model can be used to predict the possible functional failures of MTJ based circuits and to address all the issues at the early design phase. By using this model, we have carried out reliability analysis and explored some design strategies to tolerate the reliability issues and improve the circuit performance. Finally, some novel realizations of conventional specific circuits are presented to benefit from the reliability issue of MTJ.

The thesis is part of the project “ANCD2” funded by IDEX Paris-Saclay, ANR-11-IDEX-0003-02 supported by French National research Agency (ANR). The project concentrates on the control and diagnosis of components and devices in the application of nanotechnologies.

1.2 Thesis contributions

This thesis is focused on the reliability analysis of hybrid MTJ/CMOS circuits from device level to circuit level. The main research contributions are as follows:

- Investigation of reliability issues in magnetic tunnel junction (MTJ): synthesis of the physical mechanisms and quantification by theoretical deductions.

- Compact modeling of main reliability issues in MTJ, which includes process variations, stochastic switching, temperature fluctuation and dielectric breakdown.

- Proposition of a worst-case corners model for fast performance evaluation of variability-awareness. This model provides faster simulation speed while guaranteeing a high level of analysis quality, especially in very large scale circuit.

- Integration of proposed model into memory and logic circuits for reliability assessment to validate its functionality. The methods for performance estimation are presented in details for hybrid MTJ/CMOS circuits.

- A novel circuit design methodology for variability tolerant circuits and systems (Dynamic asymmetrical body bias for symmetrical structure based circuits). This design features faster operation speed and less sensing errors.

- Realization of true random number generator using stochastic switching behavior of MTJ. The functionality is well confirmed and its robustness is optimized by correction systems.

- New circuits of MTJ-based approximate computing and stochastic computing. The performance of these circuits are significantly improved in terms of area and power consumption.

1.3 Organization of the thesis

The organization of this thesis is as follows:

Chapter 2 presents the background of this thesis in details. We firstly introduce the physical mechanisms of spintronic devices and then concentrate on the working principles of magnetic tunnel junction (MTJ). An overview of MTJ based MRAM and computing circuits is also presented. Meanwhile, the research on reliability analysis of MTJ based circuits is studied and existing compact models of MTJ are investigated.

Chapter 3 proposes a compact model of STT-PMA-MTJ programmed in VerilogA language which includes the main reliability issues. The origins of the main reliability

issues in MTJ are well studied. The physical models used for describing the functional behaviors and reliability issues are well confirmed. After introducing the employed modeling method and programming language VerilogA, the simulation results of the model are demonstrated.

Chapter 4 applies the proposed model in the hybrid MTJ/CMOS circuits to study their reliability. Based on these analysis, we explore some methodologies to improve the circuit robustness and yield probability. These methodologies are implemented in certain designs to demonstrate its feasibility and performance in terms of speed, energy consumption and area.

Chapter 5 tries to explore the usage of MTJ reliability issues in special applications. As a significant functional failure issue, stochastic switching behavior of MTJ can be appropriately inserted in the security applications as an intrinsic randomness source. We carry out the detailed circuit design and execute simulations to verify the functionality and performance. MTJ is also used in approximate computing and stochastic computing to realize low power and low complexity circuit.

Chapter 6 concludes the work realized during this thesis and presents some perspectives relative to the thesis and future research directions.

Chapter 2

State of the art

This chapter presents the preliminary work relative to the reliability of MTJ device. Firstly, the physics of MTJ are introduced in details. Then, the main applications based on MTJ are discussed and compared in terms of performance and reliability. Finally, the current status of research on main reliability issues of MTJ is reviewed and the required work is synthesized.

2.1 Magnetic tunnel junction

2.1.1 MTJ working principles

Spintronics is an emerging technology which concentrates on the correlation between the two attributes of an electron: spin and charge [1]. Before the appearance of the discipline Spintronics, the research is dominated by manipulating the charge of electron from classical conductors as copper to semiconductor as silicon. In these devices, there is no spin polarization because the spin direction is naturally random. The most outstanding breakthrough of Spintronics was the discovery of Giant Magnetoresistance (GMR) effect in 1988 by Fert and Grünberg [5, 6].

GMR effect was observed in stacks composed of thin ferromagnetic (FM) and non-magnetic (NM) layers such as metal. Naturally, electrons have two spin states: spin-up and spin-down which are discovered in paired electrons. In FM layer, the number of spin-up (majority) and spin-down (minority) are totally different, resulting in the different contribution to electrical transport regarding to the amount of conducting electrons. This contribution is defined as spin polarization P:

$$P = \frac{n\uparrow - n\downarrow}{n\uparrow + n\downarrow} \tag{2.1}$$

where n↑ and n↓ are the numbers of spin-up and spin-down electrons, respectively. GMR effect can be explained by a simplest form demonstrated in the Figure 2.1. When injecting a current into a FM layer such as Fe and Co, only the electrons with specific spin direction will be able to pass through. Thus, if the two FM layers have parallel (P) magnetization direction, the electrons with one specific spin direction will travel through the sandwich nearly without scattering while those with opposite direction can not pass. The structure behaves relatively low resistance R_p. Respectively, in the case of anti-parallel (AP) magnetization direction, both spin-up and spin-down electrons will pass partially, leading to a relatively higher resistance R_{ap} [13]. With special composition of materials in some multilayers structure, the relative magnetoresistance $\Delta R/R = (R_{AP} - R_P)/R_P$ can reach 100% or more. In fact, the first discovery was aready 80% in the Fe/Cr multilayer [5]. Many applications have been realized by profiting GMR effect, such as "spin valve" which has been widely used in the hard disk drives (HDDs) as read heads [25]. With intense research interest, the areal density of spin valve based HDDs has been increased by three orders of magnitude (from ∼0.1 to ∼100 Gbit/in^2) between 1991 and 2003 [13].

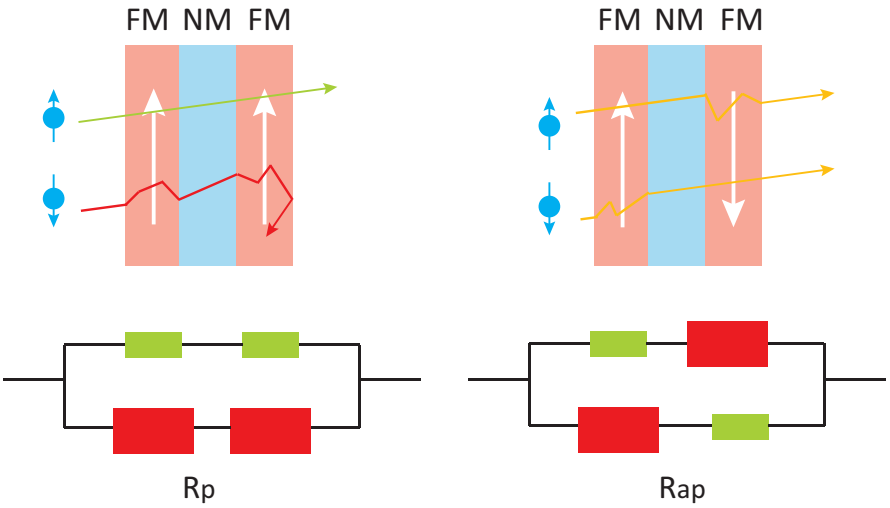


Figure 2.1: Equivalent resistance model to describe GMR effect in the structure of non-magnetic (NM) layer sandwiched by two ferromagnetic (FM) layers: Anti-parallel (AP) state presents higher resistance value than parallel (P) state.

Another important breakthrough of Spintronics is the discovery of tunnel magnetoresistance (TMR) effect by Julliere in 1975 [7], in which the non-magnetic metal layer is replaced by an insulating layer. The phenomenon can be microscopically explained from the viewpoint of band structure, which is demonstrated in Figure 2.2 [13]. In FM materials, the populations of spin-up and spin-down are different at the Fermi energy level, leading to unequal density of states available for each [26]. As a result, the FM material is

magnetized by the net magnetic moment generated by the disequilibrium. The electrons near the Fermi level act as carriers during the transport. The spin-polarized electrons pass through the oxide barrier by tunnel effect with conservation of spin state: An electron with spin-up state from one FM layer can travel across the insulator only if it can find a spin-up state at the Fermi level of the other FM layer. If the magnetization directions of the two FM layers are parallel (P), all the spin-up and spin-down electrons can easily find a corresponding state after traveling through the barrier because the band structures of two FM layers are almost the same. Inversely, if they are anti-parallel (AP), only partial electrons can act as carriers for the tunneling current, resulting in a lower conductance than AP state. Thus, the resistance of the trilayer stack is different according to the magnetization state of FM layers.

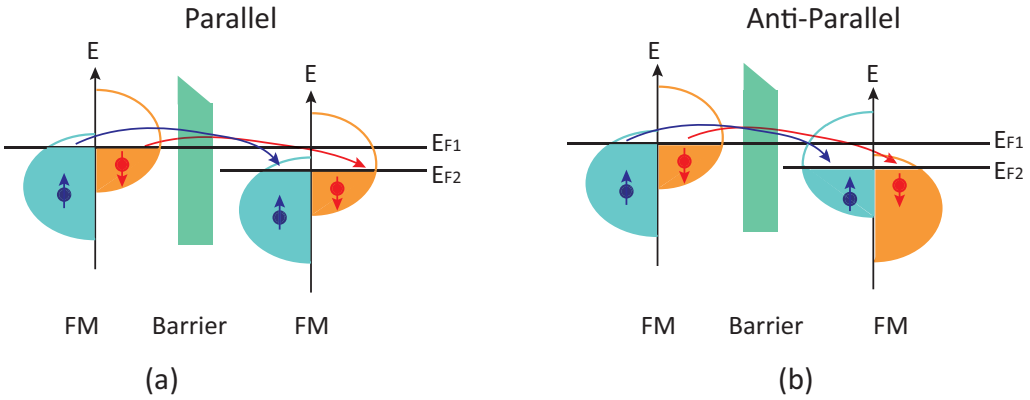


Figure 2.2: Spin-dependent tunneling of electrons in an MTJ while the magnetization directions in two FM layers are (a) parallel and (b) anti-parallel.

Magnetic tunnel junction (MTJ) is created by using this phenomenon, which induces much research effort and becomes a promising memory candidate. Figure 2.3 demonstrates a typical structure of MTJ stack which mainly consists of three layers: a thin insulator (oxide barrier such as AlxOy and MgO) sandwiched by two ferromagnetic layers (e.g., CoFe). The two FM layers are with different configurations: one with a fixed spin magnetization direction which is noted as pinned layer or reference layer; whereas the other one can be changed in two directions (storage layer, switching layer or free layer). Thus, parallel (P) and anti-parallel (AP) are usually used to describe the two different configurations of MTJ. The MTJ configuration can be tuned by switching the spin magnetization orientation in the storage layer, which can be achieved by a magnetic field above the threshold value with opposite direction.

With oxide barrier between two ferromagnetic layers, MTJ behaves resistance value which is comparable with CMOS transistor technology. This makes it possible to detect

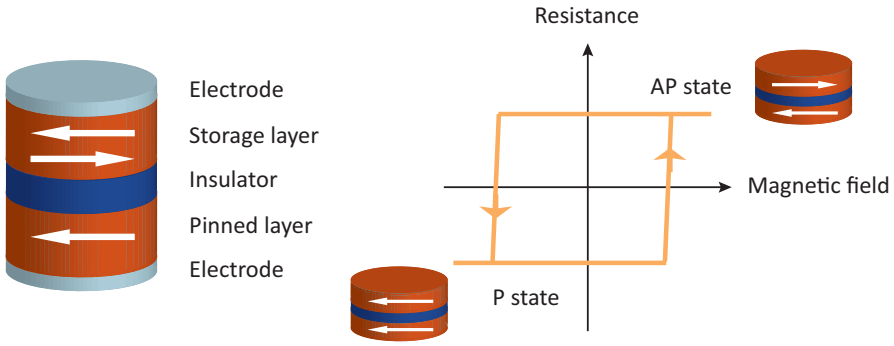


Figure 2.3: MTJ consists of three layers: two ferromagnetic layers separated by an oxide barrier. The nanopillar resistance (R_p, R_{ap}) depends on the corresponding state of the magnetization of the two ferromagnetic layers Parallel (P) or Anti-Parallel (AP). The MTJ state can be switched by modulating the magnetic field.

the state of MTJ using CMOS based sense amplifier and generate logic '0' and '1' with specific design. TMR ratio is one of the most important parameters which determines the performance of MTJ device. It is defined as follows:

$$TMR = \frac{\Delta R}{R_P} = \frac{R_{AP} - R_P}{R_P} \tag{2.2}$$

where R_P and R_{AP} are the MTJ resistances of P and AP state. From the Figure 2.2, it can be deduced that the TMR ratio is determined by the spin polarization of the FM layers, which can be expressed by 2.3:

$$TMR = \frac{2P_1P_2}{1 - P_1P_2} \tag{2.3}$$

where P_1 and P_2 are the spin-polarization in two FM layers which can be calculated by equation (2.1).

For better immunity to process variations and mismatch generated in fabrication process, high TMR value is always preferred, which has been the motivation of intense research and fast development of MTJ. Recently, new ferromagnetic materials, oxide barrier and MTJ process have been exploited to achieve higher TMR value (e.g., CoFeB as FM layer and MgO as oxide barrier).

2.1.2 MTJ switching approaches

As aforementioned, the switching of MTJ state can be realized by changing the spin magnetization orientation in the storage layer. Several switching approaches have been proposed since the appearance of MTJ. This section will review these switching methods

and evaluate their efficiency.

2.1.2.1 Field-induced magnetic switching (FIMS)

Field Induced Magnetic Switching (FIMS) is the main switching approach in the first generation of MTJ device [27]. As depicted in Figure 2.4, the magnetic state of MTJ is written by means of a magnetic field generated by currents flowing through two orthogonal write lines. To write information in the MTJ, I_b works as the bit line which generates a magnetic field to switch the spin magnetization direction of the storage layer while I_w operates as the word line to assist the above operation. Thus, the written state is determined by the polarity of I_b. The two writing lines used in FIMS allow this writing approach easy to be addressed in memory array. It can be observed in Figure 2.4 that the lines for sensing operation are entirely independent with those for writing. Thereby, the two operations can be asynchronous, resulting in better flexibility of hybrid circuit design of FIMS-MTJ and CMOS than other writing methods.

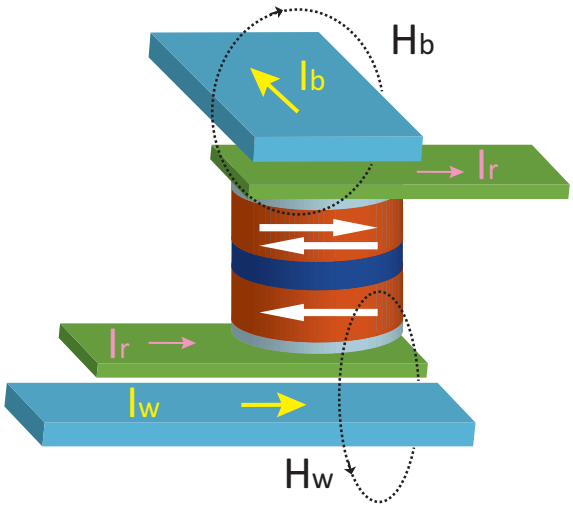


Figure 2.4: Field induced magnetic switching approach structure.

However, the combination of two perpendicular pulses of magnetic fields should be precisely configured to execute correctly the writing selectivity. This may lead to narrow operating window induced by half-selectivity disturbance [28]. Moreover, the external fields generated also have impact on the devices nearby, which limits the realization of high density FIMS-MRAM. The most severe issue of this approach is the high currents (∼10 mA) needed to generate magnetic fields, hindering its integration with conventional CMOS transistors due to the limit of electromigration issue. In 2005, Freescale proposed and patterned the toggle switching approach , which increases the energy barrier during programming and then reduces significantly the disturbance problem. In this method,

Synthetic Anti-Ferromagnetic (SAF) layers have been used to replace one storage layer. Based on this optimization, Freescale commercialized the first MRAM product in 2006 (4 Mbit). Despite the continuous optimization, this approach can not meet the increasing demand for high speed, high density and low power in large scale MRAM designs.

2.1.2.2 Thermally assisted switching (TAS)

Thermally assisted switching (TAS) was proposed by SPINTEC laboratory to improve the performances of write selectivity, power consumption and thermal stability of MTJ [15, 29]. As illustrated in Figure 2.5, an additional anti-ferromagnetic (AFM1) layer with low blocking temperature (TB2 ∼160°C) is normally added above the storage layer and the reference layer is pinned by another (AFM2) with a much higher blocking temperature TB1 (typically∼300°C). This configuration enhances the flexibility of storage layer and facilitates the switching while prevents any magnetization switching of the reference layer. For the write operation, a current is injected into MTJ (I_h) to heat up the FM layers above their magnetic ordering temperature, particularly the storage layer. When the temperature exceeds TB2, the spin magnetization direction of storage layer can be easily reversed by a small magnetic field generated by (I_b). I_h is mono-directional whereas I_b is bidirectional.

Compared with FIMS, TAS features relatively lower power, higher density and lower switching disturbance between memory cells. However, the switching speed is limited by the existence of heating and cooling duration, which make it not appropriate for high speed logic applications, such as magnetic flip-flop (MFF) and magnetic arithmetic units. In addition, the heating process increases the average temperature of the entire MTJ stack, which accelerates the breakdown of oxide barrier and results in relatively short time to failure.

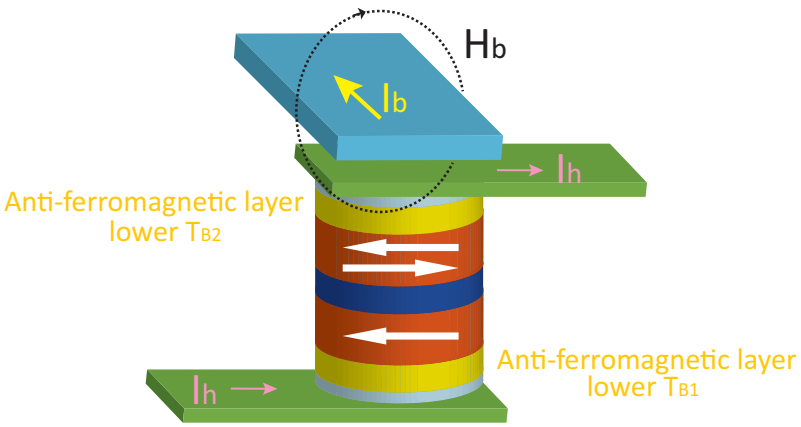


Figure 2.5: Thermally assisted switching approach structure.

2.1.2.3 Spin transfer torque (STT)

Spin transfer torque (STT) was predicted independently by Berger and Slonczewski in 1996 [16, 17], which promises much better energy efficiency and scalability than the two switching approaches presented above. From the view of electrical property, STT switching method only requires a bidirectional current I higher than the threshold current to change the state of MTJ (see Figure 2.6). It was observed that a spin-polarized current injected perpendicularly to the plane could influence the magnetization of FM layers. The transfer of spin angular momentum from a spin-polarized current to a local magnetization of the FM layer can generate a large torque (noted as spin transfer torque) to the magnetization to this FM layer. This torque efficiently facilitates the magnetic manipulations of FM layers in MTJ than the aforementioned switching methods using magnetic fields alone. If the current density exceeds the threshold value, the torque applied by the current will change the magnetization of the free layer (FL) of MTJ [30].

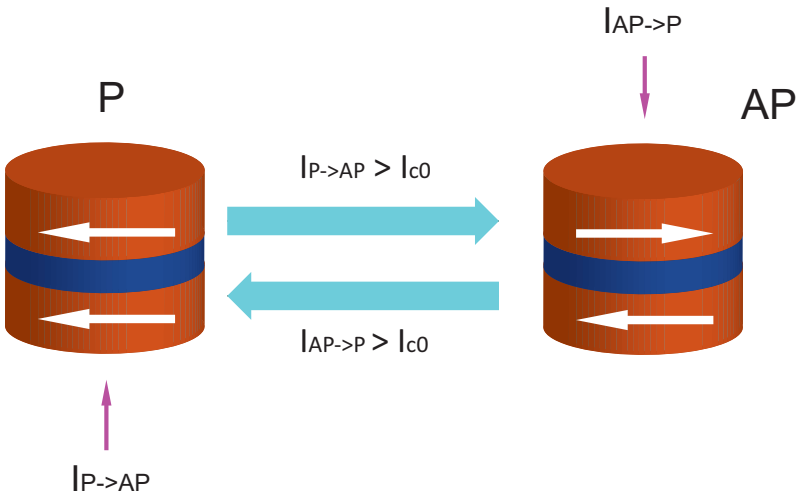


Figure 2.6: Spin transfer torque switching approach structure.

In STT-MTJ, the electrons injected into one FM layer are polarized and then transfer angular momentum by applying a torque on the magnetization of the other FM layer after tunneling across the oxide barrier. The basic considerations for spin-transfer torque devices can be illustrated in a single domain model, which assumes that the layers are uniformly magnetized [18, 19]. The dynamics of magnetization switching of free layer (FL) can be described by a Landau-Lifshitz-Gilbert (LLG) equation including the STT [31, 32] as following equation:

$$\frac{\partial \vec{m}}{\partial t} = -\gamma \mu_0 \vec{m} \times \vec{H_{eff}} + \alpha \vec{m} \times \frac{\partial \vec{m}}{\partial t} - \beta J \vec{m} \times (\vec{m} \times \vec{m}_r) \tag{2.4}$$

where $\vec{m}$ represents the unit magnetic moment of the FL magnetization under the macrospin approximation, where H_{eff} is the effective magnetic field, which is the sum of different magnetic fields, such as the external magnetic field, the demagnetization field and the anisotropy field, the magnetostatic field, the Oersted field and the exchange coupling field. γ is the gyromagnetic ratio, μ_0 is the vacuum permeability. α is the Gilbert damping constant, $\hbar$ is the reduced Planck constant, β is the is the STT coefficient depending on both the spin polarization and the geometric configuration of the spin torque efficiency, J is the switching current density, $\vec{m}_r$ is the unit vector of the reference layer (RL) magnetization.

This equation can be understood using Figure 2.7 [33]. On the right side of the equation, the first term represents the precession of the field-induced magnetization, the second describes the intrinsic Gilbert damping torque which reduces the precessional angle as a function of time and leads to the relaxation of the precession, the last is the STT term with the opposite direction of the damping vector which induces the switching of magnetization momentum. In such a current-induced magnetization switching MTJ, the switching is determined by the competition between damping term and the STT term. For instance, the STT term generated by a small current is relatively weaker than the damping term, leading to unchanged magnetization direction. Contrarily, the STT term generated by a high current is stronger than the damping term, resulting in larger precessional angles and eventual state switching. The two regimes are distinguished by the threshold current (noted as critical current I_{c0}).

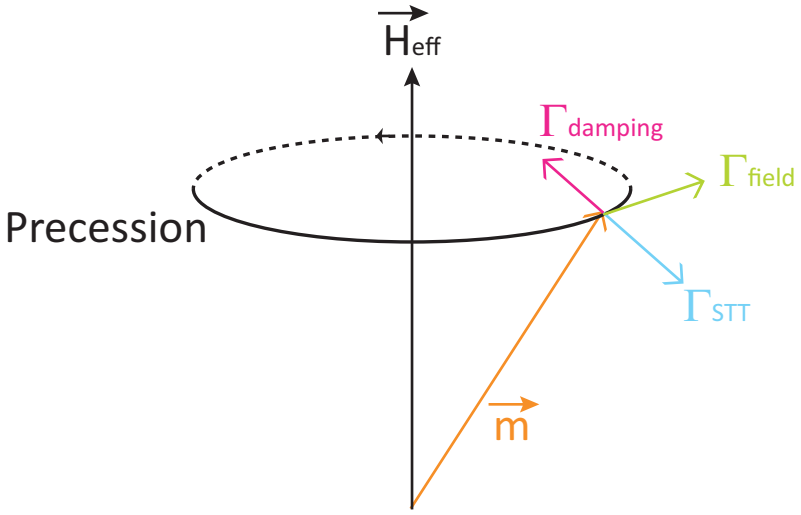


Figure 2.7: Diagram of the LLG equation: $\Gamma_{damping}$ is the Gilbert damping torque, Γ_{STT} is the STT term and Γ_{field} is the effective field torque generated by effective magnetic field H_{eff}.

With only a bi-directional current, this current-only approach simplifies drastically

the switching process. Furthermore, the magnitude of current required by STT is significantly reduced comparing with the previous switching methods (normally less by an order). Consequently, higher density and faster speed can be achieved in STT-MTJ based MRAM. Since its practical demonstration, STT switching approach is considered as the most promising candidate for the future MRAM applications.

2.1.2.4 Thermally assisted spin transfer torque (TAS+STT)

Thermally assisted spin transfer torque (TAS + STT) switching is an emerging approach combining the TAS mechanism with spin transfer torque effect [34]. Similar to TAS, an additional Anti-ferromagnetic layer is required to heat up the MTJ for easier switching. The same as in STT, this method needs only one polarized current flowing into the MTJ. As shown in Figure 2.8, this switching mechanism involves applying a low current through STT to raise the MTJ temperature above the blocking temperature (T_b) of the antiferromagnetic layer associated to the storage layer, resulting in a hysteresis loop centered about zero. T_b depends mainly on the material composition (e.g. ~423K for IrMn and ~573K for PtMn). This method benefits from the advantages of both TAS and STT technologies, which achieves the best tradeoff among data reliability, power efficiency, speed and density. However, it still requires the supplementary time for cooling and power for heating, limiting its wide use in high-speed and low-power applications.

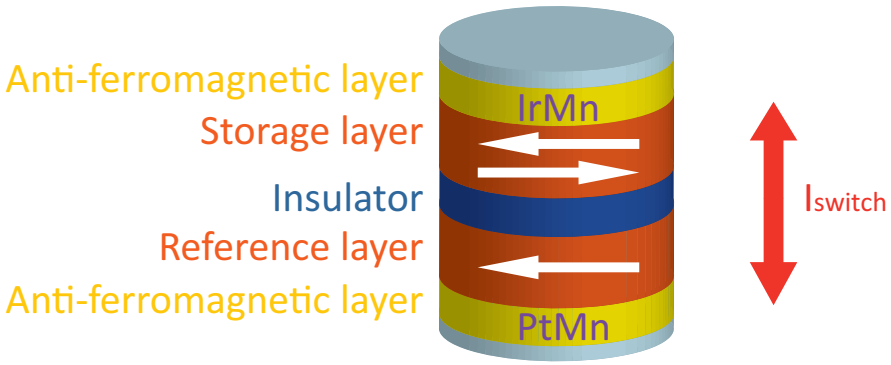


Figure 2.8: Thermally assisted spin transfer torque switching approach structure.

2.1.2.5 Spin Hall effect spin transfer torque (SHE+STT)

Spin hall effect (SHE) assisted STT switching has been experimentally demonstrated to overcome the incubation delay generated by STT switching method [35, 36]. The switching mechanism can be explained by the three-terminal SHE device composed of a typical STT-MTJ deposited on a heavy metal (e.g., tantalum) illustrated in Figure 2.9. Spin

accumulation on the lateral surfaces can be generated by injecting a charge current I_e into the heavy metal due to the spin-orbit interaction [37]. As a result, a spin-polarized current I_s along the direction orthogonal to both the charge current and electron spin is generated to pass through the MTJ which can assist the switching process. Thus, the writing operation can be realized by injecting a relatively low current I_{switch} into MTJ structure. The direction of spin current can be controlled by changing the direction of injected charge current I_e. Respectively, the state switching is determined by the charge current I_{switch}.

Compared with the STT switching approach, the SHE+STT switching method removes the undesirable incubation. In this approach, the writing and sensing operations is completely separated by the three terminals configuration. Therefore, low resistance can be realized for easier writing and high resistance can be realized for sensing. Moreover, the switching current can be reduced by nearly one order of magnitude compared with STT switching mechanism by optimizing the thickness of heavy metal layer. With these advantages, this approach features lower power, faster speed and better reliability. However, the scalability becomes a bottleneck for this approach due to the difficulty of integrating the three-terminal device into very large scale circuit which causes area efficiency degradation.

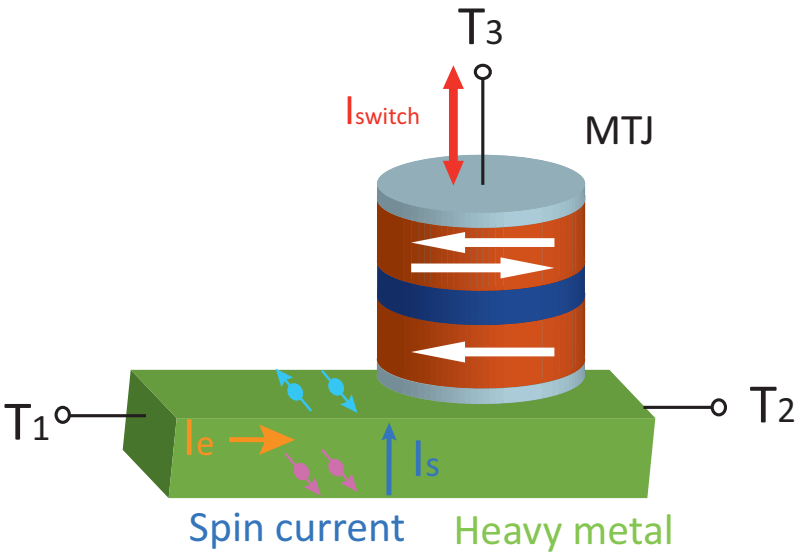


Figure 2.9: Spin Hall effect spin transfer torque switching approach.

The performance of the different switching approaches in terms of scalability, endurance, operation speed and power consumption are compared in details as demonstrated in Table 2.1. Among these switching mechanisms, STT is regarded as the most promising MRAM technology and attracted intense research attention. We will focus on investigating the reliability analysis of spintronic devices based on this switching approach in this

thesis.

Table 2.1: Comparison of performance for the different switching approaches [28]

Approaches	Scalability	Endurance (cycles)	Write time	Write Current
FIMS	Poor	10^{16}	Long (>10ns)	Very high (∼10mA)
TAS	Good	10^{12}	Very long (>20ns)	High (∼1mA)
STT	Very good	10^{16}	Short (<5ns)	Low (∼100uA)
TAS+STT	Best	10^{12}	Medium (<8ns)	Medium (∼100uA)
SHE+STT	Good	10^{12}	Best (<3ns)	Best (∼10uA)

2.2 Magnetic tunnel junction based memories and logic circuits

With the aforementioned features of MTJ, much research effort has been devoted to applying it in design of memories and specific logic functions. This section will briefly review some typical designs of MTJ based circuits.

2.2.1 Magnetic Random Access Memory

Cross point architecture was firstly proposed to realize MRAM [13, 38, 39]. As demonstrated in Figure 2.10, each MTJ is connected to the crossing points of two perpendicular arrays of parallel conducting rows and columns. To successfully program the memory cell, current pulses are sent through one line of each array and the MTJ at the crossing point of these two orthogonal lines can be switched with sufficient magnetic field (for FIMS) or current density (for STT). For reading operation, the resistance of the device between the two selected crossing lines can be sensed out, which represents the information stored in the MTJ. The cross-point architecture promises high-density integration, but it suffers from the sneak path issue and low access speed, limiting its wide application for fast and reliable reading [40].

Another more complex structure named as 1T1R was proposed to eliminate the unwanted current paths, which is one of the most widely used emerging Non-volatile array architectures [41]. As demonstrated in Figure 2.11, the elementary cell consists of one MTJ connected with one selection MOS transistor in series. The added transistor contributes to isolating the selected cell from others, removing the sneak path issue. The word line (WL) controls the gate of the transistor and the write current can be regulated

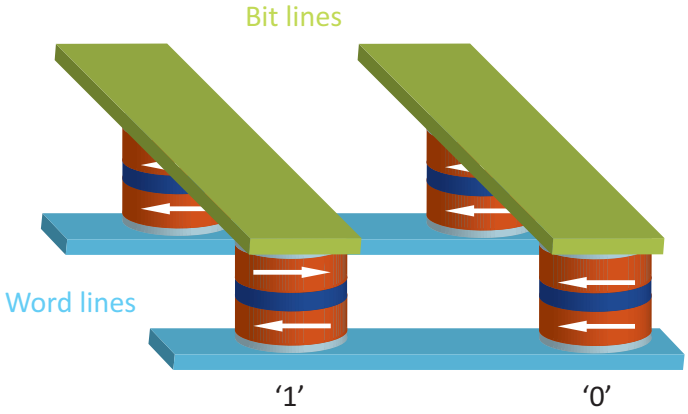


Figure 2.10: A schematic of the cross-point array. The selector is added in series with the MRAM cell at each cross-point.

by tuning the WL voltage. The bit line (BL) is connected to the drain through MTJ and the source line (SL) is connected to the source of the transistor, which serve as supply voltage according to the corresponding operation. This architecture promises fast access speed and better reliability for both writing and reading operations compared with the cross-point architecture. However, the density of 1T/1MTJ cell architecture is less due to the added transistor for each cell.

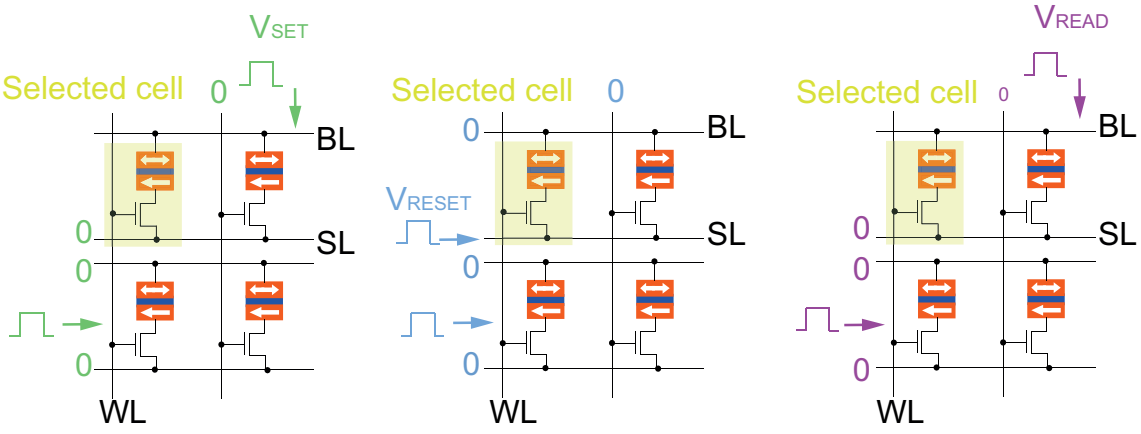


Figure 2.11: 1T/1MTJ memory cell architecture.

As data can be stored without extra power, there exists no static power in MTJ based non-volatile memories compared with the conventional memories. Thus, the endurance can also be improved as no stress voltage is necessarily applied across MTJ for maintaining its magnetization state at standby mode. Table 2.2 demonstrates the universal memory candidates which drive most of research and development [39, 42]. With comprehensive consideration, STT-MRAM is an ideal candidate for future memory which features low power consumption (no static power compared with mainstream RAMs) and fast operation

speed.

Table 2.2: Performance comparison of the universal memory candidates widely used and appeared in the last decade: F represents feature size of the lithography, the energy estimation is on the cell-level (not on the array-level), the endurance is signified by the writing cycles [39, 42].

Technology	**Mainstream Memories**				**Emerging Memories**			
	SRAM	*DRAM*	*NOR-flash*	*NAND-flash*	*STT-MRAM*	*PCRAM*	*RRAM*	*FeRAM*
Cell Area	$> 100F^2$	$6F^2$	$10F^2$	$4F^2$ (3D)	$6\sim50F^2$	$4\sim30F^2$	$4\sim12F^2$	$15\sim35F^2$
Multibit	1	1	2	3	1	2	2	1
Voltage	<1 V	<1 V	>10 V	>10 V	<1.5 V	<3 V	<3 V	∼1.8 V
Read time	∼1 ns	∼10 ns	∼50 ns	∼10 μs	<10 ns	<10 ns	<10 ns	<10 ns
Write time	∼1 ns	∼10 ns	10 μs−1 ms	100 μs−1 ms	<10 ns	∼50 ns	<10 ns	<5 ns
Retention	N/A	∼64 ms	>10 y	>10 y	>10 y	>10 y	>10 y	>10 y
Endurance	$> 10^{16}$	$> 10^{16}$	$> 10^5$	$> 10^4$	$> 10^{15}$	$> 10^9$	$10^6 \sim 10^{12}$	10^{13}
Write energy	∼fJ/bit	∼10fJ/bit	∼100pJ/bit	∼10fJ/bit	∼0.1pJ/bit	∼10pJ/bit	∼0.1 pJ/bit	∼10fJ/bit

2.2.2 Logic in Memory

The concept of logic in memory (LIM) was proposed early in 1960s [43] to reduce the power consumption and interconnection delay of the computing units. In the conventional Von-Neumann architecture, the memory and the logic circuits are spatially separated, leading to severe data-transfer traffic between them. In contrast, the memory cells are deposited over the logic circuits plane in the LIM architecture to eliminate this shortcoming. The distance between memory and logic circuits has been drastically shortened, resulting in faster speed and smaller power consumption on the interconnections. Since the appearance of spintronics devices, many researchers attempt to develop magnetic logic circuits. In order to maximize the advantage of the logic-in-memory architecture, it is necessary to implement a non-volatile memory that has a capability of short access time (sub 10 ns), infinite endurance, scalable write, and small dimension comparable to the employed CMOS technology [44]. For all of these purposes, MTJ switched by STT mechanism has become the unique available candidate [45]. Since the data has been already memorized into MTJ devices in the proposed LIM circuits, the supply voltage can be immediately cut off without data transmission into external non-volatile storage devices when the circuit changes to a standby mode. Moreover, the intrinsic long data retention time of MTJ enables instant on/off computing, namely the system can immediately continue to work after “waking up” from the “sleeping” mode. owing to these properties, power dissipation can be significantly reduced.

Figure 2.12 shows the general architecture of logic in memory based on STT-MRAM [46]. Note that STT-MRAM is deposited on the highest level of metal over CMOS transistors. It consists of three parts: a pre-charge sense amplifier (PCSA) circuit evaluates the logic result on the outputs, a write logic block programs the STT-MRAM cells and a logic data control block. Compared with the logic part, every bit of STT-MRAM costs a relatively high programming energy (∼ from 0.2 to 0.5 pJ/bit@40 nm) and low switching speed (∼ns), the logic data block contains a MOS logic tree and STT-MRAM in order to keep an area-power-efficient advantage. In this case, the logic volatile data can be driven by a high processing frequency, contrarily to analog non-volatile data, which should be changed with a relatively low frequency i.e. they are more critical data or quasi-constant for computing. Depending on the MOS state in the logic tree and the STT-MRAM element state, the discharge currents are different in both branches and the current sense amplifier latches opposite logic values on outputs.

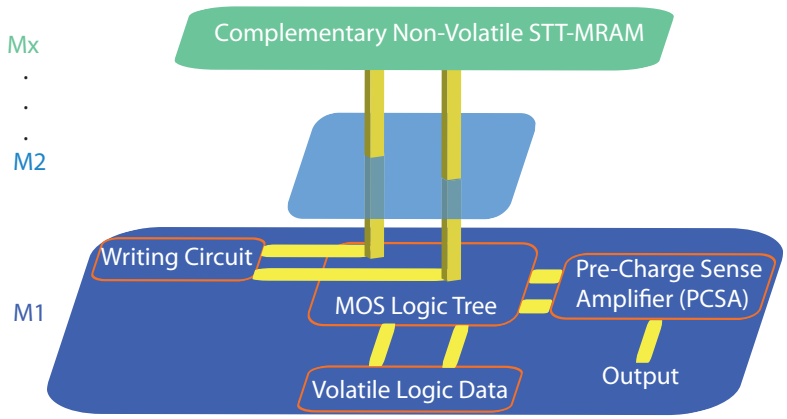


Figure 2.12: General architecture of logic in memory based on STT-MRAM: Mx represents the highest level of metal in CMOS technology.

Figure 2.13 demonstrates several typical logic gates and basic computing chip cell based on the general architecture proposed in [46, 47, 48, 49]. These circuits have been demonstrated to be advantageous in terms of area saving, energy efficiency and operation speed compared with the conventional CMOS implementations. It is remarkable that the PCSA structure as sensing circuit features perfect performance in terms of variability awareness and immunity to read disturbance. In this thesis, all of the circuits studied and designed will be based on this structure.

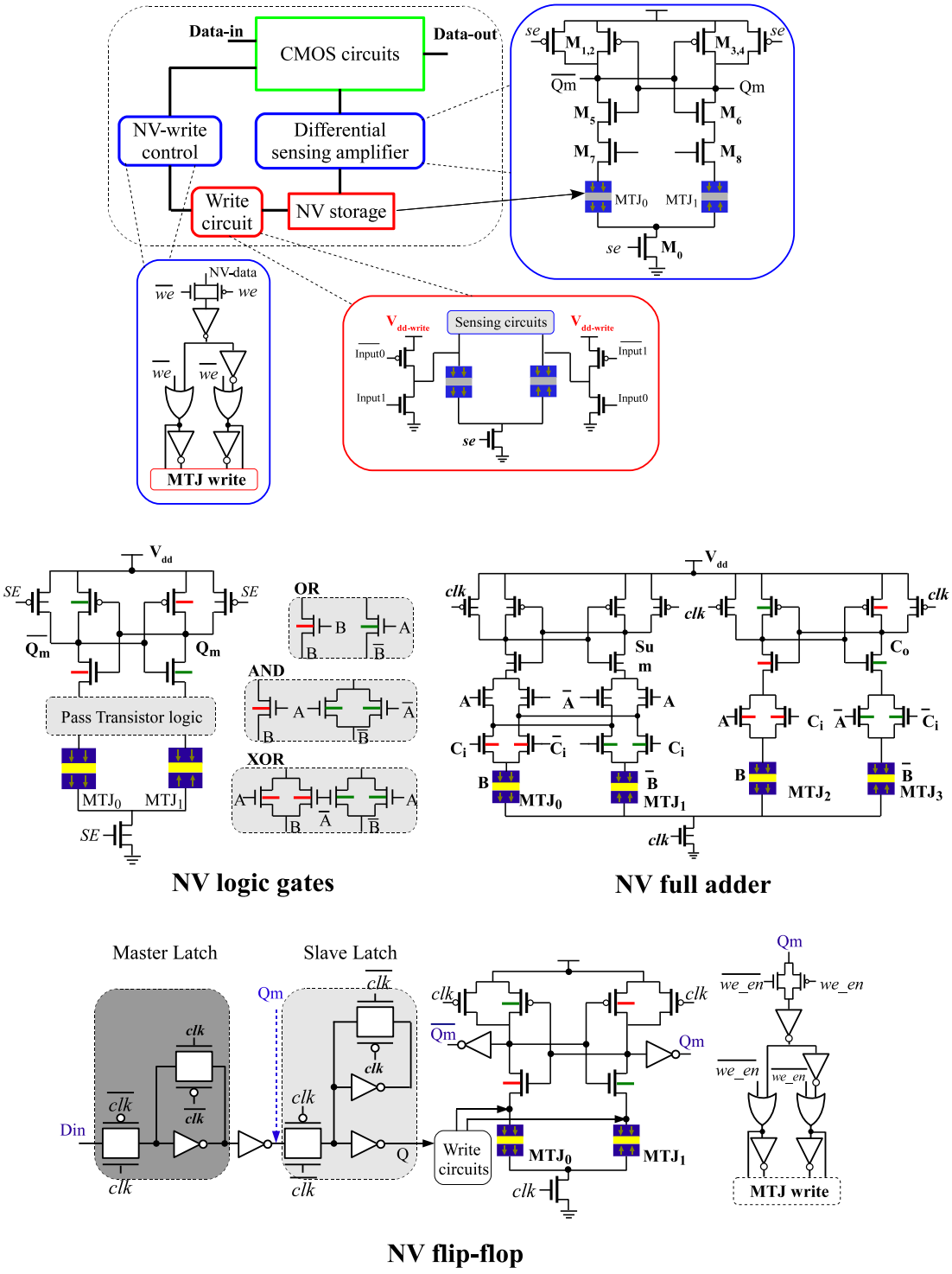


Figure 2.13: Typical MOS/MTJ NV-LIM circuits based on pre-charge sense amplifier structure: logic gates, full adder and flip-flop.

2.2.3 Other novel applications

Because of high sensitivity to magnetic field, MTJs have also been considered to work as magnetic field sensors [50]. Among the several MagnetoResistance (MR) sensor technologies available presently, MTJ based on MgO barriers are highlighted as the most competitive sensors aiming pT detection at room temperature [51]. Since the perpendicular anisotropy is significantly associated with thickness, the sensor response depends critically on the thickness of the sensing layer. The proposed sensors exhibit a large field sensitivity and a high linear field range of up to 600 Oe. In addition, the nano-scale size and simple structure of the sensors make them easy to integrate with complementary metal-oxide-semiconductor technology for nano-scale low power-consumption sensors. MTJ based MR sensors have proven to be a reliable tool in hard disk magnetic recording and biomedical device for magnetocardiography. However, noise problem limits the sensitivity of magnetic tunnel junction (MTJ) sensors for ultra-low magnetic field applications.

All-spin logic (ASL) has recently attracted much research interest because of its non-volatility, high density, lower device count, and good scalability [52]. ASL is considered as a promising post-CMOS device candidate for next generation of computing chip. An ASL device is mainly composed of input and output magnets connected by a channel medium (typically copper or graphene) as demonstrated in Figure 2.14 [53]. The logic operation can be realized by using spin injection, spin diffusion and STT switching in a lateral spin-valve (LSV) structure. The charge current flowing through the input magnet will generate spin polarized electrons which conserve the magnetization momentum. The injection and diffusion through the channel induce different electrochemical potential between parallel and anti-parallel states in the output magnet. Thus, the output magnetization orientation can be changed by the spin torque transferred by the sufficient spin current. ASL design stores information by utilizing spin direction of the magnets and communicates using pure spin current. Since no transistor is required for ASL applications and all the logic functions can be constructed with a minimal set of Boolean logic gates, ASL is generally thought to be a good post-CMOS candidate from energy efficient and scaling perspective. It has been demonstrated that ASL can potentially reduce the switching energy-delay product by a significant amount, but there are major challenges to be overcome. One is the room temperature demonstration of switching in multi-magnet networks interacting via spin currents [54]. The other is the introduction of high anisotropy magnetic materials into relevant experiments which can improve energy-delay. Issues such as current density and proper choice of channel materials also have to be carefully considered. The analog

nature of ASL communication can be efficiently coupled with median function to develop an architecture called Functionality Enhanced ASL (FEASL) to realize low-power, short delay and small area circuits. FEASL is especially suited for adder and multiplier circuits which are an integral part of arithmetic logic units (ALU). Moreover, it should be mentioned that ASL could also provide a natural implementation for Biomimetics systems with architectures that are radically different from the standard von-Neumann architecture.

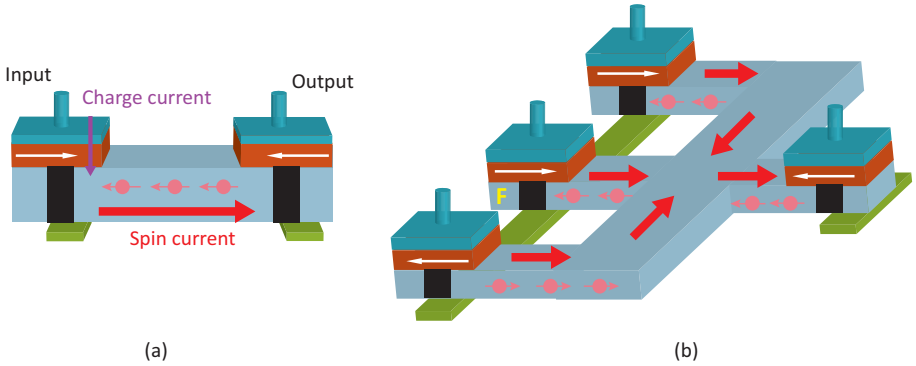


Figure 2.14: Implementation of ASL Boolean gates. Only the net spin polarization is shown for spin current. (a) Inverter. (b) NAND. "F" denotes a magnet with fixed magnetization direction.

2.3 Reliability analysis of MTJ device and MTJ based applications

Reliability is an important factor in design and operation of integrated circuits. Operational reliability is the ability of the memory and logic devices to operate reliably within their operational error tolerance given in their performance specifications [54]. The error rate of all nanoscale devices and circuits is a major concern. These errors arise from the difficulty of providing highly precise dimensional control needed to fabricate the devices and also from interference of the local environment. Error detection and correction schemes will need to be a central theme of any architecture and implementations that use nanoscale devices. With continuous scaling down of CMOS technology node, reliability becomes a critical challenge of ICs in deep sub-micron region in microelectronics applications. As for STT-MTJ of which the size is usually at the level of nm with few layers of atoms, reliability is especially challenging for wide commercialization of STT-MRAM. Reliability is defined as the ability of a circuit to conform to its specifications over a specified period of time under specified conditions [55]. The reliability issue of STT-MTJ device mainly contains process variation, stochastic switching, temperature fluctuation and dielectric breakdown.

The research on reliability issues of STT-MTJ will be reviewed in this section.

In the first experimental demonstration of STT-MRAM [41], the resistance of MTJs follow a statistical distribution (see Figure 2.15), with an approximate standard deviation σ of 4%. Even though an optimization solution using conventional MRAM production technologies has been mentioned which can suppress σ less than 1-2%, the process variation can never be removed. This indicates the existence of process variations during the device fabrication. Due to the limited process precision, many parameters are not identical with the initial target. As a result, the magnetic and electric properties are influenced, such as resistance, TMR ratio and switching delay. All of these variations may lead to functional errors during the MRAM operations. After that, many researchers have proposed various methods to model the process variations and analyze the influence on MTJ based hybrid circuits [56, 57, 58]. Some special experimental techniques have been applied in the MTJ fabrication to improve the device performance and reduce effect of process variations [59]. Meanwhile, variety of design strategies are proposed to improve the robustness of MTJ based circuit [60, 61, 62].

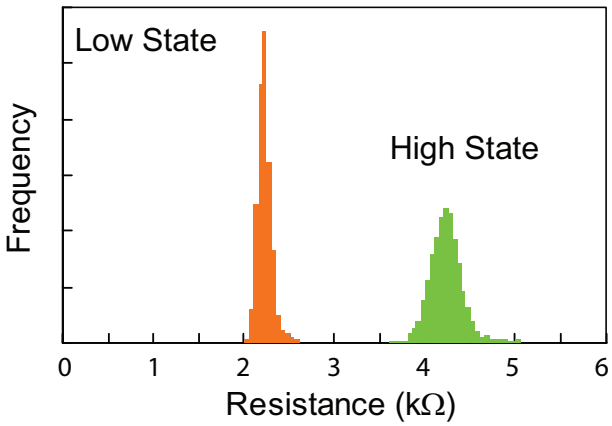


Figure 2.15: Low state resistance and high state resistance distributions of the 4kbit circuit with MTJ size of 120 x 170 nm as demonstrated in [41]. Bias voltage is kept at -0.1 V.

STT switching method has been demonstrated intrinsically stochastic [21]. As demonstrated in Figure 2.16, the switching behavior is not deterministic but follows a distribution. The reversal duration of STT writing mechanism can vary significantly from one event to the next, with a standard deviation almost as large as the average switching duration and sigmoidal distributions with exponential tails [63]. The switching success probability is a function of current flowing through MTJ and pulse duration. This is very different from the traditional electronic devices such as transistors and resistors. The stochastic behavior originates from the unavoidable thermal fluctuations of magnetiza-

tion which randomly interfere to activate or slow down magnetization reversal. After this observation, many other researchers have theoretically or experimentally verified this phenomenon [64, 65, 66, 67]. It can be concluded from the experimental measurements shown in Figure 2.16 that increasing the write current value or adding extensive margins on the driver pulse duration are the most efficient methods to avoid the writing failures. However, these may lead to significant power, speed and surface overhead, which is the provenance of the following two reliability issues.

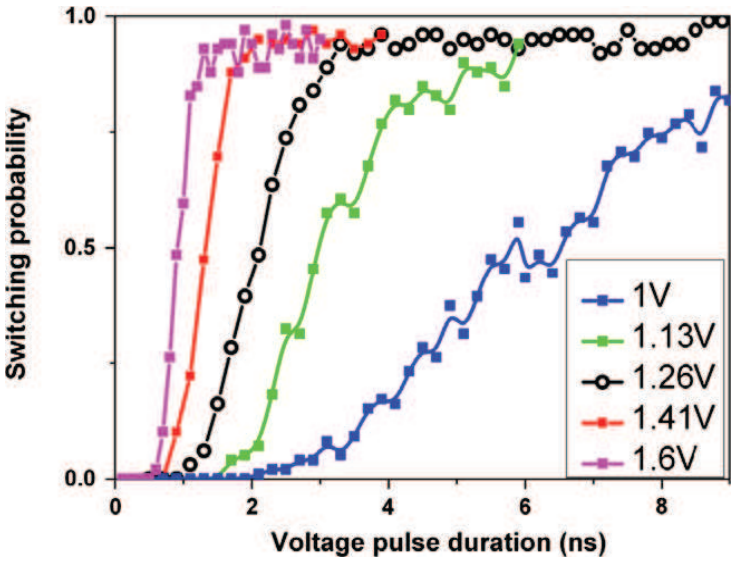


Figure 2.16: Experimental measurement of STT stochastic switching behaviors, the switching duration follows a certain distribution determined by the current and pulse duration.

With the fast development and intensive research attention, spintronic devices are used in a variety of different applications. Reliability is especially challenging in some special applications, for instance, automotive, military and aerospace applications which have extreme conditions of temperature. The self-heating effect of MTJ stack has been observed in [23] and investigated executing one-dimensional numerical simulations by solving the heat equation. Different from the TAS switching approach which heats up the MTJ by an external element, the MTJ can also be heat by itself due to Joule heating. Despite the great efforts devoted to technology optimization in the past years, a relatively high current density flowing through MTJ is always required by most of the switching mechanisms. This results in considerable self-heating effect which may cause functional errors of hybrid MTJ/CMOS circuits [68].

Moreover, the characteristics of ferromagnetic materials are very sensitive to environmental temperature, which has been already observed in many experiments as demon-

strated in Figure 2.17 [7, 8, 11, 12]. The ferromagnetic materials are very sensitive to thermal fluctuation. With different thermal conditions, the magnetic properties are totally different. The common goal of MTJ research always focuses on fabricating the higher TMR ratio at room temperature. Nevertheless, with the promising perspectives of MTJ based applications in the coming IoT era, the exact characteristics of MTJ in different thermal conditions should be carefully investigated and modeled for circuit designers. However, the existing models include either temperature dependence [69] or self-heating effect [68, 70].

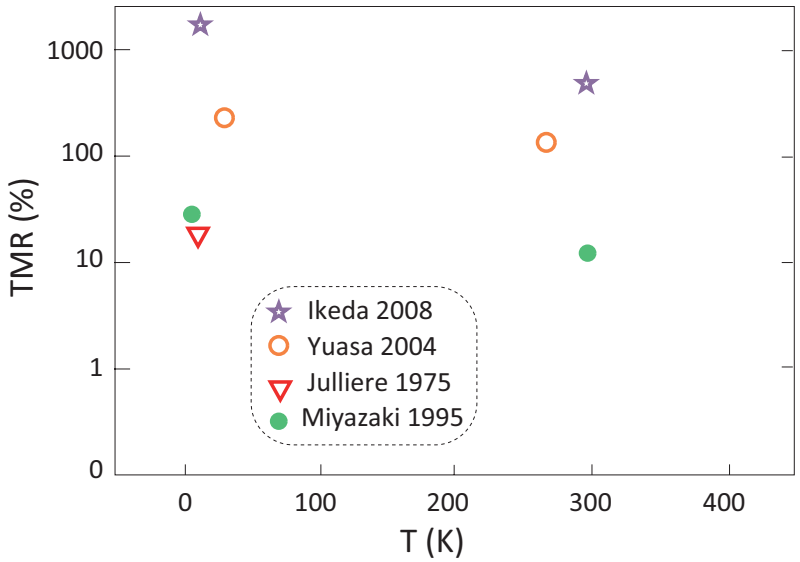


Figure 2.17: TMR ratio at different temperature in experimental measurements in [7, 8, 11, 12].

Dielectric breakdown is the most crucial reliability issue which determines the lifetime of device (transistors or MTJ). As MTJ is a memristive device and its resistance mainly comes from the oxide barrier, the voltage applied on MTJ is almost imposed on the insulator (AlxOy or MgO). With ultra-thin thickness (∼1 nm), the dielectric breakdown voltages are also scaling down, and it is necessary to avoid time dependent dielectric breakdown (TDDB) of the tunnel barrier caused by write operations [71]. Several experiments have been performed to demonstrate this phenomenon [72, 73, 74, 75, 76], others have been carried out to explore the mechanism behind the phenomenon and the factors which have impact on TDDB [24, 77, 78]. It has been found that TDDB is related to variety of factors, such as annealing temperature, oxide material purity, tunnel barrier thickness, stress voltage, temperature, stress duration, etc. Meanwhile, some models have been proposed to synthesize this phenomenon for circuit designers [79, 80].

Table 2.3 demonstrates the comparison of several recently published models in terms

of implementation, simulation time, and consideration of reliability issues. There is not yet a complete model comprising all the reliability issues.

Table 2.3: Comparison of different compact models of MTJ

Models	[81]	[82]	[68]	[80]
implementation	Verilog-A	SPICE	Verilog-A	SPICE
simulation time	Shorter	Longer	Shorter	Longer
Process variation	Yes	Yes	No	Yes
Stochastic switching	Yes	No	Yes	No
Temperature dependence	No	Yes	No	No
Self-heating	No	Yes	Yes	No
Dielectric breakdown	No	No	No	Yes

In summary, uncertainties in reliability can lead to performance, cost and time-to-market penalties. Functional failures may be induced by insufficient reliability margin which are costly to fix and damaging to reputation. Thus, it is necessary to identify and address these reliability issues at early design phase of MTJ based circuit. The requirement for more precise process technology is very important, intelligent careful designs which can tolerate these variations are also necessary. Most of the existing models only focuses on part of the reliability issues, failing to meet the increasing requirement for more accurate reliability analysis. For more realistic designs, a complete and precise model including the main reliability issues is urgently required by circuit designers.

2.4 Summary

This chapter mainly reviewed the state-of-the-art of MTJ device and reliability issues research. Firstly, the operation mechanism of MTJ and the origin of Spintronics were introduced. Then, we investigated the evolution of different switching approaches of MTJ which dominates its development. The advantages and drawbacks of each switching mechanism have been briefly analyzed. Comparing the performance of the different switching methods, we have found that STT is the most promising switching method for very large scale memory and computing chip. Thus, we will concentrate on the reliability analysis of STT-MTJ and its applications in logic circuits and memories.

In the aspect of MTJ based applications, we have studied the most widely used memory

architecture and recent logic circuit designs. Part of them will be used to investigate the reliability in the chapter 4.

Finally, the history of MTJ reliability analysis was reviewed. The reliability issues have been described and the recent works have been introduced. From the existing models of MTJ, we have found a through model including all of the four main reliability issues is urgently required for circuit designers. The chapter 3 will concentrate on the modeling of reliability issues.

