amU
Aix Marseille Université

NNT : 2024AIXM0001

THÈSE DE DOCTORAT

Co-encadrée au CEA-Leti par Laurent GRENOUILLET, Jean COIGNUS et Simon MARTIN
Soutenue à AMU — Aix-Marseille Université
le 19 décembre 2024 par

Julie LAGUERRE

Ultra-Low Power Ferroelectric Memory Arrays based on Ferroelectric HfO_2: Scalability towards 28nm Node and Below

Discipline
Sciences pour l'Ingénieur

Spécialité
Micro et Nanoélectronique

École doctorale
ED 353 Sciences pour l'Ingénieur : Mécanique, Physique, Micro et Nano électronique

Laboratoire/Partenaires de recherche
Institut Matériaux Microélectronique Nanosciences de Provence (IM2NP)
Commissariat à l'Energie Atomique et aux Energies Alternatives, Laboratoire d'Electronique et de Technologie de l'Information (CEA-Leti)

Composition du jury

Damien DELERUYELLE PR, INSA Lyon	Président du jury, Rapporteur
Quentin RAFHAY MCF, Grenoble INP Phelma	Rapporteur
Caroline BORDERON MCF, Nantes Université	Examinatrice
Cristell MANEUX PR, Université de Bordeaux	Examinatrice
Marc BOCQUET PR, Aix-Marseille Université	Directeur de thèse
Simon JEANNOT, IR, STMicroelectronics	Membre invité

im2np
cea leti

"It brings me peace of mind to know that this will all just carry on
With someone else
With something new
No need to be blue"

100% Endurance - Yard Act

Affidavit

I, undersigned, Julie Laguerre, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific supervision of Marc Bocquet, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the french national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Grenoble, 19th December 2024



This work is licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Public License

List of publications and participation in conferences

List of publications and/or patents produced within the PhD framework:

1. Barrett N, Hamouda W, Lubin C, Laguerre J, Carabasse C, Vaxelaire N, *et al.* Oxygen vacancy engineering in Si-doped, HfO_2 ferroelectric capacitors using Ti oxygen scavenging layers. Applied Physics Letters. 22 July 2024;125(4):043502.

Participation in conferences and summer schools during the PhD period:

1. Laguerre J, Bocquet M, Billoint O, Martin S, Coignus J, Carabasse C, *et al.* Memory Window in Si:HfO_2 FeRAM arrays: Performance Improvement and Extrapolation at Advanced Nodes. *2023 IEEE International Memory Workshop (IMW)*. Monterey, CA, USA: IEEE; 2023. p. 1-4. *Oral presentation*

2. Grenouillet L, Martin S, Laguerre J, Barbot J, Vaxelaire N, Deleuze PM, *et al.* Synergetic FeRAM/OxRAM memory array operation induced by oxygen vacancy interface engineering in Si:HfO_2 ferroelectric films. *2023 Solid State Devices and Materials (SSDM)*. Nagoya, Japan ; 2023. p. 1-2. *Oral presentation*

3. Grenouillet L, Barbot J, Laguerre J, Martin S, Carabasse C, Louro M, *et al.* Reliability assessment of hafnia-based ferroelectric devices and arrays for memory and AI applications (Invited). *2023 IEEE International Reliability Physics Symposium (IRPS)*. Monterey, CA, USA: IEEE; 2023. p. 1-8. *Oral presentation*

4. Grenouillet L, Laguerre J, Martin S, Barbot J, Carabasse C, Louro M, *et al.* 7nm HZO-based FeRAM array performance at 130nm node and scaling perspectives at 28nm node (Invited). *2023 Non-Volatile Memory Technology Symposium (NVMTS)*. imec, Leuven, Belgium ; 2023. p. 1-6. *Oral presentation*

5. Martemucci M, Rummens F, Hirtzlin T, Martin S, Guille O, Januel T, *et al.* Hybrid FeRAM/RRAM Synaptic Circuit Enabling On-Chip Inference and Learning at the Edge. *2023 International Electron Devices Meeting (IEDM)*. San Francisco, CA, USA: IEEE; 2023. p. 1-4. *Oral presentation*

6. Martin S, Opprecht M, Coignus J, Carabasse C, Jahan C, Louro M, *et al.* Towards wake-up free $Hf_{0.5}Zr_{0.5}O_2$-based ferroelectric capacitors thanks to Nanosecond Laser Annealing. *2023 IEEE Semiconductor Interface Specialists Conference (SISC)*. San Francisco, CA, USA: IEEE; 2023. p. 1-2. *Poster presentation*

7. Laguerre J, Martin S, Barbot J, Carabasse C, Louro M, Grimaud F, *et al.*, Insights from 16kbit HfO_2-based FeRAM array characterization at 130nm node towards FeCAP integration at 28nm node. *2024 High-K Workshop*. Dresden, Germany ; 2024. *Oral presentation*

8. Laguerre J, Martin S, Coignus J, Carabasse C, Bocquet M, Andrieu F, *et al.* Data Retention Insights from Joint Analysis on BEOL-Integrated HZO-Based Scaled FeCAPs and 16kbit 1T-1C FeRAM Arrays. *2024 IEEE International Reliability Physics Symposium (IRPS)*. Grapevine, TX, USA: IEEE; 2024. p. 1-7. *Oral presentation*; **"People's Choice for Best Paper Award"**

9. Billoint O, Martin S, Laguerre J, Hosier L, Coignus J, Carabasse C, *et al.* Charge-based Sense Demonstration in 1T-1C HZO FeRAM Arrays to Overcome C_{BL}-induced Bank Size Limitations. *2024 IEEE International Memory Workshop (IMW)*. Seoul, South Korea: IEEE; 2024. p. 1-4. *Oral presentation*

10. Laguerre J. 180s Pitch: Ferroelectric memories for ultra-low power computing in a data-centric world. *2024 Leti Innovation Days (LID)*. Grenoble, France. *Oral presentation*

11. Laguerre J, Martin S, Hosier L, Coignus J, Carabasse C, Bocquet M, *et al.* Recovery Strategies and Related Mechanisms to Overcome Fatigue-Limited Endurance on HZO-based FeCAP and FeRAM Arrays. *2024 Solid State Devices and Materials (SSDM)*. Himeji, Japan ; 2024. p. 1-2. *Oral presentation*

12. Berthaud F, Castellani N, Hosier L, Rottner J, Meli V, Grenouillet L, *et al.* New insight of ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ properties under cryogenic temperatures in integrated ferroelectric capacitors. *2024 Solid State Devices and Materials (SSDM)*. Himeji, Japan ; 2024. p. 1-2. *Oral presentation*

13. Rhetat L, Noel JP, Giraud B, Grenouillet L, Laguerre J, Marchand C, *et al.* A Novel Design Technique for Enhanced Security and New Applications of Ferroelectric-based Non-Volatile SRAM. *2024 IFIP/IEEE International Conference on Very Large Scale Integration (VLSI-SoC)*. Tanger, Morocco: IEEE; 2024. p. 1-6. *Oral presentation*

14. Martin S, Jahan C, Hosier L, Grimaud F, Louro M, Laguerre J, *et al.* $Hf_{0.5}Zr_{0.5}O_2$ FeRAM scalability demonstration at 22nm FDSOI node for embedded applications. *2024 International Electron Devices Meeting (IEDM)*. San Francisco, CA, USA: IEEE; 2024. p. 1-4. *Oral presentation*

Résumé et mots clés

Depuis leur découverte au début des années 2010, les propriétés ferroélectriques de l'oxyde d'hafnium (HfO_2) déposé en couches minces font l'objet de recherches intensives, tant par les universitaires que par les industriels. Le CEA-Leti s'intéresse particulièrement à ce matériau pour des applications de mémoire non-volatile embarquée à très basse consommation, qui pourraient dépasser les limites de miniaturisation rencontrées par les FeRAM à base de pérovskites de plomb (PZT). Dans ces dernières, les capacités ferroélectriques sont intégrées au nœud technologique 130nm, c'est-à-dire dans des circuits FeRAM avec des transistors dont la longueur de grille est de 130nm, loin de la dizaine de nanomètres développée actuellement. Grâce à leur compatibilité avec les lignes de production CMOS, les capacités ferroélectriques à base d'oxyde d'hafnium ont le potentiel d'être intégrées à des nœuds plus agressifs, tels que les nœuds 28 ou 22nm. Ce faisant, elles constitueraient une technologie très basse consommation, qui pourrait trouver une place de choix dans la hiérarchie des mémoires embarquées à l'heure du déluge de données générées par les secteurs allant de l'automobile aux produits électroniques grand public, en passant par l'industrie 4.0 et les applications d'intelligence artificielle.

L'objet de cette thèse est donc d'intégrer des matrices FeRAM performantes et fiables à base de HfO_2 ferroélectrique vers les nœuds 28nm et en-deçà, en s'appuyant sur une technologie FeRAM BEoL préalablement évaluée au nœud 130nm. Pour ce faire, une étude théorique est d'abord menée afin de déterminer les dimensions des capacités ferroélectriques requises pour assurer une marge de lecture suffisante dans des matrices FeRAM à des nœuds avancés, et présentant donc des tensions et courants limités. Les capacités ferroélectriques sont ensuite optimisées pour améliorer leurs propriétés ferroélectriques intrinsèques, et des démonstrateurs sont réalisés en vue d'intégrer des matrices FeRAM à capacité et densité élevées. Les sources de variabilité pénalisant la marge de lecture à l'échelle des matrices sont également étudiées et corrigées, lorsque cela est possible. En effet, la marge de lecture doit être suffisante pour assurer la fonctionnalité des matrices FeRAM dans toutes les conditions d'utilisation, et au cours de leur usage. En particulier, après un long temps de stockage des données, après un certain nombre de programmation / lecture à basse tension, ou après une irradiation aux ions lourds, les propriétés ferroélectriques peuvent être affectées, dégradant ainsi la marge de lecture effective. Des solutions pour améliorer la fiabilité (rétention des données, endurance à basse tension, et résistance aux irradiations) sont donc proposées. Enfin, des capacités ferroélectriques miniaturisées sont intégrées au nœud 28nm sans dégrader les performances des transistors sous-jacents, et se révèlent fonctionnelles. Cette réalisation permet de déterminer les étapes d'intégration optimales qui maximisent les chances d'intégrer avec succès des matrices FeRAM performantes et fiables aux nœuds avancés.

Mots clés: Mémoire ferroélectrique, FeCAP, FeRAM, Oxyde d'hafnium, Caractérisation électrique, Fiabilité

Abstract and keywords

Since their discovery in the early 2010s, the ferroelectric properties of hafnium oxide (HfO_2) deposited in thin films have been the subject of intensive research by both academics and industry. CEA-Leti is particularly interested in ferroelectric HfO_2 for embedded non-volatile memory applications with an ultra-low power consumption. HfO_2-based ferroelectric memories could overcome the scaling limitations encountered by commercial lead-based perovskite (PZT) FeRAMs, in which ferroelectric capacitors are integrated at the 130nm technology node, i.e. in FeRAM arrays with transistors that have a gate length of 130nm, far from the sub-10nm transistors currently being developed. Thanks to their compatibility with CMOS production lines, HfO_2-based ferroelectric capacitors have the potential to be integrated into more aggressive nodes, such as the 28nm or 22nm nodes. If successfully manufactured, they would represent an ultra-low power technology, which could find a prominent place in the embedded memory hierarchy. This is of utmost interest in the era of the data deluge generated by sectors ranging from automotive to consumer electronics, and including 4.0 industry and artificial intelligence.

The objective of this thesis is to integrate high-performance and reliable FeRAM arrays based on ferroelectric HfO_2 into the 28nm node and below, taking advantage of a BEoL 130nm FeRAM technology. To achieve this, a computational study is first conducted to determine the required dimensions of the ferroelectric capacitors to ensure sufficient read margin in FeRAM arrays at advanced nodes, which present limited available voltages and currents. The ferroelectric capacitors are then optimized to enhance their intrinsic ferroelectric properties, and intermediate demonstrators are developed with the aim of integrating high-capacity and high-density FeRAM arrays at the 28nm node. The sources of variability that degrade the read margin at the array level are also studied and corrected when possible. Indeed, the read margin must be sufficient to ensure the functionality of FeRAM arrays under all operating conditions and in all use cases. In particular, after long data storage times, after a certain number of low-voltage program/read cycles, or after heavy ion irradiation, the ferroelectric properties may be affected, thus degrading the effective read margin. Solutions to improve the reliability (data retention, endurance at low voltage, and resistance to irradiations) at FeRAM array level are therefore proposed. Finally, scaled ferroelectric capacitors are integrated at the 28nm node without degrading the performance of the front-end transistors, and are found to be functional. This step allows for the determination of optimal integration steps that maximize the chances of successfully integrating high-performance and reliable FeRAM arrays at advanced nodes.

Keywords: Ferroelectric memories, FeCAP, FeRAM, Hafnium oxide, Electrical characterization, Reliability

Acknowledgements

J'aimerais avant tout remercier mes encadrants Laurent Grenouillet, Jean Coignus et Simon Martin pour leur accompagnement tout au long de cette thèse : votre expertise, votre vision, votre aide et vos encouragements ont été inestimables pour obtenir les résultats présentés et garantir que ces trois années se déroulent sous les meilleurs auspices. Laurent, je suis très heureuse qu'on ait pu partir à SSDM ensemble, merci pour tes conseils avisés sur le Japon et les onsens, pour cette inoubliable séance de karaoké, et pour tous les autres bons moments passés ensemble. Jean, je repartirai sans hésiter avec toi à IRPS. Ça m'a fait très plaisir qu'on découvre les coins sympas de Grapevine, qu'on profite de la piscine du Hilton et qu'on participe à cette soirée poster déjantée à la mode texane : je garde précieusement les photos collectors. Simon, les heures et jours passés ensemble en carac ne remplacent pas une conférence (surtout quand rien ne marche sur l'Electroglas !), mais j'ai beaucoup aimé ces moments-là et te suis très reconnaissante pour ton aide. Cela dit, quand est-ce qu'on part en conf ensemble ?! A tous les trois, nos weeklys du mercredi matin vont me manquer, merci pour tout.

Je tiens également à remercier mon directeur de thèse Marc Bocquet : ton apport scientifique a été déterminant aux étapes-clés de la thèse, et tu as su insuffler et maintenir une dynamique salvatrice lors de la période de rédaction. Merci infiniment pour ça. Par ailleurs, je garde de très bons souvenirs de mon court séjour à Marseille : de la journée des doctorants à la visite de l'IM2NP, vous avez été très accueillants et nos échanges (scientifiques ou non) m'ont beaucoup plu ! J'en profite pour remercier la team de choc marseillaise : Hawa, Sébastien, Noémie, Khaled, Martin, Vincenzo & co !

Vielen Dank auch an das NaMLab-Team für ihren herzlichen Empfang während dieser 'Entdeckungswoche' : Uwe, Ruben, Pramoda, Roberto, Florian, Giulio, Suzanne, Jens und all die anderen! Eure Vorführungen im Reinraum und eure Verteidigungstipps haben mir sehr geholfen.

Merci également à Olivier Billoint, Nicolas Vaxelaire, et Tristan Dewolf pour vos contributions scientifiques inestimables qui ont permis d'obtenir et d'interpréter des résultats importants en simulation, caractérisation structurelle et microscopie électronique lors de cette thèse. Merci à Clément Cheung, mention excellent en Latex, ton aide a été précieuse tout au long de la rédaction. Je tiens particulièrement à remercier Maighread Gallagher-Gambarelli de TWS Editing pour ses ateliers des Vendredis de la Rédaction et son accompagnement vers une rédaction plus sereine et efficace.

Côté labo, ou plutôt labos, merci à toutes les personnes du LDMC et du LCEF que j'ai côtoyées, vous avez contribué à rendre cette thèse plus agréable ! Un merci particulier va à Cathy, Carine, Christelle, Arthur, Gabriele, Sébastien, Mathieu, Clément, Michele, Tifenn pour nos discussions (scientifiques ou non) sur le pas de la porte, dans les couloirs ou à l'extérieur que j'ai beaucoup appréciées. Bien sûr je n'oublie pas non plus les chanteurs et chanteuses du karaoké endiablé au Pachamama, ni les membres de L'"Annexe" (qui se reconnaîtront), pour tous ces joyeux moments passés ensemble ! Merci également à la team mémoire du LCEF, Niccolo, Valentina, Jean, Flavien

(petite pensée émue pour le jazz bar de Kyoto), Liam, et à Alexis sans qui les galères de carac auraient été encore pire… !

Aux anciens thésards, Ahmed, Djo, Filippo, Joel, Léo, Lucas, Théophile, merci pour les super sorties piscine, bivouac, raclette ou apéro, et pour vos sages conseils sur la thèse. Bon vent aux quatre coins de l'Europe ! Aux nouveaux thésards ou à ceux sur le point de finir, Ali, Ancelin, Elise, Hugo, Joaquin, Killian, Mario, Mohamad, Nicolo, Renzo, Simone, Tarcisius : merci pour ce barbecue inoubliable, tous ces midis à H1, et ces soirées au bar. Bon courage pour la fin de la thèse, et à bientôt !

Merci à Anthony et Clément pour votre calme à toute épreuve, vos conseils précieux et nos picnics du midi lorsqu'on rédigeait à la BU : ces moments et nos échanges ont beaucoup compté pour moi. De là à dire que je voudrais retourner rédiger à la BU, il n'y a qu'un pas… que je ne franchirai pas ! Un grand bravo également pour vos soutenances respectives, et à une prochaine pour des pizzas ?!

Un immense merci aux filles du BCC, Amanda, Camille, Justine, Martina : sans votre soutien infaillible, votre amitié et notamment nos goûters au Papapy, cette thèse aurait eu moins de saveur (sans jeu de mots !). En plus de constituer pour moi des modèles dans la vie professionnelle et de véritables sources d'inspiration à côté, vous êtes devenues au fil du temps des amies pour la vie, où que s'en déroulent les prochaines étapes. Partager les bons (et aussi moins bons) moments avec vous a été un honneur et un plaisir, et je me réjouis de ce que l'avenir nous réserve ! A très vite, ici ou ailleurs.

Enfin, je profite de cette opportunité pour remercier ma famille et mes ami.e.s non-cité.e.s plus haut, grenoblois.e.s ou non, pour leur patience, leur confiance, leurs encouragements et leur soutien inconditionnel. Merci à Ali pour toutes tes relectures et ton écoute. Merci à Héloïse, Marion, Anna et Agathe ; Zoé, Lucie R et Lucie G ; Lucile, Hugo, William ; Cléa, Mathilde ; Thomas S et Thomas V ; Agnès, Maureen ; Sibylle, Elo, Hugo & Meli ; Casey ; Maria ; la team iphy sud étendue ; les invité.e.s du weekend JAlie ; Elo & Mos ; Charlotte et toute la chorale des Rainbow Swingers : votre amitié m'est précieuse. Merci à mes parents et mon frère pour votre amour et pour avoir toujours cru en moi. Maintenant que je sors enfin de ce tunnel qu'est la thèse, je me réjouis de vous voir plus souvent et d'être à vos côtés pour les moments importants, petits et grands, qui nous attendent !

Contents

Affidavit 5

List of publications and participation in conferences 7

Résumé et mots clés 9

Abstract and keywords 11

Acknowledgements 13

Contents 15

List of Figures 17

List of Tables 20

Acronyms 21

1 Emergence of Hafnium Oxide Based Ferroelectric Random Access Memories – Context, State-of-the-Art and Objectives 24

1.1 A Favorable Context for Emerging Memories . . . 26
1.1.1 Data Explosion and Current Paradigm . . . 26
1.1.2 Memory Technologies on the Market and their Limitations . . . 27
1.1.3 Emerging Memories and their Advantages . . . 28
1.2 Ferroelectric Memories: From Physics to State-of-the-Art Devices . . . 31
1.2.1 Physical Principle of Ferroelectricity . . . 31
1.2.2 PZT-based FeRAM . . . 32
1.2.3 Advent of HfO_2 as a Promising Ferroelectric Material . . . 33
1.2.4 Tools and Methods . . . 34
1.2.5 State-of-the-Art of HfO_2-based FeRAM Arrays . . . 41
1.3 Objectives of the Thesis . . . 42

2 Computational Analysis for 1T-1C Bitcell Dimensioning towards FeRAM Array Integration at Advanced Technology Nodes 45

2.1 FeCAP Models . . . 47
2.1.1 Polarization-Based Models in the Literature . . . 47
2.1.2 Current-Based Compact Model Developed . . . 54
2.2 FeRAM Array Level Simulation . . . 60

2.3 Extrapolation at Advanced Nodes . . . 63
2.3.1 1T Scaling at the 28nm Node . . . 63
2.3.2 Towards Larger and Denser FeRAM Arrays . . . 66
2.3.3 Recommendation for 1T-1C Bitcell Scaling towards Advanced Nodes . . . 68

3 Optimization of Ferroelectric Stack and Assessment of FeRAM Array-Level Variability towards Advanced Nodes 71
3.1 Remanent Polarization Improvement . . . 73
3.1.1 Impact of Ferroelectric Layer Etching . . . 73
3.1.2 Impact of Oxygen Scavenging Layer . . . 76
3.2 Technological Achievements towards Advanced Nodes . . . 79
3.2.1 Large-Sized Arrays . . . 79
3.2.2 Charge Sensing Reading Scheme . . . 82
3.2.3 3D Capacitors for Enhanced MW at Constant Footprint . . . 84
3.2.4 Lower Ferroelectric Layer Thickness for Low Voltage Operation . . . 88
3.3 Bitcell-to-Bitcell Variability at FeRAM Array Level . . . 90
3.3.1 '0' and '1' States Distributions . . . 91
3.3.2 Intrinsic Variability Sources: Impact and Possible Fixes . . . 94

4 Reliability Study of HfO_2-based Memories, from Ferroelectric Capacitor to FeRAM Array at the 130nm Node 105
4.1 Data Retention . . . 107
4.1.1 Data Retention at Room Temperature . . . 107
4.1.2 Temperature-Accelerated Data Loss . . . 109
4.2 Endurance . . . 121
4.2.1 Performance . . . 122
4.2.2 Recovery Capability . . . 126
4.3 Radiation Hardness . . . 133
4.3.1 Effect of Heavy Ion Irradiation on BEoL-Integrated FeCAPs . . . 133
4.3.2 Effect of Heavy Ion Irradiation on BEoL-Integrated FeRAM arrays . . . 137

5 Scaling of HfO_2-based FeCAPs to the 28nm Node: Process Integration and Performance Assessment 141
5.1 Process . . . 143
5.1.1 Integration Steps . . . 143
5.1.2 Morphological Characterization . . . 146
5.2 Electrical Characterization . . . 148
5.2.1 Single Devices . . . 148
5.2.2 Capacitor Fields . . . 155

Conclusions and perspectives 161

Synthèse en français 163

List of Figures

1.1 Exponential growth of the global datasphere. . . . 26
1.2 Memory hierarchy and benchmark of established memory technologies. . . . 27
1.3 Working principle and benchmark of emerging memories. . . . 29
1.4 Ferroelectricity and principle of a ferroelectric memory. . . . 31
1.5 Demonstration of first 1T-1C PZT-based FeRAM array. . . . 32
1.6 Orthorhombic phase stabilization and crystalline structure of hafnium oxide. . . . 34
1.7 Test bench for ferroelectric performance assessment. . . . 35
1.8 PUND test sequence and associated I-V characteristics. . . . 35
1.9 P-V characteristics, evolution of $2.P_R$ with cycling and $2.P_R$ wafermap. . . . 36
1.10 Detailed steps and schematic for FeCAP BEoL-integration. . . . 38
1.11 Available test structures integrated on 130nm demonstrators. . . . 38
1.12 16kbit 1T-1C FeRAM array design and working principle. . . . 40
1.13 Illustrated state-of-the-art of HfO_2-based FeRAM arrays. . . . 42

2.1 Illustration of the Landau potential for polarization modelling. . . . 48
2.2 Illustration of the Preisach model with hysteron unit and turning point principle. . . . 50
2.3 Current-voltage characteristics evidencing FeCAP switching and back-switching behaviors. . . . 55
2.4 Calibration of the ferroelectric and dielectric contributions in the considered current-based FeCAP compact model. . . . 57
2.5 Validation of the FeCAP compact model at calibration conditions and beyond. . . . 59
2.6 Elementary circuit and associated chronograms for MW simulation. . . . 62
2.7 Validation of the elementary circuit to predict MW degradation at low voltage. . . . 62
2.8 Simulation of MW evolution in a FeRAM array at the 28nm node, impact on the FeCAP dimensions and switching speed projection. . . . 64
2.9 Projection of the MW towards high-performance, larger and denser FeRAM arrays at advanced nodes. . . . 67

3.1 Cross-section STEM imaging and EDX analysis of BEoL-integrated scaled FeCAPs with and without an etched FE layer. . . . 74
3.2 Compared ferroelectric performance and endurance of FeCAPs and FeRAM arrays with and without an etched FE layer. . . . 75
3.3 Structural analysis, defect spectroscopy and electrical characterization for the evaluation of the impact of oxygen vacancies on the FE performances of FeCAPs and FeRAM arrays. . . . 77
3.4 Footprint layout view and bit distributions measured on 16kbit, 64kbit and 256kbit arrays. . . . 80

3.5 Artificial 688kbit array constitution and performance to illustrate the benefit of limited bank sizes . . . 81
3.6 Charge-based sense amplifier (CTIA) principle and benefits for read margin sensing in larger arrays. . . . 83
3.7 Integration schematic, layout view and electrical characterization of FeRAM arrays with standard and shrunk bitcells for higher density. . . . 85
3.8 Interest, integration and literature demonstration of 3D FeCAPs. . . . 86
3.9 ShL integration and performance assessment of 2D+3D FeCAPs. . . . 87
3.10 Nanosecond Laser Anneal principle and performance to crystallize sub-5nm FE films. 89
3.11 Definition of normal distributions from a V_{ref} scan operation. . . . 92
3.12 Characterization and correction of an extrinsic 8×8 pattern. . . . 93
3.13 Measurement of FeCAP area at array level and evaluation of process-induced intrinsic variability. . . . 95
3.14 Monte-Carlo simulation to evaluate the impact of 1T variability on array-level variability. 96
3.15 Bit Line voltage gradient dependence on operation voltage. . . . 97
3.16 Post-layout simulations to determine the origin of the Bit Line voltage gradient. . . . 99
3.17 Impact of the Bit Line voltage gradient on a 16kbit, 64kbit and 256kbit. . . . 100
3.18 Analysis of the correlation between BL voltage gradient and array-level variability. . 101

4.1 Effect of long-term storage at room temperature on programmed pattern and performance at FeRAM array level. . . . 108
4.2 Specific pattern for evaluation of SS, NSS and OS retention states at FeCAP and FeRAM array level. . . . 110
4.3 Assessment of the thermal depolarization effect. . . . 112
4.4 Effect of imprint on the I-V and P-V loops measured during SS and OS retention. . . 113
4.5 Data retention model for OS degradation at FeCAP level. . . . 114
4.6 Data retention model for OS degradation at FeRAM array level. . . . 117
4.7 Design-related degradation of retention at FeRAM array level. . . . 118
4.8 Lasting memory effect of the state baked and impact on the distributions at FeRAM array level. . . . 120
4.9 Electrical recovery to overcome imprint at FeRAM array level. . . . 121
4.10 Comparison of breakdown- and fatigue-limited endurance, introduction of electrical recovery mechanisms. . . . 122
4.11 Study of various endurance strategies to preferentially induce breakdown or fatigue. 124
4.12 Demonstration of effective recovery at FeCAP level and compared performances of various strategies. . . . 127
4.13 Fatigue electrical signature on scaled BEoL-integrated FeCAPs. . . . 128
4.14 Peak splitting analysis during fatigue and recovery. . . . 130
4.15 First-Order Reversal Curve (FORC) measurement procedure. . . . 130
4.16 FORC switching density maps and associated current peaks for unipolar and bipolar strategies. . . . 131
4.17 Demonstration of an efficient recovery strategy at FeRAM array level. . . . 132
4.18 Evaluation of the effect of various heavy ion fluences on FeCAP performance and reliability. . . . 135

4.19 GIXRD analysis of samples irradiated at various heavy ion fluences. 136
4.20 Effect of moderate fluence irradiation on programmed pattern and performance at FeRAM array level. 138

5.1 Principle of a Fully-Depleted Semiconductor on Insulator (FDSOI) transistor. 144
5.2 Process steps for the integration of a FeCAP in the BEoL of 28FDSOI technology. . . 145
5.3 STEM cross-section of a FeCAP integrated in the BEoL of 28FDSOI process and in-line morphological analysis. 147
5.4 Electrical characterization of scaled elementary devices integrated on 300mm wafers for functionality assessment. 149
5.5 Process variants on scaled 1C for PoR determination. 151
5.6 Parametric testing results verifying the integrity of FEoL 28FDSOI transistors after FeCAP BEoL-integration. 153
5.7 Layout view and functionality demonstration of 1T-1C bitcells at the 28FDSOI node. 154
5.8 Electrical characterization of capacitors fields integrated on 300mm wafers for PoR determination. 157

List of Tables

2.1 Fitting parameters values and capacitor dimensions for FeCAP compact model calibration 58

4.1 Fitting parameters of ΔOS Margin and Time-to-Fail 116

Acronyms

1C Single capacitor
1T Single transistor
1T-1C One transistor - one capacitor bitcell
2.P_R Remanent polarization
ALD Atomic Layer Deposition
AR Aspect Ratio
BE Bottom Electrode
BEoL Back-End-of-Line
BER Bit Error Rate
BL Bit Line
CAGR Compound Annual Growth Rate
C_{BL} Bit Line capacitance
CCIC Cross-Coupled Invertor Comparator
CD Critical Dimension
C_d Dielectric capacitance (C_{DE} is also used)
C_{FB} Feedback capacitance
C_{GS} Gate-Source capacitance
CMOS Complementary Metal-Oxide Semiconductor
CMP Chemical Mechanical Polishing
CPU Control Processing Unit
CTIA Capacitive Trans-Impedance Amplifier
CVD Chemical Vapor Deposition
DRAM Dynamic Random Access Memory
DUT Device Under Test
DUV Deep Ultra-Violet (UV)
EDX Energy-Dispersive X-ray spectroscopy
EoP End-of-Process
FBB Forward Body Biasing
FDSOI Fully-Depleted Semiconductor On Insulator
FE Ferroelectric
FeCAP Ferroelectric Capacitor
FeFET Ferroelectric Field Effect Transistor
FEoL Front-End-of-Line
FeRAM Ferroelectric Random Access Memory
FG Floating Gate
FIB Focused Ion Beam
FORC First-Order Reversal Curve

FTJ Ferroelectric Tunnel Junction
GIXRD Grazing Incidence X-Ray Diffraction
GO1 Thin gate oxide transistor (SG)
GO2 Thick gate oxide transistor (EG)
HAADF High Angle Annular Dark Field (STEM mode)
HAXPES Hard X-Ray Photoelectron Spectroscopy
HDD Hard Drive Disk
HfO_2 Hafnium dioxide (always called hafnium oxide)
HKMG High-K Metal Gate
HSO Silicon-doped hafnium oxide (Si:HfO_2)
HZO Alloy of hafnium oxide and zirconium oxide ($Hf_{0.5}Zr_{0.5}O_2$)
ICT Information and Communication Technologies
IDM Integrated Devices Manufacturer
I_{off} Off-current of the transistor
I_{on} On-current of the transistor
IoT Internet of Things
L Gate Length of the transistor
MAD Memory Advanced Demonstrator
MCU Microcontroller Unit
MFM Metal Ferroelectric Metal
MLC Multi-Level Capability
MoL Middle-of-Line
MRAM Magnetic Random Access Memory
MW Memory Window
MW(0σ) Median Memory Window
MW(16k) Application-oriented Memory Window at a 16kbit array scale
NAND NOT AND boolean operator
nC Capacitor field made of n single capacitors
NLA Nanosecond Laser Anneal
NLS Nucleation Limited Switching
NOR NOT OR boolean operator
NSS New Same State
NVDRAM Non-Volatile Dynamic Random Access Memory
NVM Non-Volatile Memory
NVSRAM Non-Volatile Static Random Access Memory
o- Orthorhombic (crystalline phase)
OS Opposite State
OxRAM Oxide-based Random Access Memory
PCRAM Phase-Change Random Access Memory (also called PCM)
PL Plate Line
PoR Process of Reference
PTMEM Lithography level "Point mémoire"
PUND Positive Up Negative Down
PVD Physical Vapor Deposition

PZT Lead zirconate titanate ($Pb[Zr_xTi_{1-x}]O_3$ $(0 \leq x \leq 1)$)
RAM Random Access Memory
RRAM Resistive Random Access Memory (also called ReRAM)
RT Room Temperature
SA Sense Amplifier
SCM Storage Class Memory
ShL Short Loop
SL Source Line
SoC System-on-Chip
SS Same State
SSD Solid State Drive
STEM Scanning Transmission Electron Microscopy
t- Tetragonal (crystalline phase)
TE Top Electrode
TEM Transmission Electron Microscope
TiN Titanium Nitride
TP Turning Point
TTF Time To Failure
V_{th} Threshold Voltage of the transistor
W Gate Width of the transistor
WB Write-Back
WL Word Line

1 Emergence of Hafnium Oxide Based Ferroelectric Random Access Memories – Context, State-of-the-Art and Objectives

Introduction

In our hyper-connected world, the industry of semiconductors is expected to be valued over 600 billion dollars worldwide in 2024 [1]. The integrated circuits constitute over 80% of this forecast value, and include sectors like stand-alone memories, logic circuits, microcontrollers and analog circuits. The first two present the highest growth trends (77% and 11% respectively), while the microcontroller sector encompasses embedded memories and is expected to be worth nearly 80 billion dollars in 2024 [1]. The considered components feed the needs for microelectronic chips in the consumer, industrial and automotive branches [2]. In addition to being a profitable industry, or perhaps because it is, the semiconductors industry is a very competitive sector for industrial groups, and a strategic investment sector at the national scale [2, 3, 4]. Indeed, the recent Covid crisis revealed the dependence of the European semiconductors industry on South Asian Integrated Device Manufacturers (IDMs) [5, 1], which present a growing environmental and geopolitical vulnerability [6, 7, 8]. In this context, Europe is joining forces with the European Chips Act [9], which aims at reinforcing the sovereignty of the European semiconductors ecosystem. Within this framework, a large-scale pilot line project (FAMES) has been launched by the European Commission to develop energy-saving chips for greener electronics [10, 11]. A particular focus of this project is the development of embedded non-volatile memories, including the integration of Ferroelectric Random Access Memories (FeRAM) at advanced technology nodes.

This chapter aims at giving a specific description of the memory sector in the big data era, and at presenting how emerging memories could solve some of the limitations of the established memories. Especially, the mechanism responsible for ferroelectricity makes ferroelectric memories particularly promising in terms of low power consumption and high switching speed, an asset to overcome the data deluge with a high energy-efficiency. While PZT-based FeRAM has reached maturity [12], occupies a niche sector of the market and faces challenges, the advent of ferroelectric hafnium oxide has opened a new way for scalable CMOS-compatible ferroelectric memories.

Summary

1.1 A Favorable Context for Emerging Memories . . . 26
1.1.1 Data Explosion and Current Paradigm . . . 26
1.1.2 Memory Technologies on the Market and their Limitations . . . 27
1.1.3 Emerging Memories and their Advantages . . . 28
1.2 Ferroelectric Memories: From Physics to State-of-the-Art Devices . . . 31
1.2.1 Physical Principle of Ferroelectricity . . . 31
1.2.2 PZT-based FeRAM . . . 32
1.2.3 Advent of HfO_2 as a Promising Ferroelectric Material . . . 33
1.2.4 Tools and Methods . . . 34
1.2.4.1 Large Capacitors for Ferroelectric Stack Optimization . . . 34
1.2.4.2 Scaled Devices Integrated in BEoL CMOS . . . 37
1.2.4.3 1T-1C FeRAM Arrays for Statistical Evaluation . . . 39
1.2.5 State-of-the-Art of HfO_2-based FeRAM Arrays . . . 41
1.3 Objectives of the Thesis . . . 42

1.1 A Favorable Context for Emerging Memories

In this part, a general overview of the memory technology sector in the specific context of the big data era is given. The memory hierarchy is presented to highlight the need for multiple technologies, whether stand-alone or embedded, to compute and store information from the Central Processing Unit (CPU) to permanent storage. While the mainstream memories and their manufacturers dominate the market, some scaling limitations arise beyond the 28nm process node. Moreover, new environmental constraints, together with the data deluge and associated rising energy consumption [13, 14] feed the need for a change of paradigm that could benefit to emerging memories.