Chapter 3

Compact modeling of reliability issues in STT-PMA-MTJ

Reliability of integrated circuits (ICs) has attracted intensive research interest since the appearance of IC. Uncertainties in reliability may lead to performance, cost, and time-to-market penalties. Functional failures can be caused by insufficient reliability margin, which are costly to fix [54]. These issues place difficult challenges on testing and reliability modeling. As a critical factor impacting the ICs quality, the reliability issues require significant research and development. With the fast technology scaling down, reliability analysis becomes more and more important in academic and industrial ICs community. For a successful design, it is essential to understand and control the failure mechanisms associated with new materials and structures. There is no exception for the promising memory candidate, i.e., magnetic tunnel junction (MTJ). With extremely small dimension of MTJ (with ultra-thin oxide barrier ∼1nm), the impact of all the reliability issues on MTJ based circuits performance are increasing. To improve the yield of MTJ based circuit, it is necessary to take into account the reliability issues during the early design phase. However, there exists no complete model which includes most of the reliability issues of MTJ. This chapter investigates the possible reliability issues in MTJ and then quantify them using mathematical equations. Consequently, a compact model is proposed for circuit designers to consider these reliability issues.

3.1 Perpendicular magnetic anisotropy (PMA) MTJ

Our model is based on the MTJ with perpendicular magnetic anisotropy (PMA) which was discovered by Ikeda in 2010 [20]. Its structure and switching behavior are displayed in Figure 3.1. The core of MTJ mainly consists of three layers: two ferromagnetic (FM)

layers separated by an oxide barrier. The resistance (R_p, R_{ap}) depends on the relative magnetization of the two FM layers (Parallel (P) or Anti-Parallel (AP)) [7]. The resistance difference is characterized by Tunnel Magnetoresistance Ratio TMR = (R_{ap}-R_p)/R_p [11]. With STT mechanism, MTJ changes between two states when a bidirectional current I is higher than the critical current I_{c0}. The main provenance of the reliability issues are also indicated.

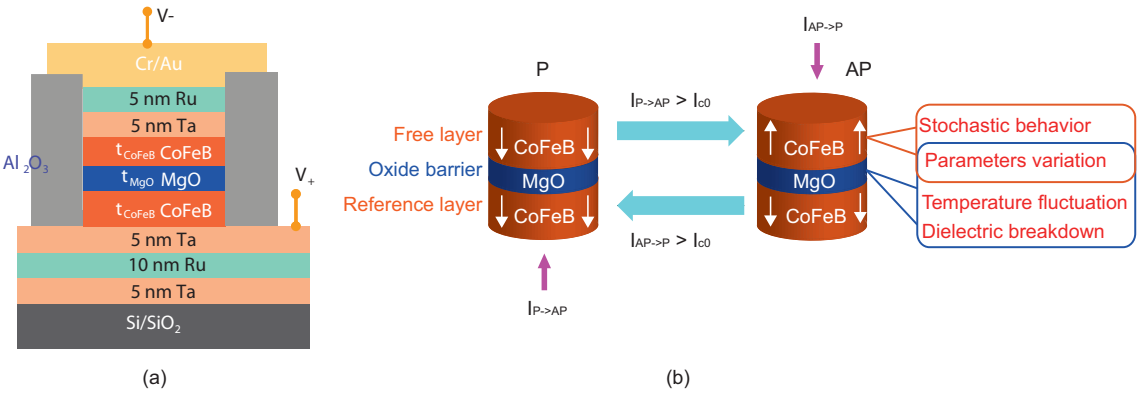


Figure 3.1: (a) Structure of PMA STT MTJ based on CoFeB/MgO stack. (b) Core of MTJ and switching mechanism.

Compared to the conventional MTJ with in-plane magnetic anisotropy as shown in the Chapter 2, new materials are used to form PMA-MTJ which features lower critical switching current, faster switching speed and higher thermal stability. All of these characteristics make it more promising for future logic and memory applications which require more compact area, lower switching current, higher TMR ratio, higher thermal stability and easier integration into existing mature semiconductor process. The following equations will theoretically demonstrate the provenance of these advantages. The barrier energy and critical current of STT switching in the materials with in-plane magnetic anisotropy can be expressed as:

$$E_i = \frac{\mu_0 M_s H_c V_{sl}}{2} \tag{3.1}$$

$$I_{c0} = \alpha \frac{\gamma e}{\mu_B g}(\mu_0 M_s)(H_{ext} \pm H_{ani} \pm \frac{H_d}{2})V_{sl} \tag{3.2}$$

where H_c is the coercive field, H_{ext} is the external field, H_{ani} is the in-plane uniaxial magnetic anisotropy field, H_d is the out-of-plane magnetic anisotropy induced by the demagnetization field, μ_0 is the permeability in the free space, M_s is the saturation magnetization, V_{sl} is the volume of the free layer, μ_B is the Bohr magneton, γ is the gyromagnetic ratio,

e is the electron charge. Whereas, the barrier energy and cirtical current in materilas with PMA are described as:

$$E_p = \frac{\mu_0 M_s H_k V_{sl}}{2} \tag{3.3}$$

$$I_{c0} = \alpha \frac{\gamma e}{\mu_B g} (\mu_0 M_s) H_k V_{sl} \tag{3.4}$$

where H_k is the perpendicular magnetic anisotropy field.

By comparing the equations (3.1) and (3.3), as H_k is higher than H_c, PMA allows obtaining relatively high barrier energy with a small size. By comparing (3.2) and (3.4), as H_k is much lower than H_d, the critical current for PMA materials can be significantly reduced.

Despite the great potential in PMA-MTJ, it also suffers from the reliability issues. In the following part, we will investigate the mechanisms behind the poor reliability.

3.2 Reliability issues of STT-PMA-MTJ

Compared with the conventional transistors, PMA-MTJ is a promising candidate for non-volatile memories thanks to its high speed, low power, infinite and easy integration with CMOS circuits. Equipped with these advantages, it solves some drawbacks of traditional transistor based ICs. With a relatively smaller size, the performance of PMA-MTJ based circuits are more severely impacted which behave poor reliability. In this section, four main reliability issues are discussed in details: process variation, stochastic switching, temperature fluctuation and dielectric breakdown.

3.2.1 Process variation

The large process variation is an intrinsic failure issue for PMA-MTJ which is based on the interfacial effects between ultra-thin films with few layers of atoms. This drawback severely limits the wide commercialization of STT-MRAM. In this part, an entire investigation of process variations origin during the nanofabrication of PMA-MTJ will be presented [83]. The nanofabrication of PMA-MTJ is based on standard back-end CMOS technology, but it needs additional specific processes. For instance, we need the growth of ultra-thin multilayers with a high quality tunnel barrier and precise crystallization matching of ferromagnetic layers to obtain giant TMR ratios and strong PMA. For this purpose, an ultra-high resolution sputtering machine is required. In case the process resolution

cannot meet the requirements, the large distribution of magnetic and electrical properties may occur, which will lead to poor performance of PMA-MTJ nanopillars. Figure 3.2 demonstrates the typical PMA-MTJ device fabrication process.

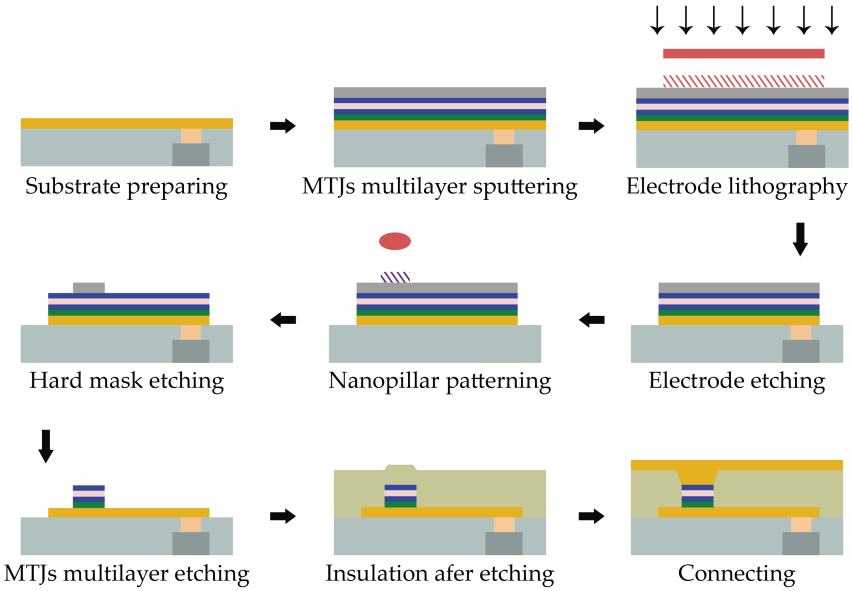


Figure 3.2: Typical flow of magnetic tunnel junction (MTJ) device fabrication, which mainly consists of stack deposition, patterning, etching dielectric encapsulation, and connecting.

Despite the optimization of deposition process in the past few years, PMA-MTJ based on interfacial effects still suffers from significant failure issues due to the variation of thickness and materials uniformity under 1 nm or with a few layers of atoms. Compared with in-plane magnetic anisotropy based MTJ, PMA-MTJ with interfacial magnetic anisotropy is more sensitive to the thickness variation, as it comes from the hybridization of atoms in the two interfaces MgO/CoFeB/Capping layer [84]. Both experiments and first-principles calculations have shown that the production of interfacial PMA matters with a certain thickness of ferromagnetic film and capping layer, which is usually a few atoms [20, 85]. For example, in order to trigger a MTJ's easy axis from in-plane to out-of-plane direction, thinner ferromagnetic film, i.e., less than 1.5 nm in the case of CoFeB/MgO structure, should be deposited [20]. In addition, other magnetic properties, including the offset field and thermal budgets, could be tunable by adjusting the relevant thickness of the individual layers in synthetic antiferromagnetic (SAF) structure, which is mainly because of a thickness-dependent co-tuning of exchange coupling of the SAF [86]. During the deposition process, uniformity or surface roughness is another critical parameter which has significant impact on the magnetic properties of PMA-MTJ . The imperfect process lead to the wide

distribution of the critical parameters of MTJ, e.g., anisotropy field H_k, Magnetization saturation M_s.

Besides, the following annealing treatment also influences strongly the device magnetic characteristics as well as electrical properties of MTJ nanopillars [20]. It has been demonstrated that the performance of MTJ improves monotonically while starting to increase the annealing parameters (such as Tex, H or annealing times). At certain condition, the best performance could be achieved, then decays when exceeding the optimum parameters [87]. Hence, we can divide annealing treatments into three stages: insufficient annealing, optimum annealing, and over-annealing. However, the optimum annealing parameters to get the best magnetic characteristics and the electrical properties do not coincide at the same time. As the magnetic curves shown in Figure 3.3, various values of magnetic parameters can be obtained by different annealing temperature and time. Reasonable annealing time (60 min, red curve) produced higher M_s and lower H_{sat}, which means stronger perpendicular magnetization in a typical PMA-MTJ structure of substrate/Ta/MgO/CoFeB/Ta. However, there are many uncertainties in most of the process which degrade the process perfection.

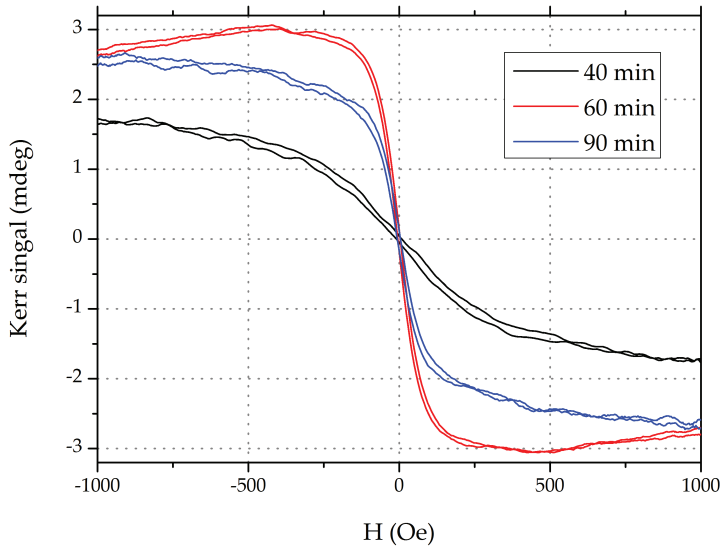


Figure 3.3: Magnetic curves (measured by NanoMOKE) of MTJ stacks annealed at different annealing times. The film stack deposited by magnetic sputtering processing are *exsitu* annealed at 300 °C for different annealing times (40, 60 and 90 min) with perpendicular H = 0.775 T in a high vacuum chamber.

After magnetic films deposition and annealing, etching also has an important impact on the quality of devices. For MTJ etching process, several issues may cause the failure:

sidewall redeposition, magnetic layer damage, or corrosion, and critical dimension (CD) control [88].

All of these imperfections generated in process are noted as process variation, which have strong impact on the magnetic and electrical properties of MTJ. Part of them are catastrophic, e.g., short, open, stuck at parallel state or anti-parallel state, while others can be masked or their impact can be mitigated at early design phase. The latter class will be quantified in the modeling part.

3.2.2 Stochastic switching behavior of MTJ

The switching of STT-MTJ has been revealed intrinsically stochastic due to thermal fluctuation of magnetization [21, 83]. As a result, the switching delay of MTJ is not a deterministic value but follows a statistical distribution. Because of this phenomenon, write errors might occur with insufficient writing current or short writing pulse, while unexpected switching may happen in sensing operation [61].

It has been well confirmed, both theoretically and experimentally that a spin-polarized current will deposit its spin-angular momentum into the magnetic system when passing through a small magnetic conductor. Consequently, it causes the magnetic moment to precess or even switch when the spin-current is sufficient [18]. Figure 3.4 illustrates the precession of magnetization under the influence of a spin current. Due to the thermal fluctuation of magnetization, the initial state of free layer magnetic moment (represented by θ) is different at each measurement. This leads to the stochastic reversal of free layer magnetization.

3.2.3 Temperature fluctuation behavior of MTJ

The same as for the conventional silicon based devices, environmental temperature has a significant impact on the performance of MTJ. The magnetic and electrical properties of MTJ can be easily influenced by operation temperature, which further leads to performance degradation and reliability issues of MTJ based memories and logic circuits. Moreover, despite the technology optimization in the past years, a relatively high current density flowing through MTJ is always required by most of the switching mechanisms. This results in considerable self-heating effect which may cause functional errors of hybrid MTJ/CMOS circuits. This section investigates the behaviors of MTJ under different temperature conditions and self-heating effect.

Several magnetic properties of MTJ are sensitive to temperature fluctuation, e.g.

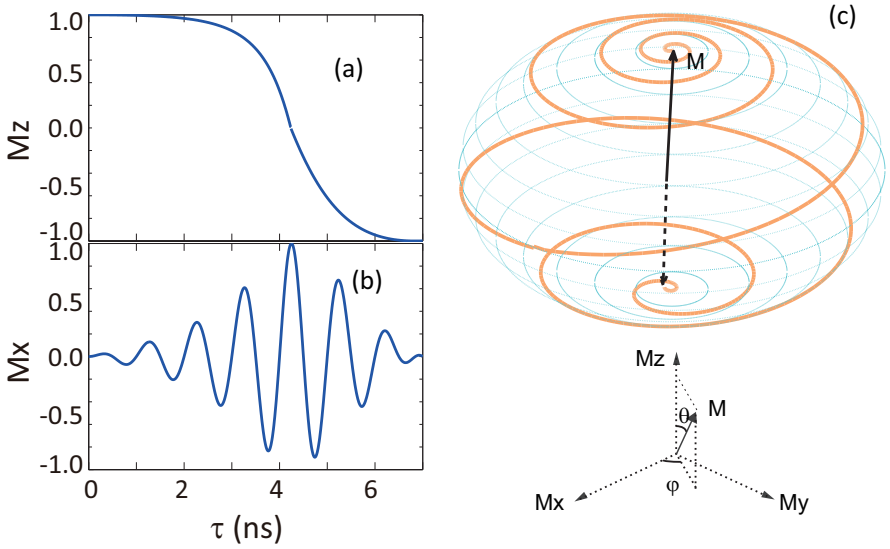


Figure 3.4: The precession of magnetization under the influence of a spin current: Time dependence of (a) M_z and (b) M_x, (c) The reversal process of magnetic moment. θ and ϕ represent the initial state of free layer magnetic moment. For PMA-MTJ, the switching behavior is mainly dependent on initial value of θ.

anisotropy field H_k, magnetization of ferromagnetic layers M_s [89, 90], tunneling magnetoresistance ratio (TMR). This leads to the unsteadiness of electrical properties of MTJ such as MTJ resistance R_P and R_{AP}, thermal stability factor ζ, critical switching current I_{c0}, as well as switching delay τ and further results in operational failures. Thus, as one of the major causes of stochastic STT switching, the environmental temperature also has an important impact on data retention [91].

Furthermore, most of the STT switching operations require a high current density flowing through the MTJ [16], which generates temperature increase due to Joule heating [92]. Therefore, a thorough study of high-temperature behaviors of MTJ is always required for reliability aware design of MTJ/CMOS circuits.

3.2.4 Dielectric breakdown

Oxide barrier breakdown represents one of the main reliability issues for advanced semiconductor memory technology. The effect of this issue which is catastrophic and nonreciprocal, which determines the lifetime of devices. In hybrid MTJ/CMOS design, the MTJ resistance must be configured comparable to the resistance of the selected transistor [73]. With shrinking of the MTJ size, a thinner tunnel barrier is essential to decrease the resistance so that a CMOS compatible design can be realized. In addition, a high current flowing through the MTJ is needed by the STT switching operations, which results in high

voltage across MTJ [92]. Consequently, dielectric breakdown becomes a key issue for the integration of STT-MRAM.

Several breakdown modes exist with large voltages applied to MTJs which cause reliability concerns. Figure 3.5 displays the three key breakdown mechanisms. Among the three, dielectric breakdown is the most detrimental effect on an MTJ bit and is one of the only unrecoverable hard breakdown faults [77]. The dielectric breakdown is induced by the intrinsic breakdown field of the insulating material (typically around 10 MV/cm), leading to dramatic drop of the resistance and TMR of the junction. Pinhole breakdown is caused by a series of minute conductive paths existing between the electrodes through the barrier. This is typically a metallic short or localized high-tunneling current region due to inhomogeneous oxidation of MgO or large surface roughness. As a result, TMR may be reduced through two mechanisms: The pinhole conduction paths run parallel to the spin-filtered tunneling path through MgO, leading to the majority of the current flowing through the non-spin polarized pinhole path and thus the overall spin-dependent transport is greatly reduced; Secondly, in case sections of the reference layer in contact and ferromagnetically coupled to the free layer, MTJ will no longer switch independently and the fully P and AP states will never be obtained. In the case of shunt breakdown, a very thin metallic particle or film surrounds the barrier area and creates a non-spin polarized parallel current path. This shunt is typically created during the barrier etching step during patterning where etched materials have a chance of redepositing onto the sidewalls. Because sidewall redeposition can be as thin as just a few monolayers, there is a chance that high current density may damage it and return the MTJ to a relatively normal state. Since these defects are detrimental to MTJ performance, it is crucial to understand their prevalence in patterned devices. Intentional application of large voltages can probe the quality of the MgO barrier and patterning process.

All of the reliability issues investigated in this section will be quantified by physical models and inserted into the final compact model in the following part.

3.3 Physical models of PMA-MTJ

This section introduces the physical models which describe the basic characteristics of MTJ as the first step of MTJ compact modeling. It includes the tunnel barrier resistance model, bias-voltage-dependent TMR model, spin polarization factor model and STT switching dynamic model.

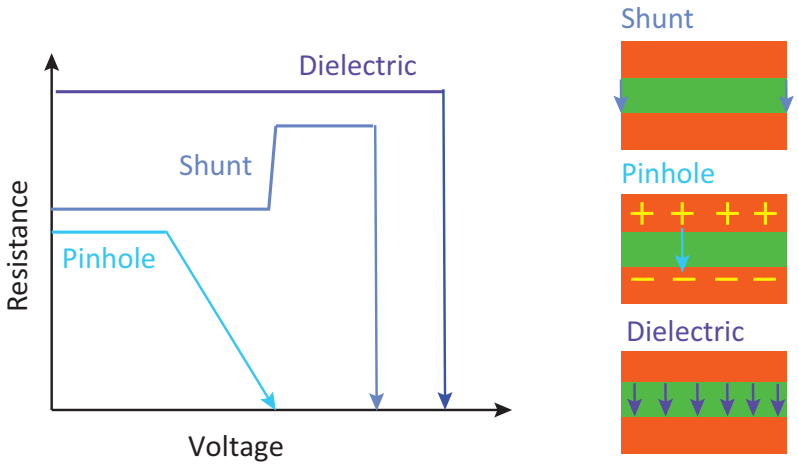


Figure 3.5: Three main breakdown mechanisms for MTJ barriers: Pinhole and shunt are soft-breakdown mechanisms that are typically lower than the intrinsic dielectric breakdown voltage [71].

3.3.1 Tunnel barrier resistance model

The physical model of the MTJ conductance was proposed by Brinkman in 1970 [93]. The conductance value is bias voltage dependent and is mainly determined by the thickness of oxide barrier and the interfacial effect between oxide barrier and the ferromagnetic layers:

$$\frac{G(V)}{G(0)} = 1 - \frac{A_0 \Delta\Phi}{16\bar{\varphi}^{3/2}} eV + \frac{9{A_0}^2}{128\bar{\varphi}} (eV)^2 \tag{3.5}$$

$$G(0) = 3.16 \cdot 10^{10} \cdot \bar{\varphi}^{1/2} \frac{exp(-coef \cdot t_{ox} \cdot \bar{\varphi}^{1/2})}{t_{ox}} \tag{3.6}$$

$$A_0 = \frac{4 \cdot (2m)^{1/2} t_{ox}}{3\bar{h}} \tag{3.7}$$

where t_{ox} is the thickness of oxide barrier $\bar{\varphi}$ is the average potential barrier height of MgO (0.4eV), coef=$1.025 nm^{-1} eV^{-1/2}$ is a fitting parameter , V is the bias voltage applied on MTJ, and h is Planck's constant. Considering the oxide barrier is symmetric, $\Delta\phi$ is equal to 0. In order to integrate this model into our compact model, simplified equations obtained from the above equations are employed to calculate the parallel state resistance of the CoFeB/MgO PMA MTJ [81]:

$$R_0 = \frac{t_{ox}}{(F \cdot \bar{\varphi}^{1/2} \cdot Area)} \cdot exp(coef \cdot t_{ox} \cdot \bar{\varphi}^{1/2}) \tag{3.8}$$

$$R_V = \frac{R_0}{1 + \frac{t_{ox}^2 e^2 m}{4\bar{h}^2 \bar{\varphi}} V^2} \tag{3.9}$$

where $Area$ is the MTJ area, F is a fitting parameter corresponding to t_{ox} and RA product which depends on the material composition of the three thin layer [11, 93]. For instance, if RA is defined as 10 $\Omega\mu m^2$ which gives F = 332.2. This model was based originally on MTJ using amorphous AlxOy, it has been demonstrated to be also suitable for the MgO oxide barrier based MTJ. Meanwhile, it is noteworthy that there is no evident dependence between the resistance at parallel state and bias voltage for the most advanced MgO based MTJ. Thus, the resistance in the parallel state R_p is defined equal to $R(0)$.

3.3.2 Bias-voltage-dependent TMR model

The value of TMR ratio is a key parameter determining the sensing operation performance of MTJ based memory and logic circuits. It has been well confirmed that the TMR ratio is highly dependent on the bias voltage imposed on MTJ. The strong dependence can be described as follows [81]:

$$TMR(V) = TMR(0) \cdot (1 + \frac{V^2}{V_h^2})^{-1} \tag{3.10}$$

$$R_{ap} = R_p \cdot (1 + TMR(V)) \tag{3.11}$$

where real $TMR(V)$ is the real value of TMR ratio depending on bias voltage, $TMR(0)$ is the TMR ratio with zero bias voltage, V_h is the bias voltage as $TMR(V) = 0.5TMR(0)$. Figure 3.6 displays the dependence of resistance on bias voltage.

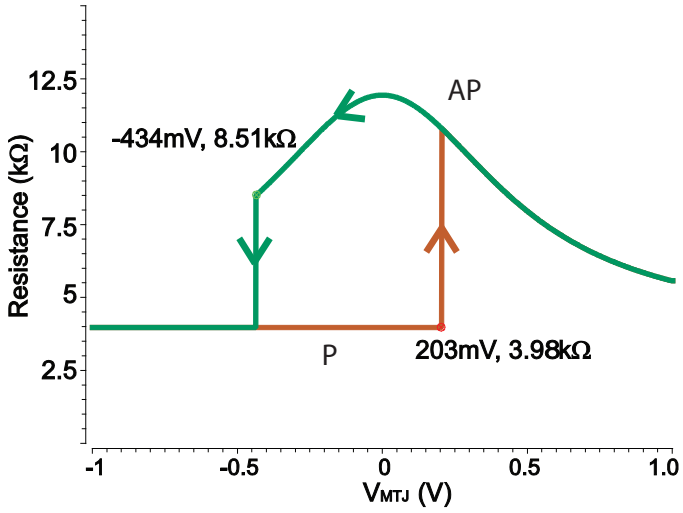


Figure 3.6: The TMR ratio of MTJ is dependent on the bias voltage.

3.3.3 Model of static behavior

The static behavior of PMA-MTJ signifies the calculation of the threshold value of switching current. Critical switching current is another important prameter of PMA-MTJ, which mainly determines the performance of writing operation of MTJ based memory and logic circuits. The calculation of critical current I_{c0} was expressed by equation (3.4) in the first section of this chapter. In this compact model, the spin accumulation effects are neglected and the spin polarization efficiency factor g is firstly obtained with the following equation to describe the asymmetric current case [94]. It shows great consistency with the experimental results illustrated in [20]:

$$g = g_{sv} \pm g_{Tunnel} \tag{3.12}$$

where the sign depends on the free-layer alignment. g_{sv} and g_{Tunnel} are respectively the spin polarization efficiency in a spin valve and tunnel junction nanopillars. They are both predicted by Slonczewski in [16],

$$g_{sv} = [-4 + (P^{-1/2} + P^{1/2})^3(3 + cos\theta)/4]^{-1} \tag{3.13}$$

$$g_{Tunnel} = (P/2)/(1 + P^2 cos\theta) \tag{3.14}$$

where P is the spin polarization percentage of the tunnel current, θ is the angle between the magnetization of the free and the pinned layers. Furthermore, more recent experimental progress of IBM shows that an MTJ involving symmetric electrodes provides a single spin polarization efficiency factor g for both state change processes (anti-parallel state to parallel state process or parallel state to anti-parallel state process) of MTJ [43], which allows the same critical current for both parallel and anti-parallel states. In this mechanism, g is only related to TMR ratio and can be calculated by the following equation:

$$g = \frac{\sqrt{TMR(TMR+2)}}{2(TMR+1)} \tag{3.15}$$

This simplified equation can be easily integrated in our model to calculate the critical current.

3.3.4 STT switching dynamic model

The dynamic model is mainly composed of calculating the average switching delay τ_{sw} (with 50% of switching probability). Depending on the magnitude of switching current,

the dynamic behavior of MTJ can be divided into two regimes [91]: Sun model (I > I_{c0}) [95] and Neel-brown model (I < $0.8I_{c0}$) [96]. The former is also called precessional switching which addresses fast switching (until sub 3ns) but consumes more energy with high current density. Reversely, the latter consumes less energy with low current density but leads to a slower switching which is called thermally-assisted switching. The two regimes are derived from the Landau-Lifshitz-Gilbert equation [91]. τ_{sw} can be calculated as follows [95, 96]:

$$\tau_{sw} = \tau_0 \cdot exp[\frac{\Phi_b}{k_B T}(1 - \frac{I}{I_{c0}})], \quad when\ I < 0.8I_{c0} \tag{3.16}$$

$$\frac{1}{\tau_{sw}} = [\frac{2}{C + ln(\frac{\pi^2 \zeta}{4})}]\frac{\mu_B P_{ref}(I - I_{c0})}{em_m(1 + P_{ref}P_{free})}, \quad when\ I > I_{c0} \tag{3.17}$$

where τ_0 is the attempt period, T is the temperature, k_B is the Boltzmann constant, C is the Euler's constant, ζ is the thermal stability factor, m_m is the magnetization moment, P is the tunneling spin polarizations. Usually a high current (I > I_{c0}) is applied to guarantee fast writing in memory. Meanwhile, MTJ can also be switched erroneously by relatively low current (I < $0.8I_{c0}$) during a long period of reading operation, which determines the data retention time. The switching behavior with $0.8I_{c0} \leq I \leq I_{c0}$ is very complex and there is no clear physical picture due to the competing effects of spin transfer torque and thermal fluctuation, which was described in the relative literatures [91, 95, 96]. As a result, there is no confirmed mathematical equation to allow electrical modeling.

3.4 Physical models of reliability issues in MTJ

The physical models introduced in the previous section will constitute an ideal model which is just appropriate to the experiment of [20]. For instance, if the environmental temperature or experimental equipment are changed, the model will not be fit. Thus, it is essential to add the reliability issues into the model to make it more realistic and provide the opportunity to predict the possible failures to the users. This section will introduce the models of reliability issues.

3.4.1 Process variation

The major sources of process variations of MTJ arise from two parts: (1) Variable geometrical parameters due to surface roughness and inherent film variations (cross-sectional area A, thickness of oxide barrier t_{ox}, and thickness of free layer t_{sl}); (2) Inexact magnetic

properties due to inhomogeneity of materials induced by imperfect process (Anisotropy field H_k, magnetization saturation M_s). These variations have severe impact on electrical properties of MTJ (Resistance Rp and Rap, TMR ratio, critical current I_{c0} and switching delay τ) and further lead to performance degradation.

The prameters variation are usually considered to follow approximately gaussian distribution [56]. In our model, the process variations are integrated by using the random functions and statistical block, which are provided by Verilog-A under Cadence environment. For instance, $\$rdist_uniform$ generates a uniform distribution in a limited range and $\$rdist_normal$ generates a normal distribution with fixed mean value and standard deviation. The users are free to choose different types of statistical distributions for different parameters (t_{sl}, t_{ox}, TMR).

3.4.2 Stochastic switching

The thermal fluctuation of environment introduces the randomness in the switching process (stochastic switching) [21]. For I > I_{c0}, the switching probability can be described as follows [97]:

$$P_{sw} = exp\{-4 \cdot f \cdot \zeta \cdot exp[-\frac{2 \cdot (t_{pulse} - delay)}{\tau_{sw}}]\} \tag{3.18}$$

$$f = (\frac{2}{1 - \frac{I_{c0}}{I}})^{(\frac{-2}{1+\frac{I}{I_{c0}}})} \tag{3.19}$$

where t_{pulse} is the voltage pulse width, $delay$ is a fitting parameter. As the experimental results have shown that the distribution of switching delay is nearly identical to gaussian distribution, the stochastic switching is also integrated into the model by using the random functions in Verilog-A language [98]. The users are free to reconfigure the simulation conditions by choosing different types of statistical distributions for switching delay τ_{sw}). The variables used in the model are adjusted to achieve good agreement with the equation (3.18). Figure 3.7 demonstrates the switching probability as a function of pulse width. It is obvious that the model result fits well the theoretical values.

3.4.3 Temperature fluctuation behavior of MTJ

Several magnetic properties of MTJ are sensitive to temperature fluctuation, e.g. anisotropy field, magnetization of ferromagnetic layers [89, 90]. This leads to the unsteadiness of electrical properties of MTJ such as tunneling magnetoresistance ratio (TMR), thermal stability factor ζ, critical switching current I_{c0}, as well as switching delay τ and further

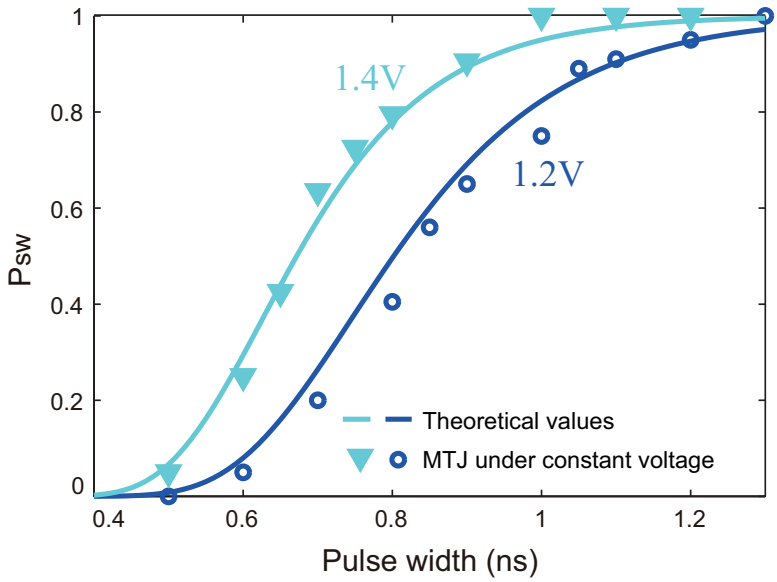


Figure 3.7: Switching probability P_{sw} as a function of pulse width: the lines are theoretical values plotted from (3.18) and the markers are statistical results from 1000 times of Monte Carlo simulation under Cadence.

results in operational failures. As one of the major causes of stochastic STT switching, the initial temperature variation also has an important impact on data retention [91]. Furthermore, most of the STT switching operations require a high current density flowing through the MTJ [16], which generates temperature increase due to Joule heating [92]. This part proposes the models of temperature sensitive parameters and the self-heating effect. As typical industrial temperature range is from -40 °C to 125 °C, the simulations in this work will consider temperatures between 233K and 400K.

3.4.3.1 Models of temperature sensitive parameters

This section concentrates on the impact of temperature on tunneling magnetoresistance ratio (TMR), thermal stability factor ζ, critical switching current I_{c0} and switching delay τ.

a. Temperature dependence of TMR

Experimental results show that resistance at antiparallel state reduces faster with temperature increase than that at parallel state, which originates from the degradation of TMR [99]:

$$TMR(T) = \frac{TMR_0 + 1}{1 + 2Q \cdot \beta_{AP} \cdot ln(\frac{k_B T}{E_c})} - 1 \tag{3.20}$$

where TMR_0 is at zero temperature, E_c is the magnon cutoff energy, Q describes the probability of a magnon involved in the tunneling process, $\beta_{AP} = Sk_BT/E_m$, S is the spin parameter, while E_m is related to the Curie temperature T_C of the ferromagnetic electrodes $E_m = 3k_BT_C/S + 1$. In addition, TMR is also dependent on bias voltage [81]:

$$TMR(V) = TMR(0,T) \cdot (1 + \frac{V^2}{V_h^2})^{-1} \tag{3.21}$$

where $TMR(0,T)$ is at zero bias, V_h is a voltage for which TMR becomes half of $TMR(0,T)$. Thus, a complete model of TMR can be deduced:

$$TMR(V,T) = TMR(T) \cdot (1 + \frac{V^2}{V_h^2})^{-1} \tag{3.22}$$

b. Temperature dependence of thermal stability factor ζ and data retention

Thermal stability factor ζ is often used to quantify the reliable data retention of magnetic data storage [100] and its value should be as large as possible to ensure a low reading failure rate. It can be calculated as follows:

$$\zeta = \frac{E_p}{k_BT} \tag{3.23}$$

The impact of reading operation on the required ζ while keeping an acceptable failure rate of MTJ based memory can be expressed as follows [61]:

$$F_{chip} = 1 - exp[-N\frac{\tau_r}{\tau_{r0}}exp(-\zeta(1 - \frac{I_R}{I_{c0}}))] \tag{3.24}$$

where F_{chip} is the switching error rate due to the cell read current I_R, N is the number of bits per word in the memory array, τ_{r0} is the attempt period = 1 ns and τ_r is accumulated read duration. Figure 3.8 demonstrates that effective anisotropy filed H_k and saturation magnetization M_s decrease with increasing temperature which show good agreement with experimental data in [90]. Thus, critical switching current Ic0 has strong temperature dependence. This results in the deep temperature dependence of ζ as shown in Figure 3.9. The chip failure rate of 8 bit per word with different reading duration ratio in data retention (1% or 10%) and different ratio of read current/critical current (1/5 and 1/15) is also presented in Figure 3.9. We can conclude that high temperature reduces read duration and the read current.

Temperature dependence of switching delay τ

Figure 3.10 depicts the voltage needed for a fixed average switching delay. It can be observed that the critical current and the average switching delay are reduced by increasing

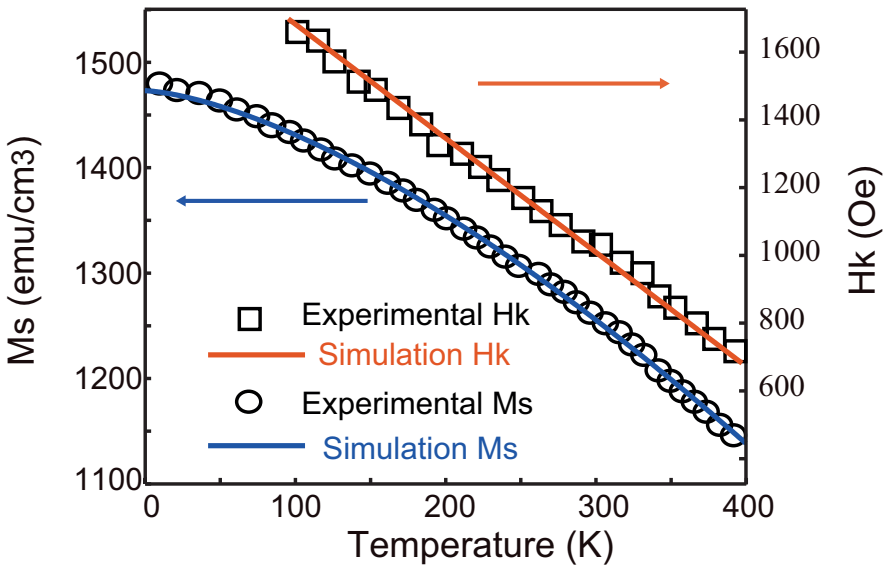


Figure 3.8: Temperature dependence of effective anisotropy field H_k and saturation magnetization M_s.

temperature. As a result, high temperature enlarges the writing margin and reduces the reading margin, which should be taken into account at the phase of circuit design.

3.4.3.2 Temperature fluctuation due to Joule heating

In spite of optimization in the past several years, a large current density of several MA/cm2 is always needed for current-induced magnetization switching [92], which heats up the MTJ due to Joule heating. The equations presented above all demonstrate that thermal fluctuation affects deeply the characteristics of MTJ. Thus, it is necessary to investigate the temperature increase due to Joule heating effect during current pulses. The self-heating effect can be described by [23, 98]:

$$T_{heat} = T_0 + \frac{V_s \cdot j}{\lambda / t_{ox}} \cdot [1 - exp(-\frac{D_{heat}}{\tau_{th}})] \tag{3.25}$$

$$T_{cool} = T_0 + (T_{heat} - T_0) \cdot exp(-\frac{D_{cool}}{\tau_{th}}) \tag{3.26}$$

$$\tau_{th} = \frac{C_v thick_s}{\lambda / t_{ox}} \tag{3.27}$$

where T_{heat} and T_{cool} represent respectively the temperature increase during current pulses and decrease with no current, T_0 is room temperature, V_s is stress voltage, j is current density, λ is thermal conductivity, D_{heat} is the heating pulse duration and D_{cool} is the

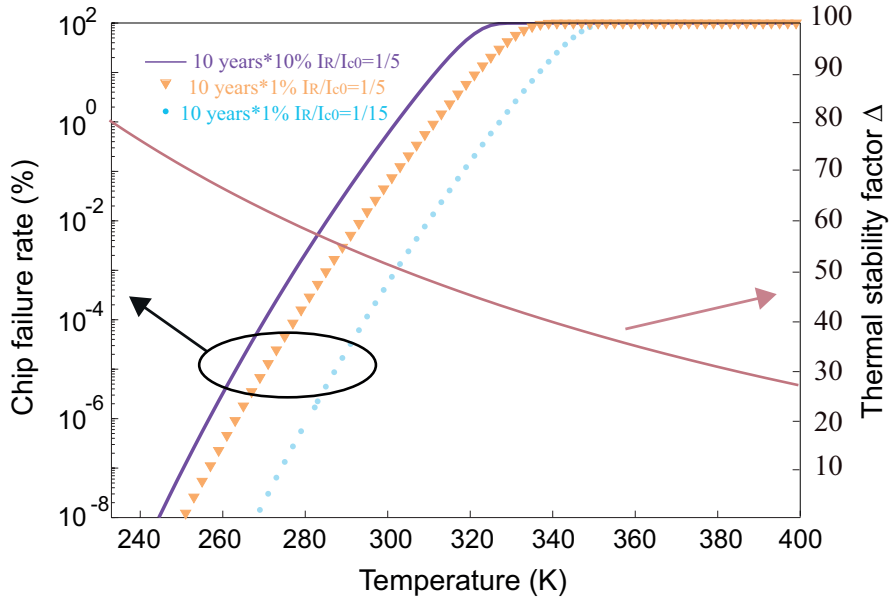


Figure 3.9: Temperature dependence of thermal stability factor and of MTJ based chip failure rate with 8 bits per word, different reading duration and different reading current.

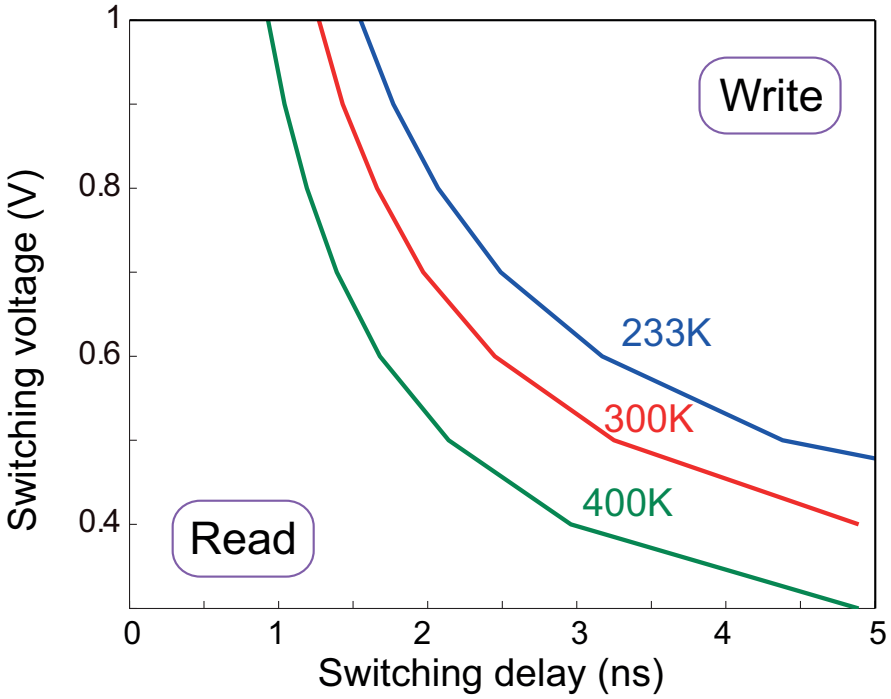


Figure 3.10: Switching voltage versus average switching time at different temperature conditions.

cooling duration, C_v is heat capacity per unit volume and $thick_s$ is total thickness of MTJ, τ_{th} is the characteristic heating/cooling time.

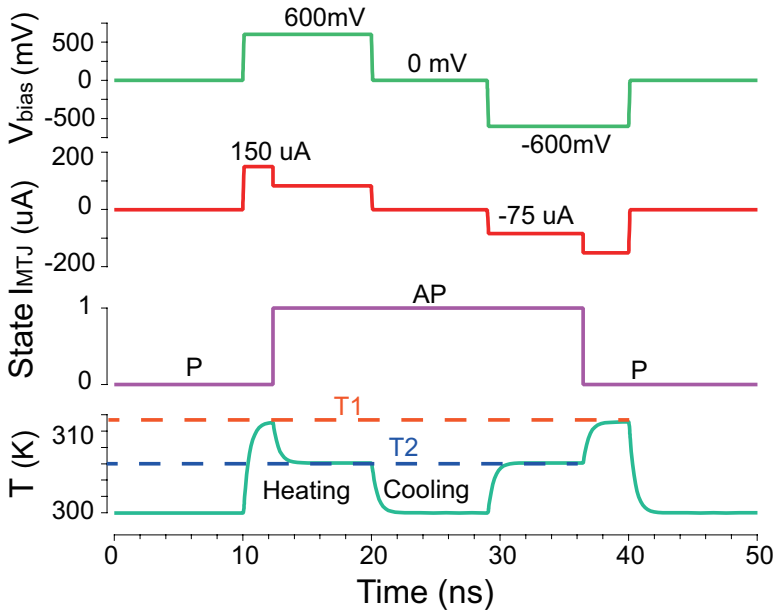


Figure 3.11: Temperature evaluation of MTJ during current pulses.

Figure 3.11 shows the temperature evaluation of MTJ. During a current pulse, temperature increases to T1 and then decreases to T2 after switching from P to AP. Temperature saturates at a maximum value T1 or T2 corresponding to the MTJ state. In this model, the impact of ambient heating due to thermal neighborhood (e.g., self-heating effect of CMOS devices) can be taken into account (by regulating the initial temperature T according to the requirements of designers).

3.4.4 Dielectric breakdown

This part presents the models involving dielectric breakdown phenomenon, which can be used to predict the breakdown voltage, time to failure and breakdown probability. With this model, the circuit designers can evaluate the aging effect and optimize the endurance.

3.4.4.1 Breakdown voltage

Dielectric breakdown of MTJ can be induced by a critical electric field applied across MgO barrier. Consequently, the MTJ resistance decreases abruptly ($\sim$10Ω in [101]) due to the formation of microscopic shorts in the barrier [72].

In the high-quality devices, the breakdown field is intrinsic to the barrier structure and independent of process conditions, defects or other uncontrollable variables [75]. The

breakdown voltage V_b can be calculated by:

$$V_b = E_b \cdot t_{ox} + V_{off} \tag{3.28}$$

where the value of E_b is obtained from the experimental data in [72], t_{ox} is the thickness of MgO, V_{off} is a fitting parameter. V_b for different MTJ configuration (P or AP) and polarity (positive or negative) versus MgO thickness is displayed in Figure 3.12. It is shown that V_b is different for positive and negative stress voltage, as well as for P and AP configuration, which matches exactly the experimental data presented in [72, 102]. In this model, the behavior of MTJ after breakdown is with ultra low resistance and fixed state, leading to permanent functional failure.

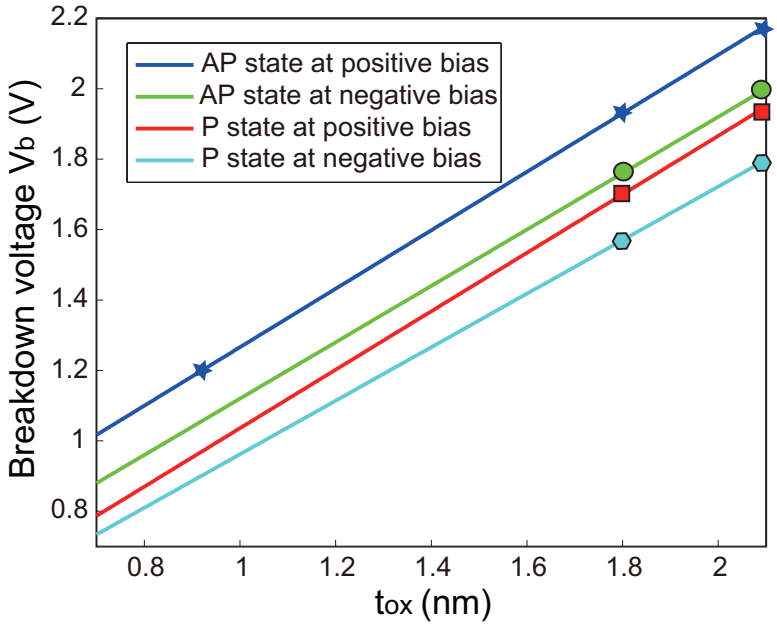


Figure 3.12: Breakdown voltage for different configuration (P or AP) and stress voltage (positive or negative) versus MgO thickness, the markers are experimental results from [72] (t_{ox}=1.8nm, 2.1nm) and [21] (t_{ox}=0.9nm).

Figure 3.13 demonstrates the switching voltage margin between the critical switching voltage V_c and the breakdown voltage V_b. V_c at P state can be defined as:

$$V_c = R_p \cdot I_{c0} \tag{3.29}$$

$$R_p = \frac{t_{ox}}{(F \cdot \bar{\varphi}^{1/2} \cdot Area)} \cdot exp(coef \cdot t_{ox} \cdot \bar{\varphi}^{1/2}) \tag{3.30}$$

where $\bar{\varphi}$ is the average potential barrier height of MgO (0.4eV), coef=$1.025nm^{-1}eV^{-1/2}$ is a fitting parameter and F is a fitting parameter corresponding to t_{ox} and RA product

which depends on the material composition of the three thin layers [11, 93]. The equations (3.4), (3.29), (3.30) indicate that V_c increases exponentially with t_{ox}. As V_b increases linearly with the t_{ox}, it is obvious that the switching operation margin (V_b-V_c) can be enlarged by reducing t_{ox}.

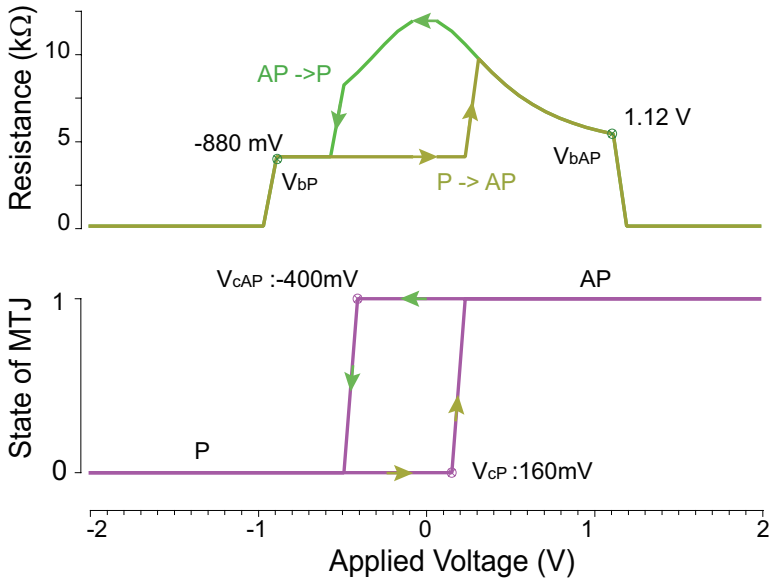


Figure 3.13: The MTJ state is changed as the applied voltage is higher than the critical switching voltage (V_{cP}, V_{cAP}). The resistance is steeply degraded beyond breakdown voltage (V_{bP}, V_{bAP}).

3.4.4.2 Prediction of lifetime

In condition of stress voltage below V_b, dielectric breakdown can also be induced by localized heating with high current flowing through the oxide barrier.

There exist numerous models for the lifetime estimation of thin dielectrics, e.g. E model, 1/E model, $E^{1/2}$ model, V model and power-law model [74, 102]. Among these models, E model is outstanding in explaining the majority of phenomena found in experiments of TDDB. Moreover, E model is the most conservative [102], which gives a shorter time-to-failure than other models (This is important due to the intrinsic nature of approximations in compact models). The lifetime of MTJ is described as:

$$ln(TF) \propto \frac{\Delta H_0}{k_B T} - \Gamma \cdot E_{ox} \tag{3.31}$$

where TF is the time to failure, ΔH_0 is the activation energy, $E_{ox} = V_{ox}/t_{ox}$ is the electric field across the oxide barrier, and Γ is the field acceleration parameter.

As ΔH_0 and Γ are intrinsic parameters of material MgO, we applied their values (fitting results of the experimental measurements in [74] and (3.31)) for predicting the lifetime of thinner MTJ. As shown in (3.31), temperature has an important impact on MTJ lifetime. One of the essences of E-model is, due to large local electric field with high dielectric constant (k), polar molecular-bonds inside MgO are weakened and will be broken by additional stress with applied heat. From the aspect of extrinsic mechanism, the risen temperature accelerates the area growth of pre-existing pinholes (e.g. boron diffusion in MgO) until the occurrence of breakdown [103].

3.4.4.3 Breakdown probability

Based on the MTJ lifetime, breakdown probability as a function of time can be consequently deduced. It is widely accepted that breakdown probability fits well the Weibull distribution for both Al_2O_3 and MgO based MTJ [73, 74, 76, 78, 80]. The cumulative distribution of breakdown probability can be described by:

$$F(t) = 1 - exp[-(\frac{t}{\alpha})^{\beta}] \tag{3.32}$$

where β is the shape parameter (decided by the MTJ oxide process [78]), α is equivalent to TF calculated in the precedent section. Figure 3.14 demonstrates the Weibull plot of this model and the experimental results of [74]. The lines are obtained by theoretical calculation. Taking into account the uncertainties in experiments (e.g. process variations, variable experimental conditions), the tendency shows great agreement. Breakdown probability increases fast with time propagation.

3.4.4.4 TDDB phenomena submitted to voltage pulse stress

The normal working condition of MTJ in MRAM is under pulsed voltage [73]. Thus, it is necessary to integrate the breakdown behavior of MTJ submitted to successive voltage pulses into the model. During voltage pulse δ, each trapped electron in MgO appears with a screening positive charge in the metallic electrodes, thus yielding a large electrostatic force between these two opposite charges. These electrons will escape during the interval of pulses Δt, which generates modulation of charge in MgO. The number of electrons trapped in the oxide barrier for three different cases is illustrated in Figure 3.15

The breakdown probability is composed of three mechanisms: electric field and heating (presented above F(t)), charges trapped in barrier (P_c) and modulation of trapped charges

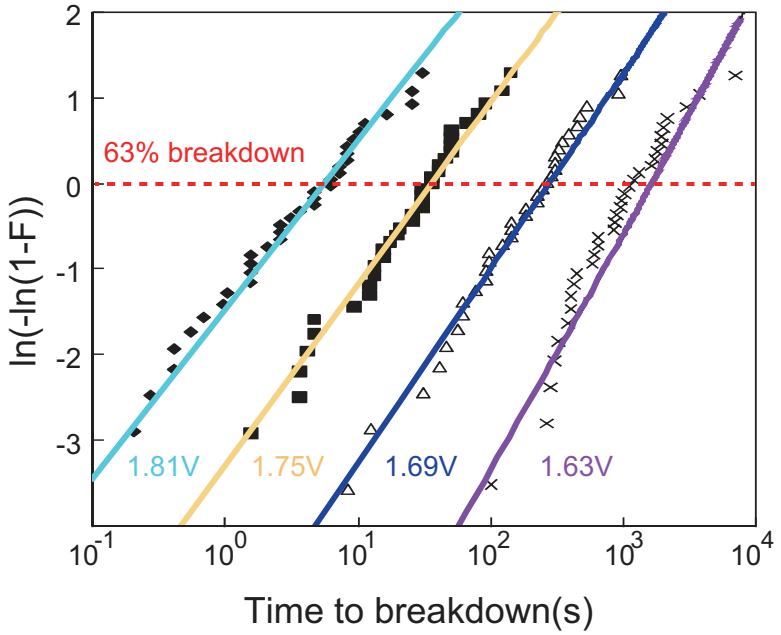


Figure 3.14: Time-to-failure statistics of MTJ at different stress voltages (t_{ox}=1.25nm). The time value corresponding to 0 of Weibull function represents 63% failure time.

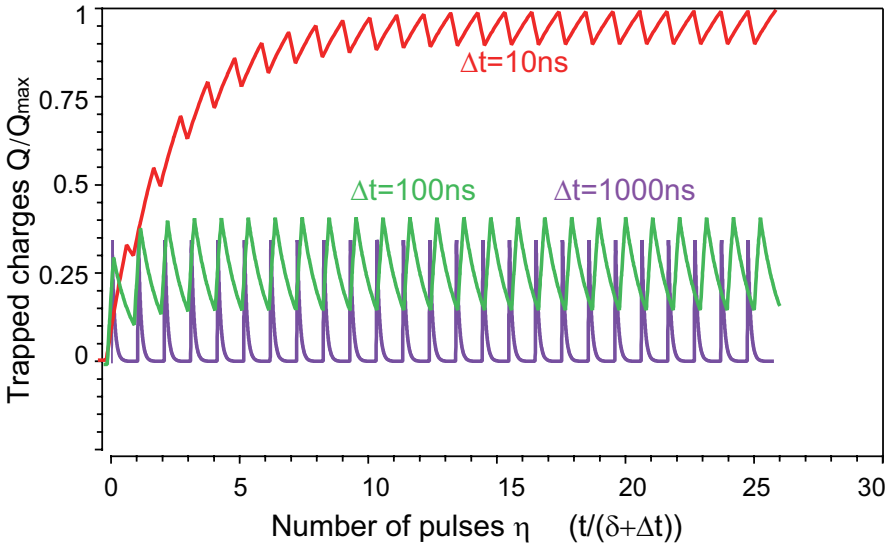


Figure 3.15: Number of voltage pulses before breakdown (η) versus interval between voltage pulses (Δt) with δ=30ns and τ=100ns.

(P_m). P_c and P_m can be calculated as [104]:

$$P_c = \frac{1 + exp(\Delta t/\tau)}{2\{1 + [e/(AI\delta)][exp(\Delta t/\tau) - 1]\}} \tag{3.33}$$

$$P_m = \frac{exp(\Delta t/\tau) - 1}{1 + [e/(AI\delta)][exp(\Delta t/\tau) - 1]} \tag{3.34}$$

where δ is pulse width, τ is the time constant for electron escaping, A is a dimensionless constant representing an effective normalized cross section of electron trapping. Hence the total breakdown probability P is given by:

$$1 - P = (1 - P_c) \cdot (1 - P_m) \cdot (1 - F(t)) \tag{3.35}$$

where α=$\eta\delta$ in $F(t)$, η is the number of pulses after which 63.2% of MTJs have failed and $\log(\eta)$ is proportional to electric field E_{ox}. η is also dependent on the interval between pulses Δt. Simulation results for the evolution of η and the experimental data in [73] are demonstrated in Figure 3.16. With $\Delta t \ll \tau$, the short lifetime results from the high level of charge trapping in the barrier (large P_c) and significant heating with relatively high frequency. The trapped charges enhance the electric field applied on MgO and the heating increases the temperature, which render the barrier fragile. In contrast, for $\Delta t \gg \tau$, the electrons have enough time to escape but exhibit a strong time dependent modulation (large P_m). This modulation induces a strong mechanical stress in the barrier and renders it fragile. The generated alternating stress on the oxide barrier facilitates atomic mobility through the barrier, i.e., pinhole formation. In the most favorable situation with $\Delta t \sim \tau$, an optimum trade-off regime is obtained between the density of trapped electrons and the charge modulation, thus yielding a maximum endurance [104].