1.1.1 Data Explosion and Current Paradigm

In the current era of big data, there is an exponential increase in demand for compute [15]. While the mobile and cloud era that started in the 2000s enabled the worldwide connection of individuals [16], it also launched a Fourth Industrial Revolution that resulted in an unprecedented increase in the number of connected devices and an explosion of the global datasphere [3, 17]. This effect is characterized by the so-called three V's: volume, variety and velocity [18].

- Today, over 150 zettabytes (1.5×10^{23} bytes) of data are created and replicated globally each year and the forecast trend is evolving exponentially, as shown in Figure 1.1.a. from [17].
- In addition to the data generated by the transport, smart industry and smart city sectors, individuals increasingly generate high-resolution photos and videos for example [3].
- The flow of data transfer is also increasing exponentially, especially driven by on-demand video [3].

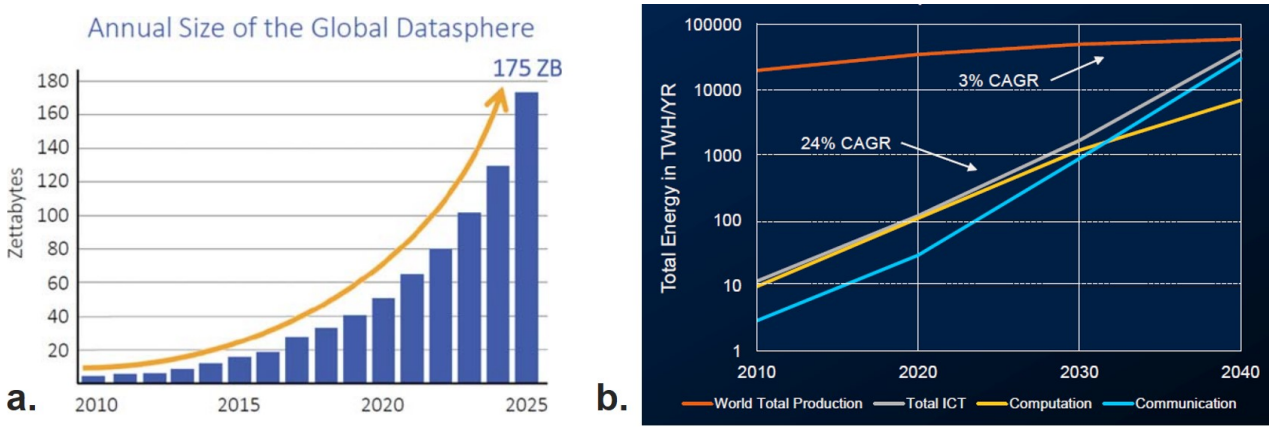
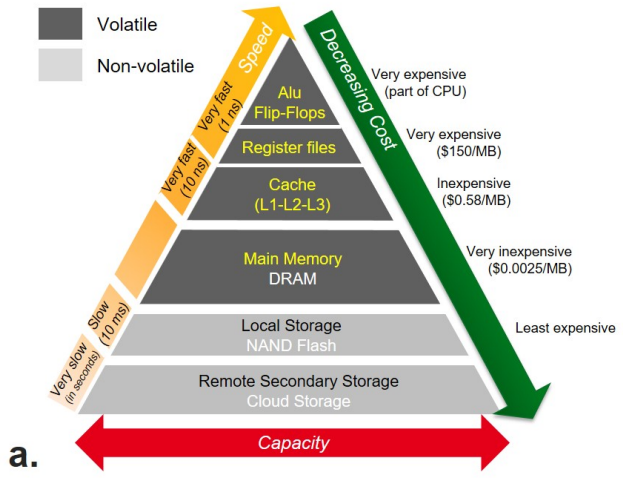


Figure 1.1: a. Exponential growth of the global datasphere from [17] ; b. Associated ICT energy demand from [16, 19], where CAGR is the Compound Annual Growth Rate.

Together with this data deluge, the energy consumption of Information and Communication Technologies (ICT) is rising exponentially, a significant portion of which is being dissipated by Joule heating. Moreover, as shown in Figure 1.1.b. from [16], the combined computation and communication energies associated progress toward the global energy production according to projections from [19]. Therefore, the need for high-performance, energy-efficient and scalable memory technologies is stronger than ever.

1.1.2 Memory Technologies on the Market and their Limitations

In the memory sector, several technologies co-exist today to address different requirements: as shown in Figure 1.2.a., the memory pyramid aims at classifying the existing computing systems depending on their speed, cost and capacity. From top to bottom, the memory capacity (*i.e.* the amount of storable data) increases as the speed (proportional to latency) and fabrication cost decrease. Traditional memory architectures are organized with the Central Processing Unit (CPU) at the top for computing, followed by SRAM, DRAM, Flash and Hard Drive Disk (HDD) or Magnetic Tape at the bottom. There is therefore a tradeoff between cost and performance, and no 'universal' memory emerges. The aforementioned memory technologies are either volatile (*i.e.* the stored information is lost when the power is shut down) or non-volatile, and can be divided into embedded and stand-alone groups: in the former, on-chip memory integrated into CPUs, microcontrollers (MCUs) or Systems-on-Chips (SoC) are particularly relevant when targeting the market of consumer electronics, automotive, Internet of Things (IoT) or wearable devices [20, 21]; the latter is constituted of discrete chips dedicated to a memory / storage function and are intended for data centers and servers, or cloud computing [20, 18].



b.

	eSRAM	eDRAM	FG Flash
Mechanism	Cross-coupled inverter+charge	Charge on capacitor	Charge on FG
Cell Structure	6T	1T1C	1.5T
Cell Size	120-150 F^2	40 F^2	50 F^2
MLC	No	No	Yes
R_{on}/R_{off} ratio	N/A	N/A	$>10^4$
Integration Node	7nm FinFET	22nm FinFET	40nm
Additional Masks	0	5+	13+
Energy/bit	~1 fJ	~1 pJ	100 pJ
Latency	<1 ns	>10 ns	0.1-1 ms
Endurance	10^{16}	10^{16}	10^4-10^5
Retention	volatile	Refresh	10 yrs

Figure 1.2: a. Memory pyramid describing the hierarchy from CPU to memory to storage ; b. Comparative benchmark of embedded SRAM, embedded DRAM and Floating Gate Flash memory technologies from [22, 23].

This current paradigm presents some limitations and it is widely accepted that two bottlenecks arise: the first one is the latency gap between DRAM and NAND Flash leading to high energy consumption and long latency to transfer data across the system. Within this gap, a new tier called Storage Class Memory (SCM) has emerged and could be addressed with emerging memories [24, 18, 25]. Secondly, higher in the memory pyramid, a so-called memory wall appears between the high-performing CPU and the memory that has not kept pace with its performance and speed growth, creating the need for innovative computing paradigms [26, 18].

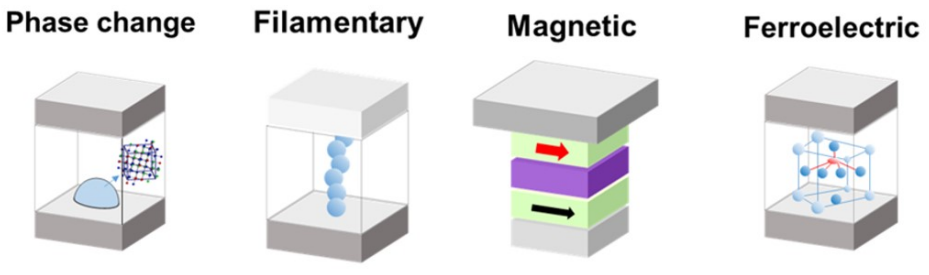
Moreover, the established memories like DRAM or Flash are faced with challenges in terms of scaling and energy consumption. Indeed, according to Yole Group, the lack of a cost-competitive embedded Flash solution at geometries smaller than the 28nm node is the main driver of top foundries and IDMs for active investment in embedded emerging Non-Volatile Memory (NVM) technologies to be integrated at more aggressive nodes [2]. What's more, as can be seen in Figure 1.2.b. from [22], the need for refresh operations in DRAM and the high energy per bit of Floating Gate (FG) Flash are concerning with regards to the growing energy consumption of ICTs. Therefore, embedded emerging NVM technologies such as MRAM, PCM, RRAM, and recently FeRAM, continue to gain momentum especially for code and data storage in microcontrollers and low-power devices [2].

1.1.3 Emerging Memories and their Advantages

Contrary to the very concentrated market of established memories, the variety of emerging memories is reflected in the number of different technologies investigated and their IDM contributors. According to TechInsights [27], emerging memories are likely to constitute interesting non-volatile memory devices for embedded applications rather than emerge as stand alone memory devices. Figure 1.3.a. from [18] presents the schematics of several unit cells that aim at describing the mechanisms driving the storage of information in various emerging memory technologies.

- In Phase-Change Memories (PCM or PCRAM), two levels of resistance encode the '0' and '1' logic bits and physically stem from an amorphous (insulating) chalcogenide alloy that can be heated up by Joule effect until crystallization (conductive), thus modulating the electrical conductivity [28, 29, 30].
- Filamentary memories (encompassing OxRAM) rely on two levels of resistance brought by the reversible formation and disruption of a conductive filament between top and bottom electrode [31, 32, 33]. They are most generally known as Resistive Memories (RRAM or ReRAM).
- Magneto-resistive memories (MRAM) are based on the magnetism of electron spins in Magnetic Tunnel Junctions that exhibit a high / low level of resistance when the magnetization of the fixed layer and the free layer are anti-parallel / parallel, respectively [34, 35, 36].
- Contrary to the aforementioned memories, ferroelectric memories encode the binary information in the meta-stable position of ions within a polar crystalline material, which is then evaluated through the displacement current rather than the resistance itself [37, 38, 39].

a.



b.

Current Integration & developments focus @CEA-Leti

	NOR FLASH	MRAM	PCRAM	ReRAM (OxRAM/CBRAM)	FeRAM (PZT)	FeRAM (HfO_2)
Programming power	~200pJ/bit	~20pJ/bit	~300pJ/bit	~100pJ/bit	**~10fJ/bit**	**~10fJ/bit**
Write speed	20 µs	**20 ns**	**10-100 ns**	**10-100 ns**	<100ns	14ns @ 2.5V
Endurance	10^5 - 10^6	**10^6-10^{15}**	10^8	10^5 - 10^6	**> 10^{15}**	**> 10^{11}**
Retention	> 125°C	85°C - 165 °C	**165°C**	> 150°C	125°C	125°C
Extra masks	Very high (>10)	Limited (3-5)	Limited (3-5)	**Low (2)**	**Low (2)**	**Low (2)**
Process flow	Complex	Medium	Medium	**Simple**	**Simple**	**Simple**
Multi-Level Cell	Yes	No	**Yes**	**Yes**	No	No
Scalability	Bad	Medium	**High**	**High**	Medium	**Poor (2D) High (3D)**

c.

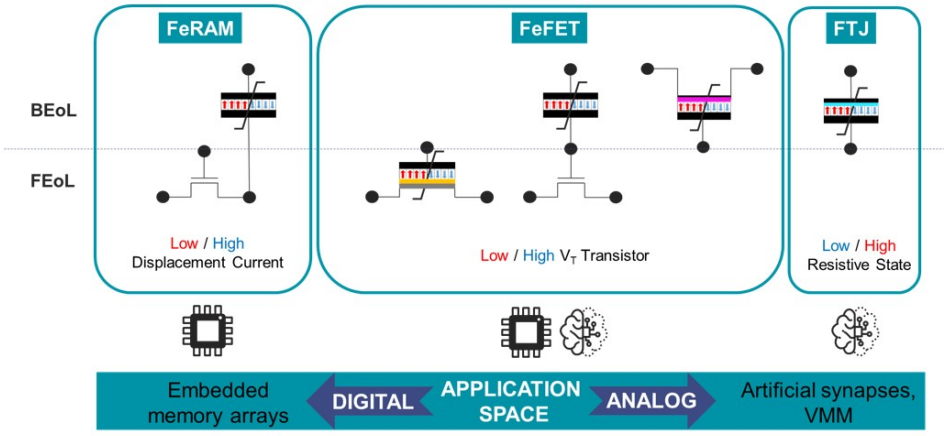


Figure 1.3: a. Schematic of PCM, RRAM, MRAM and FeRAM emerging memories and their working principle from [18] ; b. Benchmark of emerging memories for embedded NVM applications ; c. Schematic of the ferroelectric memories zoology and related application space

Figure 1.3.b. is a benchmark of these emerging memories, and compares the main key metrics of interest in terms of performance (such as programming energy, write speed), integration process (masks, flow) and specific features to target artificial intelligence (Multi-Level Capability, MLC) or aggressive process nodes (scalability). This table also compares reliability metrics such as retention and endurance, that quantify the temporal and thermal conditions in which stored information can be safely retrieved, and the number of write and read cycles that can be performed until both logic states cannot be discriminated, respectively [40]. Desired performances and reliability characteristics are highly dependent on the application targeted, and the level of maturity of these emerging memories differs. PCM products have been commercialized by Intel and Micron under the 3D XPoint brand [41, 42], but their production stopped in 2022 due to their lack of cost-effectiveness [43, 44]. PCM are still explored now mainly by STMicroelectronics (with Samsung foundry) [45, 46], and are particularly well-suited to automotive applications thanks to their good retention at 165°C [2]. RRAM are fast and easy to manufacture (hence cost-effective), and therefore attract a lot of interest, in particular from Infineon: recently, Infineon reported the adoption of the 28nm RRAM from TSMC foundry for automotive MCU applications [47]. Spin-transfer torque MRAM, thanks to its high endurance, is and has been the focus of numerous foundries [48, 49, 50] for firms such as NXP, Avalanche, Everspin [51] and IBM. However, one possible aspect limiting the widespread adoption of MRAM is the need for numerous rare materials [52] that make the integration process complex and costly.

Regarding ferroelectric memories, various device structures exist and address different application spaces from digital to analog, as shown in Figure 1.3.c. Perovskite-based FeRAM is a mature and commercialized emerging technology (see Paragraph 1.2.2), whereas hafnium oxide based FeRAM is currently under active research and development (see Paragraph 1.2.3). The structure of a FeRAM array resembles that of a DRAM, in that a scaled capacitor is connected to the drain of a selection (also called access) transistor. In the case of FeRAM, the oxide layer of the capacitor is also ferroelectric, thus forming a ferroelectric capacitor (FeCAP) and enabling non-volatility thanks to the remanent polarization properties. A realistic application option for FeRAM is to target digital embedded memory arrays. As can be seen on Figure 1.3.c. on the other end of the spectrum, Ferroelectric Tunnel Junctions (FTJs) rely on tunnelling currents in a non-ferroelectric oxide layer sandwiched between a ferroelectric layer and a metallic electrode, which can therefore provide an analog output well-suited to neuromorphic applications. In-between, a ferroelectric layer connected to the gate of a transistor constitutes a Ferroelectric Field-Effect Transistor (FeFET), whose memory states are encoded in the threshold voltage of the transistor.

Recap

Emerging memories are being developed for embedded NVM applications to specifically target aggressive nodes and low-power requirements that established memory technologies struggle to tackle. In this view, ferroelectric memories (and more specifically 1T-1C FeRAM arrays) constitute a promising technology whose potential for application needs to be further assessed.

1.2 Ferroelectric Memories: From Physics to State-of-the-Art Devices

1.2.1 Physical Principle of Ferroelectricity

Ferroelectrics belong to a subset of crystalline materials that always exhibit piezoelectric and pyroelectric properties, and present a polar non-centrosymmetric structure [53]. They are non-linear dielectric materials that have a high polarizability and therefore possess a spontaneous polarization. In a ferroelectric, the polarization is not proportional to the electric field but rather follows a hysteresis loop versus the electric field as shown in Figure 1.4.a.: in P-E loops, the metrics of interest are the coercive field, the polarization at saturation (*i.e.* at very high electric field) and the remanent polarization (*i.e.* at null electric field). Additionally to the electronic polarization present in every material, ionic and orientation contributions to the polarization exist in ferroelectric materials [54]. Two stable polarization states originate from the double-well potential of ferroelectrics, which distorts under the application of a non-zero electric field, thus allowing to switch states if the coercive field is reached. One of the first ferroelectric material to be discovered was Rochelle salt by Valasek in 1921 [55]. In analogy to ferromagnetism [56], the fundamental physics of ferroelectricity was presented, and the hysteresis loop of the material is presented in Figure 1.4.a.

Some time afterwards, Buck proposed the concept of ferroelectric memories in 1952 [37]. A crossbar array structure made of barium titanate ($BaTiO_3$) ferroelectric capacitors was fabricated. Indeed, the aforementioned spontaneous and switchable polarization allows for ferroelectric materials to be used as memory devices. A binary information ('0' and '1' logic states) can be encoded in the material's remanent polarization by applying a positive or negative voltage to a ferroelectric capacitor. The reading operation consists in applying a positive pulse to the capacitor, which has the effect of switching the ferroelectric state in the case of a previously programmed '1', thus destroying the stored information in the process. The reading operation is therefore destructive. Both logic states can be discriminated depending on whether or not they generate ferroelectric switching when read. As proposed by Buck [37], the transient current can be measured at the terminals of the capacitor to detect a '0' or a '1', as shown in Figure 1.4.b.

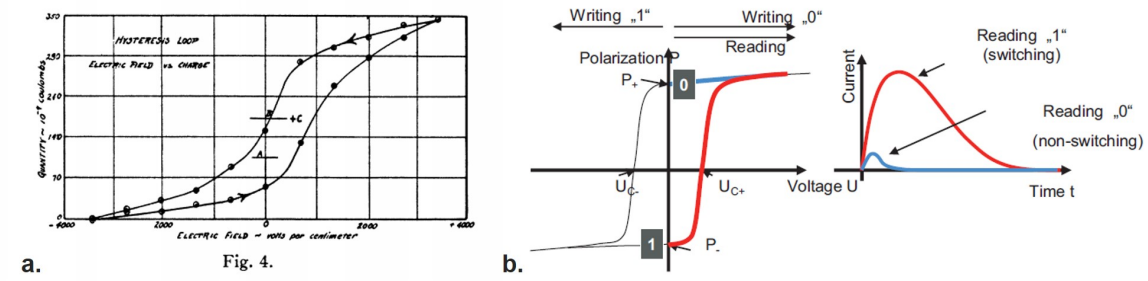


Figure 1.4: a. Rochelle salt polarization versus electric field hysteresis loop from [55] ; b. Principle of a ferroelectric memory (writing and reading) based on the hysteresis loop of a ferroelectric capacitor from [54].

1.2.2 PZT-based FeRAM

Several decades after the invention of ferroelectric memories and the demonstration of $BaTiO_3$-based crossbar arrays, a Ferroelectric (called ferroelectronic at the time) Random Access Memory FRAM was launched in 1990. Based on lead zirconate titanate (PZT) ferroelectric capacitors integrated in the Back-End-of-Line (BEoL) of a CMOS process at the 1.5µm node, as shown in Figure 1.5.a., it adopted a 1T-1C structure (see Figure 1.5.b.) rather than the crossbar structure preferred beforehand. Indeed, in this DRAM-like configuration, where a selection transistor acts like a switch to select only a bitcell of choice, the previously critical issue of write disturb of unselected cells was solved. Moreover, the choice of PZT material, based on its high thermal stability, low coercive voltage and high remanent polarization (10-20 $\mu C/cm^2$) yielded good retention and endurance reliability characteristics. Several FRAM array sizes were hence demonstrated, and a top view of the 4kbit FRAM is shown in Figure 1.5.c. These products were commercialized by Ramtron and were considered the world's first integrated non-volatile static RAM.

Since then, PZT-based FeRAM was further developed and commercialized by Infineon, Texas Instruments, Lapis (Rohm), SKHynix and Fujitsu down to the 180nm and 130nm process nodes. Their performances were optimized thanks to material breakthroughs and design innovations, which lead to the attractive features of low energy consumption (< 10fJ/bit), low latency (< 100ns), low voltage (< 3V) and high endurance (up to 10^{15} cycles). In fact, it is has been a commercial success for Fujitsu, which sold over 4 billion unit since 1999, more than any other emerging technology.

However, PZT is not CMOS-compatible as lead is a contamination source, and is not scalable below the 130nm node, limiting PZT-based FeRAM to niche applications and relaxed nodes. Considering the ongoing extreme scaling of the transistor combined with the need for emerging memories for embedded NVM applications at aggressive nodes, PZT-based FeRAM is not an ideal candidate.

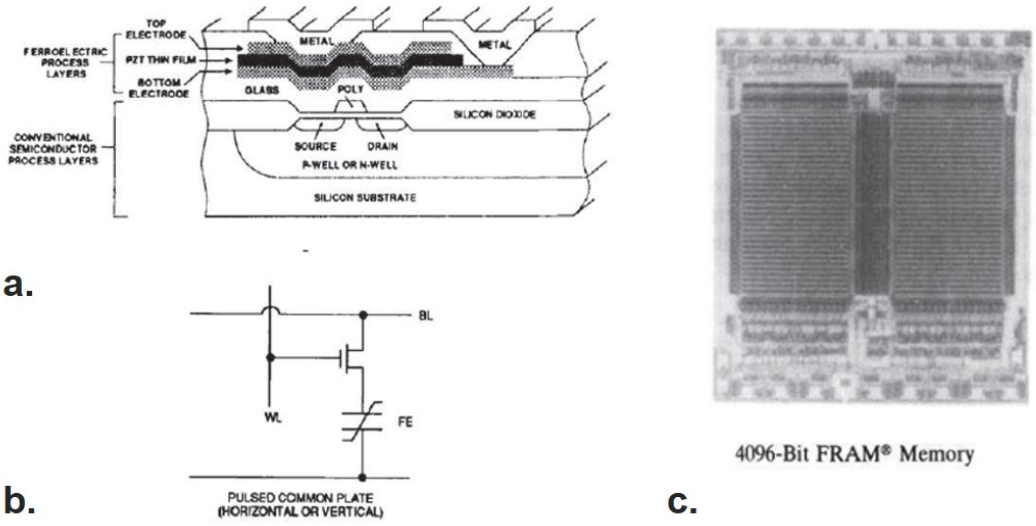


Figure 1.5: From [38]: a. Cross-section schematic of the integration of a BEoL 1.5µm PZT-based FeCAP ; b. Electrical configuration of a 1T-1C bitcell ; c. Top-view of a 4kbit FRAM array by Ramtron

1.2.3 Advent of HfO_2 as a Promising Ferroelectric Material

For these reasons, the discovery of ferroelectric properties in hafnium oxide, a well-known material in the semiconductor industry as a high-k dielectric for High-K Metal Gate (HKMG) scaled transistors, has sparked renewed interest and a new wave of research and development in the field of ferroelectric memories. Contrary to perovskite-based ferroelectrics (such as $BaTiO_3$ or PZT), hafnium oxide has a fluorite structure and does not exhibit ferroelectric properties in bulk configuration. In addition, very specific fabrication conditions are required in order to generate ferroelectricity in hafnium oxide based thin films, which partly explains the late discovery of hafnium oxide's potential for NVM applications. Indeed, as reported in [57], the authors fabricated thin films based on hafnium oxide that exhibited ferroelectric features when doped at an optimal silicon content of 3.1 mol%, and when annealed at a sufficient temperature in the presence of a top electrode (which is known as capped crystallization, or clamping effect). The origin for ferroelectricity in hafnium oxide was identified by structural analysis to be linked to the presence of a meta-stable non-centrosymmetric crystalline phase, namely the orthorhombic phase (o-phase) $Pca2_1$ [58]. In this phase, the movement of oxygen atoms along the c-axis allows for a spontaneous polarization to occur. Note that a ferroelectric rhombohedral phase has also been evidenced in epitaxially grown $Hf_{0.5}Zr_{0.5}O_2$ [59], but will not be studied here. Figure 1.6 describes the schematic phase diagram of hafnium oxide and its dependence on the annealing temperature and dopant (mainly silicon or zirconium) content on the left-hand side, and the two preferential positions of oxygen ions in the orthorhombic crystalline structure on the right-hand side. These positions determine two saturated polarization states that encode the information as binary bits.

The physical origin of the formation and stabilization of the meta-stable FE o-phase is still under investigation [60]. Indeed, according to the bulk phase diagrams of hafnium and zirconium oxides, the polar orthorhombic phase cannot be stabilized under reasonable pressure and temperature conditions. In addition to that, several studies have reported that the free energy difference between the meta-stable FE o-phase and the stable monoclinic phase (m-phase) is too large to be overcome by doping alone [61, 62]. The role of surface energy [63], or a combined effect of doping, surface energy, strain, and electric field [61] have been thoroughly investigated as possible root causes for formation and stabilization of the meta-stable FE o-phase. As of today, no consensus emerges and this remains an open research field.

Apart from theoretical and fundamental work into the origin of ferroelectricity in doped hafnium oxide thin films, several studies have pointed out several features that need to be taken into consideration when assessing the potential of hafnium oxide based material for FeRAM integration. First of all, the coercive field of doped hafnium oxide is quite high compared to PZT-based films, which leads to a higher operating voltage and an increased susceptibility to dielectric breakdown, thus limiting the endurance. Moreover, similarly to PZT-based ferroelectric capacitors, the remanent polarization is not stable throughout the life of the device, as a wake-up operation is initially needed to boost $2.P_R$ to its maximum level, and a ferroelectric fatigue can occur after an elevated number of write / read cycles. The origin of wake-up is hypothesized to be phase transition, defect-induced phase transition, or charge de-trapping or redistribution. On the other hand, fatigue is thought to occur because of domain pinning or charges in-plane inhomogeneity.

Due to its high speed, potential low power consumption and low cost, HfO_2-based FeRAM is currently researched in various applications from memory in MCU for IoT devices up to DRAM

replacement or slow SRAM replacement. Currently, however, the endurance needs to be improved, operating voltages must be reduced, and the scalability towards more aggressive nodes must be demonstrated.

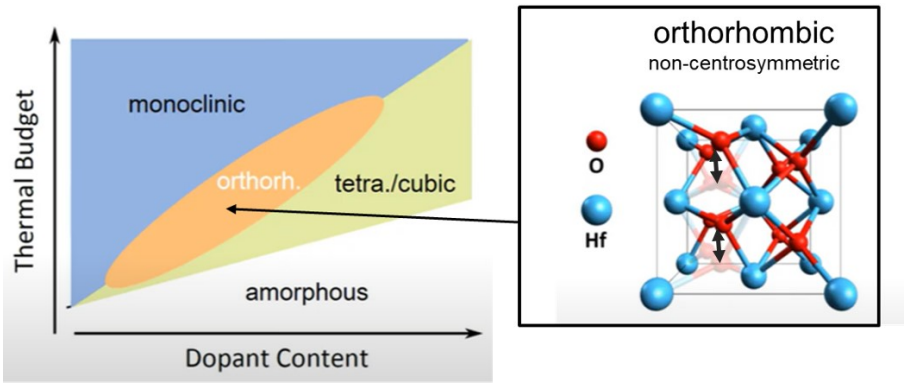


Figure 1.6: a. Schematic phase diagram featuring the monoclinic, orthorhombic and cubic/tetragonal phases as a function of annealing temperature and dopant content, as well as crystalline structure of the non-centrosymmetric orthorhombic phase in hafnium oxide from [64]

1.2.4 Tools and Methods

In order to develop HfO_2-based ferroelectric memories, several intermediary steps are necessary: optimizing the ferroelectric performances of the material through ferroelectric stack iterations, demonstrating the integrity of the ferroelectric performances of scaled capacitors integrated in the BEoL of a CMOS technology, and determining a preferred process of reference. Finally, designing and demonstrating the functionality of a FeRAM array advanced demonstrator is needed from a statistical and circuit design standpoint. Those steps were initiated in previous works at CEA-Leti and this thesis benefits from the thus-developed tools. The following parts will detail the integration steps and electrical characterization procedures of various test structures which are relevant to meet the aforementioned goals.

1.2.4.1 Large Capacitors for Ferroelectric Stack Optimization

In this section, the devices of interest are 100µm diameter capacitors manufactured on 300mm wafers with a very limited number of fabrication steps and only one lithography level. This fabrication process is called Short Loop (ShL) as it is much quicker and less costly than that for more advanced demonstrators, especially those with FEoL transistors and BEoL-integrated scaled capacitors. Its main utility is to provide a quick feedback regarding the influence of some process steps variations (*e.g.* ferroelectric layer composition, thickness ; electrode materials, deposition technique ; interface engineering) on the ferroelectric performances of the large capacitors. Note that the conclusions drawn from electrical results obtained on ShL wafers can mainly apply to

scaled integrated capacitors. However, a major difference in terms of mechanical stress is observed between large and scaled capacitors due to the differing active areas, as will be illustrated in Chapter 3. Moreover, one also has to keep in mind that material discrepancies can occur between 300mm ShL and 200mm BEoL-integrated FeCAPs as clean room fabrication tools may differ for 200mm and 300mm wafers.

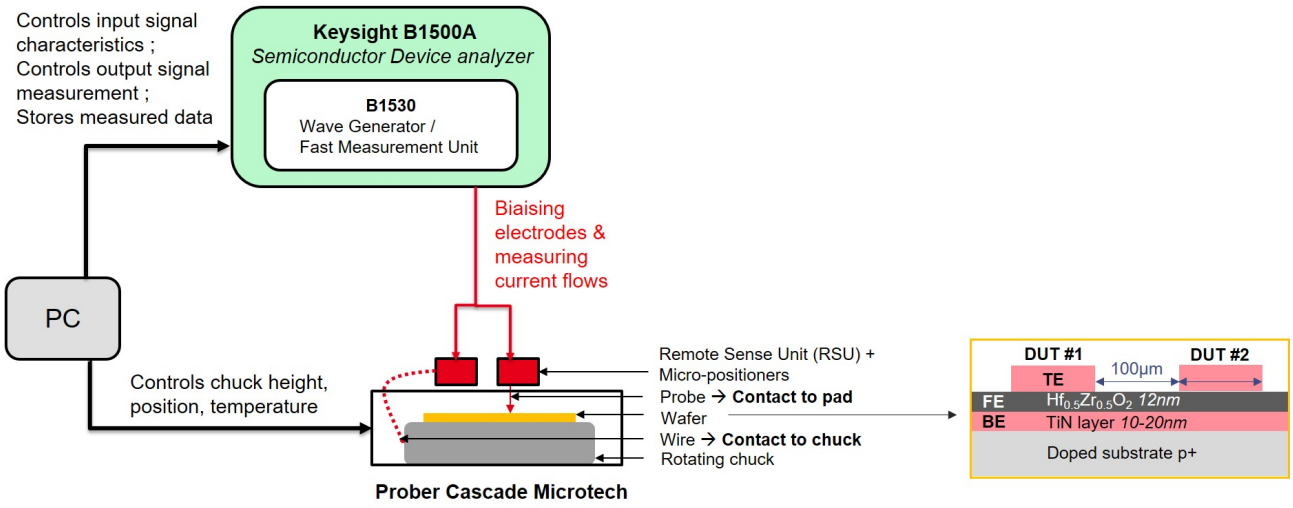


Figure 1.7: Schematic of the testing setup for the ferroelectric performance assessment of 100µm diameter capacitors manufactured on 300mm Short Loop wafers.

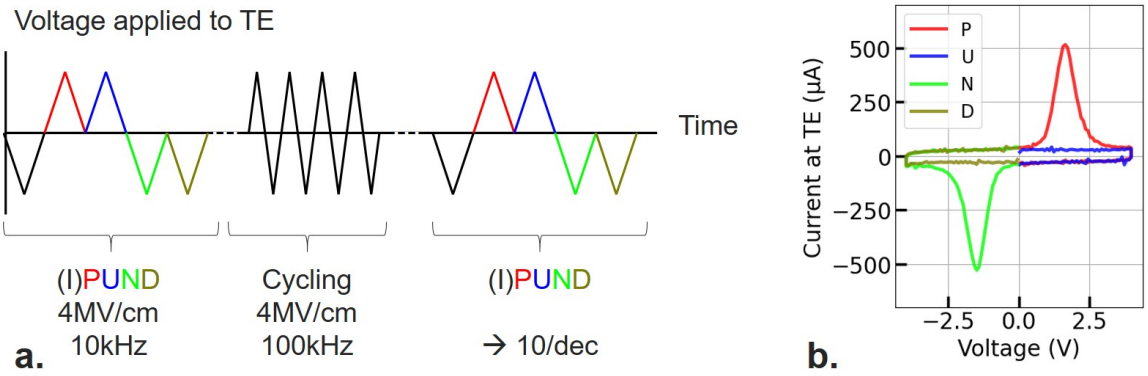


Figure 1.8: a. Positive Up Negative Down (PUND) sequence applied to a ferroelectric capacitor between cycling occurrences to measure switching and electric currents for the extraction of remanent polarization and coercive voltage. ; b. I-V characteristics measured on a 100µm diameter capacitor for the P, U, N and D voltage pulses, here for a TiN/HZO (12nm)/TiN ferroelectric stack.

As the aforementioned capacitors are two-terminal devices, only the top and bottom electrodes need to be contacted for the evaluation of the ferroelectric performances. Figure 1.7 details the test bench used for characterizing ShL devices: a schematic of the semi-automated probe station and associated analyzer is shown, as well as a drawing of a ShL wafer with a TiN/HZO (ALD-deposited, 12nm, 50% HfO_2 - 50 %ZrO_2)/TiN ferroelectric stack. In this specific case, the top electrode is biased using the tip of a probe, and the current flow is measured at the bottom electrode through a wire connected to the chuck. Otherwise (*i.e.* for integrated two-terminal devices with no available back-face contact), the same test bench is used with a second probe connected to the bottom electrode pad on the top face of the wafer.

Figure 1.8.a. presents the default voltage sequence applied to a ferroelectric capacitor in order to obtain a hysteresis loop and extract the remanent polarization and coercive voltages. The principle of this characterization is independent of the Device Under Test (DUT), whether it be a large capacitor from a ShL wafer or a BEoL-integrated scaled capacitor. As explained in [65], the Positive Up Negative Down (PUND) sequence is applied, and consists in alternatively pulsing positively the top electrode twice (P and U), and then negatively pulsing the top electrode twice (N and D). A negative initialization pulse was previously applied to pre-pole the dipoles in the P_{up} polarization state. The P and N pulses generate a switching current in the ferroelectric capacitor, as well as a dielectric current and possibly a leakage current across the ferroelectric layer. U and D pulses are non-switching pulses as the domains are already poled in the P_{down} and P_{up} states, respectively. However, dielectric and leakage currents are measured, as can be seen in Figure 1.8.b. Therefore, subtracting P-U and N-D currents isolates the ferroelectric switching contribution, which is then integrated over time to obtain the PUND-corrected polarization using Equation 1.1

$$P = \frac{1}{S}\int I\,\mathrm{d}t \tag{1.1}$$

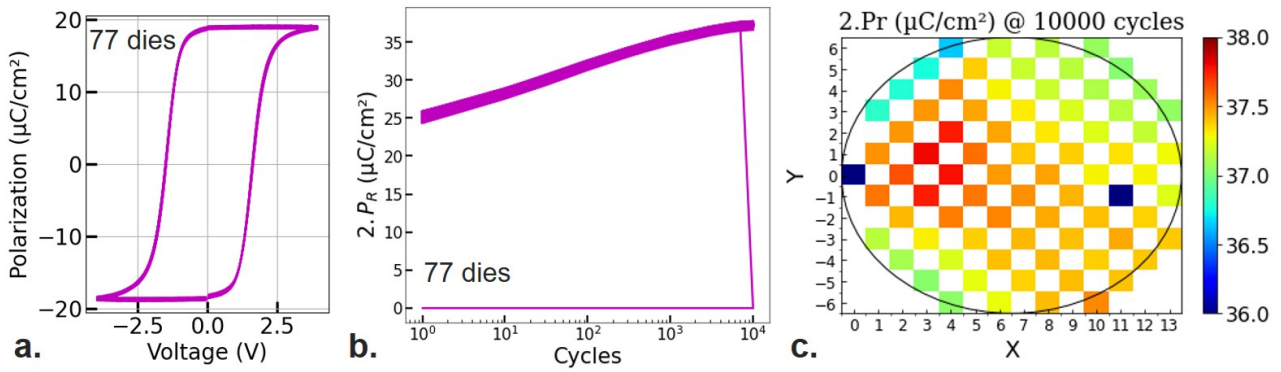


Figure 1.9: For a 300mm Short Loop wafer with a TiN/HZO (12nm)/TiN stack: a. P-V characteristics measured for 77 devices with a 4MV/cm - 10kHz PUND sequence after 10^4 wake-up cycles. ; b. $2.P_R$ evolution with cycling measured for 77 devices with 10 PUND sequences per decade. ; c. Wafermap of $2.P_R$ after 10^4 wake-up cycles for the measured devices evidencing $2.P_R$ values between 36 and 38$\mu C/cm^2$.

An example of electrical characterization results obtained on such a ShL wafer is given in Figure 1.9. In this case, 77 100µm diameter capacitors with a TiN/HZO (12nm 50-50)/TiN stack were measured with the PUND technique at 4MV/cm - 10kHz during cycling at 4MV/cm - 100kHz. Figure 1.9.a. presents the P-V hysteresis loops of the 77 devices measured after 10^4 cycles. As all loops overlap, it can be inferred that a high device-to-device homogeneity is obtained thanks to the ShL fabrication process and the 300mm tools. Moreover, the obtained remanent polarization nears $40\mu C/cm^2$, which indicates very good ferroelectric performances. As can be seen on Figure 1.9.b., the remanent polarization increases from 25 to over $35\mu C/cm^2$ with cycling. Although no initial wake-up procedure was applied, the pristine $2.P_R$ values are quite elevated, and remain very similar for all 77 devices during cycling. Note that two devices out of 77 present breakdown either from the start or after cycling, as can be seen on Figure 1.9.c. After 10^4 cycles, the remanent polarization wafermap presented in Figure 1.9.c. shows high device-to-device homogeneity with a slight center-edge pattern.

Recap

The ShL approach is relevant for fast assessment of the FE performance of large capacitors, and provides a very large number of nominally identical devices with state-of-the-art $2.P_R$. The behaviors observed at large capacitor scale can mainly be extrapolated to scaled capacitors, but care has to be taken for material comparison between 200mm versus 300mm diameter wafers as the fabrication tools may differ for the two wafer sizes.

1.2.4.2 Scaled Devices Integrated in BEoL CMOS

In this section, the focus is directed towards scaled and BEoL-integrated ferroelectric capacitors, which constitute initial demonstrators in view of integrating 1T-1C FeRAM arrays. In this case, Front-End-of-Line (FEoL) transistors are dimensioned to the 130nm technology node on 200mm diameter wafers, and manufactured by STMicroelectronics. In the Back-End-of-Line (BEoL), Metal 1 to Metal 4 interconnections are fabricated by STMicroelectronics as well. At CEA-Leti, the ferroelectric capacitor is integrated between Metal 4 and Metal 5, as is presented in Figure 1.10. The critical dimensions of scaled ferroelectric capacitors are comprised between 300nm and 600nm, and the ferroelectric material can be silicon-doped hafnium oxide (HSO) [66] or an alloy of hafnium and zirconium oxides (HZO). The ferroelectric layer thickness is typically 10nm, but interest in and attempts at reducing the layer thickness will be presented in Chapter 2 and Chapter 3, respectively. As shown in Figure 1.11, several typologies of scaled BEoL-integrated devices are available. Single 1C and 1T-1C are equivalent to the FeCAPs integrated in the bitcells of 1T-1C FeRAM arrays, but are challenging to characterize individually because of their limited area, as reported in [67]. Therefore, capacitors fields (nC) made of parallely routed single 1Cs provide a larger effective area and allow for easier electrical characterization while preserving the electrical signature of realistic scaled FeCAPs. They will be the main focus of this thesis, and the characterization method adopted is similar to the one presented in Paragraph 1.2.4.1 and Figure 1.7 and Figure 1.8. In the case of 1C and nC two-terminal devices, the TE and BE pads are contacted by two probes on the top face of

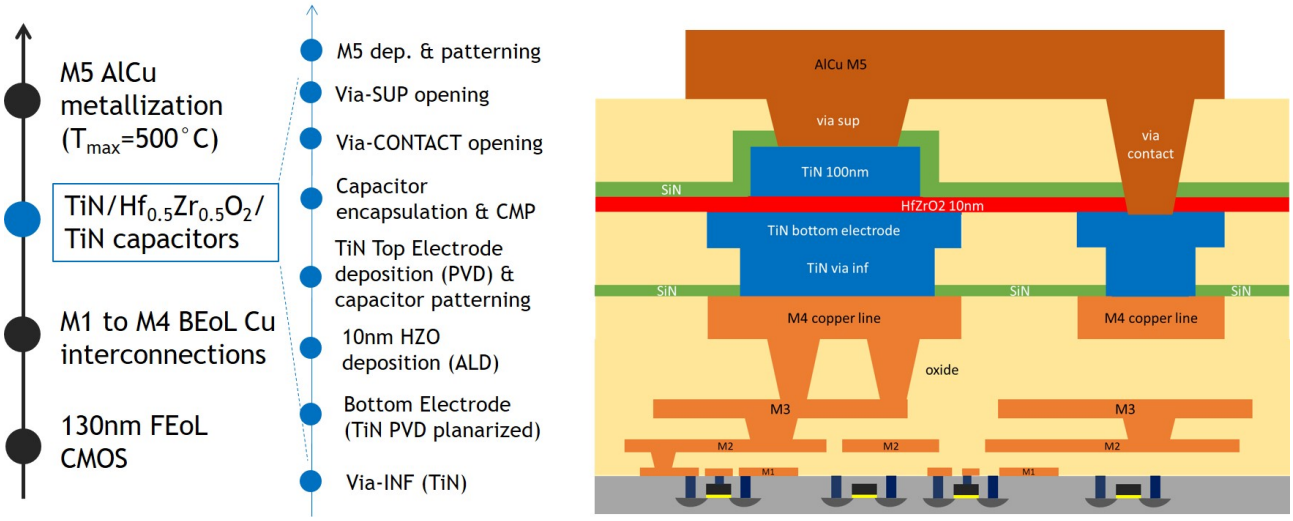


Figure 1.10: Detailed steps and schematic for the integration of ferroelectric capacitors between Metal 4 and Metal 5 in the Back-End-of-Line of ST Microelectronics CMOS process at the 130nm technology node on 200mm diameter wafers.

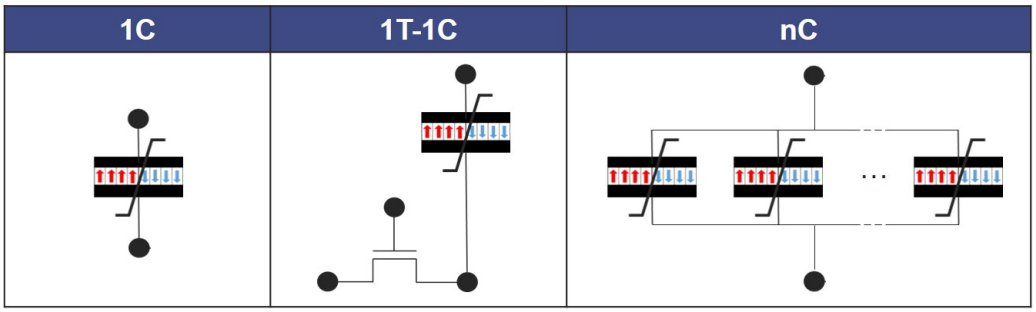


Figure 1.11: Schematics of the BEoL-integrated structures available for electrical characterization on 200mm wafers at the 130nm node: scaled capacitors (1C), 1 Transistor - 1 Capacitor bitcells (1T-1C), and capacitors fields made of n nominally identical scaled capacitors (nC).

the wafers ; in the case of the 1T-1C four-terminal device, four probes are necessary to contact the top electrode of the 1C, as well as the source, gate and bulk of the 1T.

Recap

Scaled FeCAPs are integrated in the BEoL of 130nm transistors and consistute a first step towards FeRAM array integration, as well as a useful tool to analyze the ferroelectric performances and reliability properties of devices. In this thesis, the starting point Process of Reference (PoR) is a TiN/ HSO or HZO (10nm)/TiN stack, and the preferred devices for extensive electrical characterization are capacitor fields made of individual scaled 1Cs.

1.2.4.3 1T-1C FeRAM Arrays for Statistical Evaluation

The main interest for a FeRAM array demonstrator is two-fold: on the one hand, it provides a very high number of nominally identical capacitors and is therefore a precious statistical tool whose median and dispersion give useful information about the homogeneity of the processed devices. On the other hand, the FeRAM array demonstrator tends towards application-oriented circuits with dedicated programming and reading schemes and can therefore provide relevant design-related insights. In this part, a specific description of the working principle, characterization process, and metrics of interest of a 16kbit FeRAM array developed at CEA-Leti and reported in [68] are presented.

A possible FeRAM array organization is proposed in Figure 1.12.a., and corresponds to the FeRAM reported in [68]. Each of the 16384 (128 × 128) FeCAP is connected to the drain of a selection transistor at the 130nm node, thus creating 1 Transistor - 1 Capacitor (1T-1C) bitcells. Bitcells can be addressed individually by four Scan Chains and a Clock, which also handle circuit control. A bitcell can be programmed in the '0' / '1' state by positively pulsing the Source Line (SL) / Bit Line (BL), respectively. While essentially controlled by external instrumentation via dedicated analog pads, pulse widths and delays can also be finely tuned by internal Pulse Generators over more than 3 decades from the nanosecond to the microsecond range.

Each BL has a specific capacitance depending on the BL layout and the BL size, which is defined by the number of transistors connected to it (*i.e.* the number of WLs connected to it). The static read operation of a ferroelectric state is impossible on a FeCAP. Therefore, the read operation implemented consists in attempting to program the bitcell in the '0' state by pulsing SL while leaving the BL potential floating (following a pre-set at 0V). A BL voltage elevation V_{BL} is subsequently induced through the capacitive voltage divider formed by the BL and FeCAP capacitances, and depends on the programmed state of the FeCAP:

- In the case of a programmed '0', the SL pulse will generate a dielectric response (and possibly some leakage), resulting in a low ${V_{BL}}^0$ value.
- In the case of a programmed '1', the SL pulse will additionally generate a ferroelectric response (switching of the polarization), resulting in a high ${V_{BL}}^1$ value.

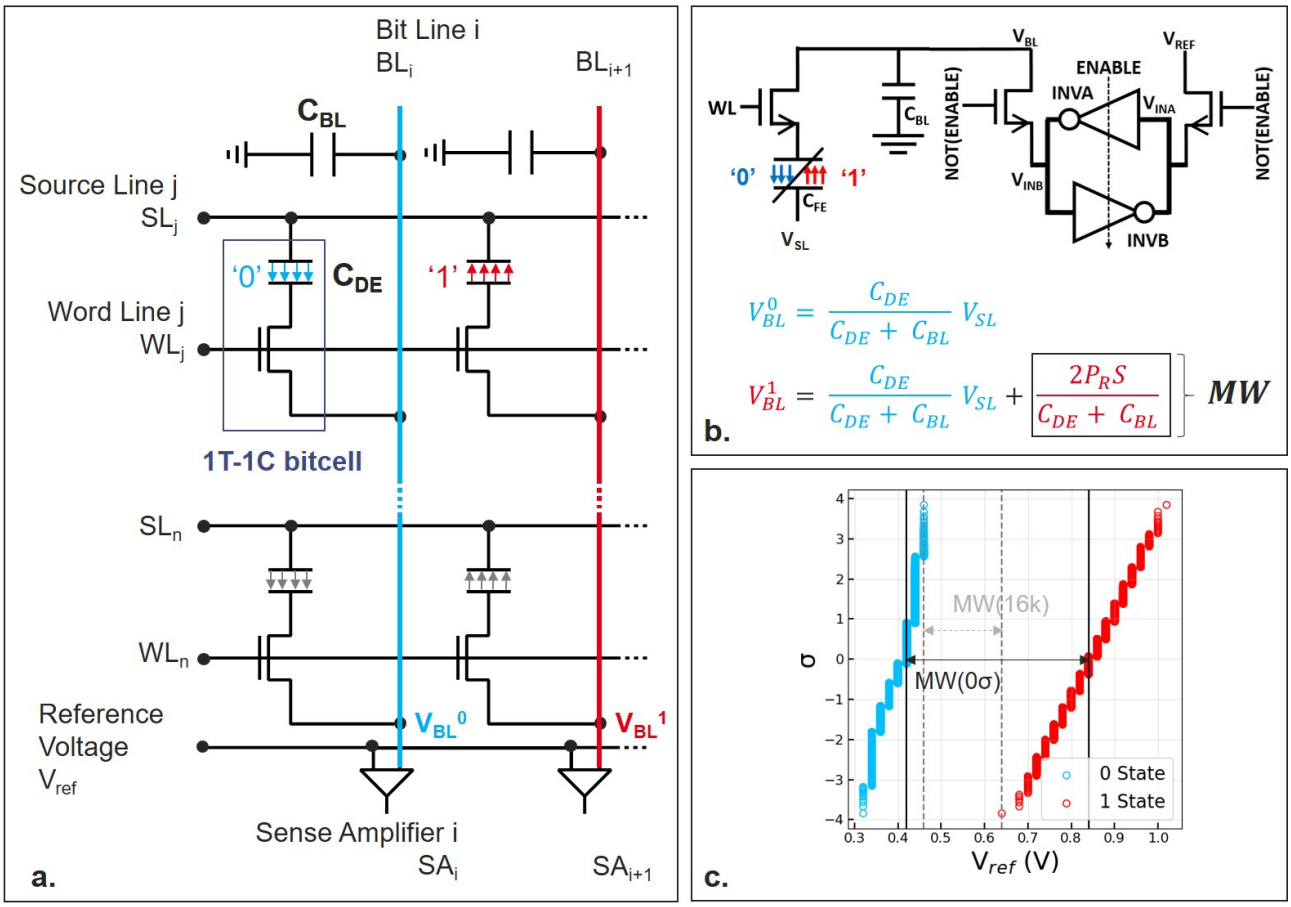


Figure 1.12: a. 1T-1C FeRAM array organization with Bit Lines perpendicular to Word Lines and Source Lines, featuring a voltage-based Sense Amplifier at the end of each BL. ; b. Schematic of the Cross-Coupled Inverters Comparator (CCIC) Sense Amplifier (SA) and associated BL voltages following a read operation. ; c. Quantile-quantile plot of '0' and '1' states from a V_{ref} scan operation showing MW(0σ) and MW(16k).

A specific Sense Amplifier (SA) designed at the end of each BL is then used for capacitive bitcell reading and consists in a Cross-Coupled Inverters Comparator (CCIC) as presented in Figure 1.12.b. Upon reading, the CCIC pre-charges V_{INB} input at V_{BL} and V_{INA} input at the reference threshold voltage V_{ref} before turning on the SA using the ENABLE signal. If V_{INB} (at V_{BL}) is higher than V_{INA} (at V_{ref}), inverter INVB will switch its output to the ground faster than inverter INVA, leading to a new steady state where V_{INB} is equal to the power supply V_{DD} and V_{INA} is equal to the ground voltage supply. As V_{INB} was chosen as the output node, and connected to a digital buffer, it outputs a '1' when $V_{BL} > V_{ref}$. Conversely, a '0' is output when $V_{BL} < V_{ref}$. The read operation is therefore destructive. Note that, in this case, '0' and '1' do no represent the logic states of the ferroelectric memory but rather the result of the comparison between V_{BL} and V_{ref}.

A well-chosen V_{ref} thus enables to discriminate ${V_{BL}}^0$ and ${V_{BL}}^1$. Write-Back capability is implemented to enable the re-writing of state '1' affected by the destructive read operation. Digital

characterization performed at various V_{ref} voltages thus enables the extraction of ${V_{BL}}^0$ and ${V_{BL}}^1$ for each bitcell. '0' and '1' states distributions are measured at the cost of 200 programming and reading operations, and thus result in a combined contribution of cell-to-cell and cycle-to-cycle variability. Cumulative '0' and '1' states distributions are calculated and quantile-quantile plots as shown in Figure 1.12.c. are plotted. The metric of interest is the Memory Window (MW): MW(0σ) describes the median behavior across the array, while MW(16k) corresponds to the read margin at FeRAM array level and accounts for bitcell-to-bitcell variability. Detailed chronograms for write and read operations, as well as a full description of the array pads can be found in [69].

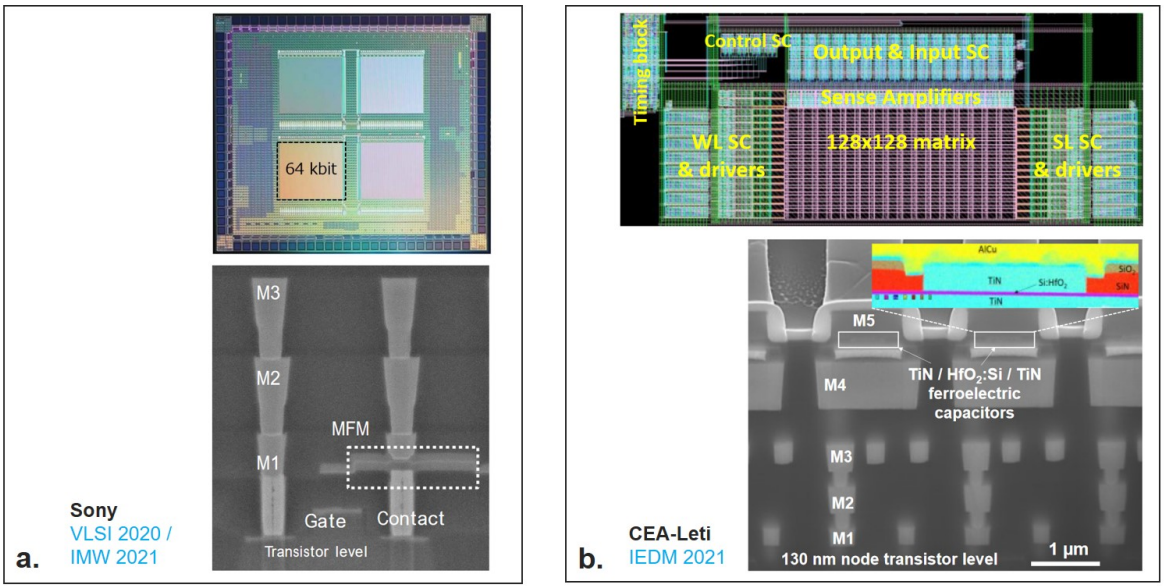
Recap

1T-1C FeRAM array demonstrators provide a statistical and circuit design perspective that is useful for performance assessment before targeting a specific application. In the considered FeRAM arrays, digital read can be performed by a Cross-Coupled Inverters Comparator at the end of each BL. An incremental V_{ref} scan process is therefore needed to extract '0' and '1' states distributions across the array, and determine MW(0σ) and MW(16k).

1.2.5 State-of-the-Art of HfO_2-based FeRAM Arrays

In this part, the state-of-the-art of HfO_2-based FeRAM arrays at the beginning of this thesis is presented. Although many academic groups actively research elementary ferroelectric capacitors, the fabrication of advanced demonstrators was principally made by two groups: Sony and CEA-Leti. While it has been evoked that ferroelectric properties in HfO_2 thin films pave the way for the integration of FeRAM arrays at technology nodes more aggressive than the 130nm node at which PZT-based FeRAMs are capped, early demonstrations of a 1T-1C 16kbit by Francois et al. [67] [68] and of a 1T-1C 64kbit by Okuno et al. [70] [71] also relied on bitcell and periphery transistors at the 130nm node. Figure 1.13.a. and Figure 1.13.b. present an optical microscope image or a layout view, as well as electron microscope cross-section images of the integrated stacks from Sony and CEA-Leti, respectively. Very similarly, planar scaled capacitors are integrated between metal lines (in the MoL and BEoL, respectively) and connected to the drain of the FEoL selection transistors. For both arrays, the voltage-sensing strategy is adopted for read operations, and consists in comparing the C_{BL}-dependent Bit Line voltage to a given reference voltage.

A benchmark adapted from [68] is presented at the bottom of Figure 1.13, and details the achieved performances of the aforementioned arrays. Similar performances in terms of read voltage and switching speed were reported, and the exploration of reliability features such as endurance or data retention were only at their starting point. In this thesis aiming for the integration of FeRAM arrays at more aggressive nodes, read margin performance assessments and reliability investigations will be at the heart of the conducted studies.



Group Publication	Structure / technology node	Array capacity	FE layer thickness & material	FeCAP integration level	Minimal write voltage	Write speed	Endurance at FeRAM array level	Retention at FeRAM array level	Solder reflow compatibility (SMT)
CEA-Leti IEDM 2019	1T-1C / 130nm	16kbit	10nm HZO	BEoL (M4-M5)	-	-	-	-	-
Sony VLSI 2020	1T-1C / 130nm	64kbit	10nm HZO	MoL (M1)	2.5V	14ns at 2.5V *(one state)*	-	-	-
Sony IMW 2021	1T-1C / 130nm	64kbit	8nm HZO	MoL (M1)	2V	16ns at 2V *(one state)*	> 10^{15} cycles at 2V (extrapolated)	-	-
CEA-Leti IEDM 2021	1T-1C / 130nm	16kbit	10nm HSO	BEoL (M4-M5)	2.5V	4ns at 4.8V *(both states)*	10^{7} cycles at 4V – 40ns (measured)	125°C during 10^{4} s with no bitfail	Yes (3x SMT with no bitfail)

Figure 1.13: a. Picture of a 64kbit array (top) and SEM cross-section of a capacitor integrated in the MoL of 130nm CMOS (bottom) from [70] ; b. Detailed layout view of a 16kbit array (top) and STEM cross-section of a capacitor integrated in the BEoL of 130nm CMOS, as well as EDX analysis of the FE stack in inset (bottom) from [68] ; Benchmark of HfO_2-based FeRAM arrays at the beginning of the thesis, including [67], [70], [71] and [68] publications, and adapted from [68].

1.3 Objectives of the Thesis

As described in the previous parts, the discovery of ferroelectric properties in CMOS-compatible HfO_2-based thin films paves the way for ultra-low power and scalable ferroelectric memory devices. Previous work at CEA-Leti has demonstrated the successful integration of silicon-doped hafnium oxide-based scaled ferroelectric capacitors in the Back-End-of-Line of CMOS technology at the 130nm node. Functional 1T-1C 16kbit FeRAM arrays were reported and key metrics of interest such as endurance, speed, or solder reflow compatibility were evaluated. These preliminary results constitute the starting point of this thesis, which is incorporated within the framework of research

and development on such 16kbit FeRAM arrays, in view of optimizing their performances, as well as integrating them in larger and denser arrays at advanced transistor nodes. With this in mind, the main goals of this thesis are two-fold:

- Assessing and optimizing the intrinsic and reliability performances of 16kbit FeRAM arrays at the 130nm node, in order to evaluate the potential of this technology from an application point of view. This necessitates to delve deeper into mechanisms occurring at the single device scale and impacting the performances and reliability at FeRAM array level.

- Evaluating the potential for scaling of HfO_2-based FeRAM arrays towards systems at more advanced nodes. This includes computational and dimensional studies for the determination of the constraints brought by transistors at advanced nodes, as well as the development of intermediary demonstrators towards advanced nodes.

To this end, this dissertation is divided into four chapters.

Chapter 2 aims at extrapolating the MW performance in large FeRAM arrays at the 28nm node. Several tools are developed to accurately estimate the advantages and drawbacks of scaling down transistors dimensions, and the implications for ferroelectric capacitor dimensioning, as well as FeRAM array design and integration. A compact model describing the ferroelectric current response of a FeCAP to an arbitrary input voltage is used to issue recommendations regarding the ferroelectric layer thickness for a readable MW in FeRAM arrays at advanced nodes, and hence with a limited available voltage. Moreover, while scaled transistors limit the current flow compared to transistors at the 130nm node, simple order-of-magnitude calculations are performed in order to evaluate how this limitation impacts the switching speed of a FeCAP. Further calculations based on an analytical equation for the MW in a FeRAM array enable the projection towards arrays with a high capacity and density, evidencing design and integration issues to be overcome in order to maintain a readable MW.

Chapter 3 addresses some of the recommendations issued in Chapter 2 for a readable MW in larger and denser arrays at advanced nodes, and specifically focuses on two aspects of the MW, namely median behavior and bitcell-to-bitcell variability at FeRAM array level. Indeed, both these parts contribute to the MW at array level, and need to be optimized for a readable MW in the most extreme reliability conditions. After demonstrating the role of mechanical stress and oxygen vacancies in the median ferroelectric performance of scaled BEoL-integrated FeCAPs, a few technological achievements are presented. Consisting of process integration and design-related demonstrations, they tend to confirm that the recommendations issued in Chapter 2 are not only achievable, but also provide the expected performance gain towards functional arrays at advanced nodes. Finally, sources of bitcell-to-bitcell variability are identified and their contribution to the overall variability at FeRAM array level is assessed. This enables making further design and layout recommendations for an increased MW at array level.

Before integrating FeCAPs and FeRAM arrays at advanced nodes, Chapter 4 focuses on the reliability properties of 16kbit FeRAM arrays at the 130nm process node. Three key aspects are evaluated: data retention, endurance, and resistance to heavy ion irradiation. In the first part, the experimental protocol of data retention for the evaluation of thermal depolarization and imprint is transferred from single FeCAP to FeRAM array level, and enables the comparison of data retention dynamics in FeCAPs and FeRAM arrays. For both data retention and fatigue-limited endurance,

the efficiency of different electrical recovery strategies can then be evaluated and optimized to overcome strong limitations. These strategies enable the 16kbit FeRAM arrays to achieve promising reliability performances. Moreover, the resistance to irradiation of FeCAPs and FeRAM arrays is evaluated in order to verify that ferroelectric memories are more resilient in harsh environments than charge-based memory technologies like Flash for example.

Finally, Chapter 5 presents the approach adopted to integrate FeCAPs and FeRAM arrays at the 28nm node. More precisely, it details the integration process, from FeCAP BEoL-integration to sanity check of the FEoL 28FDSOI transistors after the end of the whole process. On single device structures, it is verified that the functionality of capacitors, 1T-1C bitcells and capacitor fields remains unaltered compared to devices at the 130nm node, creating a first-of-a-kind demonstration. Several process options are investigated to determine a new process of reference for the ferroelectric stack, so as to maximize the chances of adequate performances for future FeRAM arrays integrated at advanced technology nodes.

To conclude this dissertation, a final part summarizes the obtained results, and explores the perspectives of this work towards the integration of high-performance and reliable FeRAM arrays with an extended capacity and a scaled footprint for ultra-low power memory applications.

2 Computational Analysis for 1T-1C Bitcell Dimensioning towards FeRAM Array Integration at Advanced Technology Nodes

Introduction

As was presented in the Introduction Chapter, the main objective of this work is to bring forward the integration of the 1T-1C FeRAM technology towards advanced transistor nodes, and specifically 28 or 22nm FDSOI. To this end, it is helpful to identify the constraints and assets brought by transistors (1T) at advanced nodes, and the implications for the ferroelectric capacitors (1C) as well as the 1T-1C bitcell as a whole. While some quick calculations (*e.g.* I_{on} of a 28nm 1T, switching charge in a 1C,...) and basic considerations (*e.g.* 28nm 1T will limit the available voltage) can easily be very helpful to define global trends, a more quantitative approach is necessary to project FeRAM array performance at advanced technology nodes. In this chapter, a computational analysis enables the simulation of the Memory Window (MW) evolution in FeRAM arrays at advanced nodes, and constitutes one of the tools used to make recommendations regarding 1T and 1C dimensioning. A first part of this chapter is dedicated to presenting the computationally-efficient compact model of a ferroelectric capacitor (FeCAP) developed at the IM2NP laboratory, calibrated on CEA-Leti FeCAP electrical characteristics, and validated at various operating voltages. This model enables the accurate simulation of the remanent polarization ($2.P_R$) evolution within boundaries defined according to die-to-die variability at wafer level. In the second part, the presented FeCAP model is coupled with an adequate foundry transistor model, and the MW is simulated using an elementary circuit designed for pure analog simulation in a 16kbit-like environment. The MW evolution as a function of the operating voltage is accurately simulated, and the MW decrease below the 2.5V experimental limitation faced at 130nm node FeRAM array level can be projected at the same technology node. Finally, the third part of this chapter uses the elementary circuit with the FeCAP model to evaluate the impact of 1T and 1C scaling on the MW in view of integrating dense FeRAM arrays at advanced technology nodes. The recommendations brought by these simulations are supported and extended by additional analytical results.

Summary

2.1 FeCAP Models . . . 47
2.1.1 Polarization-Based Models in the Literature . . . 47
2.1.1.1 Landau-Inspired Models . . . 47
2.1.1.2 Preisach-Inspired Models . . . 49
2.1.1.3 Switching Dynamics Models . . . 53
2.1.2 Current-Based Compact Model Developed . . . 54
2.1.2.1 Definition . . . 54
2.1.2.2 Calibration with FeCAP Experimental Results . . . 56
2.1.2.3 Validation at FeCAP Level . . . 58
2.2 FeRAM Array Level Simulation . . . 60
2.3 Extrapolation at Advanced Nodes . . . 63
2.3.1 1T Scaling at the 28nm Node . . . 63
2.3.1.1 Memory Window Simulation . . . 63
2.3.1.2 Switching Dynamics Considerations . . . 65
2.3.2 Towards Larger and Denser FeRAM Arrays . . . 66
2.3.3 Recommendation for 1T-1C Bitcell Scaling towards Advanced Nodes . . . 68

2.1 FeCAP Models

In order to accurately simulate the MW evolution in FeRAM arrays integrated at advanced nodes, a model for ferroelectric capacitor is calibrated and an adequate surrounding circuit is developed to enable array-like simulation. At the beginning of this section, a brief review of the literature focuses on various models enabling the simulation of the polarization as a function of the applied electric field: while historical models were inspired from ferromagnetism, they were developed and adapted to describe ferroelectrics in a physically-relevant or computationally-efficient way. This review determines several necessary and optional modelling features, that can be implemented depending on the model's purpose. Although the models considered focus on simulating the polarization evolution, the proposed compact model solution enables the simulation of the switching current response of a FeCAP in time after an externally applied electrical stress. In the following parts, polarization-based models from the literature and the developed model are presented.

2.1.1 Polarization-Based Models in the Literature

In the field of ferroelectric behavior modelling, there have been historically two distinct trends that originate from concepts developed in the early 1930's, namely Preisach model and Landau theory. From there, various adaptations or simplifications of these concepts have led to a large number of models that serve two main purposes: on the one hand, so-called computationally-efficient models enable designers to simulate the electrical response of circuits featuring ferroelectric elements ; on the other hand, physically-relevant models allow for theorists to link observed macroscopic behaviors to physical mechanisms occurring at the microscopic scale. In both cases, accurate reproduction of basic ferroelectric features as well as predictive capability are key.

2.1.1.1 Landau-Inspired Models

As introduced in the Introduction Chapter (Chapter 1), Landau theory describes a system's equilibrium behavior near a phase transition. In the case of ferroelectrics, it can be applied to phenomenologically describe the thermodynamic phase transition from a high symmetry paraelectric phase to a low symmetry ferroelectric phase. This application was initially proposed by Devonshire [74] in the case of bulk ferroelectrics, therefore the denomination Landau-Devonshire can be encountered. Models based on Landau theory are sometimes referred to as "lattice models" as they aim at describing microscopic phenomena based solely on lattice symmetry considerations [73]. During the aforementioned phase transition, Landau potential distorts from a single-well to a double-well shape, whose energy minima correspond to two metastable polarization states P_{up} and P_{down}, as represented in Figure 2.1.a. from [56]. The Gibbs free energy $F(P)$ of the system in the vicinity of the phase transition can be described in Equation 2.1 by a series expansion of the order parameter, here the polarization P:

$$F(P) = \frac{1}{2}a_0(T - T_0)P^2 + \frac{1}{4}bP^4 + \frac{1}{6}cP^6 - EP \tag{2.1}$$

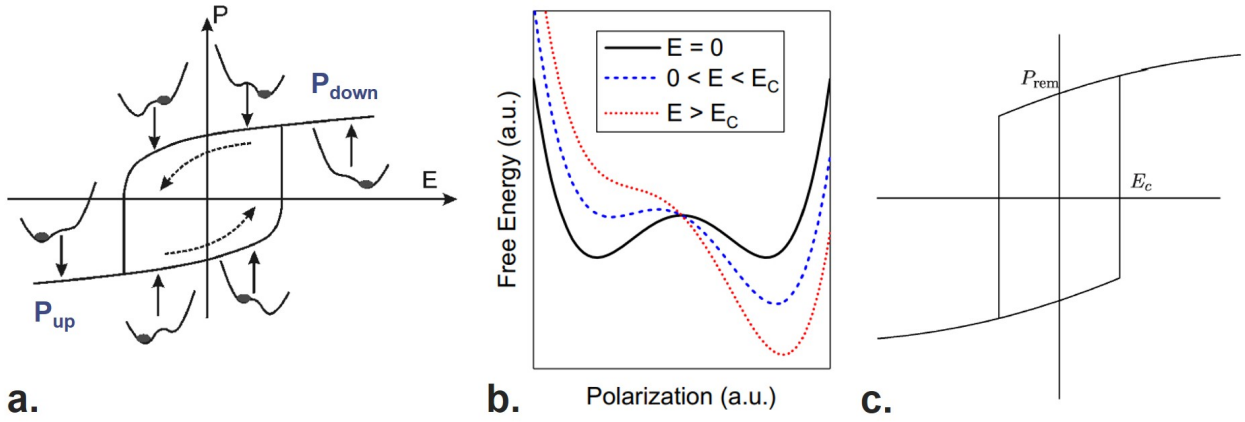


Figure 2.1: a. Ferroelectric hysteresis loop representation based on Landau-Devonshire theory linking the different polarization states at equilibrium with the minimization of the free energy potential, from [56]. ; b. Free energy potential calculated from Landau theory for different ranges of applied field, evidencing one or two energy minima, from [72]. ; c. Hysteresis loop simulated using Landau equations in a lattice model, featuring the remanent polarization and coercive field and exhibiting characteristically sharp dP/dE slopes, from [73].

where E is the applied electric field, T is the temperature, T_0 is close to the Curie temperature[1] and a_0, b, c are Landau coefficients. a_0 and c are positive, and the sign of b determines whether the phase transition is second-order (continuous) or first-order (discontinuous) [56]. These coefficients need to be calibrated on experimental data, or derived from microscopic calculations. To account for strain phenomenon, and its coupling with polarization, additional terms can be added to Equation 2.1 (see [56]), but will not be detailed further here. When no electric field is applied, two degenerate polarization states co-exist, as shown in black in Figure 2.1.b. As the electric field departs from 0, one of the minimum becomes energetically more favorable while the other increases in energy (Figure 2.1.b., blue) until its disappearance when the externally applied voltage reaches the coercive voltage (Figure 2.1.b., red) [72]. The polarization state at equilibrium P_{eq} can be derived by minimizing the system's free energy described in Equation 2.1 with regard to polarization, *i.e.* by solving $\frac{\partial F}{\partial P}(P = P_{eq}) = 0$. The polarization can be saturated in either state according to the energy minimum. In order for the polarization to switch to the other state, the application of a sufficient field in the opposite polarity is needed to distort the Landau energy potential so that a new minimum appears at the expense of the previous minimum. Thus, a sharp P-E hysteresis loop can be well reproduced, as shown in Figure 2.1.c. taken from [73]. Ideally, Landau-Devonshire theory can therefore predict the polarization at any given time, and accounts for the history effect.

However, this model does not consider any switching dynamics, as is the case in real ferroelectric systems. A viscosity coefficient ρ can be added to account for the lag time between the application of an electric field and the effective polarization reversal, and Landau-Khalatnikov model follows

[1] The Curie temperature is the temperature above which the electric polarization drops to zero and phase transition from ferroelectric to paraelectric occurs.

Equation 2.2 at equilibrium:

$$\rho \times \frac{dP}{dt} = -\frac{\partial F}{\partial P} \tag{2.2}$$

and therefore using Equation 2.1 (and minimization of free energy):

$$\rho \times \frac{\mathrm{d}P}{\mathrm{d}t} + a_0(T - T_0)P + bP^3 + cP^5 - E = 0 \tag{2.3}$$

Since we consider polycrystalline thin films, inhomogeneities in the polarization distribution across the film can be encountered. Accounting for spacial anisotropy of the polarization and boundary effects stemming from low film thicknesses enables a more accurate reproduction of the tilted shape of the hysteresis loop [56]. To this end, Landau-Ginzburg theory makes use of a polarization gradient ($|\nabla P|^2$) associated with the Ginzburg coefficient g [72]. Moreover, to account for the polycrystalline nature of the considered capacitors, the ensemble of grains (or domains) are represented by a distribution of a_0, b and c coefficients: this enables a multi-domain approach that can reproduce well the sub-cycling behavior occurring at voltages below saturation observed in fabricated devices.