3.5 Compact modeling in EDA tool Cadence

3.5.1 Modeling language: Verilog-A

The first step of modeling is to choose an appropriate language which can meet all the requirements. Modeling language is the interface between component or system physical models and electrical simulator. In the ICs designs, there exist four frequently used languages, e.g., C, Matlab, VHDL-AMS and Verilog-A [105]. Most of the analog or digital components and systems are described by these four languages. Although C language is outstanding in fast simulating speed and direct access to simulator, it has no standard interface. As a result, its applications in general macro modeling is limited because intimate

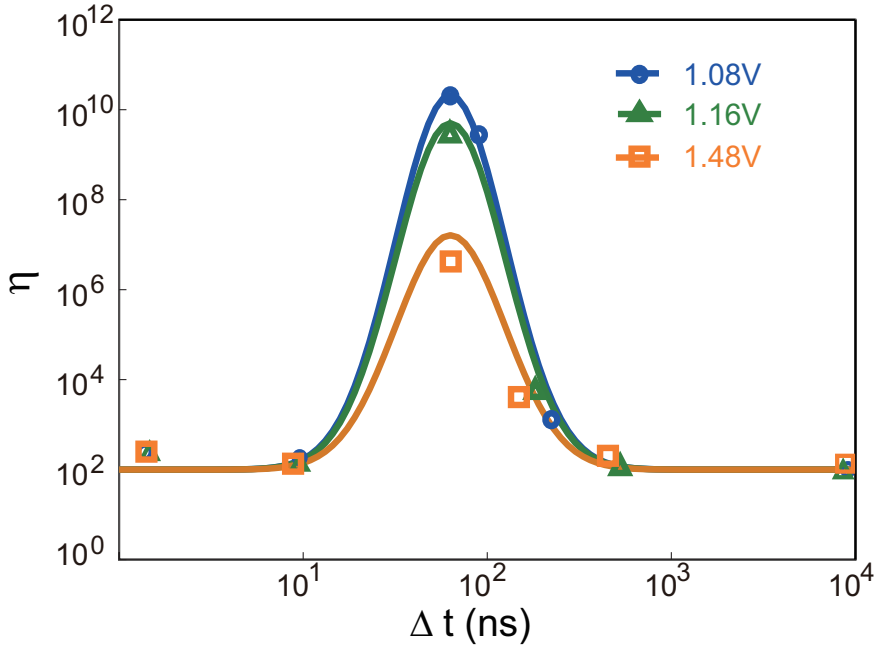


Figure 3.16: Number of voltage pulses before breakdown (η) versus interval between voltage pulses (Δt) with δ=30ns and τ=100ns.

knowledge of simulator is mandatory. Matlab is perfectly appropriate and efficient for data fitting and data processing, but it can not run directly in any analog simulator. As the first analog behavioral modeling language, VHDL-AMS is able to run in AMS designer of Cadence and ADVance MS of Mentor graphics etc. Nevertheless, there are only AMS simulators and no clear definition of VHDL-A. Moreover, the simulation time of VHDL-AMS language is relatively long, which limits its efficiency in analog designs. Verilog-A was an excellent language for compact model development and a dramatic improvement over others [106].

Verilog-A is high level modeling language for writing most of behavioral models of analog systems, which is an analog-only subset of the Verilog-AMS language developed by Accellera [106]. Verilog-A can run in the same AMS simulators as VHDL-AMS such as Spectre [107], Eldo [108] and ADS [109], as well as internal simulators of foundries such as STMicroelectronics, IBM and TSMC, etc. Compared with general-purpose programming languages, VerilogA is advantageous in compact modeling because it frees the model developer from the burden of handling the simulator interface. Moreover, it features excellent capability of differential-algebraic equations, conservative or signal-flow systems and mixed disciplines (mechanical, electrical, rotational ...) and enables the feasibility of parameterization, hierarchy, analog operators as delay, transition, slew, noise and analog events as cross, timer, initial/final steps etc. Verilog-A also provides a strong system for

defining model parameters [106]. The declaration statement includes the default value and can specify the range of valid values. The default value may also be a function of other (previously declared) parameters. Furthermore, Verilog-A models can be shared which promises global standardization. Large high-level systems including mixed-discipline and non-electrical systems could be quickly investigated with deeper design exploration.

Generally speaking, Verilog-A is perfect as the programming language for creating an easy, efficient, accurate, fast and compatible compact model of STT-MTJ. Besides all the advantages mentioned above, it is very readable, both characterization engineers and circuit designers can easily comprehend it, which improves the continuity of this work and simplifies the development of next version of model.

3.5.2 Electrical Modeling of MTJ under Cadence

3.5.2.1 Hierarchy of the physical models integrated in the compact model

The hierarchy of the physical models which are integrated in the compact model is illustrated in Figure 3.17. All the parameters, constants and variables are defined at the beginning of the model. The values of parameters can be reconfigured by circuit designers. In computer-aided design tools, each run of DC/transient simulation is automatically divided into many small steps, e.g. a simulation of 10ns can be divided as 0-1ns, 1-1.3ns, 1.3-1.5ns etc. and a simulation from 0 to 1V can be segmented as 0-0.1V, 0.1V-0.2V etc. For each iteration, the simulation starts from extracting the bias voltage value V_{bias} and all the parameters relative to the precedent state: *State* represents the MTJ relative magnetization orientation (0 for parallel and 1 for anti-parallel), *Break* signifies the occurrence of dielectric breakdown after last iteration and T_{ini} signifies the temperature at the end of the last iteration. If the breakdown has already taken place, the simulation is finished immediately. If not, the calculations of the performance parameters are executed while the reliability terms are injected. At the end, the breakdown status and the temperature are refreshed and some performance parameters can be obtained at the output.

3.5.2.2 Parameters of the compact model and Component Description Format (CDF)

Tremendous parameters are integrated in the model to match the experimental measurements. In order to get a configuration corresponding to experimental data, the users can set the MTJ physical parameters using the default values in Table 3.1. There are mainly four types including general constants, device technology parameters, device specification parameters and reliability issues control flag. The technology parameters is determined

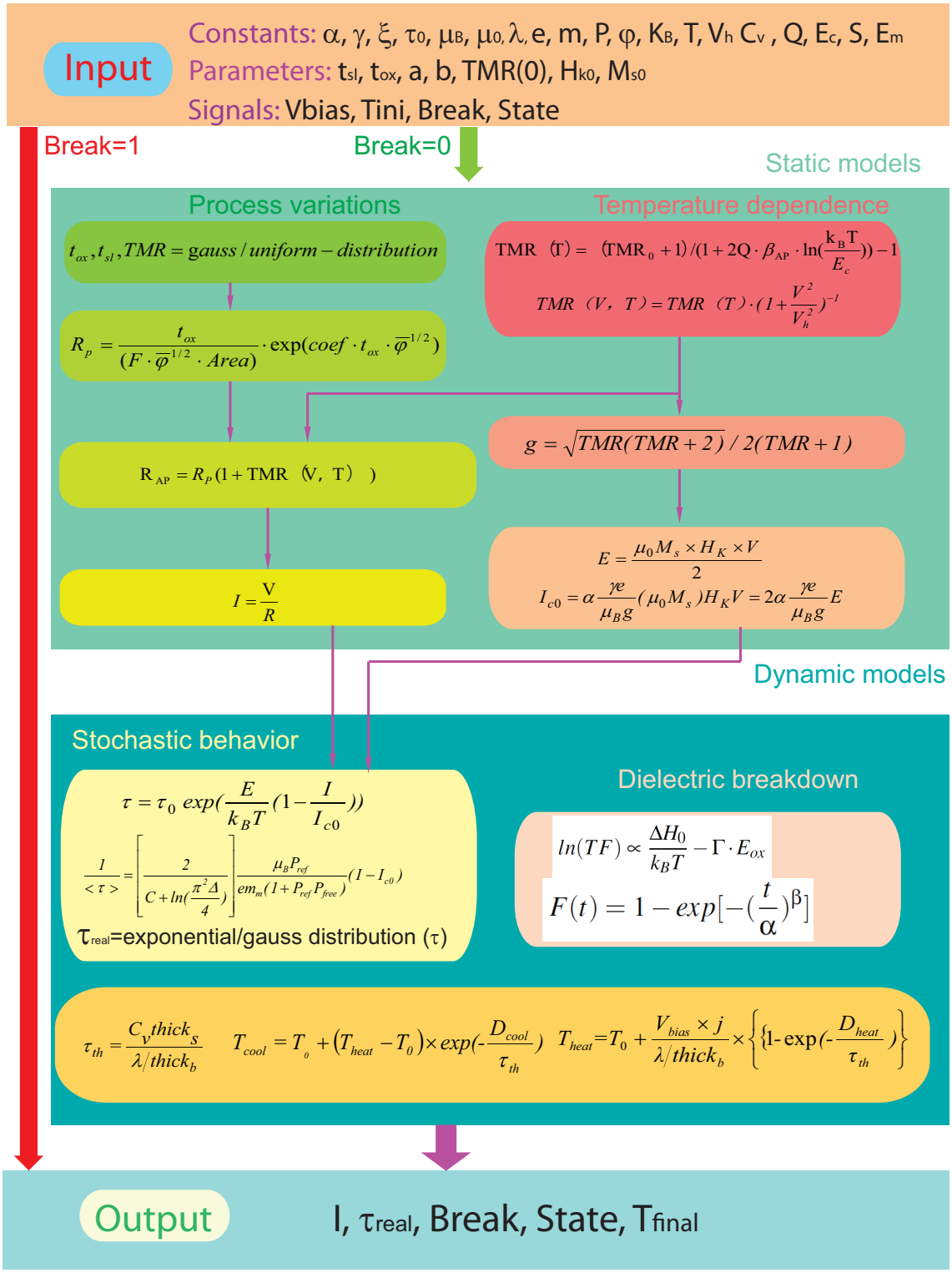


Figure 3.17: Architecture of PMA STT MTJ compact model integrating physical models of reliability issues.

by the material composition and homogeneity. The device parameters depend mainly on the process and mask design. Model users can modify these two types of parameters according to different magnetic process and material composition. The values of parameters also have constraints, which is limited to the mainstream technology, for instance, the thickness of oxide barrier is ranged between 1 nm and 3nm.

In order to facilitate the utilization of model, we have created the graphical user interface (GUI) using the function of Component Description Format (CDF) in Cadence design environment. CDF provides the method to define the parameters and the attributes of individual component or libraries. Users can reconfigure the device by only entering the values at the interface and the system will transfer these values to the simulator (e.g. spectre) for simulations. As demonstrated in Figure 3.18, the initial configuration of MTJ (device parameters, P or AP state, considered reliability issues) can be set by modifying the properties in the cases. As the configuration method is exactly the same with that of conventional transistors,it is very convenient for circuit designers to concept more complex hybrid MTJ/CMOS circuits.

3.5.2.3 Schematic view of model and relative circuit in the design environment

In the simulation tool Cadence, a symbol can be created by representing the model programmed in VerilogA language. This symbol will be visible to users and facilitate the simulation settings. Figure 3.19 shows the symbol of model and a test circuit in the simulation tool Cadence. The symbol has five terminals: 'T1' and 'T2' represent the two real pins of MTJ while other three are virtual which are used to demonstrate the key information to circuit designers. 'State' represents the relative magnetization configuration of MTJ ('0' for P and 1 for AP), 'Temp' signifies the real temperature of MTJ which fluctuates with real time, 'Break' demonstrates the occurrence of dielectric breakdown. As this symbol has two many terminals, it is not easy to integrate into the circuit. So we have created a simplified symbol in the right figure which only demonstrates the two real pins and the MTJ state. If the users need more information, they can descend to the basic model. Figure 3.20 illustrates the schematic view of pre-charge sense amplifier (PCSA) circuit which is usually used in hybrid MTJ/CMOS design. The currents are injected from the two real pins and the 'State' can be used to determine if the writing/reading operation has changed or not the MTJ state.

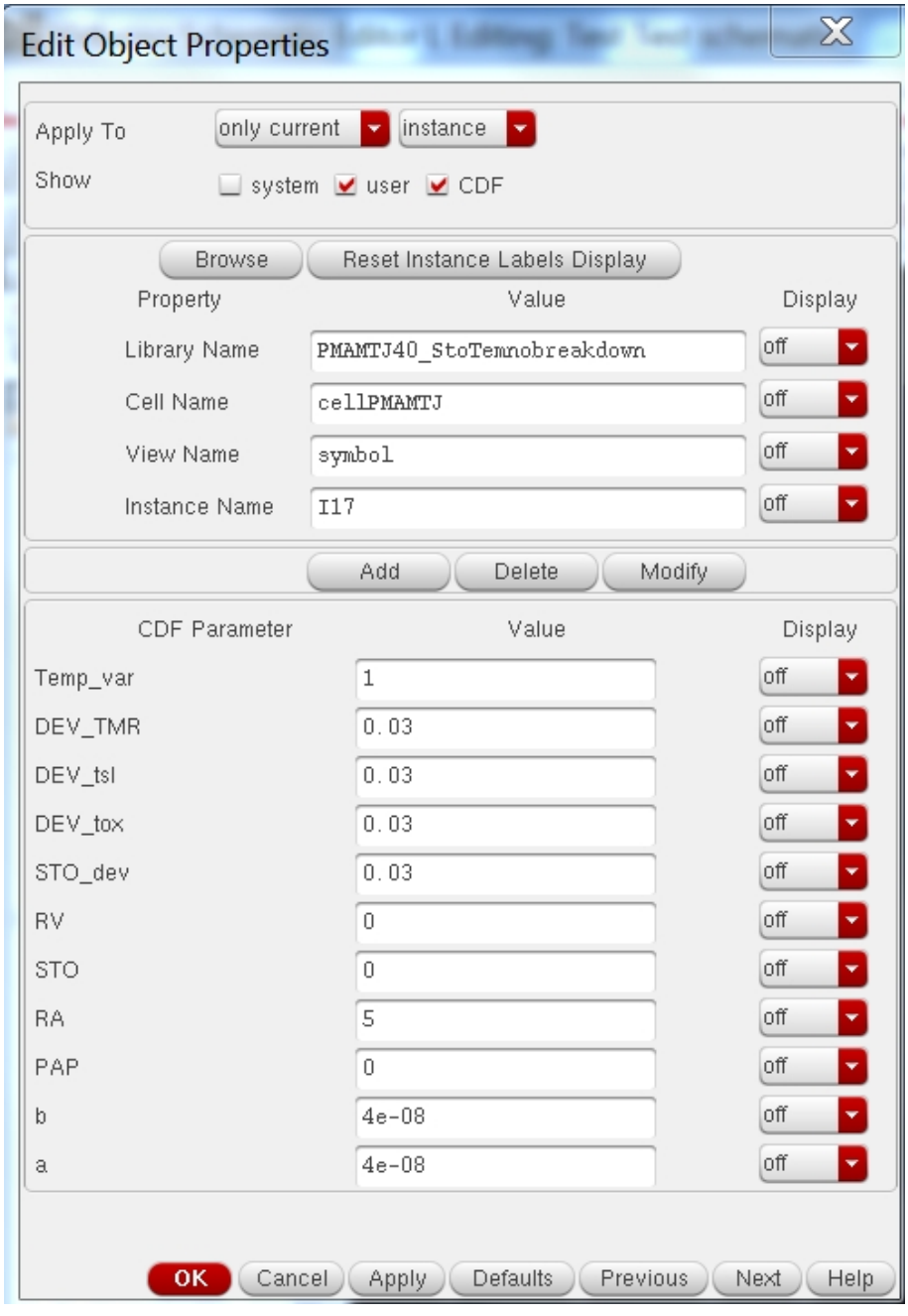
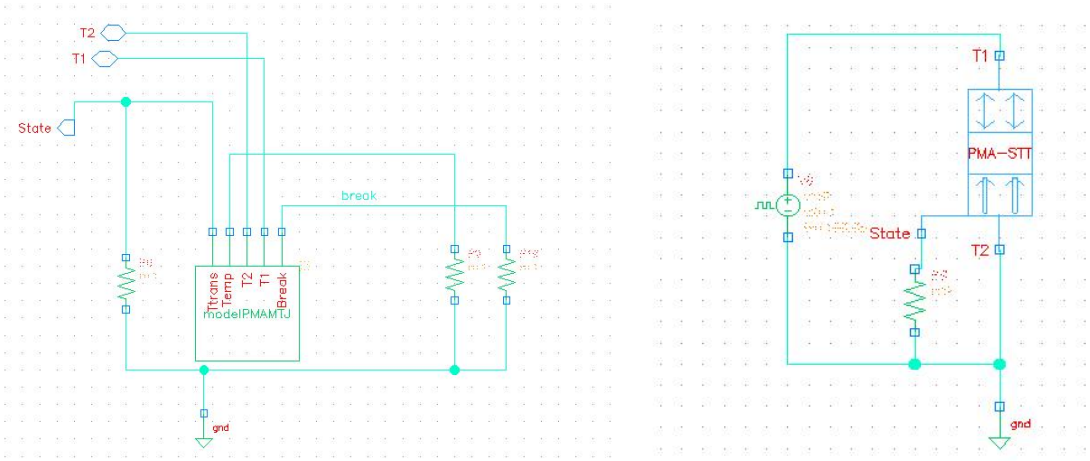


Figure 3.18: Component Description Format (CDF) in Cadence.



(a) (b)

Figure 3.19: (a)Symbol of the PMA-STT-MTJ compact model (b) Symbol at circuit level.

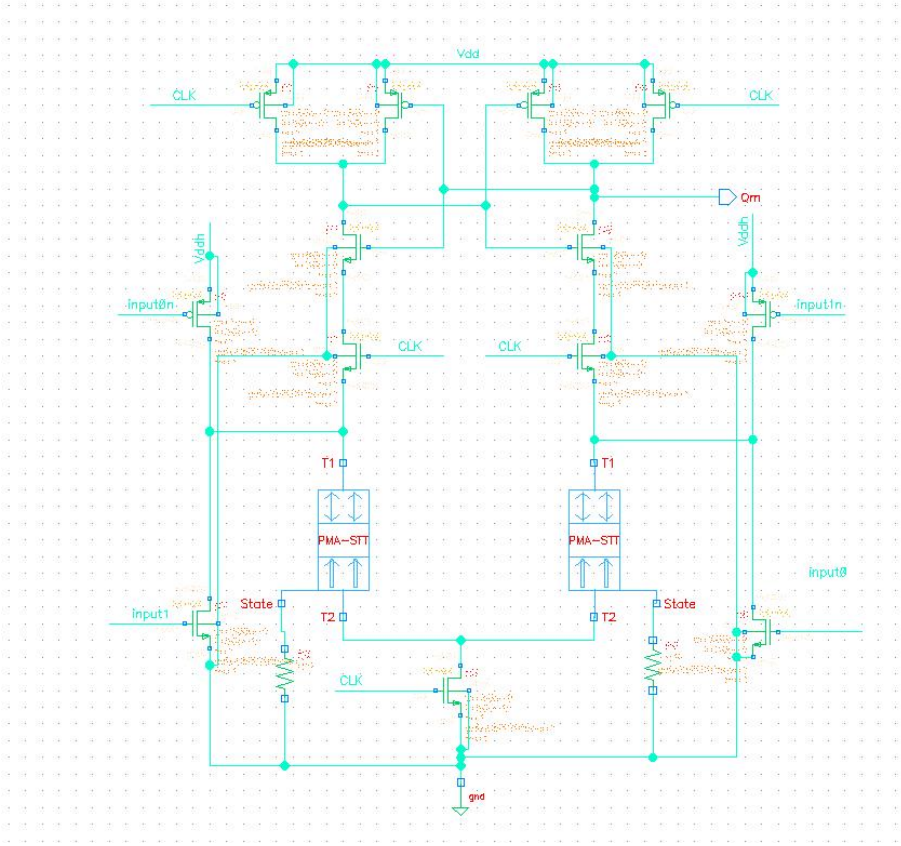


Figure 3.20: Schematic of pre-charge sense amplifier circuit.

Table 3.1: Parameters integrated in the compact model including constants, technology parameters, device parameters and reliability control flags.

Constants	**Description**	**Default Value**
C	Euler's constant	0.577
e	Elementary charge	$1.60 \cdot 10^{-19} C$
m	Electron mass	$9.1 \cdot 10^{-31} kg$
$\hbar$	Reduced Plank constant	$1.0545 \cdot 10^{-34} J \cdot s$
α	Magnetic damping constant	0.027
γ	gyromagnetic ratio	$1.76 \cdot 10^{11} rad/(s \cdot T)$
μ_0	permeability in free space	$1.257 \cdot 10^{-6} T/(m \cdot A)$
μ_B	Bohr magneton constant	$9.27 \cdot 10^{-28} J/Oe$
k_B	Boltzmann constant	$8.625 \cdot 10^{-5} eV/K$
T_0	Room temperature(TR)	300 K
Technology	**Description**	**Default Value**
H_k	Anisotropy filed	$113.0 \cdot 10^3 A/m$
M_s	Saturation magnetization	$1257.0 \cdot 10^3 A/m$
ΔH_0	Activation energy	0.8eV
Γ	Field acceleration parameter	84.897 W/(m K)
β	Shape parameter	1.5
C_v	Heat capacity per unit volume	$2.74 \cdot 10^6 J/(m^3 K)$
λ	Thermal conductivity	1.5
Device	**Description**	**Default Value**
t_{ox}	Thickness of oxide barrier	0.85nm
TMR(0)	TMR ratio with 0 stress voltage	150%
Area	MTJ surface	40nm $\cdot$ 40nm $\cdot$ $\pi/4$
t_{sl}	Thickness of free layer	1.3nm
$thick_s$	Thickness of MTJ nanopillar	33.55nm
V_{sl}	Volume of free barrier	Area $\cdot$ t_{sl}
a,b,r	Dimensions of MTJ	40 nm, 40nm, 20nm
RA	Resistance-area product	5 $\Omega \cdot \mu m^2$
Flags	**Description**	**Default Value**
PAP	State of MTJ (P or AP)	0/1
RV	Process variations	0/1/2
DEV_TMR	Variation percentage of TMR when RV=1,2	0.03
DEV_t_{ox}	Variation percentage of t_{ox} when RV=1,2	0.03
DEV_t_{sl}	Variation percentage of t_{sl} when RV=1,2	0.03
STO	Stochastic switching	0/1/2
DEV_STO	Variation percentage of switching duration when STO=1,2	0.03
Temp_var	Self-heating effect	0/1
Break	Occurrence of dielectric breakdown	0/1

3.5.3 Functionality validation of model

This section demonstrates some simulation results to validate the functionality of the compact model including reliability issues.

The function of process variations can be demonstrated by two kinds of simulations: DC and transient. As illustrated in Figure 3.21, the first one can reflect the different resistances value caused by the variations of t_{ox} and TMR. As the TMR value nearly has no impact on the resistance of parallel (P) state. The anti-parallel (AP) state resistance has a larger scale distribution than that of P state. In the latter one, the current of MTJ follows a distribution with the same bias voltage. Note that the parameters (t_{sl}, t_{ox}, TMR follow a gaussian distribution with deviation of 1%.

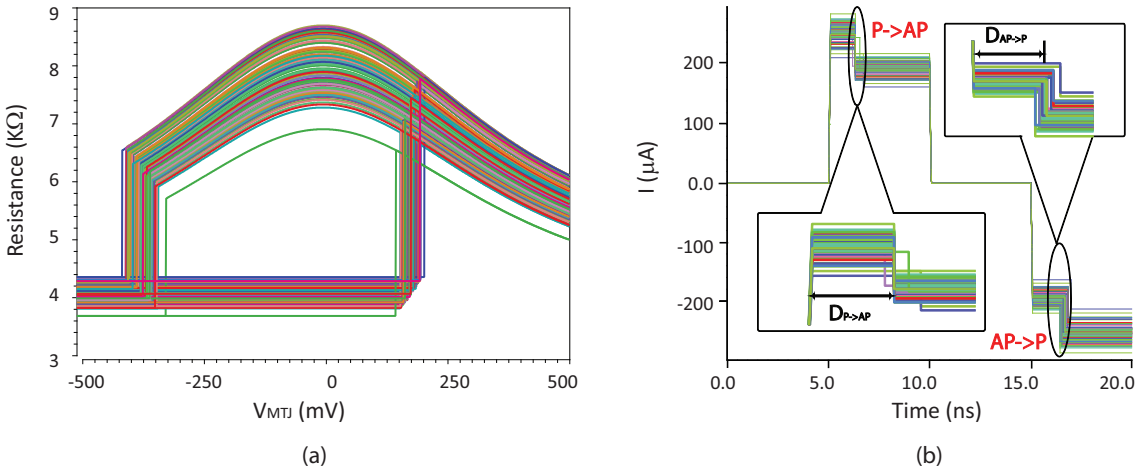


Figure 3.21: MC simulations of (a)bias voltage dependent resistance and (b) 1000 complete writing process with process variations.

Using the writing circuit shown in Figure 2.13, Monte-Carlo simulations of 1000 writing processes are performed in which the switching duration follows a normal distribution with the mean value of $\tau_{p->ap}$ or $\tau_{ap->p}$ and variation of 0.02. The results in Figure 3.22 demonstrate that the average switching duration (without stochastic behavior) is $\tau_{p->ap}$=1.4716 ns and $\tau_{ap->p}$=2.4898 ns. As expected all the values of switching duration for parallel (P) state to antiparallel (AP) state are in the interval [$0.98\tau_{p->ap}$, $1.02\tau_{p->ap}$]. It follows a normal distribution around the average switching delay time $\tau_{ap->p}$ and the variation set. On the other hand, for AP to P state, there are 98% values of switching duration in the interval [$0.98\tau_{ap->p}$, $1.02\tau_{ap->p}$]. As the voltage is identical for both states, the current for AP state is lower than P state, which leads to a longer switching duration [21] and higher error rate of switching.

Figure 3.23 illustrates the switching probability as a function of applied switching

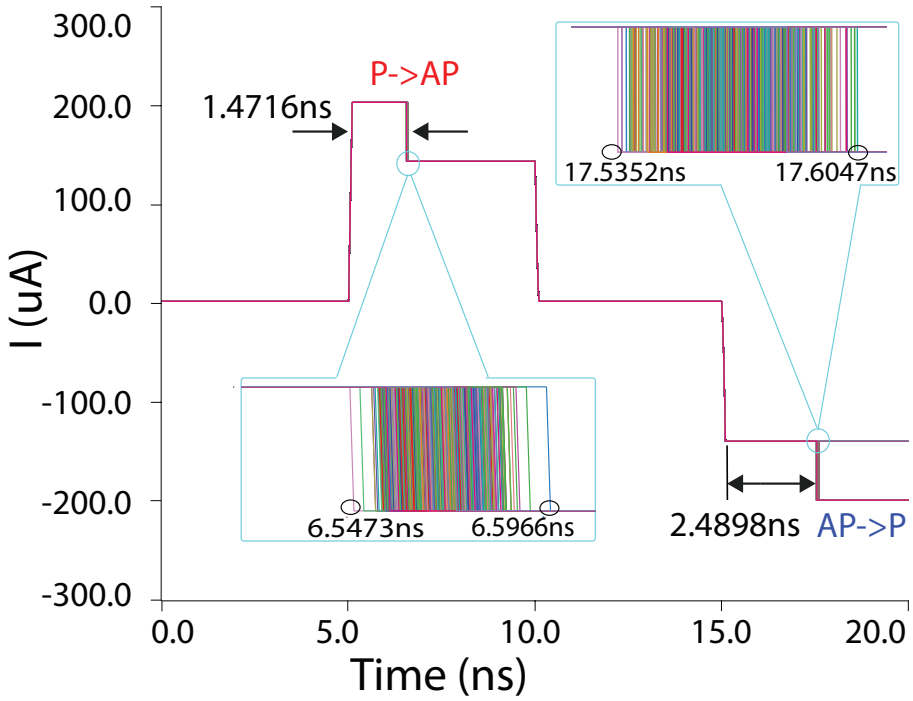


Figure 3.22: MC simulations of 1000 complete writing process with the stochastic behaviors. The switching duration is set following a normal distribution with variation of 0.02.

voltage and switching delay. The dark red zone is considered as the reliable writing zone whereas the dark blue zone is the reliable reading zone.

Figure 3.24 displays the temperature dependence of TMR, which is consistent with the experimental results [99].

The temperature dependence of TMR ratio and critical switching current can be observed in Figure 3.25. The mechanism behind this phenomenon is as follows: As temperature increases, the barrier energy goes down and the magnetic spins have a larger thermal energy, which helps the spins cross over the barrier more easily.

As the temperature tolerance of FPGA circuits is usually limited (218K to 398K) and the temperature increase of MTJ can reduce the operating temperature range of MTJ-based integrated logic circuits, the temperature should not exceed the limit (e.g. 388 K 90nm) [68]. Thus, the parameters of MTJ should be carefully chosen to assure the best performance. An analysis is executed to find the dependence of the maximal increase of the temperature on the thickness of oxide barrier tox and the area of MTJ (10 x 10 nm^2, 20 x 20 nm^2, 40 x 40 nm^2). From the results shown in Figure 3.26, we find that t_{ox} should be small enough to guarantee the sufficient temperature tolerance of MTJ. As expected, the temperature is proportional to the power density j x V_s [23] and the area has no

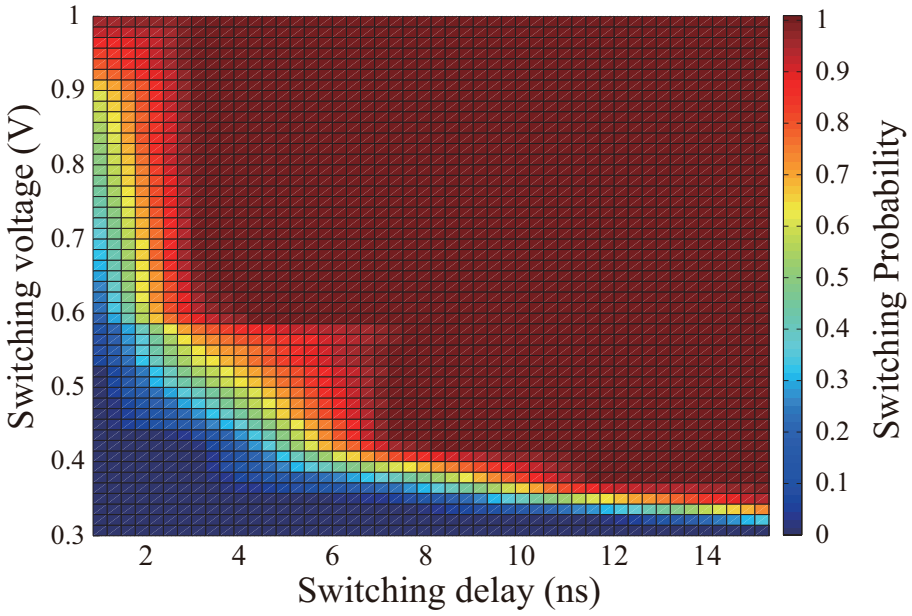


Figure 3.23: Switching probability as a function of applied switching voltage and switching time.

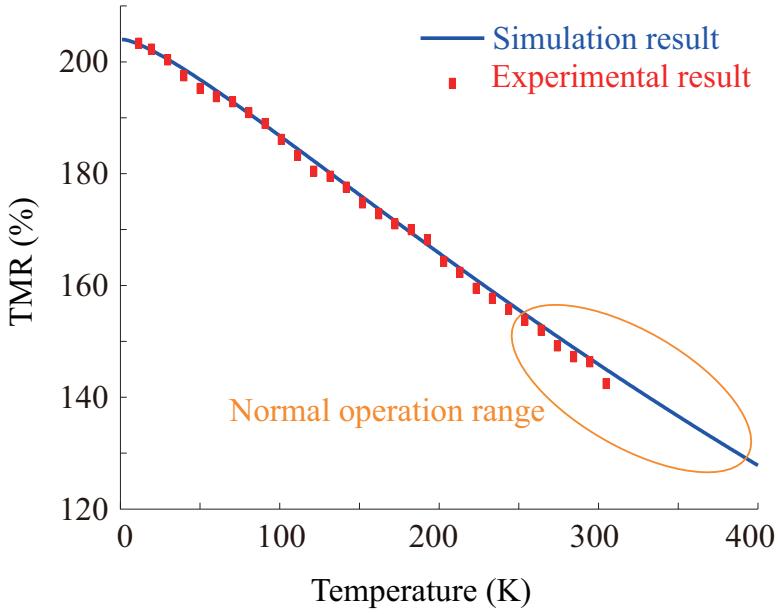


Figure 3.24: TMR evolution with temperature increase and the experimental data (red points) in [99].

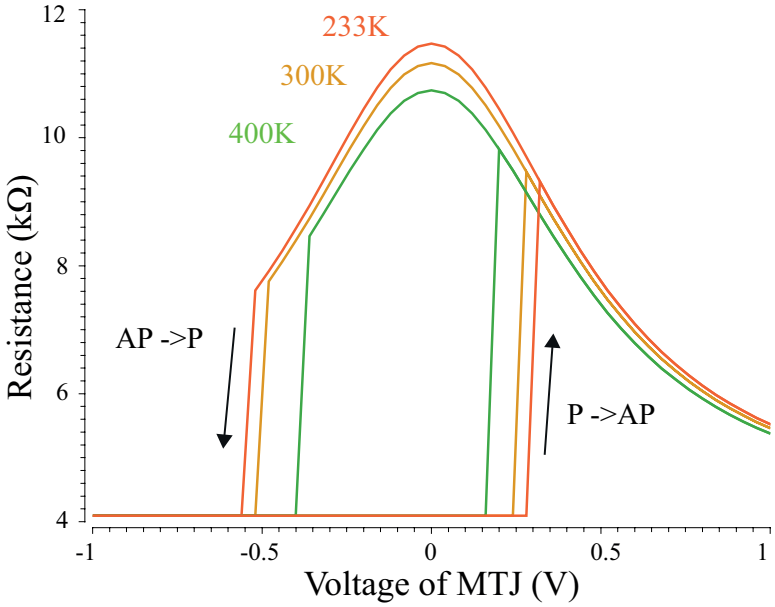


Figure 3.25: Resistance of MTJ versus bias voltage with different temperatures. Critical current is reduced by increasing temperature.

impact on temperature increase. From equations (3.25) and (3.26), the maximal increase value of the temperature can be described by

$$\Delta T_{max} = \frac{V_s \cdot j}{\lambda / t_{ox}} = \frac{Area \cdot R \cdot j^2}{\lambda / t_{ox}} \tag{3.36}$$

where R is the magnetoresistance of MTJ as demonstrated in (3.30):

$$R \propto \frac{F(t_{ox})}{Area} \tag{3.37}$$

where F(t_{ox}) is a function of t_{ox} and some constants, then

$$R \propto \frac{F(t_{ox}) \cdot j^2}{\lambda / t_{ox}} \tag{3.38}$$

Simultaneously, the Equation (3.38) shows that a higher value of tox results in a more obvious temperature increase.

Figure 3.27 shows the MTJ lifetime for different oxide barriers. In consideration of the self-heating, the MTJ lifetime with 1 nm thick barrier can be estimated to 10 years for a typical operating voltage of 420 mV. This result meets excellent agreement with the value referred in [110].

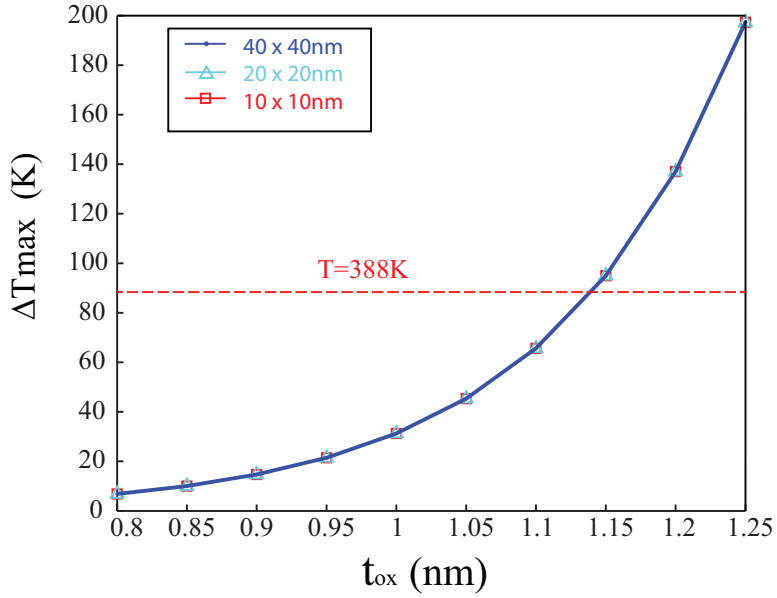


Figure 3.26: Dependence of reading error rate on the thickness of oxide barrier tox and area of MTJ.

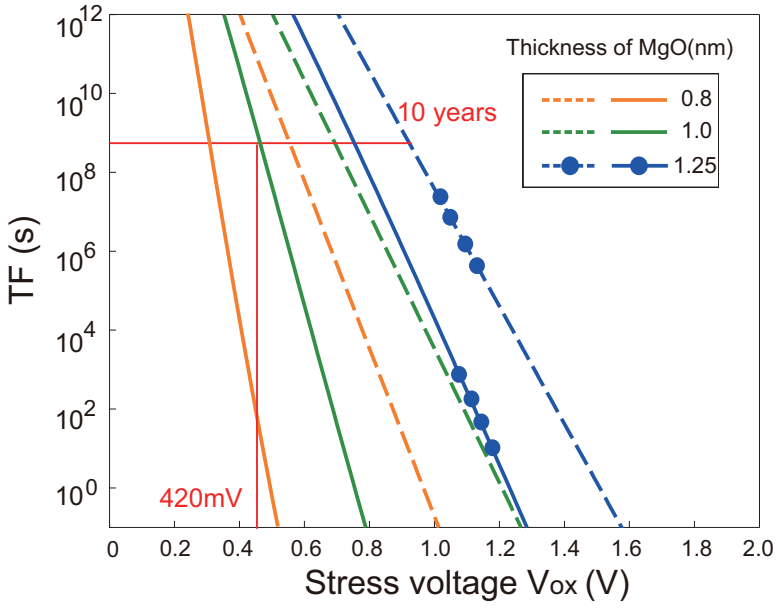


Figure 3.27: Lifetime of MTJ without (dashed lines) and with (lines) consideration of self-heating. The dots are experimental data in [74].

3.6 Fast simulation model using worst-case corners

3.6.1 Introduction of worst-case fixed corners model

For better variability aware design, tremendous compact models have been proposed to mimic the process variations in MTJs [56, 98]. Monte-Carlo (MC) methodology is usually used in most of the MTJ compact models for variability analysis [111]. MC method is used to predict the parameter fluctuations with the probability distribution. Without consideration of simulation times, MC techniques are inherently accurate as they do not involve any approximation of simulation results. In practice, MC simulation performs at a low hierarchical level, demands excessive amounts of computation time.

In traditional MC sampling, a large number of simulation iterations are required to achieve a reasonably precise estimation of ICs fabrication yield. Advanced sampling techniques such as the stratified sampling, Latin Hypercube Sampling (LHS) and Quasi Monte Carlo (QMC) are applied to some digital circuits [112]. They can achieve a faster convergence rate comparing with MC-based timing analysis. Even though LHS is a type of stratified MC method with less number of samplings, considerable simulation time is always required for an effective conclusion.

Nonetheless, all these MC relative methods take too long time for an effective conclusion, especially when the circuit is in very large scale, e.g., high density memory arrays. Moreover, the simulation tool may not be able to handle the sophisticated task with a complex circuit structure and even break down, leading to unnecessary time loss. With the fast increasing number of devices integrated in one chip, the requirement of compact models of MTJ and CMOS transistor with fast simulation speed becomes more intensive. While models of transistors include worst-case corners which are efficient in performance evaluation [113, 114, 115, 116], it is essential to integrate a fast and efficient methodology in the MTJ model as well for hybrid MTJ/CMOS design. In this modeling approach, the standard deviation limits are preset pessimistically to include any potential process variability over a wide range. Compared with MC simulation, worst-case analysis has much faster speed while demonstrating the performance boundary. This boundary covers most of the possible cases of process variations and demonstrates to designers an outline of power-delay, so that they can adjust the design vectors (design kit, architecture, bias voltage, devices size) to achieve an optimization concept [117].

This section proposes a new model including worst-case corners for fast simulation requirement of hybrid CMOS/MTJ circuits [118].

3.6.2 Worst-case fixed corners model of MTJ

For the CMOS transistors models, a compact surface-potential-based (PSP) MOSFET SPICE model is applied into circuit simulation. In order to know the variability and yield information, design corners with letter acronyms F, S and T (F: fast, S: slow and T: typical) are used to describe NMOS and PMOS performance characteristics, which aim to show the general trends in the design quantities caused by the manufacturing process. For example, 'TT' is the typical compact model extracted from the golden die of the golden wafer representing the center-line process technology. Transistors with the minimum oxide thickness, threshold voltage (V_{th}) and ΔW, as well as maximum ΔL are represented with FF (fast NMOS, fast PMOS). In PSP model for FDSOI technology, these maximum and minimum values are accounted by several key process parameters, which are varied to reflect parametric process variation effects on circuit performance. As shown in Table 3.2, V_{th} is determined by flat-band voltage V_{FBO} and $NSubO$. UO and CSO are mobility parameters. CFL is setup for short-channel effects [119].

Table 3.2: Parameters settings of CMOS transistors

PSP model	Description	Default Value
VFBO	Geometry-independent flat-band voltage	-65.65mV
NSUBO	Substrate doping	$2 \cdot 10^{18} m^{-3}$
UO	Zero-field mobility at TR	15.05 m^2/V/s
CSO	Geometry-independent flat-band voltage	-65.65mV
CFL	Length dependence of DIBL-parameter	14u V^{-1}

In the conventional transistor models, NMOS and PMOS are modeled by four worst-case corners: slow nMOS and slow pMOS (SS) for worst-case speed, fast nMOS and fast pMOS (FF) for worst-case power, fast nMOS and slow pMOS (FS) for worst-case '1', and slow nMOS and fast pMOS (SF) for worst-case '0' [114]. Based on this method, we try to model the worst cases of MTJ from the two different states (Parallel (P) and anti-parallel (AP)). Note that MTJ at P and AP state are the two states of the same device, resulting in strong correlation (All of the process parameters are identical). Consequently, only fast P state and fast AP state (FF) are generated to model the worst-case power and slow P state and slow AP state (SS) are generated to model the worst-case speed. Note that the FF signifies the minimum resistance value and the fastest state switching of MTJ. Reversely, SS signifies the maximum resistance value and the slowest switching. The other two worst

cases can be configured with the exact application of MTJ in the circuit.

This modeling approach presets the standard deviation (σ) limits pessimistically to include any potential process variability over a wide range [114]. The worst-case corners are generated by offsetting the selected parameters, P of the typical (TT) compact model by worst-case distance $\pm dP = n\sigma$ to account for the process variability window, where n is the number of σ for P so that $3 \leq n \leq 6$ is selected to set the fixed lower and upper limits (LL and UL), respectively of the worst-case models. Note that n, which represents the worst-case distance, should be very carefully chosen after tremendous simulations of small circuit with considering the exact simulation conditions to avoid over-estimation and under-estimation [116]. The principal is that the corners should cover most of the cases obtained by statistical simulations (e.g., Monte-Carlo). Usually the percentage of cases covered should be more than 99%. From the Eequation (3.30), it can be deduced that the current I_p and I_{ap} are initially calculated by [98]:

$$I_p = \frac{V_{MTJ}(F \cdot \bar{\varphi}^{1/2} \cdot A)}{t_{ox} \cdot exp(coef \cdot t_{ox} \cdot \bar{\varphi}^{1/2})} \tag{3.39}$$

$$I_{ap} = \frac{V_{MTJ}(F \cdot \bar{\varphi}^{1/2} \cdot A)}{t_{ox} \cdot exp(coef \cdot t_{ox} \cdot \bar{\varphi}^{1/2})(1 + TMR)} \tag{3.40}$$

where V_{MTJ} is the voltage applied across MTJ. Then, the UL (or LL) is set by taking the appropriate offset of the device parameters to maximize (or minimize) the value of I_p and I_{ap}. Thus, the UL of MTJ current can be expressed as:

$$I_p(UL) = \frac{V_{MTJ}(F \cdot \bar{\varphi}^{1/2} \cdot (A + dA))}{(t_{ox} - dt_{ox}) \cdot exp(coef \cdot (t_{ox} - dt_{ox}) \cdot \bar{\varphi}^{1/2})} \tag{3.41}$$

$$I_{ap}(UL) = \frac{I_p(UL)}{(1 + TMR - dTMR)} \tag{3.42}$$

$$I_p(LL) = \frac{V_{MTJ}(F \cdot \bar{\varphi}^{1/2} \cdot (A - dA))}{(t_{ox} + dt_{ox}) \cdot exp(coef \cdot (t_{ox} + dt_{ox}) \cdot \bar{\varphi}^{1/2})} \tag{3.43}$$

$$I_{ap}(LL) = \frac{I_{ap}(UL)}{(1 + TMR + dTMR)} \tag{3.44}$$

Within the same principal, other parameters of worst-case corners can be deduced, e.g., the switching delay τ_p ($I > I_{c0}$) of P state can be calculated as:

$$\frac{1}{\tau_{sw}(UL)} = [\frac{2}{C + ln(\frac{\pi^2 \zeta}{4})}] \frac{\mu_B P_{ref}(I - I_{c0}(UL))}{em_m(1 + P_{ref} P_{free})} \tag{3.45}$$

$$I_{c0}(UL) = \frac{\mu_0 e\alpha\gamma(M_s - dM_s)(H_k - dH_k)(A - dA)(t_{sl} - dt_{sl})}{\mu_B(g + dg)} \tag{3.46}$$

Note that the upper limit for τ_p and critical current I_{c0} is the minimum value while the lower limit is the maximum value.

Fig. 3.28 demonstrates the current and switching delay in the two states of MTJ obtained from the worst-case model. The distribution of MC simulation results (1000 runs) is generated from statistical model proposed in the previous sections. Note that the standard deviation is σ=0.01and n=3. It is obvious that the fixed worst-case corners can exactly cover the majority of process variation conditions (> 99%). Meanwhile, the worst-case simulations cost 3 runs (4.59s) while the MC simulations cost 1000 runs (668s).

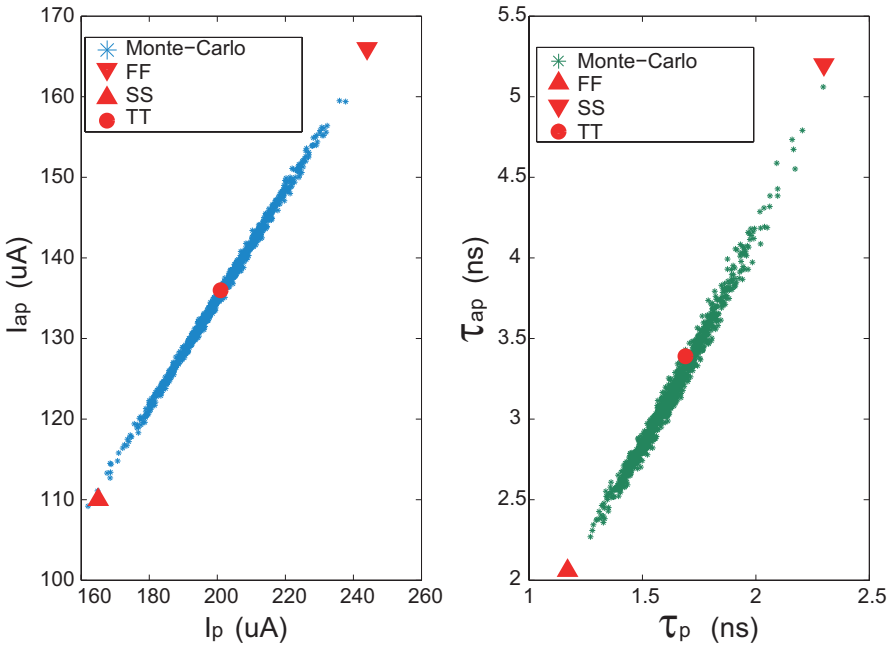


Figure 3.28: Simulation results of statistical model (1000 MC simulations) and worst-cases model (TT, FF, and SS): (a) current and (b) switching delay of MTJ in different states (P or AP) with a voltage pulse(σ=0.01 and n=3).

3.7 Conclusion

This chapter presented a compact model of PMA STT MTJ including reliability issues. Some reliability effects are firstly investigated and modeled at physical level, such as process variations, stochastic switching, temperature fluctuation and dielectric breakdown. Then, several physical models are utilized to constitute the compact model. Comprehensive study of modeling languages has conducted us to choose Verilog-A language for com-

pact modeling, which features high compatibility with standard CMOS computer-aided design tools and easy interface settings. Simulation results using realistic material parameters show excellent agreement with experimental data. Different kinds of simulations (DC, transient) were performed to validate the basic functionality and the temperature dependence of the parameters and temperature fluctuation due to Joule heating effect. Monte-Carlo simulations of CMOS/MTJ circuit were performed to validate its functionality for success possibility of writing process, reading operation and probabilistic occurrence of breakdown.

Besides, a fast simulation model using worst case corners has been proposed and validated which features high simulation efficiency and accuracy in very large scale circuits. It can be used to obtain an outlook of the very large scale circuits performance with few time.

The developed models can be used to design MRAM and non-volatile logic circuits with an enhanced performance of operation speed, power consumption, reliability and endurance. Completely implemented in Verilog-A language, they have high compatibility with different dimensions of CMOS design kit under Cadence. This will significantly contribute to realizing future non-volatile logic and memory applications.

In the following chapters, these compact models will be used with CMOS technology design kit to study and analyze the reliability of more complex hybrid logic and memory circuits and to concept novel circuit designs of specific applications.

Chapter 4

Reliability analysis and variability-aware design of hybrid MTJ/CMOS circuits

This chapter will concentrate on the reliability analysis and exploration for reliability optimization methodology using the compact model developed in the previous chapter. Firstly, some typical circuits for memory and logic will be analyzed to validate the model functionality and find the key parameters determining the reliability. Based on the results, we will propose some methodologies for reliability-aware designs.

4.1 Reliability analysis of MTJ based circuits

The pre-charge sense amplifier (PCSA) circuit is very important in MTJ based memory and logic circuits. In PCSA, the dynamic sensing method allows the amplification from analog data to digital with ultra-low power. Moreover, the read disturbance induced by sensing operations can be significantly decreased [61]. The latter is very important for embedded STT-MRAM as it is an intrinsic constraint limiting the reliability of logic circuit where complex error correction circuit (ECC) is necessary to ensure fast computing speed (e.g. 1 GHz). Thus, PCSA structure is widely used in logic gates, arithmetic unit cells and memory cells [46, 47, 49, 120]. In this section, the circuit presented in Figure 4.1 will be used to carry out reliability analysis.

This circuit consists of two blocks: writing control block for sending writing signals to MTJs and pre-charge sense amplifier for sensing out the data stored in MTJs. It can not only serve as writing and reading block of STT-MRAM but also function as Flip-Flop with an extra register at the output. When the writing control signal WE is disabled, the circuit functions at standby mode (no input but the data stored in MTJs can be sensed). Contrarily, the circuit operates in two phases:

- *Writing phase (Pre-charge phase)* : CLK='0' and WE= '1', if the input Data is set to logic '0', the transistors MP4 and MN6 are turned on while the transistors MP5 and MN5 are turned off. Thus, the MTJ0 is switched to AP state and MTJ1 is switched to P state. At the same time, the complementary outputs of sensing part are pre-charged to V_{dd}.

- *Sensing phase* : CLK = '1', the two branches begin to discharge, the data stored in the previous phase is sensed at the output Q_m. For example, after the storage of Data= '0', MTJ0 is in AP state and MTJ1 is in P state, the discharge current in right side is higher and will be finally pulled up to V_{dd} (logic '1') , while the output Q_m will be pulled down to the ground with the effect of the inverter. Inversely, the output Q_m will be pulled down to the ground (logic '0'). Figure 4.2 displays the waveform of 4T-2M writing and PCSA circuit. The design vectors are demonstrated in Table 4.1.

Table 4.1: Design parameters settings

Circuit	Description	Default Value
V_{dd}	Supply voltage of PCSA	1V
V_{ddh}	Supply voltage of writing circuit	1.6V
W_{min}/L_{min}	Minimum transistor dimension	80nm/30nm

4.1.1 Variability analysis of MTJ based circuits

With the scaling down of device size and requirement for low power, the tolerance against process variations becomes very important. Here, a variability analysis based on the PCSA circuit is carried out to investigate the dependence of reading error rate on area of MTJ and oxide barrier thickness t_{ox}. The results in Figure 4.3 show that both reducing the transistor size and enlarging the thickness of oxide barrier can sufficiently improve the reliability of hybrid CMOS/MTJ circuits [98].

4.1.2 Influence of MTJ stochastic switching behavior on MTJ/CMOS circuits

In order to validate the stochastic switching behavior in compact model, Monte Carlo simulations have been performed for a 2T-1M writing circuit. Figure 4.4 demonstrates that the switching probability increases with the growth of stress voltage and pulse width.

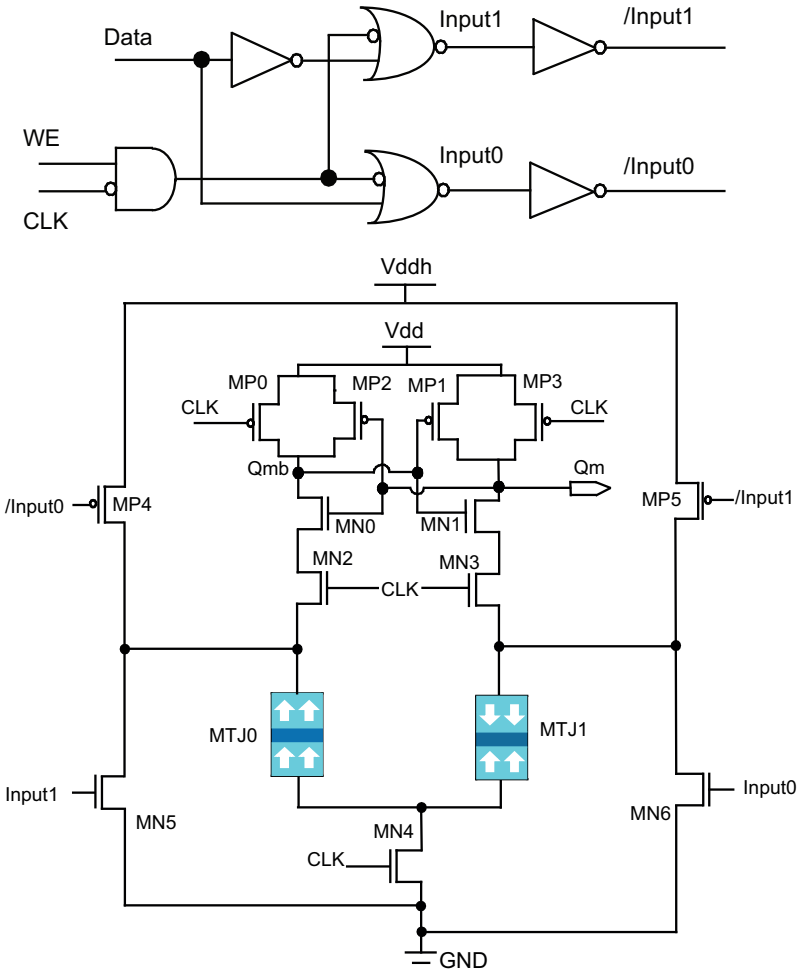


Figure 4.1: Architecture of pre-charge sense amplifier based STT-MRAM cell circuit proposed in [49]. It consists of two parts: writing control part and PCSA part.

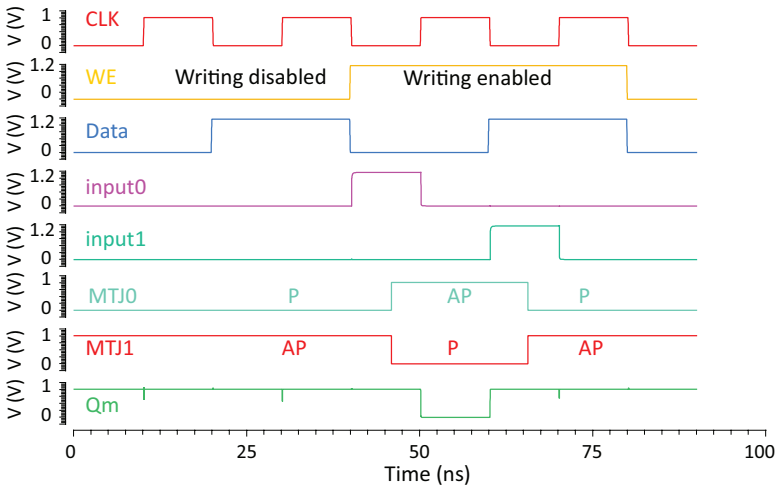


Figure 4.2: Transient simulations of 4T-2M writing circuit and PCSA circuit.

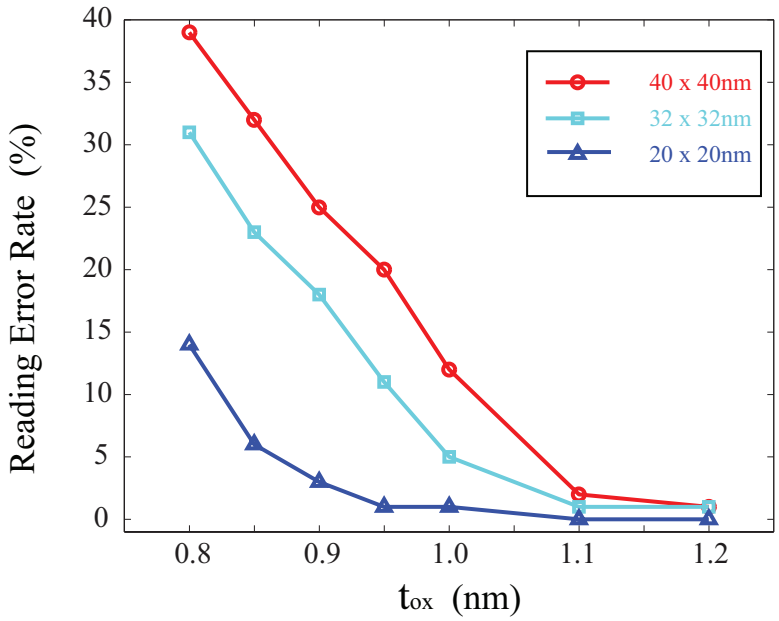


Figure 4.3: Dependence of reading error rate on the thickness of oxide barrier tox and area of MTJ.

Compared with Figure 3.7, the deviation is much higher in Figure 4.4 due to process variations of MTJ and transistors. The effective voltage across MTJ fluctuates due to the process variations of transistors and MTJ (resistance variation). Meanwhile, the average switching delay τ_{sw} floats due to the process variations of MTJ. With 1.4V stress voltage, 40 of 1000 samples are not successfully switched because of fast breakdown, thus leading to the final switching probability of 96%. This simulation can be used to find a tradeoff between operation frequency and power consumption of MTJ based MRAM.

4.1.3 Temperature impact on MTJ based circuits

In order to validate this model, we have conducted performance analysis including temperature dependence of switching delay, energy consumption using the conventional 1T-1M writing circuit [121]. As depicted in Figure 4.5, high temperature results in a faster speed and lower power writing process, but accelerates the dielectric breakdown of MTJ and leads to a shorter time to failure. Thus, there exists a tradeoff of design between writing performance (power consumption and frequency) and endurance in consideration of operation temperature condition.

For the purpose of investigating the temperature impact on reading circuit, a reliability analysis of the pre-charge sense amplifier (PCSA) circuit [49] has also been carried out. As the size of CMOS has an important impact on the error rate of this circuit, we have

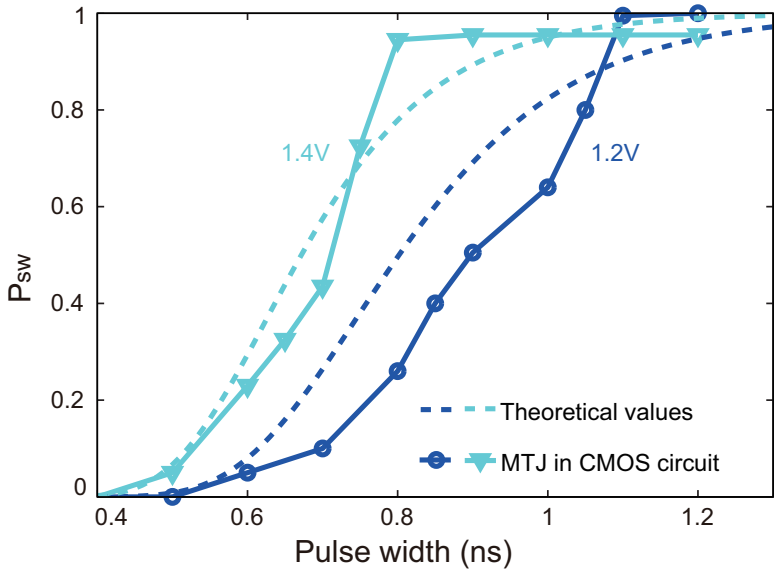
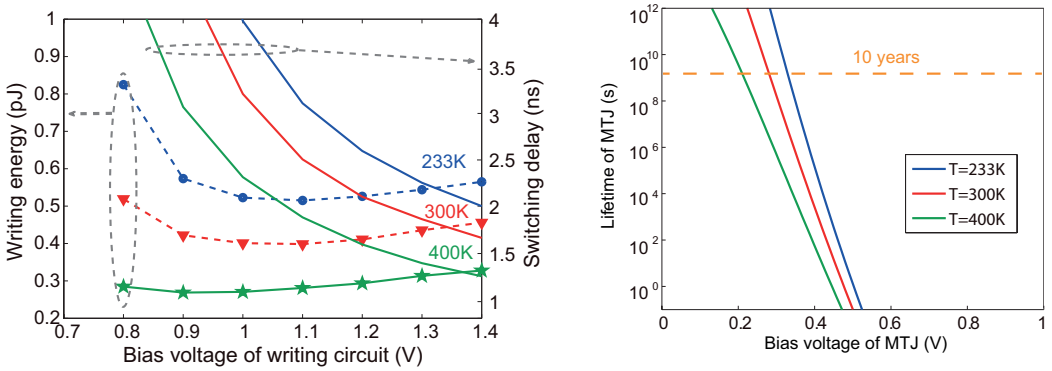


Figure 4.4: Switching probability with different writing voltages: the dashed lines are theoretical values plotted from equation (3.18) and the markers are statistical results of MC simulation.

changed their size to analyze the temperature impact on this circuit. Meanwhile, different stored data (0 or 1) has also been taken into account to ensure the conclusion. The result in Figure 4.6 demonstrates that reading error rate increases slowly with temperature due to the temperature dependence of TMR and CMOS. Because of the relatively low current flowing through MTJ and CMOS, temperature has very weak impact on reading behavior. A larger size of PCSA circuit can be efficient to ensure a reliable reading operation under high temperature operation conditions.

4.1.4 Ageing of MTJ based circuits

Figure 4.7 presents the breakdown probability distribution for theoretical case (dashed lines), the simulation results of MTJ under constant voltage (circles) and MTJ integrated in CMOS circuit (stars). While the theoretical case is obtained by deterministic equations without any imperfections, the two other cases are statistical simulation results in consideration of process variations. The circles show good agreement with lower deviation due to the impact of oxide barrier t_{ox} variation on breakdown behavior of MTJ (implied by (3.31) and (3.32)), while the stars have higher deviation as the process variations of MTJ and CMOS are both taken into account, in which the effective voltage across MTJ fluctuates around the indicated values (1.4V, 1.3V and 1.2V). Comparing the Figures 4.4 and 4.7, high stress voltage facilitates the switching but induces short time to breakdown.



(a) (b)

Figure 4.5: (a) Switching probability as a function of applied switching voltage and switching time. (b) Switching voltage versus average switching time at different temperature conditions.

Thus an optimum tradeoff of power design can be obtained according to the requirements of writing probability and endurance.

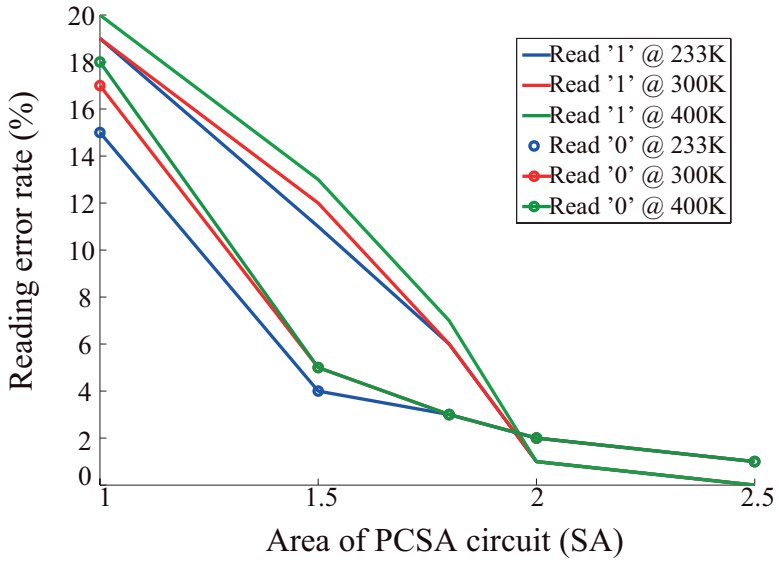


Figure 4.6: Reading error rate of PCSA with different area of circuit (SA is the minimum size of PCSA circuit) under different temperature conditions.

In order to have a complete understanding of hybrid MTJ/CMOS circuits, it is essential to take into account the dielectric breakdown of transistors. With continuous scaling down of MOS technology, the transistors may also suffer from oxide breakdown. Dielectric breakdown behavior has been extensively studied in nanometer digital circuits at different CMOS nodes, e.g., 45nm [122], 40nm [123, 124], 32nm [125] and 28nm [126, 127]. In

general, digital circuits firstly suffer from soft-breakdown (SBD), which is represented by performance parameter fluctuations/degradations. During the circuit lifetime, the probability to suffer SBD is increased. When these fluctuations or degradations are accumulated, hard-breakdown (HBD) can be induced as circuit functional failure. Oxide breakdown and its impact on memory cell (e.g., SRAM) was studied in [123, 128]. It is concluded that SRAMs are sensitive to SBD and the severity is dependent on the breakdown location.

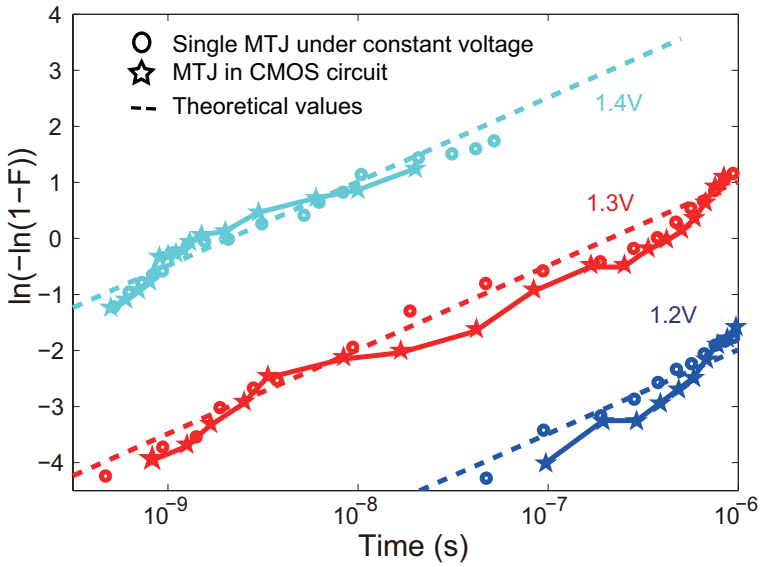


Figure 4.7: Cumulative breakdown probability distribution for theoretical case (dashed lines), the simulation results of MTJ under constant voltage (circles) and MTJ integrated in CMOS circuit (stars).

Figure 4.8 shows the schematic view of an NV-MFF with multiplexing sense amplifier (pulse generator followed by slave latch) topology of which the breakdown analysis is performed [129]. It consists of two non-volatile MTJs, a sense amplifier (SA), a differential write block and a SR latch stage. We try to find the weakest spot where breakdown may firstly occur.

The breakdown severity is highly dependent on stress condition. The writing circuit is sensitive to SBD because of MTJ operation current requirement (In order to ensure write success, transistor W/L ratios are significantly increased). Due to analogical characteristic of sense amplifier, its SBD behaviors induce MFF timing performance degradation, whereas other digital parts suffer output level degradation.

Figure 4.9 shows maximum SBD gate current density for main transistors in MFF circuit. Their HBD occurrence threshold is lower than sense amplifier and MTJ writing circuits. Notice that in SR latch, the less sensitive MOS transistor is connected to output

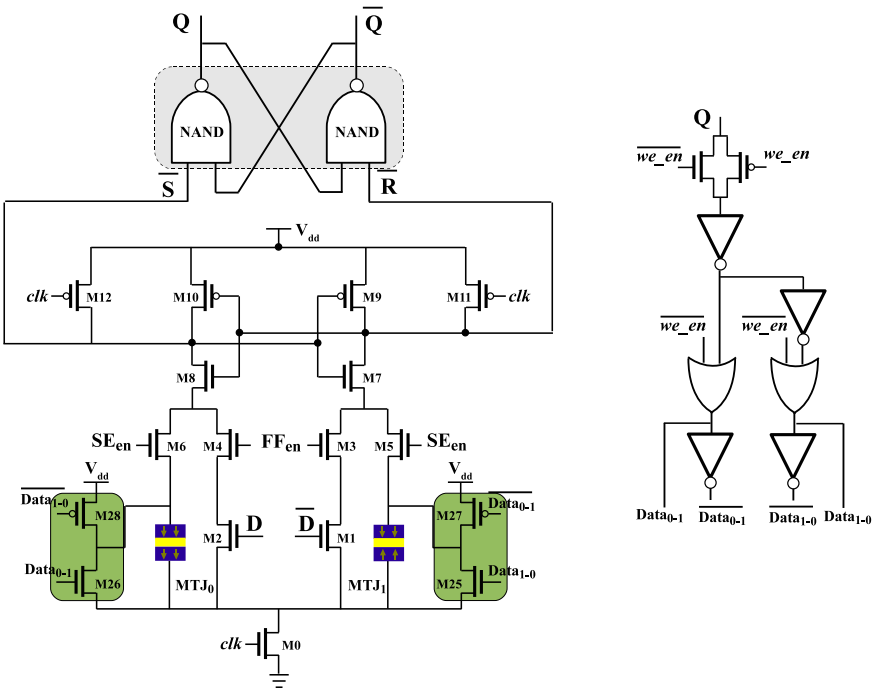


Figure 4.8: The studied symmetrical MFF is composed of two parallel MTJs, the writing block, a clocked sense amplifier, NAND-based slave SR latch and feedback loop.

signal $Q/\bar{Q}$ (the other one is with sense amplifier output). Final HBD condition can be proceeded by progressive SBD. The breakdown condition is sensitive to stress and transistor characteristics. We find that logic circuits, e.g., SR latch and feedback loop are sensitive to increased gate current. The weakest link characteristic of oxide breakdown is validated in NV-MFF circuits that any building block suffers HBD, the entire circuit functionality is failure.

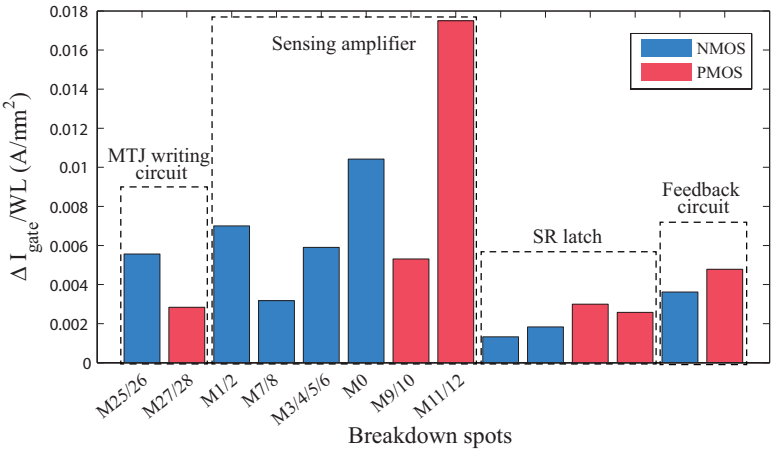


Figure 4.9: HBD failure gate current density: the breakdown sensitivity of transistor in MFF circuit.

4.1.5 Application of non Monte-Carlo Methodology in hybrid MOS/MTJ Circuits

Based on the worst-case corner models of STT-MTJ and transistors, we proposed a non-Monte-Carlo methodology for variability analysis of spin transfer torque (STT) magnetic tunnel junction (MTJ) based circuits. The proposed methodology is integrated into the 1 transistor-1 MTJ (1T-1M) memory array [130], pre-charge sense amplifier (PCSA) based STT-MRAM cell [49] and magnetic full adder circuit [46] to validate its functionality. The circuits are implemented with a 28nm fully depleted silicon on insulator (FDSOI) design kit, both statistical and worst-case compact models of MTJ and FDSOI transistors are considered in the simulation [131]. As simulation speed is the key improvement, it should be mentioned that all of the simulations are performed on the same machine with AMD Opteron quad-core processor at 2.3 GHz and 16-GB memory.

4.1.5.1 Switching delay and time to failure estimation of 1T-1M memory array

As shown in Figure 2.11, a typical STT-MRAM cell is composed of an MTJ connected in series with a CMOS transistor. In this section, the conventional performance of 1T-1M memory cell is evaluated using different modeling approaches.

Despite excellent potential in STT-MTJ, the switching of MTJ has been revealed intrinsically stochastic due to thermal fluctuation of magnetization [83, 65]. As a result, the switching delay of MTJ is not a deterministic value but follows a statistical distribution. Because of this phenomenon, write errors might occur with insufficient writing current or short writing pulse, while unexpected switching may happen in sensing operation [83]. Thus, a relatively high current density is always used in most of the memory designs to guarantee reliable writing operation [110]. This may generate high electric field across MTJ oxide barrier and induce significant self-heating effect [131]. Consequently, dielectric breakdown of the ultra-thin (∼1nm) oxide barrier in MTJ may occur, which leads to functional errors of hybrid CMOS/MTJ circuits. It is necessary to evaluate the switching performance and the lifetime of MTJ.

The simulation results of switching delay and time-to-failure in memory arrays are displayed in Figure 4.10. Note that F in the Weibull function signifies the breakdown probability of oxide barrier under voltage stress as function of time. It is demonstrated that most of the elements can be covered by the worst-case corners. Thus, it is possible to replace the statistical simulation (1000 runs cost 1800s) by worst-case analysis (3 runs of simulation cost 7.2s) and thus obtain the design margin of MTJ-based memory arrays.

The model can be used to find efficiently the optimized tradeoff between the performance such as switching speed and the reliability such as time-to-failure in the memory designs.

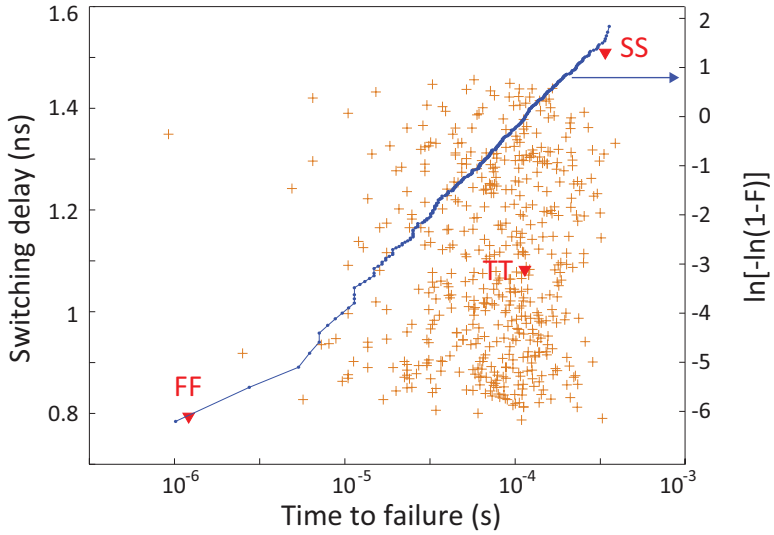


Figure 4.10: Switching delay and time to failure of memory arrays: the cross are from 1000 elements of memory arrays with statistical model; the blue dot line is the Weibull function of the cross in which F signifies the failure probability of the memory elements; the triangles are from 1 element with the worst-case model (σ=0.01 and n=3).

4.1.5.2 Variability-aware energy-delay analysis of PCSA based STT-MRAM cell

Except 1T-1M structure, there exist also many other structures, such as differential pair type which behaves improved robustness and design degrees of freedom [49, 132]. This section integrates the proposed model into the pre-charge sense amplifier based MTJ memory cell proposed in [49].

Dynamic power and circuit latency are evaluated for both writing and sensing operations using worst-cases models and statistical models of MTJ and transistors. The parameter settings of worst-case models are listed in Table 4.2. For instance, the worst case power (FF) of entire circuit signifies the best speed, which needs the best speed of all the devices (FF). Meanwhile, the FS corner favors the writing as well as sensing of '0' but degrades the speed of writing and sensing of '1', which is more complicated. From the view of writing, *input*0 should be more easily written in MTJ than *input*1. Thus, the current generated by input0 is more significant, leading to $R_p(MTJ0) + R_{ap}(MTJ1) < R_{ap}(MTJ0) + R_p(MTJ1)$. This can be realized by using SS corner for MTJ0 and FF corner for MTJ1. For the sensing part, the performance depends mainly on the resistance difference (RD) of the two branches. Then, sensing '0' should have larger RD value than

sensing '1', i.e., $R_{ap}(MTJ0) - R_p(MTJ1) > R_{ap}(MTJ1) - R_p(MTJ0)$. This can be maximized by setting SS corner for MTJ0 and FF corner for MTJ1.

Table 4.2: Worst-case corners setting of transistor and MTJ models for worst-case performance analysis of STT-MRAM cell

Worst-case performance	Worst case corners of devices model		
	MOS	MTJ0	MTJ1
Power (FF)	FF	FF	FF
Speed (SS)	SS	SS	SS
'1' (FS)	FS	SS	FF
'0' (SF)	SF	FF	SS

Figure 4.11 shows the performance of power and delay for the two operation phases of STT-MRAM cell. Most of the statistical results are distributed in the area fixed by worst-case corners, i.e., worst-case analysis can be used to estimate completely the circuit performance. Furthermore, worst-case analysis takes 28.2s while statistical analysis takes 5195s. The outside values can also be covered by assigning a higher value for n. The design vectors (bias voltage, devices parameters, and correction block) can be modulated to achieve an optimized tradeoff of performance terms. For instance, if the writing operation of the worst-case speed fails, the current should be reinforced by increasing the bias voltage or scaling device.

4.1.5.3 Worst-case analysis of magnetic full-adder dynamic performance

MTJ-based full-adder has been proposed for the first time in [46], as shown in Figure 2.13. The proposed model is applied into this circuit to investigate the circuit performance. As the delay time and dynamic energy are generally two crucial parameters to evaluate the performance of computation system, we have performed the simulation under different sizing conditions. Note that the dimension of discharge transistor determines the discharge current which is critical for sensing performance. Figure 4.12 illustrates the simulation results with different dimensions of discharge transistor. From both simulation methods, it can be concluded that larger discharge transistor size drives faster sensing operation while consuming more energy. This model can be used to obtain the best device dimensions for certain design specification. Note that 1000 runs of MC simulations cost 3267 seconds and the worst-case analysis costs 14.45 seconds. The simulation speed is improved to 226x.

4.1.5.4 Results discussion

From comparison of the results presented above, it can be identified that the value of worst-case distance n (number of σ) is different between the two PCSA based circuits. After careful study of the circuit structures and device dimensions, the difference is probably generated from the dimension of the transistors.

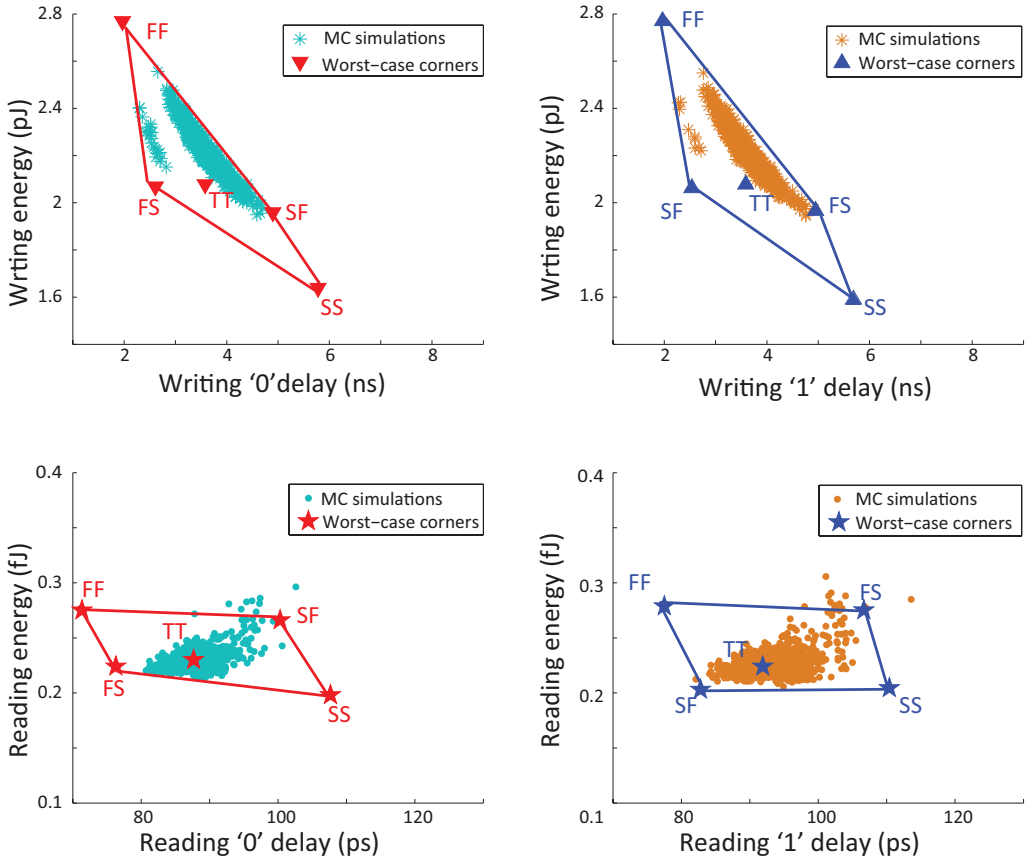


Figure 4.11: Writing and sensing performance of STT-MRAM cell: the stars and dots are from statistical model (1000 MC simulations); the frames are from worst-case model of MTJs (σ=0.01 and n=4.5) and CMOS transistors.

Further study has been carried out to verify this suppose. We performed simulations with scaling the circuit size to investigate the model precision. The results are demonstrated in Table 4.3. $A1$ signifies the area of circuit with default values and the percentages represent the fraction of the statistical results which are covered by the worst-case corners. It can be deduced that the precision is increased with the growth of n and the transistors dimension. This is consistent with the conclusion drawn in [133]. The circuit performance becomes more sensitive to process variations with CMOS technology scaling down. With the growth of n value, the area covered by the worst-case corners is enlarged, resulting in the increasing number of cases included.

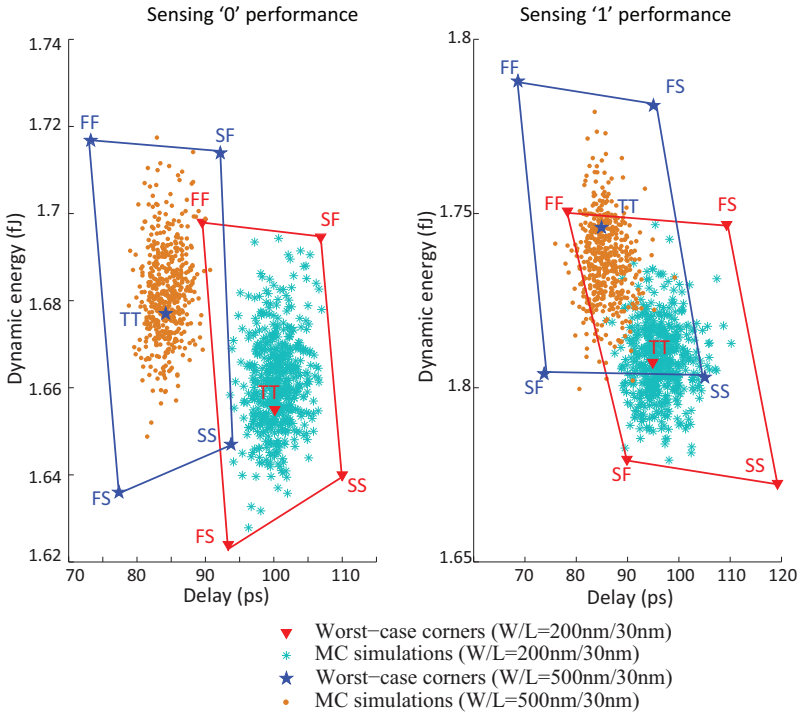


Figure 4.12: Performance of delay time and dynamic energy in MFA circuit: the stars and dots are from statistical model (1000 MC simulations); the frames are from worst-case model of MTJs and CMOS (σ=0.01 and n=3). Two different discharge transistor sizes are considered: W/L=200nm/30nm and W/L=500nm/30nm.

4.2 Reliability-aware design of MTJ-based circuits

With continuous scaling of MTJ and transistors, the impact of process variation on the functionality, performance and reliability of non-volatile circuits and systems becomes increasingly severe [83, 134]. Thus, intensive attention has been driven to variability induced performance degradation for circuit designers [135]. In this section, a circuit design is proposed to improve the robustness to variability.

Symmetrical architecture is widely used in the design of non-volatile flip-flop (NVFF) owing to its high immunity to read disturbance [49, 135, 136, 137]. However, its circuit performance is significantly impacted by the process variation. Based on these symmetrical structures, we propose a methodology of dynamic asymmetrical body bias (DABB) to optimize the variability immunity of NVFF [138]. This methodology is implemented with a 28nm planar ultra thin body and buried oxide (UTBB) fully depleted silicon on insulator (UTBB-FDSOI) design kit, both statistical and worst-case compact models of MTJ and FDSOI transistors are considered in the simulation.

Table 4.3: Model precision on function of n and transistors size

Worst-case distance	Area of circuit		
n	A1	2*A1	4*A1
3	97.9%	99.0%	99.2%
3.5	98.5%	99.0%	99.3%
4	98.8%	99.2%	99.6%
4.5	99.2%	99.3%	99.6%

4.2.1 Transistors with UTBB-FDSOI technology

Shrinking horizontal (gate length) and vertical (gate dielectric thickness) device parameters as well as increasing channel doping concentration are no longer practical beyond 28nm with bulk-MOSFET technology [54]. Thus, traditional bulk-MOSFET cannot obtain expected benefits from technology scaling down. In order to overcome scaling down limitations beyond 28nm node, UTBB-FDSOI technology has been proposed and validated which brings power-speed improvement to integrated circuits (ICs) [139, 140, 141].

Figure 4.13 shows the thin film devices with FDSOI technology. Transistors with UTBB FDSOI technology are designed with high-κ metal-gate (HKMG) dielectric stacks. In 28nm FDSOI technology, this HKMG material has been used to replace Poly-Si-SiO2 in dielectric layer of transistor, with an equivalent oxide thickness equals to 1.35nm [124] [127]. Low channel doping in FDSOI technology reduces band-to-band tunneling and increase the source/drain breakdown voltage [142].

FDSOI process offers better tolerance of very low channel doping concentration, so that random dopant fluctuations (RDF) can be significantly reduced [141],[143]. By using the flexible body bias, circuits in FDSOI technology can achieve tradeoff between performance and variability. Moreover, as the effects of reliability issues, circuit performance parameters are fluctuated or degraded, even functional failures may emerge. For instance, aging mechanisms of deep sub-micron MOS transistor include negative bias temperature instability (NBTI), hot carrier injection (HCI), time dependent dielectric breakdown (TDDB) and electromigration (EM) [144, 145]. The NBTI and HCI can cause the generation of the interface traps which result in transistor parameters shift over time (e.g., mobility, V_{th}). It has been reported that aging induced degradation can be well alleviated by transistors with FDSOI forward body bias [146].

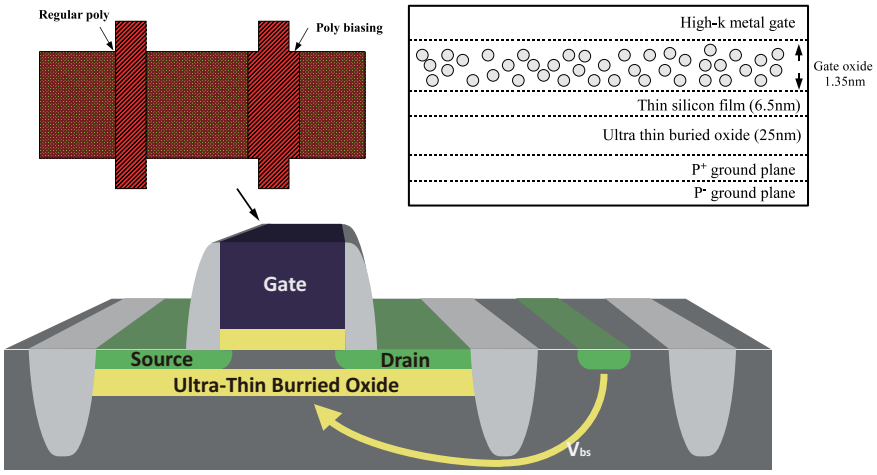


Figure 4.13: The thin film devices in FDSOI technology with a cross-section view of planar/2D structure FDSOI CMOS. Body bias voltage can impact transistor performance. Poly bias is achieved by additional gate length.

Transistors with UTBB FDSOI technology have an extended body bias range than traditional bulk-MOSFET. The back gate bias is provided by other supply point, e.g., 1V forward biased NMOS is with a 1V at transistor body (normally 0V). Forward body bias used in back-gate can decrease transistor threshold voltage (V_{th}) and increase transistor drain current (I_d). Thus, transistor performance is improved (e.g., drain current, switch speed), whereas the reverse body biased (RBB) transistor achieves part of performance-robustness tradeoff [143]. As shown in Figure 4.14, when transistor works in saturation, V_{th} variation is independent of varied body bias. Considering the coefficient of V_{th} variation (the ratio between standard deviation and mean), it is three times higher in transistors with FBB than those with RBB. However, due to the amplifying of the variation in gate overdrive through the RDF caused fluctuation [143], I_d fluctuation is increased with FBB (2V body bias), whereas RBB can reduce 24.3% of I_d variability compared to FBB. This property can be appropriately combined with MTJ-based symmetrical circuits to explore optimization in performance and reliability [117, 135, 147].

4.2.2 Circuit Design of non-volatile Flip-Flop using dynamic asymmetrical body bias of FDSOI

The methodology is implemented with the circuit demonstrated in Figure 4.15. The writing control part is the same as illustrated in Figure 4.1. We focus on the PCSA circuit in which all the transistors are with minimum size (W/L=80nm/30nm). The resistance

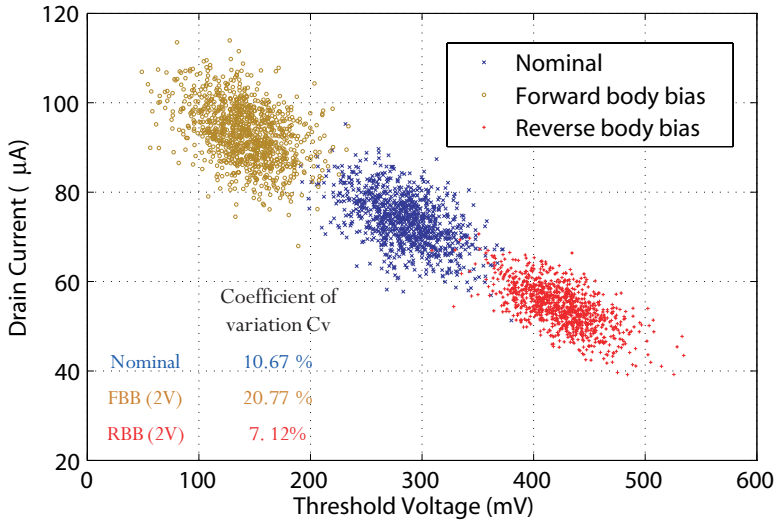


Figure 4.14: Variability FDSOI: A single NMOS transistor works in saturation region. The coefficient of V_{th} variation is analyzed among FBB, nominal design (no body bias) and FBB.

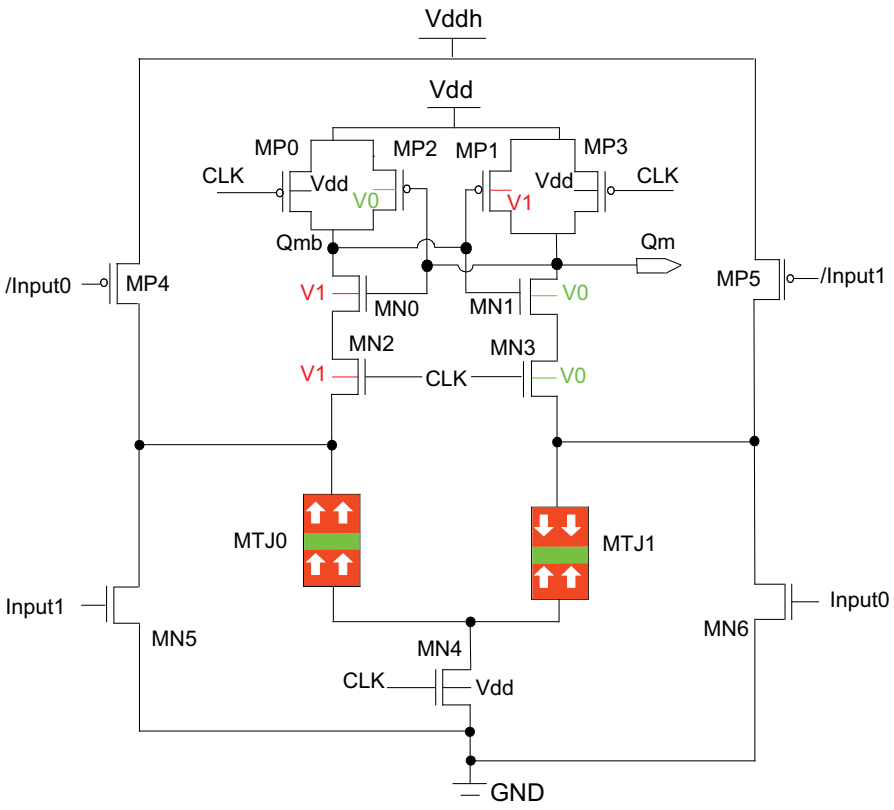
difference of the two sides is critical to the reading performance. In order to enlarge the resistance difference, two RC circuits are inserted to generate bias voltages for the two branches of PCSA. During the writing phase, only one of the two bias voltages: V_0 or V_1 is charged to V_{dd}. For instance, after writing '1', V_1 is charged to V_{dd} and V_0 is discharged to the ground. During the sensing phase, with the bias voltage V_0 and V_1, the resistance of transistors in MTJ1 side are reduced while that in MTJ0 side are not changed. Consequently, the resistance difference of the two sides is enlarged, resulting in better sensing performance.

4.2.3 Reliability analysis and performance evaluation

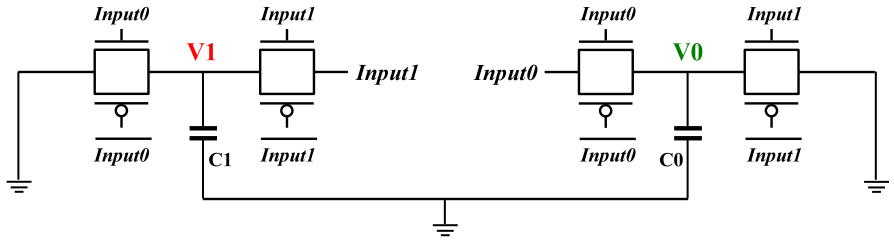
Figure 4.16 shows the waveform of the proposed NVFF. During the first two cycles, the output Qm keeps constant while writing is disabled. From the third cycle, writing is activated and V0 is charged to Vddh during the writing phase. Meanwhile, the output Qm obtained during the sensing phase is identical with the Data input during writing phase.

From the circuit operation mechanism detailed above, it can be deduced that TMR value is critical for the performance of symmetrical sensing circuit. TMR is determined by many factors, e.g., the proportion of different materials, the purity, surface roughness and size of FM layers. The fabricated MTJ usually has a wide range of TMR value. Thus, it is essential to study the circuit performance corresponding to different TMR

value. Figure 4.17 shows the sensing error rate as function of TMR. Nominal body bias (NBB) means all the substrates of NMOS are connected to ground and those of PMOS are connected to Vdd. It is obvious that the proposed circuit with asymmetrical forward body bias (AFBB) can improve the sensing success probability with different TMR value.



(a)



(b)

Figure 4.15: (a) Pre-charge sense amplifier with dynamic asymmetrical bias bias (b) RC circuits generate the body bias voltages for transistors in PCSA.

Monte-Carlo simulations have been performed to investigate the reliability of the proposed circuit taking into account process variations, voltage scaling and temperature fluctuation.

Figure 4.18 demonstrates the reading error rate of PCSA circuit in different conditions. It shows that reading errors have been almost removed by the proposed method with

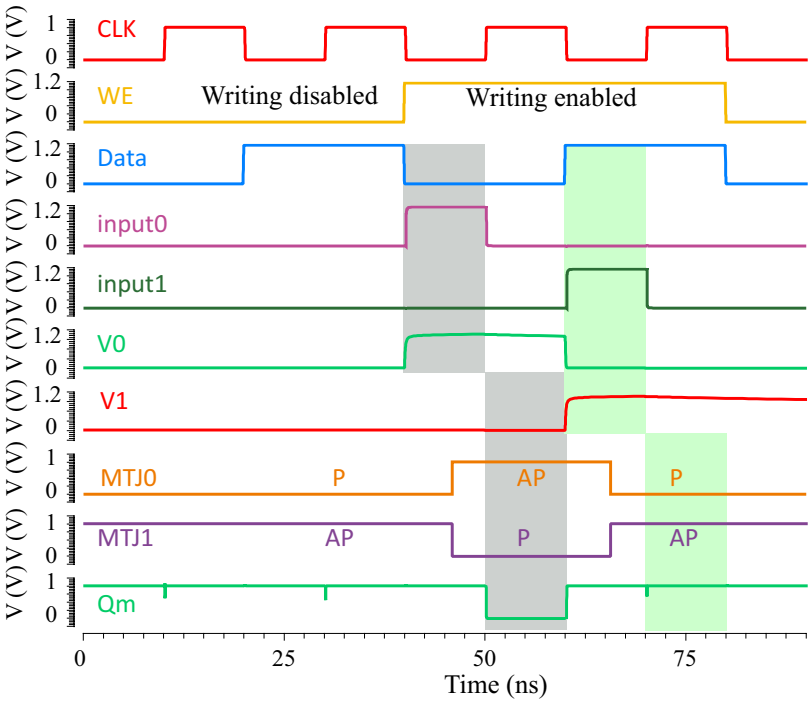


Figure 4.16: Waveform of proposed Non-volatile flip-flop circuit using dynamic asymmetrical body bias.

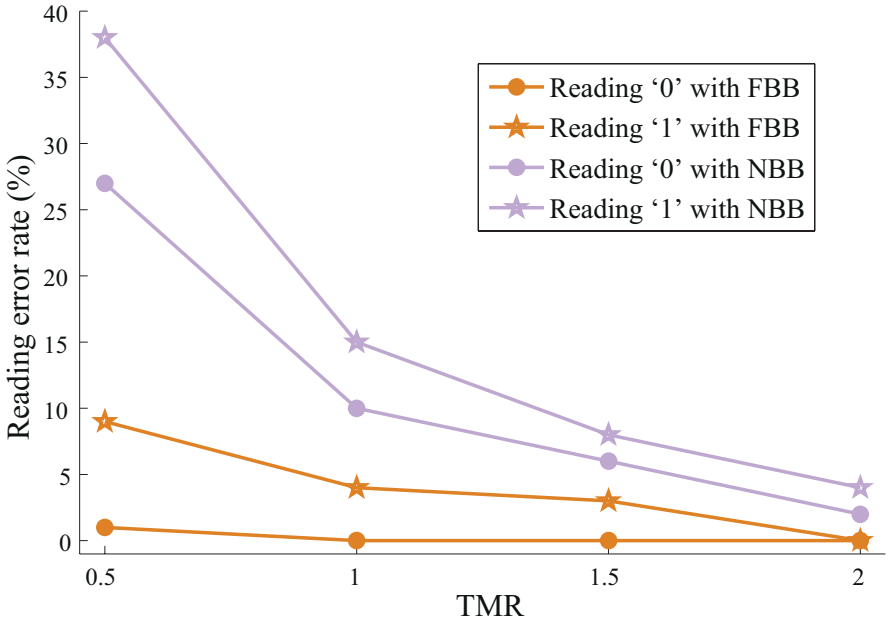


Figure 4.17: Reading error rate of the proposed NVFF versus different TMR value: FBB means dynamic asymmetrical body bias, and NBB means nominal body bias.

AFBB. As the resistance difference of the two branches is enlarged, the variability can be partly masked by the sufficient current difference. Figure 4.19 displays the reading error rate with supply voltage scaling. With Vdd superior to 0.8V, the reading error rate is independent to the supply voltage, because higher supply voltage is no more efficient for increasing the current difference between the two sides. This property is efficient in low power design.

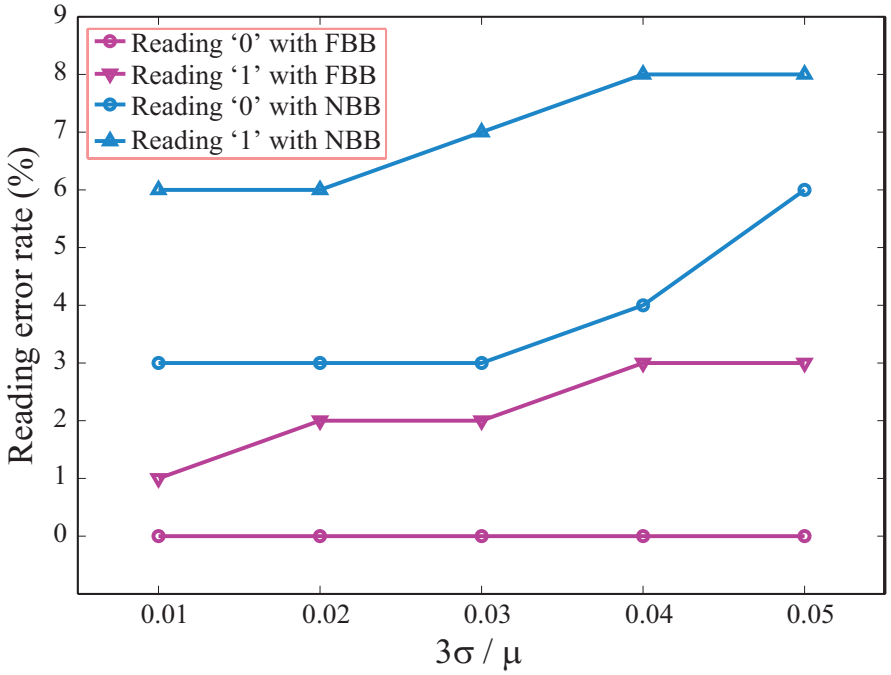


Figure 4.18: Reading error rate of the NVFF versus process variations: MTJ parameters t_{sl} , t_{ox}, TMR follow normal distribution around the mean value μ with the deviation σ.

As the value of TMR is very sensitive to thermal condition, the circuit with NBB is severely influenced by temperature fluctuation. Figure 4.20 presents the reading error rate versus different thermal conditions. With DABB, the sensing failure is nearly independent of the temperature fluctuation. As capacitors are also highly temperature dependent, the body bias voltage values V_0 and V_1 change with temperature. The effect of changes in V_0 and V_1 on transistors has almost compensated the impact of temperature on TMR. Consequently, the reading error rate of proposed circuit is nearly immune to temperature variation.

In order to estimate the cost of this method, we have carried out a comparison concerning layout area, power consumption and flip-flop latency. As shown in Figure 4.18, it can be deduced that the variability of transistors has more important impact on sensing reliability than MTJs. Thus, the conventional method to guarantee the sensing success is

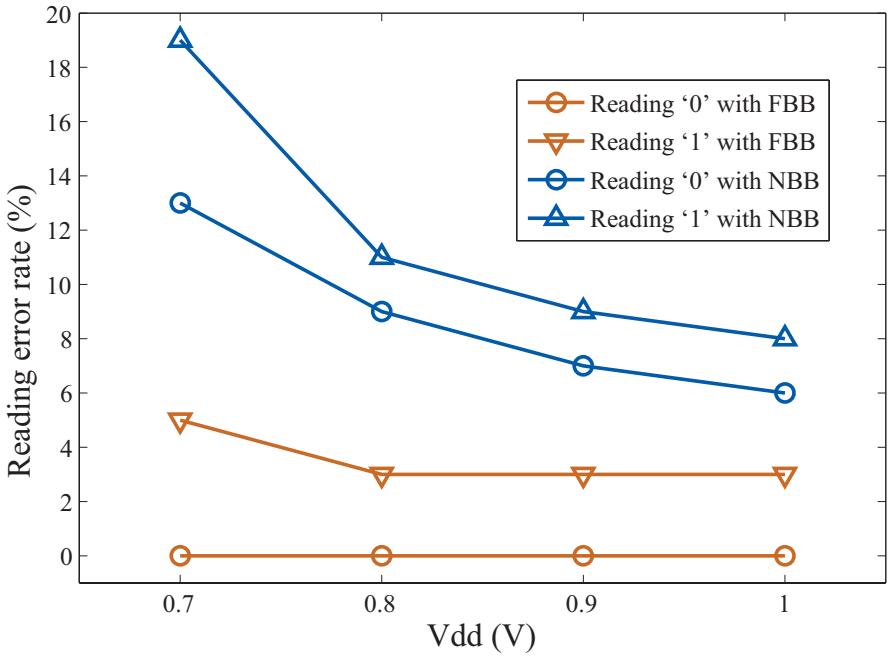


Figure 4.19: Reading error rate of the proposed NVFF versus different supply voltage: FBB means asymmetrical forward body bias, and NBB means nominal body bias.

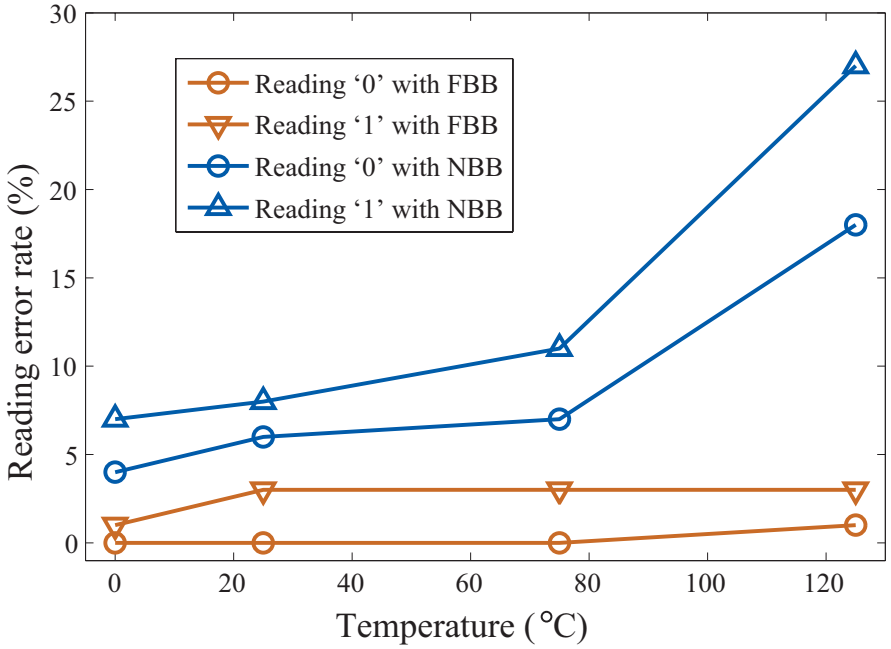


Figure 4.20: Reading error rate of the proposed NVFF versus different thermal conditions: FBB means asymmetrical forward body bias, and NBB means nominal body bias.

Table 4.4: Comparison of performance between proposed methodology (DABB) and conventional method (NBB)

	NBB	DABB
Writing power	131.54 μW	132.63 μW
Writing delay	1.66 ns	1.66 ns
Sensing power	54.4 nW	50.2 nW
Sensing delay	91.5 ps	82.7 ps
Area of PCSA	1.42 A_m	1.21 A_m

increasing the size of transistors MN0-MN4 and MP1-MP2. With increased size of these transistors, the resistances are all reduced. This further increases the total current flowing through the two sides at the beginning of sensing phase. Consequently, the current difference is largely increased, which yields better sensing performance. We define A_m as the minimum size of PCSA circuit in which all the transistors are at the minimum size (W/L=80/30 nm). Various sizing conditions have been attempted to find the error free PCSA circuit. The simulation results have shown that the minimum area for an error free PCSA circuit is estimated to be 1.42 A_m, whereas the proposed architecture with asymmetrical forward body bias voltage generating circuit is estimated to be 1.21 A_m. With this design vector, the characteristics of operation speed and energy consumption are listed in Table 4.4. As the writing control part is not changed, the writing power indicates the part in Figure 4.15. It is shown that the proposed method has better sensing performance, at the expense of more writing power.

This methodology has been implemented into non-volatile full adder [148], which has successfully optimized the sensing circuit performance including circuit latency, dynamic power, variability and sensing probability. It proves the applicability and efficiency of this methodology.

4.3 Conclusion

This chapter carried out overall reliability analysis of MTJ/CMOS circuits to validate the functionality of compact models developed in the previous chapter. The analysis is based on the basic STT-MRAM cell and arithmetic unit. For the writing operation, we have found a tradeoff between the power, operation speed and endurance. To ensure

the successful writing, a sufficient bias voltage should be applied, which will lead to high power consumption and fast dielectric breakdown. For the reading operation, the reading error rate can be influenced by many factors, such as the size of MTJ, supply voltage, environmental temperature (PVT), etc. Thanks to the intrinsic advantage of PCSA, there is nearly no read disturbance because the sensing operation terminates in less than 1 ns which is too short for MTJ switching. As the most wide application of MTJ is in memories, reading operation is much more frequent than writing. Thus, it is very important to improve the robustness of reading circuit to PVT variations.

Then, we proposed a methodology to alleviate the impact of PVT variations on the performance of MTJ based applications. The methodology is presented by carrying out a novel design of PCSA circuit using dynamic asymmetrical body bias (DABB) of transistors with fully depleted silicon on insulator (FDSOI) technology. Simulation results show that the sensing errors have been almost removed by this method with the minimum size of circuit. In addition, the thermal robustness of this circuit has also been dramatically improved. This methodology can be applied to other symmetrical circuits, which has already been proven in [148].

Moreover, a non Monte-Carlo Methodology in hybrid MOS/MTJ Circuits was proposed to validate the functionality of the fast simulation model. The worst-case corners model is applied into 1 transistor-1 MTJ memory array, PCSA based STT-MRAM cell and magnetic full adder. Different performance parameters are investigated according to the design specifications of each circuit. The simulation results are compared with the model using MC method, demonstrating that this methodology can drastically improve the simulation efficiency with a perfect evaluation quality.

Chapter 5

Novel applications of MTJ in conventional circuits

Although the circuits function of memory and computing chip may be bothered by the reliability issues, some other applications can appropriately profit from these issues. This chapter focuses on the novel realization of the conventional applications using stochastic switching behavior.

5.1 A novel circuit design of MTJ based true random number generator

Random numbers are widely used in the cryptography and security systems. However, most of the true random number generators (TRNG) which use physical randomness are with high complexity and high power consumption. This section proposes a new TRNG circuit using magnetic tunnel junction (MTJ) [149]. As one of the reliability issues in MTJ based circuit, the stochastic switching behavior provides a perfect physical source of randomness.

5.1.1 Traditional true random number generators

Random numbers are always necessary in many traditional areas, e.g., Monte Carlo simulations, cryptography, statistical sampling and many other security applications [150, 151]. Moreover, with the rapid development of digital ecosystems, random numbers are essential for the security of online transactions and mobile applications. Thus, algorithms with fast speed, low power and high reliability are required to design true random number generators (TRNG). Physical randomness is usually used as entropy sources in conventional

TRNG, such as thermal noise, metastability, and oscillator jitter [152]. However, these TRNGs all require extensive post-processing to guarantee a high level of random output, which degrades the performance in terms of speed, power, and area [153]. Consequently, it is urgent to explore new methods of generating random numbers for low complexity design, compact area, high randomness, and reliable operation.

The stochastic behavior of emerging non-volatile devices have been considered as a promising physical noise source for TRNGs [154, 155]. With the rapid development of non-volatile devices, many novel TRNG designs have been proposed by using spin dice [156], memristor [157], and magnetic tunnel junction (MTJ) [158, 159, 160, 161]. Compared with the conventional CMOS based TRNGs, the magnetic devices based technology designs can effectively achieve simplified structure, more compact area, higher speed and better energy-efficiency. However, the process variations of MTJ and transistors have not been taken into account in these designs, thus the robustness remains doubtful. As the exact probability is critical for TRNG, it is essential to guarantee its variability awareness. The next parts will introduce the novel circuit design of TRNG, validate the design by performing transient simulation and evaluate the design by comparing with other work in terms of performance and variability awareness.

5.1.2 Circuit design of true random number generator using MTJ

The MTJ stochastic switching provides a new randomness source for TRNG. Based on unpredictable physical phenomenon, it can supply real random bitstreams by special circuit design. As shown in Figure 3.7, continuous switching probability can be obtained by tuning either the applied current or the stress time. A tunable switching current I_{sw} with 5ns pulse is applied in this work to investigate the switching probability. The simulation result presented in Figure 5.1 indicates that $I_{sw} = 84.5\mu A$ is required for a switching probability of 50%. Our compact model is based on a symmetrical MTJ (the critical current for P to AP is the same with that of AP to P).

With continuous scaling down of semiconductor devices, performance degradation induced by process variation becomes a critical issue in CMOS circuits and systems design [162]. Meanwhile, the limited fabrication precision of MTJ induces variable device parameters like the oxide barrier thickness t_{ox}, free layer thickness t_{sl} and TMR ratio [61]. In order to guarantee an accurate probability and high level of randomness, it is necessary to take into consideration the process variations.

In the considered 28nm FDSOI design kit, process variability is modeled by four worst-

case corners [163]. TT is the typical compact model extracted from the "golden die" of the "golden wafer" representing the center-line process technology. On the one hand, the corner models are generated from slow nMOS and slow pMOS (SS) to model the worst-case speed, and from fast nMOS and fast pMOS (FF) to model the worst-case power. On the other hand, the corner models are generated from fast nMOS and slow pMOS (FS) to model the worst-case '1', and from slow nMOS and fast pMOS (SF) to model the worst-case '0' [114]. The worst-case corner models offer for designers capability to simulate the pass/fail results of a typical design and are usually pessimistic. The process variations of MTJ are also integrated in the compact model [98]. Using these models, two current values of worst cases (maximum(FF) and minimum (SS)) are obtained for the switching probability of 50% (shown in Figure 5.1).

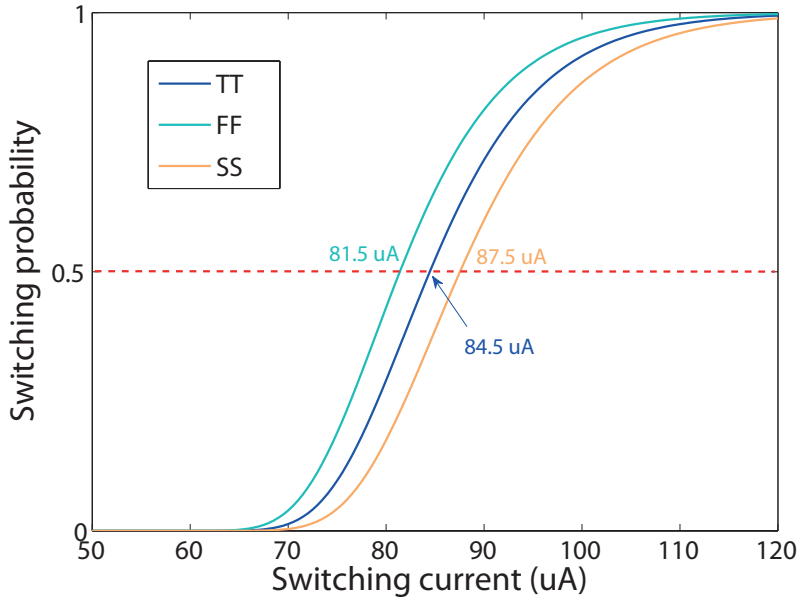


Figure 5.1: Switching probability of MTJs on function of switching current with 10ns pulse. The current with 50% switching success is indicated above. This figure is obtained by 1000 runs of Monte-Carlo simulation with the same MTJ under voltage pulses.

With the obtained switching current, we propose a novel design of TRNG circuit. The general architecture is illustrated in Figure 5.2. It is composed of MTJ random writing part, pre-charge sense amplifier (PCSA) and correction block. By an appropriate choice of transistors dimensions, a particular switching probability can be obtained to get a real random bitstream. In order to improve the reliability, a correction logic block composed of counters and comparator is implemented. This block generates a control signal to modulate the switching current, which guarantees the exact given probability of obtained random number bitstream (ideally with 50% of '1' and 50% of '0').

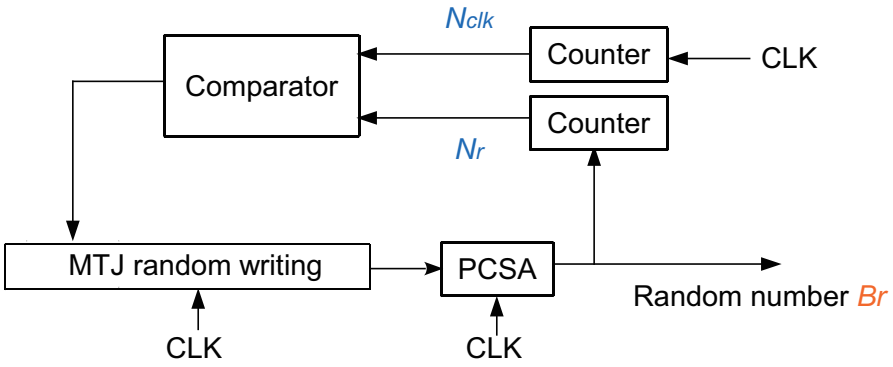


Figure 5.2: Architecture of proposed MTJ-based true random number generator: The random writing circuit generates a switching current to write the MTJs with 50% success and it is controlled by the correction block; The MTJ writing part enables MTJ switching (generating random number or resetting to initial state); The correction block composed of counter and comparator is used to execute real-time output probability tracking and send feedback to writing block.