A joint Landau-Ginzburg-Khalatnikov approach is for example used in the commercial Sentaurus TCAD simulator [75], in which Equation 2.4 is solved in three dimensions to determine P at all times:

$$\rho \times \frac{\mathrm{d}P_i}{\mathrm{d}t} + \alpha_i P_i + \beta_i {P_i}^3 + \gamma_i {P_i}^5 - E_i - 2g_{ij}\frac{\partial^2 P_i}{\partial {x_j}^2} = 0 \tag{2.4}$$

where α, β, γ are equivalent to $a_0(T - T_0)$, b and c in Equation 2.1 and Equation 2.3, respectively, and index i describes cartesian coordinates x, y and z.

The derived Landau-based models, when correctly fitted with experimental data, constitute a powerful tool to model physical behaviors and verify microscopic hypotheses. However, from a computational point of view, they are rather intensive and are therefore preferred for physically complex but small scale simulations.

Recap

Landau theory can be used to define a physical model for ferroelectricity based on a description of the phase transitions in a lattice. While Khalatnikov and Ginzburg adaptations to the initial theory have led to an accurate prediction of polarization hysteretic behavior in thin films, Landau-based models remain computationally intensive and are preferred for small scale physical simulations rather than complex circuit design.

2.1.1.2 Preisach-Inspired Models

For simulations on a larger scale, Preisach(-based) models are preferentially adopted. Preisach model was developed in 1935 [77] to describe the hysteretic behavior of the magnetization induced in a ferromagnet by the application of an external magnetic field. This model is based on the

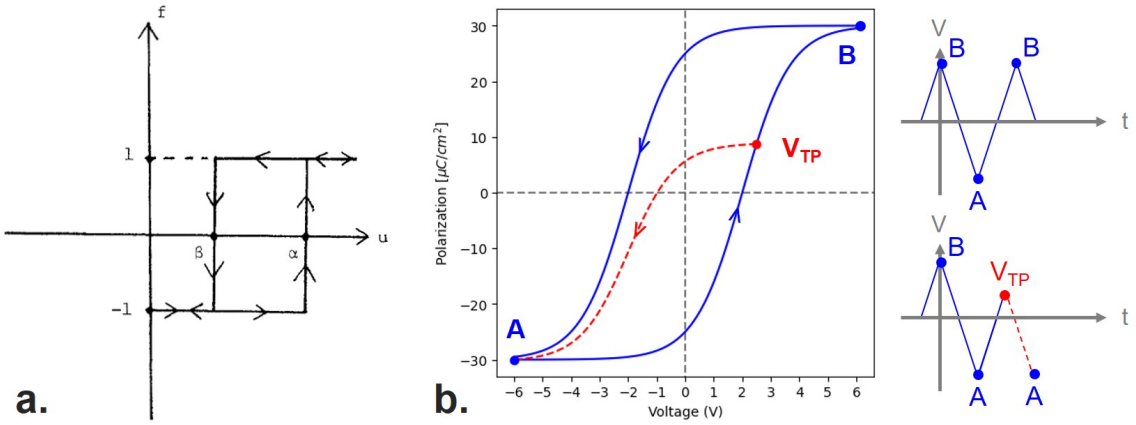


Figure 2.2: a. Representation of a hysteron from Preisach theory in the form of a rectangular shaped hysteresis loop with α and β coercive fields, from [76]. ; b. P-V hysteresis loop simulated with $tanh$ mathematical function to illustrate major loop (blue) and minor loop (red) principles, accompanied by indicative voltage waveforms on the right-hand side to present the notion of turning point (TP).

assembly of non-interacting units (called hysterons) defined by rectangular shaped hysteresis loops. As shown in Figure 2.2.a. from [76] the hystereses are defined by a couple of coercive fields (α and β) that follow a certain distribution $\mu(\alpha,\beta)$. The output described by Equation 2.5 is computed as the double integral of the hysterons over their distributed coercive fields α and β:

$$f(t) = \int_{\alpha}\int_{\beta} \mu(\alpha,\beta)\cdot\gamma_{\alpha\beta}\cdot u(t)\cdot d\alpha \mathrm{d}\beta \tag{2.5}$$

where u and f describe the input and output of a hysteresis transducer, respectively, and $\gamma_{\alpha\beta}$ is the hysteresis operator. In the case of ferromagnetism, u is the applied magnetic field and f is the induced magnetization.

In the 1990's and with the advent of PZT-based ferroelectrics for memory applications, this model was adopted more widely to reproduce the hysteretic behavior of polarization in a ferroelectric capacitor. Indeed, despite Preisach model being initially intended for phenomenologically describing ferromagnetic hysteresis loops, it was discovered by Krasnoselskii [78] that this model was deeply rooted in a mathematical paradigm that could be extrapolated to other hysteretic phenomena, for example in the field of ferroelectrics. Mayergoyz [76] furthered this approach by proposing a geometrical interpretation of Preisach model in the α - β plane which lead to a simplification of Equation 2.5, as well as a precise analysis of the history effect governing the polarization at any given time with regard to previous polarization states. From a modelling point of view, Preisach model intrinsically enables the reproduction of a hysteretic P-E loop at saturation and subloops below saturation conditions. Minor loops can be described thanks to the memory effect that stores the relevant previous states susceptible to influence future polarization states, followed by the wipe out effect that erases previous polarization states that have become irrelevant to determine future

states.

While Preisach model and Mayergoyz interpretation and simplification enable the accurate description of the hysteretic behavior and memory effect occurring in a ferroelectric capacitor, the calculation of a double integral remains computationally intensive. What's more, if no analytical solution for the double integral of the coercive fields distribution function can be found, a costly numerical double integration over a large range of switching thresholds will be required to determine a single polarization step. In [79], Meyer et al. used Mayergoyz simplification of Preisach double integral in the α - β plane, and proposed an analytical solution for the coercive fields distribution function. After exploring the adequacy of Gaussian and Lorentzian distributions, it preferred exponential functions (simple and double) that had the two-fold advantage of very accurately fitting P-E hysteresis loops, and of providing an analytical solution to simplified Equation 2.5. In this example, a purely Preisach approach was taken and was demonstrated to enable the computation of saturated P-E loops, subloops, as well as minor loops through the memory / wipe out effect of previously stored states (or turning points) that are relevant / irrelevant for the determination of future polarization points.

Beside this pure Preisach-based approach, which requires the determination of an analytical function to compute the ferroelectric polarization from Equation 2.5 consistently with experimental measurements, different modelling methods inspired by Preisach and Mayergoyz can be mentioned as follows.

The model developed by Miller et al. [80], often referred to as a simplification of Preisach model (although not explicitly defined as such), makes use of electrostatics and Maxwell equations to determine an expression for the polarization at a given time depending on previous polarization value and relying on $\mathrm{d}P/\mathrm{d}E$ evolution. In the absence of a physical law predicting $\mathrm{d}P/\mathrm{d}E$ variation, a convenient mathematical function was used to fit the hysteretic P-E loop: the hyperbolic tangent, or $tanh$. Equation 2.6 proposes the following equation to describe polarization variation as a function of the applied field:

$$P(E) = P_s \tanh\left[\left(\frac{E - E_c}{2\delta}\right)\right] \tag{2.6}$$

with

$$\delta = E_c \left[\log\left(\frac{1 + P_r/P_s}{1 + P_r/P_s}\right)\right]^{-1} \tag{2.7}$$

where P_s is the saturation polarization, P_r is the remanent polarization and E_c is the coercive field. This model is physics-based, and although it relies on an arbitrary mathematical tanh function to describe $\mathrm{d}P/\mathrm{d}E$, it enables the correct description of subloops. Moreover, this model allows for an evaluation of the hypothesis according to which the polarization loss observed during fatigue stems from a decreased number of switching domains, potentially caused by domain pinning [80]. However, one of its major limitations is that no history effect is taken into account, in the sense that the ferroelectric response to an arbitrary input (a minor loop for example) cannot be accurately described by Equation 2.6 alone.

From Miller's approach [80], Bo Jiang et al. [81] developed a more computationally efficient model that enables the modeling of minor loops. This model makes use again of two tanh functions decentered from the point of null electric field by the addition of positive and negative coercive field values to reproduce the positive and negative branches of the hysteresis P-E loop, respectively. In this

compact model, no physics-based equation is used, and the link to Miller is only the $tanh$ function to reproduce $\mathrm{d}P/\mathrm{d}V$ (or $P(V)$) variation in the form of a hysteresis loop. Similarly to a discrete approach of Preisach model (*i.e.* with a finite, or limited, number of hysterons), Jiang's approach requires the modelling of a large number of cells. Although not physics-based, this Preisach-based compact model still aims at reproducing the history behavior needed to model minor loops that stem from asymmetric sub-saturation cycling: to this end, Mayergoyz interpretation [76] was implemented in a specific algorithm. As described in [81], the ascending (+) and descending (-) branches $F^{\pm}$ of the major loop are described similarly to Miller [80] by Equation 2.8,

$$F^{\pm}(V) = Q_s \tanh[a(V - V_c^{\pm})] \tag{2.8}$$

where Q_s is the maximal charge contribution from saturated ferroelectric switching, V is the applied voltage and $V_c^{\pm}$ the positive and negative coercive voltages, and a describes the slope of $\mathrm{d}P/\mathrm{d}V$. As shown in Figure 2.2.b., these equations enable the representation of the blue polarization hysteresis loop as a result of applying a voltage similar to the upper right waveform. On the other hand, the branches of minor loops $f_i^{\pm}$ at time i are represented by major (or saturated) loops scaled and shifted by m_i and b_i coefficients, respectively, following Equation 2.9,

$$f_i^{\pm}(V) = m_i F^{\pm}(V) + b_i \tag{2.9}$$

m_i and b_i coefficients are dependent on previous polarization state at time i-1 and $f_i^{\pm}$ describes the polarization evolution inside the saturated P-V loop. Specifically, when starting from a saturated state and applying a voltage below the voltage enabling saturation of the polarization but above the coercive voltage, only a limited number of domains will be able to switch, thus causing the overall polarization to be reduced. When subsequently applying a saturating voltage in the other polarity, only the previously switched domains should switch back: at the turning point (V_{TP}), the branch of a minor loop is formed. This is illustrated in red in Figure 2.2.b., and can correspond to the lower right waveform. When alternating sub-saturation steps at different voltages, previous states need to be "remembered" until the application of higher voltages makes the future states independent of some (or all) past states. Therefore, the model is required to store enough turning points to accurately determine the next polarization state (memory effect). However, in order to avoid storing an unnecessary high number of turning points, couples of turning points are erased when becoming irrelevant to predict next polarization points (wipe out effect). To put it differently, the model keeps track of the turning points and constantly updates and/or wipes them out to determine future polarization states based on relevant past sub-saturated polarization states. This scheme is implemented as an algorithm in Jiang's compact model [81] to accurately model the polarization after any given voltage sequence and to reproduce minor loops, thus making up for the fact that it is not physics-based.

As a final note, a few observations regarding the Preisach-based models presented can be made.

- These scalar models can be extended as vector models to fully account for three-dimensional polarization.
- In all cases, the linear part of the dielectric displacement is not described, but can be easily added to the model to account for the dielectric behavior.

- To all Preisach-based models presented, it can be interesting to add some switching dynamics model, or computational equivalent (RC delay, viscosity coefficient) to accurately describe the effective polarization.

Recap

The saturated P-E loop, sub-saturation P-E loops (also called subloops), minor loops (memory and wipe out effects) occuring in ferroelectrics can be well reproduced by a Preisach approach. In the original Preisach model, all aforementioned features were encompassed in a double integral computation over a distribution of the coercive fields for single hysteresis units. In derived models, the tanh mathematical function is used to visually reproduce the saturated P-E loops and memory and wipe out effects can be added in carefully designed algorithms inspired by Mayergoyz analysis of the history effect. Preisach-based models are considered computationally-efficient and enable larger scale circuit simulations.

2.1.1.3 Switching Dynamics Models

To give a full picture of published models describing the ferroelectric response of a ferroelectric capacitor (FeCAP) after an arbitrary voltage input, it is necessary to mention two additional models by Alessandri et al. [82] and Tung et al. [83]. Indeed, the recent interest for hafnium-oxide based ferroelectrics launched a new wave of hybrid models for ferroelectricity, aiming at computational efficiency to simulate or design larger circuits featuring ferroelectric components, while maintaining close links to physical mechanisms like the switching behavior.

Alessandri et al. [82] proposed a physics-based model making use of the Nucleation Limited Switching (NLS) model [84] in a Monte-Carlo simulation framework. The NLS model is the switching kinetics model of choice when considering polycrystalline materials, like HfO_2-based thin films. Indeed, it describes the polarization from the integral of the switching probability in an ensemble of grains that follows a distribution of nucleation time constants depending on the effective applied field. To extend the NLS model, a Weibull process is considered to govern switching probability from one logic state to the other. Therefore, using a Monte-Carlo simulation framework it is possible to compute a system's switching probability and simulate its polarization response to an arbitrary waveform input. Such simulations based on a small number of grains enable the reproduction of the increased variability observed in scaled devices, as was demonstrated in [85].

In a similar manner, Tung et al. [83] defined the polarization as a function of the variable area of different grains depending on their switching rate, according to NLS theory. In this case, the polarization history was tracked to reproduce transitions from major loops to minor loops. The calibrated model was able to reproduce the accumulative behavior of polarization in which successive pulse trains enable the gradual increase of the polarization until saturation.

Recap

In recent years, hybrid models have been developed to account for switching dynamics of ferroelectric materials. While intended for large scale simulations using commercial tools, they have the additional feature of very accurately simulating the time-dependent switching behavior of multiple grains based on the physical Nucleation Limited Switching (NLS) model.

2.1.2 Current-Based Compact Model Developed

The model developed at the IM2NP laboratory and published in [86] is Preisach-inspired, in the vein of the compact models presented in Paragraph 2.1.1.2. However, contrary to these polarization-oriented models, the developed model focuses on precisely describing the ferroelectric switching current evolution. The main goal of this section is to present the developed FeCAP compact model to be used for the projection of the MW performance of FeRAM arrays at advanced nodes.

2.1.2.1 Definition

The compact model developed at the IM2NP laboratory aims at reproducing the complete current variation measured when applying a Positive Up Negative Down (PUND, [65]) sequence to a FeCAP. In this entire chapter, experimental data have been measured on a TiN/HSO (10nm)/TiN stack integrated in the BEoL of CMOS process at the 130nm node. The ferroelectric layer was etched in the process of etching the top electrode, as reported in [86]. Test structures available on the wafer include capacitor fields made of 1290 FeCAPs routed in parallel for electrical signal enhancement, and 16kbit 1T-1C FeRAM arrays for statistical reinforcement. The demonstration and functionality of such FeRAM arrays was presented in the Introduction Chapter (Chapter 1) and previously reported in [87]. Figure 2.3.a. presents the variation of the current measured on a BEoL-integrated scaled FeCAP field with individual capacitors of 550nm diameter, together with the applied voltage waveform at 4.8MV/cm. It consists of two positive polarity pulses (P, U), followed by two negative polarity pulses (N, D) to enable the extraction of the ferroelectric, dielectric and leakage currents. A negative polarity pulse was initially applied to generate a fully polarized state. The dielectric current (I_{DE}), ferroelectric switching current (I_{FE}) and static leakage current ($I_{leakage}$) all contribute to the total current measured and can be modelled by the following analytical equations:

$$I_{DE} = C_{DE}\frac{dV}{dt} \tag{2.10}$$

with C_{DE} the dielectric constant defined as:

$$C_{DE} = \frac{\epsilon_0 \cdot \epsilon_R \cdot S}{t_{FE}} \tag{2.11}$$

where ϵ_0 and ϵ_R are the vacuum permittivity and FE layer relative permittivity, respectively, and S and t_{FE} the capacitor area and FE layer thickness, respectively.

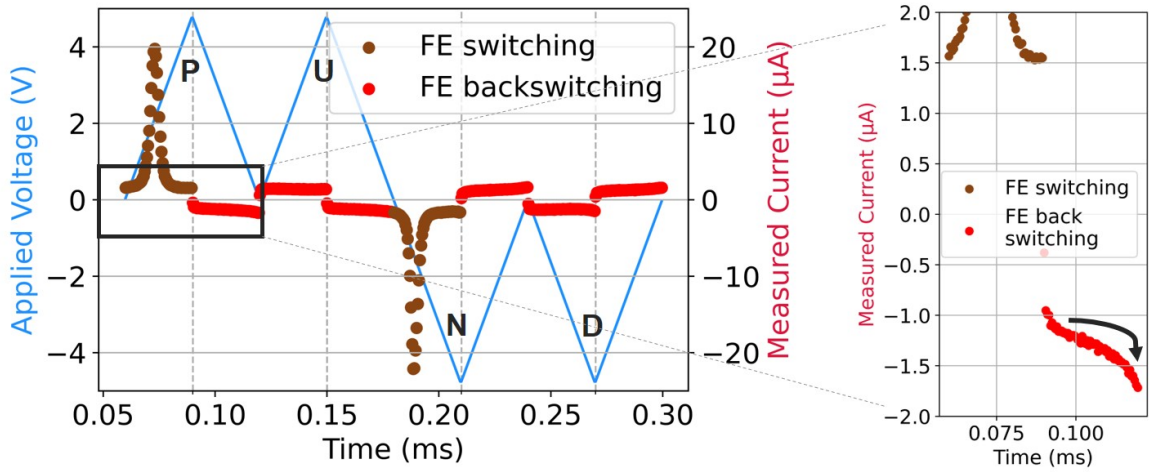


Figure 2.3: Example of a post wake-up I-t characteristics (red) measured on a 306μm^2 FeCAP field following the application of a PUND sequence (blue) at 4.8V where switching and backswitching behaviors are evidenced, and zoom on P-pulse induced currents.

$$I_{FE} = S\frac{dP}{dE}\frac{dE}{dt} \tag{2.12}$$

with $E = \frac{V_{app}-V_{off}}{t_{FE}}$ the electric field depending on the applied voltage V_{app} and offset voltage V_{off} that accounts for FeCAP possible coercive voltages asymmetry and fixed charges. $\frac{dP}{dE}$ will be determined to match the measured ferroelectric current distributions.

$I_{leakage}$ is modelled by Equation 2.13, which enables the description of the main trap-assisted conduction mechanisms, including the Poole-Frenkel effect [88], and so:

$$I_{leakage} = S \cdot A_{leak} \cdot E^{B_{leak}} \tag{2.13}$$

where A_{leak} and B_{leak} are fitting parameters.

Similarly to Preisach-based models described in Paragraph 2.1.1.2, an adequate function needs to be determined for $\mathrm{d}P/\mathrm{d}E$; contrary to these models, and as shown from Equation 2.12, the primary focus will be to fit the measured ferroelectric current rather than the polarization. Moreover, a relevant algorithm has to be developed in order to store and wipe out turning points that are mandatory to predict the future polarization based on past history in case of sub-saturation stress.

As can be seen on Figure 2.3 for a FeCAP stressed at saturation between ±4.8V, the ferroelectric part of the current can be separated into switching and backswitching behaviors.

- In the switching case, a sharp current peak is measured and indicates the switching of a vast majority of domains, leading to saturated polarization. This occurs during the rising edge of the first voltage peak that follows saturation in the opposite polarity, for voltages higher than the minimal coercive voltage over the ensemble of domains. It can occur in both polarities, as is shown for P and N pulses of the PUND sequence.

- In the backswitching case, as can be seen in light red in Figure 2.3, non-flat current contributions are added to an otherwise purely dielectric behavior outside of the ferroelectric switching regions. Indeed, because of the experimental distribution of coercive voltages (translated in the broadness of the switching current peak), backswitching phenomenon occurs during the falling edge of a switching peak for both polarities, and also for the rising and falling edges of the U and D pulses as a consequence.

These considerations enable the determination of main cases for the ferroelectric current (switching or backswitching), as well as sub-cases regarding positive and negative polarities, to anticipate and account for potential ferroelectric current asymmetry. An algorithm is adopted to account for the history effect (see Paragraph 2.1.1.2), thus allowing to determine whether present and past voltage conditions give rise to ferroelectric switching or backswitching currents. In the switching case, the peak of the current distribution is the main contribution, whereas in the backswitching case, the tail of the current distribution is the main contribution. This compact model does take into account subloops, in the strict sense of symmetric sub-saturation voltage waveform applied to a FeCAP with symmetric I-V characteristics. Moreover, as the current response accounts for a potential asymmetric sub-saturation switching at a previous step, the model is expected to accurately reproduce any minor loop, especially in the case of asymmetric applied voltage or I-V characteristics. Therefore, an accurate determination of the switching current (and hence of the polarization) on a wide range of applied voltage conditions is possible for any sub-saturation switching. Indeed, similarly to work presented by Jiang [81], turning points are stored, and wiped out when becoming irrelevant to predict future polarization states (or here, currents). This compact model developed allows for the storage of a limited number of turning points, and has been verified to accurately reproduce the minor loop behavior. More details on the calibration and validation of the model will be given in the following sections.

Recap

The model developed is Preisach-based and consists in describing the current response to a voltage waveform of a FeCAP as a function of ferroelectric, dielectric and leakage currents. The ferroelectric response is divided in switching and backswitching behaviors depending on the applied voltage and its polarity compared to a previous voltage extremum. The model stores and wipes out turning points so as to reproduce minor loop behavior.

2.1.2.2 Calibration with FeCAP Experimental Results

In the previous section, the Preisach-based compact model and its associated algorithm enabling the simulation of the current following a PUND-like voltage sequence were presented. In this part, the model is calibrated with FeCAP experimental results. Here, the FeCAP considered for model calibration experienced a 16kbit-like bipolar wake-up with square pulses at 4.8V - 2µs. This approach aimed at mimicking stress conditions experienced by a FeCAP in a FeRAM array, so that the calibrated model for FeCAPs would be relevant in a FeRAM array environment. To measure

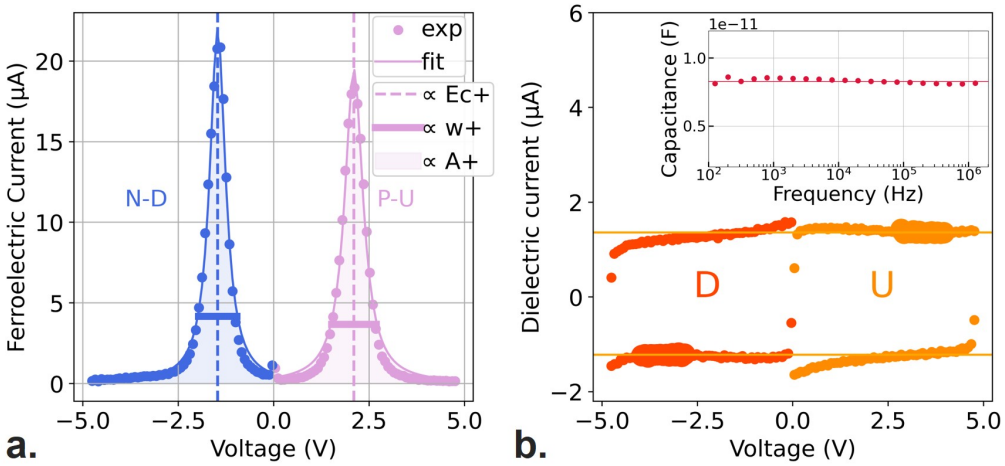


Figure 2.4: On an etched FeCAP of median behavior: a. Experimental FE current from P-U and N-D voltage pulses at 4.8V - 30µs and Lorentzian distribution calibrated to fit I_{FE} ; b. DE current from U and D voltage pulses at 4.8V - 30µs. Inset: Capacitance measured over the 100Hz - 2MHz frequency range at 0V. The relative permittivity extracted from C-f determines the DE current area of reliable ϵ_R extraction, highlighted on the graph by bigger dot symbols.

ferroelectric currents, standard triangular PUND sequences were still used. Based on Equations 2.10, 2.12 and 2.13, the model is calibrated to match the total current measured with 4.8V - 30µs pulses on a 550nm diameter etched capacitor field chosen to be representative of wafer median behavior (Figure 2.4).

Regarding the dielectric current, as can be seen from Equation 2.10 and Equation 2.11, the only material-related parameter I_{DE} depends on is the relative permittivity of the ferroelectric material, ϵ_R, which has therefore been chosen as a fitting parameter. Extracting ϵ_R from the DE current measured during the U and D voltage pulses is challenging due to backswitching and leakage phenomena that can possibly occur in FeCAPs. In order to verify the accuracy of such extraction, a capacitance versus frequency (C-f) measurement was performed on the same device. Using E4980A Precision LCR meter at 0V on a 100Hz - 20MHz frequency range enables the determination of a median value for C and hence for ϵ_R according to Equation 2.11, as is demonstrated in Figure 2.4, inset. This ϵ_R value was then converted into a dielectric current, and reported on Figure 2.4.b. The largest dots markers determine a safe area for reliable ϵ_R extraction, and therefore validate the approach of relying on PUND-measured I_{DE} to determine ϵ_R to be used as a fitting parameter.

To calibrate the ferroelectric current part of the model, an adequate dP/dE function has to be determined and adjusted to fit the measured data. As shown in Figure 2.4.a., a Lorentzian distribution allows a good fit of both positive and negative ferroelectric current peaks so $\frac{dP}{dE}^{\pm}$ can be defined as:

$$\frac{dP^{\pm}}{dE} = \frac{2 \cdot \frac{A^{\pm}}{\pi} \cdot w^{\pm}}{4 \cdot (E - E_c^{\pm})^2 + w^{\pm 2}} \tag{2.14}$$

To accurately describe both positive (P-U) and negative (N-D) ferroelectric current peaks, three

fitting parameters from Equation 2.14 are adjusted to the experimental dots of Figure 2.4.a. for each polarity through a non-linear least squares numerical method. Automated and systematic parameters extraction from the measured current has been demonstrated to be reliable and repeatable for different process options. Although Equation 2.14 is not physics-based, these parameters can be approached as physical characteristics of the FeCAP: E_c, the coercive field, w, the dispersion of the domain-scale coercive fields and A, the peak area, or domains density.

Table 2.1 lists the 12 fitting parameters extracted from the etched FeCAP experimental data presented in Figure 2.4 complemented with the FeCAP dimensions to calibrate the model. In addition to the fitting parameters for the ferroelectric switching current described just above, ϵ_R enables the modelling of the dielectric current, and the FeCAP area and FE layer thickness dimensions are used in Equations 2.10, 2.12 and 2.13. In the case studied, no leakage was measured, so the fitting parameters A_{leak} and B_{leak} are fixed to 0, but it has been verified that non-zero parameters enable the precise reproduction of the leakage-induced I-V deformation at higher voltages.

Table 2.1: Fitting parameters values and capacitor dimensions for FeCAP compact model calibration

Ferroelectric						
$A(C/m^2)$		$E_c(V/nm)$		$w(V/nm)$		$V_{off}(V)$
A^+	A^-	E_c^+	E_c^-	w^+	w^-	
0.365	0.354	0.179	-0.179	0.058	0.050	0.32
Dielectric		**Leakage**		**Dimensions**		
$\epsilon_R(-)$		A_{leak}	B_{leak}	$S(\mu m^2)$	$t_{FE}(nm)$	
29.7		0	0	306	10	

2.1.2.3 Validation at FeCAP Level

In order to validate the accuracy of the FeCAP compact model, two aspects need to be evaluated: firstly the combined currents (including ferroelectric and dielectric contributions) have to fit the measured data in the calibration conditions ; secondly, the model has to accurately simulate the current evolution beyond the calibration conditions.

The simulation results presented were obtained thanks to the thus-calibrated FeCAP compact model using the purely analog Eldo simulator commercialized by Siemens. Figure 2.5.a. and Figure 2.5.b. respectively present the I-t and I-V characteristics of the experimental current (red dots) and of the current simulated using the previously presented and calibrated model (blue line). Figure 2.5.c. presents the P-V loops extracted from the measured and simulated I-V characteristics. The combined currents including ferroelectric and dielectric contributions simulated when applying a PUND signal at 4.8V to the FeCAP perfectly reproduce the measured data. It is especially noticeable that the combination of the ferroelectric switching and backswitching currents with the dielectric current enable a very accurate description of the measured characteristics across the entire sequence. When integrating both the measured and simulated current to assess the

accuracy of the model to reproduce P-V hysteresis loops, it can be seen that a perfect overlap between experimental and simulated data is obtained at calibration conditions.

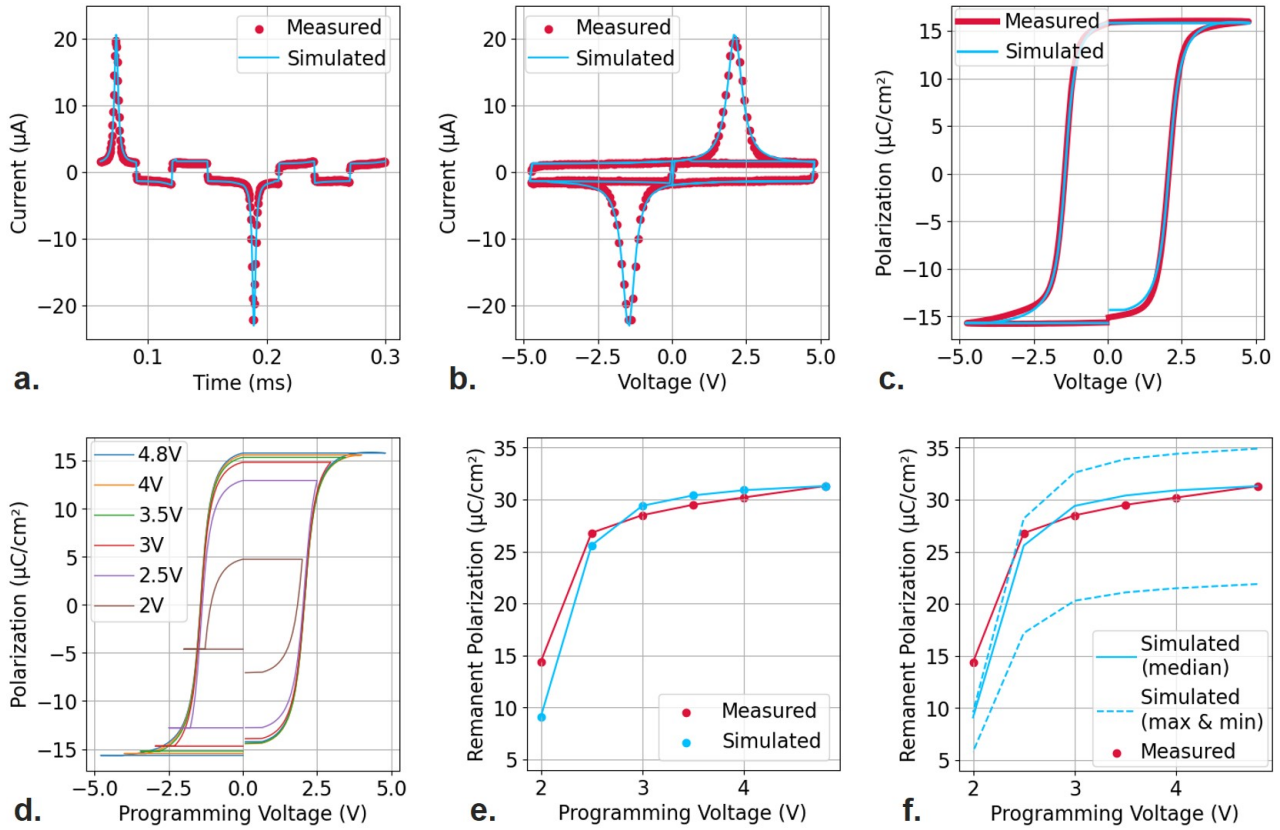


Figure 2.5: On a 550nm diameter etched FeCAP of median behavior after wake-up with square pulses: a. Experimental I-t measured for a PUND sequence at 4.8V - 30µs (red) and simulated I-t using the developed current-based model (blue) ; b. Corresponding I-V characteristics ; c. Corresponding experimental and simulated P-V hysteresis loops numerically integrated from a. ; d. P-V hysteresis major loop simulated within calibration conditions at 4.8V, and nested subloops simulated at different operating voltages ; e. Corresponding $2.P_R$ values extracted from measured and simulated P-V subloops ; f. Simulated $2.P_R$ versus programming voltage where dashed lines result from simulations calibrated on measured FeCAP of maximal and minimal $2.P_R$ across the wafer.

To assess the model accuracy beyond the calibration conditions, the simulated FeCAP is subjected to a wider range of voltage operation. It is important to note that the model was not re-calibrated to fit experimental data at different operating voltages, but that the initial parameters calibrated on experimental results at 4.8V and listed in Table 2.1 remained unchanged. As can be observed in Figure 2.5.e. in red, the remanent polarization extracted from measured current follows a slightly decreasing trend as the operating voltage is scaled down, until an abrupt $2.P_R$ loss occurs below 2.5V. This trend is expected since the coercive voltage distribution observed in Figure 2.5.b. indicates

that fewer domains will be switchable below 2.5V, hence reducing the overall remanent polarization. As reported earlier, the compact model takes into account the history effect by relying on stored turning points to compute the ferroelectric current [81] and so reproduces well the low voltage nested hysteretic subloops behavior of FeCAPs in the case of asymmetric I-V characteristics, as shown in Figure 2.5.d. The simulated $2.P_R$ along programming voltage successfully follows the same decreasing trend as was observed on measured data (Figure 2.5.d.). Boundaries determined by simulations using the compact model calibrated on dies of extreme $2.P_R$ behavior across the wafer surround the experimental $2.P_R$ (Figure 2.5.f.), thus validating the model accuracy at different voltages at the single-cell level.

Recap

The Preisach-inspired model presented considers a Lorentzian distribution of ferroelectric currents, and can be fitted to experimental currents measured on a scaled BEoL-integrated FeCAP after the application of a PUND sequence at 4.8V. Relevant parameters for ferroelectric, dielectric and leakage currents modelling are safely extracted from the PUND measurement, and dimensional parameters complete the set of 12 parameters needed to calibrate the model. Beyond excellent current prediction in calibration conditions at 4.8V, the model was shown to provide accurate description of $2.P_R$ dependence with programming voltage, and especially subloop behavior below saturation.

2.2 FeRAM Array Level Simulation

This FeCAP model can be coupled with a model provided by foundries for FEoL transistors at the 130nm or 28nm node. Using Eldo simulator, the current output from a classic PUND operation on a 1T-1C bitcell depends on the voltage applied to the transistor gate. When the gate is fully polarized (*e.g.* at V_{dd}=4.8V for a GO2 130nm transistor), the I-V characteristics simulated on a 1T-1C are strictly identical to characteristics simulated on a single 1C, as is observed experimentally. However, in order to simulate the median Memory Window (MW(0σ)) of a 16kbit array, a more complex circuit reproducing the array sensing scheme is needed. In a full 16kbit FeRAM array circuit simulated by fast-spice commercial tools, as was used for the design of the 16kbit FeRAM array that was published in [68], the combination of complex transistor models and the test-case based VerilogA compact model for FeCAP leads to elevated simulation runtimes. Therefore, an elementary circuit model is developed in order to simulate MW(0σ) evolution in a 16kbit FeRAM array-like environment, and assess the scalability of FeRAM technology towards advanced nodes.