The Analog part of writing and sensing blocks are displayed in Figure 5.3. The voltage-current converter (in the blue frame) generates a writing current for MTJ according to the random number probability obtained in the previous clock cycle. The transistors are configured to guarantee a switching probability of P_{sw}=50%. This circuit operates in three phases:

1) *Reset phase*: CLK='0', P_6 and N_6 are open while all of other transistors are closed. The two MTJs are switched to the initial state (MTJ is with low resistance) with a relatively high current (usually $>3I_{c0}$ which guarantees 100% switching probability).

2) *Random writing phase*: CLK='0', P_5 and N_5 are open to enable the MTJ switching with a certain current according to the required probability. When N_{c1} is open and N_{c0}, N_{c2} are closed, the switching probability of MTJ in the next step will keep 50%. If $P_{sw} > 50\%$, N_{c0} is open while the two others are closed to reduce writing success (with I_{sw}=81.5μA), resulting in decreased random probability. If $P_{sw} < 50\%$, N_{c2} is open and N_{c0}, N_{c1} are closed, thus the MTJ is more easily switched with a higher current flow (with I_{sw}=87.5μA). Consequently, the random probability will increase until 50%. During the first two phases, the sensing circuit is always at pre-charge phase and both outputs of PCSA (Q_m and $\bar{Q_m}$) are charged to V_{dd}.

3) *Sensing phase*: CLK='1', N_0, N_1, N_2, N_3, N_4 are open to drive sensing current flowing to the ground through MTJ0 or MTJ1. With resistance difference between the MTJ and the reference resistor (R_{ref}=(R_p+R_{ap})/2), the unbalanced current generates

different discharge speeds. The lower resistance side discharges more quickly, and its output (Q_m or $\bar{Q_m}$) voltage will be pulled down to the ground, whereas that of the other branch will be pulled up to V_{dd}. Thus, the random number is obtained at the output. In order to guarantee the right probability, the random number bitstream is evaluated using counters and comparator. The evaluation result will generate a control signal which tunes the writing current in the next switching phase. The detailed phase transition diagram is illustrated in Figure 5.4.

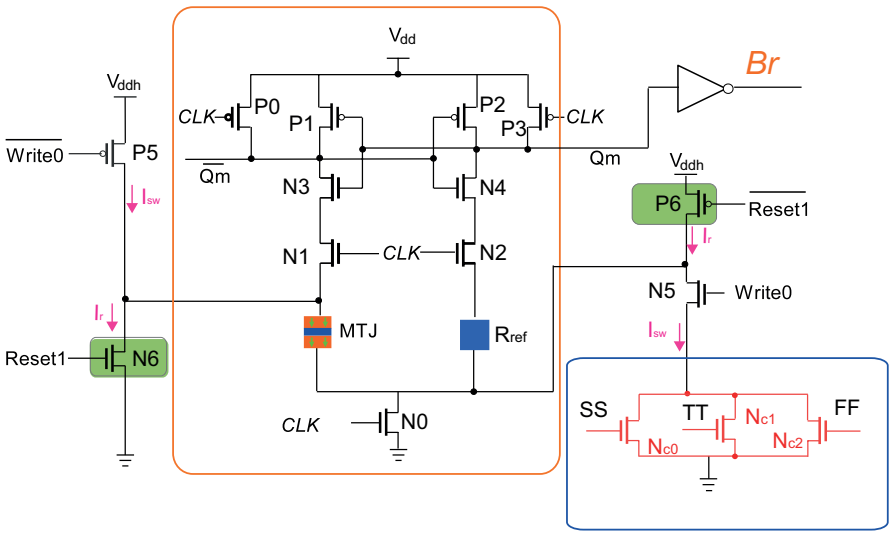


Figure 5.3: MTJ writing circuit and PCSA: I_{sw} is the switching current flowing through MTJ during random writing phase and I_r is the switching current during reset phase. N_{c0}, N_{c1} and N_{c2} modulate the switching current according to the random number probability obtained in the precedent cycle.

As a reliable sensing circuit is required, PCSA is utilized in this work because of its perfect performance in sensing latency and reliability [164, 165]. Normally, sensing errors may be induced by the process variation of FDSOI CMOS and MTJ. In order to get an error-free PCSA, the dimension of all the transistors in PCSA is validated by 1000 runs of Monte Carlo simulations. Transistors N_1 and N_2 are implemented to avoid the crosstalk between writing and sensing circuit.

5.1.3 Simulation results

By using the FDSOI 28nm design kit and MTJ compact model, we carried out transient simulation of the proposed TRNG circuit. The corresponding time-domain diagram is presented in Figure 5.5. Firstly, the switching circuit starts to write with a relatively low current ($84.5\mu A$) with the condition of N_r=$N_{clk}/2$ (switching probability P_{sw}=50%). With $N_r > N_{clk}/2$, the switching probability is decreased while the control transistor

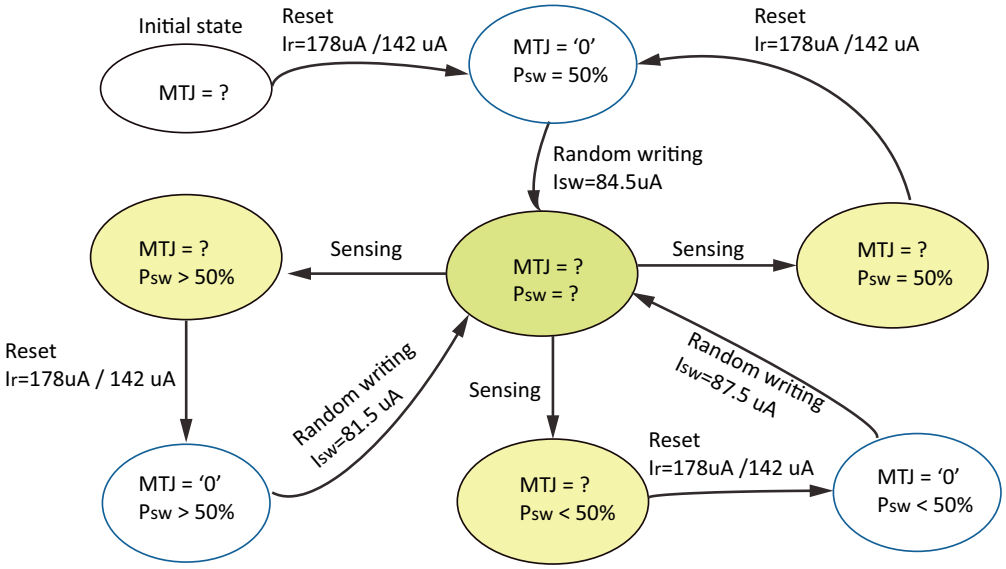


Figure 5.4: The phase transition diagram of proposed circuit design: The three states in blue frame are with different output random number probability after reset phase; The state in green signifies the unknown switching probability after random writing phase; The three states in yellow are with different known random number probability after sensing phase. For MTJs, '0' represents P state and the resistance is relatively low. '?' represents unknown information.

N_{c0} is opened by the correction block. When N_r is smaller than $N_{clk}/2$, the switching current is increased with the control transistor N_{c2} activated by the correction block. This simulation result matches well the aforementioned design goal and the functionality is well confirmed.

Based on the validated functionality, the circuit stability has also been estimated. The process variation of MTJs and transistors is taken into account by simulations with different cases of process variation. Figure 5.6 displays the probability of '1' in the obtained random bitstream during 100 cycles under different conditions. It is observed that the random number probability with different fixed-corner models becomes stable around 50% after 30 cycles. It is neccessary that the output bitstream probability needs to keep stable around 50%, which should pass the NIST test [166]. Our design has been proven to pass at least 12 tests among 15 using a 100 kbits sequence with different conditions of process variability.

5.1.4 Performance evaluation and optimization

The detailed comparison of performance with the works in [159] is shown in table 5.1. It demonstrates the proposed design has smaller area (DAC is not used), shorter tuning steps

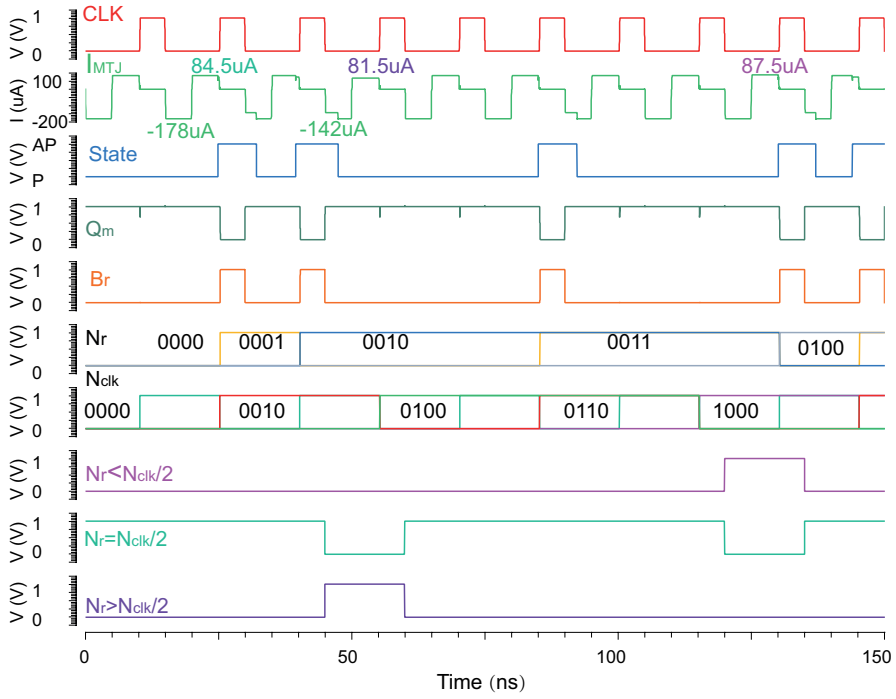


Figure 5.5: Time-domain diagram of proposed true random number generator. During each cycle, the MTJs are firstly reset to the initial state (with I_r=178μA for P state and I_r=142μA for AP state), then randomly switched, and finally sensed at the output. The initial current is set for 50% of switching success.

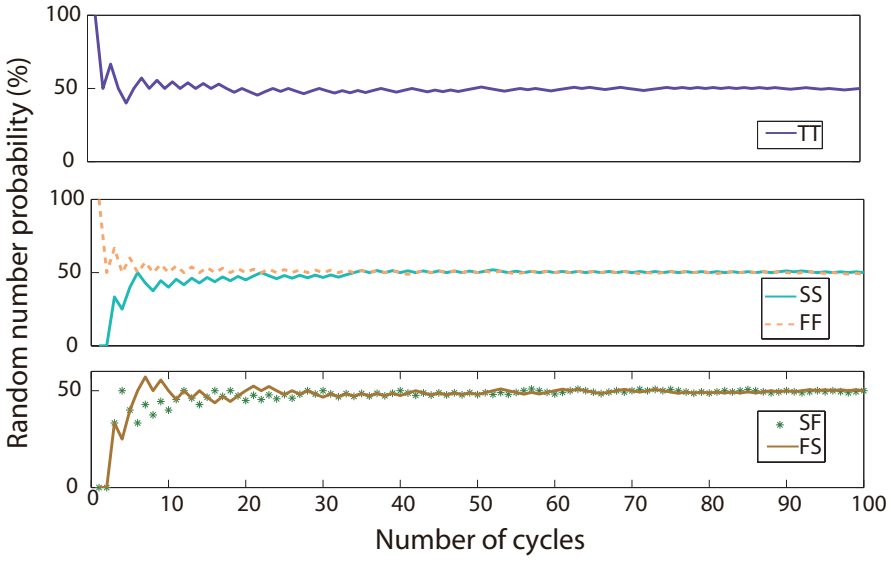


Figure 5.6: Output '1' probability versus number of clock cycles: The output random number probability becomes stable after 30 cycles (The probability of '1' occurrence stays around 50%) for all the five corner models.

Table 5.1: Comparison of performance in TRNG

	[159]	This work
Technology node	CMOS 90nm	FDSOI 28nm
Probability control	Digital	Analog
Estimated area	N/A	5.88 μm^2
Tuning steps	300	~30
Energy Efficiency	Not reported	1.25pJ/bit
Operation Frequency	66.6MHz	66.6MHz
Variability tolerance	N/A	high
NIST	N/A	pass

(with short delay before generating stable random bitstreams). Moreover, the switching current of the MTJ in [159] is much higher than that of the MTJ in this paper, our PCSA is ultra low power, and the block DAC consumes much energy, our design can be estimated to have more power efficiency.

5.2 Realization of Stochastic computing using MTJ

Stochastic Computing (SC) with random bit streams has been used to replace binary radix encoding. SC-based logic cicuits take advantage of area minimization, fast and accurate operation and inherent fault tolerance. In this section, the inherent stochastic characteristics in Spin Transfer Torque Magnetic Tunnel Junction (STT-MTJ) bring on an innovative stochastic number generator (SNM) circuit [167]. The hybrid MOS-MTJ process allows to design a 4T1M structure SNM with 1.98μm*1.46μm layout area, using 28 nm ultra thin body and buried oxide fully depleted silicon-on-insulator (UTBB FD-SOI) technology. A case study of designed SNM is performed by polynomial function synthesis, which significantly reduces area. The proposed circuit also takes advantage of non-volatility and infinite endurance from STT-MTJs, which can be applied to reliability-aware circuits and systems.

5.2.1 Introduction of stochastic computing

Stochastic Computing (SC) was first proposed by Gaines in 1967 for certain complex arithmetic operations [168]. In this method, the stochastic bit stream is used to replace conventional binary radix encoding, which achieves fast and accurate operation, hardware

minimization, and also inherent fault-tolerance. Nowadays, stochastic computation has been applied to different domains such as digital filter design [169], image signal processing [170] [171], stochastic logic circuits [172] as well as reliability evaluation [173].

The emerging microelectronic device with new material, e.g., half-metallic ferromagnetics, have been major breakthroughs after its discovery in 2007 [13] [174]. STT-MTJ is used because it features high power efficiency, speed and infinite endurance [13]. STT effect has made magnetic nano-devices realistic candidates for active elements of memory devices and applications [175]. Spin torque building blocks such as magnetic Non-Volatile Flip-Flop (NVFF) [176] [120] [177], non-volatile logic circuits have been implemented with transistor/MTJ hybrid structure [47].

Normally, the random number generator is composed by Liinear Feedback Shift Register (LFSR), counters and comparators [169] [171] [172]. In this work, the intrinsic stochastic behavior of STT-MTJ is profited to realize an innovative stochastic number generator (SNM) circuit, which can be used to generate stochastic bit streams. This method is applied to polynomial function synthesis.

5.2.2 Stochastic computation with combinational logic

Unlike deterministic computation with binary radix, stochastic computation uses signal probability to describe input and output signals. The occurrence probability for the signal to be of logic '1' or logic '0' is encoded from random bit stream. For example, both 5 bits '01010' and 10 bits '1001100001' are stochastic codes for a signal probability as 0.4.

Stochastic logic gate operations such as AND, OR, XOR and multiplication (scaled addition) are shown in Figure 5.7. Independent inputs are assumed. Compared with conventional logic gates, stochastic logic gates achieve high speed, fault tolerance and power efficiency. However, lack of general design implementations, especially at circuit level, limits the development of SC based circuits.

5.2.3 Stochastic computing using STT-MTJ

A new stochastic number generator is proposed to generate probabilistic output as stochastic bit stream by using STT-MTJ. STT switching mechanism of MTJ has been demonstrated intrinsically stochastic due to the thermal fluctuations of magnetization [21, 64]. Caused by this phenomenon, the switching probability (SP) is related to switching time, MTJ operation current and critical switching current (I_{c0}). Normally, a deterministic data is obtained when MTJ operation current and pulse width is sufficient. For example, to

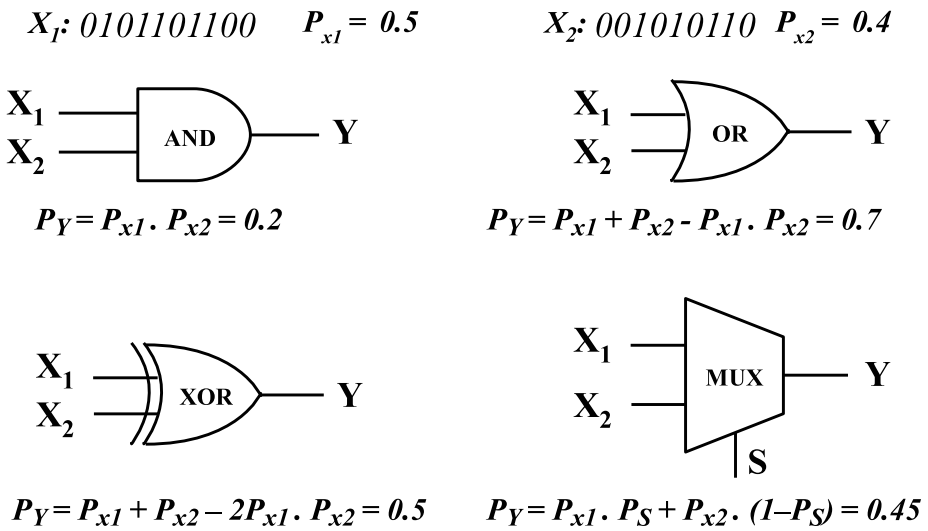


Figure 5.7: Examples: SC based on two input combinational logic (AND, OR, XOR and scaled addition).

store a '1' into MTJ, the signal $Data_{1-0}$ is '1', $Data_{0-1}$ bar is '0'. In order to obtain a probabilistic behavior, MTJ operation current is set up as lower than sufficient current. MTJ SP is estimated on the basis of MC simulation at transistor level. The switching probabilities versus MTJ operation current are shown in Figure 5.9. The stochastic behavior is also determined by operation frequency. 50 MHz (MTJ writing duration = 10 ns) and 100 MHz (MTJ writing duration = 5 ns) input signals are simulated with MC method. MTJ generated signal probability can be used for stochastic logic computation.

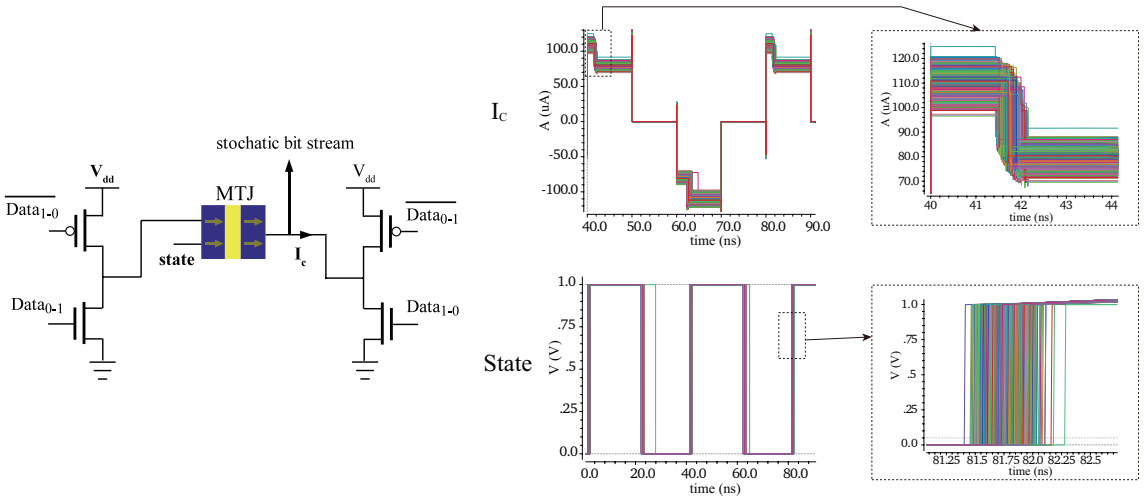


Figure 5.8: Proposed stochastic bit generator with 4T1M structure. 1000 runs monte-carlo simulation illustrates the stochastic behavior of MTJ.

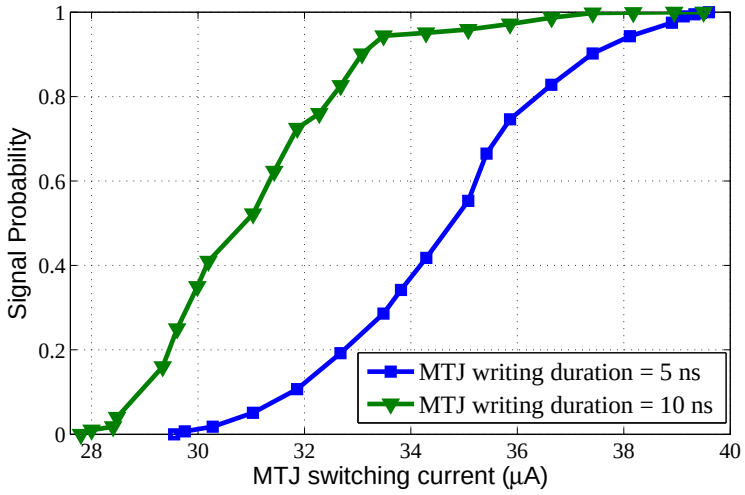


Figure 5.9: Simulation result: switching probability versus MTJ operation current.

Figure 5.10 presents the layout view of a 4T1M SNG circuit. The size of the designed SNG is about 2.9 μm^2. In order to separate MTJs during chemical mechanical polish processing, the contact points of MTJ are between two highest adjacent metal layers of overall layout. In this design, it locates at the back-end of the CMOS process from the metal layers 3 (metal3).

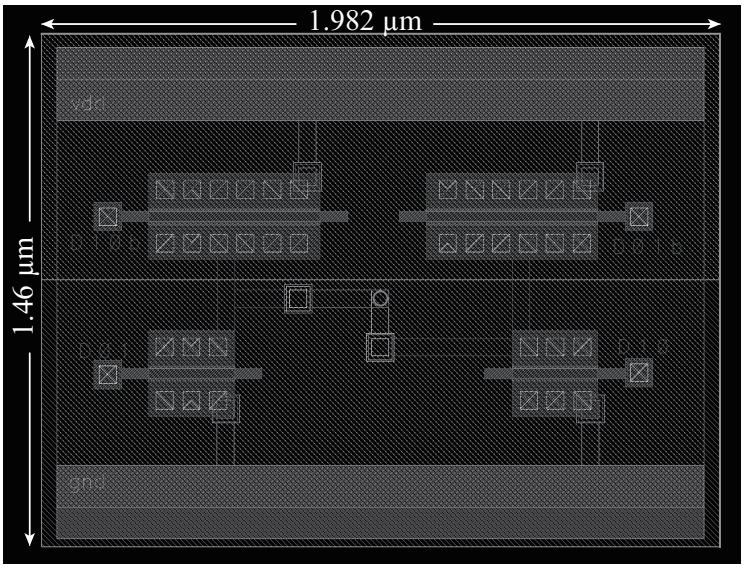


Figure 5.10: Layout of 4T1M SNM with 28nm FDSOI process.

5.2.4 Case Study: Polynomial function RTL synthesis

STT-MTJ based SNG can be used to synthesize polynomial function at Register Transfer Level (RTL). In this work, we consider the synthesis of the function y described in (5.1).

$$y = 0.06x^2 + 0.19x + 0.25 \tag{5.1}$$

With a 10-bit length input signal, we compare the RTL synthesis results of traditional synthesis flow (binary signal) to those from a stochastic synthesis flow. Circuit area is evaluated with 28nm FD-SOI design-kit.

Figure 5.11 illustrates the realized schematic based on binary signal. The total circuit area is 1122.98 μm^2. It is composed of Carry Select Adder (CSA) trees and multiplexers.

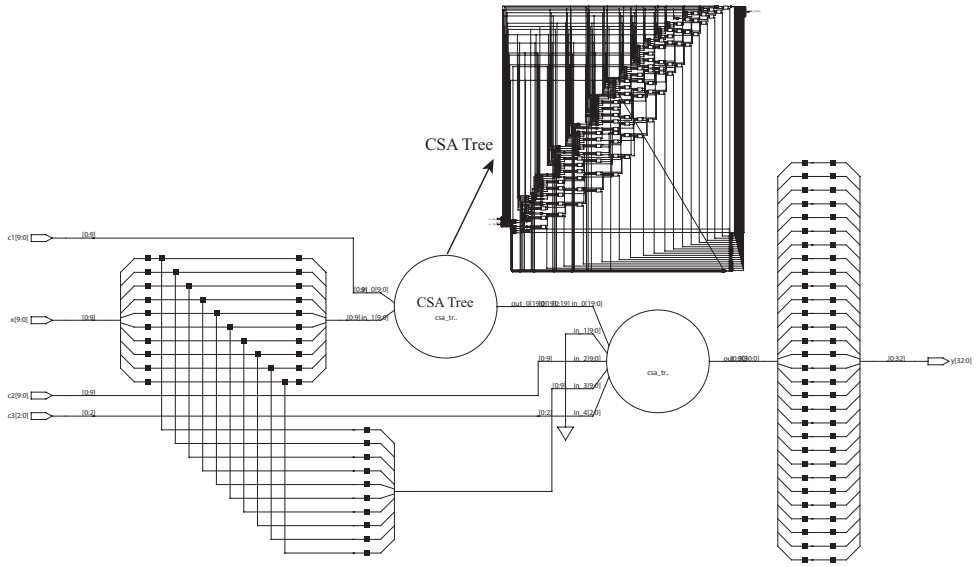


Figure 5.11: Polynomial function synthesis with traditional binary signal.

Figure 5.12 shows the circuit using STT-MTJ based SNG for polynomial function. It only consists of 5 STT-MTJ based generators, 3 MUXES and 2 AND logic gates. The total cell area should be less than 100 μm^2 according to 28nm design kit. Considering low power family transistors, the estimated propagation delay for stochastic method is 52 ps, which is far below 360.8 ps achieved by traditional method synthesis. Detail description of Bernstein polynomial synthesis can be found in [178]. It is noted that de-randomizer unit is required to reconstruct stochastic results to binary values.

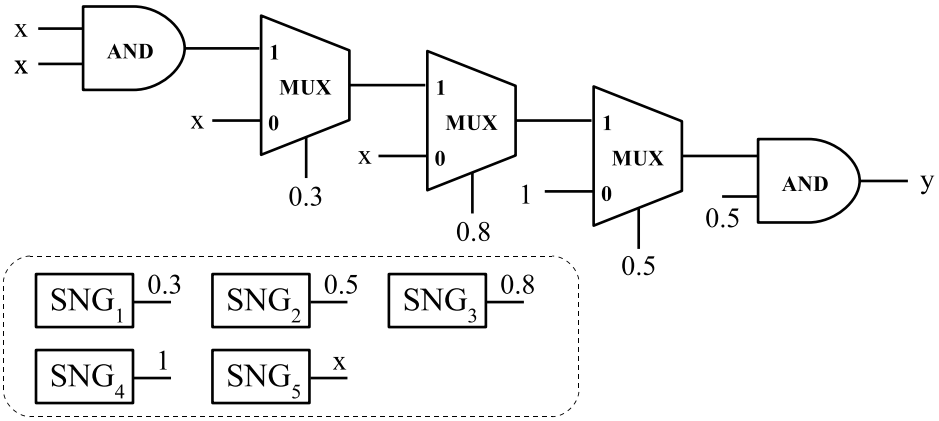


Figure 5.12: An example of polynomial function synthesis.

5.3 Approximate computing method using MTJ

Approximate computing and its related topics have shown the potential in next generation computing systems. In this part, new circuit level design for approximate computing is proposed based on non-volatile (NV) logic-in-memory structure. Two types of NV approximate adders are implemented with circuit reconfiguration and insufficient writing current. Spin torque transfer magnetic tunnel junction (STT-MTJ) is used as NV memory element in magnetic full adder (MFA). The proposed approximate MFAs are implemented with 28nm ultra thin body and buried oxide (UTBB) fully depleted silicon-on-insulator (FD-SOI) technology. Simulation results are presented including power consumption, circuit latency, leakage power, error distance and reliability performance. Low V_{dd} design strategies are discussed as well.

5.3.1 Introduction of approximate computing

In order to improve energy efficiency compared to conventional numerical calculations with high precision, several computation methods have been considered e.g., approximate computing and probabilistic computing [179, 180, 181]. Approximate computing in hardware and software is promising in energy-efficiency which is regarded as a new degree of freedom to improve next generation computing systems [182]. For instance, an approximate multi-bit adder can be implemented with accurate circuit for more significant bits, whereas inexact circuit for less significant bits [183, 181].

Emerging non-volatile (NV) technology has shown the potential to be commercially used in these years. Energy efficiency in hybrid magnetic-MOS circuits is of great importance in circuit level implementation. The spintronic devices have been well-developed in

logic circuits [49, 184] and memories [120, 135, 177]. Related to this work, non-volatile magnetic full adders (NV-MFA) have been investigated in [46, 164, 47, 185], with improved performances in terms of dynamic power, layout area and operation speed. Recently, approximate storage with spintronic devices has been reported showing favorable energy-quality trade-offs [186].

In this section, we explore novel design for approximation methods in hybrid circuits based on 28nm ultra thin body and buried oxide (UTBB) fully depleted silicon-on-insulator (FD-SOI) technology and spin transfer torque magnetic tunnel junction (STT-MTJ). Approximate NV-MFA is realized by applying either circuit reconfiguration (reduced logic complexity) or insufficient writing current in MTJ. The major contributions of this section are as follows:

- Low power dual-mode MFAs based on differential sensing amplifier are implemented considering both dynamic and leakage power as weel as process variations.
- Reliability-aware design strategies including poly bias and single-well device are investigated. Sub-V_t and near-V_t operations of MTJ based MFA are explored.

5.3.2 Design for Approximation

Design for approximation methods are discussed in this section. Non-volatile full adders based on MTJs are designed as approximate adders by reducing logic complexity and implementing inexact MTJ writing.

5.3.2.1 Reduced Logic Complexity

The accurate 1-bit full adder has three inputs A, B and C_i (carry in). The logic functions of Sum and C_o (carry out) are given by:

$$Sum = A \otimes B \otimes C_i \tag{5.2}$$

$$C_{out} = AB + AC_i + BC_i \tag{5.3}$$

An important metric named error distance (ED) is introduced into approximate adder evaluation besides the power and latency performance [187]. In any approximate FA, the inexact output a and accurate output b is compared arithmetically for all possible

combination FA input bit by bit:

$$ED(a,b) = |a-b| = |\Sigma_i a[i] * 2^i - \Sigma_j b[j] * 2^j| \tag{5.4}$$

where i and j are the indices for the bits in a and b [183, 187].

Conventional approximate adders with CMOS technology are designed with reduced logic complexity. Figure 5.13 shows some developed CMOS approximate adders [181, 183] based on logic simplification. The AXA1 described in [181] consists of an inverter and two pass transistors connected to two inputs. AXA2 and AXA3 are implemented based on the XNOR pass transistor network [188].

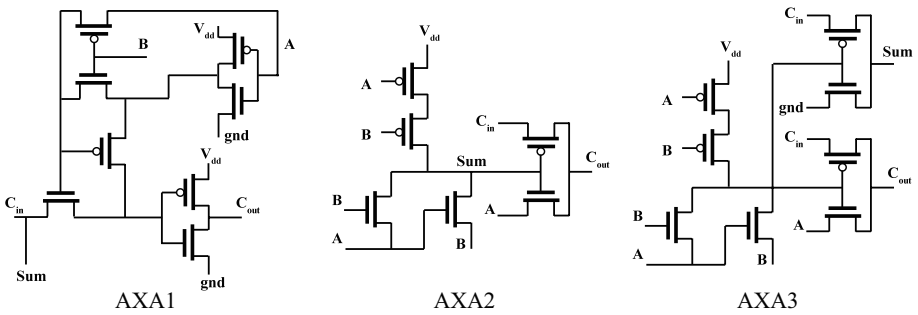


Figure 5.13: Conventional CMOS approximate adders: AXA1, AXA2 and AXA3 [183].

Traditional simplified logic method is also applicable to MTJ based logic-in-memory circuits, in which pass transistor logic is an essential building block. As shown in Figure 5.14 (AX-MFA1), input C_i can be eliminated in logic-in-memory network (dashed rectangle) to approximate the function of Sum operation, with $Sum = A \otimes B$, whereas C_o function is still with accurate computation. Input B is non-volatile data stored in MTJs by using the 4T2M write block (see Figure 2.13).

5.3.2.2 The Dual-mode MFA

In this work, a new approximate computing method is proposed based on insufficient MTJ writing current. In order to guarantee MTJ operation, the threshold switching current is required in conventional hybrid MOS/MTJ circuits. Usually a high current ($I > 2 * I_{c0}$) is applied to guarantee the fast writing speed (sub $3ns$) in memory. Here, the insufficient MTJ writing operation is used to generate an approximate signal (input B). The circuit implementation of AX-MFA2 (see Figure 5.14) is similar to AX-MFA1. Note that AX-MFA2 is implemented with the whole schematic.

AX-MFA2 can adaptively operate in the accurate mode with a 0.8V above V_{dd}, while MTJ switching is deterministic. On the other hand, we use MTJ switching behavior as the

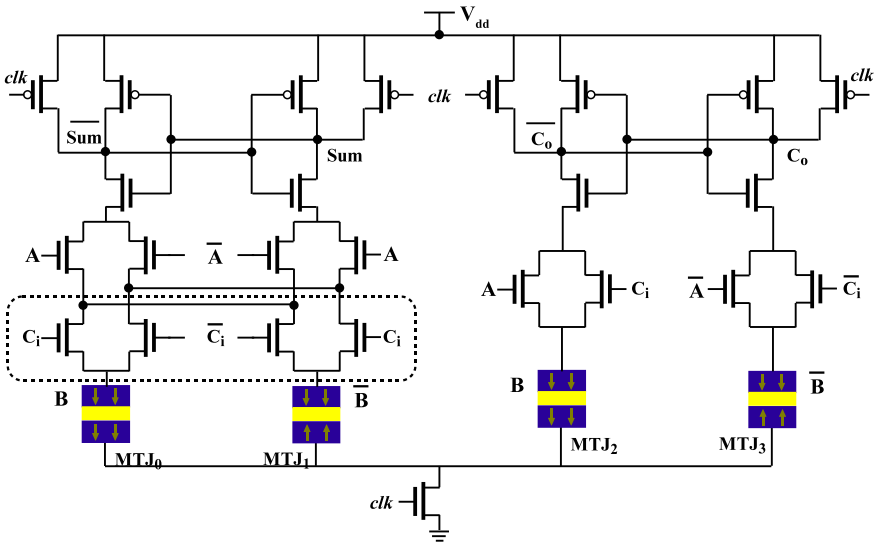


Figure 5.14: Circuit implementation of two approximate MFAs: AX-MFA1 (without dashed line box) and AX-MFA2. The first approximate AX-MFA1 is implemented with conventional simplified logic: input C_i in dashed rectangle is eliminated to get an approximate $Sum = A \otimes B$. The second dual-mode approximate AX-MFA2 is implemented with the whole schematic.

selection mechanism between accurate mode and approximate mode: when V_{dd} is lower than 0.8V, there is insufficient switching current of MTJ to process new input. Thus, an approximate MFA can be implemented with approximate data B.

5.3.2.3 Functional Simulation

A low V_t 28nm commercial FDSOI technology is applied into approximate adder design. The SPICE compact model of MTJ [98] programmed in Verilog-A is used to simulate static and dynamic behaviors of MTJ based circuits. Minimum dimension transistors are used in sensing circuit. Large dimension transistors are used in MTJ writing circuit, where W/L of pMOS transistor is 1000nm/30nm, nMOS transistor is 500nm/30nm.

Figure 5.15 shows the timing waveform of the approximate adder with reduced logic complexity. All possible combination of inputs A, B (NV data stored in MTJs) and C_{in} are designated. The approximate adder becomes active for computation when clock signal is high. Note that inexact output Sum is marked whereas output C_{in} is accurate. The total error distance is 4 in simplified logic based AX-MFA1.

Since the proposed dual-mode MFA architecture is the same as previous MFAs [47, 164], functional simulation of MFA accurate mode (above 0.8V supply) is not presented in detail. When V_{dd} is lower than 0.8V, the current flowing in MTJ is lower than MTJ critical

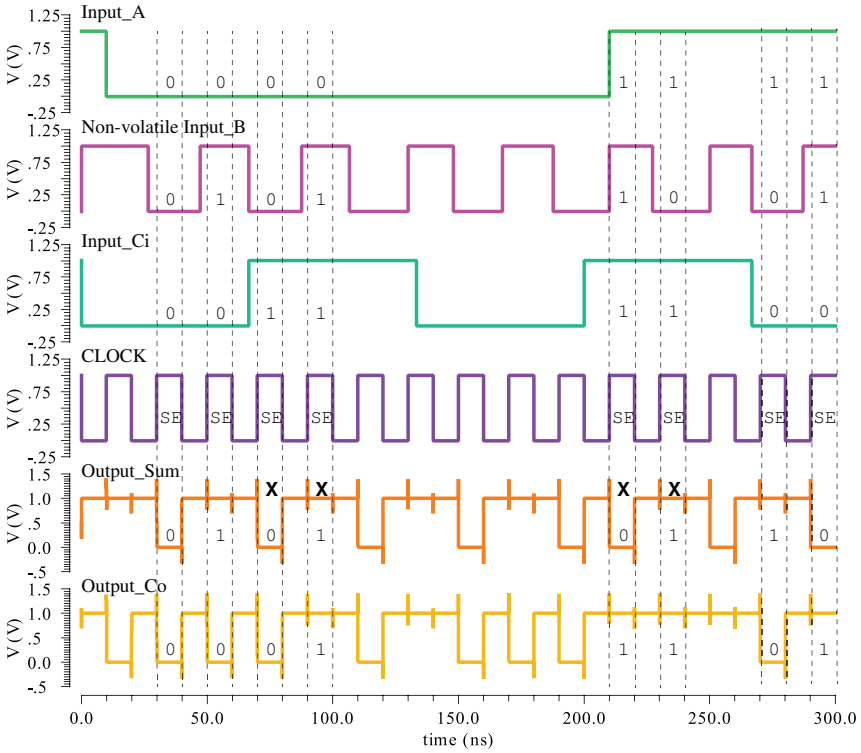


Figure 5.15: The transition simulation waveforms of approximate adder with reduced logic complexity (AX-MFA1). Output *Sum* is with errors.

switching current and MTJ states can not be changed between P and AP. Thus, power consumption in both MTJ sensing and writing operation are greatly reduced. Meanwhile, inexact output occurs at adder output (both Sum and C_{in}). Approximate mode can be executed over a supply ranging from 0.3V to 0.75V. The transient simulation results of dual-mode approximate adder are shown in Figure 5.16. A 0.5V supply is applied as an example in this MFA structure. The total error distance is 6 in this simplified approximate MFA.

5.3.3 Design Considerations

Although simplified logic based approximate MFA has 4 less transistor count and 2 less error distance, the dual-mode MFA (AX-MFA2) is with a flexible operation with both approximate and accurate modes. Based on FDSOI technology, a supply scaling strategy for MTJ based PCSA circuit is proposed with single well doping and poly bias [117]. Low power design related performance tradeoff is presented.

5.3.3.1 Supply Scaling Strategy

Previous MTJ based circuits are designed with minimum 1V V_{dd} considering successful MTJ writing and sensing operations. With a 0.8V above supply, the proposed MFA

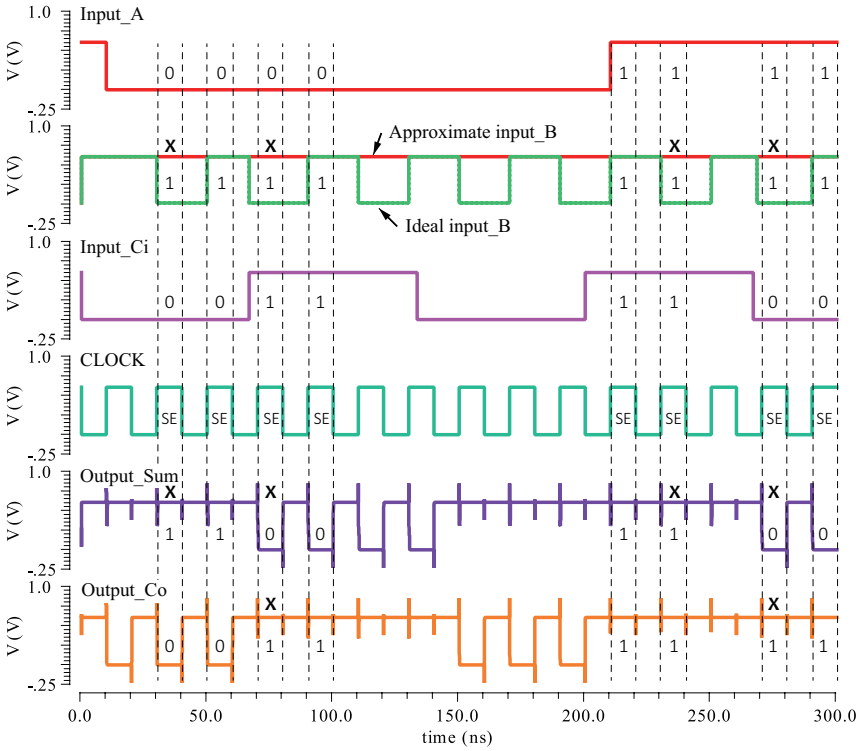


Figure 5.16: The transient simulation waveforms of approximate adder by insufficient writing current (AX-MFA2).

operates at accurate mode since MTJ switching can be nominally guaranteed. MFA approximate mode can be executed over a supply ranging from 0.36V to 0.75V. Figure 5.17 demonstrates the power strategy applied to MTJ based dual-mode MFA. Four well doping types are included into supply scaling strategy. Approximate mode can be executed over a supply ranging from 0.3V to 0.75V. A proper selection on V_t level is important.

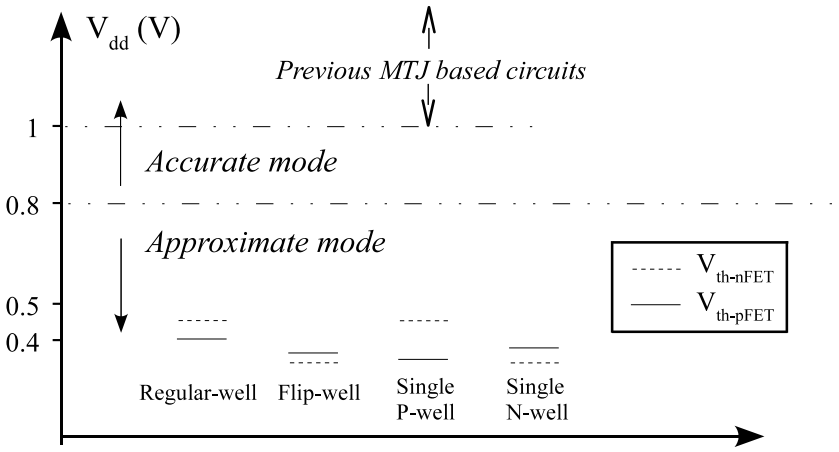
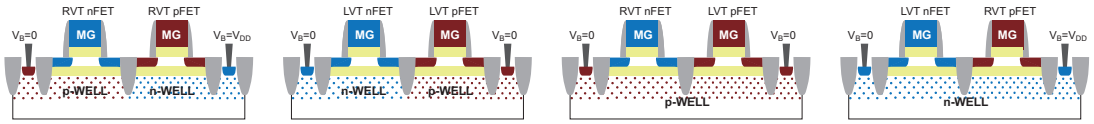


Figure 5.17: Supply voltage strategy in bi-mode MFA.

Figure 5.18 shows dynamic transistor implementation in FDSOI technology: the regular well, flip well, single P-well (SPW) and single N-well (SNW). Flip well implementation has improved performance with the tradeoff in leakage power and variability. The SPW

has been validated in FDSOI based bit-cell and SRAM circuit [189, 190, 191]. It has been reported that minimum V_{dd} can be 70mV to 100mV lower than regular well/flip well implementation, the circuit stability (e.g., failure probability) is greatly improved around 0.5V supply. The drawback of SPW is that a deep-N-well must be inserted between the P-well and substrate for isolation, whereas no isolation is required in SNW structure. SNW enables energy efficient design with controlled leakage power [192]. We find single P-well doping method can increase pMOS V_t (RVT implementation), whereas nMOS V_t is reduced by LVT implementation. In PCSA based logic-in-memory circuits, the single P-well doping method can boost logic operation, whereas weaken SA performance.



(a) Regular well (b) Flip well (c) Single P-well (d) Single N-well

Figure 5.18: Cross-sectional view of dynamic well FDSOI MOS devices. Different well configurations impact circuits performance.

Figure 5.17 demonstrates the four V_t configurations. The proposed approximate MFA can work at a minimum 0.36V supply, where both nMOS and pMOS transistors work at sub-V_t region. However, several problems impact the sub-V_t and near-V_t circuit performance (e.g., process variations, leakage power and low speed). The supply scaling strategy of the proposed approximate MFA must include optimization techniques to overcome these problems. The single P-well transistor implementation is used to provide variability enhancement near-V_t circuit.

5.3.3.2 Performance Analysis

A minimum 360mV V_{dd} can be realized in the proposed 1 bit MFA for sub-V_t operation, where nMOS transistors are with 366.5mV V_t, pMOS transistors are with 416.5mV V_t. Near-V_t operation can be realized with 420mV V_{dd}.

Figure 5.19 illustrates the relative size and layout of proposed NV-MFA. As MTJ is usually placed over the highest metal level in the back-end of the CMOS process, the die area cost can be reduced.

Figure 5.20 shows latency performance of proposed dual-mode MFA. In accurate mode with 1V supply, 19.3 ps latency is achieved. A 152.7 ps latency is realized in approximate Sum operation when V_{dd}=0.5V. Continuous supply scaling down to sub-V_t region leads to large latency (1.27 ns when V_{dd}=0.36V).

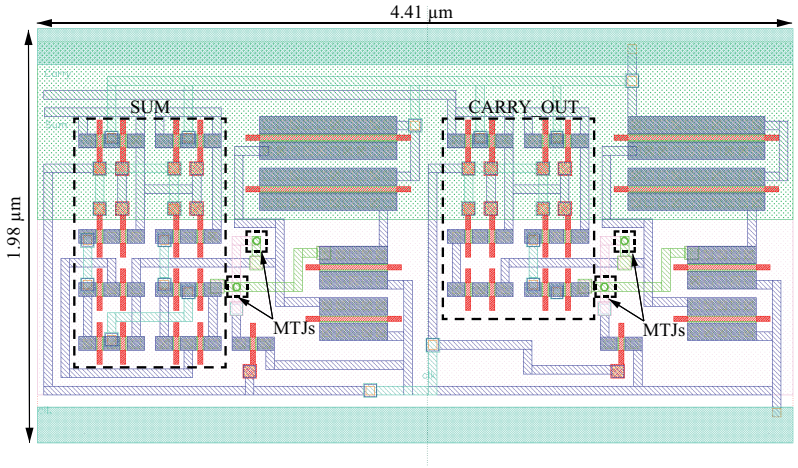


Figure 5.19: 4.41μm*1.98 μm Layout with planar 28nm FDSOI technology. A 16nm poly bias is used to reduce leakage power and enhance yield. LVT-RVT strategy is performed in layout, single P-well covers nMOS RVT transistor (in sense amplifier) and pMOS LVT transistor (in logic network).

5.3.3.3 Reliability-aware Simulation

The global and local process variations of transistor and MTJ device, as well as the stochastic behaviors of MTJ are evaluated by Cadence ADE-XL, with 500 runs Monte-Carlo methods. 1-sigma transistor variability is considered, whereas the Gaussian distribution is realized in MTJ at the range [0.97, 1.03].

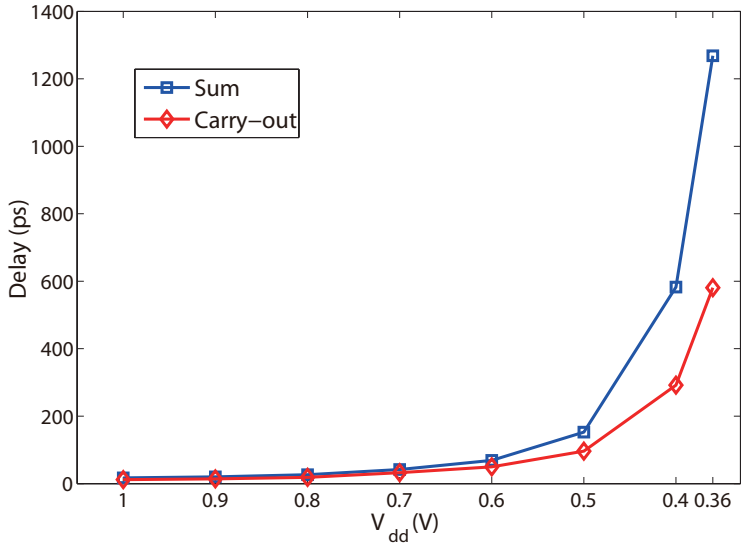


Figure 5.20: Latency simulation of dual-mode MFA. A 152.7 ps latency is realized in approximate Sum operation when V_{dd}=0.5V. Continuous supply scaling down to sub-V_t region leads to large latency (1.27 ns when V_{dd}=0.36V).

Figure 5.21 presents these impacts in proposed dual-mode MFA. Supply voltage scaling

leads to increased error probability. In accurate mode, reliability problems can be suppressed with a 1V above supply. In approximate mode, since NV data stored in MTJs is treated as approximate input, both random variation and stochastic effect in MTJ device can be neglected. The variability aware study of transistor shows 89% and 95% probability with 0.5V and 0.55V supply separately. Notice that single N-well doping is used.

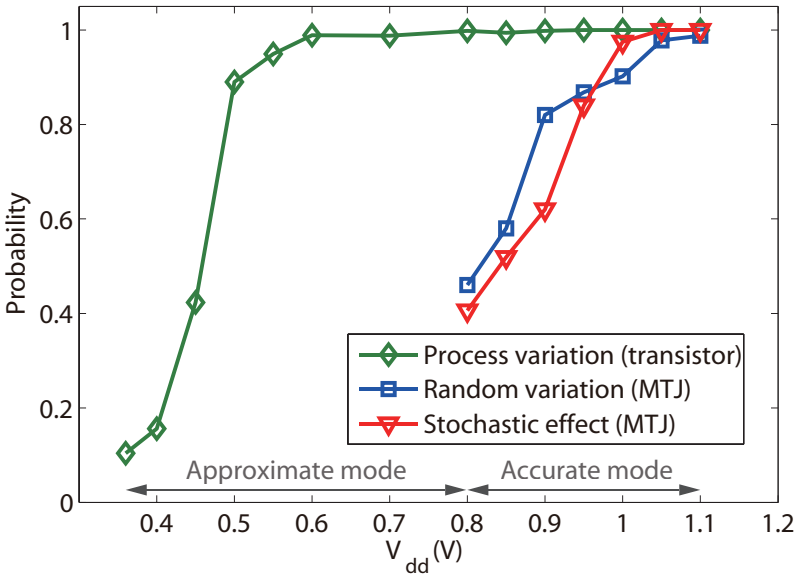


Figure 5.21: Probability with respect to V_{dd} scaling. MOS/MTJ process variations and MTJ stochastic effect influence MFA probability.

Another failure analysis is performed in the proposed MFA with different well doping methods. As shown in Figure 5.22, single N-well doping method reduce variability induced circuit failure in low power approximate mode (between 0.36V to 0.65V). Comparing with regular well and single P-well, a maximum $130mV$ V_{dd} margin is achieved for a given probability. Comparing with flip-well implementation, single N-well doping method achieves $70mV$ V_{dd} margin, as well as 17% leakage reduction in circuit active mode.

Table 5.2 summarizes simulation results of proposed approximate MFA with simplified logic (AX-MFA1) and dual-mode MFA (AX-MFA2). We compare the performance with CMOS approximate full adders (CMOS AX-FA) and previous MFAs with different CMOS nodes. Approximate adder with simplified logic achieves less power than conventional MFA. It also accelerates the speed by 17%. Approximate adder with insufficient writing can operate with low supply voltage. Both sensing and writing energy is reduced by nearly 70%. Another advantage of this technique is that designers can select MFA operation mode between accurate and approximate. Meanwhile, 0.8V V_{dd} is the threshold value to achieve approximate MFA.

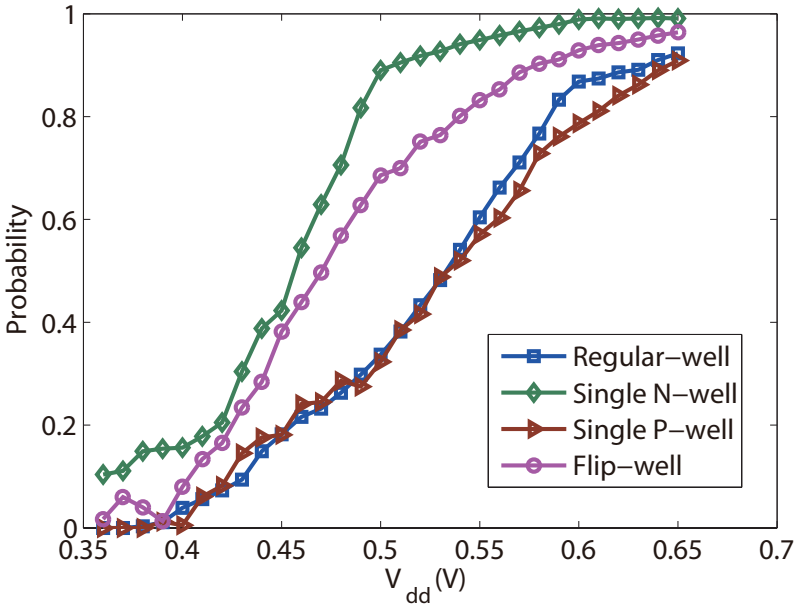


Figure 5.22: Sensing probability with respect to V_{dd} considering process variation. Different well configurations impact sensing error rate. Single N-well doping method achieves the extra V_{dd} margin.

Table 5.2: Performance comparison of conventional MFA and proposed approximate adder.

	1-bit adder Type	V_{dd} (V)	Delay (ps)	Error dist. Sum	Error dist. C_o	Dynamic power(nW)	Leakage (pW)	Device count	Layout (μm^2)
CMOS AX-FA [1)]	AXA1	1	20.14	4	4	9.697	1073	8T	0.81
	AXA2	1	69.83	4	0	6.984	1362	6T	0.64
	AXA3	1	48.6	2	0	9.041	1397	8T	0.77
Previous MFA	[47] 65nm bulk	1	170	0	0	2950	0 [3)]	30T+4M	20
	[46] 40nm bulk	1.2/1.5	87.4	0	0	1980	<1 nW [3)]	38T+4M	20
	[164] 28nm bulk	1	150 (8 bit)	0	0	$0.68pJ/8bits$	0 [3)]	25T+4M	24.81 [2)]
This work	AX-MFA1	1	16.22	4	0	8.69	329.5	21T+4M	8.51
	AX-MFA2(accurate)	1	19.3	0	0	9.46	401.6	25T+4M	8.74
	AX-MFA2(approximate)	0.5	152.7	4	2	2.112 [4)]	77.91/5.06 [5)]	25T+4M	8.74

[1)] CMOS AX-FAs are implemented with 28nm FDSOI technology.

[2)] 1 bit MFA layout area includes magnetic flip-flop (MFF).

[3)] Leakage power in active mode is not considered. Zero leakage is achieved only in standby mode.

[4)] 2.112nW, $0.00195pj/bit$.

[5)] 77.91pW is achieved without poly bias, 5.06pW is realized with 16nm poly bias.

5.4 Conclusion

In this chapter, we have realized novel design of three traditional circuits using the stochastic switching behavior of STT-MTJ.

Firstly, a circuit design for TRNG based on STT-MTJ was proposed which use the intrinsic stochastic switching behavior as a randomness source. The proposed solution has been implemented with FDSOI CMOS circuits. Monte-Carlo simulations were performed to prove its feasibility and transient simulation has validated its functionality. Furthermore, the simulation with process variation of MTJs and transistors has proved the variability awareness of this design. The comparison with another similar work shows that the proposed design has better performance in terms of area, power efficiency, operation speed and variability tolerance.

Secondly, inherent stochastic characteristics of STT-MTJ were used to design stochastic number generators. A 4T-1M circuit with hybrid MOS-MTJ process is implemented with 28nm UTBB-FDSOI technology, with only 2.9 μm^2 layout area. The probability of stochastic bit stream is determined by supply voltage and signal input frequency. A case study of designed SNM is performed by polynomial function synthesis, which achieves greatly area minimization compared with traditional binary synthesis method.

Finally, two approaches to achieve approximate computing in a non-volatile logic-in-memory architecture are realized. Approximate method with simplified logic and insufficient writing of non-volatile storage were implemented in magnetic full adder based on 28nm FD-SOI technology. The approximate approach with insufficient writing is more effective compared with conventional logic simplification. A dual-mode MFA is proposed for ultra low power consideration. Its approximate mode with insufficient writing current can save 78% energy compared with accurate MFA. Dynamic well-doping method and poly bias are used to overcome variability and leakage problems. The proposed approximate MFA can be described as standard cell in logic synthesis for ultra low power hybrid magnetic-MOS circuit.

Chapter 6

Conclusions and Perspectives

6.1 Conclusions

This thesis is dedicated to the reliability analysis, enhancement and exploration of new applications based on PMA-STT-MTJ. The work mainly includes three parts: compact modeling of the key reliability issues existing in PMA-MTJ; reliability analysis of typical memory and logic circuits as well as proposal of reliability aware design methodologies; new designs of traditional specific applications using PMA-MTJ benefiting from the stochastic switching behavior.

Starting from the state of the art, we have reviewed the development of Spintronics devices and its application in memory and logic circuits. The evolution of magnetic tunnel junction was driven by the switching approach optimization and improvement of tunnel mangetoresistance ratio. With the detailed comparison of different switching methods, it has been concluded that STT is the most suitable candidate for future memory and logic circuits with the best tradeoff between power consumption, operation speed, scalability, endurance and 3D-integration into conventional CMOS circuits. Compared with other memories widely used or emerging in the last decade, STT-MRAM is the most promising candidate. However, it suffers from considerable reliability issues which limit its wide application. The history of reliability analysis has been reviewed. However, the current work fails to meet the urgent requirement of high reliability designs, which motivate us to investigate the reliability issues of MTJ and create an accurate compact model for circuit designers.

In the part of compact modeling, the PMA-MTJ working principles and its advantages have been introduced and theoretically demonstrated. Then, we have entirely analyzed the provenance of the reliability issues including process variations, stochastic switching,

temperature fluctuation and dielectric breakdown. For instance, the influence of every step in fabrication process has been presented. Based on the comprehensive study of physical mechanisms of reliability issues, different physical models are considered to constitute compact modeling. With all the modeling elements well prepared, we begin to present our modeling language and hierarchy. Meanwhile, the modeling and simulation results are presented to validate its functionality. These models can be used to execute a more realistic design according to the constraints obtained from simulation. With the confirmed reliability model, circuit designers are able to predict the circuit performance accurately. For example, they can investigate the robustness against process variation, find the tradeoff between switching performance and power, predict the lifetime of hybrid MTJ/CMOS circuits and estimate the tolerance to temperature fluctuation. Considering these information, the model users can find solutions to adjust their designs to the target at the early design phase, resulting in less unnecessary loss and higher yield rate. The compact model is developed in SPICE-compatible language and can be used in all the environments for circuit level simulations.