To simulate the programming operations and capacitive reading mode implemented on 16kbit FeRAM arrays, a 130nm 1T-1C bitcell routed in an elementary circuit is simulated. The elementary circuit developed is presented in Figure 2.6.a. and features the standard 1T and 1C components constituting the bitcell, that are respectively modelled using a foundry 1T model and the previously presented FeCAP behavioral compact model. Length (L) and width (W) of the channel area are both fixed at 500nm for the bitcell transistor. To the periphery of this bitcell, four transistors that emulate the environment of a bitcell in a 16kbit array are added. Similarly to the bitcell access transistor,

they are STMicroelectronics HCMOS9 transistors at the 130nm node. On each port of the bitcell, via the Source Line (SL, also called Plate Line PL) and Bit Line (BL), two transistors on each side (hence the four periphery transistors) enable the connection to either a voltage source or the ground. In addition to the 1T-1C bitcell and the four periphery transistors, a capacitance component is implemented in order to emulate the Bit Line capacitance, C_{BL}. In FeCAP-based circuits, it is of paramount importance to account for C_{BL}, as it directly impacts the voltage dynamics during a read operation. Indeed, as will be shown in Equation 2.16, MW(0σ) is inversely proportional to C_{BL}, and variations in C_{BL} drastically affect MW(0σ). The estimation of C_{BL} is complex as it intrinsically depends on many parameters of the circuit itself: node and dimensions of the bitcell access transistors, metal level at which the FeCAPs are integrated, and size and layout of the entire FeRAM array. In a pragmatic approach to estimate C_{BL}, two contributions are considered and evaluated.

- On the one hand, the gate-source capacitance (C_{GS}) of 130nm GO2 (thick oxide gate) transistors with L = W = 500nm was evaluated using the adequate foundry model. In Eldo simulations varying dV_{GS}/dt on a single 1T, C_{GS} was extrapolated from the slope of I_G versus dV_{GS}/dt. Multiplied by 128 to account for all the access transistors connected to a given BL in a 16kbit array (with 128 WL * 128 BL), a total C_{GS} contribution of 108fF was obtained. Note that this contribution depends on the BL size, the technology node of the bitcell transistors, as well as their dimensions.

- On the other hand, the parasitic capacitance coupling between metal lines plays a significant role in C_{BL} as well. From the complete layout of a 16kbit array designed at CEA-Leti, it is possible to extract resistive and capacitive parasitic elements using Synopsis StarRCXT tool. In a 16kbit array at the 130nm node, the parasitic capacitance extraction outputs a value of 80fF, in the same order of magnitude as C_{GS} of the 128 transistors placed on the same Bit Line. Note that this contribution depends on the layout of the array, that itself is adjusted to the technology node of the bitcell transistors.

Therefore, C_{BL} is estimated at 108+80=188fF to account for the combined effect of the cumulative gate-source capacitance brought by the access transistors, together with the parasitic capacitance brought by the Bit Line layout in the considered 16kbit array at the 130nm node. Both effects are significant and will affect MW(0σ) performance at FeRAM array level. In Figure 2.6.a., the elementary circuit is summarized. In order to validate this elementary circuit and its relevance to accurately simulate MW(0σ) evolution in a 16kbit FeRAM array at the 130nm node, programming and reading operations were simulated by applying specific voltage waveforms to the circuit ports. Figure 2.6.b. presents a chronogram corresponding to a realistic voltage sequence applied to the WL, SL and BL to simulate the following operations on a 16kbit array: programming a bitcell in the '0' state (W0), reading (R), programming a bitcell in the '1' state (W1), reading (R), and reading again (R) to evaluate the efficiency of the write-back operation. The Bit Line voltage elevation is monitored during read phases, and the simulated MW(0σ) is defined as $V_{BL}^{1} - V_{BL}^{0}$. As can be seen in Figure 2.7, the simulated MW(0σ) follows the same trend along programming voltage as the MW(0σ) measured with 4.8V - 2µs pulses on a 16kbit array with square 600nm etched capacitors, on the previously used wafer with a 10nm thick HSO stack. Note that the model parameters remain unchanged, except for the FeCAP area. This result is consistent with $2.P_R$ results at FeCAP level,

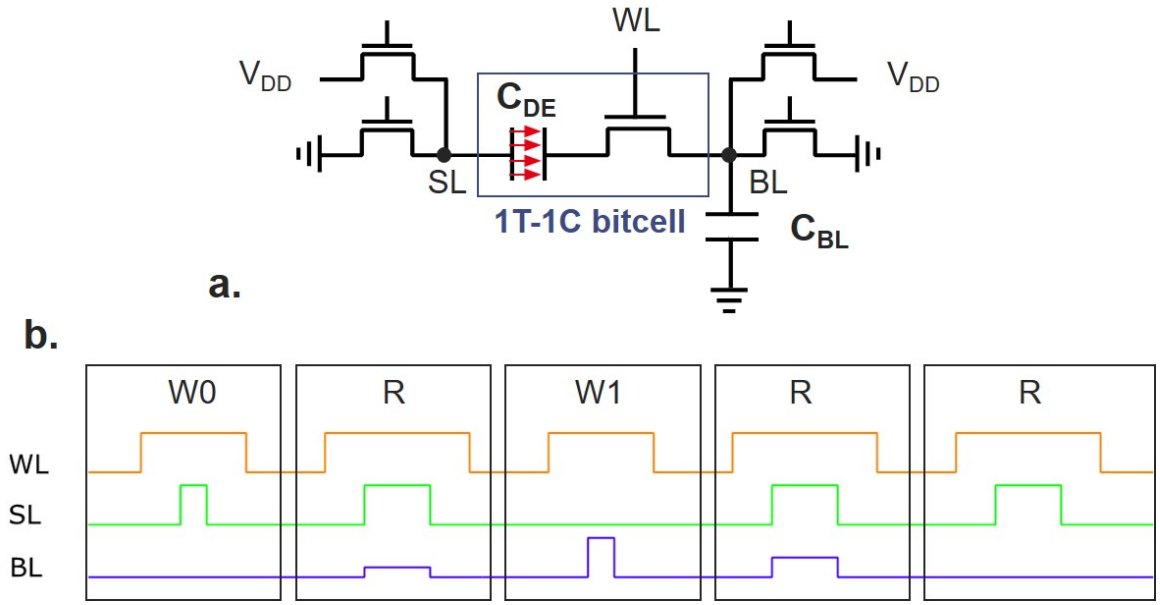


Figure 2.6: a. Elementary circuit featuring the 1T-1C bitcell, 4 surrounding transistors connected to the ground and VDD on the SL and BL sides, and a capacitive component to emulate Bit Line capacitance. Calibrated FeCAP model is used for the 1C, and foundry models are used for the access device and periphery. ; b. Example of a chronogram for quick assessment of Bit Line voltage elevation when reading a FeCAP programmed in '0' and '1' states, where input signal evolution is similar to FeRAM array operation.

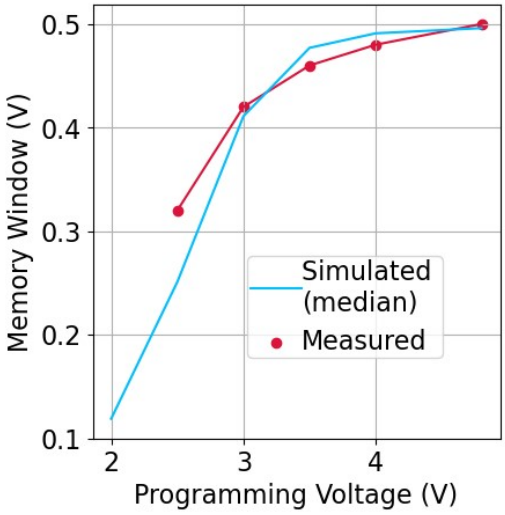


Figure 2.7: MW(0σ) values measured (red) on a 16kbit FeRAM array and MW(0σ) simulated (blue) on the elementary circuit (see Figure 2.6.a.) combining the FeCAP compact model and a 1T foundry model emulating a full 16kbit circuit, both with square 600nm etched capacitors.

presented in Figure 2.5.f. The decreasing trend of the MW(0σ) at lower voltages is well reproduced by the simulation, thus validating the elementary circuit and proving that it can correctly simulate MW(0σ) variation along voltage.

Recap

The full 16kbit FeRAM array circuit for design enablement delivers precious information about the parasitic capacitance brought by the Bit Line layout. The gate-source capacitance of the access transistors, which constitutes the other major contribution to C_{BL}, can be assessed with simple Eldo simulations based on a 1T foundry model. This complete C_{BL} contribution was implemented in an elementary circuit developed to fit the experimental MW(0σ) measured on 16kbit FeRAM arrays at the 130nm node. Relying on the FeCAP compact model, the reduction of MW(0σ) with decreasing voltage is consistently simulated, thus enabling the precise and flexible simulation of MW(0σ) in FeRAM arrays.

2.3 Extrapolation at Advanced Nodes

The goal for developing a FeCAP model and an associated elementary circuit emulating the 16kbit FeRAM array environment was to assess the median memory window evolution at advanced technology nodes. In this section, the impact of 1T and 1C scaling on MW(0σ) is simulated using the computational tools developed, allowing to issue global recommendations. In order to focus the integration efforts in specific directions, challenges and advantages brought by transistors at advanced nodes are studied. More broadly, the double requirement for array size expansion and high density and its consequences on FeCAP and transistors dimensioning are explored.

2.3.1 1T Scaling at the 28nm Node

2.3.1.1 Memory Window Simulation

In order to demonstrate the potential for 1T-1C FeRAM scaling towards arrays with thinner gate oxide transistors on more aggressive nodes like 28FDSOI [89], one has to evaluate the impact of 1T scaling on the 1T-1C operation. In this part, the elementary circuit consisting of the 1T-1C bitcell, its periphery transistors and the Bit Line capacitance is used to assess the MW(0σ) performance in various conditions. In Figure 2.8.a., the MW(0σ) evolution in a 16kbit array with transistors at the 130nm node (with C_{BL} = 188fF) and 0.36μm^2 FeCAPs made of a 10nm thick ferroelectric layer is explored. In extension to the results presented in Figure 2.7, the simulated response to a wider range of operating voltages is reported in this graph. It can be seen in Figure 2.8.a. (black squares) that below 2V operating voltage, the simulated MW(0σ) does not exceed 0.1V: therefore, a very precise sensing scheme is needed to sense this read margin and discriminate a '0' from a '1' state. In order to evaluate the MW(0σ) performance at advanced nodes, GO2 transistors at the 28nm node are substituted to the 130nm node transistors used so far in both experiments and

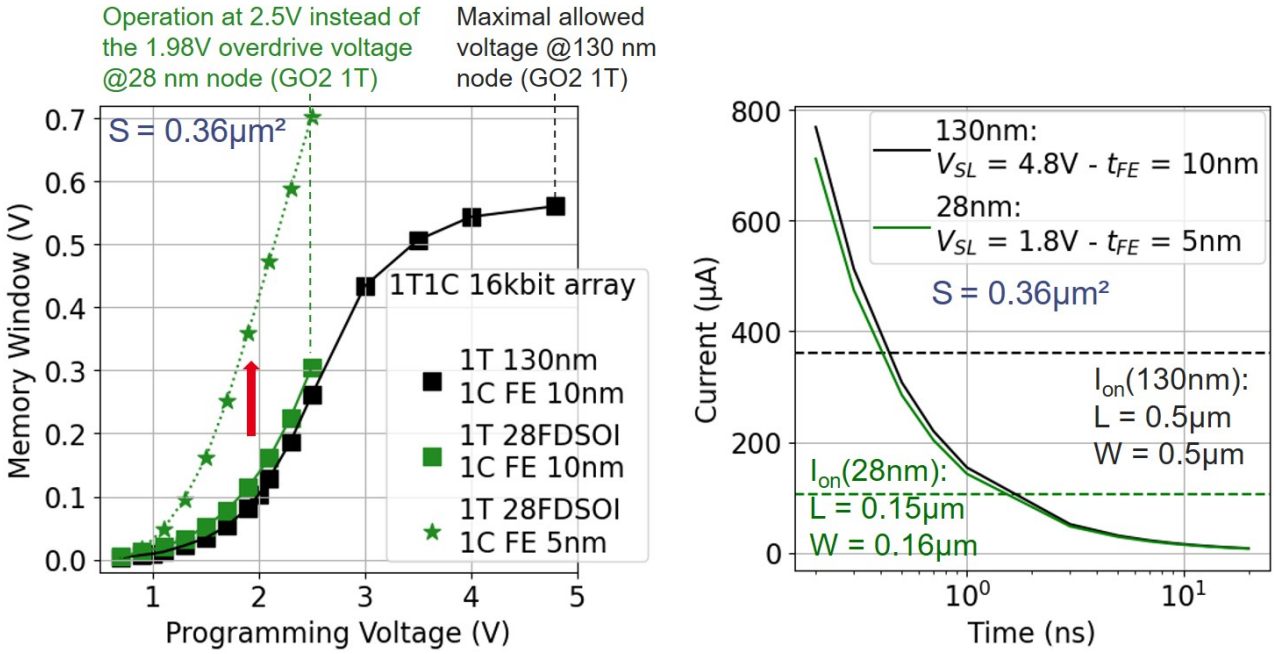


Figure 2.8: a. Simulated MW(0σ) versus voltage in an elementary circuit emulating a 16kbit 1T-1C FeRAM array with $0.36\mu m^2$ FeCAPs and 1T at the 130nm and 28nm nodes. At the 28FDSOI node, MW(0σ) improvement at low voltage with a 5nm thick FeCAP is shown. 28FDSOI 1T gate length and gate width are 150nm and 160nm, respectively (GO2 type). ; b. Calculations of the current flow generated by a combined dielectric and ferroelectric charge in time. On-current of 1T at the 130nm and 28nm nodes in their minimal dimensions are represented crossing the charge-induced current at 0.5ns and 2ns, respectively.

simulations. Similarly to Section 2.2, both C_{GS} and the parasitic capacitance are assessed using the foundry model from STMicrolectronics for their 28FDSOI transistor and a 16kbit array designed at the 22nm node, respectively. Considering 28FDSOI 1T in 16kbit arrays scales C_{BL} down to 89fF for thick gate oxide transistors (GO2). This represents roughly half the value obtained at the 130nm node and accounts for a reduction in both the 1T dimensions (down to 10 times) and the parasitic capacitance (38%) in a 16kbit circuit at advanced 1T nodes. In such circuits at advanced nodes, a lower operating voltage is imposed (overdrive voltage 2.5V is used at the 28nm node, instead of the 4.8V nominal voltage at the 130nm node for GO2 transistors) to protect the gate dielectric against breakdown. As can be observed in Figure 2.8.a. (black squares), scaling the 1T to the 28FDSOI node (C_{BL} = 89fF) increases only slightly the simulated MW(0σ) compared to the circuit with 130nm 1T. C_{BL} reduction at advanced nodes does not compensate for a rapidly decreasing MW(0σ) at low voltage due to the FeCAP subloops behavior (see Figure 2.5.e.). Improving the FeCAP performances at operating voltages below 2.5V is therefore needed.

To efficiently increase the memory performance in FeRAM arrays at advanced nodes, an increase of the FeCAP remanent polarization could be targeted. However, and since operation at low voltage is imposed at advanced nodes, the gain at the edge of the ferroelectric current distribution is

expected to be limited, thus limiting the achievable memory window. A more interesting strategy would be to reduce FeCAP coercive voltages, so as to obtain a significant $2.P_R$ gain at low voltage. Indeed, given the current Lorentzian shape and sharp dI/dt slope reported in Figure 2.3.a., a small shift of the current peak towards lower voltages would result in a largely increased integrated charges area. To verify the effect of coercive voltage scaling on the MW(0σ) at low voltage, a FeCAP with 5nm HfO_2 layer and unchanged $2.P_R$ is simulated in the circuit with advanced 1T. Indeed, from [90], coercive fields (E_c) are constant with ferroelectric layer thickness (t_{FE}). Hence, lower coercive voltages (V_c) can be obtained by reducing t_{FE} ($E_c = V_c/t_{FE}$). All other fitting parameters of the FeCAP model remain unchanged compared to the values reported in Table 2.1. In particular, the simulated FeCAP areas remain constant at 130nm and 28nm nodes. As shown in Figure 2.8.a. (green stars), the simulated MW(0σ) is improved at low voltage compared to a 10nm HfO_2 layer, thus relaxing the constraints on the reading system at low voltage in 16kbit arrays at the 28nm node. Therefore, overcoming the challenge of developing 5nm thick ferroelectric HfO_2-based films at BEoL-compatible temperature with large $2.P_R$ would lead to devices and arrays fully functioning at 2.5V and below with an increased MW(0σ).

Recap

Simulations in the elementary circuit with constant FeCAP coercive voltage (10nm thick) show that no significant MW(0σ) gain is obtained when adopting transistors at the 28nm node in the bitcell and periphery.Taking full advantage of bitcell shrink requires to scale FE layer thickness. Indeed, simulations show that reducing the FE layer thickness to 5nm would lead to an enhanced MW(0σ) down to 1V operating voltage, thanks to effective coercive voltage scaling.

2.3.1.2 Switching Dynamics Considerations

In order to account for FeCAP switching dynamics, this second part focuses on the tradeoff between ferroelectric switching current and time at the 130nm node, and a projection of this tradeoff at the 28nm node. The objective is to assess whether the on-current of scaled transistors at advanced nodes allows for rapid and saturated polarization switching, or rather limits the FeCAP switching time (or switched charge in a given time frame). To this end, basic calculations are conducted to determine the current flow (I) generated by a ferroelectric and dielectric charge (Q_{tot}) over a large range of switching times (t from 0.5ns to 50ns), using Equation 2.15:

$$I = \frac{Q_{tot}}{t} = \frac{2.P_R S + C_{DE} V_{SL}}{t} \tag{2.15}$$

In Figure 2.8.b., V_{SL} was fixed at 4.8V or 1.8V to represent the nominal voltage of GO2 transistors at the 130nm or 28nm node, respectively. Consistently with simulation results from Figure 2.8.a., the FeCAP ferroelectric layer thickness was chosen at either 10nm or 5nm depending on the transistor node emulated (130nm or 28nm). In both cases, the remanent polarization was fixed at $30\mu C/cm^2$ and the FeCAP area was $0.36\mu m^2$. Since the V_{SL}/t_{FE} ratio does not vary much between the two nodes emulated, the current calculated from Equation 2.15 is very similar in the two FeCAP

dimensioning conditions. However, the calculated current has to flow through a transistor, whose I_{on} current depends on its technology node, as well as its dimensions. In the case of an NMOS transistor at the 130nm node with L = W = 500nm, the I_{on} current calculated determines that the total charge (encompassing ferroelectric switching charge and dielectric charge) can flow in half a nanosecond time frame. However, in a transistor at the 28nm node with minimal dimensions (L = 150nm and W = 160nm), an equivalent charge will require a nanosecond to flow through the channel. According to latest results from Sony [91] and Micron [92], the reported FeCAP switching speed in FeRAM arrays are not faster than 20ns. While there clearly is an impact of 1T scaling on FeCAP switching dynamics, the switching times considered are still very fast, therefore 1T scaling does not limit ferroelectric switching times. Moreover, if faster switching time is required at the 28nm node, I_{on} can always be increased by increasing 1T width.

Recap

While the on-current of a scaled transistor at the 28nm node is reduced compared to a transistor at the 130nm node, it nonetheless allows for a FeCAP switching time in the nanosecond range. Therefore, 1T scaling is not a limiting factor for FeRAM in terms of switching dynamics, even when using a GO2 28nm transistor in the bitcell.

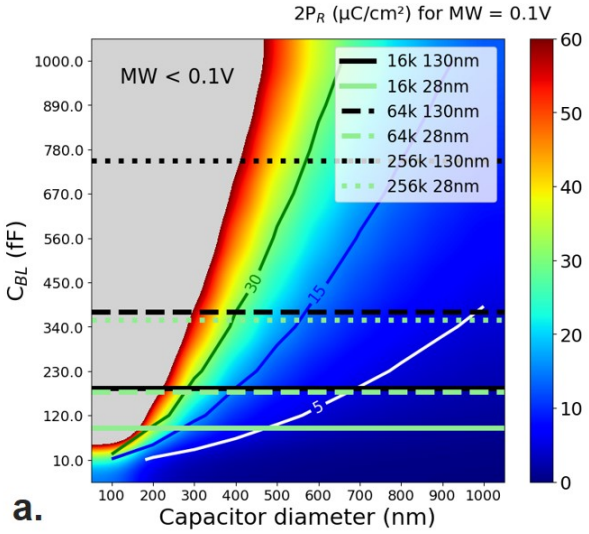
2.3.2 Towards Larger and Denser FeRAM Arrays

In Paragraph 2.3.1, the specific case of a 16kbit FeRAM array was simulated with transistors at 130nm or 28nm technology nodes, and FeCAPs with 10nm or 5nm thick ferroelectric layers. The limited voltage available at the 28nm node led to the global recommendation of reducing coercive voltages (via FeCAP thickness scaling) to improve the MW performance. It was then calculated that adopting scaled transistors at the 28nm node was not limiting for FeCAP switching dynamics. This second part is focused on assessing the advantages and constraints brought by transistors at advanced nodes in larger and/or denser FeRAM arrays. Here, the FeCAP compact model and elementary circuit are not used, and the dimensional analysis provided comes from simple analytical calculations based on the following MW equation:

$$MW = \frac{2.P_R \cdot S}{C_{DE} + C_{BL}} \tag{2.16}$$

As indicated already, Equation 2.16 describes a median behavior, therefore this part focuses on median MW performance and does not account for any process- or design-induced variability across the FeRAM array.

Figure 2.9.a. represents the $2.P_R$ (colorbar) needed to maintain a constant 0.1V MW in a square FeRAM array of given Bit Line size (described by C_{BL}, y-axis) with FeCAPs of given diameter (x-axis). In this graph, $2.P_R$ is computed from Equation 2.16, with MW being equal to 0.1V. In a 16kbit array at the 130nm (28nm) node, C_{BL} is 188fF (89fF), as assessed in Section 2.2. From this graph, two general observations can be made:



b.

130 nm node			
BL size	128 WLs	256 WLs	512 WLs
C_{BL} (fF)	188	376	752
$\emptyset_{FeCAP}$ (nm)	400	550	800
28 nm node			
BL size	128 WLs	256 WLs	512 WLs
C_{BL} (fF)	89	178	356
$\emptyset_{FeCAP}$ (nm)	250	400	550

Figure 2.9: a. Projection of the remanent polarization needed to maintain a 0.1V MW depending on capacitor diameter and array size (via C_{BL}). C_{BL} values at different nodes and for different array sizes were computed with C_{BL}(16k, 130nm) = 188fF and C_{BL}(16k, 28FDSOI) = 89fF. ; b. Corresponding table presenting an approximated minimal FeCAP area to maintain a 0.1MV with 2.P_R fixed at 15$\mu C/cm^2$ and variable 1T technology node and array sizes.

- Increasing the Bit Line size (*i.e.* increasing the number of Word Lines) at the same technology node increases the total gate-source capacitance and thus C_{BL}. This is represented by black lines in Figure 2.9.a. for arrays at the 130nm nodes. Therefore, FeCAPs with larger diameters are needed at constant 2.P_R to maintain a 0.1V MW in larger arrays, which is in drastic contradiction with FeCAP footprint scaling required in high density arrays. The top part of the table presented in Figure 2.9.b. summarizes this trend for a 2.P_R chosen at 15$\mu C/cm^2$ from experimental results reported in Figure 2.5.a. at 2V operating voltage. Therefore, in the current operation mode, larger and denser arrays seem to constitute opposing trends.

- Comparing black (130nm 1T) and green (28nm 1T) lines in Figure 2.9.a., it can be inferred that the reduced C_{BL} in 28FDSOI FeRAM arrays enhances FeCAP scaling capability and favors array size expansion compared to 130nm arrays. A 0.1V MW in 256kbit arrays based on 550nm diameter FeCAPs with a realistic 2.P_R is achievable at the 28FDSOI node (Figure 2.9.b.). However, this kind of calculation being quite simplistic, it does not enable the reproduction of the limited MW gain obtained at low voltage with 28nm 1T, as was simulated in Paragraph 2.3.1.

Recap

Simple calculations using an analytical MW equation indicate the existence of a tradeoff between FeRAM array size (more specifically BL size) and density (here FeCAP footprint). The development of 3D capacitors and/or banked arrays could solve this tradeoff. Adopting transistors at the 28nm node should not negatively impact MW performance, especially in larger arrays with scaled FeCAPs.

2.3.3 Recommendation for 1T-1C Bitcell Scaling towards Advanced Nodes

The objective of this part is to issue recommendations for 1T and 1C scaling, while bearing in mind the specific constraints faced by FeRAM array for integration at advanced nodes. Consistently with results from previous parts, three main axes are considered: effective adoption of transistors at the 28nm node, expansion of FeRAM array size, and improvement of FeRAM array density. While the exploration of these axes leads to different recommendations for 1T and 1C dimensioning that can be interconnected (and sometimes incompatible), an attempt at giving clear guidelines for each axis is pursued.

Adopting transistors at the 28nm node will have the two-fold effect of limiting the available voltage in order to prevent the dielectric breakdown of their thin gate oxide, and, according to their dimensions, will exhibit a reduced on-current compared to their 130nm counterparts. In the 1T-1C bitcell, this effect can negatively impact the FeCAP performance: limited voltage can cause unsaturated switching, and therefore reduced remanent polarization and MW in a FeRAM array ; reduced on-current can degrade the FeCAP ferroelectric switching speed, which otherwise follows intrinsically very rapid dynamics. To address the first point, simulation results have shown that reducing FeCAP coercive voltages via ferroelectric layer thickness scaling constitutes a promising strategy to increase the achievable remanent polarization at limited voltages. In this view, FeCAP stack optimization becomes of high interest to maintain a reasonable ferroelectric performance in very thin layers. Regarding the second point, quick calculations have indicated that 1T-limited maximal FeCAP switching speed is roughly multiplied by 10 when replacing transistors at the 130nm node by transistors at the 28nm node. Although this potential reduction in switching speed is significant, it remains below the tens of nanoseconds range and can still be improved by increasing the transistor active area.

Expanding the array size is necessary to advance from a 16kbit demonstrator to larger product-oriented arrays. The chosen reading scheme relies on capacitive transfer between the FeCAP and the Bit Line capacitance, C_{BL}, so that the memory window is inversely proportional to C_{BL}. Increasing the array size along Bit Lines leads to an increased C_{BL} value, and therefore a degraded MW. To explore the potential for array size expansion, different strategies can be envisioned: array expansion across Bit Lines (*i.e.* along Word Lines), or C_{BL} artificial reduction through sub-division of the FeRAM array into smaller memory banks are realistic solutions developed in the literature [91][93]. It is worth noting that adopting transistors at the 28nm has the happy benefit of reducing C_{BL} altogether, thus favoring FeRAM expansion compared to arrays integrated at the 130nm node.

High-density is a critical target for FeRAM technology in order to reduce the array footprint and fabrication cost. As mentioned before, adopting transistors at the 28nm node will dramatically reduce their dimensions with a controlled impact on FeCAP performance. However, to aim for high-density, FeCAP footprint needs to be scaled, which has the downside of reducing the switching charge, and therefore the achievable memory window. It is therefore necessary to increase the FeCAP area while reducing its footprint, hence the interest of developing 3D capacitors. While intensive design efforts should be directed at scaling the 1T-1C bitcells and their pitch, parallel integration efforts should aim at depositing the FeCAP in a 3D structure, as was initially proposed by [94] and further demonstrated for antiferroelectric capacitors in [95].

Conclusions

This chapter aimed at proposing well-argued guidelines for transistor and FeCAP dimensioning towards 1T-1C FeRAM array integration at advanced nodes. To this end, several tools were adopted and enabled to explore quantitatively and qualitatively the consequences of 1T scaling to the 28nm node, and to assess the feasibility of array size expansion and improved array density requirements.

The first tool developed was a Preisach-inspired FeCAP compact model. While the history of modelling hysteretic behavior dates back to the 1930's Landau and Preisach theories, its wider adoption for ferroelectric capacitor modelling followed the successive development of PZT- and HfO_2-based FeCAPs for ferroelectric memories. Similarly to Preisach-based models in the literature, the model developed aimed at being computationally-efficient. Contrary to most Preisach-based models, it aimed at describing the ferroelectric current evolution rather than the polarization. The complete model, calibrated on experimental data, enables accurate current simulations for a large variety of input signals. In particular, it reproduces very well the remanent polarization scaling with decreasing operating voltage, due to sub-saturation switching. An elementary circuit was accurately dimensioned around the 1T-1C bitcell to emulate 16kbit array operation, which allowed for a good reproduction of the MW degradation at low voltage at the 130nm node. Adopting transistors at the 28nm node has been simulated to have a limited impact on the MW in the available voltage range. Recommendation has been issued to reduce the FeCAP FE layer thickness to 5nm to maintain a 0.1V MW below 2V operating voltage. A considerable challenge raised by this recommendation is to achieve the fabrication of very thin FE layers without degrading their remanent polarization. Through basic calculations, it has been assessed that scaling of 28nm 1T to minimal dimensions would slightly impede ferroelectric fast switching capability, but that the nanosecond range achievable was still reasonable for FeRAM array operation. Depending on the application targeted, the transistor dimensions can be adjusted to allow for quicker current flow.

Transistors at advanced nodes (and consequently of reduced dimensions and gate-source capacitance) have a largely beneficial effect on the development of denser and larger FeRAM arrays. However, the constraints then fall on the FeCAP, and specifically on its footprint reduction to achieve higher density. As a general rule, further analytical calculations have hinted that attempts at developing denser (scaled FeCAP diameter) and larger (increased C_{BL}) arrays would lead to an unrealistically high $2.P_R$ required to maintain a 0.1V MW. One solution envisioned to maintain a large memory window in larger arrays while favoring high density is to develop 3D FeCAPs.

As a conclusion, it can be stated that stack optimization towards lower ferroelectric layer thicknesses, as well as joint design and integration efforts, can lead to the promising development of FeRAM arrays at the 28nm node. In the following chapters, both median and array-level memory windows will be optimized in this regard, and FeRAM array reliability will be explored before integration at advanced nodes can be demonstrated.

3 Optimization of Ferroelectric Stack and Assessment of FeRAM Array-Level Variability towards Advanced Nodes

Introduction

The previous chapter aimed at proposing recommendations regarding the transistor (1T) and ferroelectric capacitor (1C, or FeCAP) dimensions for high-performance FeRAM arrays at advanced technology nodes. Thanks to different tools, such as dimensional analysis and projections with either relevant analytical equations or circuit simulations using a FeCAP compact model, the median memory window (MW(0σ)) evolution was studied. To improve MW(0σ) in FeRAM arrays, an interesting and direct axis can be to improve the remanent polarization (2.P_R) of the FeCAP. In the literature, several options have been explored and lead to a boosted 2.P_R in HfO_2-based FeCAPs. Among them, varying the doping element and concentration, the deposition technique or the electrodes has been shown to enhance the ferroelectric (FE) performances. Exploring other levers, a first part of this chapter will be dedicated to the evaluation of two techniques that have been demonstrated to improve 2.P_R: etching the ferroelectric layer together with the Top Electrode (TE) during the fabrication of the FeCAP, and inserting an oxygen scavenging layer between the ferroelectric layer and the top electrode. In addition to the material engineering and improved process steps that can boost 2.P_R at FeCAP level, and bearing in mind the requirements towards FeRAM array integration at advanced nodes, some technological achievements were realized. Hence, a second part of this chapter will focus on the compared performances of 1T-1C FeRAM arrays integrated in 130nm CMOS, with a classic voltage-based Sense Amplifier (SA), and with an alternative charge-based SA. In addition to that, Short Loop (ShL) integration with 3D capacitors or with a laser annealing step will continue to offer an overview of the considered options for a readable MW(0σ) in large FeRAM arrays at advanced nodes. Lastly, from a product perspective, it is important to consider not only MW(0σ) but also the array-level MW, or MW(16k). Indeed, MW(16k) can be reduced compared to MW(0σ) due to the effective bitcell-to-bitcell variability at FeRAM array level. A third part of this chapter will thus evaluate the different factors that contribute to this variability, and propose fixes when possible.

Summary

3.1 Remanent Polarization Improvement . . . 73
3.1.1 Impact of Ferroelectric Layer Etching . . . 73
3.1.2 Impact of Oxygen Scavenging Layer . . . 76
3.2 Technological Achievements towards Advanced Nodes . . . 79
3.2.1 Large-Sized Arrays . . . 79
3.2.2 Charge Sensing Reading Scheme . . . 82
3.2.3 3D Capacitors for Enhanced MW at Constant Footprint . . . 84
3.2.3.1 Integration . . . 84
3.2.3.2 Performance . . . 86
3.2.4 Lower Ferroelectric Layer Thickness for Low Voltage Operation . . . 88
3.2.4.1 Voltage and Annealing Temperature Limitations . . . 88
3.2.4.2 Role of Nanosecond Laser Anneal for Crystallization of Thin Films . . . 89
3.3 Bitcell-to-Bitcell Variability at FeRAM Array Level . . . 90
3.3.1 '0' and '1' States Distributions . . . 91
3.3.1.1 Assumption of Normality . . . 91
3.3.1.2 A Specific Counter-Example: Extrinsic 8×8 Pattern . . . 92
3.3.2 Intrinsic Variability Sources: Impact and Possible Fixes . . . 94
3.3.2.1 Process-Induced Variability . . . 94
3.3.2.2 Access Devices and Sense Amplifier Variability Along a Bit Line . . . 96
3.3.2.3 Bit Line Voltage Gradient Across Bit Lines . . . 97
3.3.2.4 Double Issue for Large-Sized Arrays . . . 100

3.1 Remanent Polarization Improvement

As FeCAP $2.P_R$ is a key contributor to MW(0σ) in a FeRAM array, increasing $2.P_R$ through stack engineering is a goal many groups work towards. As was presented in [57], the presence of ferroelectric properties in ALD-deposited silicon-doped hafnium oxide (HSO) was conditioned to an optimized doping content, and a high-temperature crystallization anneal in the presence of a top electrode. The polar orthorhombic phase $Pca2_1$ was identified to be responsible for ferroelectricity. Later works on HfO_2-based ferroelectric materials have evidenced the critical role of many parameters in the presence of this phase, and thus a ferroelectric response and its intensity. To name a few of these multiple factors, one can cite the doping element(s) and content [96, 97, 98, 99], oxygen vacancies content [100, 101], deposition or doping technique [102, 66], electrodes and interfaces [103, 104, 105], strain [106, 107, 108, 109] and annealing procedure [110, 111, 112]. Many of these factors probably contribute to the ferroelectricity in an intertwined way, which makes it difficult to determine the optimal process for enhanced $2.P_R$ in a FeCAP ; moreover, $2.P_R$ improvement may lead to a degraded reliability, and therefore stack engineering has to be carefully tailored for the targeted application.

With that in mind, some results regarding the optimization of the FeCAP stack for an improved $2.P_R$ are presented in the following sections, and their impact on FeCAP endurance is assessed. Starting from a TiN/HSO (10nm)/TiN Process of Reference (PoR) ferroelectric stack, variations of strain and oxygen vacancies content were aimed for in order to provide additional boosts to the FeCAP $2.P_R$. Therefore the primary focus of this study is directed towards two main levers: strain modulation by etching the ferroelectric layer, and scavenging oxygen from the ferroelectric layer.

3.1.1 Impact of Ferroelectric Layer Etching

During the integration process of a FeCAP, as is the case in the considered 200mm Memory Advanced Demonstrators (MAD), the top electrode is patterned down to the ferroelectric layer after deposition to define the capacitor electrical area. This step enables to electrically separate different capacitors, while the ferroelectric layer remains unpatterned in the PoR stack. For FeCAPs scaled at sub-µm dimensions, processing steps may induce mechanical strain variations impacting device properties. In order to evaluate the influence of the processing steps (and therefore the possible influence of strain) on the intensity of the ferroelectric response, the effect of ferroelectric layer etching on FeCAPs integrated in the Back-End-of-Line (BEoL) of transistors the 130nm node is assessed.

As a reference, a STEM cross-section image of a non-etched FeCAP (*i.e.* a stack with an unpatterned FE layer) comprising a 10nm-thick silicon-doped HfO_2 layer between TiN electrodes and integrated into 130nm process is presented in Figure 3.1.a. It can be seen in white that the FE layer is not etched contrary to top electrode, and that it is continuous between adjacent integrated FeCAPs. On a different wafer with an identical TiN/ HSO (10nm)/TiN stack, both the Top Electrode (TE) and the ferroelectric layer of BEoL-integrated FeCAPs and FeRAM arrays were patterned. Final anneal was performed after M5 metallization to enable crystallization in the polar orthorhombic phase. STEM cross-section imaging (Figure 3.1.b.) reveals a clear discontinuity of the Si:HfO_2 layer, indicating that the etching of both TE and FE layers was successfully performed. To further analyze the elemental composition of the stack, the Energy-Dispersive X-ray (EDX) spectroscopy technique

was used in thin cross-sections of the ferroelectric capacitors. [1] The EDX spectrum of the FeCAP with an etched ferroelectric layer (also called etched FeCAP) is presented in Figure 3.1.c. and reveals clean etched walls, elemental composition consistent with the desired TiN/Si:HfO_2/TiN stack, as well as homogeneous TiN and Si:HfO_2 layers. Similarly, several wafers with a TiN/$Hf_{0.5}Zr_{0.5}O_2$ (ALD-deposited) /TiN stack were manufactured. Among them, the FE layer thickness was varied (7nm and 10nm), and both etched and non-etched FeCAP process options were pursued.

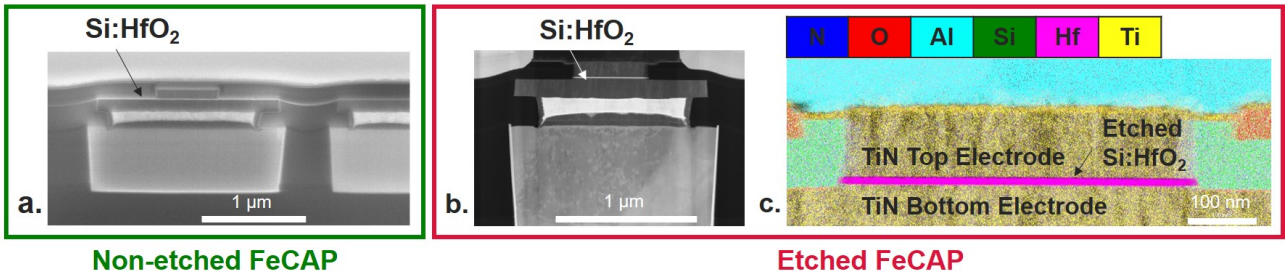


Figure 3.1: a. Close-up STEM imaging of BEoL-integrated TiN/Si:HfO_2 (10nm)/TiN stack with etched top electrode to define 550nm diameter FeCAP; b. Close-up HAADF-STEM imaging of etched FeCAP showing vertical etched walls of Si:HfO_2 layer ; c. EDX analysis of etched FeCAP shows uniformity of Si:HfO_2 layer

The ferroelectric performance of non-etched and etched FeCAPs in capacitor fields and FeRAM arrays integrated in BEoL 130nm was electrically assessed at wafer level on different stacks: TiN/HSO (10nm)/TiN, TiN/HZO (7nm)/TiN and TiN/HZO (10nm)/TiN. For each stack, capacitor fields consisting of 1290 parallely-routed scaled FeCAPs were measured on over 30 dies across the wafers. PUND sequences at 4MV/cm - 10kHz were applied to extract $2.P_R$ during cycling at 4MV/cm - 100kHz. On HSO-based wafers, the '0' and '1' logic states distributions of 16kbit 1T-1C FeRAM arrays were measured by square pulses at 4.8V - 2µs after the application of 10^4 wake-up cycles with similar pulses.

Figure 3.2.a. represents the post-wake up (10^4 cycles) Polarization-Voltage hysteresis loops measured on HSO-based non-etched (green) and etched (red) FeCAP fields. They were measured at the same die position on two different wafers and are representative of the ferroelectric performance at wafer scale. In these Si:HfO_2-based BEoL-integrated scaled FeCAPs, the ferroelectric layer etching enables the improvement of $2.P_R$ from 20 to $27\mu C/cm^2$. On these HSO-based wafers, etching the ferroelectric layer also reduced $2.P_R$ die-to-die variability after wake-up. At FeRAM array level, the improved ferroelectric performances evidenced on FeCAPs have been confirmed, as shown in Figure 3.2.b. Indeed, etching the FeCAP has been demonstrated to increase MW(0σ) from 0.22V (green) to 0.34V (red), as well as triple MW(16k) from 0.04V to 0.12V. A preferred hypothesis for these improved performances is that etching generates a more favorable and localized microscopic strain on the FE layer, thus favoring crystallization in the orthorhombic phase. The role of strain in ferroelectricity enhancement has otherwise been evidenced in many publications, for example

[1] This technique enables the identification of elements in a sample based on the X-rays they generate following exposition to an incident electron beam.

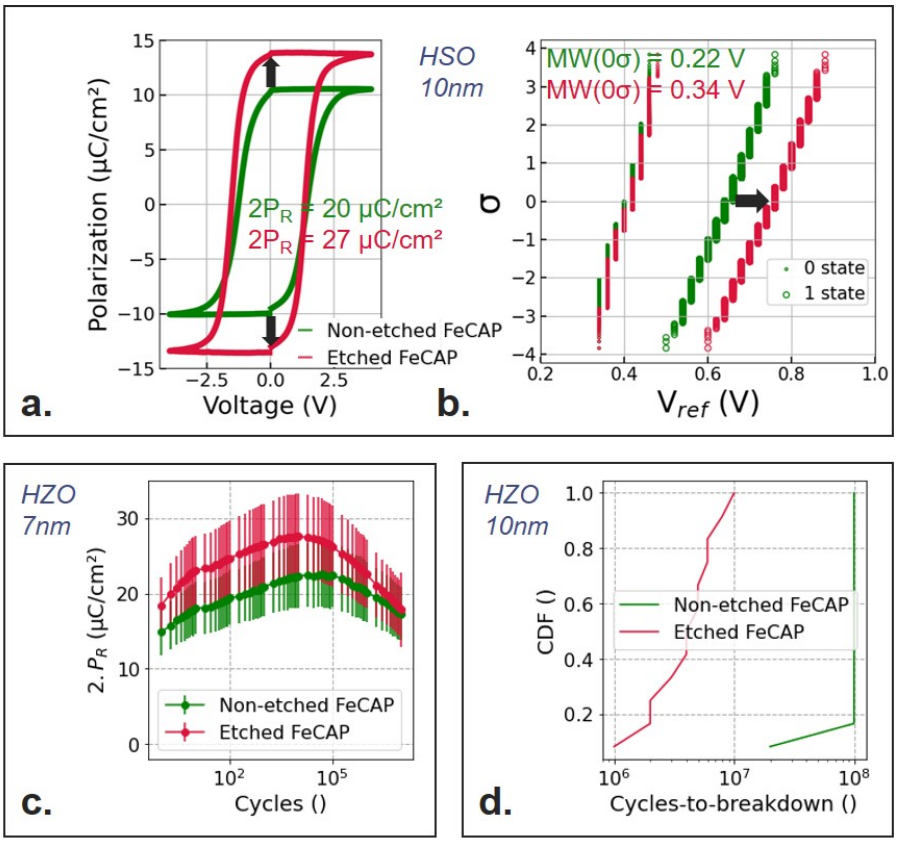


Figure 3.2: On non-etched (green) and etched (red) FeCAPs integrated in BEoL 130nm: a. Polarization-Voltage hysteresis loops measured after 10^4 wake-up cycles at 4MV/cm on 10nm HSO-based 306μm^2 FeCAP fields ; b. '0' and '1' distributions measured at 4.8V on 10nm HSO-based 16kbit FeRAM arrays with 0.24μm^2 FeCAPs ; c. 2.P_R-Cycles median and standard deviation measured at 4MV/cm on 76 7nm HZO-based 306μm^2 FeCAP fields ; d. Cycles-to-breakdown distributions measured at 4MV/cm on 10nm HZO-based 250μm^2 FeCAP fields.

through the so-called 'clamping effect' [57, 108] provided by the top electrode, or else through the existence of an optimal strain level to stabilize the ferroelectric phase [107, 109].

In Figure 3.2.c., an equivalent improvement in FeCAP 2.P_R with ferroelectric layer etching is observed on another lot comprising wafers with a 7nm-thick HZO layer deposited by ALD between TiN electrodes. Indeed, in similar testing conditions, the median 2.P_R after 10^4 wake-up cycles across 76 dies on the wafer was measured at 22$\mu C/cm^2$ or 28$\mu C/cm^2$ for the devices with a non-etched (green) or an etched (red) ferroelectric layer, respectively. No significant difference in die-to-die variability was evidenced, probably caused by the fact that a higher homogeneity can generally be reached with HZO compared to HSO thanks to HZO's lower crystallization temperature, thus reducing variability. Up to 10^7 cycles were applied at 100kHz to evaluate the endurance of non-etched versus etched devices. While no dielectric breakdown is reached, it is worth noting that a similar 2.P_R level is reached after 10^7 cycles in both cases, indicating a lower endurance in

etched devices. To assess the impact of etching the FE layer on dielectric breakdown, 10^8 cycles were applied to FeCAPs of another lot with a 10nm-thick HZO ferroelectric layer. As a thicker film, 10nm HZO is expected to exhibit a slightly lower breakdown field than 7nm HZO [90], and dielectric breakdown should be facilitated. As shown in Figure 3.2.d., cycles-to-breakdown is measured on 12 nominally identical FeCAP fields for non-etched (in green) and etched (in red) wafers. While breakdown occurs in only one non-etched FeCAP, all etched FeCAPs break down between 10^6 and 10^7 cycles, confirming a reduced endurance in etched devices. This could originate from degraded HZO sidewalls favoring the creation of percolation paths that increase leakage currents and eventually causes dielectric breakdown. If confirmed, this endurance-limiting factor could be largely reduced by improving the etching and/or the FeCAP encapsulation process.

An interesting point to note is that a similar etching process of the ferroelectric layer was performed on 6nm- and 10nm-thick HZO films integrated in Short Loops wafers. 100µm diameter capacitors were cycled at 4MV/cm - 100kHz up to 10^5 times. For both thicknesses, the devices measured on wafers with non-etched and etched FeCAPs, respectively, exhibit absolutely identical properties in terms of remanent polarization, die-to-die homogeneity and endurance. Following our first hypothesis, this could highlight very different levels of strain for different FeCAP sizes, *i.e.* between 100µm and 550nm diameter FeCAPs, thus urging to caution when measuring strain-related metrics on large capacitors for extrapolation to scaled integrated FeCAPs.

Recap

In HSO- and HZO-based scaled FeCAPs, etching the ferroelectric stack during process of top electrode patterning results in an improved remanent polarization. This is thought to be linked to a strain-related effect in the ferroelectric layer, favorable to crystallization in the polar orthorhombic phase following an End-of-Process anneal. This hypothesis is further strengthened by the fact that this effect is not observed on very large (100µm diameter) FeCAPs. However, FE layer etching causes a reduced endurance that could be explained by degraded capacitor sidewalls favoring leakage and eventually leading to dielectric breakdown. This effect is expected to be corrected by proper etching process optimization.

3.1.2 Impact of Oxygen Scavenging Layer

The role oxygen vacancies (V_O) play in ferroelectricity is the focus of active research. Mittmann et al. [100] showed that an optimized V_O content is needed to stabilize the orthorhombic ferroelectric phase and enhance FeCAP 2.P_R. It is worth noting that Resistive Random Access Memory (RRAM) also relies on V_O management in HfO_2, but HfO_2 in its amorphous phase and not the polar orthorhombic phase as is the case for FeRAM. The reversible creation (set) or destruction (reset) of a conductive filament between top and bottom electrodes gives rise to two stable resistance states that can be used as binary bits. The filament is formed by neutral or charged oxygen vacancies that enable controlled current flow through HfO_2 layer: it is therefore easily understandable that a high amount of oxygen vacancies in FeCAP devices could lead to dielectric breakdown.

V_O can be induced in the stack during the process in a specifically controlled environment [113], but are usually present at the interfaces between TiN electrodes and the HfO_2 layer [104, 101]. Indeed, TiO_x or TiO_xN_y interfacial layers can be formed during deposition depending on the process flow steps performed [114]. In the literature, different proportions of oxygen vacancies have been reported at the top electrode and bottom electrode interfaces, highlighting a very process-dependent V_O generation [115, 104, 114].

In Si:HfO_2-based MFM stacks, $2.P_R$ improvement through fine-tuning of oxygen vacancies was attempted. To this end, a Ti metallic layer was deposited between HSO and the top electrode, and was expected to oxidize through oxygen scavenging from HSO. Oxygen vacancies were engineered on purpose in different proportions using different Ti thicknesses, and the effect on ferroelectricity was assessed by structural analysis, defect spectroscopy and electrical characterization. The results shown were presented in [116] and published in [117].

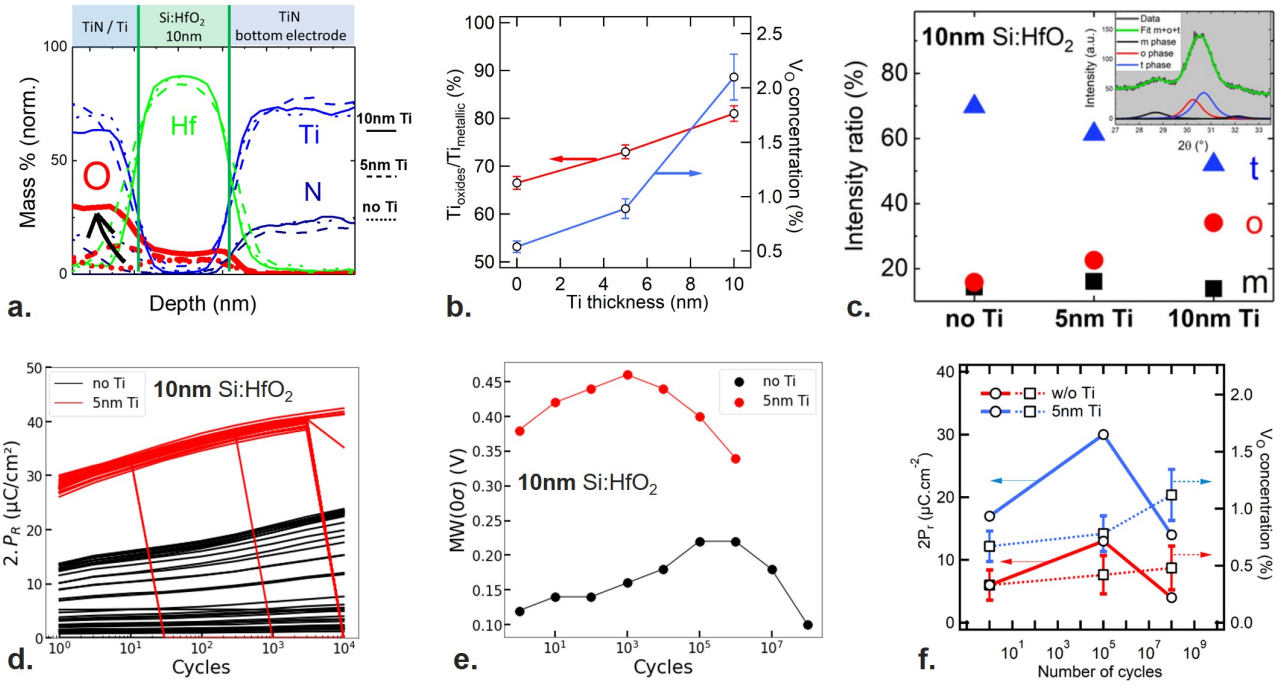


Figure 3.3: From TiN/HSO (10nm)/Ti (0nm-5nm-10nm)/TiN stacks: a. Mass proportions of Hf, Ti, N and O elements across the stack from EDX elemental spectroscopy ; b. Ti oxides and V_O concentrations as a function of Ti layer thickness from HAXPES defect spectroscopy ; c. Relative intensity ratios of tetragonal (t), orthorhombic (o) and monoclinic (m) phases as a function of Ti layer thickness from GIXRD measurement ; d. $2.P_R$-Cycles measured at 4MV/cm on over 30 306μm^2 BEoL-integrated FeCAPs with no Ti (black) or a 5nm-thick Ti (red) layer ; e. MW(0σ)-Cycles measured at 3V on 16kbit FeRAM arrays with 0.36μm^2 FeCAPs with no Ti (black) or a 5nm-thick Ti (red) layer ; f. $2.P_R$-Cycles measured at 4MV/cm on large FeCAPs with no Ti (black) or a 5nm-thick Ti (red) layer and V_O concentration evolution with field cycling from in-situ HAXPES measurement.

To confirm the scavenging effect of the Ti layer in a TiN/Si:HfO_2 (10nm)/Ti/TiN stack, an EDX analysis was performed on full sheet samples with 0nm-, 5nm- and 10nm-thick Ti layers, respectively. The elemental proportions of Hf, O, Ti and N were extracted across the entire stack depth. In Figure 3.3.a., an increase in the oxygen content at the interface with the top electrode is observed as Ti thickness increases, which can be attributed to the formation of a TiO_x layer. To further analyze the content of oxygen vacancies generated by O-scavenging from the Ti layer and its effect on the polar orthorhombic phase, Hard X-ray Photo-Electron Spectroscopy (HAXPES) and Grazing Incidence X-ray Diffraction Spectroscopy (GIXRD) measurements are performed. [2] From HAXPES spectra at two different energies probing the Ti 1s and Hf $3d_{5/2}$ electrons, respectively, it can be seen in Figure 3.3.b. that the proportion of Ti oxides (compared to metallic Ti), together with the content of positively charged oxygen vacancies in HSO increases with Ti thickness, confirming the oxygen scavenging effect of the Ti layer. GIXRD measurements were performed on the same stacks to evaluate the impact of the Ti layer presence and thickness on the monoclinic (m), tetragonal (t) and orthorhombic (o) crystalline phases. The obtained spectra shown in Figure 3.3.c. were fitted using a Voigt deconvolution of the o/t peak visible at 30.5°. The intensity ratios extracted demonstrate that o-phase proportion increases at the expense of t-phase with increasing Ti thicknesses. Combining the results from the HAXPES and GIXRD measurements, it can be inferred that the introduction of an O-scavenging Ti layer generates oxygen vacancies that play a key role in the t- to o- phase transition during the crystallization anneal.

For a definite intake on the ferroelectric performances of MFM stacks with a Ti layer inserted between HSO and the TiN top electrode, electrical assessment needs to be conducted. A reference wafer with a TiN/Si:HfO_2 (10nm)/TiN stack integrated in the BEoL of 130nm CMOS process is compared to another wafer comprising a 5nm-thick Ti layer between HSO and the TiN TE, but was otherwise similarly integrated. Capacitor fields across the wafers were measured using a PUND sequence at 4V - 10kHz during cycling at 4V - 100kHz up to 10^4 cycles. In Figure 3.3.d., the thus-extracted $2.P_R$ was plotted as a function of cycles for all FeCAP fields measured on each wafer. $2.P_R$ is generally much improved when adding Ti, both in the pristine state ($30\mu C/cm^2$ with Ti, compared to $15\mu C/cm^2$ without Ti) and after wake-up ($40\mu C/cm^2$ with Ti, compared to $25\mu C/cm^2$ without Ti). The Ti layer tends to reduce wake-up in the FE stack, without significantly modifying the coercive voltages. It can also be seen immediately that the dispersion of $2.P_R$ across the wafer is much lower for the stack with the Ti layer, indicating an improved die-to-die homogeneity with Ti. However, and this is a an important drawback, FeCAPs with the 5nm Ti layer exhibit a reduced endurance: most (over 70%) of the devices failed due to dielectric breakdown before 10^4 cycles. At FeRAM array level, the '0' and '1' logic states distributions of 1T-1C 16kbit arrays with $0.36\mu m^2$ FeCAPs were measured with a V_{ref} scan procedure at 3V operating voltage. Voltage was reduced compared to FeCAP characterization (4V) to avoid early dielectric breakdown. MW(0σ) evolution as a function of cycles was plotted in Figure 3.3.e. Consistently with FeCAP $2.P_R$ measurements, it can be observed that both pristine and maximal MW(0σ) are increased by the insertion of an O-scavenging Ti layer, but this MW(0σ) enhancement is obtained at the expense of a reduced endurance. In the FeRAM array with a Ti layer, dielectric breakdown occurs at 10^6 cycles whereas a 10^8 cycles endurance is achieved without Ti.

[2] HAXPES relies on synchrotron-generated X-rays to probe defects across the depth of devices like an MFM ferroelectric stack, while GIXRD enables to evidence the crystalline phases present in the stack and their relative proportions.

To understand the mechanisms causing a reduced endurance in the presence of the O-scavenging Ti layer, in-situ HAXPES measurements were performed on large capacitors without Ti and with a 5nm-thick Ti layer. As can be seen in Figure 3.3.f., the decrease in $2.P_R$ during cycling is correlated with an increase in oxygen vacancies concentration, potentially favoring the creation of a conductive filamentary path. On capacitors without Ti, this increase is very limited, whereas the insertion of a Ti layer almost doubles V_O from 10^5 to 10^8 cycles, explaining breakdown-limited endurance.

Recap

The insertion of a Ti layer between the FE and TE layers of a standard TiN/HSO/TiN stack enables oxygen scavenging and leads to the generation of oxygen vacancies in HSO. Ti thickness modulates the concentration of V_O to the point of favoring the orthorhombic phase at pristine state. Electrical characterization on BEoL-integrated FeCAPs and FeRAM arrays confirms improved pristine FE performance with Ti, and evidences early breakdown that is correlated with V_O concentration increase with cycling.

3.2 Technological Achievements towards Advanced Nodes

As was presented in Chapter 2, various limitations regarding the development of large FeRAM arrays integrated at advanced process nodes were identified through the use of dimensional calculations, as well as MW(0σ) projections using either a FeCAP compact model or analytical equations. Among the critically limiting factors identified, the voltage-based reading scheme relying on capacitive transfer from the FeCAP to the Bit Line capacitance (C_{BL}) is expected to degrade the read margin in large arrays. Moreover, the high-density requirement weights heavily on the FeCAP planar footprint reduction, further degrading MW(0σ), and thus the read margin. Finally, the transistors at aggressive nodes tolerate a much lower operating voltage than is the case for 130nm transistors in order to protect their very thin gate oxide from dielectric breakdown. This therefore forces the capacitor in a sub-saturation regime, hence leading to a reduction in $2.P_R$ – and therefore MW(0σ) at FeRAM array level – unless the FeCAP coercive voltages can be sufficiently scaled to reach saturated $2.P_R$ at low voltage.

The purpose of the following sections is therefore to propose effective solutions to overcome the aforementioned limitations. To this end, we present several achievements that demonstrate the feasibility of a few technological options and show the next steps towards array integration at advanced nodes.

3.2.1 Large-Sized Arrays

To move from a 16kbit FeRAM array at the 130nm node as was published in [68] [87] towards a larger 1T-1C FeRAM array at a more advanced process node, it is necessary to first consider the challenges of MW(0σ) performance in larger arrays at the 130nm node. Then, MW(0σ) evolution at the 28nm node will be projected.

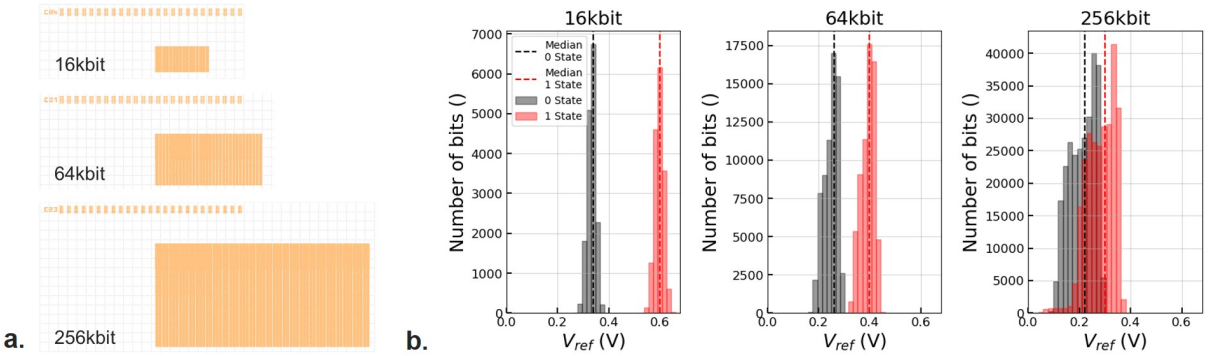


Figure 3.4: a. Layout view of 16kbit (128×128), 64kbit (256×256), 256kbit (512×512) 1T-1C FeRAM arrays at the same scale. Total area comprises the memory array and its periphery, and expands both vertically and horizontally as the number of BL and WL increases, respectively. ; b. Bit density for '0' and '1' logic states measured with 4MV/cm - 2µs pulses after the application of 10^4 wake-up cycles on a 16kbit, 64kbit and 256kbit with 550nm diameter FeCAP on a TiN/HZO (7nm)/TiN stack.

As was mentioned in Chapter 1, the reading scheme adopted consists in sensing the Bit Line (BL) voltage elevation following a capacitive transfer between the FeCAP and C_{BL}. Two main factors contribute to C_{BL}: the largest contribution comes from the gate-source capacitance (C_{GS}) of the selection transistors connected to each considered Bit Line. The size of a BL is defined as the number of transistors connected to it, and is therefore equivalent to the number of WLs. The second contribution stems from a parasitic capacitance coupling occurring between different metal lines in the array, and is related to the design geometry and the integration level, *i.e.* the metal levels between which the FeCAP is integrated. Both C_{GS} and the parasitic capacitance coupling depend on the transistor technology node.

At the 130nm node, FeRAM array expansion along BLs is expected to increase C_{BL}, and hence to decrease MW(0σ) (which is inversely proportional to C_{BL}). This analytical result was verified in 1T-1C FeRAM arrays integrated in the BEoL of 130nm CMOS. Within the framework of this PhD, FeRAM arrays larger than the previously developed 16kbit could be evaluated. A 16kbit (128WLs × 128BLs), a 64kbit (256WLs × 256BLs) and a 256kbit (512WLs × 512BLs) FeRAM arrays were integrated on the same wafer. The FeCAP stack consists of a 7nm HZO layer deposited by ALD on a TiN bottom electrode, followed by the deposition of a TiN top electrode at low temperature (T = 100°C). Only the top electrode was etched to define circular capacitors with a 550nm diameter, and a final End-of-Process thermal annealing was performed at 500°C. All arrays were designed in a similar way, and the layout view of the designed 16kbit, 64kbit and 256kbit arrays is shown in Figure 3.4.a. To assess MW(0σ) performance evolution with array size, V_{BL}^0 and V_{BL}^1 distributions were measured through a V_{ref} scan procedure with square pulses at 4MV/cm - 2µs after the array was programmed entirely in the '0' state and then in the '1' state. Figure 3.4.b. presents the '0' and '1' states bit densities in arrays of all three sizes, and it can be seen that the Bit Line voltages decrease and the distributions broaden as the array size increases. Moreover, MW(0σ) measured

on each array decreases in an inverse trend with BL size, as can be seen between dashed lines. As predicted from the analytical equations and design considerations presented in Chapter 2, this confirms that voltage sensing using C_{BL} limits the achievable MW(0σ) in large arrays at the 130nm technology node.

To overcome this limitation, a possible solution at the 130nm node is to decrease the value of C_{BL}. It was simulated by Sony [70] that reducing C_{BL} from 250fF to 120fF in a 64kbit FeRAM array would increase the read margin by 0.2V for a FeCAP area of 0.6µm². In [91], this solution was executed on a Mbit by subdividing the whole array into smaller 4kbit arrays, enabling an overall C_{BL} reduction to 90fF. Smaller banks to reduce the BL size and further design optimizations to reduce the parasitic capacitance constitute promising paths towards C_{BL} reduction. Note that this solution comes at the expense of FeRAM array footprint scaling, and therefore at the expense of density. As a first step to assess the expected performance of larger arrays with a reduced C_{BL}, a pseudo 688kbit was emulated out of the 16kbit arrays measured on 42 dies of the same wafer at 4MV/cm - 2µs. Their individual '0' (black) and '1' (red) state distributions are presented on a wafermap in Figure 3.5.a. and show a non-zero MW(16k) (green) for all devices. Aggregating these arrays forms a cumulative larger array that tends toward an artificial Mbit array with no C_{BL} increase compared to a 16kbit. It is equivalent to designing a near-Mbit array subdivided into 16kbit banks, apart from the fact that die-to-die variability is expected to limit the overall memory window. The resulting '0' and '1' logic states distributions of the pseudo 688kbit are plotted in Figure 3.5.b., together with those of an individual 16kbit for comparison. The pseudo 688kbit exhibits a 0.28V MW(0σ) and a Bit Error Rate (BER) around 10^{-5} despite still being impacted by die-to-die variability at wafer scale. As a visual representation, Figure 3.5.c. gathers MW(0σ) results from the designed 16kbit, 64kbit and 256kbit (Figure 3.4.b.) and from the pseudo 688kbit showing the clear advantage of the latter.

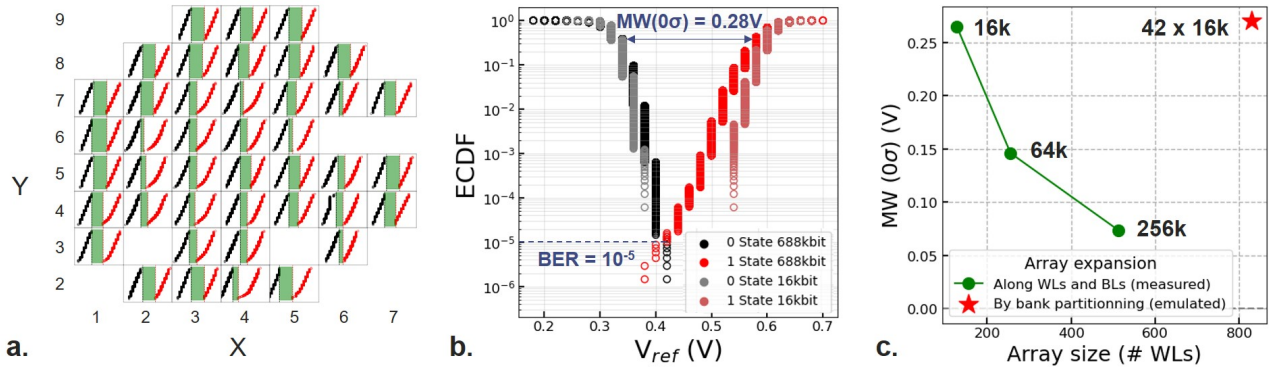


Figure 3.5: a. Wafermap showing '0' and '1' logic states distributions between 0.2V and 0.7V measured with 4MV/cm - 2µs pulses after 10^4 wake-up cycles on 42 1T-1C 16kbit arrays with 550nm diameter FeCAPs across the TiN/HZO (7nm)/TiN wafer ; b. '0' and '1' logic states empirical cumulated distribution function for the aggregated 688kbit showing a 0.28V MW(0σ) and 10^{-5} BER. Comparison with an individual 16kbit evidences the closing of MW(16kbit) due to die-to-die variability. ; c. MW(0σ) versus number of WLs measured on the 16kbit, 64kbit and 256kbit arrays described in Figure 3.4, and an extra point describing the pseudo 688kbit.

Therefore, high-performance large FeRAM arrays could be achieved if MW(0σ) dependence on C_{BL} was reduced (*e.g.* opting for limited bank sizes, or a C_{BL}-independent reading scheme).

To mention a few other parameters that could favor MW(0σ) in larger arrays, it can be expected that, at the same technology node, the parasitic contribution to C_{BL} would decrease in integration options where the FeCAP is closer to the transistor level, *e.g.* Middle-of-Line (MoL) integration would be more favorable than Back-End-of-Line (BEoL) integration. Another consideration is that at advanced nodes, the transistors scaled dimensions result in a reduced gate-source capacitance, as well as a decrease in the overall parasitic capacitance at FeRAM array level. For example, as detailed in Chapter 2, swapping the transistors at the 130nm node in a 16kbit array for transistors at the 28nm FDSOI node will practically halve C_{BL}. Note that, as described in Chapter 2, the allowed operating voltages will be much lower in order to limit the risk of dielectric breakdown of the transistor gate, such that new constraints on the ferroelectric layer thickness will arise at the 28nm node to decrease the coercive voltages.

Recap

The reading scheme measuring BL voltage elevation is highly dependent on C_{BL}. In larger arrays, up to 256kbit, C_{BL} increase is shown to be detrimental to MW(0σ). C_{BL} reduction with design optimizations, such as minimization of parasitic capacitance, subdivision into memory banks, and adoption of transistors at advanced nodes could help maintain a readable MW(0σ) in large arrays. First demonstration of a pseudo 688kbit with no C_{BL} increase compared to a 16kbit augurs very good performances in larger arrays at advances nodes with smart bank partitioning.

3.2.2 Charge Sensing Reading Scheme

We have seen that C_{BL} limitations due to the voltage sensing reading scheme can be pushed back by design innovations and/or the adoption of transistors at advanced technology nodes. A radically different approach would be to completely overcome the dependence on C_{BL}, that is to say not rely anymore on a capacitive transfer between the FeCAP and C_{BL} during the read operation. Within the framework of this PhD, FeRAM arrays with an alternative reading system called a Capacitive Trans-Impedance Amplifier (CTIA) (contrary to the Cross-Coupled Invertor Comparator (CCIC) SA previously developed) could be evaluated. In these CTIA-based FeRAM arrays, a different SA directly senses the switching charge, and is inspired from the reading schemes used for photodetectors [118]. It relies on a two-stage operational amplifier with a feedback capacitance (C_{FB}) and a reset switch in parallel, which provides an analog voltage output that depends on the sensed charge. The CTIA-based reading circuit developed by Billoint et al. [118] is presented in Figure 3.6.a.