A non-Monte-Carlo methodology for variability analysis of MTJ based circuits was also presented. The methodology is implemented by using worst-case corner models of PMA-MTJ and transistor. The MTJ compact model of worst-case corners is proposed for the first time. The design specifications are detailed in non-volatile memory cells and arithmetic unit, e.g., 1 transistor-1MTJ memory array, pre-charge sense amplifier based STT-MRAM and magnetic full adder circuit. The methodology is implemented by using a 28 nm FDSOI design kit and this compact model of MTJ. Results show that the proposed methodology is much more efficient than conventional Monte-Carlo (MC) method while keeping the same target of performance evaluation. The simulation speed has been improved up to 226x for an effective evaluation of circuit performance (compared with 1000 times MC simulation). MTJ based circuit designers can assess the impact of process variation on circuit performance without time consuming MC simulations by using this method.

Based on the validated models, we have carried out reliability analysis of commonly used MTJ based memory and logic circuits. The first step is demonstrating the effect of each reliability issue on the whole circuit. Meanwhile, the impact of MTJ parameters on the circuits performance were also investigated. As a result, we have found the correlation of the impact of different parameters on the performance terms. For example, high supply voltage can efficiently prevent the writing operation from stochastic switching effect but

lead to fast dielectric breakdown, high temperature assists the switching process as well as increases also the sensing error rate, etc. During the design phase, all of the temrs should be considered for a best tradeoff. Based on the reliability analysis, we presented a methodology of PVT variation immune design using dynamic asymmetrical body bias of 28nm FDSOI transistors. This methodology is implemented in PCSA circuit to demonstrate its feasibility. Simulation results have shown a significant improvement of reading success under different simulation conditions with the minimum circuit area. The proposed circuit behaves perfect robustness to process, voltage and temperature variability awareness. As symmetrical structure is widely used in the MTJ based applications, this method can be applied in many other MTJ/CMOS circuits and systems for a variability-aware design.

Stochastic switching behavior has been a performance degradation factor in common memory and arithmetic circuits. However, it can also be useful or even advantageous in some specific applications. We have proposed novel designs of true random number generator (TRNG), stochastic computing (SC) and approximate computing using this phenomenon. The proposed TRNG demonstrates perfect variability awareness, which is taken into account for the first time in the MTJ based TRNGs. Compared with other works, it features lower power, higher operation speed, better robustness and more compact area. In the SC design, a case study of designed SNM is performed by polynomial function synthesis, which achieves greatly area minimization compared with conventional binary synthesis method. Finally, two types of NV approximate adders are implemented with circuit reconfiguration and insufficient writing current. Compared with traditional logic simplification, insufficient writing of MTJ is more effective which can save most of the energy. Moreover, the variability and leakage problems are overcome by dynamic well-doping method and poly bias of FDSOI transistors.

We are convinced that our work is beneficial to the development of STT-MRAM and Logic in Memory. The compact model of reliability issues and the methodologies can be used by circuit designers for realizing more robust designs with less time loss and higher yield rate. The novel designs using stochastic switching issue provide new insights into more wide applications of MTJ and explore its potentials in the future applications.

6.2 Perspectives

The thesis comes to the end while the work never stops. As an extension of this thesis, there are some points which can further improve our work.

1. Optimization of compact model including reliability issues

From the view of precision, the proposed compact model includes the main but not all reliability issues. With the widespread of MTJ applications, other effects should also be taken into account, such as radiation effect. Otherwise, breakdown is a very complicated phenomenon which includes many factors as we mentioned in the thesis. We have just considered the most severe effect (dielectric breakdown), the model will be more precise if other mechanisms can be integrated, such as performance degradation caused by soft breakdown. To further facilitate the model utilization and accelerate the simulation speed, it is also possible to combine the model of CMOS transistors and MTJ device in only one file, which has the potential to increase the design efficiency. Moreover, a library file of MTJ compatible with that of transistors for register transfer level (RTL) synthesis is also required for very large scale memory and logic circuits.

2. Better methodology for reliability-aware designs

In the design of variability aware methodology using dynamic asymmetrical body bias, the body bias voltage generator needs much improvement. In very large scale circuits, generator composed of resistor and capacitance means considerable area loss and energy cost. Consequently, the scalability and power efficiency will be drastically degraded. Thus, a new voltage generator should be developed for widespread of this methodology. If this problem is solved, the dynamic asymmetrical body bias can be used in many other symmetrical circuits.

3. Security applications based on MTJ

Besides the stochastic switching behavior employed in true random number generator, the probabilitic breakdown behavior can also be used in another security application: physical unclonable function (PUF). The unpredictable exact time for breakdown provides excellent randomness source. As hard dielectric breakdown is catastrophic and irreversible, it can meet the requirement of one-way function in PUF. With the feature of low power and high speed operation, MTJ based security applications will play an important role in the coming era of Internet-of-Things where both instant data processing and big data storage are necessary.

Bibliography

[1] "Nanomagnetism and spintronics," Amsterdam: Elsevier, 2009.

[2] O. www.itrs.net/Links/2015ITRS/Home2015.htm, "International technology roadmap for semiconductors (ITRS)," 2015.

[3] O. www.itrs.net/Links/2011ITRS/Home2011.htm, "International technology roadmap for semiconductors (ITRS)," 2011.

[4] N. S. Kim, T. Austin, D. Baauw, T. Mudge, K. Flautner, J. S. Hu, M. J. Irwin, M. Kandemir, and V. Narayanan, "Leakage current: Moore's law meets static power," *Computer*, vol. 36, pp. 68–75, Dec 2003.

[5] M. N. Baibich, J. M. Broto, A. Fert, F. N. Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, and J. Chazelas, "Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices," *Phys. Rev. Lett.*, vol. 61, pp. 2472–2475, Nov 1988.

[6] G. Binasch, P. Grünberg, F. Saurenbach, and W. Zinn, "Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange," *Phys. Rev. B*, vol. 39, pp. 4828–4830, Mar 1989.

[7] M. Julliere, "Tunneling between ferromagnetic films," *Physics Letters A*, vol. 54, no. 3, pp. 225 – 226, 1975.

[8] T. Miyazaki and N. Tezuka, "Giant magnetic tunneling effect in Fe/ Al_2O_3/Fe junction," *Journal of Magnetism and Magnetic Materials*, vol. 139, no. 3, pp. L231 – L234, 1995.

[9] J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, "Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions," *Phys. Rev. Lett.*, vol. 74, pp. 3273–3276, Apr 1995.

[10] D. Wang, C. Nordman, J. M. Daughton, Z. Qian, and J. Fink, "70% at room temperature for SDT sandwich junctions with CoFeB as free and reference layers," *IEEE Transactions on Magnetics*, vol. 40, pp. 2269–2271, July 2004.

[11] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, "Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions," *Nature Materials*, vol. 3, pp. 868–871, 2004.

[12] S. Ikeda, J. Hayakawa, Y. Ashizawa, Y. M. Lee, K. Miura, H. Hasegawa, M. Tsunoda, F. Matsukura, and H. Ohno, "Tunnel magnetoresistance of 604% at 300 K by suppression of ta diffusion in CoFeB/MgO/CoFeB pseudo-spin-valves annealed at high temperature," *Applied Physics Letters*, vol. 93, no. 8, 2008.

[13] C. Chappert, A. Fert, and F. N. Van Dau, "The emergence of spin electronics in data storage," *Nature Materials*, vol. 6, no. 11, pp. 813–823, 2007.

[14] W. J. Gallagher and S. S. P. Parkin, "Development of the magnetic tunnel junction MRAM at IBM: From first junctions to a 16-Mb MRAM demonstrator chip," *IBM J. Res. Dev.*, vol. 50, pp. 5–23, Jan. 2006.

[15] J. Nozieres, B. Dieny, O. Redon, R. Sousa, and I. Prejbeanu, "Magnetic memory with a magnetic tunnel junction written in a thermally assisted manner, and method for writing the same," Sept. 15 2005.

[16] J. C. Slonczewski, "Current-driven excitation of magnetic multilayers," *Journal of Magnetism and Magnetic Materials*, vol. 159, no. 1, pp. L1 – L7, 1996.

[17] L. Berger, "Emission of spin waves by a magnetic multilayer traversed by a current," *Phys. Rev. B*, vol. 54, pp. 9353–9358, Oct 1996.

[18] J. Z. Sun, "Spin-current interaction with a monodomain magnetic body: A model study," *Phys. Rev. B*, vol. 62, pp. 570–578, Jul 2000.

[19] J. A. Katine, F. J. Albert, R. A. Buhrman, E. B. Myers, and D. C. Ralph, "Current-driven magnetization reversal and spin-wave excitations in Co/Cu/Co pillars," *Phys. Rev. Lett.*, vol. 84, pp. 3149–3152, Apr 2000.

[20] S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. D. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, and H. Ohno, "A perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction," *Nature Materials*, vol. 9, pp. 721–724, 2010.

[21] T. Devolder, J. Hayakawa, K. Ito, H. Takahashi, S. Ikeda, P. Crozat, N. Zerounian, J.-V. Kim, C. Chappert, and H. Ohno, "Single-shot time-resolved measurements of nanosecond-scale spin-transfer induced switching: Stochastic versus deterministic aspects," *Phys. Rev. Lett.*, vol. 100, p. 057206, Feb 2008.

[22] H. Cai, *Fiabilisation de convertisseurs analogique-numérique à modulation Sigma-Delta*. PhD thesis, 2013. Thèse de doctorat dirigée par Naviner, Jean-François et Petit, Hervé Electronique et communications Paris, ENST 2013.

[23] R. C. Sousa, I. L. Prejbeanu, D. Stanescu, B. Rodmacq, O. Redon, B. Dieny, J. Wang, and P. P. Freitas, "Tunneling hot spots and heating in magnetic tunnel junctions," *Journal of Applied Physics*, vol. 95, no. 11, pp. 6783–6785, 2004.

[24] W. Oepts, H. J. Verhagen, W. J. M. de Jonge, and R. Coehoorn, "Dielectric breakdown of ferromagnetic tunnel junctions," *Applied Physics Letters*, vol. 73, no. 16, pp. 2363–2365, 1998.

[25] B. Dieny, V. S. Speriosu, S. S. P. Parkin, B. A. Gurney, D. R. Wilhoit, and D. Mauri, "Giant magnetoresistive in soft ferromagnetic multilayers," *Phys. Rev. B*, vol. 43, pp. 1297–1300, Jan 1991.

[26] G. A. Prinz, "Magnetoelectronics," *Science*, vol. 282, p. 1660–1663, 1998.

[27] S. A. Wolf, D. D. Awschalom, R. A. Buhrman, J. M. Daughton, S. von Molnár, M. L. Roukes, A. Y. Chtchelkanova, and D. M. Treger, "Spintronics: A spin-based electronics vision for the future," *Science*, vol. 294, p. 1488–1495, 2001.

[28] Z. Wang, *Modélisation compacte et conception de circuit à base de jonction tunnel ferroélectrique et de jonction tunnel magnétique exploitant le transfert de spin assisté par effet Hall de spin*. PhD thesis, 2015. Thèse de doctorat dirigée par Klein, Jacques-Olivier Physique Paris Saclay 2015.

[29] I. L. Prejbeanu, M. Kerekes, R. C. Sousa, H. Sibuet, O. Redon, B. Dieny, and J. P. Nozières, "Thermally assisted MRAM," *Journal of Physics: Condensed Matter*, vol. 19, no. 16, p. 165218, 2007.

[30] J. Xiao, A. Zangwill, and M. D. Stiles, "Macrospin models of spin transfer dynamics," *Phys. Rev. B*, vol. 72, p. 014446, Jul 2005.

[31] T. L. Gilbert, "A lagrangian formulation of the gyromagnetic equation of the magnetization field," *Phys. Rev.*, vol. 100, p. 1243, 1955.

[32] J.-M. L. Beaujour, D. B. Bedau, H. Liu, M. R. Rogosky, and A. D. Kent, "Spin-transfer in nanopillars with a perpendicularly magnetized spin polarizer," *SPIE Proceedings*, vol. 7398, p. 73980D, 2009.

[33] D. Ralph and M. Stiles, "Spin transfer torques," *Journal of Magnetism and Magnetic Materials*, vol. 321, no. 16, p. 2508, 2009.

[34] J. Deak, "Thermomagnetically assisted spin-momentum-transfer switching memory," 2008. EP Patent App. EP20,050,805,862.

[35] L. Liu, O. J. Lee, T. J. Gudmundsen, D. C. Ralph, and R. A. Buhrman, "Current-induced switching of perpendicularly magnetized magnetic layers using spin torque from the spin hall effect," *Phys. Rev. Lett.*, vol. 109, p. 096602, Aug 2012.

[36] L. Liu, C.-F. Pai, Y. Li, H. W. Tseng, D. C. Ralph, and R. A. Buhrman, "Spin-torque switching with the giant spin hall effect of tantalum," *Science*, vol. 336, pp. 555–558, May 2012.

[37] J. E. Hirsch, "Spin hall effect," *Phys. Rev. Lett.*, vol. 83, pp. 1834–1837, Aug 1999.

[38] W. Zhao, S. Chaudhuri, C. Accoto, J. O. Klein, C. Chappert, and P. Mazoyer, "Cross-point architecture for spin-transfer torque magnetic random access memory," *IEEE Transactions on Nanotechnology*, vol. 11, pp. 907–917, Sept 2012.

[39] S. Yu and P. Y. Chen, "Emerging memory technologies: Recent trends and prospects," *IEEE Solid-State Circuits Magazine*, vol. 8, no. 2, pp. 43–56, 2016.

[40] E. Linn, R. Rosezin, C. Kugeler, and R. Waser, "Complementary resistive switches for passive nanocrossbar memories," *Nat Mater*, vol. 9, pp. 403–406, April 2010.

[41] M. Hosomi, H. Yamagishi, T. Yamamoto, K. Bessho, Y. Higo, K. Yamane, H. Yamada, M. Shoji, H. Hachino, C. Fukumoto, H. Nagao, and H. Kano, "A novel nonvolatile memory with spin torque transfer magnetization switching: spin-ram," in *IEEE InternationalElectron Devices Meeting, 2005. IEDM Technical Digest.*, pp. 459–462, Dec 2005.

[42] T. Endoh, H. Koike, S. Ikeda, T. Hanyu, and H. Ohno, "An overview of nonvolatile emerging memories-spintronics for working memories," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, pp. 109–119, June 2016.

[43] W. H. Kautz, "Cellular logic-in-memory arrays," *IEEE Transactions on Computers*, vol. C-18, pp. 719–727, Aug 1969.

[44] S. Matsunaga, J. Hayakawa, S. Ikeda, K. Miura, T. Endoh, H. Ohno, and T. Hanyu, "MTJ-based nonvolatile logic-in-memory circuit, future prospects and issues," in *2009 Design, Automation Test in Europe Conference Exhibition*, pp. 433–435, April 2009.

[45] S. Ikeda, J. Hayakawa, Y. M. Lee, F. Matsukura, Y. Ohno, T. Hanyu, and H. Ohno, "Magnetic tunnel junctions for spintronic memories and beyond," *IEEE Transactions on Electron Devices*, vol. 54, pp. 991–1002, May 2007.

[46] E. Deng, Y. Zhang, J. O. Klein, D. Ravelsona, C. Chappert, and W. Zhao, "Low power magnetic full-adder based on spin transfer torque MRAM," *IEEE Transactions on Magnetics*, vol. 49, pp. 4982–4987, Sept 2013.

[47] Y. Gang, W. Zhao, J.-O. Klein, C. Chappert, and P. Mazoyer, "A high-reliability, low-power magnetic full adder," *Magnetics, IEEE Transactions on*, vol. 47, pp. 4611–4616, Nov 2011.

[48] W. Zhao, E. Belhaire, C. Chappert, F. Jacquet, and P. Mazoyer, "New non-volatile logic based on spin-MTJ," *Phys. Stat. Sol.*, vol. (a) 205, no. 6, pp. 1373–1377, 2008.

[49] W. Zhao, C. Chappert, V. Javerliac, and J. P. Noziere, "High speed, high stability and low power sensing amplifier for MTJ/CMOS hybrid logic circuits," *IEEE Transactions on Magnetics*, vol. 45, pp. 3784–3787, Oct 2009.

[50] Z. M. Zeng, P. Khalili Amiri, J. A. Katine, J. Langer, K. L. Wang, and H. W. Jiang, "Nanoscale magnetic tunnel junction sensors with perpendicular anisotropy sensing layer," *Applied Physics Letters*, vol. 101, no. 6, 2012.

[51] S. Cardoso, D. C. Leitao, L. Gameiro, F. Cardoso, R. Ferreira, E. Paz, and P. P. Freitas, "Magnetic tunnel junction sensors with ptesla sensitivity," *Microsystem Technologies*, vol. 20, no. 4, pp. 793–802, 2014.

[52] B. Behin-Aein, D. Datta, S. Salahuddin, and S. Datta, "Proposal for an all-spin logic device with built-in memory," *Nat Nano*, vol. 4, pp. 266–270, February 2010.

[53] J. Kim, A. Paul, P. A. Crowell, S. J. Koester, S. S. Sapatnekar, J. P. Wang, and C. H. Kim, "Spin-based computing: Device concepts, current status, and a case study on a high-performance microprocessor," *Proceedings of the IEEE*, vol. 103, pp. 106–130, Jan 2015.

[54] O. www.itrs.net/Links/2011ITRS/Home2011.htm, "International technology roadmap for semiconductors (ITRS)," 2013.

[55] G. Gielen, P. De Wit, E. Maricau, J. Loeckx, J. Martin-Martinez, B. Kaczer, G. Groeseneken, R. Rodriguez, and M. Nafria, "Emerging yield and reliability challenges in nanometer CMOS technologies," *Proc. Design, Automation and Test*, pp. 1322–1327, 2008.

[56] J. Li, C. Augustine, S. Salahuddin, and K. Roy, "Modeling of failure probability and statistical design of spin-torque transfer magnetic random access memory (STT MRAM) array for yield enhancement," in *Design Automation Conference, 2008. DAC 2008. 45th ACM/IEEE*, pp. 278–283, June 2008.

[57] J. F. Kong, K. Eason, K. P. Tan, and R. Sbiaa, "Parameter variation investigation of magnetic tunnel junctions," in *2012 Digest APMRC*, pp. 1–2, Oct 2012.

[58] A. Vatankhahghadim, S. Huda, and A. Sheikholeslami, "A survey on circuit modeling of spin-transfer-torque magnetic tunnel junctions," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, pp. 2634–2643, Sept 2014.

[59] E. Y. Chen, R. Whig, J. M. Slaughter, D. Cronk, J. Goggin, G. Steiner, and S. Tehrani, "Comparison of oxidation methods for magnetic tunnel junction material," *Journal of Applied Physics*, vol. 87, no. 9, pp. 6061–6063, 2000.

[60] M. Natsui, T. Nagashima, and T. Hanyu, "Process-variation-resilient OTA using mtj-based multi-level resistance control," in *2012 IEEE 42nd International Symposium on Multiple-Valued Logic*, pp. 214–219, May 2012.

[61] W. S. Zhao, Y. Zhang, T. , J. O. Klein, D. Ravelosona, C. Chappert, and P. Mazoyer, "Failure and reliability analysis of STT-MRAM," *Microelectronics Reliability*, vol. 52, pp. 1848–1852, 2012.

[62] K. Ono, T. Kawahara, R. Takemura, K. Miura, H. Yamamoto, M. Yamanouchi, J. Hayakawa, K. Ito, H. Takahashi, S. Ikeda, H. Hasegawa, H. Matsuoka, and

H. Ohno, "A disturbance-free read scheme and a compact stochastic-spin-dynamics-based MTJ circuit model for Gb-scale SPRAM," in *2009 IEEE International Electron Devices Meeting (IEDM)*, pp. 1–4, Dec 2009.

[63] T. Devolder, C. Chappert, and K. Ito, "Subnanosecond spin-transfer switching: Comparing the benefits of free-layer or pinned-layer biasing," *Phys. Rev. B*, vol. 75, p. 224430, Jun 2007.

[64] M. Marins de Castro, R. C. Sousa, S. Bandiera, C. Ducruet, A. Chavent, S. Auffret, C. Papusoi, I. L. Prejbeanu, C. Portemont, L. Vila, U. Ebels, B. Rodmacq, and B. Dieny, "Precessional spin-transfer switching in a magnetic tunnel junction with a synthetic antiferromagnetic perpendicular polarizer," *Journal of Applied Physics*, vol. 111, no. 7, p. 07C912, 2012.

[65] H. Tomita, S. Miwa, T. Nozaki, S. Yamashita, T. Nagase, K. Nishiyama, E. Kitagawa, M. Yoshikawa, T. Daibou, M. Nagamine, T. Kishi, S. Ikegawa, N. Shimomura, H. Yoda, and Y. Suzuki, "Unified understanding of both thermally assisted and precessional spin-transfer switching in perpendicularly magnetized giant magnetoresistive nanopillars," *Applied Physics Letters*, vol. 102, no. 4, 2013.

[66] J. Z. Sun, R. P. Robertazzi, J. Nowak, P. L. Trouilloud, G. Hu, D. W. Abraham, M. C. Gaidis, S. L. Brown, E. J. O'Sullivan, W. J. Gallagher, and D. C. Worledge, "Effect of subvolume excitation and spin-torque efficiency on magnetic switching," *Phys. Rev. B*, vol. 84, p. 064413, Aug 2011.

[67] H. Zhao, Y. Zhang, P. K. Amiri, J. A. Katine, J. Langer, H. Jiang, I. N. Krivorotov, K. L. Wang, and J. P. Wang, "Spin-torque driven switching probability density function asymmetry," *IEEE Transactions on Magnetics*, vol. 48, pp. 3818–3820, Nov 2012.

[68] L. B. Faber, W. Zhao, J. O. Klein, T. Devolder, and C. Chappert, "Dynamic compact model of spin-transfer torque based magnetic tunnel junction (MTJ)," in *2009 4th International Conference on Design Technology of Integrated Systems in Nanoscal Era*, pp. 130–135, April 2009.

[69] H. Lim, S. Lee, and H. Shin, "Advanced circuit-level model for temperature-sensitive read/write operation of a magnetic tunnel junction," *IEEE Transactions on Electron Devices*, vol. 62, pp. 666–672, Feb 2015.

[70] S. Chatterjee, S. Salahuddin, S. Kumar, and S. Mukhopadhyay, "Impact of self-heating on reliability of a spin-torque-transfer RAM cell," *IEEE Transactions on Electron Devices*, vol. 59, pp. 791–799, March 2012.

[71] J. J. Kan, "Engineering of metallic multilayers and spin transfer torque devices," *UC San Diego: Materials science and engineering*, no. b8163492, 2014.

[72] A. A. Khan, J. Schmalhorst, A. Thomas, O. Schebaum, and G. Reiss, "Dielectric breakdown in Co-Fe-B/MgO/Co-Fe-B magnetic tunnel junction," *Journal of Applied Physics*, vol. 103, no. 12, 2008.

[73] S. Amara-Dababi, R. C. Sousa, M. Chshiev, H. Bea, J. Alvarez-Herault, L. Lombard, I. L. Prejbeanu, K. Mackay, and B. Dieny, "Charge trapping-detrapping mechanism of barrier breakdown in MgO magnetic tunnel junctions," *Applied Physics Letters*, vol. 99, no. 8, 2011.

[74] K. Hosotani, M. Nagamine, T. Ueda, H. Aikawa, S. Ikegawa, Y. Asao, H. Yoda, and A. Nitayama, "Effect of self-heating on time-dependent dielectric breakdown in ultrathin MgO magnetic tunnel junctions for spin torque transfer switching magnetic random access memory," *Japanese Journal of Applied Physics*, vol. 49, no. 04DD15, 2010.

[75] D. V. Dimitrov, Z. Gao, X. Wang, W. Jung, X. Lou, and O. G. Heinonen, "Dielectric breakdown of MgO magnetic tunnel junctions," *Applied Physics Letters*, vol. 94, no. 12, 2009.

[76] C. Yoshida, M. Kurasawa, Y. M. Lee, K. Tsunoda, M. Aoki, and Y. Sugiyama, "A study of dielectric breakdown mechanism in CoFeB/MgO/CoFeB magnetic tunnel junction," in *Reliability Physics Symposium, 2009 IEEE International*, pp. 139–142, April 2009.

[77] B. Oliver, Q. He, X. Tang, and J. Nowak, "Dielectric breakdown in magnetic tunnel junctions having an ultrathin barrier," *Journal of Applied Physics*, vol. 91, no. 7, pp. 4348–4352, 2002.

[78] K.-S. Kim, Y. M. Jang, C. H. Nam, K.-S. Lee, and B. K. Cho, "Stress polarity dependence of breakdown characteristics in magnetic tunnel junctions," *Journal of Applied Physics*, vol. 99, no. 8, 2006.

[79] S. Y. Kim, G. Panagopoulos, C.-H. Ho, M. Katoozi, E. Cannon, and K. Roy, "A compact SPICE model for statistical post-breakdown gate current increase due to TDDB," in *Reliability Physics Symposium (IRPS), 2013 IEEE International*, pp. 2A.2.1–2.4, April 2013.

[80] C.-H. Ho, G. Panagopoulos, S. Y. Kim, Y. Kim, D. Lee, and K. Roy, "A physics-based statistical model for reliability of STT-MRAM considering oxide variability," in *Simulation of Semiconductor Processes and Devices (SISPAD), 2013 International Conference on*, pp. 29–32, 2013.

[81] Y. Zhang, W. S. Zhao, Y. Lakys, J. O. Klein, J.-V. Kim, D. Ravelosona, and C. Chappert, "Compact modeling of perpendicular-anisotropy CoFeB/MgO magnetic tunnel junctions," *Electron Devices, IEEE Transactions on*, vol. 59, pp. 819–826, March 2012.

[82] G. D. Panagopoulos, C. Augustine, and K. Roy, "Physics-based spice-compatible compact model for simulating hybrid MTJ/CMOS circuits," *IEEE Transactions on Electron Devices*, vol. 60, pp. 2808–2814, Sept 2013.

[83] W. Zhao, X. Zhao, B. Zhang, K. Cao, L. Wang, W. Kang, Q. Shi, M. Wang, Y. Zhang, Y. Wang, S. Peng, J.-O. Klein, L. A. de Barros Naviner, and D. Ravelosona, "Failure analysis in magnetic tunnel junction nanopillar with interfacial perpendicular magnetic anisotropy," *Materials*, vol. 9, no. 1, p. 41, 2016.

[84] H. X. Yang, M. Chshiev, B. Dieny, J. H. Lee, A. Manchon, and K. H. Shin, "First-principles investigation of the very large perpendicular magnetic anisotropy at Fe|MgO and Co|MgO interfaces," *Phys. Rev. B*, vol. 84, p. 054401, Aug 2011.

[85] M. Yamanouchi, R. Koizumi, S. Ikeda, H. Sato, K. Mizunuma, K. Miura, H. D. Gan, F. Matsukura, and H. Ohno, "Dependence of magnetic anisotropy on MgO thickness and buffer layer in $Co_{20}Fe_{60}B_{20} - MgO$ structure," *Journal of Applied Physics*, vol. 109, no. 7, p. 07C712, 2011.

[86] M. Gottwald, J. J. Kan, K. Lee, X. Zhu, C. Park, and S. H. Kang, "Scalable and thermally robust perpendicular magnetic tunnel junctions for STT-MRAM," *Applied Physics Letters*, vol. 106, no. 3, p. 032413, 2015.

[87] H. Maehara, K. Nishimura, Y. Nagamine, K. Tsunekawa, T. Seki, H. Kubota, A. Fukushima, K. Yakushiji, K. Ando, and S. Yuasa, "Tunnel magnetoresistance

above 170% and resistance–area product of 1 Ω ($\mu\ m^2$) attained by In situ annealing of ultra-thin MgO tunnel barrier," *Applied Physics Express*, vol. 4, no. 3, p. 033002.

[88] E. Chen, B. Schwarz, C. J. Choi, W. Kula, J. Wolfman, K. Ounadjela, and S. Geha, "Magnetic tunnel junction pattern technique," *Journal of Applied Physics*, vol. 93, no. 10, pp. 8379–8381, 2003.

[89] M. Nakayama, T. Kai, N. Shimomura, M. Amano, E. Kitagawa, T. Nagase, M. Yoshikawa, T. Kishi, S. Ikegawa, and H. Yoda, "Spin transfer switching in TbCoFe/CoFeB/CoFeB/TbCoFe magnetic tunnel junctions with perpendicular magnetic anisotropy," *Journal of Applied Physics*, vol. 103, no. 7, p. 07A710, 2008.

[90] J. G. Alzate, P. Khalili Amiri, G. Yu, P. Upadhyaya, J. A. Katine, J. Langer, B. Ocker, I. N. Krivorotov, and K. L. Wang, "Temperature dependence of the voltage-controlled perpendicular anisotropy in nanoscale MgO/CoFeB/Ta magnetic tunnel junctions," *Applied Physics Letters*, vol. 104, no. 11, p. 112410, 2014.

[91] R. H. Koch, J. A. Katine, and J. Z. Sun, "Time-resolved reversal of spin-transfer switching in a nanomagnet," *Phys. Rev. Lett.*, vol. 92, p. 088302, Feb 2004.

[92] D. H. Lee and S. H. Lim, "Increase of temperature due to joule heating during current-induced magnetization switching of an MgO-based magnetic tunnel junction," *Applied Physics Letters*, vol. 92, no. 23, 2008.

[93] W. F. Brinkman, R. C. Dynes, and J. M. Rowell, "Tunneling conductance of asymmetrical barriers," *Journal of Applied Physics*, vol. 41, no. 5, pp. 1915–1921, 1970.

[94] G. D. Fuchs, I. N. Krivorotov, P. M. Braganca, N. C. Emley, A. G. F. Garcia, D. C. Ralph, and R. A. Buhrman, "Adjustable spin torque in magnetic tunnel junctions with two fixed layers," *Applied Physics Letters*, vol. 86, no. 15, 2005.

[95] D. C. Worledge, G. Hu, D. W. Abraham, J. Z. Sun, P. L. Trouilloud, J. Nowak, S. Brown, M. C. Gaidis, E. J. O'Sullivan, and R. P. Robertazzi, "Spin torque switching of perpendicular Ta/CoFeB/MgO-based magnetic tunnel junctions," *Applied Physics Letters*, vol. 98, no. 2, 2011.

[96] R. Heindl, W. H. Rippard, S. E. Russek, M. R. Pufall, and A. B. Kos, "Validity of the thermal activation model for spin-transfer torque switching in magnetic tunnel junctions," *Journal of Applied Physics*, vol. 109, no. 7, 2011.

[97] H. Tomita, T. Nozaki, T. Seki, T. Nagase, K. Nishiyama, E. Kitagawa, M. Yoshikawa, T. Daibou, M. Nagamine, T. Kishi, S. Ikegawa, N. Shimomura, H. Yoda, and Y. Suzuki, "High-speed spin-transfer switching in GMR nano-pillars with perpendicular anisotropy," *Magnetics, IEEE Transactions on*, vol. 47, no. 6, pp. 1599–1602, 2011.

[98] Y. Wang, Y. Zhang, E. Y. Deng, J. O. Klein, L. Naviner, and W. S. Zhao, "Compact model of magnetic tunnel junction with stochastic spin transfer torque switching for reliability analyses," *Microelectronics Reliability*, vol. 54, pp. 1774–1778, 2014.

[99] V. Drewello, J. Schmalhorst, A. Thomas, and G. Reiss, "Evidence for strong magnon contribution to the TMR temperature dependence in MgO based tunnel junctions," *Phys. Rev. B*, vol. 77, p. 014440, Jan 2008.

[100] R. Takemura, T. Kawahara, K. Miura, H. Yamamoto, J. Hayakawa, N. Matsuzaki, K. Ono, M. Yamanouchi, K. Ito, H. Takahashi, S. Ikeda, H. Hasegawa, H. Matsuoka, and H. Ohno, "A 32-Mb SPRAM with 2T1R memory cell, localized bi-directional write driver and '1'/'0' dual-array equalized reference scheme," *IEEE Journal of Solid-State Circuits*, vol. 45, pp. 869–879, April 2010.

[101] M. Schafers, V. Drewello, G. Reiss, A. Thomas, K. Thiel, G. Eilers, M. Munzenberg, H. Schuhmann, and M. Seibt, "Electric breakdown in ultrathin mgo tunnel barrier junctions for spin-transfer torque switching," *Applied Physics Letters*, vol. 95, no. 23, 2009.

[102] J. W. McPherson, "Time dependent dielectric breakdown physics - models revisited," *Microelectronics Reliability*, vol. 52, pp. 1753–1760, 2012.

[103] B. Oliver, G. Tuttle, Q. He, X. Tang, and J. Nowak, "Two breakdown mechanisms in ultrathin alumina barrier magnetic tunnel junctions," *Journal of Applied Physics*, vol. 95, no. 3, pp. 1315–1322, 2004.

[104] S. Amara-Dababi, H. Bea, R. Sousa, K. Mackay, and B. Dieny, "Modelling of time-dependent dielectric barrier breakdown mechanisms in MgO-based magnetic tunnel junctions," *Journal of Physics D: Applied Physics*, vol. 45, p. 295002, 2012.

[105] W. Zhao, *Conception, evaluation and development of the non-volatile programmable logic circuits using the Magnetic Tunnel Junction (MTJ).* PhD thesis, 2008. Thèse de doctorat dirigée par Belhaire, Eric et Chappert, Claude Physique Université de Paris-Sud. Faculté des Sciences d'Orsay (Essonne) 2008.

[106] G. J. Coram, "How to (and how not to) write a compact model in verilog-A," in *Proceedings of the 2004 IEEE International Behavioral Modeling and Simulation Conference, 2004. BMAS 2004.*, pp. 97–106, Oct 2004.

[107] *Virtuoso Spectre circuit simulator datasheet, Cadence.*

[108] *Eldo custom Design & Simulation simulator datasheet, Mentor-Graphics.*

[109] *Advanced Design System (ADS) simulator datasheet, Agilent.*

[110] T. Min, Q. Chen, R. Beach, G. Jan, H. Cheng, W. Kula, T. Torng, R. Tong, T. Zhong, D. Tang, P. Wang, M.-m. Chen, J. Sun, J. DeBrosse, D. Worledge, T. Maffitt, and W. Gallagher, "A study of write margin of spin torque transfer magnetic random access memory technology," *Magnetics, IEEE Transactions on*, vol. 46, pp. 2322–2327, June 2010.

[111] H. Mostafa and Y. Ismail, "Process variation aware design of multi-valued spintronic memristor-based memory arrays," *IEEE Transactions on Semiconductor Manufacturing*, vol. 29, pp. 145–152, May 2016.

[112] V. Veetil, D. Sylvester, and D. Blaauw, "Efficient monte carlo based incremental statistical timing analysis," in *Design Automation Conference, 2008. DAC 2008. 45th ACM/IEEE*, pp. 676–681, June 2008.

[113] C. H. Lin, M. V. Dunga, D. D. Lu, A. M. Niknejad, and C. Hu, "Performance-aware corner model for design for manufacturing," *IEEE Transactions on Electron Devices*, vol. 56, April 2009.

[114] S. K. Saha, "Compact mosfet modeling for process variability-aware VLSI circuit design," *IEEE Access*, vol. 2, pp. 104–115, 2014.

[115] M. B. Yelten, P. D. Franzon, and M. B. Steer, "Surrogate-model-based analysis of analog circuits-part I: Variability analysis," *IEEE Transactions on Device and Materials Reliability*, vol. 11, pp. 458–465, Sept 2011.

[116] Y. L. Chen, W. R. Wu, C. N. J. Liu, and J. C. M. Li, "Simultaneous optimization of analog circuits with reliability and variability for applications on flexible electronics," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 33, pp. 24–35, Jan 2014.

[117] H. Cai, Y. Wang, L. A. de Barros Naviner, and W. Zhao, "Low power magnetic flip-flop optimization with FDSOI technology boost," *IEEE Transactions on Magnetics*, vol. 52, no. 8, p. 3401807, 2016.

[118] Y. Wang, H. Cai, L. A. de Barros Naviner, and W. Zhao, "A non-monte-carlo methodology for variability analysis of magnetic tunnel junction based circuits," *IEEE Transactions on Magnetics*, 2017.

[119] B. Cheng, D. Dideban, N. Moezi, C. Millar, G. Roy, X. Wang, S. Roy, and A. Asenov, "Statistical-variability compact-modeling strategies for BSIM4 and PSP," *IEEE Design Test of Computers*, vol. 27, pp. 26–35, March 2010.

[120] D. Chabi, W. Zhao, E. Deng, Y. Zhang, N. B. Romdhane, J. O. Klein, and C. Chappert, "Ultra low power magnetic flip-flop based on checkpointing/power gating and self-enable mechanisms," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, pp. 1755–1765, June 2014.

[121] Y. Wang, H. Cai, L. Naviner, Y. Zhang, J. Klein, and W. Zhao, "Compact thermal modeling of spin transfer torque magnetic tunnel junction," *Microelectronics Reliability*, vol. 55, no. 9–10, pp. 1649 – 1653, 2015.

[122] F. Ahmed and L. Milor, "Analysis and on-chip monitoring of gate oxide breakdown in SRAM cells," *Very Large Scale Integration (VLSI) Systems, IEEE Trans. on*, vol. 20, pp. 855–864, May 2012.

[123] S. Cheffah, V. Huard, R. Chevallier, and A. Bravaix, "Soft oxide breakdown impact on the functionality of a 40 nm SRAM memory," in *Reliability Physics Symposium (IRPS), 2011 IEEE International*, pp. CR.3.1–CR.3.2, April 2011.

[124] M. Saliva, F. Cacho, V. Huard, D. Angot, X. Federspiel, M. Durand, M. Parra, A. Bravaix, and L. Anghel, "New insights about oxide breakdown occurrence at circuit level," in *Reliability Physics Symposium, 2014 IEEE International*, pp. 2D.5.1–2D.5.6, June 2014.

[125] H. Nan and K. Choi, "TDDB monitoring and compensation circuit design for deeply scaled CMOS technology," *Device and Materials Reliability, IEEE Trans. on*, vol. 13, pp. 18–25, March 2013.

[126] S. Knebel, S. Kupke, U. Schroeder, S. Slesazeck, T. Mikolajick, R. Agaiby, and M. Trentzsch, "Influence of frequency dependent time to breakdown on high-k/metal

gate reliability," *Electron Devices, IEEE Trans. on*, vol. 60, pp. 2368–2371, July 2013.

[127] M. Saliva, F. Cacho, C. Ndiaye, V. Huard, D. Angot, A. Bravaix, and L. Anghel, "Impact of gate oxide breakdown in logic gates from 28nm FDSOI CMOS technology," in *Reliability Physics Symposium (IRPS), 2015 IEEE International*, pp. CA.4.1–CA.4.6, April 2015.

[128] R. Rodriguez, R. Joshi, J. Stathis, and C. Chuang, "Oxide breakdown model and its impact on SRAM cell functionality," in *Simulation of Semiconductor Processes and Devices, 2003. SISPAD 2003. International Conference on*, pp. 283–286, Sept 2003.

[129] H. Cai, Y. Wang, L. A. d. B. Naviner, and W. Zhao, "Breakdown analysis of magnetic flip-flop with 28-nm UTBB FDSOI technology," *IEEE Transactions on Device and Materials Reliability*, vol. 16, pp. 376–383, Sept 2016.

[130] Y. Chen, H. Li, X. Wang, W. Zhu, W. Xu, and T. Zhang, "A 130 nm 1.2 V/3.3 V 16 Kb spin-transfer torque random access memory with nondestructive self-reference sensing scheme," *IEEE Journal of Solid-State Circuits*, vol. 47, pp. 560–573, Feb 2012.

[131] Y. Wang, H. Cai, L. A. d. B. Naviner, Y. Zhang, X. Zhao, E. Deng, J. O. Klein, and W. Zhao, "Compact model of dielectric breakdown in spin-transfer torque magnetic tunnel junction," *IEEE Transactions on Electron Devices*, vol. 63, pp. 1762–1767, April 2016.

[132] T. Ohsawa, S. Ikeda, T. Hanyu, H. Ohno, and T. Endoh, "Trend of tunnel magnetoresistance and variation in threshold voltage for keeping data load robustness of metal-oxide -semiconductor/magnetic tunnel junction hybrid latches," *Journal of Applied Physics*, vol. 115, no. 17, p. 17C728, 2014.

[133] S. Saxena, C. Hess, H. Karbasi, A. Rossoni, S. Tonello, P. McNamara, S. Lucherini, S. Minehane, C. Dolainsky, and M. Quarantelli, "Variation in transistor performance and leakage in nanometer-scale technologies," *IEEE Transactions on Electron Devices*, vol. 55, pp. 131–144, Jan 2008.

[134] K. J. Kuhn, "Considerations for ultimate CMOS scaling," *IEEE Transactions on Electron Devices*, vol. 59, pp. 1813–1828, July 2012.

[135] H. Cai, Y. Wang, W. Zhao, and L. de Barros Naviner, "Multiplexing sense-amplifier-based magnetic flip-flop in a 28-nm FDSOI technology," *Nanotechnology, IEEE Transactions on*, vol. 14, no. 4, pp. 761–767, 2015.

[136] I. Kazi, P. Meinerzhagen, P. E. Gaillardon, D. Sacchetto, Y. Leblebici, A. Burg, and G. D. Micheli, "Energy/reliability trade-offs in low-voltage ReRAM-based non-volatile flip-flop design," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, pp. 3155–3164, Nov 2014.

[137] K. Ryu, J. Kim, J. Jung, J. P. Kim, S. H. Kang, and S. O. Jung, "A magnetic tunnel junction based zero standby leakage current retention flip-flop," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, pp. 2044–2053, Nov 2012.

[138] Y. Wang, H. Cai, L. Naviner, X. Zhao, Y. Zhang, M. Slimani, J. Klein, and W. Zhao, "A process-variation-resilient methodology of circuit design by using asymmetrical forward body bias in 28 nm FDSOI," *Microelectronics Reliability*, vol. 64, pp. 26 – 30, 2016.

[139] T. Skotnicki, C. Fenouillet-Beranger, C. Gallon, F. Buf, S. Monfray, F. Payet, A. Pouydebasque, M. Szczap, A. Farcy, F. Arnaud, S. Clerc, M. Sellier, A. Cathignol, J.-P. Schoellkopf, E. Perea, R. Ferrant, and H. Mingam, "Innovative materials, devices, and CMOS technologies for low-power mobile multimedia," *Electron Devices, IEEE Transactions on*, vol. 55, pp. 96–130, Jan 2008.

[140] N. Planes, O. Weber, V. Barral, S. Haendler, D. Noblet, D. Croain, M. Bocat, P. Sassoulas, X. Federspiel, A. Cros, A. Bajolet, E. Richard, B. Dumont, P. Perreau, D. Petit, D. Golanski, C. Fenouillet-Beranger, N. Guillot, M. Rafik, V. Huard, S. Puget, X. Montagner, M.-A. Jaud, O. Rozeau, O. Saxod, F. Wacquant, F. Monsieur, D. Barge, L. Pinzelli, M. Mellier, F. Boeuf, F. Arnaud, and M. Haond, "28nm FDSOI technology platform for high-speed low-voltage digital applications," in *VLSI Technology (VLSIT), 2012 Symposium on*, pp. 133–134, June 2012.

[141] P. Magarshack, P. Flatresse, and G. Cesana, "UTBB FD-SOI: A process/design symbiosis for breakthrough energy-efficiency," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2013*, pp. 952–957, March 2013.

[142] S. Vitale, P. Wyatt, N. Checka, J. Kedzierski, and C. Keast, "FDSOI process technology for subthreshold-operation ultralow-power electronics," *Proceedings of the IEEE*, vol. 98, pp. 333–342, Feb 2010.

[143] Y. Yang, S. Markov, B. Cheng, A. Zain, X. Liu, and A. Cheng, "Back-gate bias dependence of the statistical variability of FDSOI MOSFETs with thin BOX," *Electron Devices, IEEE Transactions on*, vol. 60, pp. 739–745, Feb 2013.

[144] H. Cai, H. Petit, and J.-F. Naviner, "Reliability aware design in low power continuous-time sigma-delta modulator," *Microelectronics reliability journal*, vol. 51, no. 9-11, pp. 1449–1453, 2011.

[145] E. Maricau and G. Gielen, "Computer-aided analog circuit design for reliability in nanometer CMOS," *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on*, vol. 1, pp. 50–58, March 2011.

[146] T. Ishigaki, R. Tsuchiya, Y. Morita, N. Sugii, and S. Kimura, "Effects of device structure and back biasing on HCI and NBTI in silicon-on-thin-BOX (SOTB) CMOSFET," *Electron Devices, IEEE Transactions on*, vol. 58, pp. 1197–1204, April 2011.

[147] H. Cai, Y. Wang, L. Naviner, and W. Zhao, "Ultra wide voltage range consideration of reliability-aware STT magnetic flip-flop in 28 nm FDSOI technology," *Microelectronics Reliability*, vol. 55, no. 9–10, pp. 1323 – 1327, 2015.

[148] H. Cai, Y. Wang, L. A. D. B. Naviner, and W. Zhao, "Robust ultra-low power non-volatile logic-in-memory circuits in FDSOI technology," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. PP, no. 99, pp. 1–11, 2016.

[149] 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, July 2016.

[150] M. Bucci, L. Germani, R. Luzzi, A. Trifiletti, and M. Varanonuovo, "A high-speed oscillator-based truly random number source for cryptographic applications on a smart card IC," *IEEE Transactions on Computers*, vol. 52, pp. 403–409, April 2003.

[151] R. F. W. Coates, G. J. Janacek, and K. V. Lever, "Monte carlo simulation and random number generation," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 58–66, Jan 1988.

[152] K. Yang, D. Fick, M. B. Henry, Y. Lee, D. Blaauw, and D. Sylvester, "A 23Mb/s 23pJ/b fully synthesized true-random-number generator in 28nm and 65nm CMOS,"

in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, pp. 280–281, Feb 2014.

[153] N. Liu, N. Pinckney, S. Hanson, D. Sylvester, and D. Blaauw, "A true random number generator using time-dependent dielectric breakdown," in *VLSI Circuits (VLSIC), 2011 Symposium on*, pp. 216–217, June 2011.

[154] J. Rajendran, R. Karri, J. B. Wendt, M. Potkonjak, N. McDonald, G. S. Rose, and B. Wysocki, "Nano meets security: Exploring nanoelectronic devices for security applications," *Proceedings of the IEEE*, vol. 103, pp. 829–849, May 2015.

[155] X. Fong, Y. Kim, K. Yogendra, D. Fan, A. Sengupta, A. Raghunathan, and K. Roy, "Spin-transfer torque devices for logic and memory: Prospects and perspectives," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 35, pp. 1–22, Jan 2016.

[156] 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, no. 8, p. 083001, 2014.

[157] Y. Wang, W. Wen, H. Li, and M. Hu, "A novel true random number generator design leveraging emerging memristor technology," in *Proceedings of the 25th Edition on Great Lakes Symposium on VLSI*, GLSVLSI '15, (New York, NY, USA), pp. 271–276, ACM, 2015.

[158] X. Fong, M. C. Chen, and K. Roy, "Generating true random numbers using on-chip complementary polarizer spin-transfer torque magnetic tunnel junctions," in *72nd Device Research Conference*, pp. 103–104, June 2014.

[159] 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 *New Circuits and Systems Conference (NEWCAS), 2015 IEEE 13th International*, pp. 1–4, June 2015.

[160] K. Lee, T. Kim, X. Zhu, D. Jacobson, R. Madala, W. Wu, J. Kim, and S. Kang, "Magnetic tunnel junction based random number generator," Apr. 17 2014. US Patent App. 13/651,954.

[161] E. I. Vatajelu, G. D. Natale, and P. Prinetto, "Security primitives (PUF and TRNG) with STT-MRAM," in *2016 IEEE 34th VLSI Test Symposium (VTS)*, pp. 1–4, April 2016.

[162] J. Mazurier, O. Weber, F. Andrieu, A. Toffoli, O. Rozeau, T. Poiroux, F. Allain, P. Perreau, C. Fenouillet-Beranger, O. Thomas, M. Belleville, and O. Faynot, "On the variability in planar FDSOI technology: From MOSFETs to SRAM cells," *IEEE Transactions on Electron Devices*, vol. 58, pp. 2326–2336, Aug 2011.

[163] G. Cesana, *The FD-SOI technology for very high-speed and energy efficient SoCs*. STMicroelectronics, july 2014.

[164] E. Deng, Y. Zhang, W. Kang, B. Dieny, J. O. Klein, G. Prenat, and W. Zhao, "Synchronous 8-bit non-volatile full-adder based on spin transfer torque magnetic tunnel junction," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, pp. 1757–1765, July 2015.

[165] W. Zhao, M. Moreau, E. Deng, Y. Zhang, J. M. Portal, J. O. Klein, M. Bocquet, H. Aziza, D. Deleruyelle, C. Muller, D. Querlioz, N. B. Romdhane, D. Ravelosona, and C. Chappert, "Synchronous non-volatile logic gate design based on resistive switching memories," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, pp. 443–454, Feb 2014.

[166] J. Soto, *The NIST Statistical Test Suite*. National Institute Of Standards and Technology, april 2010.

[167] L. A. de Barros Naviner, H. Cai, Y. Wang, W. Zhao, and A. B. Dhia, "Stochastic computation with spin torque transfer magnetic tunnel junction," in *New Circuits and Systems Conference (NEWCAS), 2015 IEEE 13th International*, pp. 1–4, June 2015.

[168] B. R. Gaines, "Stochastic computing," in *Proceedings of the April 18-20, 1967, Spring Joint Computer Conference*, AFIPS '67 (Spring), pp. 149–156, 1967.

[169] N. Saraf, K. Bazargan, D. Lilja, and M. Riedel, "Iir filters using stochastic arithmetic," in *Design, Automation and Test in Europe Conference and Exhibition (DATE), 2014*, pp. 1–6, March 2014.

[170] B. Moons and M. Verhelst, "Energy-efficiency and accuracy of stochastic computing circuits in emerging technologies," *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on*, vol. 4, pp. 475–486, Dec 2014.

[171] A. Alaghi, C. Li, and J. Hayes, "Stochastic circuits for real-time image-processing applications," in *Design Automation Conference (DAC), 2013 50th ACM / EDAC / IEEE*, pp. 1–6, May 2013.

[172] W. Qian, X. Li, M. Riedel, K. Bazargan, and D. Lilja, "An architecture for fault-tolerant computation with stochastic logic," *Computers, IEEE Transactions on*, vol. 60, pp. 93–105, Jan 2011.

[173] J. Han, H. Chen, J. Liang, P. Zhu, Z. Yang, and F. Lombardi, "A stochastic computational approach for accurate and efficient reliability evaluation," *Computers, IEEE Transactions on*, vol. 63, pp. 1336–1350, June 2014.

[174] "International technology roadmap for semiconductors (2012)."

[175] N. Locatelli, V. Cros, and J. Grollier, "Spin-torque building blocks," *Nature Materials*, vol. 13, pp. 1476–1122, 2014.

[176] K. Ryu, J. Kim, J. Jung, J. Kim, S. Kang, and S.-O. Jung, "A magnetic tunnel junction based zero standby leakage current retention flip-flop," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 20, pp. 2044–2053, Nov 2012.

[177] N. Sakimura, T. Sugibayashi, R. Nebashi, and N. Kasai, "Nonvolatile magnetic flip-flop for standby-power-free socs," *Solid-State Circuits, IEEE Journal of*, vol. 44, pp. 2244–2250, Aug 2009.

[178] W. Qian, X. Li, M. Riedel, K. Bazargan, and D. Lilja, "An architecture for fault-tolerant computation with stochastic logic," *Computers, IEEE Transactions on*, vol. 60, pp. 93–105, Jan 2011.

[179] R. Xiao and C. Chen, "Towards power optimization and implementation of probabilistic circuits using single-electron technology," *IEEE Transactions on Nanotechnology*, vol. 14, pp. 513–523, May 2015.

[180] N. R. Shanbhag, R. A. Abdallah, R. Kumar, and D. L. Jones, "Stochastic computation," in *Design Automation Conference (DAC), 2010 47th ACM/IEEE*, pp. 859–864, June 2010.

[181] J. Han and M. Orshansky, "Approximate computing: An emerging paradigm for energy-efficient designs," in *IEEE 18th European Test Symposium (ETS)*, pp. 1–6, January 2013.

[182] S. Venkataramani, S. T. Chakradhar, K. Roy, and A. Raghunathan, "Approximate computing and the quest for computing efficiency," in *Proceedings of the 52Nd Annual Design Automation Conference*, DAC '15, pp. 120:1–120:6, 2015.

[183] Z. Yang, A. Jain, J. Liang, J. Han, and F. Lombardi, "Approximate XOR/XNOR-based adders for inexact computing," in *Nanotechnology (IEEE-NANO), 2013 13th IEEE Conference on*, pp. 690–693, Aug 2013.

[184] M. Natsui, D. Suzuki, N. Sakimura, R. Nebashi, Y. Tsuji, A. Morioka, T. Sugibayashi, S. Miura, H. Honjo, K. Kinoshita, S. Ikeda, T. Endoh, H. Ohno, and T. Hanyu, "Nonvolatile logic-in-memory LSI using cycle-based power gating and its application to motion-vector prediction," *Solid-State Circuits, IEEE Journal of*, vol. 50, pp. 476–489, Feb 2015.

[185] K. Huang, R. Zhao, and Y. Lian, "A low power and high sensing margin non-volatile full adder using racetrack memory," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 62, pp. 1109–1116, April 2015.

[186] A. Ranjan, S. Venkataramani, X. Fong, K. Roy, and A. Raghunathan, "Approximate storage for energy efficient spintronic memories," in *Proceedings of the 52nd Annual Design Automation Conference*, DAC '15, pp. 195:1–195:6, 2015.

[187] J. Liang, J. Han, and F. Lombardi, "New metrics for the reliability of approximate and probabilistic adders," *Computers, IEEE Transactions on*, vol. 62, pp. 1760–1771, Sept 2013.

[188] J.-F. Lin, Y.-T. Hwang, M. hwa Sheu, and C.-C. Ho, "A novel high-speed and energy efficient 10-transistor full adder design," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 54, no. 5, pp. 1050–1059, 2007.

[189] B. Nikolic, M. Blagojevic, O. Thomas, P. Flatresse, and A. Vladimirescu, "Circuit design in nanoscale FDSOI technologies," in *Microelectronics Proceedings - MIEL 2014, 2014 29th International Conference on*, pp. 3–6, 2014.

[190] O. Thomas, B. Zimmer, S. Toh, L. Ciampolini, N. Planes, R. Ranica, P. Flatresse, and B. Nikolic, "Dynamic single-p-well SRAM bitcell characterization with back-

bias adjustment for optimized wide-voltage-range SRAM operation in 28nm UTBB FD-SOI," in *Electron Devices Meeting (IEDM), 2014 IEEE International*, pp. 3.4.1–3.4.4, 2014.

[191] F. Abouzeid, A. Bienfait, K. Akyel, A. Feki, S. Clerc, L. Ciampolini, F. Giner, R. Wilson, and P. Roche, "Scalable 0.35 V to 1.2 V SRAM bitcell design from 65 nm CMOS to 28 nm FDSOI," *Solid-State Circuits, IEEE Journal of*, vol. 49, no. 7, pp. 1499–1505, 2014.

[192] A. Valentian, Y. Thonnart, B. Pelloux-Prayer, and P. Flatresse, "Single-well design in FDSOI technology: Towards energy-efficient ultra-wide voltage range digital circuits," in *SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), 2015 IEEE*, pp. 1–2, 2015.

Appendix A

Source code of STT-PMA-MTJ compact model

In this model, it takes into account the static, dynamic and stochastic behavoirs of PMA MTJ nanopillar

1.MTJ resistance calculation based on brinkman model
2.TMR dependence on the bias voltage
3.Spin polarity calculation model for magnetic tunnel junction
4.Critical current calculation
5.Dynamic model (>critical current, also sun's model)
6.Stochastic model
7.Resistance variation
8.Temperature evaluation
9.Breakdown voltage
10.Lifetime (Time to failure)
11.Breakdown probability
12.Temperature dependent parameters

```
/*--------------------The parameters are from the prototypes of Univ. Tohuku--------------------*/
`resetall
`include "constants.vams"
`include "disciplines.vams"
`define explimit 85.0
`define exp(x) exp(min(max((x),-`explimit),`explimit))
`define sqrt(x) pow( (x), 0.5)

`define rec 1          //Shape definition
`define ellip 2
`define circle 3

/*-----------------------------Electrical Constants-----------------------------*/
/*----------Elementary Charge----------*/
`define e 1.6e-19
/*----------Bohr Magnetron Costant----------*/
`define ub 9.27e-28
/*----------Boltzmann Constant---------- */
`define Kb 1.38e-23
/*----------Electron Mass---------- */
`define m 9.10e-31
/*----------Euler's constant----------*/
`define C 0.577

module Model(T1,T2,Ttrans,Temp,Break);
inout T1, T2;
electrical T1, T2;
electrical n1,n2;    //virtual terminals of RC circuit for temperature evaluation
/*----------Ttrans=store the state of the MTJ with time influence, non-volatile way---------- */
/*----------Temp=store the temperature, Break stores the appearance of breakdown---------- */
inout Ttrans,Temp, Break;
electrical Ttrans,Temp, Break;

/*--------------------MTJ Technology Parameters(Corresponds to the HITACHI MTJ Process)----------*/
/*----------Gilbert Damping Coefficient----------*/
parameter real  alpha=0.027;
/*----------GyroMagnetic Constant in Hz/Oe----------*/
parameter real  gamma=1.76e7;
/*----------Electron Polarization Percentage % ----------*/
parameter real  P=0.52;
/*----------Out of plane Magnetic Anisotropy in Oersteds----------*/
parameter real  Hk0=1433;
```

```
/*----------Saturation Field in the Free Layer in Oersteds----------*/
parameter real  Ms0=15800;
/*----------The Energy Barrier Height for MgO in electron-volt----------*/
parameter real  PhiBas=0.4;
/*----------Voltage bias when the TMR(real) is 1/2TMR(0) in Volt----------*/
parameter real  Vh=0.5;                 //experimental value with MgO barrier

/*----------Device Parameters(Corresponds to the HITACHI 240 x 80 MTJ)----------*/
/*----------Height of the Free Layer  in nm----------*/
parameter real  tsl=1.3e-9 from[0.7e-9:3.0e-9];
/*----------Length in nm----------*/
parameter real  a=40e-9;
/*----------Width in nm----------*/
parameter real  b=40e-9;
/*----------Radius in nm----------*/
parameter real  r=20e-9;
/*----------Height of the Oxide Barrier in nm----------*/
parameter real  tox=8.5e-10 from[8e-10:15e-10];
/*----------TMR(0) with Zero Volt Bias Voltage ----------*/
parameter real  TMR=0.7;

/*----------Shape of MTJ----------*/
parameter real    SHAPE=2  from[1:3];  //SQUARE
/*----------Neel-Brown model parameter ----------*/
parameter real    tau0=8.7e-10;     //experiental value, prototype Hitachi 2007m with CoFe layer
/*----------Error probability Ps=1-Pr(t) ----------*/
parameter real    Ps=0.999999;
/*----------Threshold for Neel-Brown model----------*/
parameter real brown_threshold=0.0;

/*----------MTJ State Parameters----------*/
/*----------Initial state of the MTJ, 0 = parallele, 1 = anti-parallele----*/
parameter integer PAP=1 from[0:1];
/*----------Room temperature in Kelvin----------*/
parameter real T= 300;
/*----------Resistance area product in ohmum2----------*/
parameter real                RA=5 from[5:15];

/*----------Parameters of RC circuit for time modelisation for temperature----------*/
parameter integer Temp_var=0 from[0:1];    //choice of temperature fluctuation
/*----------Heat capacity per unit volume in J/m3*K----------*/
parameter real                Cv= 2.74e6  from[2.735e6:2.7805e6];
/*----------Thermal conductivity of the thermal barrier(MgO) in W/m*K----------*/
parameter real                lam= 84.897 from [84.8912:84.9449];
/*----------Total thickness of MTJ nanopillar in nm----------*/
parameter real                thick_s= 3.355e-8;
/*----------RC circuit for time modelisation for temperature----------*/
parameter real resistor=100e6;
parameter real coeff_tau=12; //Coefficient to increase tau_th
real capacitor;   //virtual capacitor
real tau_th;      //characteristic heating/cooling time
real temp;    //real temperature of MTJ
real temp_init;   //temperature initialised
real R;               //resistance of MTJ

/*----------Parameters for real TMR ratio----------*/
parameter real S=1.5;
```

```
parameter real Em0=1.936e-20;   //121 meV
parameter real epsilon=0.305;         //1/3.279;
parameter real Q=0.025;
parameter real Ec=4.32e-23;               //0.27e-3*1.6e-19;
real Ms, Hk, Beta;

/*---------Parameters for stochastic behaviors----------*/
parameter integer STO=0 from[0:2];       // stochastic dynamic, 0 no stochastic, 1 random exponential distribution, 2 random gauss distribution
parameter integer RV=0 from[0:2];        //choice of process variation intrinsically, 0 no stochastic, 1 random uniform distribution,2 random gauss distribution
parameter real DEV_tox=0.03;       // standard deviation of gauss distribution for tox when RV=2
parameter real DEV_tsl=0.03;       // standard deviation of gauss distribution for tsl when RV=2
parameter real DEV_TMR=0.03;       // standard deviation of gauss distribution for TMRwhen RV=2
parameter real STO_dev=0.03;       // standard deviation of stochastic dynamic gauss distribution when STO=2

/*--------variables----------*/
//Polaristion constant for the two states of STT-MTJ
real PolaP;          //Polarization state parallel of STT-MTJ
real PolaAP;         //Polarization state anti-parallel of STT-MTJ
real surface;        //Surface of MTJ
real gp;  //Critical current density for P state
real gap;   //Critical current density for AP state
real Em,EE;          //Variable of the Slonczewski model
real TMRR;           //TMR real value for P state
real TMRRT;          //TMR real value for AP state
real Ro;  //Resistance of MTJ when bias voltage = 0V
real Rap;            //Resistance value for AP state
real Rp;  //Resistance value for P state

//Voltage of MTJ
real Vb;  //V(T1,T2)
real Vc;  //V(T2,T1)
real Id;   //Current of MTJ

//critial current for the two states of STT-MTJ
real IcAP;           //Critial current for AP state
real IcP;            //Critial current for P state
real ix;   //Current used to store the state of the MTJ
real tau;//Probability parameter
real FA; //Factor for calculating the resistance based on RA

/*--------Stochastic effects----------*/
integer seed;       //Used to initialize the random number generator
real durationstatic, duration;         //time needed to be sure that the switching is effected

real toxreal;   //real thickness of oxide layer
real tslreal;        //real thickness of free layer
real TMRreal;       //real TMR
(*cds_inherited_parameter*)parameter real seedin = 0;  //generation of a random value of seed for random distribution function
(*cds_inherited_parameter*)parameter real seed1 = 0;  //generation of a random value of seed for breakdown
integer seed2;
/*----------switching delay----------*/
real P_APt;
real AP_Pt;
real NP_APt,NAP_Pt;
```

```
/*----------breakdown voltage----------*/
real Vbp_p,Vbp_n,Vbap_p, Vbap_n;

/*----------parameters for calculating the lifetime----------*/
parameter real acc=1.53e-8;  //acceleration parameter  1.53e-8
parameter real H=0.8e-19;  //activation energy parameter 0.8e-19
parameter real beta=1.5;  //shape parameter 1.5

/*----------parameters for calculating the breakdown probability----------*/
real possibilite; //random probability
real F;      //breakdown probability
real xF;    //Weibull function
real TF;  //lifetime
real break;  //breakdown has already occurred or not

analog begin

        if (SHAPE==1)
        begin
                surface=a*b;      //SQUARE
        end
        else if (SHAPE==2)
        begin
                surface=`M_PI*a*b/4;      //ELLIPSE
        end
        else
        begin
                surface=`M_PI*r*r;   //ROUND
        end

        Vc=V(T2,T1);      //potential between T2 and T1
        Vb=V(T1,T2);      //potential between T2 and T1
//initial conditions
@(initial_step)
Begin
        break=0;     //Breakdown doesn't occur at the beginning of simulation
        seed=1000000000*seedin;          //initialization of seed modified 20140516

        seed2=100000000*seed1;

        FA=3322.53/RA;  //initialization of resistance factor according to RA product

        if (RV==1)
        begin  //real thinkness of oxide layer, free layer and real TMR with uniform distribution
                toxreal=$rdist_uniform(seed,(tox-tox*DEV_tox),(tox+tox*DEV_tox));
                tslreal=$rdist_uniform(seed,(tsl-tsl*DEV_tsl),(tsl+tsl*DEV_tsl));
                TMRreal=$rdist_uniform(seed,(TMR-TMR*DEV_TMR),(TMR+TMR*DEV_TMR));
        end
        else if (RV==2)
        begin    //real thinkness of oxide layer, free layer and real TMR with gauss distribution
                        toxreal=abs($rdist_normal(seed,tox,tox*DEV_tox/3));
                        tslreal=abs($rdist_normal(seed,tsl,tsl*DEV_tsl/3));
                        TMRreal=abs($rdist_normal(seed,TMR,TMR*DEV_TMR/3));
        end
        else
        begin
```

```
                    toxreal=tox;
                    tslreal=tsl;
                    TMRreal=TMR;
        end
        temp=T;             //parameters for temperature
        temp_init=T;
        tau_th= Cv*thick_s / (lam/thick_s);
        capacitor=coeff_tau*tau_th/resistor;        //tau_th=resistor*capacitor
        Ro=(toxreal*1.0e10/(FA*`sqrt(PhiBas)*surface*1.0e12))*exp(1.025*toxreal*1.0e10*`sqrt(PhiBas));
        //resistance

        Vbp_p=toxreal*7.6e8+0.202;   //breakdown voltage of parallel, positive bias
        Vbp_n=toxreal*8.3e8+0.206;   //parallel, negative bias
        Vbap_p=toxreal*8.3e8+0.436; //antiparallel, positive bias
        Vbap_n=toxreal*8e8+0.32;    // antiparallel, negative bias

        Em=Ms*tslreal*surface*Hk/2;     //parameters for calculating switching delay
        duration=0.0;
        P_APt=1000000000;
        AP_Pt=1000000000;
        NP_APt=1000000000;
        NAP_Pt=1000000000;
        if(analysis("dc"))          //States inititialisation
        begin
            ix=PAP;
        end

        else
        begin
            ix=-PAP;
        end
end

if(Temp_var==0)
begin
    temp=temp_init;     //temperature is constant
end
else
begin
    temp=V(Temp);       //temperature actualisation
end

Ms=18342*(1-(temp/1120)*sqrt(temp/1120));
Hk=-3*temp+2333;
Em=Ms*tslreal*surface*Hk/2;
EE=Em/(`Kb*temp*40*`M_PI);
Beta=S*`Kb*temp/(Em0*epsilon);

 /*----calculation of real current------*/
 TMRR=1/(1+Vb*Vb/(Vh*Vh))*((TMRreal+1)/(1+2*Q*Beta*log(`Kb*temp/Ec))-1);  //real TMR ratio
 Rp=Ro;
 Rap=Rp*(1+TMRR);

if(break==1)
begin
    R=10;
end
```

```
else if(break==0&&ix==0)
begin
	R=Rp;
end
else
begin
	R=Rap;
end
Id=Vb/R;

 /*----calculation of rcritical current------*/
 PolaP=`sqrt(TMRR*(TMRR+2))/(2*(TMRR+1));	//Polarization state parallel
 gp=alpha*gamma*`e*Ms*tslreal*Hk/(40*`M_PI*(`ub*PolaP));	//Critical current density
 IcP=gp*surface;		// Critical current for P state

 PolaAP=`sqrt(TMRR*(TMRR+2))/(2*(TMRR+1));	//Polarization state anti-parallel
 gap=alpha*gamma*`e*Ms*tslreal*Hk/(40*`M_PI*(`ub*PolaAP));	//Critical current density
 IcAP=gap*surface;		// Critical current for AP state

 /*------Counter of time when real current is higher than critical current */
@(above(Id-IcP,+1))
begin
		P_APt = $abstime;
		NP_APt=1000000000;
end

@(above(-Id-IcAP,+1))
begin
		AP_Pt = $abstime;
		NAP_Pt=1000000000;
end
@(above(Vb-brown_threshold,+1))
begin
		NP_APt = $abstime;
		AP_Pt=1000000000;
		NAP_Pt=1000000000;
end
@(above(Vc-brown_threshold,+1))
begin
		NAP_Pt = $abstime;
		P_APt=1000000000;
		NP_APt=1000000000;
end

if(analysis("dc"))   //dc analysis
begin
		if(ix==0) //Case which the magnetizations of the two layers are parallel
		begin
			if(Vb>=Vbp_p||Vb<=-Vbp_n)
			begin
				R=10;
			end
			else
			begin
				if(Vb>=(IcP*Rp))
				begin
					ix=1.0;
```

```
                        end
                end
        end
        else
        begin
                if(Vb>=Vbap_p||Vb<=-Vbap_n)
                begin
                        R=10;
                end
                else
                begin
                        if(Vc>=(IcAP*Rap))
                        begin
                                ix=0.0;
                        end
                end
        end
        V(Ttrans)<+ix;
        Id=Vb/R;
        I(T1,T2)<+Id;       //Actualisation of the current of MTJ with the value calculated
end
else                //transient analysis
begin

    if(break==0)    // breakdown hasn't occured
    begin
            if(Vb>=Vbp_p||Vb<=-Vbp_n||Vb>=Vbap_p||Vb<=-Vbap_n)
            begin
                break=1;
            end
            possibilite=$rdist_uniform(seed2,0,1);   //a probability between 0 and 1
            TF= exp(H/(`Kb*T)-acc*abs(Vb)/toxreal);    //lifetime of breakdown
            if($abstime<=1e-8)
            begin
                xF=beta*(log(1e-20)-ln(TF));  //If abstract time is too small, the value is defined to avoid bug
            end
            else
            begin
                xF=beta*(log($abstime-1e-8)-ln(TF)+log(exp(1)));    //weibull distribution
            end
            F=1-exp(-exp(xF));    //probability of breakdown
            if(F>=possibilite)
            begin
                break=1;    // random probability <breakdown probability, breakdown occurs
            end
            else
            begin
                break=0;
            end
            if(STO==1||STO==2)    //considering the stochastic behaviors
            begin
                if(ix==0)  //Case which the magnetizations of the two layers are parallel
                begin
                        if(Vb>=IcP*Rp)
```

```
                    begin       //Current higher than critical current, dynamic behavior: Sun model
        durationstatic=(`C+ln(`M_PI*`M_PI*(Em/(`Kb*temp*40*`M_PI))/4))*`e*1000*Ms*surface*tslreal*(1+
P*P)/(4*`M_PI*2*`ub*P*10000*abs(Id-IcP));                   //Average time needed for switching
                        if(STO==1) //stochastic effect(exponential distribution)
                        begin
                            duration=abs($rdist_exponential(seed, durationstatic));
                        end
                        else if(STO==2) //stochastic effect(gauss distribution)
                        begin
                            duration=abs($rdist_normal(seed,durationstatic,durationstatic*STO_dev/3));
                        end
                        else
                        begin
                            duration=durationstatic;
                        end
                        if(duration<=($abstime-P_APt))
                        begin          //Switching of the free layer always occurs
                            ix=-1.0;   //change the current state of MTJ
                        end
                        else
                        begin
                            ix=0.0;
                        end
                    end
                    else
                    begin       //Current smaller than critical current  :  Neel-Brown model
                        ix=0.0;        //save the current state of MTJ
                        tau=tau0*exp(Em*(1-abs(Id/IcP))/(`Kb*temp*40*`M_PI));
                        if(Vb>brown_threshold)
                        begin
                            if (Vb<0.8*IcP*Rp)
                            begin
                                if(STO==1)
                                begin
                                    duration=abs($rdist_exponential(seed, tau));  //stochastic effect
                                end
                                else if(STO==2)
                                begin
                                    duration=abs($rdist_normal(seed,tau,tau*STO_dev/3));
        //stochastic effect(gauss distribution)
                                end
                                else
                                begin
                                        duration=tau;
                                end
                                if (($abstime-NP_APt) >= duration)
                                begin
                                        ix=-1.0;   //change the current state of MTJ
                                end
                                else
                                        begin
                                                ix=0.0;
                                        end
                            end
                            end
                        end
        end  //end of parallel state
```

```
        else      //Case which the magnetizations of the two layers are antiparallel
        begin
            if(Vc>=(IcAP*Rap))
            begin         //Current higher than critical current, dynamic behavior : Sun model
durationstatic=(`C+ln(`M_PI*`M_PI*(Em/(`Kb*temp*40*`M_PI))/4))*`e*1000*Ms*surface*tslreal*(1+P*P)/(4*`
M_PI*2*`ub*P*10000*abs(-Id-IcAP));      //Average time needed for switching
                if(STO==1)
                begin
                        duration=abs($rdist_exponential(seed, durationstatic));        //stochastic effect
                end
                else if(STO==2)
                begin
                       duration=abs($rdist_normal(seed,durationstatic,durationstatic*STO_dev/3.0));
                                //stochastic effect(gauss distribution)
                end
                else
                begin
                        duration=durationstatic;
                end
                if(duration<=($abstime-AP_Pt))
                begin   //Switching of the free layer always occurs
                        ix=0.0;   //change the current state of MTJ
                end
                else
                begin
                        ix=-1.0;
                end
            end
            else
            begin         //Current smaller than critical current, dynamic behavior : Neel-Brown model
                tau=tau0*exp(Em*(1-abs(Id/IcAP))/(`Kb*temp*40*`M_PI));
                if(Vc>brown_threshold)
                begin
                        if (Vc<0.8*IcAP*Rap)
                        begin
                                if(STO==1)
                                begin
                                        duration=abs($rdist_exponential(seed, tau)); //stochastic effect
                                end
                                else if(STO==2)
                                begin
                                        duration=abs($rdist_normal(seed,tau,tau*STO_dev/3));// gauss
                                end
                                else
                                begin
                                        duration=tau;
                                end
                                if (duration<=($abstime-NAP_Pt))
                                begin
                                        ix=0.0;   //change the current state of MTJ
                                end
                                else
                                begin
                                        ix=-1.0;
                                end
                        end
                end
```

```
                end
            end   // end of antiparallel state
        end   //end of module with consideration of stochastic behaviors
        else                    //without consideration of stochastic behaviors
        begin
                if(ix==0) //Case which the magnetizations of the two layers are parallel
                begin
                        if(Vb>=IcP*Rp)      //Current higher than critical current, dynamic behavior : Sun model
                        begin
durationstatic=(`C+ln(`M_PI*`M_PI*(Em/(`Kb*temp*40*`M_PI))/4))*`e*1000*Ms*surface*tslreal*(1+P*P)/(4*`
M_PI*2*`ub*P*10000*abs(Id-IcP)); //Average time needed for switching
                        duration=durationstatic;
                                if(duration<=($abstime-P_APt))
                                begin       //Switching of the free layer always occurs
                                        ix=-1.0;   //change the current state of MTJ
                                end
                                else
                                begin
                                        ix=0.0;
                                end
                          end
                          else
                          begin    //Current smaller than critical current
                                tau=tau0*exp(Em*(1-abs(Id/IcP))/(`Kb*temp*40*`M_PI));
                                if(Vb>brown_threshold)
                                begin
                                        if (Vb<0.8*IcP*Rp)
                                        begin
                                                duration=tau;
                                                if (($abstime-NP_APt) >= duration)
                                                begin
                                                        ix=-1.0;   //change the current state of MTJ
                                                end
                                                else
                                                begin
                                                        ix=0.0;
                                                end
                                        end
                                    end
                                end
                        end
                        else          //Case which the magnetizations of the two layers are antiparallel
                                begin
                                        if(Vc>=(IcAP*Rap))
                                        begin  //Current higher than critical current, dynamic behavior : Sun model
durationstatic=(`C+ln(`M_PI*`M_PI*(Em/(`Kb*temp*40*`M_PI))/4))*`e*1000*Ms*surface*tslreal*(1+P*P)/(4*`
M_PI*2*`ub*P*10000*abs(-Id-IcAP));
                                                duration=durationstatic;  //Average time needed for switching
                                                if(duration<=($abstime-AP_Pt))
                                                begin    //Switching of the free layer always occurs
                                                        ix=0.0;   //change the current state of MTJ
                                                end
                                                else
                                                begin
                                                        ix=-1.0;
                                                end
                                    end
```

Appendix B

List of publications

International Journals:

1. **Y. Wang**, H. Cai, L. Naviner, WS. Zhao. "A non-Monte-Carlo Methodology for Variability Analysis of Magnetic Tunnel Junction Based Circuits", accepted by IEEE Trans. on Magnetics 53(3), 2017.
2. H. Cai, **Y. Wang**, L. Naviner, WS. Zhao. "Robust Ultra-Low Power Non-Volatile Logic-in-Memory Circuits in FD-SOI Technology", IEEE Trans. on Circuits and Systems, PP (99), 2016.
3. **Y. Wang**, H. Cai, L. Naviner, Y. Zhang, XX. Zhao, E. Deng, JO. Klein, WS. Zhao. "Compact model of dielectric breakdown in spin transfer torque magnetic tunnel junction", IEEE Trans. on Electron Devices 63 (4), 1762-1767, 2016.
4. **Y. Wang**, H. Cai, L. Naviner, Y. Zhang, XX. Zhao, M. Slimani, JO. Klein, WS. Zhao. "A process-variation-resilient methodology of circuit design by using asymmetrical forward body bias in 28nm FDSOI", Microelectronics Reliability 64, 26-30, 2016.
5. H. Cai, K. Liu, L. Naviner, **Y. Wang**, M. Slimani, JF. Naviner, "Efficient reliability evaluation for combinational circuits", Microelectronics Reliability 64, 19-25, 2016.
6. M. Slimani, P. Butzen, L. Naviner, **Y. Wang** and H. Cai, "Reliability analysis of hybrid spin transfer torque magnetic tunnel junction/CMOS majority voters", Microelectronics Reliability 64, 48-53, 2016.
7. H. Cai, **Y. Wang**, WS. Zhao, L. Naviner. "Breakdown Analysis of Magnetic Flip-flop With 28nm UTBB FDSOI Technology", IEEE Trans. Device and Materials Reliability, 2016: 52(8): 3401807, 2016.
8. H. Cai, **Y. Wang**, WS. Zhao, L. Naviner. "Low Power Magnetic Flip-flop Optimization with FDSOI Technology Boost", IEEE Trans. on Magnetics, 2016: 8: 52.
9. WS. Zhao, XX. Zhao, MX. Wang, SZ. Peng, BY. Zhang, K. Cao, L. Wang, Q. Shi, W. Kang, Y. Zhang, **Y. Wang**, JO. Klein, L. Naviner and D. Ravelosona. "Failure Analysis in Magnetic Tunnel Junction Nanopillar with Interfacial Perpendicular Magnetic Anisotropy", Materials, 2016: 9(1): 41.
10. **Y. Wang**, H. Cai, L. Naviner, Y. Zhang, JO. Klein, WS. Zhao. "Compact thermal modeling of spin transfer torque magnetic tunnel junction", Microelectronics Reliability 55 (9), 1649-1653, 2015.
11. P. Butzen, M. Slimani, **Y. Wang**, H. Cai, and L. Naviner. "Reliable majority voter based on spin transfer torque magnetic tunnel junction device", IEEE Electronics Letters 52 (1), 47-49, 2015.
12. H. Cai, **Y. Wang**, WS. Zhao, L. Naviner. "Multiplexing Sense Amplifier Based Magnetic Flip Flop in 28nm FDSOI Technology", IEEE Transactions on Nanotechnology 14 (4), 761-767, 2015.

13. H. Cai, **Y. Wang**, WS. Zhao, L. Naviner. “Ultra-wide voltage range consideration of reliability-aware STT magnetic flip-flop in 28nm FDSOI technology”, Microelectronics Reliability 55 (9), 1323-1327, 2015.
14. H. Cai, **Y. Wang**, K. Liu, L. Naviner, H. Petit, JF. Naviner, “Cross-layer investigation of continuous-time sigma–delta modulator under aging effects”, Microelectronics Reliability 55 (3), 645-653, 2015.
15. **Y. Wang**, Y. Zhang, E. Deng, JO. Klein, L. Naviner, W. Zhao, “Compact model of magnetic tunnel junction with stochastic spin transfer torque switching for reliability analyses”, Microelectronics Reliability 54 (9), 1774-1778, 2014.

Conferences papers:

1. **Y. Wang**, H. Cai, L. Naviner, JL. Yang, JO. Klein, WS. Zhao. “A novel circuit design of true random number generator using magnetic tunnel junction”, 2016 IEEE/ACM NANOARCH, Beijing, China, 18-20 July 2016.
2. **Y. Wang**, H. Cai, L. Naviner, Y. Zhang, XX. Zhao, M. Slimani, JO. Klein, WS. Zhao. “A process-variation-resilient methodology of circuit design by using asymmetrical forward body bias in 28nm FDSOI”, ESREF, Halle, Germany, 2016 (Best paper nomination).
3. H. Cai, **Y. Wang**, L. Naviner, WS. Zhao. “Approximate computing in MOS/Spintronic Non-Volatile Full-Adder”, 2016 IEEE/ACM NANOARCH, Beijing, China, 18-20 July 2016.
4. **Y. Wang**, H. Cai, L. Naviner, Y. Zhang, JO. Klein, WS. Zhao. “Compact thermal modeling of spin transfer torque magnetic tunnel junction”, ESREF2015, Toulouse, France.
5. E. Deng, **Y. Wang**, ZH. Wang, JO. Klein, B. Dieny, G. Prenat, WS. Zhao, “Robust Magnetic Full-Adder with Voltage Sensing 2T2MTJ Cell”, 2015 IEEE/ACM NANOARCH, Boston, USA, 2015.
6. L. Naviner, H. Cai, **Y. Wang**, WS. Zhao, “Stochastic Computation With Spin Torque Transfer Magnetic Tunnel Junction”, IEEE NEWCAS, Grenoble, France, 07-10 June 2015.
7. **Y. Wang**, H. Cai, L. Naviner, WS. Zhao, “Reliability analyse based on a compact model of spin transfer torque magnetic tunnel junction”, JNRDM2015, Bordeaux, France, 2015.
8. **Y. Wang**, Y. Zhang, E. Deng, JO. Klein, L. Naviner, W. Zhao, “Compact model of magnetic tunnel junction with stochastic spin transfer torque switching for reliability analyses”, ESREF2014, Berlin, Germany (Best poster award).

Résumé Français

I Introduction

Grâce à l'évolution rapide de la technologie des semi-conducteurs, conformément à la Loi de Moore, le nombre de transistors dans un circuit intégré (CI) a doublé tous les deux ans environ depuis plusieurs décennies. En effet, la prévision de Moore a été le moteur dans l'industrie des semi-conducteurs pour guider la planification à long terme et pour fixer des objectifs pour la recherche et le développement.

La technologie CMOS est la plus utilisée actuellement et très emblématique de cette évolution. L'ère numérique s'est développée en même temps que la réduction des dimensions (réduction du noeud) des CI basés sur cette technologie. Cependant, plusieurs contraintes et notamment les exigences en consommation d'énergie sont devenues trop importantes pour poursuivre la réduction d'échelle [2]. Il a été prédit par l'ITRS (*International Technology for Semiconductors*) que la consommation statique des mémoires en 2026 serait le triple de celles en 2016 [3]. Cette tendance est due à la contribution croissante du courant de fuite à la consommation totale pour les nœuds de technologie CMOS en dessous de 90 nm [4]. Ainsi, la consommation statique est considérée comme l'obstacle critique pour la poursuite de la réduction du nœud de technologie CMOS.

Les dispositifs spintroniques émergents qui combinent les deux attributs de l'électron (charge et spin) sont considérés comme une solution prometteuse en raison de la non-volatilité et du fonctionnement à vitesse élevée. Par rapport aux mémoires classiques composées de transistors CMOS, les mémoires à base de spintronique peuvent conserver les informations mémorisées sans alimentation. En outre, grâce à une intégration 3D facile, ces dispositifs peuvent être déposés au dessus des unités arithmétiques, évitant ainsi les échanges des données avec l'architecture classique de Von-Neumann. Cela réduit la latence de l'opération et améliore l'efficacité énergétique.

En tant que l'un des dispositifs les plus représentatifs de spintronique, la jonction tunnel magnétique (MTJ) est un candidat prometteur pour la prochaine génération de mémoires non volatiles. MTJ se compose de deux couches ferromagnétiques séparées par une couche non magnétique dans lesquelles a lieu l'effet de magnétorésistance tunnel (*Tunnel MagnetoResistance, TMR*), phénomène démontré en 1975 [7]. La résistance de MTJ dépend de l'orientation relative de l'aimantation des deux couches ferromagnétiques (R_p à l'état parallèle et R_{ap} à l'état antiparallèle). Cette résistance peut être intégrée dans les mémoires et les circuits logiques pour représenter la logique '0' ou '1', de manière comparable à celle de transistors CMOS.

Parmi toutes les approches de commutation entre l'état parallèle et l'état antiparallèle, le couple de transfert de spin (*spin transfer torque, STT*) simplifie le processus de la commutation et réduit l'énergie dissipée pendant l'écriture. Cette méthode utilise un courant relativement faible ($\sim$ 100uA) parcourant le MTJ pour changer son état. Sans la nécessité de champ magnétique, STT permet d'atteindre haute densité de mémoire magnétique (MRAM) et faible puissance. Le MTJ avec l'anisotropie magnétique perpendiculaire (*Perpendicular Magnetic Anisotropy, PMA*) combinant faible courant de commutation (49uA) et stabilité thermique élevée a été découvert [20]. La figure B.1 illustre les évènements les plus significatifs de la recherche et du développement en spintronique.

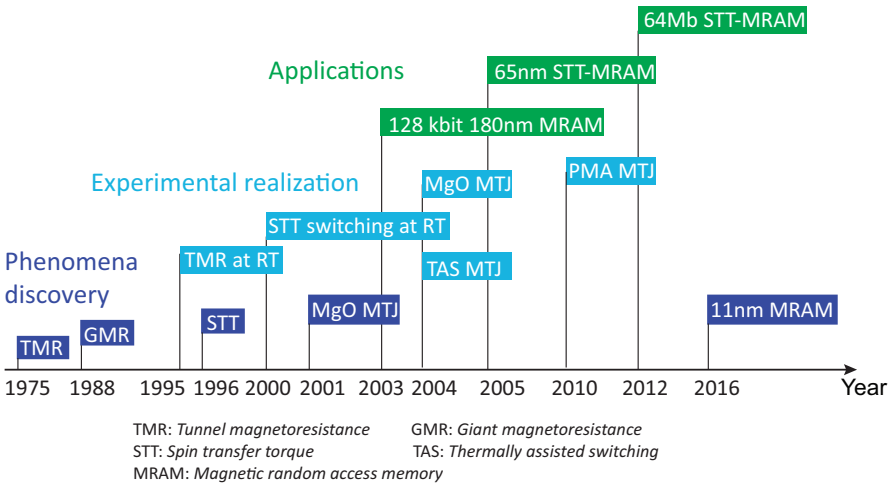


Figure B.1: Évènements importants de la recherche et du développement en spintronique.

Malgré les potentiels exceptionnels de STT-MRAM, sa large commercialisation demeure très difficile à cause de sa faible fiabilité. En effet, il a été démontré que la méthode de commutation STT est intrinsèquement stochastique [21] et une densité de courant relativement élevée est nécessaire pour commuter avec succès dans le processus d'écriture. Par ailleurs, du fait des couches ultra-minces ($\sim$1nm) et de sa petite surface, le composant MTJ

opère dans des conditions extrêmes (ex : le champ électrique intense à travers la barrière d'oxyde et forte densité de courant qui le traverse). Sa performance peut être dégradée de manière significative pour différentes raisons, telles que l'effet d'auto-échauffement, les variations de processus et les mécanismes de vieillissement. Les problèmes de fiabilité peuvent intervenir de la conception initiale et la fabrication jusqu'à l'usure finale. Tous les problèmes de fiabilité auront un impact significatif sur la qualité et le rendement des circuits basés sur MTJ.

Les travaux de recherche sur la fiabilité se concentrent principalement sur la modélisation des défauts, l'analyse de la fiabilité, la méthodologie de la prise en compte de fiabilité et la prévision de défaillance [22]. La modélisation caractérise les défauts physiques et cartographie la dégradation aux paramètres au niveau du dispositif (par exemple, modèle BSIM4), qui est le travail de base de ces trois derniers. Avec l'évolution rapide de STT-MRAM, la fiabilité a attiré l'attention des chercheurs [21, 23, 24]. Les problèmes de fiabilité des dispositifs MTJ ont toujours été bien caractérisés théoriquement et expérimentalement. Cependant, il n'existe pas encore un modèle compact combinant les différents éléments connus pouvant influer sur la fiabilité pour les concepteurs de circuits. Compte tenu des coûts importants de la fabrication des MTJ, l'existence d'un tel modèle serait très efficace surtout dans les premières phases de la conception.

Cette thèse vise à fournir une compréhension approfondie des sources des problèmes de fiabilité dans les dispositifs MTJ et à proposer un modèle compact précis pour les concepteurs de circuits basés sur ce type de dispositif. Ce modèle a pour objectifs de prédire les défaillances fonctionnelles éventuelles des circuits basés sur MTJ et d'aider à identifier les solutions lors de la phase de conception. Le modèle proposé a été utilisé pour des études de fiabilité et l'exploration de certaines stratégies de conception en vue de la tolérance aux fautes et l'amélioration de performance des circuits. Enfin, des architectures alternatives permettant des fonctions classiques ont été mises en œuvre pour tirer profit des problèmes de la fiabilité des MTJ .

II État de l'art

Ce chapitre présente les travaux préliminaires relatifs à la fiabilité des dispositifs MTJ. Il commence par la présentation détaillée des MTJ. Ensuite, les applications principales basées sur MTJ sont discutées et comparées en termes de performance et de fiabilité. Enfin, l'état actuel de la recherche sur les problèmes principaux de fiabilité de MTJ est examiné et le travail requis est synthétisé.

II.a Principes d'opération MTJ

Le développement de MTJ provient de la découverte de l'effet magnétorésistance tunnel (TMR) par Jullière en 1975 [7], dans lequel deux couches ferromagnétiques (FM) sont séparées par une barrière isolante. Le phénomène peut être microscopiquement expliqué par la Figure B.2 [13]. Dans les matériaux FM, les populations de spin-up et de spin-down sont différentes au niveau de l'énergie de Fermi, conduisant à une densité inégale des états disponibles pour chacun [26]. En conséquence, le matériau FM est magnétisé par le moment magnétique net produit par le déséquilibre. Les électrons proches du niveau de Fermi servent comme porteurs pendant le transport. Les électrons à spin polarisé passent à travers la barrière d'oxyde par effet tunnel avec la conservation de l'état de spin. Un électron avec l'état de spin-up d'une couche FM peut voyager à travers l'isolant seulement s'il peut trouver un état de spin-up au niveau de Fermi de l'autre couche FM. Si les directions d'aimantation des deux couches FM sont parallèles (P), tous les électrons de spin-up et de spin-down peuvent facilement trouver un état correspondant après avoir voyagé à travers la barrière parce que les structures de bandes des deux couches FM sont presque les mêmes. Inversement, si elles sont antiparallèles (AP), seule une partie des électrons peut agir en tant que support pour le courant d'effet tunnel, ce qui entraîne une conductance inférieure à l'état AP. Ainsi, la résistance de l'empilement des trois couches est différente en fonction de l'état d'aimantation des couches FM.

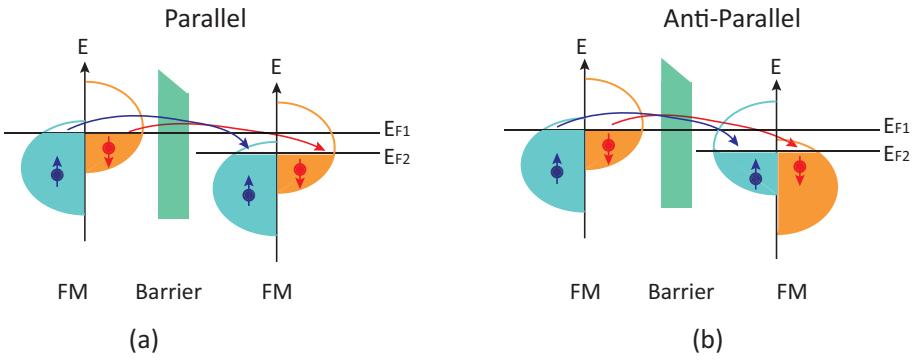


Figure B.2: Effet tunnel dépendant du spin des électrons dans un MTJ, tandis que les directions d'aimantation dans les deux couches FM sont (a) parallèles et (b) en antiparallèle.

La Figure B.3 montre une structure typique de la MTJ qui se compose essentiellement de trois couches: Un isolant mince (barrière d'oxyde tel que AlxOy et MgO) séparé par deux couches ferromagnétiques (par exemple, CoFe) Les deux couches FM ont différentes configurations: celle avec une direction d'aimantation de rotation fixe est appelée couche

de référence, tandis que l'autre peut être changée dans deux directions (la couche de stockage). Ainsi, les termes parallèle (P) et antiparallèle (AP) sont généralement utilisés pour décrire les deux configurations différentes de MTJ. La configuration du MTJ peut être réglée en changeant l'orientation d'aimantation dans la couche de stockage.

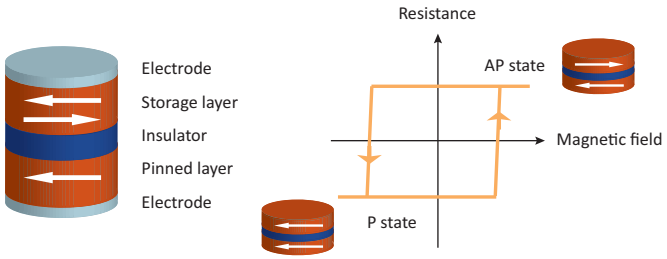


Figure B.3: Structure standard de MTJ.

Avec barrière d'oxyde entre deux couches ferromagnétiques, le MTJ présente une valeur de résistance qui est comparable à la technologie des transistors CMOS. TMR (Tunnel magnetoresistance ratio) est l'un des paramètres les plus importants pour déterminer la performance du dispositif MTJ. Il est défini comme suit:

$$TMR = \frac{\Delta R}{R_P} = \frac{R_{AP} - R_P}{R_P} \tag{B.1}$$

où R_P et R_{AP} sont les résistances de l'état P et AP de MTJ.

Pour une meilleure immunité contre les variations et inadéquation générées dans le processus de fabrication, il est toujours préférable d'avoir une la valeur élevée de TMR. Cela a constitué la motivation de recherches intenses et du développement rapide de MTJ.

Du point de vue de la propriété électrique, la méthode de commutation STT (spin transfer torque) ne nécessite qu'un courant bidirectionnel I supérieur au courant de seuil pour changer l'état de MTJ. STT promet une bonne efficacité énergétique. Il a été observé que le courant polarisé en spin injecté perpendiculairement au plan peut influer l'aimantation de couches FM. Le transfert du moment angulaire de spin à partir d'un courant spin-polarisé à une aimantation locale de la couche FM peut générer un couple avec l'aimantation de la présente couche FM. Comparée aux autres méthodes, la commutation par ce couple simplifie les manipulations magnétiques des couches FM. Si la densité de courant est supérieure à la valeur de seuil, le couple appliqué par le courant change l'aimantation de la couche libre (FL) de MTJ [30].

Cette approche simplifie considérablement le processus de commutation. En outre, l'intensité du courant requis par STT est significativement réduite par rapport aux méthodes de commutation précédentes (normalement une différence d'un ordre). En conséquence,

une densité plus élevée et une vitesse plus élevée peuvent être obtenues dans les MRAM basées sur STT-MTJ. Depuis sa démonstration pratique, STT est considérée comme l'approche la plus prometteuse pour les futures applications de MRAM.

II.b Circuits de mémoire et de logique basés sur jonction tunnel magnétique

Du fait des caractéristiques susmentionnées de MTJ, beaucoup d'efforts de recherche ont été consacrés à son application dans la conception de mémoires et de fonctions logiques spécifiques. Cette section fera une brève description de certains modèles typiques de circuits MTJ.

L'architecture *cross-point* a d'abord été proposée pour réaliser MRAM [13, 38, 39]. Comme montré dans la Figure B.4 (a), chaque MTJ est relié aux points de croisement de deux rangées perpendiculaires de lignes et de colonnes parallèles conductrices. Pour programmer la cellule de mémoire avec succès, les impulsions de courant sont envoyées par une ligne de chaque réseau et le MTJ au point de croisement de ces deux lignes orthogonales peut être commuté avec un champ magnétique ou une densité de courant suffisante. Pour la lecture, la résistance du dispositif entre les deux lignes de croisement sélectionnées peut être détectée, ce qui représente les informations stockées dans le MTJ. Ce dispositif promet une intégration à haute densité mais souffre de la problématique du chemin de fuite et de vitesse d'accès faible, ce qui limite sa large application pour une lecture rapide et fiable [40].

Une autre structure plus complexe appelée 1T1R a été proposée pour éliminer les courants indésirables [41]. Comme montré dans Figure B.4 (b), le transistor ajouté contribue à isoler la cellule sélectionnée des autres cellules, en supprimant le problème de chemin d'accès. Cette architecture permet une vitesse d'accès élevée et une meilleure fiabilité pour les opérations d'écriture et de lecture par rapport à l'architecture *cross-point*. Cependant, elle conduit à une moindre densité à cause du transistor ajouté pour chaque cellule.

Le concept de logique en mémoire (LIM) a été proposé au début des années 1960 [43] pour réduire la consommation d'énergie et le délai d'interconnexion des unités de calcul. Dans l'architecture Von-Neumann classique, la mémoire et les circuits logiques sont spatialement séparés, conduisant à un transfert important de données entre eux. Contrairement à cela, les cellules de mémoire sont déposées sur les circuits logiques dans l'architecture LIM. La distance entre la mémoire et les circuits logiques est considérablement raccourcie, entraînant une vitesse de transfert plus élevée et une consom-

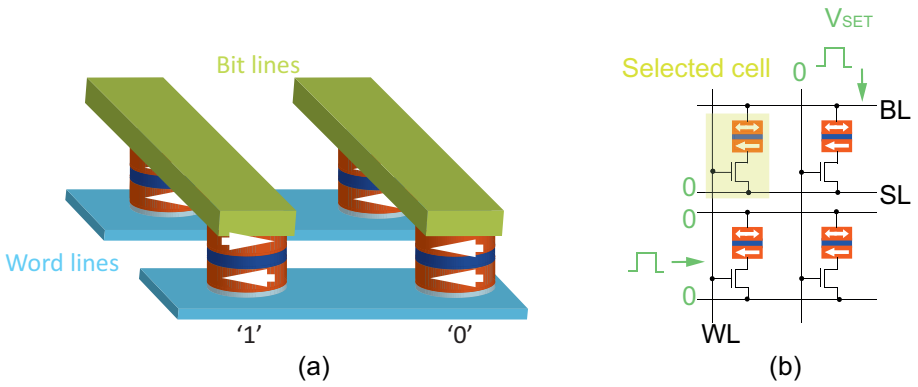


Figure B.4: Schématique de (a) cross-point array et (b) 1T/1MTJ mémoire cellule architecture

mation d'énergie plus faible dans les interconnexions. Les dispositifs MTJ commutés par le mécanisme STT apparaissent comme une solution prometteuse pour LIM [45]. Les données ayant déjà été mémorisées dans des dispositifs MTJ dans les circuits LIM proposés, la tension d'alimentation peut être coupée sans besoin de transmission de données vers des dispositifs de stockage non-volatils externes lorsque le circuit passe en mode veille. De plus, le temps de rétention de données intrinsèque aux MTJ permet un calcul instantané on/off, à savoir que le système peut immédiatement continuer à fonctionner après sa sortie du mode "endormi". En raison de ces propriétés, la dissipation de puissance peut être considérablement réduite.

Dans l'architecture générale de la logique en mémoire basée sur STT-MRAM [46], le STT-MRAM est déposé sur le plus haut niveau de métal sur les transistors CMOS. Il se compose de trois parties : un circuit d'amplification de détection de pré-charge (PCSA) évalue le résultat logique sur les sorties, un bloc logique d'écriture programme les cellules STT-MRAM et un bloc assure le contrôle de données logiques. La Figure B.5 montre plusieurs portes logiques typiques et une puce de calcul basée sur l'architecture générale proposée dans [46, 47, 48, 49]. Il a été démontré que ces circuits étaient avantageux en termes de surface, d'efficacité énergétique et de vitesse d'opération par rapport aux implémentations CMOS classiques.

II.c Analyse de fiabilité des applications basées sur MTJ

Le taux d'erreur de tous les dispositifs et circuits basés sur les technologies nanométriques est devenu une préoccupation majeure. Ces erreurs résultent de la difficulté à fournir un contrôle dimensionnel très précis qui est nécessaire à la fabrication des dispositifs et également à l'interférence de l'environnement local. Quant à STT-MTJ, dont la taille

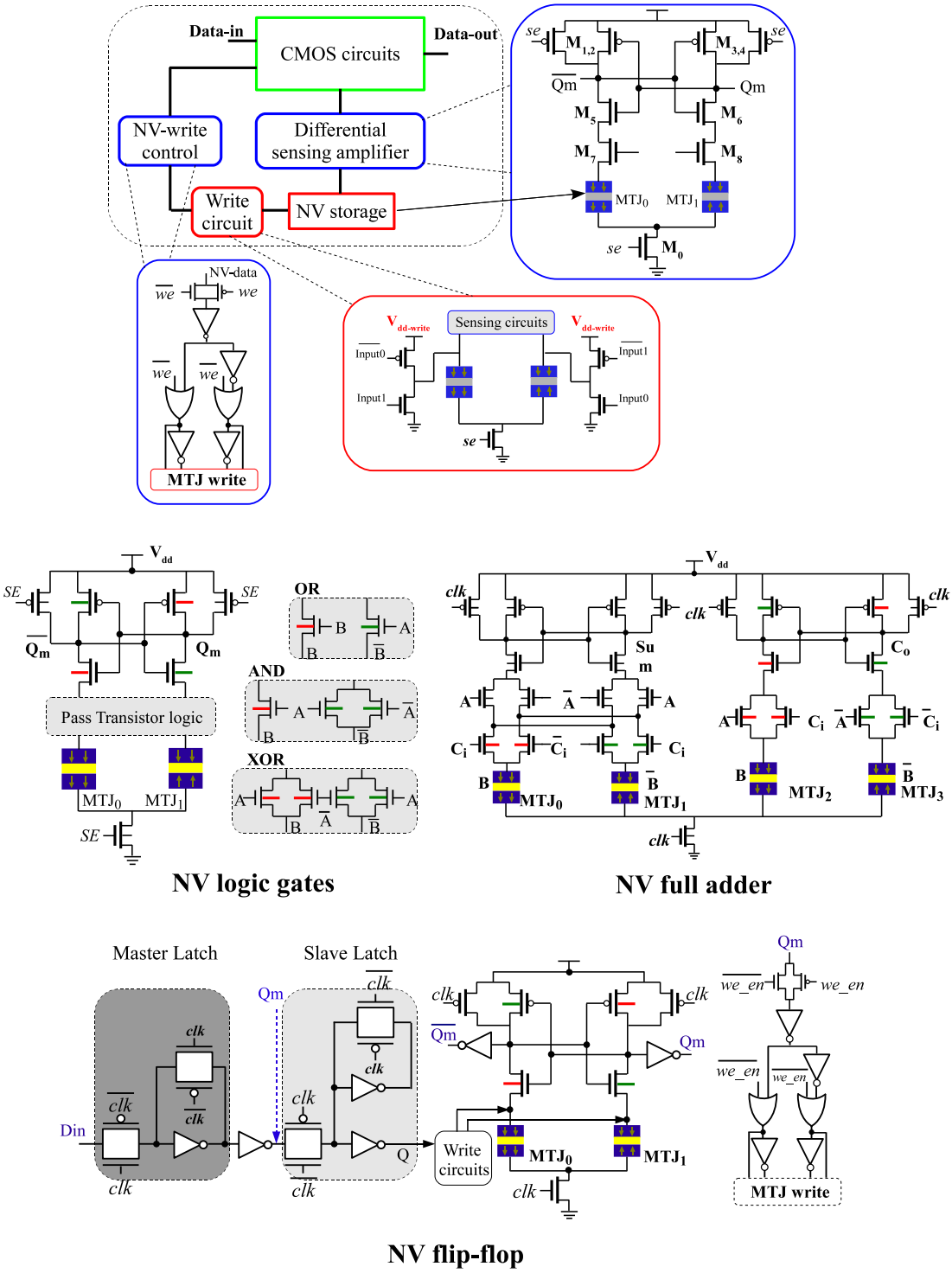


Figure B.5: Circuits typiques de MOS/MTJ NV-LIM basés sur une structure d'amplificateur de détection de pré-charge : portes logiques, additionneur complet et bascule.

est habituellement au niveau de nm avec peu de couches d'atomes, la fiabilité constitue particulièrement un obstable pour sa commercialisation large. Le problème de fiabilité du dispositif STT-MTJ comprend principalement la variation du processus, la commutation stochastique, la fluctuation de température et le claquage diélectrique.

Dans la première démonstration expérimentale de STT-MRAM [41], la résistance des MTJ suit une distribution statistique, avec un écart standard approximatif σ de 4%. Même si une solution d'optimisation utilisant des technologies de production MRAM conventionnelles a été mentionnée, ce qui peut supprimer σ moins de 1-2%, la variation du processus ne peut jamais être supprimée. En raison de la précision limitée du processus, de nombreux paramètres ne sont pas identiques à la cible initiale. En conséquence, les propriétés magnétiques et électriques sont influencées, telles que la résistance, le rapport TMR et le délais de commutation. Toutes ces variations peuvent entraîner des erreurs fonctionnelles lors des opérations de MRAM. Après cela, de nombreux chercheurs ont proposé diverses méthodes pour modéliser les variations de processus et analyser l'influence sur les circuits hybrides basés sur MTJ [56, 57, 58]. Certaines techniques expérimentales spéciales ont été appliquées dans la fabrication de MTJ pour améliorer la performance du dispositif et réduire l'effet des variations de processus [59]. En attendant, une variété de stratégies de conception sont proposées pour améliorer la robustesse des circuits à base de MTJ [60, 61, 62].

La méthode de commutation STT a été démontrée intrinsèquement stochastique [21]. La durée d'inversion du mécanisme d'écriture STT peut varier considérablement d'un événement à l'autre, avec un écart typique presque aussi grand que la durée moyenne de commutation et des distributions sinusoïdales avec des queues exponentielles [63]. La probabilité de succès de commutation est une fonction du courant circulant dans le MTJ et la durée d'impulsion. Le comportement stochastique provient des fluctuations thermiques inévitables de l'aimantation qui interfèrent de façon aléatoire pour activer ou ralentir l'inversion de l'aimantation. Beaucoup d'autres chercheurs ont théoriquement ou expérimentalement vérifié ce phénomène [64, 65, 66, 67]. On peut conclure des mesures expérimentales que l'augmentation de la valeur de courant d'écriture ou l'ajout de marges étendues sur la durée d'impulsion du pilote sont les méthodes les plus efficaces pour éviter les échecs d'écriture. Cependant, cela peut conduire à un surcroît important de puissance, vitesse et surface, ce qui est la provenance des deux problèmes de fiabilité suivants.

L'effet d'auto-échauffement des MTJ a été observé dans [23] et étudié avec l'exécution de simulations numériques à une dimension en résolvant l'équation de la chaleur. Différent

de l'approche de commutation TAS qui chauffe le MTJ par un élément externe, la MTJ peut également être chauffée par elle-même en raison de chauffage Joule. Malgré les efforts considérables consacrés à l'optimisation de la technologie au cours des dernières années, une densité de courant relativement élevée passant par MTJ est toujours requise par la plupart des mécanismes de commutation. Il en résulte un effet d'auto-échauffement considérable qui peut provoquer des erreurs fonctionnelles des circuits MTJ/CMOS hybrides [68].

De plus, les caractéristiques des matériaux ferromagnétiques sont très sensibles à la température de l'environnement, ce qui a déjà été observé dans de nombreuses expériences [7, 8, 11, 12]. Les matériaux ferromagnétiques sont très sensibles aux fluctuations thermiques. Avec des conditions thermiques différentes, les propriétés magnétiques sont totalement différentes. L'objectif commun de la recherche de MTJ se concentre toujours sur la fabrication du ratio TMR plus élevé à température ambiante. Néanmoins, avec les perspectives prometteuses des applications basées sur MTJ à l'ère de l'IOT prochaine, les caractéristiques exactes de MTJ dans différentes conditions thermiques devraient être soigneusement étudiées et modélisées pour les concepteurs de circuits.

Le claquage de diélectrique est le problème de fiabilité le plus crucial qui détermine la durée de vie du dispositif (transistor ou MTJ). Comme MTJ est un dispositif résistif et sa résistance vient principalement de la barrière d'oxyde, la tension appliquée sur MTJ est presque imposée à l'isolateur (AlxOy ou MgO). Avec une épaisseur ultra-fine ($\sim$ 1 nm), les tensions de claquage diélectrique diminuent également, et il est nécessaire d'éviter la claquage diélectrique de la barrière tunnel en fonction du temps (TDDB) due aux opérations d'écriture [71]. Plusieurs expériences ont été réalisées pour démontrer ce phénomène, d'autres ont été effectuées pour explorer le mécanisme derrière le phénomène et les facteurs qui ont un impact sur la TDDB [24, 77, 78]. On a découvert que TDDB est lié aux facteurs divers, tels que la température de recuit, la pureté du matériau d'oxyde, l'épaisseur de la barrière tunnel, la tension de contrainte, la température, la durée de polarisation, etc.

En résumé, les incertitudes de fiabilité peuvent entraîner des pénalités de performance, de coût et de délai de mise en marché. Les défaillances fonctionnelles peuvent être induites par une marge de fiabilité insuffisante qui est coûteuse à réparer et à endommager la réputation. Ainsi, il est nécessaire d'identifier et d'aborder ces problèmes de fiabilité lors de la phase de conception. L'exigence d'une technologie de processus plus précise est très importante, des conceptions prudentes et intelligentes qui tolèrent ces variations sont

également nécessaires. La plupart des modèles existants ne se concentrent que sur une partie des problèmes de fiabilité, ce qui n'est pas suffisant pour l'exigence croissante d'une analyse de fiabilité plus précise. Pour des conceptions plus précises, un modèle complet et précis incluant les principaux problèmes de fiabilité devient indispensable et urgent pour les concepteurs de circuits.

III Compact modeling of reliability issues in STT-PMA-MTJ

Ce chapitre étudie les problèmes de fiabilité liés aux dispositifs MTJ, puis les quantifie à l'aide d'équations mathématiques. Ensuite, un modèle compact est proposé aux concepteurs de circuits pour tenir compte de ces problèmes de fiabilité.

III.a Modélisation compacte des problèmes de fiabilité dans STT-PMA-MTJ

Verilog-A est un langage de programmation parfait pour créer un modèle compact, simple, efficace, précis, rapide et compatible de STT-MTJ. Comme il est très lisible, les ingénieurs de caractérisation et les concepteurs de circuits peuvent facilement le comprendre, ce qui facilite la continuité de ce travail et simplifie le développement de versions futures du modèle.

La hiérarchie des modèles physiques qui sont intégrés dans le modèle compact est illustrée dans Figure B.6. Tous les paramètres, constantes et variables sont définis au début du modèle. Les valeurs des paramètres peuvent être reconfigurées par les concepteurs de circuits.

Afin de faciliter l'utilisation du modèle, nous avons créé l'interface utilisateur graphique (GUI) en utilisant la fonction de Component Description Format (CDF) dans l'environnement de conception Cadence. Les utilisateurs peuvent reconfigurer le périphérique en saisissant les valeurs à l'interface et le système va transférer ces valeurs au simulateur (par exemple Spectre) pour les simulations. Comme la méthode de configuration est identique à celle des transistors classiques, il est très pratique pour les concepteurs de circuits de concevoir des circuits MTJ/CMOS hybrides plus complexes. Dans l'outil de simulation Cadence, un symbole peut être créé en représentant le modèle programmé en langage VerilogA. Ce symbole sera visible pour les utilisateurs et facilitera les paramètres de simulation.

III.b Validation fonctionnelle du modèle

Cette section présente quelques résultats de simulation pour valider les fonctionnalités du modèle compact, y compris les problèmes de fiabilité.

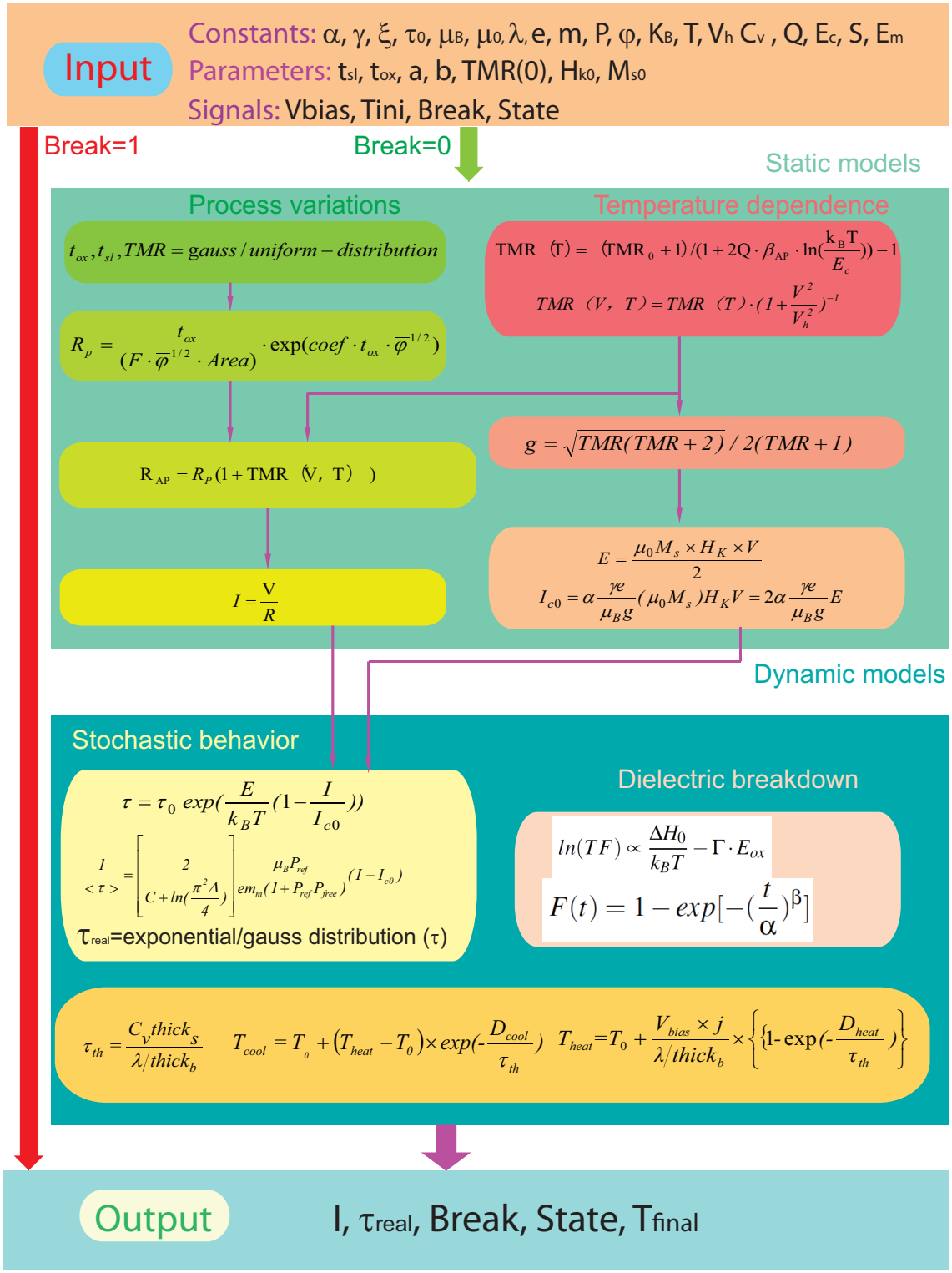


Figure B.6: Architecture du modèle compact de PMA STT MTJ intégrant des modèles physiques de problèmes de fiabilité.

La fonction des variations de processus peut être démontrée par deux types de simulations: DC et transitoire. Comme montré dans la Figure B.7, la première peut refléter les différentes valeurs de résistance causées par les variations de t_{ox} et TMR. Comme la valeur de TMR n'a pratiquement aucun impact sur la résistance de l'état parallèle (P), la résistance d'état anti-parallèle (AP) a une distribution d'échelle plus grande que celle de l'état P. Dans ce dernier, le courant de MTJ suit une distribution gaussienne avec la même tension de polarisation. Notez que les paramètres (t_{sl}, t_{ox}, TMR) suivent une distribution gaussienne avec un écart de 1%.

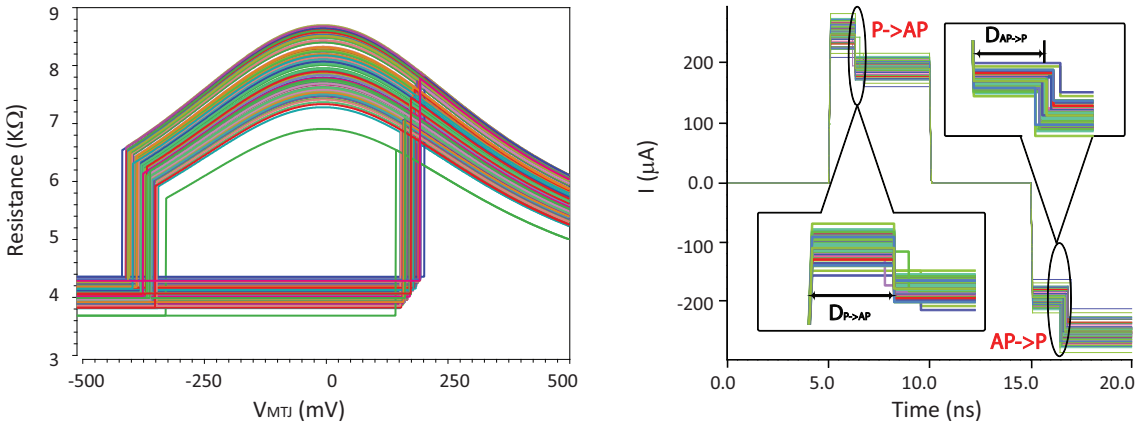
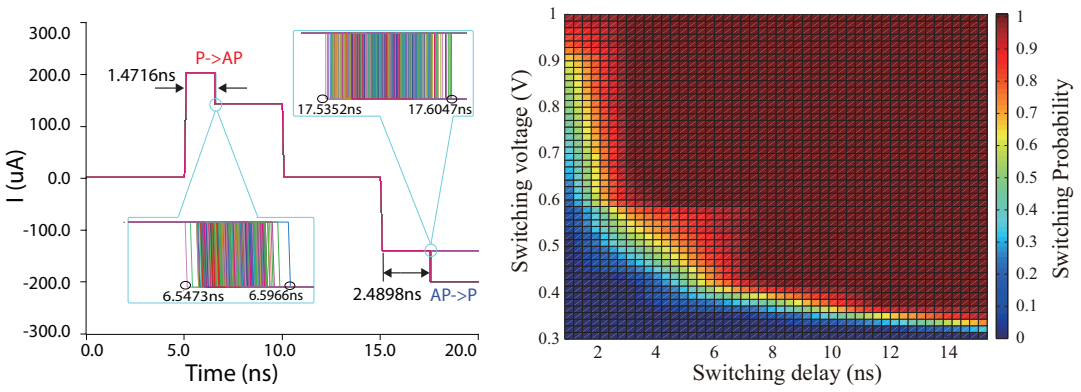


Figure B.7: Les simulations MC de (a) la résistance dépendante de la tension de polarisation et (b) 1000 processus d'écriture complet avec des variations de processus.