During a read operation, the reset switch is first enabled to discharge C_{FB} ($V_{FB} = 0$) and to polarize CTIA output to $V_{ref} - V_{FB} = V_{ref}$. Then, a PL pulse enables a charge transfer from the FeCAP to C_{FB}, and the operational amplifier output is modified from V_{ref} to $V_{ref} - V_{FB}$, where $V_{FB} = Q_{tot}/C_{FB}$. The total charge Q_{tot} depends on the state read, thus enabling the discrimination of '0' and '1' logic states. A chronogram of this operation is presented in Figure 3.6.b. Since a reduction from the

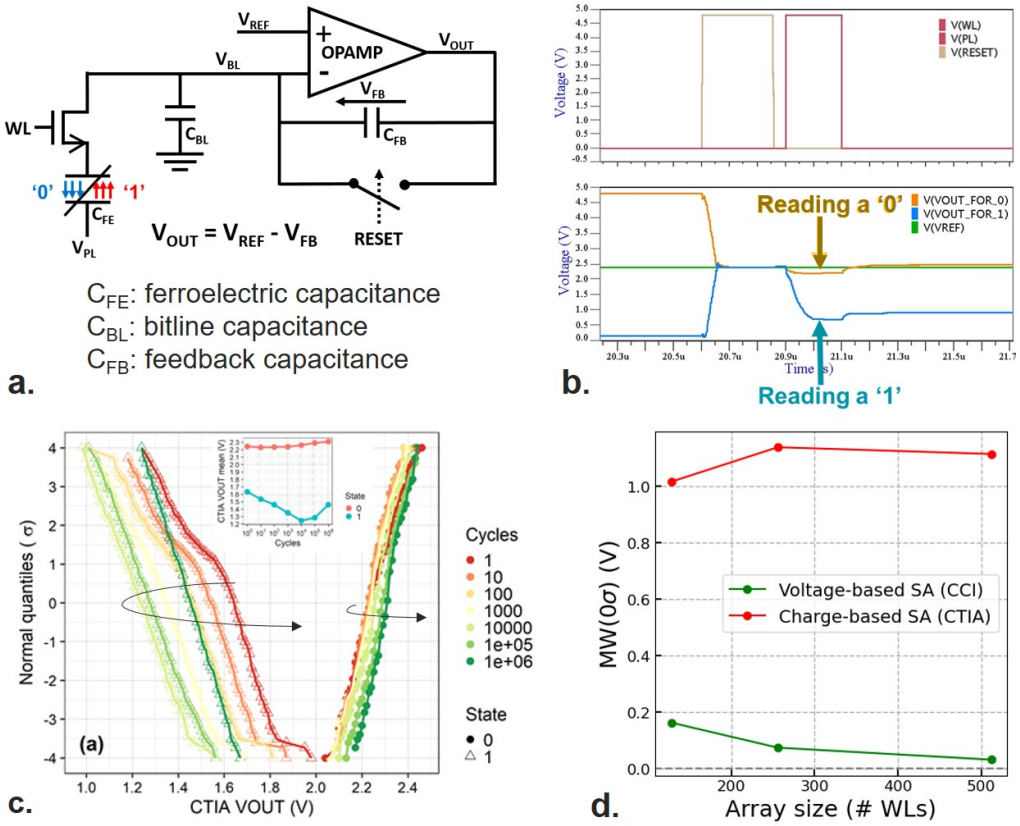
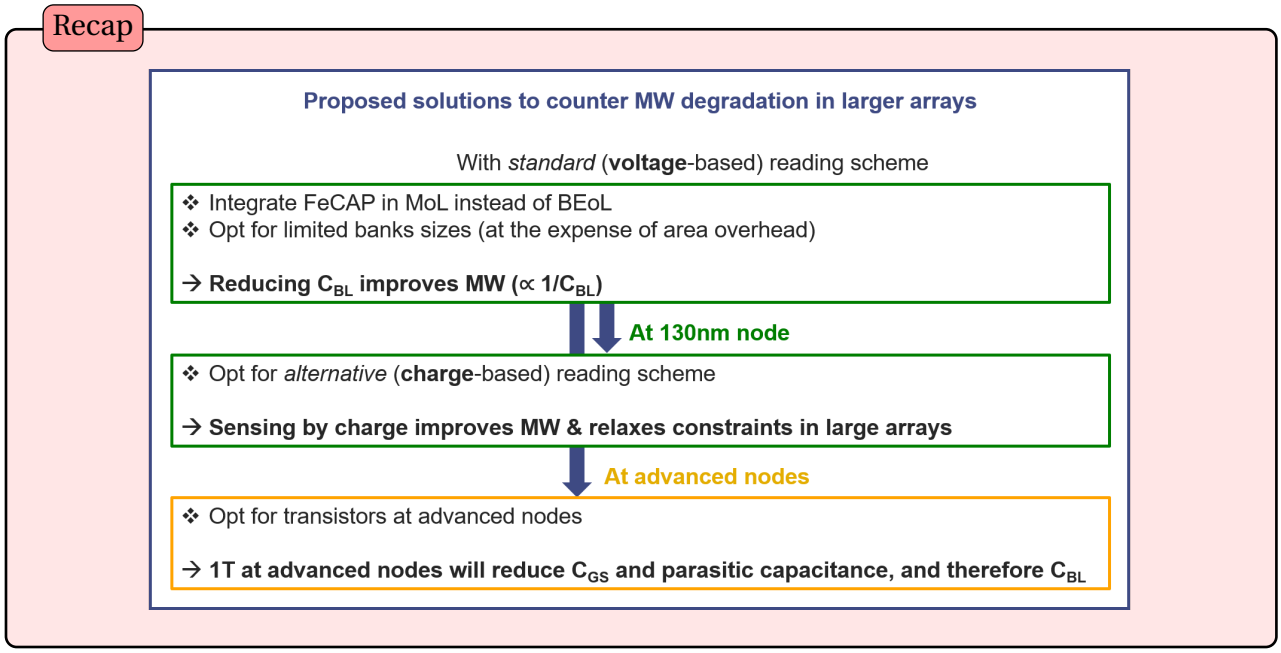


Figure 3.6: a. Electrical circuit of a Capacitive Trans-Impedance Amplifier (CTIA) element. ; b. Chronogram of a read operation in a CTIA-based circuit. ; c. '0' and '1' logic states distributions measured with 2.4V - 62ns pulses during cycling on a 1T-1C 16kbit FeRAM array with 550nm diameter FeCAPs on a TiN/HZO (10nm)/TiN wafer. Inset shows the median output voltage for each state versus cycling. ; d. Post 10^4 cycles MW(0σ) versus number of WLs measured on 16kbit, 64kbit and 256kbit arrays with CCIC and CTIA sensing schemes.

reference voltage is sensed, the '0' and '1' state distributions are reversed compared to the standard CCIC SA. Distributions are presented in Figure 3.6.c. on a 16kbit FeRAM array with 550nm diameter FeCAPs with a TiN/HZO (10nm)/TiN stack. The inset presents the median output voltage for both logic states, and evidences classical wake-up and fatigue behaviors.

To verify the efficiency of the CTIA in terms of reading performance in large arrays, 16kbit, 64kbit and 256kbit arrays based on a 10nm HZO stack were programmed at 4V and characterized at 2.4V with the standard CCIC SA and the novel CTIA for read operation to remain within the CTIA specifications. Figure 3.6.d. presents MW(0σ) extracted from the distributions measured with both sensing elements. Similarly to the results presented before at reading field of 4MV/cm, MW(0σ) degrades with array size expansion when reading with a CCIC SA. However, reading with a CTIA enables a much larger MW(0σ) that is almost constant with array size, demonstrating an almost null memory window dependence on C_{BL}.

There are many advantages to implementing a charge sensing reading system rather than a voltage sensing reading system. The main one, and the reason for this study, is that MW(0σ) does not degrade with BL size expansion. Another major benefit is that the CTIA provides a greater amplification of the output signal, hence increasing MW(0σ) in similar testing conditions. Therefore, operation at lower voltages should be achievable in CTIA-based arrays, thus relaxing the constraints on stack optimization and variability control at the current node. While its efficiency at more advanced technology nodes will need to be verified, it is important to note the CTIA as reported in [118] occupies a $630\mu m^2$ (compared to $40\mu m^2$ for a CCIC SA) footprint area, which is detrimental to high density scaled arrays. Moreover, as a unique reading element (compared to one CCIC SA per Bit Line in the voltage-based reading scheme), it could increase bitcell-to-bitcell variability at array level as the electrical path between two distant bitcells is likely to be very different.



3.2.3 3D Capacitors for Enhanced MW at Constant Footprint

3.2.3.1 Integration

As was briefly presented in Chapter 2, high-density is required when expanding the capacity of FeRAM arrays in order to avoid increasingly high (and therefore costly) footprints. A first approach to increase the density of FeRAM arrays without modifying the dimensions of 1T or 1C components is to directly scale the 1T-1C bitcell. Indeed, the current layout is suboptimized since each bitcell comprises the FeCAP and its Via-SUP interconnection, as well as a Via-CONTACT contacting the bottom electrode level. This corresponds to the configuration shown in Figure 3.7.a., and to the standard integration process presented in the Introduction Chapter (Chapter 1). A more optimized array layout enabled the centralization of the Via-CONTACT for each WL, therefore clearly scaling of 1T-1C bitcell as can be seen in Figure 3.7.b. Without modifying the periphery, the overall array area was divided by two compared to the regular configuration thanks to the adoption of shrunk bitcells.

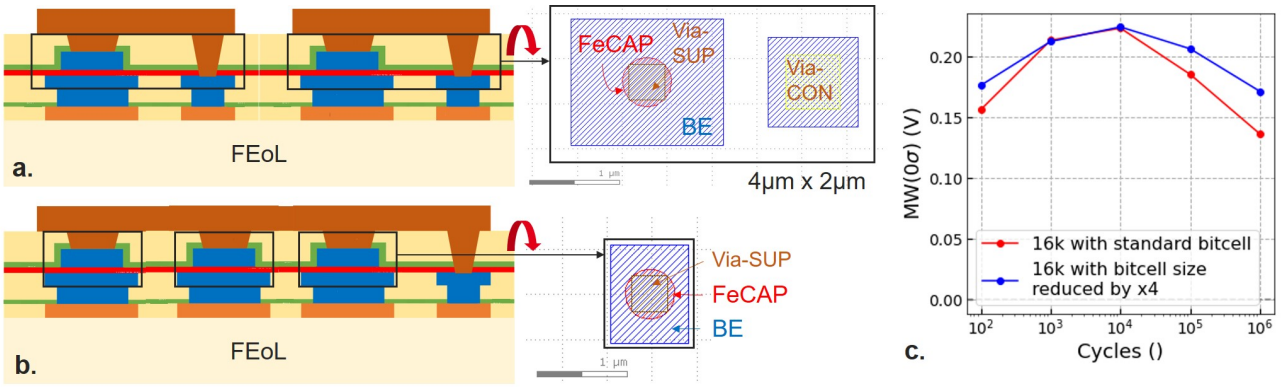


Figure 3.7: Integration schematic and layout view of: a. 2× standard (relaxed) bitcells ; b. 3× shrunk bitcells evidencing a footprint gain. ; c. MW(0σ) during cycling measured with 4MV/cm - 2µs pulses on a 16kbit array with standard and shrunk bitcells on a TiN/HZO (7nm)/TiN wafer demonstrating very similar performances.

To assess the electrical performance of the developed structure, 16kbit arrays with shrunk bitcells have been fabricated and tested. A V_{ref} scan operation on arrays (130nm node, 7nm HZO) with standard bitcells and with shrunk bitcells reveals little to no difference in MW(0σ) (and MW(16k)) during cycling, as shown in Figure 3.7.c. However, although this characterization demonstrated the integrity of the ferroelectric performances in a denser 16kbit array, the obtained scaling is still not considered as competitive according to industry standards.

Therefore, to effectively scale the whole array footprint, a reduction in transistors and FeCAPs dimensions is also required. From Chapter 2, and FeCAP switching dynamics considerations, it has been demonstrated that the selection transistor does not limit the bitcell area compared to current-driven memories such as OxRAM and PCRAM. Indeed, adopting scaled transistors at the 28nm node will not critically limit the FeCAP switching current, and consequently the FeCAP will be the limiting element in bitcell area scaling. From evident area considerations, the absolute switching charge will be reduced in scaled FeCAPs, thus leading to a decrease in MW(0σ). To overcome this issue, and still obtain a readable MW in larger and denser arrays at the 28nm node, Chapter 2 proposed the fabrication of 3D FeCAPs [94]. To this end, a possible integration technique consists in depositing the Metal / Ferroelectric / Metal stack in a 3D cavity (compared to the current 2D planar integration). It has the advantage of limiting the FeCAP footprint increase while still benefiting from an extended electrical area.

The colormap proposed in Figure 3.8.a. presents basic computation results of MW(0σ), and describes the tradeoff between array size (C_{BL}), FeCAP area and $2.P_R$ to maintain a 0.1V MW(0σ). Different FeCAP areas for 2D and 3D integration (with multiple Aspect Ratios, AR) and a unique top view footprint are projected. It can be seen that, for a given array size (fixed C_{BL}), a 0.1V MW can be maintained with a much lower remanent polarization with 3D capacitors than 2D capacitors. Conversely, at a fixed $2.P_R$, a greater MW can be achieved with 3D capacitors. Especially, 3D capacitors with a high aspect ratio, *i.e.* when the depth dimension is much greater than the width dimension, yield the highest MW gain at array level. The cost of integrating 3D FeCAPs in

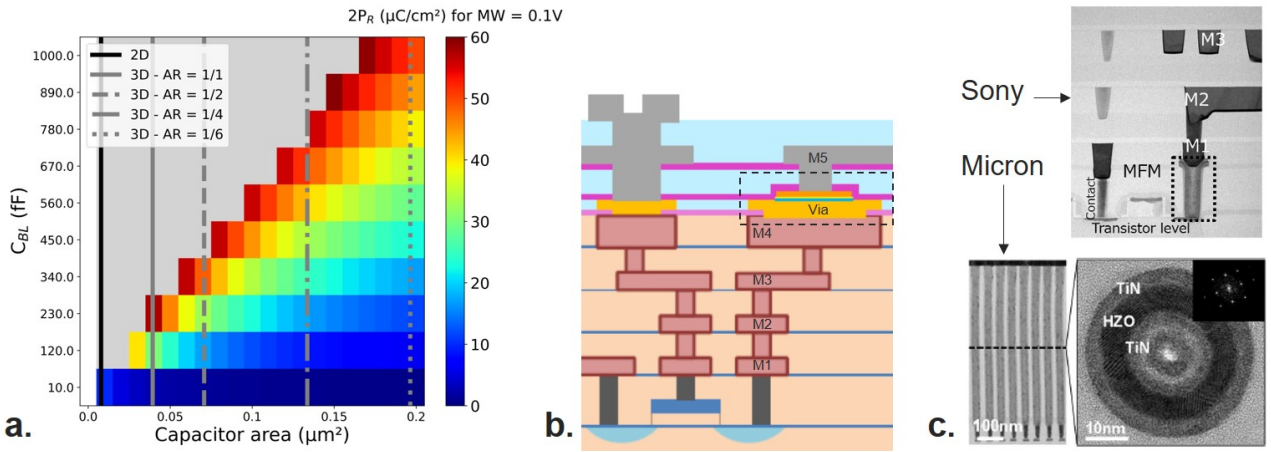


Figure 3.8: a. Remanent polarization ($2.P_R$) necessary to maintain a 0.1V MW(0σ) as a function of array size (C_{BL}) and FeCAP area (S) based on $MW(0\sigma) = 2.P_R \cdot S/(C_{DE} + C_{BL})$. Vertical lines represent different FeCAP developed areas for 2D or 3D integration with a unique planar footprint. ; b. Schematic of BEoL integration of 2D FeCAPs evidencing via space to be used for 3D integration. ; c. Micrographs of 3D scaled FeCAPs integrated in vias by Sony [91] and Micron [92].

the BEoL of CMOS process is limited as the space for 3D MFM can be found in interconnection vias, as shown in Figure 3.8.b.: no additional masks are required compared to 2D FeCAP integration. ALD deposition technique is well adapted to 3D FeCAP integration and, if full crystallization of the ferroelectric layer in the polar orthorhombic phase can be ensured on the sidewalls and in the corners, maintaining a good MW while reducing the 1C footprint should be achievable.

Some results from the literature presented in Figure 3.8.c. demonstrate the feasibility of FeCAP 3D integration and showcase realizations up to 1/20 AR. While most published structures are not integrated in arrays but as single bits [94, 119, 120], recent publications by Sony [91] and Micron [92] have reported the successful fabrication of 1Mbit and 32 Gbit arrays with 3D capacitors, respectively.

3.2.3.2 Performance

As a first demonstration, and for fabrication time and cost concerns, it was preferred to launch a ShL on 300mm wafers rather than a adopt a MAD approach. Taking advantage of two available lithography masks with no additional mask needed, a ShL process with 2D+3D FeCAPs based on a 10nm-thick HZO stack was developed. On base wafers, a first mask enabled the lithography and patterning of trenches with a sub-300nm diameter. A second mask enabled the electrical connection of the trenches with a 100μm diameter top electrode. A view of the 2D+3D structure processed on 300mm ShL wafers is drawn in Figure 3.9.a. As a reference for 2D capacitors, the same process route was followed on non-patterned base wafers, and the area ratio between 2D and 2D+3D structures has been calculated to be 1.4. The electrical characterization of 10 nominally identical devices for each wafer was performed with PUND sequences at 4MV/cm during endurance cycling,

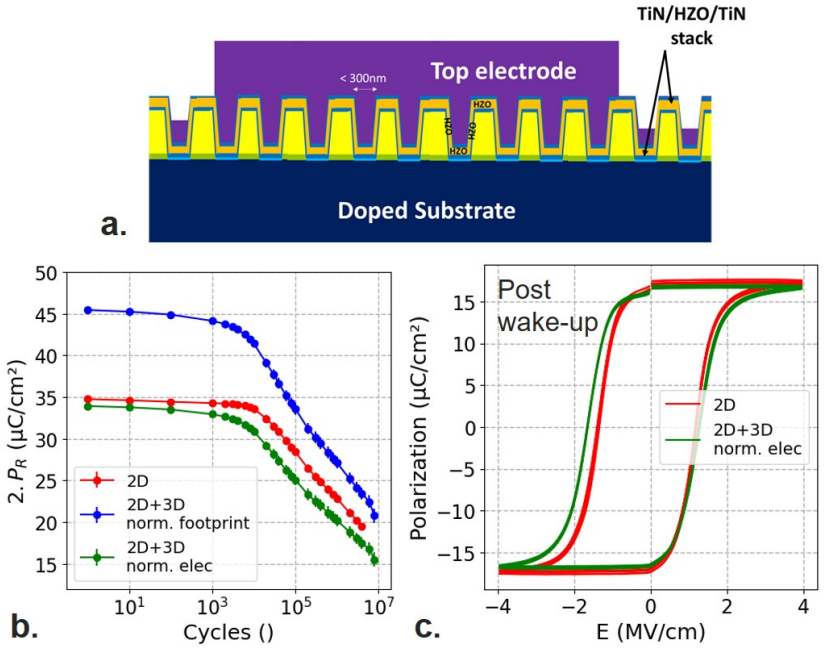


Figure 3.9: a. Schematic of the 2D+3D integration of FeCAPs in sub-300nm vias electrically connected by an overlapping top electrode. ; On a TiN/HZO (10nm)/TiN ShL measured at 4MV/cm - 10kHz after wake-up: b. Median and standard deviation of 2.P_R-Cycles for 2D and 2D+3D FeCAPs. 2.P_R is normalized either by the apparent 2D footprint (blue) or by the total electrical area (green). ; c. P-V hysteresis loops of 2D and 2D+3D structures (normalized by electrical area) for 10 devices.

following an initial wake-up procedure. To obtain a polarization metric (*i.e.* charge density per unit area), switching FE charge was normalized by two different areas: on-wafer surface occupancy (*i.e.* top-view footprint) and developed area (corresponding to the total electrical area). Figure 3.9.b. shows a 35% boost of the FE performance for the 2D+3D structure when considering the footprint-normalized 2.P_R, as expected from the theoretical area ratio. Using the geometrical aspect ratio, it can be seen that the 2D (red) and 2D+3D (green) structures present close characteristics, indicating a good crystallization of the sidewalls and corners. This is illustrated by the overlapping hysteresis loops presented Figure 3.9.c. Moving towards BEoL-integrated 3D FeCAP with high aspect ratios, likely integration challenges will be to successfully deposit the electrodes (and to a lesser extent the FE material) in the cavity, as well as control the mechanical strain so as to achieve homogeneous crystallization in the polar orthorhombic phase.

Recap

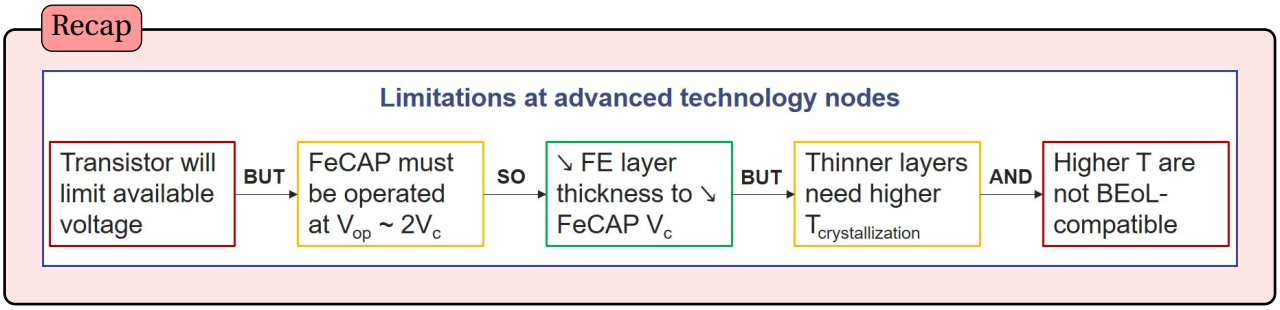
To enable the design of dense FeRAM arrays with a large MW, 3D capacitors are promising to reduce the FeCAP footprint while maintaining a large electrical area. A first realization of 2D+3D HZO-based FeCAPs shows a performance gain, consistently with literature results. The integration of 3D capacitors with high aspect ratios in BEoL 130nm or beyond would constitute an important step towards achieving high-performance FeRAM arrays at advanced nodes.

3.2.4 Lower Ferroelectric Layer Thickness for Low Voltage Operation

3.2.4.1 Voltage and Annealing Temperature Limitations

As was shown in Chapter 2, the 1T scaling to more aggressive process nodes will result in a low voltage requirement in order to avoid the dielectric breakdown of the transistor thin gate oxide. More precisely, the nominal voltage specifications of ST Microelectronics 28FDSOI transistors are 0.9V for thinner oxide gate 1T (GO1) and 1.8V for thicker oxide gate 1T (GO2). Operation at overdrive voltages of 1.1V and 1.98V, respectively, is permitted and leads to only little degradation of the transistors reliability. Considering these voltage limitations and the recommended operation at $V_{op} = 2V_c$ from [90] to enable saturated switching of ferroelectric domains, coercive voltages below 1V have to be achieved in the FeCAP.

As demonstrated in [121, 90, 122], the coercive field of FeCAPs is constant with the ferroelectric layer thickness. It follows that the coercive voltage of FeCAPs decreases with the ferroelectric film thickness, hence providing a clear advantage of low-thickness FE films for low voltage operation. Moreover, quasi-static current-voltage measurements on HZO-based FeCAPs with various thicknesses demonstrate that the breakdown field increases when the FE layer thickness decreases, therefore providing an increased operation space in low FE thickness films. An explanation for this non-intuitive result is proposed by [90]: from the band-diagram of an MFM stack it can be seen that the excess energy of electron injection from the electrode through the ferroelectric layer is weaker in thinner films, thus causing less degradation of the FE film and hence leading to a higher breakdown field. This higher breakdown field in thinner films constitutes another advantage because the device reliability is expected to be improved. However, $2.P_R$ measured by Toprasertpong et al. in [90] on large HZO-based MFM capacitors with various thicknesses and annealing temperatures ranging from 300°C to 650°C indicate that a higher crystallization temperature is needed for thinner films to maintain strong ferroelectric response (here $2.P_R > 10\mu C/cm^2$). Temperatures above 500°C are not BEoL-compatible, therefore alternative solutions need to be developed to ensure complete crystallization of thinner ferroelectric films for low-voltage operation.



3.2.4.2 Role of Nanosecond Laser Anneal for Crystallization of Thin Films

It is important to have an as large as possible MW at FeRAM array level for reliability reasons (including a better immunity to ferroelectric fatigue and imprint) that will be detailed in Chapter 4. With the PoR TiN/HZO/TiN stack, the coercive fields are around 1.5MV/cm, resulting in a ~1V coercive voltage for 7nm-thick HZO layers. Operating at twice the coercive voltage [90] would still necessitate a voltage higher than the 1.1V overdrive voltage allowed for GO1 1T at the 28FDSOI node for example. In this part, lower film thicknesses (4nm) are investigated on ShL demonstrators, and the impact of two different types of annealing steps is studied. For 4nm-thick HZO layers, a standard furnace annealing step was performed at 500°C for 15 minutes on ShL wafers with a TiN/HZO/TiN stack. Electrical characterization at 4MV/cm results in a sub-5$\mu C/cm^2$ remanent polarization, that slightly increases up to 7 $\mu C/cm^2$ after 10^5 cycles, as presented in Figure 3.10.b. This 2.P_R is small (below 10 $\mu C/cm^2$) and would result in a very limited MW at FeRAM array level. Therefore, boosting the crystallization with another annealing technique needs to be investigated.

Nanosecond Laser Anneal (NLA) (Figure 3.10.a.) consists in laser pulses applied at the end of the

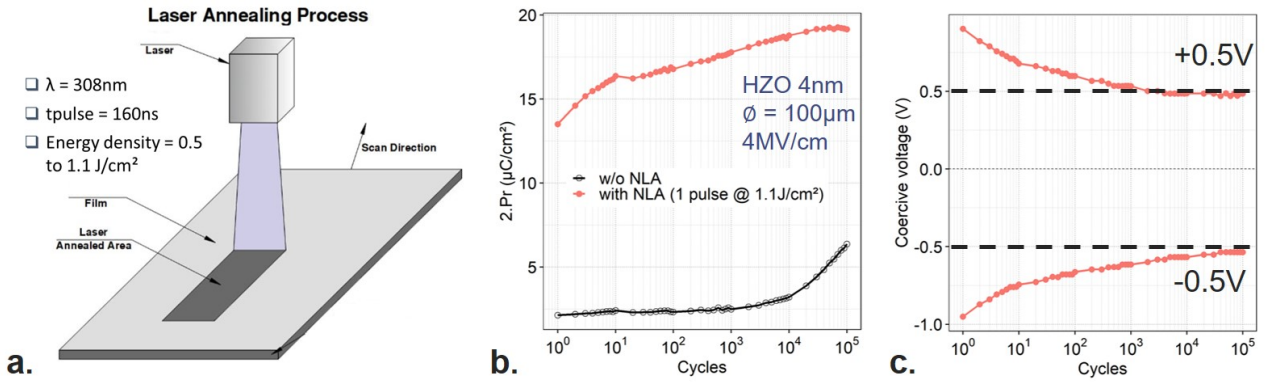


Figure 3.10: a. Schematic of laser annealing setup for local temperature elevation with the metrics of interest. ; b. 2.P_R-Cycles measured at 4MV/cm on 100µm diameter FeCAPs with TiN/HZO (4nm)/TiN stack for a wafer with (red) and without (black) laser annealing ; c. V_c-Cycles for stack with laser annealing, showing stabilization of coercive voltages around ±0.5V after wake-up.

process to locally increase the temperature for crystallization enablement of very thin ferroelectric layers while preserving FEoL integrity. 1D simulations on BEoL-integrated devices estimate the temperature elevation up to 2000°C with the highest laser energies (0.5 J/cm²) in the HZO layer, but below 500°C in the deeper layers, making it a promising solution to crystallize thinner films while respecting BEoL constraints [111]. NLA was performed on the same ShL (half-)wafer with a 4nm thick HZO layer, following the application of a first standard furnace anneal. The results of the electrical characterization are presented in Figure 3.10.b. and evidence an initial $2.P_R$ around 15 $\mu C/cm^2$ that increases up to 20 $\mu C/cm^2$ with cycling at 4MV/cm (corresponding to 1.6V). Those very good ferroelectric performances are coupled with a coercive voltage of 0.5V (Figure 3.10.c.), that could enable operation at 1V with full saturation of the hysteresis loop and thus a readable MW in arrays with GO1 or GO2 transistors at the 28nm node.

However, while Nanosecond Laser Anneal enables the crystallization in the ferroelectric phase in ShL integration, further investigations need to be conducted on BEoL-integrated wafers to evaluate the potential of this strategy towards advanced technology nodes. This can be the focus of future studies, but has not been investigated in this work.

Recap

Constraints regarding available voltages at advanced process nodes can be solved by decreasing the ferroelectric layer thickness to decrease FeCAP coercive voltages. This can result in a higher MW at lower operating voltages. However, crystallization of thinner ferroelectric layers in the polar orthorhombic phase requires high annealing temperatures that are incompatible with BEoL requirements. Nanosecond Laser Anneal is known to increase the temperature only locally and to preserve FEoL integrity, and has been demonstrated to boost the polarization in ShLs with 4nm HZO films. Further investigation at BEoL-integrated wafer level are needed to confirm the potential of this technique.

3.3 Bitcell-to-Bitcell Variability at FeRAM Array Level

While MW(0σ) gives a precise picture of the median behavior of all 16384 bits across the 16kbit FeRAM array and so evaluate the impact of process improvements on the overall array performance, it is important to take into consideration the bitcell-to-bitcell variability. Indeed, from an application-oriented perspective, read operation with a given V_{ref} voltage should single-handedly enable the discrimination of bits programmed in the '0' state from all the bits programmed in the '1' state. In an ideal case, the BL voltage elevation subsequent to reading a '0' or a '1' is the same for all bits as it depends on fixed variables as described analytically in:

$$V_{BL}^{0} = \frac{C_d}{C_d + C_{BL}} \times V_{SL} \text{ where } C_d = \epsilon_0 \epsilon_R \times \frac{S}{t_{FE}} \tag{3.1}$$

and

$$V_{BL}^{1} = \frac{C_d}{C_d + C_{BL}} \times V_{SL} + \frac{2.P_R \cdot S}{C_d + C_{BL}} \tag{3.2}$$

In fact, however, process- and design-related inhomogeneities at FeRAM array level cause a dispersion of $V_{BL}{}^{0}$ and $V_{BL}{}^{1}$, leading to tilted '0' and '1' states distributions across V_{ref}, as can be observed in quantile-quantile plots. One major downside of this variability is that the effective MW, or MW(16k), is reduced compared to MW(0σ), thus calling for ever improving median ferroelectric performances to maintain a sufficient read margin. Another way is to reduce bitcell-to-bitcell variability, and this section is devoted to identifying the sources of variability, evaluating the significance of their contribution to the distribution slopes, and proposing solutions to reduce them.

3.3.1 '0' and '1' States Distributions

In order to discuss the bitcell-to-bitcell variability at FeRAM array level, it is important to define a few terms. This first part aims at presenting how the quantile-quantile plots of '0' and '1' states distributions can serve as metrics for variability assessment.

3.3.1.1 Assumption of Normality

As explained in the Introduction Chapter (Chapter 1), the read operation in a CCIC SA-based FeRAM array consists in pulsing the SL so as to program a '0', and comparing the voltage elevation at the BL floating node to a chosen reference voltage V_{ref}. If $V_{BL} < V_{ref}$, the output voltage will be switched to the ground of the circuit ; if $V_{BL} > V_{ref}$, the output voltage will be switched to the V_{DD} of the circuit. This digital read operation (outputing a binary information regarding the state read) is highly dependent on a well-chosen V_{ref}, *i.e.* a V_{ref} chosen between $V_{BL}{}^{0}$ and $V_{BL}{}^{1}$ that can accurately differentiate a '0' state from a '1' state. Digital characterization performed at various V_{ref} voltages on a 16kbit programmed entirely in the '0' or '1' state enables the extraction of $V_{BL}{}^{0}$ or $V_{BL}{}^{1}$ for all 16384 bits and hence the count of occurrences, as presented in Figure 3.11.a. $V_{BL}{}^{0}$ and $V_{BL}{}^{1}$ are distributed across the 16384 bitcells of the array: this bit density can be fitted with a normal distribution featuring different medians and standard deviations for both logic states, as shown on the graph. From this data (taken from [68]), a relevant V_{ref} voltage for digital operation can be chosen around 0.82V, as shown by the purple horizontal line in Figure 3.11.a. In Figure 3.11.b., a quantile-quantile plot is used to represent the cumulative distributions for both logic states, which is equivalent to the probability for a bitcell of having a V_{BL} lower than the considered V_{ref}. The standard deviation at 1σ is the preferred metric to assess the intrinsic variability at FeRAM array level. For normal distributions, this metric is equivalent to the inverse slope of the linear regression fitting these distributions (in mV/σ). In the data presented, slopes around $26mV/\sigma$ and $35mV/\sigma$ are extracted for the '0' and '1' states, respectively. In the 64kbit array reported in [123], sub-$20mV/\sigma$ slopes are obtained for both logic states, a very interesting result that enables the improvement of the total MW at a given array size. These steep distributions originate from a limited bitcell-to-bitcell variability at FeRAM array level, thanks to the combined optimization of both process integration and circuit design. A very low value for slope distributions in mV/σ (*i.e.* distributions as steep as possible, exhibiting the least possible tilt) is therefore targeted.

Finally, to visually assess the normal quality of distribution slopes, an interesting option is to

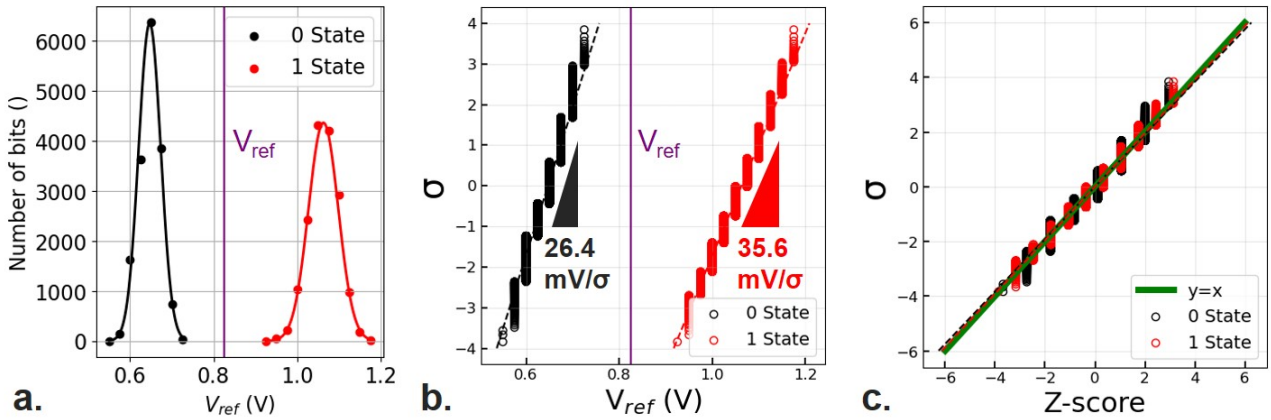


Figure 3.11: From a V_{ref} scan operation on a 16kbit 1T-1C FeRAM array constituted of $0.36\mu m^2$ FeCAPs with a TiN/HZO (10nm)/TiN ferroelectric stack: a. Number of bits with a given $V_{BL}{}^0$ and $V_{BL}{}^1$ fitted using a normal distribution. ; b. Quantile-quantile plot of cumulative '0' and '1' states distributions, fitted linearly with dashed lines. Slopes of the distributions at 1σ are $26.4mV/\sigma$ and $35.6mV/\sigma$, respectively. ; c. Z-score (normalized and centered) distributions allowing to assess normal quality of the '0' and '1' states distributions compared to y=x line.

make use of the Z-score, as shown in Figure 3.11.c. When centering (subtracting the median μ) and normalizing (dividing by the standard deviation σ) the '0' and '1' states distributions ($Z = (X-\mu)/\sigma$), it can be evaluated whether the thus-obtained Z-score distributions align with the green line labelled 'y=x' and representing a perfectly normal distribution without extrinsic element at the 16kbit scale. A non-normal distribution would not follow the 'y=x' line, and the outliers of a normal distribution (*i.e.* extrinsic points that do not follow the intrinsic distribution) would diverge at the edges. For the data considered, it can be seen that the quality of the normal distributions is very high, with '0' and '1' states distributions departing only slightly from the green line, even for larger projected statistics at 6σ. Note that Figure 3.11.c. enables the visual comparison of normality (and not variability) between two distributions.

3.3.1.2 A Specific Counter-Example: Extrinsic 8×8 Pattern

While the previous part has discussed the definition and characteristics of a normal distribution, this part focuses on a specific counter-example that has been observed to occur occasionally on some wafers, independently of the ferroelectric material (HSO or HZO) and the FeCAP area. In particular from Figure 3.12.a., it can be seen that the bit-wise bitmap of ΔV_{BL} (extracted from $V_{BL}{}^1$-$V_{BL}{}^0$ through a V_{ref} scan operation on a 16kbit array of such a wafer) can exhibit a grid pattern along WLs and BLs. ΔV_{BL} is not homogeneously distributed across the array, and a repeatable pattern of increased ΔV_{BL} at certain WLs and BLs can be evidenced. This non-homogeneous bitmap indicates a higher bitcell-to-bitcell variability and can result in a loss of normality in the corresponding distributions. Figure 3.12.b. and Figure 3.12.c. analyze closely the related BL voltage elevation along WLs and BLs, respectively.

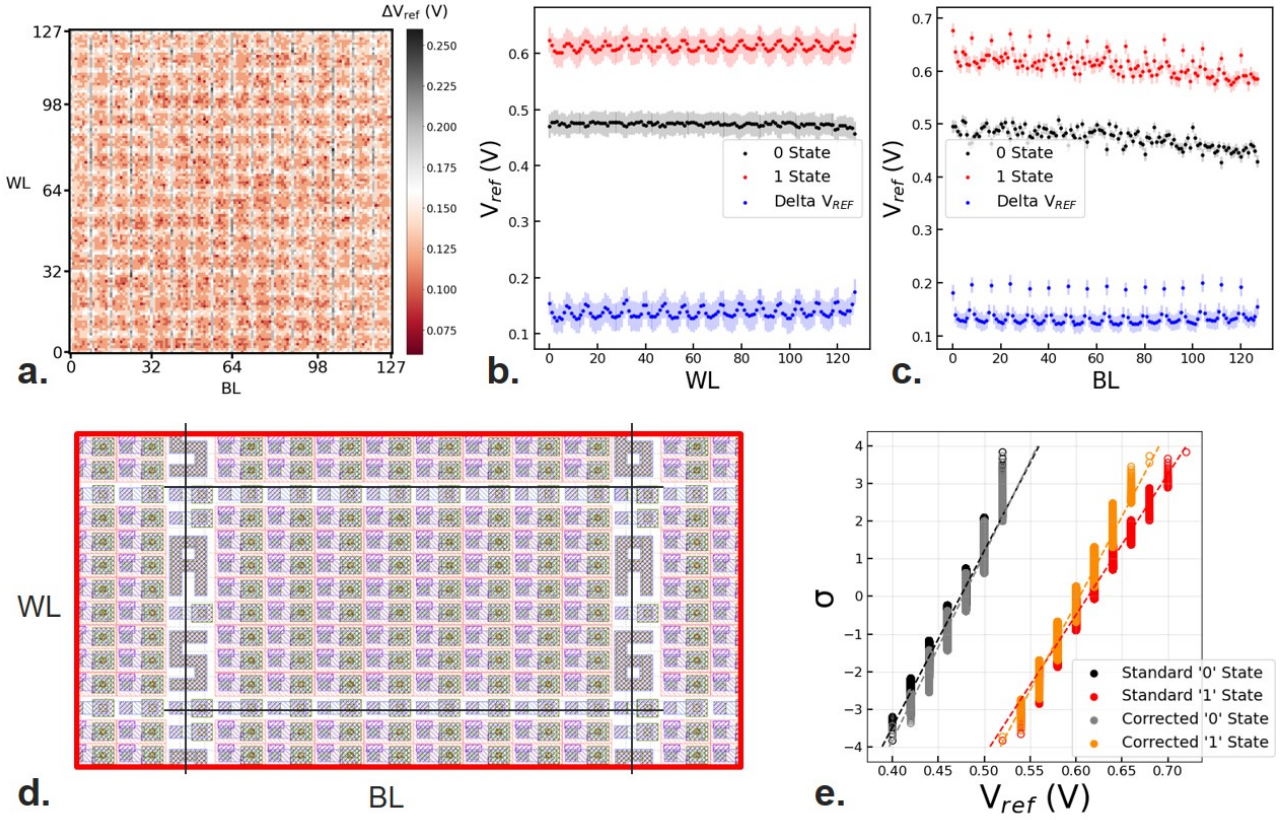


Figure 3.12: From a V_{ref} scan operation on a 16kbit 1T-1C FeRAM array constituted of $0.36\mu m^2$ FeCAPs with a TiN/HZO (10nm)/TiN ferroelectric stack: a. Bitmap of ΔV_{ref} presenting a grid pattern. ; b. V_{ref} for the '0' and '1' states, as well as ΔV_{ref}, averaged over BLs as a function of WLs. ; c. V_{ref} for the '0' and '1' states, as well as ΔV_{ref}, averaged over WLs as a function of BLs. ; d. Layout view of an 8 WLs by 8 BLs unit, with WLs separated by dummies and BLs separated by labels. ; e. '0' and '1' states distributions for the entire 16kbit array ('standard') and corrected by removing every 8 BL and 7 and 8 WLs ('corrected').

In blue, ΔV_{BL} follows a noticeable wave pattern in which a voltage maximum is reached every 7 and 8 WLs, and every 8 BL. This pattern observed on ΔV_{BL} originates mainly from a ferroelectric state contribution, as can be seen on V_{BL}[1] in red. To understand how this specific pattern occurs, it is useful to look at the layout of the 16kbit array presented in Figure 3.12.d. Indeed, as can be seen on the graph, the 16kbit array is divided into subarrays constituted by 8 WLs and 8 BLs units, separated by so-called dummies. WLs of two neighboring subarrays are separated by dummies replicating the size and shape of a 1T-1C bitcell but are not electrically connected. Equivalently, BLs are separated by larger labels. The devices exhibiting a high ΔV_{BL} are located on each side of dummies lines, and on the right-hand side of labels lines, *i.e.* further from the label than on the other side. Given that ΔV_{BL} represents the MW, which is itself equal to $(2.P_R \cdot S)/(C_{DE} + C_{BL})$, it can be hypothesized that process-induced variations of FeCAP area or etching quality arise due to a differentiated environment at the edges of subarrays. For example, the Chemical Mechanical

Polishing (CMP) step could result in inhomogeneities of the edge devices. However, an additional contribution of electrical origin cannot be excluded. To get a quick estimation of the influence of such a pattern on the variability, the distributions slopes metric is used. In the case of the standard distributions presented in Figure 3.12.e., the slopes for the '0' and '1' states are $20.4 mV/\sigma$ and $26.0 mV/\sigma$, respectively. When suppressing the peak voltages for both $V_{BL}{}^0$ and $V_{BL}{}^1$ (every 7 and 8 WL, and every 8 BL), the normality of the thus-corrected distributions increases and the slopes decrease to $19.1 mV/\sigma$ and $20.7 mV/\sigma$, respectively. Note that while MW(16k) is not impacted by the correction in this case, the Bit Error Rate (BER) is. Since this pattern occurs sparsely on some wafers and can be easily corrected numerically, it does not constitute a major variability concern. However, the future FeRAM arrays will be designed without dummies and labels to avoid the recurrence of this 8×8 pattern and steepen the distributions. Note that, from a pure MW standpoint, the most interesting option would have been to design dummies around every bitcell, which is unreasonable from a density perspective.

Recap

Normality and variability of the Bit Line voltage elevation following a read operation on a FeRAM array can be evaluated from the shape of the distribution in a quantile-quantile plot. Apart from specific cases that can indicate process- and/or design-induced inhomogeneities, V_{BL} is considered to follow a normal distribution, whose slope in mV/σ is a metric of interest when assessing bitcell-to-bitcell variability.

3.3.2 Intrinsic Variability Sources: Impact and Possible Fixes

3.3.2.1 Process-Induced Variability

As has been seen in the previous section, process-induced inhomogeneities can occasionally lead to loss of normality in distributions and increased variability. From this on, the wafers considered are specifically chosen so as to exhibit no 8×8 pattern. Now considering only normal distributions and focusing on intrinsic variability sources contributing to the distribution slopes, it can be proposed that the process yields some variability, across the wafer as well as within the array. As especially seen from Equation 3.1 and Equation 3.2, variations in dimensions such as the FE material thickness t_{FE} and FeCAP area S, and in material-related metrics such as $2.P_R$ and ϵ_R impact the BL voltage elevation during a read operation, and can therefore contribute to array-level variability.

Regarding the ferroelectric layer thickness, it has been measured by the ellipsometry technique on full sheet 300mm diameter wafers that HZO thickness non-uniformity did not exceed 2%. Therefore, at 200mm wafer level, and more importantly at 16kbit FeRAM array level, this variation is not expected to significantly impact $V_{BL}{}^0$ and $V_{BL}{}^1$ variability, and its influence on the distribution slopes can be neglected. To evaluate the dispersion of FeCAP area across a 16kbit array, a procedure was defined in the clean room to measure the ferroelectric stack area in the middle of the mesa structure after the etching of the top electrode. The measurement was divided across subarrays (numbered from 1 to 16 in Figure 3.13.a.) to enable frequent re-alignment of the SEM for improved precision. As can be seen in Figure 3.13.a., the FeCAP areas measured are reported in a bitmap,

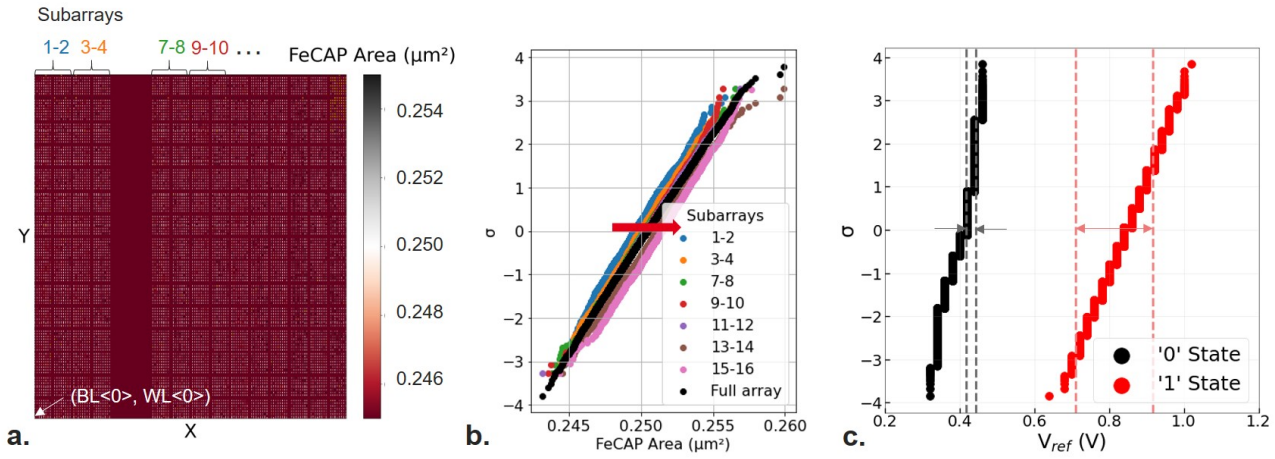


Figure 3.13: a. Bitmap of FeCAP area measured by SEM during the manufacturing process across a 16kbit 1T-1C array, where X represent the BL direction and Y the WL direction. ; b. Corresponding normal distributions of FeCAP area for each of the subarrays measured, as well as for the entire array. ; c. '0' and '1' states distributions for a 16kbit FeRAM array constituted of $0.24\mu m^2$ FeCAPs with a TiN/HSO (10nm, etched) /TiN stack. Vertical lines represent minimal and maximal 0σ values for $V_{BL}{}^0$ and $V_{BL}{}^1$ calculated from simulations using the compact model calibrated with minimal and maximal area measured across a FeRAM array, and minimal and maximal remanent polarization and relative permittivity measured across the wafer considered.

evidencing the 8 WLs by 8 BLs units presented in the previous part. The quantile-quantile plot of the FeCAP area across the array is shown in Figure 3.13.b. for the different subarrays and in the full array. The area follows a normal distribution presenting a slight left-right gradient, with a median of 0.250µm² and a standard deviation of $0.002\mu m^2/\sigma$, corresponding to less than 4% of FeCAP area variability when accounting for the 16kbit 4σ. FeCAP area variation is therefore too small to account alone for the distribution slopes of the Bit Line voltage for '0' and '1' states. To assess the contribution to the distribution slopes of material-related metrics $2.P_R$ and ϵ_R, a worst-case scenario is adopted: it is considered that $2.P_R$ and ϵ_R variation at 16kbit array level is the same as is measured by PUND on FeCAPs at wafer level. Using minimal and maximal measured values for FeCAP area, $2.P_R$ and ϵ_R, a Write 0 - Read - Write 1 - Read (W0 - R - W1 - R) operation is simulated with the FeCAP model and elementary circuit presented in Chapter 2. The $V_{BL}{}^0$ and $V_{BL}{}^1$ voltages extracted are represented by dashed vertical lines on Figure 3.13.c., together with the '0' and '1' states distributions measured on a 10nm HSO 16kbit array. Information on this wafer and model calibration can be found in [86]. In fact, measured process-induced variations in dimensions and material at array and wafer level (in the worst case) do not fully represent the distribution slopes obtained from 16kbit FeRAM array measurements. It is therefore necessary to evaluate the contribution of other possible variability sources.

Recap

Process-induced variability (FeCAP area, ferroelectric layer thickness, remanent polarization and relative permittivity) has been assessed at FeRAM array level. Even in the least favorable conditions (considering the wafer-level variability as representative of the array-level variability), '0' and '1' states variability is not completely reproduced by simulations using the compact model and elementary circuit presented in Chapter 2. Other variability sources must be assessed.

3.3.2.2 Access Devices and Sense Amplifier Variability Along a Bit Line

In this part, it is attempted to evaluate the contribution of the 128 access transistors (transistors of the 1T-1C bitcells) and the CCIC SA element (and its constituting transistors) to the variability along a Bit Line. To this end, a single BL structure as shown in Figure 3.14.a. is simulated using Eldo simulator. The FeCAP element is modelled by a Preisach-inspired compact model based on a $tanh$ description of the P-V hysteresis loop, calibrated to fit standard 10nm HZO data. Its characteristics are not varied, as the main goal of this part is to assess transistors-related variability rather than FeCAP-induced variability. To emulate a dispersion of transistors characteristics that can reasonably occur at single Bit Line level, a Monte-Carlo approach is adopted with 128 iterations. The transistors characteristics are varied within boundaries (or corners) from the HCMOS9 transistor model provided by STMicroelectronics. Classical V_{ref} scan operation is simulated on the 128 FeCAPs of this Bit Line, and cumulated distributions for $V_{BL}{}^0$ and $V_{BL}{}^1$ can be extracted (see Figure 3.14.b.).

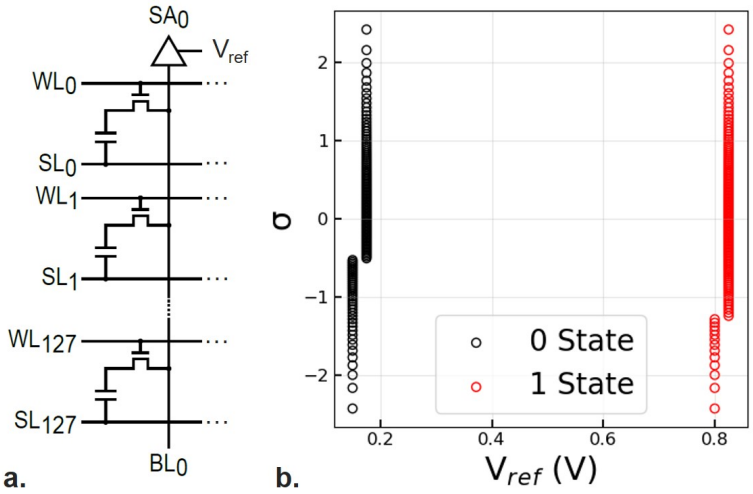


Figure 3.14: a. Schematic from [69] of a BL with 128 PTMEM connected to 128 access transistors and a single CCIC Sense Amplifier used in Monte-Carlo simulation ; b. '0' and '1' states distributions simulated on a BL with FeCAPs calibrated on HZO 10nm experimental data and transistors characteristics varied in the specified corners.

The obtained slopes are around $10mV/\sigma$, which corresponds to 50% of the usual distributions slopes obtained on 10nm HZO wafers. Therefore, the variability of access devices and CCIC SA contributes to half the BL voltage variability at 16kbit FeRAM array level.

3.3.2.3 Bit Line Voltage Gradient Across Bit Lines

In order to evaluate the contribution of design to intrinsic variability at FeRAM array level, the specific pattern of BL voltage gradient reported in [68] is investigated in this part. It consists in a progressive decrease of the output Bit Line voltage values for the '0' and '1' states as the number of Bit Lines increases. In [68], a $28mV/\sigma$ slope was extracted for the '0' state in a 16kbit FeRAM array operated at 4.8V. When artificially correcting the BL voltage gradient by subtracting the median

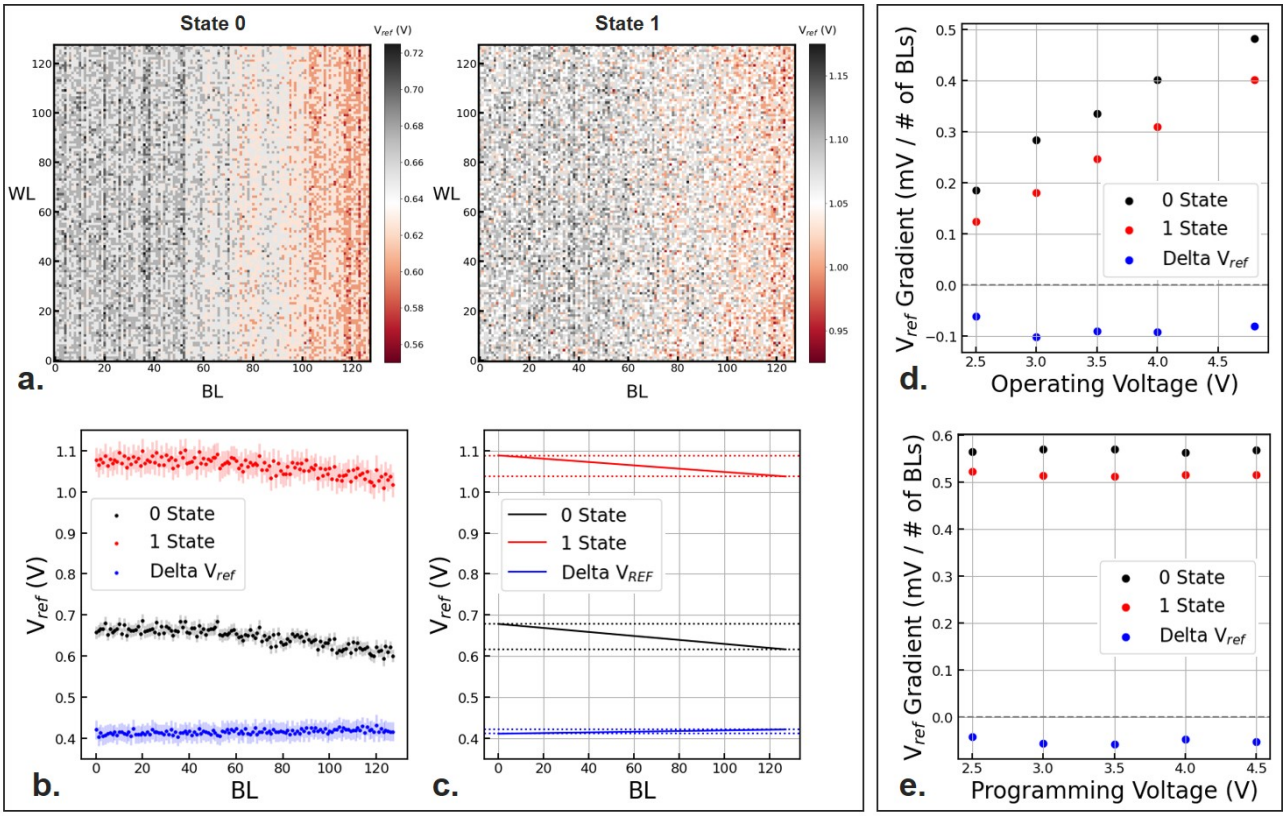


Figure 3.15: (a. to d.) From a V_{ref} scan operation on a 16kbit 1T-1C FeRAM array constituted of $0.36\mu m^2$ FeCAPs with a TiN/HSO (10nm)/TiN ferroelectric stack: a. Bitmap of $V_{BL}{}^0$ (left) and $V_{BL}{}^1$ (right) presenting a gradient across BLs. ; b. Corresponding V_{ref} for the '0' and '1' states, as well as ΔV_{ref}, averaged over WLs as a function of BLs. ; c. Linear fit of V_{ref} versus BLs and gradient defined by dashed lines between BL<0> and BL<127>. ; d. Variation of Bit Line voltage gradient for the '0' and '1' states, as well as ΔV_{ref}, as a function of operating voltage (V_{prog}=V_{read}) for the aforementioned 10nm HSO stack. ; e. Variation of Bit Line voltage gradient for the '0' and '1' states, as well as ΔV_{ref}, as a function of programming voltage (V_{read}= constant = 4V) for a 10nm HZO stack.

value for $V_{BL}{}^0$, the obtained slope was down to $19mV/\sigma$. This correction of the BL voltage gradient therefore had the beneficial effect of increasing MW(16k). In Figure 3.15.a.and Figure 3.15.b., bitmaps for $V_{BL}{}^0$ and $V_{BL}{}^1$, respectively, are extracted from a V_{ref} scan operation at 4.8V on a 16kbit array with a TiN/HSO (10nm)/TiN stack. Across BL, it can be seen that the gradient of BL voltage occurs homogeneously for '0' and '1' states. While it had been shown earlier that ΔV_{BL} evolution across BL was quite linear (except in the case of an 8×8 pattern), it is clear that $V_{BL}{}^0$ and $V_{BL}{}^1$ exhibit a strong dependence on BL number, as shown in Figure 3.15.c. This decrease can be modelled by a linear regression for the '0' and '1' states, as well as for ΔV_{BL}, as represented in Figure 3.15.c. The value for BL voltage gradient is then defined as the voltage drop over the 128 BLs for a 16kbit array (or more in the case of a 64kbit or 256kbit array) divided by the number of BLs, as shown between the dashed lines in Figure 3.15.c.

To try to understand the origin of this gradient, and especially determine if it comes from a programming or reading variability across the BLs, its dependence on voltage was studied. In Figure 3.15.d., an extension of the data set presented by François et al. in [68] is used to evaluate the evolution of this BL gradient for the '0' and '1' states, as well as for ΔV_{BL}, as a function of operating voltage. Note that, in this case, the term operating voltage encompasses BL and SL voltages modulated for both programming and reading operations. As can be seen on the graph, the BL gradient for the '0' and '1' states decreases when operating voltage decreases, meaning that the gradient is less severe (and therefore the distribution slopes less tilted) at low operating voltage. On a nominally identical 16kbit FeRAM array on another wafer (10nm HZO), the reading voltage was kept constant at 4V, regardless of the programming voltage. As can be observed in Figure 3.15.e., the BL gradient for the '0' and '1' states remains unchanged with programming voltage, and fixed around 0.5 and 0.6mV/BL. Therefore, it can be concluded that the reading operation is responsible for the BL gradient dependence on voltage, and that reading at low voltage reduced the BL voltage gradient.

From there, and because the BL voltage gradient is much likely design-induced, Eldo fast spice simulations are performed on a full 16kbit array with the $tanh$ compact model already used in Paragraph 3.3.2.2. More specifically, it is important to mimic the simultaneous read operation on 128 FeCAPs of a single WL that corresponds to the standard electrical protocol for FeRAM arrays. These simulations are run post-layout and therefore take into account possible resistive and capacitive parasitic effects. Figure 3.16.a. shows the structure of a single WL on which the operations W0 - R - W1 - R - R are simulated of each one of the 128 BLs. The last two reads are of importance to assess $V_{BL}{}^1$ and post Write-Back (WB) state, respectively. The former is expected to be an analog value consistent with Equation 3.2, and the latter should be to the ground if $V_{ref} > V_{BL}{}^1$, and to V_{DD} if $V_{ref} < V_{BL}{}^1$. A first analysis therefore consists in monitoring the BL voltage elevation during the read pulse (and before SA enabling). Simulation results shown in Figure 3.16.b. reveal fluctuations of $V_{BL}{}^1$, that could originate from insufficient simulation precision, but no gradient is observed. This therefore strongly suggests that the experimentally observed BL voltage gradient comes from the reading scheme, and more precisely the SA, rather than from the WL or SL. In particular, a voltage drop occurring along the V_{ref} line could cause a degradation of V_{ref} and hence a gradient in output voltage. To verify this hypothesis, several read operations are simulated on the entire WL (128 BLs) while varying V_{ref} close to $V_{BL}{}^1$. Indeed, if the V_{ref} value is not uniform along the WL, the effective comparison threshold is different for each SA (and hence for each BL) and a BL voltage gradient is expected to be measured. The post-WB state is monitored for

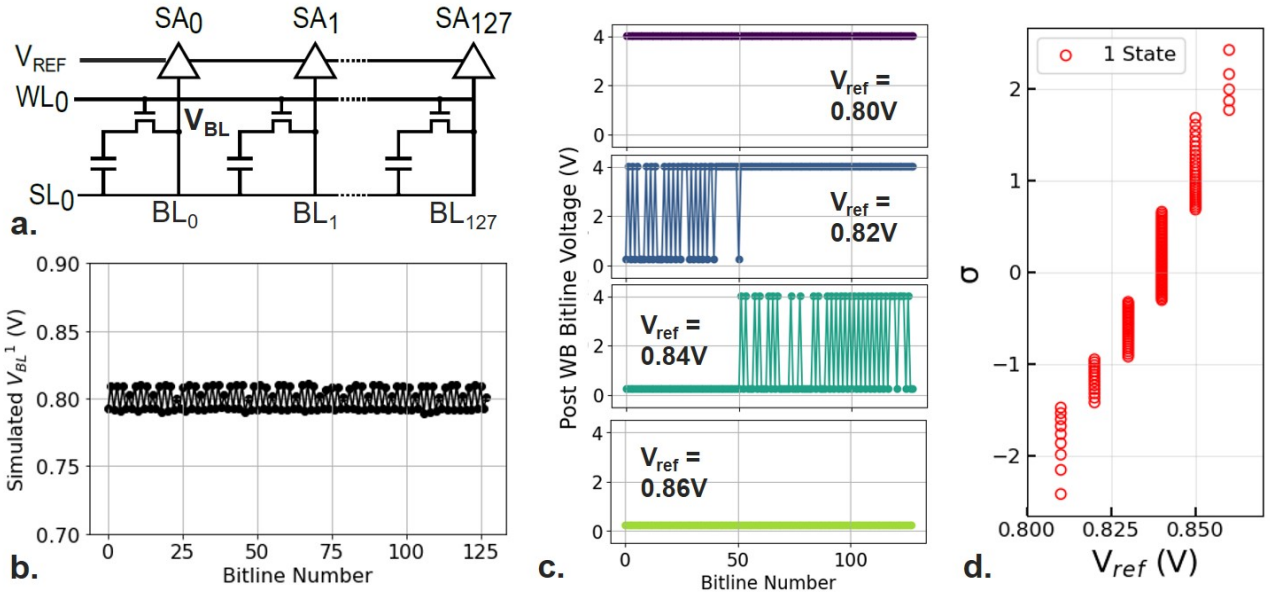


Figure 3.16: a. Schematic from [69] of a WL with 128 FeCAPs connected to 128 access transistors and 128 CCIA SA used in simulation of simultaneous read operation on all FeCAPs previously programmed in the '1' state. ; b. Simulated BL voltage elevation around 0.80V with regard to BL number exhibiting noise from limited simulation precision. ; c. Simulated post Write-Back BL voltage with regard to BL number for V_{ref} values between 0.80V to 0.86V resulting in a GND or VDD voltage depending on the comparator output. ; d. Corresponding simulated distribution taking into account reading variability.

each BL and is indicative of the outcome of the digital sense operation. Figure 3.16.c. gathers the simulation results for four different V_{ref} values. As expected, at high V_{ref}, every bit is read as a '0', regardless of the BL number. Similarly, at low V_{ref}, every bit is read as a '1', regardless of the BL number. For intermediate V_{ref} values, as can be seen from the color gradation, there is indeed a left-right gradient: on the left (*i.e.* for small BL numbers), low V_{ref} values enable the detection of a '1' ; on the right (*i.e.* for large BL numbers), higher V_{ref} values are required to detect '1'. During the commutation of the SA comparator, charge injection occurs during FeCAP read, which can cause a voltage drop along the V_{ref} line. This drop can be significant as the designed V_{ref} line has been observed to be thinner than SL or BL, for example. In further designs of FeRAM arrays, the thickness of the V_{ref} line will be corrected. There might also be a problem in the operation delays, with a parasitic short-circuit current generated by the transistors driving the ENABLE of the CCIC SA. Nonetheless, a non-negligible part of the variability at 16kbit FeRAM array level comes from this BL voltage gradient, as shown in Figure 3.16.d. This simulated '1' state distribution can be plotted with the results from the aforementioned simulations at seven different V_{ref} values, and exhibits a $12.5 mV/\sigma$ slope due to the BL voltage gradient, which represents roughly half the overall slope, and could be corrected easily.

Recap

To understand the Bit Line voltage gradient observed experimentally when reading a 16kbit array WL by WL, a simulation is run in the full 16kbit array circuit taking into account parasitic capacitance: WL<0> is programmed (in '0' and then '1' state) and read. While no gradient is observed before SA commutation, it can be detected afterwards and is probably due to an IR drop during SA commutation in a too thin V_{ref} line.

3.3.2.4 Double Issue for Large-Sized Arrays

As can be anticipated from the previous section regarding the BL voltage gradient on a 16kbit FeRAM array, the gradient will only accentuate in larger arrays, thus increasing the variability.

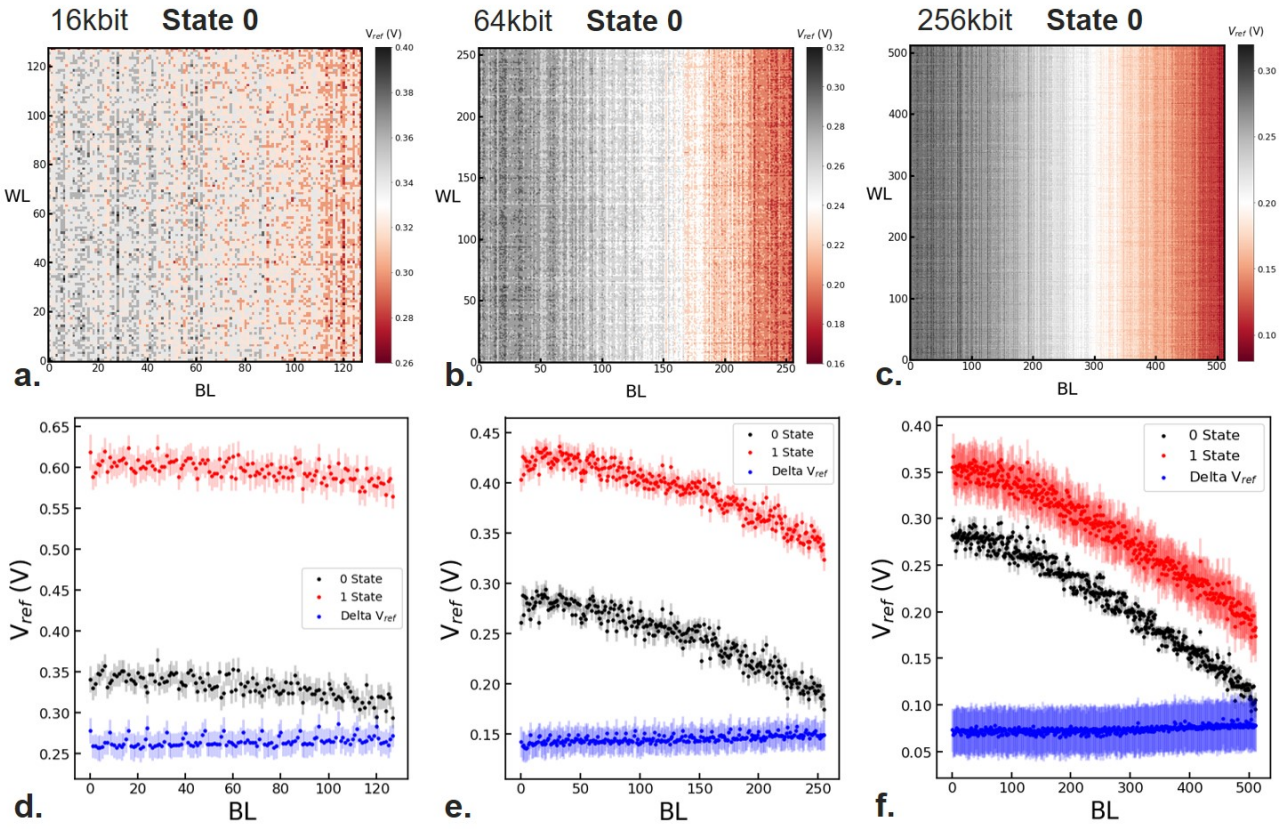


Figure 3.17: From a V_{ref} scan operation at 4MV/cm (2.8V) operating voltage on 1T-1C FeRAM arrays constituted of 0.24μm^2 FeCAPs with a TiN/HZO (7nm)/TiN ferroelectric stack: Bitmap of $V_{BL}{}^0$ for a 16kbit (a.), 64kbit (b.), 256 kbit (c.) ; V_{ref} for the '0' and '1' states, as well as ΔV_{ref}, averaged over WLs as a function of BLs for a 16kbit (d.), 64kbit (e.), 256 kbit (f.)

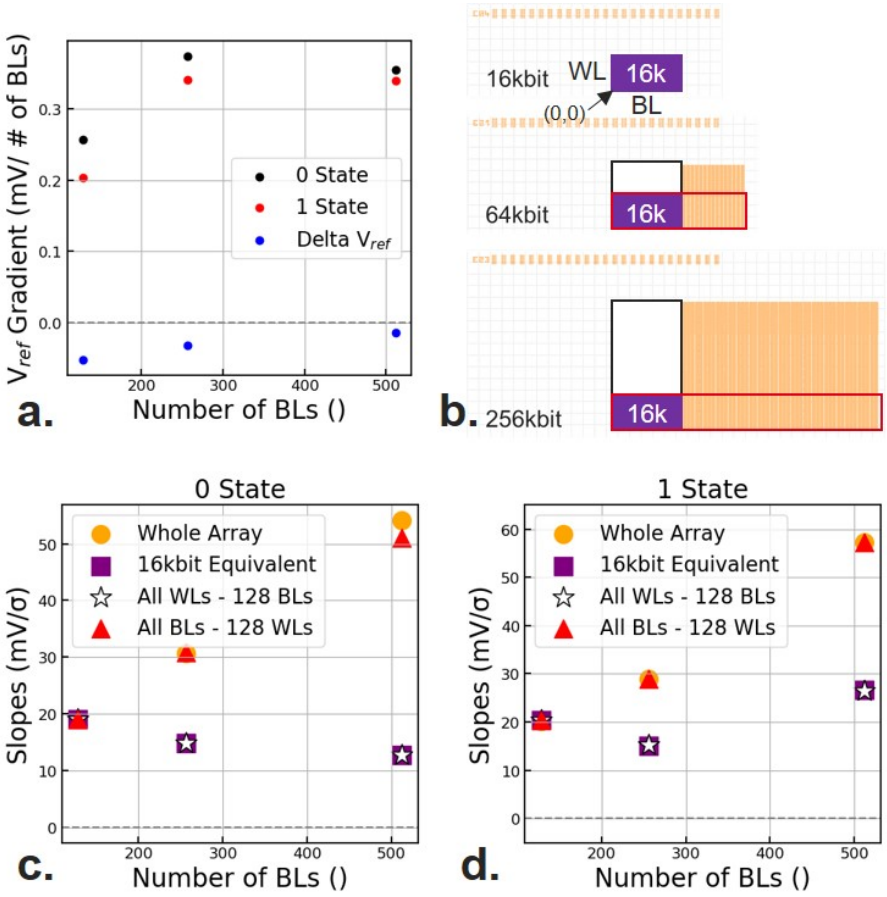


Figure 3.18: From the data presented in Figure 3.17: a. Variation of Bit Line voltage gradient for the '0' and '1' states, as well as ΔV_{ref}, as a function of Bit Line size for the 7nm HZO stack. ; b. Top-view of 16kbit, 64kbit and 256kbit arrays from layout presenting different subdivisions: whole array (orange), equivalent 16kbit array (purple), all WLs and 128 first BLs (white), and all BLs and 128 first WLs (red). ; Corresponding slopes of the '0' (c.) and '1' (d.) states distributions versus array size for the different subarrays.

While this might seem critical, it is important to keep in mind that the gradient comes from a design-related issue and should be easily fixed. Nonetheless, this part evaluates the impact of the BL voltage gradient on large-sized arrays. While Chapter 2 proposed to extend the array size along WLs rather than along BLs for larger arrays in order to avoid increasing C_{BL} and therefore decreasing MW, Figure 3.17 shows that this strategy could be severely detrimental. Indeed, while the upper row presents the '0' state bitmaps of a 16kbit, a 64kbit and a 256kbit array integrated on the same die on a TiN/HZO (7nm)/TiN wafer, the BL voltage gradient is clearly visible. On the bottom row of Figure 3.17, it can be seen that $V_{BL}{}^0$ and $V_{BL}{}^1$ values decrease with increasing array size in the WL direction (effect of C_{BL}), and that the value of gradient increases with increasing array size in the BL direction (effect of V_{ref}), thus reducing both MW(0σ) and MW(16k). Figure 3.18.a. demonstrates that the '0' and '1' states gradient depends linearly on the array size. As is shown in Figure 3.18.c.

and Figure 3.18.d. with the yellow dots, the distribution slopes follow the same trends as the Bit Line voltage gradient, indicating that the latter constitutes a major contributor to array-level variability. To confirm this result, Figure 3.18.b. proposes to compute the distributions slopes independently for different subdivisions of the arrays: a 16kbit array equivalent (purple), all WLs but only the 128 first BLs, starting from BL<0> (white), all BLs but only the 128 first WLs, starting from WL<0> (red). As can be seen in Figure 3.18.c. and Figure 3.18.d. for '0' and '1' states, the BLs are the main contributors to the distribution slopes, whereas considering only the 128 first BLs (as in a 16kbit array) is beneficial for the distribution slopes, and therefore for variability. It is therefore crucial to correct the source of the BL voltage gradient.

Conclusion

The goal of this chapter was to explore variations to the ferroelectric stack, technology innovations and process- and design-induced variability in view of improving the raw performances of FeCAPs and FeRAM arrays at the 130nm node for future integration at more aggressive nodes. To this end, the MW performance was studied in depth, and the focus was split between median memory performance and array-level variability.

For the first part, it was demonstrated that either etching the ferroelectric layer, or inserting a