En utilisant le circuit d'écriture représenté sur la Figure B.5, on effectue des simulations Monte-Carlo de 1000 processus d'écriture dans lesquelles la durée de commutation suit une distribution normale avec la valeur moyenne de $\tau_{p->ap}$ ou $\tau_{ap->p}$ et variation de 0,02. Les résultats dans la Figure B.8 (a) démontrent que la durée moyenne de commutation (sans comportement stochastique) est $\tau_{p->ap}$=1,4716 ns et $\tau_{ap->p}$=2,4898 ns. Comme prévu, toutes les valeurs de la durée de commutation pour l'état parallèle (P) à l'état antiparallèle (AP) sont dans l'intervalle [0,98 $\tau_{p->ap}$, 1,02 $\tau_{p->ap}$]. La Figure B.8 (b) illustre la probabilité de commutation en fonction de la tension de commutation appliqué et du délai de commutation. La zone rouge foncé est considérée comme la zone d'écriture fiable tandis que la zone bleu foncé est la zone de lecture fiable.

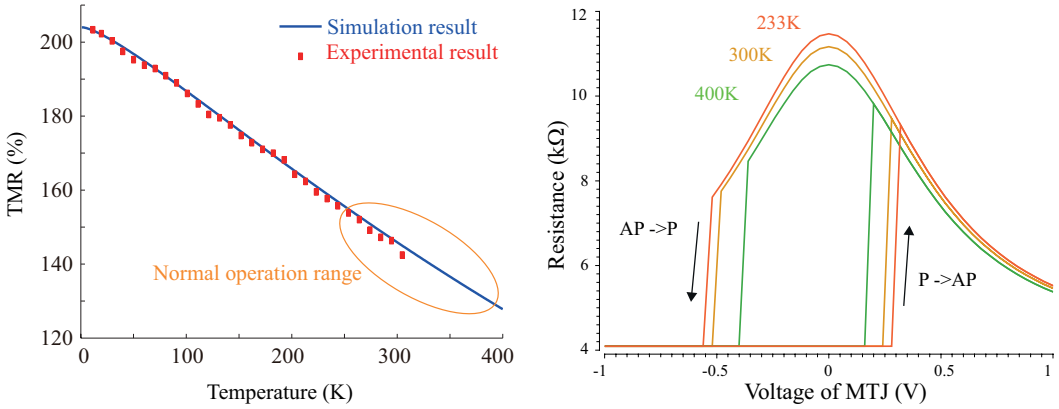
La Figure B.9 (a) affiche la dépendance en température du TMR, ce qui est cohérent avec les résultats expérimentaux [99]. La dépendance de la température du ratio TMR et du courant critique de commutation peut être observée dans la Figure B.9 (b). Le mécanisme derrière ce phénomène est le suivant: à mesure que la température augmente,



(a) (b)

Figure B.8: (a) Simulations de MC de 1000 processus d'écriture complète avec les comportements stochastiques. (b) Probabilité de commutation en fonction de la tension de commutation et du temps de commutation.

l'énergie de barrière diminue et les spins magnétiques ont une plus grande énergie thermique, ce qui aide les spins à traverser plus facilement la barrière.



(a) (b)

Figure B.9: (a) Evolution du TMR avec augmentation de la température et données expérimentales (points rouges) dans [99]. (b) Résistance du dispositif MTJ par rapport à la tension de polarisation à différentes températures. Le courant critique est réduit en augmentant la température.

La Figure B.10 montre la durée de vie de MTJ pour différentes barrières d'oxyde. En tenant compte de l'auto-échauffement, la durée de vie MTJ avec une barrière de 1 nm d'épaisseur peut être estimée à 10 ans pour une tension de fonctionnement typique de 420 mV. Ce résultat rencontre un excellent accord avec la valeur mentionnée dans [110].

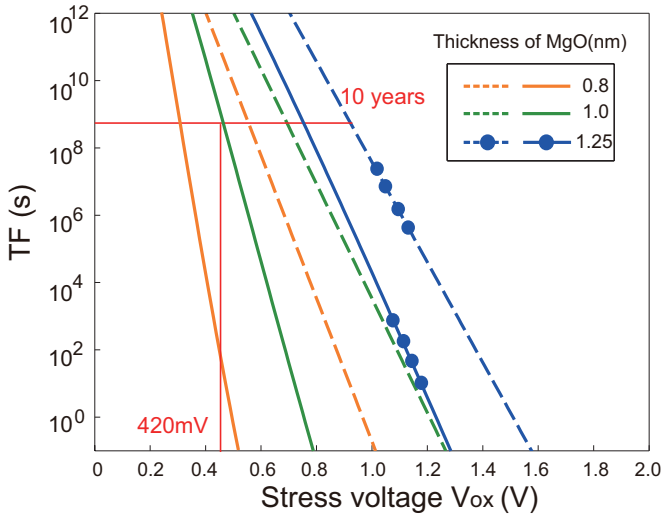


Figure B.10: Durée de vie de MTJ sans (lignes pointillées) et avec (lignes) prise en considération de l'auto-échauffement. Les points sont des données expérimentales dans [74].

IV Analyse de la fiabilité et conception adaptée à la variabilité des circuits hybrides MTJ/CMOS

Ce chapitre se concentre sur l'analyse de la fiabilité et l'exploration de la méthodologie d'optimisation de la fiabilité à l'aide du modèle compact développé au chapitre précédent.

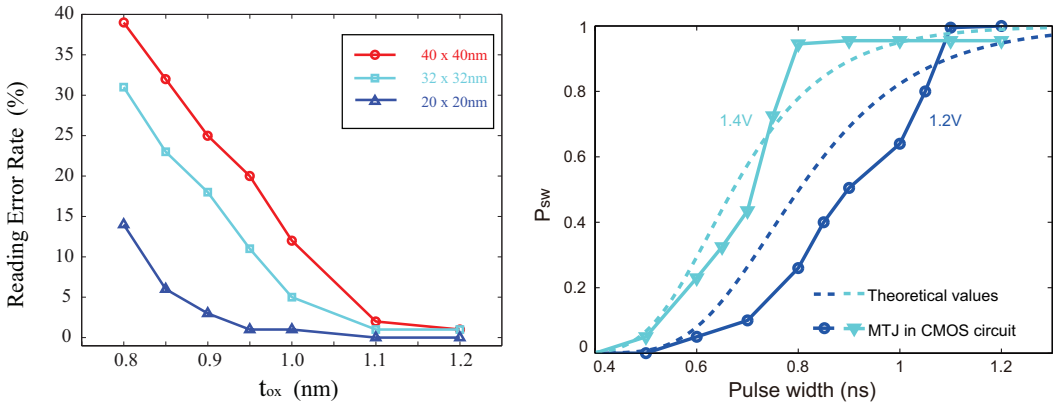
IV.a Analyse de fiabilité des circuits basés sur MTJ

Le circuit de l'amplificateur de détection de pré-charge (PCSA) illustré dans Figure B.5 est très important dans la mémoire et les circuits logiques basés sur MTJ. Dans PCSA, la méthode de détection dynamique permet l'amplification des données analogiques au numérique avec une puissance ultra-faible. De plus, la perturbation de lecture induite par les opérations de détection peut être considérablement diminuée [61]. Ce dernier est très important pour le STT-MRAM intégré car il s'agit d'une contrainte intrinsèque limitant la fiabilité du circuit logique où le circuit de correction d'erreur complexe (ECC) est nécessaire pour assurer une vitesse de calcul élevée (par exemple 1 GHz). Ainsi, la structure PCSA est largement utilisée dans les portes logiques, les cellules unitaires arithmétiques et les cellules de mémoire [46, 47, 49, 120].

Premièrement, une analyse de variabilité basée sur le circuit PCSA est réalisée pour étudier la dépendance du taux d'erreur de lecture sur la surface de MTJ et l'épaisseur de barrière d'oxyde t_{ox}. Figure B.11 (a) montre que la réduction de la taille du transistor et l'agrandissement de l'épaisseur de la barrière aux oxydes peuvent améliorer suffisamment

la fiabilité de circuit PCSA [98].

Afin de valider le comportement de commutation stochastique dans le modèle compact, des simulations Monte Carlo ont été réalisées pour un circuit d'écriture 2T-1M. La Figure B.11 (b) démontre que la probabilité de commutation augmente avec la croissance de la tension de contrainte et de la largeur d'impulsion. La tension effective à travers MTJ fluctue en raison des variations de processus des transistors et MTJ (variation de résistance). Pendant ce temps, le délai de commutation moyen τ_{sw} flotte en raison des variations de processus de MTJ. Avec une tension de 1,4V, 40 de 1000 échantillons ne sont pas commutés avec succès en raison d'une panne rapide, conduisant ainsi à la probabilité finale de commutation de 96%. Cette simulation peut être utilisée pour trouver un compromis entre la fréquence de fonctionnement et la consommation d'énergie d'une MRAM basée sur MTJ.

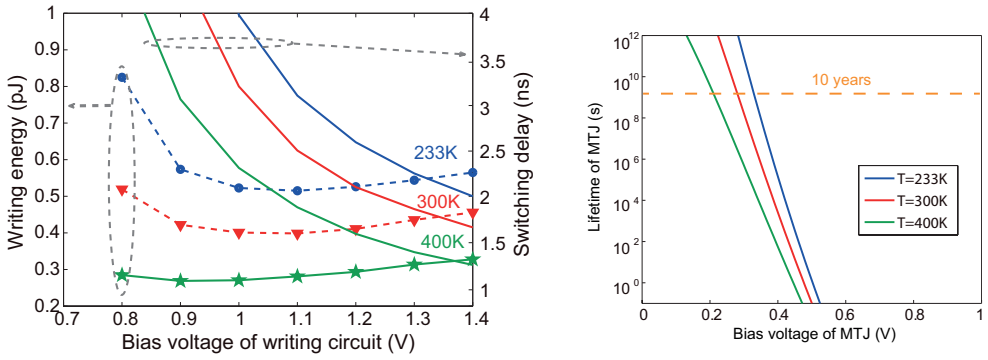


(a) (b)

Figure B.11: (a) Taux d'erreur de lecture versus l'épaisseur de MgO et la surface de MTJ. (b) Probabilité de commutation avec différentes tensions d'écriture.

Comme cela est présenté dans Figure B.12, une température élevée entraîne une vitesse plus grande et un processus d'écriture de puissance plus faible, mais accélère la rupture diélectrique de MTJ et conduit à un temps plus court pour la panne. Ainsi, il existe un compromis de conception entre la performance d'écriture (consommation et fréquence) et l'endurance en considération de l'état thermique.

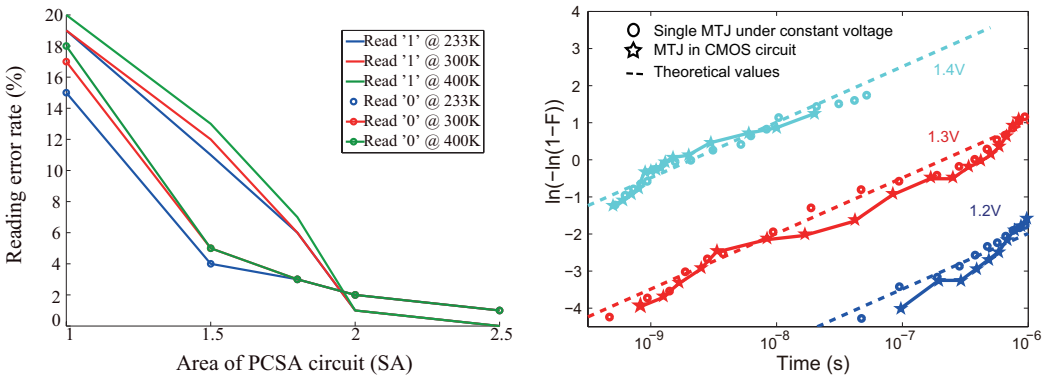
Pour étudier l'impact de la température sur le circuit de lecture, une analyse de fiabilité du PCSA a également été réalisée. Le résultat dans Figure B.13 (a) montre que le taux d'erreur de lecture augmente lentement avec la température en raison de la dépendance à la température du TMR et du CMOS. En raison du courant relativement faible traversant les dispositifs MTJ et CMOS, la température a un impact très faible sur le comportement de lecture.



(a) (b)

Figure B.12: (a) Probabilité de commutation en fonction de la tension de commutation et du temps de commutation appliqués. (b) Tension de commutation en fonction du temps moyen de commutation à des températures différentes.

La Figure B.13 (b) présente la distribution de probabilité de claquage pour le cas théorique, les résultats de simulation de MTJ sous tension constante et MTJ intégré dans le circuit de CMOS. Les cercles montrent un bon accord avec un écart plus faible en raison de l'impact de la variation de la barrière aux oxydes sur le comportement de claquage de MTJ, tandis les étoiles ont une déviation plus importante car les variations de MTJ et CMOS sont prises en compte, dans lesquelles la tension effective à travers MTJ fluctue autour des valeurs indiquées (1,4V, 1,3V et 1,2V).



(a) (b)

Figure B.13: (a) Taux d'erreur de lecture du PCSA avec la surface différente du circuit (SA est la taille minimum du circuit de PCSA) dans différentes conditions thermiques. b) Distribution cumulative de probabilité de claquage.

IV.b Application of non Monte-Carlo Methodology in hybrid MOS/MTJ Circuits

En se basant sur le modèle des pires cas, nous avons proposé une méthode non Monte-Carlo pour l'analyse de la variabilité des circuits basés sur STT-MTJ. Cette méthode proposée est intégrée dans la cellule STT-MRAM pour valider sa fonctionnalité. La puissance dynamique et la latence du circuit sont évaluées à la fois pour les opérations d'écriture et de détection en utilisant les modèles de pires cas et les modèles statistiques de MTJ et de transistors.

La Figure B.14 montre les performances en termes de puissance et de délai pour les deux phases de fonctionnement. La plupart des résultats statistiques sont répartis dans la zone fixée par les pires angles, c'est-à-dire que l'analyse du pire cas peut être utilisée pour estimer complètement la performance du circuit. En outre, l'analyse du pire cas prend 28.2s tandis que l'analyse statistique prend 5195s. Les valeurs externes peuvent également être couvertes par l'attribution d'une valeur plus élevée pour n. Les vecteurs de conception (tension de polarisation, paramètres des dispositifs et bloc de correction) peuvent être modulés pour obtenir un compromis optimisé parmi les termes de performance. Par exemple, si l'opération d'écriture de la vitesse la plus défavorable échoue, le courant doit être renforcé en augmentant la tension de polarisation ou modifiant les tailles des dispositifs.

IV.c Conception adaptable à la variabilité avec polarisation du substrat asymétrique dynamique des transistors FDSOI

L'architecture symétrique est largement utilisée dans la conception du flip-flop non volatile (NVFF) en raison de son immunité élevée à la perturbation de lecture [49, 135, 136, 137]. Cependant, la performance du circuit est fortement influencée par la variation du processus. Sur la base de ces structures symétriques, nous avons proposé une méthode de polarisation du substrat asymétrique dynamique (DABB) pour optimiser l'immunité de variabilité de NVFF [138]. Cette méthode est mise en œuvre avec un kit de conception UTBB-FDSOI (Ultra Thin Body and Box Fully Depleted Silicon On Insulator), les deux modèles compacts et les modèles les plus pessimistes des transistors MTJ et FDSOI sont considérés dans la simulation.

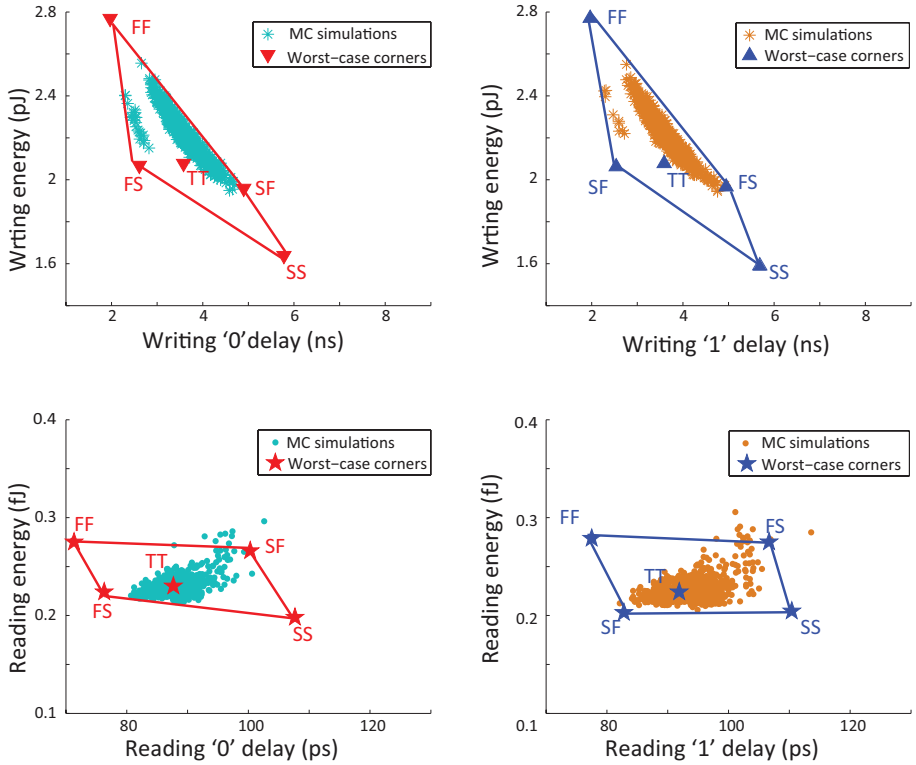


Figure B.14: Performance d'écriture et de lecture de la cellule STT-MRAM: les étoiles et les points sont issus du modèle statistique (1000 simulations MC); Les trames viennent du modèle le plus défavorable de MTJs ($sigma = 0.01$ et $n = 4.5$) et de transistors CMOS.

IV.c1 Transistors avec technologie UTBB-FDSOI

Les transistors à technologie UTBB FDSOI ont une gamme de polarisation plus étendue que celle du MOSFET traditionnel. La polarisation de grille arrière est fournie par un autre point d'alimentation, par exemple, 1V à polarisation directe par le biais de NMOS avec un 1V au substrat de transistor (normalement 0V). La polarisation du substrat utilisé dans la grille arrière peut diminuer la tension de seuil du transistor (V_{th}) et augmenter le courant de drain du transistor (I_d). Ainsi, les performances des transistors sont améliorées (par exemple, le courant de drain, la vitesse de commutation), alors que le transistor à polarisation du substrat inverse (RBB) réalise une partie du compromis de robustesse de performance [143]. Comme montré dans la Figure B.15, lorsque le transistor fonctionne en saturation, la variation de V_{th} est indépendante des différentes polarisations du substrat. En ce qui concerne le coefficient de variation V_{th} (le rapport entre l'écart typique et la moyenne), il est trois fois plus élevé dans les transistors avec FBB que ceux avec RBB. Cependant, en raison de l'amplification de la variation de la porte par la polarisation de la fluctuation du RDF [143], la fluctuation de I_d augmente avec FBB, alors que RBB peut réduire 24.3% de variabilité de I_d par rapport à FBB. Cette propriété peut être

combinée de façon appropriée avec des circuits symétriques basés sur MTJ pour explorer l'optimisation de la performance et de la fiabilité [117, 135, 147].

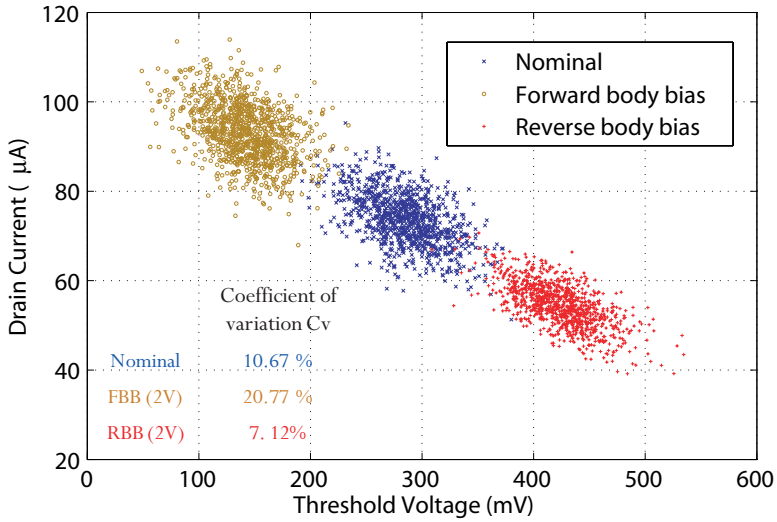


Figure B.15: Variabilité FDSOI: Un seul transistor NMOS fonctionne dans la région de saturation. Le coefficient de V_{th} variation est analysé entre FBB et la conception nominale (sans biais corporel) et FBB.

III.c2 Conception d'une bascule non volatile utilisant polarisation du substrat asymétrique dynamique de FDSOI

L'approche est mise en œuvre avec le circuit illustré dans Figure B.16. On se focalise sur le circuit PCSA dans lequel tous les transistors sont avec une taille minimale (W/L=80nm/30nm). Afin d'augmenter la différence de résistance, deux circuits RC sont insérés pour générer des tensions de polarisation pour les deux branches de PCSA. Pendant la phase d'écriture, une seule des deux tensions de polarisation: V_0 et V_1 est chargée à V_{dd}. Par exemple, après avoir écrit "1", V_1 est chargé à V_{dd} et V_0 est déchargé à la masse. Pendant la phase de détection, avec la tension de polarisation V_0 et V_1, la résistance des transistors à côté de MTJ1 est réduite. Par conséquent, la différence de résistance des deux côtés est agrandie, ce qui donne une meilleure performance de détection.

III.c3 Analyse de fiabilité et évaluation de performance

Des simulations de Monte-Carlo ont été réalisées pour étudier la fiabilité du circuit proposé en tenant compte des variations du processus, de la tension et de la fluctuation de température. La Figure B.17 montre le taux d'erreur de lecture du circuit PCSA sous différentes conditions. Elle montre également que les erreurs de lecture ont été quasiment

éliminées par la méthode proposée avec DABB. Comme la différence de résistance des deux branches est agrandie, la variabilité peut être en partie masquée par la différence de courant suffisante.

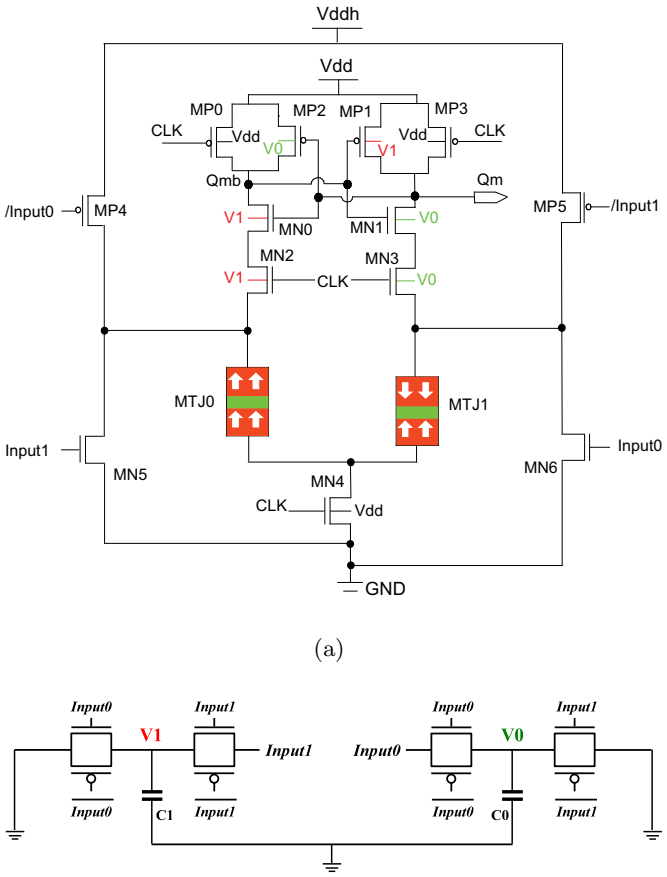
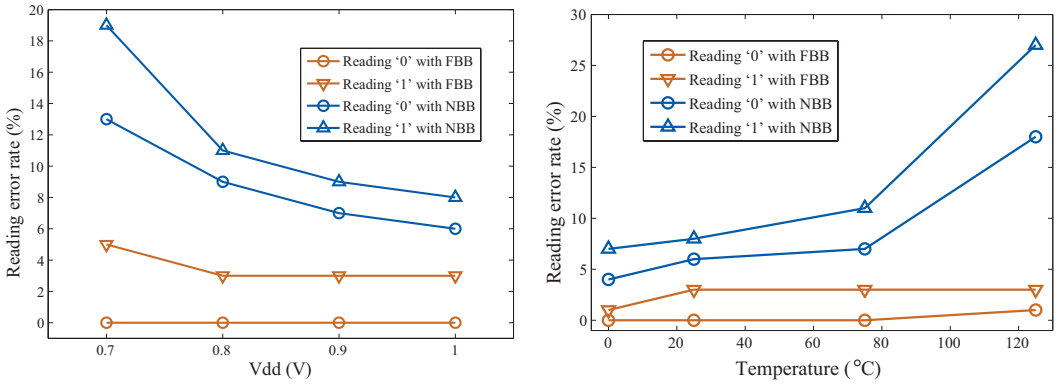


Figure B.16: (a) Amplificateur de détection de pré-charge avec polarisation du substrat asymétrique dynamique (b) Les circuits RC génèrent les tensions de polarisation du substrat pour les transistors dans PCSA.

Avec Vdd supérieur à 0,8 V, le taux d'erreur de lecture est indépendant de la tension d'alimentation, car une tension d'alimentation plus élevée n'est plus efficace pour augmenter la différence de courant entre les deux branches. Cette propriété est utile dans la conception orientée faible consommation. De plus, la défaillance de détection est presque indépendante de la fluctuation de température. Comme les condensateurs sont également fortement dépendants de la température, les valeurs de tension de polarisation du corps V_0 et V_1 changent avec la température. L'effet des variations de V_0 et V_1 sur les transistors a presque compensé l'impact de la température sur le TMR. Par conséquent, le taux d'erreur de lecture du circuit proposé est pratiquement à l'abri de variations de température.



(a) (b)

Figure B.17: Taux d'erreur de lecture du NVFF en fonction de différentes (a) tensions d'alimentation et (b) conditions thermiques.

Cette méthode a été mise en œuvre pour un additionneur total non volatile [148], conduisant à optimisation de la performance du circuit de détection, y compris la latence du circuit, la puissance dynamique, la variabilité et la probabilité de réussite de détection.

V Nouvelles applications du MTJ dans les circuits conventionnels

Bien que les problèmes de fiabilité puissent perturber le fonctionnement de certains circuits et fonctions, d'autres applications peuvent en profiter. Ce chapitre se concentre sur de nouvelles mises en œuvre d'applications classiques dans lesquelles il est tiré profit du comportement de commutation stochastique.

V.a Une nouvelle conception de générateur de nombres aléatoires vraie basée sur MTJ

Les nombres aléatoires sont largement utilisés dans les systèmes de cryptographie et de sécurité. Cependant, la plupart des générateurs de nombres aléatoires vrais (TRNG) qui utilisent le caractère aléatoire physique sont de haute complexité et de forte consommation d'énergie. Cette section propose un nouveau circuit de TRNG utilisant la jonction tunnel magnétique (MTJ) [149]. Le comportement de commutation stochastique de MTJ fournit une nouvelle source de hasard pour le TRNG. Basé sur un phénomène physique imprévisible, il peut fournir de vrais flux de bits aléatoires par la conception appropriée de circuit. Un courant de commutation adaptable I_{sw} à une impulsion de 5ns est appliqué pour étudier la probabilité de commutation. $I_{sw} = 84.5\mu A$ est requis pour une proba-

bilité de commutation de 50%. Deux valeurs de courant des pires cas (maximum (FF) et minimum (SS)) sont obtenues pour la commutation avec probabilité de 50%.

L'architecture générale proposée pour le circuit est illustrée dans la Figure B.18. Elle est composés d'une partie d'écriture aléatoire MTJ, d'un amplificateur de détection de pré-charge (PCSA) et d'un bloc de correction. Avec un choix approprié de dimensions de transistors, une probabilité de commutation particulière peut être obtenue pour obtenir un flux binaire réel aléatoire. Afin d'améliorer la fiabilité, un bloc logique de correction composé de compteurs et de comparateur est implémenté. Ce bloc génère un signal de commande pour moduler le courant de commutation, ce qui garantit la probabilité exacte de flux binaire de nombre aléatoire obtenu (idéalement avec 50% de '1' et 50 % de '0').

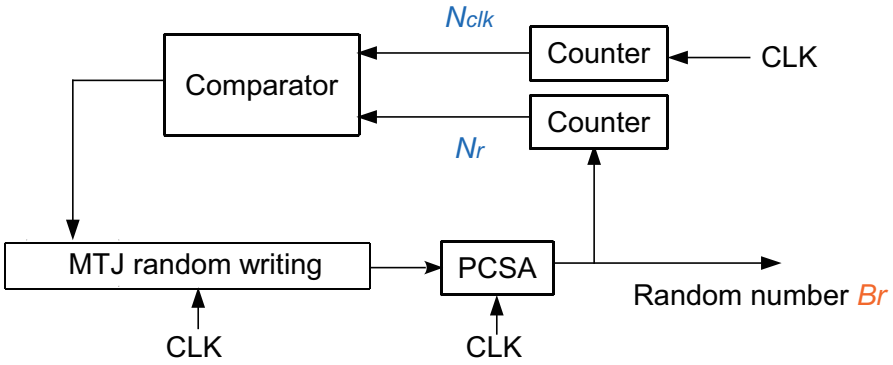


Figure B.18: Architecture du circuit de TRNG proposée.

Le chronogramme correspondant est présenté dans la Figure B.19. Tout d'abord, le circuit de commutation commence à écrire avec un courant relativement faible (84.5 μA) et la condition de N_r=$N_{clk}/2$ (probabilité P_{sw}=50%). Avec $N_r > N_{clk}/2$, la probabilité de commutation est diminuée tandis que le transistor de contrôle N_{c0} est ouvert par le bloc de correction. Lorsque N_r est inférieur à $N_{clk}/2$, le courant de commutation est augmenté avec le transistor de commande N_{c2} activé par le bloc de correction. Ce résultat de simulation correspond bien à l'objectif de conception susmentionné et la fonctionnalité est bien confirmée.

En comparaison avec les travaux présentés dans [159], la conception proposée conduit à une surface plus petite (DAC n'est pas utilisé) et des étapes d'accord plus courtes (avec un court délai avant de générer des flux de bits aléatoires stables). De plus, le courant de commutation du MTJ dans [159] est beaucoup plus élevé, alors que le PCSA proposé dans cette thèse est ultra basse consommation, et le bloc DAC consomme beaucoup d'énergie.

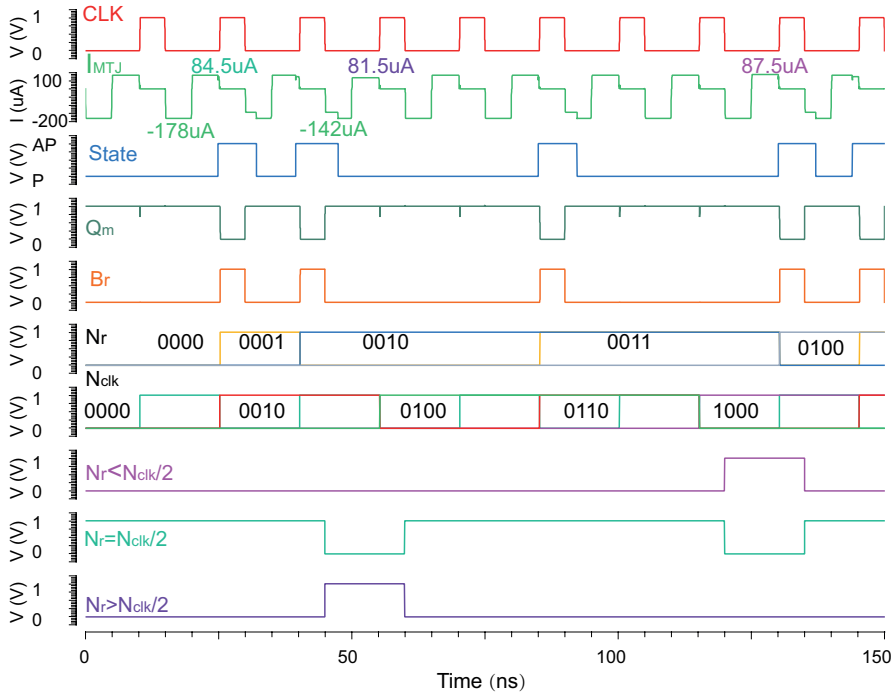


Figure B.19: Diagramme temporel du circuit proposé.

V.b Réalisation de calcul stochastique avec MTJ

Le calcul stochastique (SC) avec des flux binaires aléatoires a été utilisé pour remplacer le codage binaire. Les circuits logiques basés sur SC bénéficient de la minimisation de surface, du fonctionnement rapide et précis et de la tolérance de pannes inhérente. Dans cette section, les caractéristiques stochastiques inhérentes des STT-MTJ conduisent à un circuit innovateur de générateur de nombres stochastique (SNG) [167]. Le processus de MOS-MTJ hybride permet de concevoir une structure SNG de 4T1M avec une surface de 1,98 μm * 1,46 μm, en utilisant la technologie FDSOI de 28 nm. Une étude de cas de SNG conçu a été réalisée pour la synthèse de fonction polynomiale, ce qui a réduit considérablement la surface requise. Le circuit proposé profite également de la non-volatilité et de l'endurance infinie des STT-MTJ, qui peuvent être appliquées à des circuits et des systèmes fiables.

Le nouveau générateur de nombres stochastiques est proposé pour générer une sortie probabiliste en tant que flux binaire stochastique en utilisant STT-MTJ. Pour obtenir un comportement probabiliste, le courant de fonctionnement MTJ est réglé comme étant inférieur à un courant suffisant. MTJ SP est estimée sur la base de la simulation MC au niveau du transistor. Les probabilités de commutation par rapport au courant de fonctionnement MTJ sont représentées dans la Figure B.20. Le comportement stochastique est également déterminé par la fréquence de fonctionnement. Les signaux d'entrée de 50 MHz (durée d'écriture MTJ = 10 ns) et 100 MHz (durée d'écriture MTJ = 5 ns) sont

simulés avec la méthode MC. La probabilité du signal généré par le MTJ peut être utilisée pour le calcul logique stochastique.

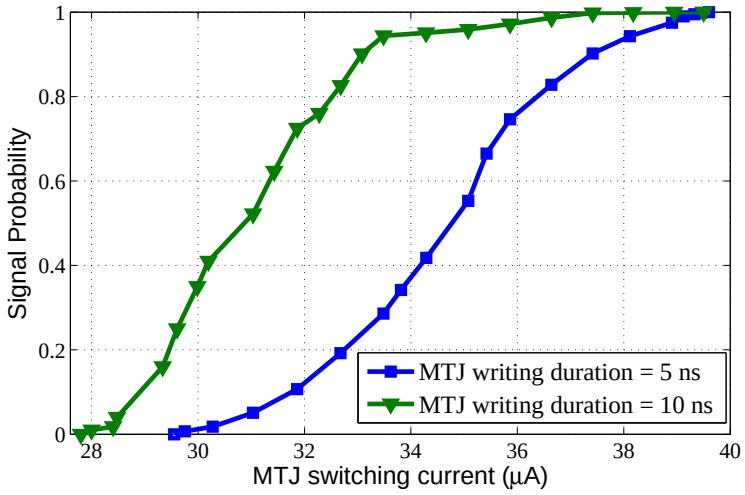


Figure B.20: Résultat de simulation: probabilité de commutation par rapport au courant de fonctionnement MTJ.

Le SNG basé sur STT-MTJ peut être utilisé pour synthétiser la fonction polynomiale au niveau de Transfert de Registre (RTL). Dans ce travail, nous considérons la synthèse de la fonction y décrite dans (B.2).

$$y = 0,06x^2 + 0,19x + 0,25 \tag{B.2}$$

Avec un signal d'entrée de 10 bits, nous comparons les résultats de synthèse RTL du flux de synthèse traditionnel (signal binaire) à ceux d'un flux de synthèse stochastique. La surface du circuit est évaluée avec un kit de conception FD-SOI de 28 nm.

La Figure B.21 montre le circuit de SNG utilisant STT-MTJ pour la fonction polynomiale. Il ne se compose que d'un générateur basé sur 5 STT-MTJ, 3 MUXES et 2 AND portes logiques. La surface totale par cellule est inférieure à 100 mum^2 selon le kit de conception de 28 nm. Si l'on considère les transistors dans la gamme de faible puissance, le délai de propagation estimé pour la méthode stochastique est de 52 ps, ce qui est bien inférieur à délai de 360,8 ps obtenu par synthèse de la méthode traditionnelle. La description détaillée de la synthèse polynomiale de Bernstein se trouve dans [178].

V.b Méthode de calcul approximatif en utilisant MTJ

Le calcul approximatif a montré son potentiel dans les systèmes informatiques de la prochaine génération. Dans cette section, une nouvelle conception de niveau de circuit

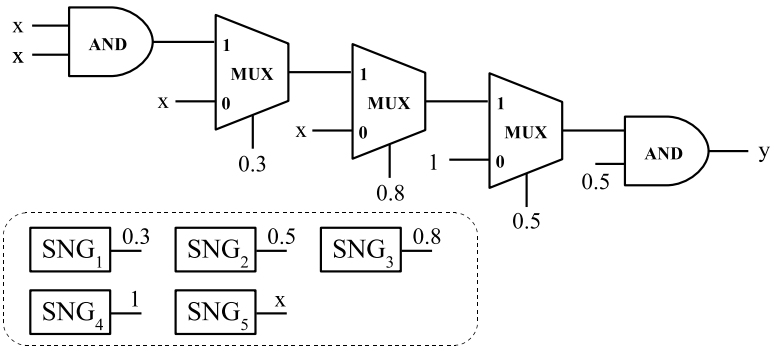


Figure B.21: Un example de synthèse de fonction polynomiale.

pour calcul approximatif est proposée, en se basant sur la structure logique en mémoire non volatile (NV). Deux types d'additionneurs approximatifs (AX-MFA1 et AX-MFA2) sont implémentés avec une reconfiguration de circuit et un courant d'écriture insuffisant. Les résultats de la simulation sont présentés, y compris la consommation d'énergie, la latence du circuit, la consommation de fuite, la distance d'erreur et la fiabilité.

Afin de garantir le fonctionnement de MTJ, le seuil du courant de commutation est requis dans les circuits hybrides MOS/MTJ classiques. Habituellement, un courant élevé ($I > 2 * I_{c0}$) est appliqué pour garantir la vitesse d'écriture élevée (sub 3 ns) en mémoire. Ici, l'écriture de MTJ insuffisante est utilisée pour générer un signal approximatif (entrée B).

AX-MFA2 peut fonctionner de manière adaptative dans le mode précis avec un V_{dd} supérieur à 0,8V, alors que la commutation de MTJ est déterministe. D'autre part, nous utilisons le comportement de commutation de MTJ comme mécanisme de sélection entre le mode précis et le mode approximatif: lorsque V_{dd} est inférieur à 0,8V, le courant de commutation MTJ est insuffisant pour traiter une nouvelle entrée. Ainsi, un MFA approximatif peut être mis en œuvre avec des données approximatives B.

La Figure B.22 montre les simulations de synchronisation de l'additionneur approximatif avec une complexité logique réduite. Toutes les combinaisons possibles d'entrées A, B (données NV stockées dans MTJs) et C_{in} sont désignées. L'additionneur approximatif devient actif pour le calcul lorsque le signal d'horloge est élevé. Notez que la sortie inexacte Sum est marquée alors que la sortie C_{in} est exacte. La distance d'erreur totale est de 4 en logique simplifiée basée sur AX-MFA1.

Étant donné que l'architecture MFA à double mode proposée est la même que celle des MFA précédentes [47, 164], la simulation fonctionnelle du mode précis pour MFA (au-dessus de l'alimentation 0,8V) n'est pas présentée en détail. Lorsque V_{dd} est inférieur à

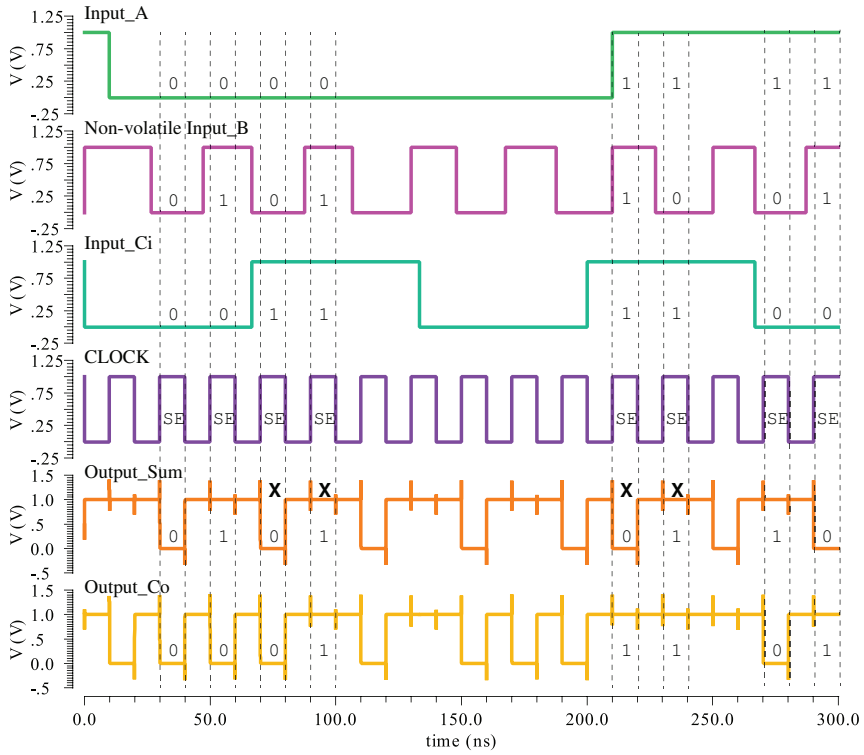


Figure B.22: Les simulations de transition de l'additionneur approximatif avec une complexité logique réduite (AX-MFA1). La sortie *Sum* est avec des erreurs.

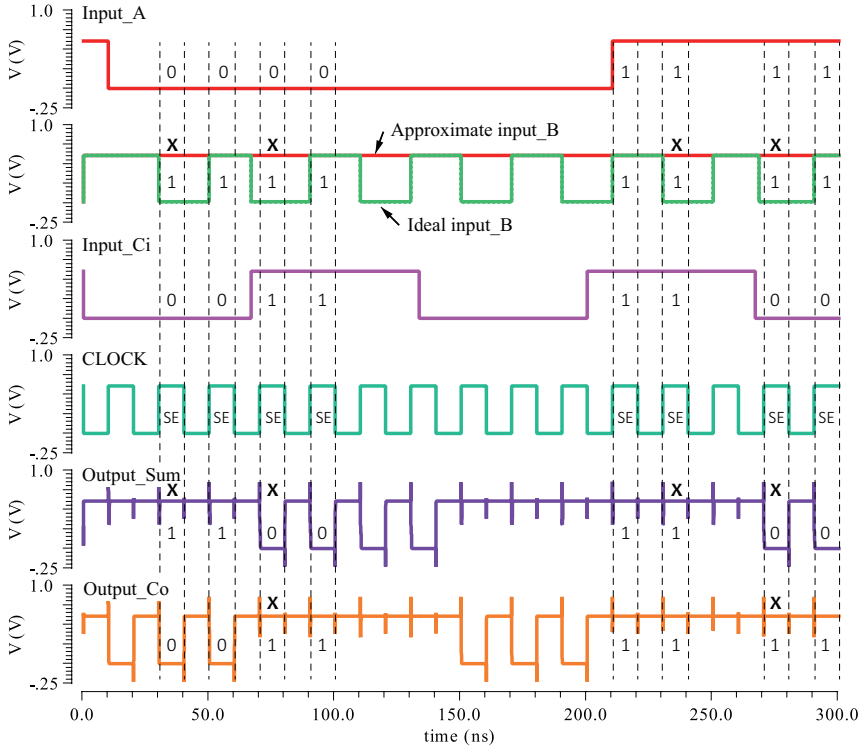


Figure B.23: Les simulations de transition de l'additionneur approximatif avec double mode (AX-MFA2). La sortie *Sum* et C_o est avec des erreurs.

0,8V, le courant dans MTJ est inférieur au seuil et les états de MTJ ne peuvent pas être changés entre P et AP. Ainsi, la consommation d'énergie dans les opérations de détection et d'écriture MTJ est considérablement réduite. Ainsi, la sortie inexacte se produit à la sortie de l'additionneur (Sum et C_{in}). Le mode approximatif peut être exécuté sur une alimentation allant de 0,3 V à 0,75 V. Les résultats de simulation transitoire de l'additionneur approximatif à double mode sont représentés dans la Figure B.23. Une alimentation de 0,5 V est appliquée dans cette structure. La distance d'erreur totale est de 6 dans ce MFA approximatif simplifié.

Un minimum de V_{dd} à 360mV peut être utilisé dans le MFA de 1 bit proposé pour l'opération sous-V_t, où les transistors nMOS sont à 366.5mV V_t, et les transistors pMOS sont à 416,5mV V_t. L'opération près de V_t peut être réalisée avec 420mV V_{dd}. La Figure B.24 montre la performance en latence pour le MFA à bimode proposé. En mode précis avec alimentation 1V, une latence de 19,3 ps est atteinte. Une latence d'approximativement 152,7 ps est obtenue en Sum lorsque V_{dd} = 0,5V. L'abaissement continu de l'alimentation jusqu'à la région sous-V_t entraîne une latence importante (1,27 ns lorsque V_{dd} = 0,36V).

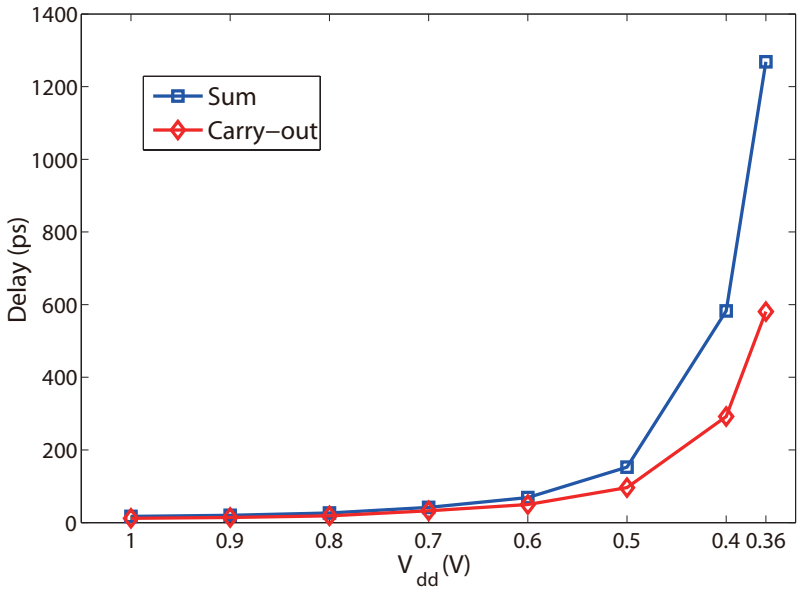


Figure B.24: Simulation de latence de MFA à double mode.

La Table B.1 résume les résultats de la simulation du MFA approximatif proposé avec une logique simplifiée (AX-MFA1) et une AMF bimode(AX-MFA2). Nous comparons les performances des structures CMOS approximatives completes (CMOS AX-FA) et les MFA précédents pour différents nœuds CMOS. L'additionneur approximatif avec une logique simplifiée consomme moins d'énergie que MFA classique. Il accélère également la vitesse

de 17%. L'additionneur approximatif avec l'écriture insuffisante peut fonctionner avec une faible tension d'alimentation. L'énergie de détection et d'écriture est réduite de près de 70%. Un autre avantage de cette technique est que les concepteurs peuvent sélectionner le mode de fonctionnement MFA entre précis et approximatif.

Table B.1: Comparaison de performance de MFA conventionnel et MFA approximatif propsés.

	1-bit adder Type	V_{dd} (V)	Delay (ps)	Erreur dist.		Dynamique power(nW)	Fuite (pW)	Dispositif count	Surface (μm^2)
				Sum	C_o				
CMOS AX-FA [1)]	AXA1	1	20,14	4	4	9,697	1073	8T	0,81
	AXA2	1	69,83	4	0	6,984	1362	6T	0,64
	AXA3	1	48,6	2	0	9,041	1397	8T	0,77
Previous MFA	[47] 65nm bulk	1	170	0	0	2950	0[3)]	30T+4M	20
	[46] 40nm bulk	1,2/1,5	87,4	0	0	1980	<1 nW[3)]	38T+4M	20
	[164] 28nm bulk	1	150 (8 bit)	0	0	$0{,}68pJ/8bits$	0[3)]	25T+4M	24,81[2)]
This work	AX-MFA1	1	16,22	4	0	8,69	329,5	21T+4M	8,51
	AX-MFA2(accurate)	1	19,3	0	0	9,46	401,6	25T+4M	8,74
	AX-MFA2(approximate)	0,5	152,7	4	2	2,112[4)]	77,91/5,06[5)]	25T+4M	8,74

[1)] CMOS AX-FAs sont implémenté avec 28nm FDSOI technologie.

[2)] 1 bit MFA layout surface comprends bascule magnétique.

[3)] Fuite en mode active n'est pas considéré. Zéro fuite est attente seulement en mode stand-by.

[4)] 2,112nW, $0{,}00195pj/bit$.

[5)] 77,91pW est atteint sans polarisation de poly, 5,06pW est réalisé par 16nm polarisation de poly.

VI Conclusions

Cette thèse a été dédiée à l'analyse et l'amélioration de la fiabilité, ainsi qu'à l'exploration de nouvelles applications basées sur des dispositifs PMA-STT-MTJ. Le travail comprend principalement trois parties: Modélisation compacte des principaux problèmes de fiabilité existant dans PMA-MTJ; Analyse de fiabilité des circuits logiques et de mémoire typiques ainsi que proposition de méthodologies de conception fiables; Nouvelles conceptions d'applications spécifiques traditionnelles en bénéficiant de la commutation stochastique des PMA-MTJ.

A travers l'étude de l'état de l'art, nous avons revu le développement des dispositifs spintroniques et leur application dans les circuits de mémoire et logique. L'évolution de la jonction tunnel magnétique est due à l'optimisation de l'approche par commutation et à l'amélioration du TMR. Avec la comparaison détaillée de différentes méthodes de commutation, il a été conclu que STT est le candidat le plus approprié pour les futures mémoires et les circuits logiques grâce à leur meilleur compromis entre consommation d'énergie, vitesse d'opération, miniaturisation, endurance et intégration 3D dans les circuits CMOS conventionnels. Cependant, ces dispositifs souffrent de problèmes de fiabilité considérables qui limitent leur large commercialisation. Le passage en revue de la littérature sur la fiabilité des MTJ a montré que les travaux actuels ne répondent pas à l'exigence urgente de

conception de haute fiabilité, ce qui nous a motivés à étudier les problèmes de fiabilité de MTJ et à créer un modèle compact précis pour les concepteurs de circuits.

Dans la partie de la modélisation compacte, nous avons analysé les sources des problèmes de fiabilité, y compris les variations de processus, la commutation stochastique, la fluctuation de température et le claquage diélectrique. Sur la base de l'étude approfondie des mécanismes physiques générant des problèmes de fiabilité, différents modèles physiques ont été considérés pour constituer la modélisation compacte. Ensuite, les résultats de la simulation ont été présentés pour valider sa fonctionnalité. Ces modèles peuvent être utilisés pour exécuter une conception plus réaliste en fonction des contraintes obtenues à partir de la simulation. Grâce à la validation de ce modèle, les concepteurs de circuits sont en mesure de prédire la performance du circuit avec précision. Le modèle proposé a été développé dans un langage compatible avec SPICE et peut être utilisé dans tous les environnements pour des simulations au niveau de circuit.

Une approche non Monte-Carlo pour l'analyse de la variabilité des circuits à base de MTJ a également été présentée. Elle a été mise en œuvre en utilisant le pire cas de PMA-MTJ et transistor. Les spécifications de conception ont été détaillées pour les cellules de mémoire non volatiles et l'unité arithmétique. Les résultats de la simulation montrent que la solution proposée est beaucoup plus efficace que la méthode conventionnelle Monte-Carlo (MC) tout en gardant la même cible d'évaluation de la performance.

En s'appuyant sur les modèles validés, nous avons effectué une analyse de fiabilité des circuits de mémoire et de logique MTJ couramment utilisés. La première étape a consisté à démontrer l'effet de chaque problème de fiabilité sur l'ensemble du circuit. Parallèlement, l'impact des paramètres MTJ sur la performance des circuits a également été étudié. En conséquence, nous avons trouvé l'impact de différents paramètres sur les termes de performance. Pendant la phase de conception, tous les paramètres doivent être considérés pour un meilleur compromis. De même, l'analyse de la fiabilité,nous conduits à proposer une méthode de conception tolérante à la variation PVT en utilisant la polarisation du substrat asymétrique dynamique des transistors FDSOI en technologie 28 nm. Cette approche a été mise en œuvre dans le circuit PCSA pour démontrer sa faisabilité. Les résultats de simulation ont montré une amélioration significative du succès de la lecture dans différentes conditions à surface de circuit minimum. Le circuit proposé se caractérise par une robustesse parfaite avec la prise en compte de la variabilité du processus, de la tension et de la température.

Le comportement stochastique de la commutation est normalement un facteur de

dégradation de la performance dans la mémoire et les circuits arithmétiques. Cependant, il peut être utile ou même avantageux dans certaines applications spécifiques. Nous avons proposé de nouvelles conceptions du générateur de nombres aléatoires vrais (TRNG), du calcul stochastique (SC) et du calcul approximatif en utilisant ce phénomène. Le TRNG proposé démontre une parfaite stabilité contre la variabilité, qui est prise en compte pour la première fois dans les TRNG basés sur MTJ. Comparé à d'autres travaux, il offre une puissance plus faible, une vitesse de fonctionnement plus élevée, une meilleure robustesse et une surface plus compacte. Dans la conception de SC, une étude de cas de SNM conçu a été réalisée par la synthèse d'une fonction polynomiale et a conduit à la minimisation de surface considérablement comparée à la méthode conventionnelle de synthèse binaire. Enfin, deux types d'additionneurs approximatifs NV ont été implémentés avec une reconfiguration de circuit et un courant d'écriture insuffisant. Par rapport à la simplification logique traditionnelle, l'écriture insuffisante de MTJ est plus efficace, ce qui permet de réduire drastiquement la consommation de l'énergie. De plus, les problèmes de variabilité et de fuite ont été surmontés grâce à la méthode de dopage dynamique et la polarisation des transistors FDSOI.

Les résultats obtenus dans le cadre de cette thèse nous ont convaincus que les propositions (modélisation, analyse, solutions architecturales) effectuées contribuent de manière significative au développement de STT-MRAM et de la logique en mémoire. Le modèle compact prenant en compte les problèmes de fiabilité et les méthodologies peuvent être utilisés par les concepteurs de circuits pour réaliser, en moins de temps, des conceptions plus robustes et avec un rendement plus élevé. Les nouvelles conceptions tirant profit de la commutation stochastique fournissent de nouvelles pistes pour l'utilisation de MTJ et élargissent ses potentiels dans les applications futures.

Analyse de Fiabilité de Circuits Logiques et de Mémoire basés sur Dispositif Spintronique

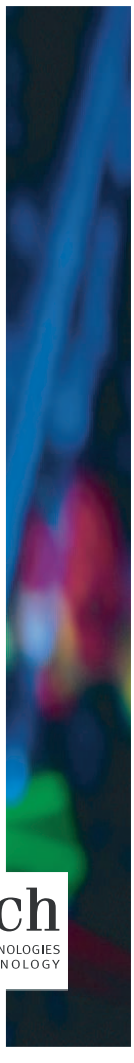
You WANG

RESUME: La jonction tunnel magnétique (JTM) commutée par la couple de transfert de spin (STT) a été considérée comme un candidat prometteur pour la prochaine génération de mémoires non-volatiles et de circuits logiques, car elle fournit une solution pour surmonter le goulet d'étranglement de l'augmentation de puissance statique causée par la mise à l'échelle de la technologie CMOS. Cependant, sa commercialisation est limitée par la fiabilité faible, qui se détériore gravement avec la réduction de la taille du dispositif. Cette thèse porte sur l'étude de la fiabilité des circuits basés sur JTM. Tout d'abord, un modèle compact de JTM incluant les problèmes principaux de fiabilité est proposé et validé par la comparaison avec des données expérimentales. Sur la base de ce modèle précis, la fiabilité des circuits typiques est analysée et une méthodologie d'optimisation de la fiabilité est proposée. Enfin, le comportement de commutation stochastique est utilisé dans certaines nouvelles conceptions d'applications classiques.

MOTS-CLEFS: Jonction tunnel magnétique, Analyse de fiabilité, Modèle compact, Polarization du corps asymétrique dynamique, Générateur de nombre aléatoire vrai, Calcul approximatif, Calcul stochastique

ABSTRACT: Spin transfer torque magnetic tunnel junction (STT-MTJ) has been considered as a promising candidate for next generation of non-volatile memories and logic circuits, because it provides a perfect solution to overcome the bottleneck of increasing static power caused by CMOS technology scaling. However, its commercialization is limited by the poor reliability, which deteriorates severely with device scaling down. This thesis focuses on the reliability investigation of MTJ based non-volatile circuits. Firstly, a compact model of MTJ including main reliability issues is proposed and validated by the comparison with experimental data. Based on this accurate model, the reliability of typical circuits are analyzed and reliability optimization methodology is proposed. Finally, the stochastic switching behavior is utilized in some new designs of conventional applications.

KEY-WORDS: Magnetic tunnel junction, Reliability analysis, Compact model, Dynamic asymmetrical body bias, True random number generator, Approximate computing, Stochastic computing



TELECOM ParisTech

EDiF

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY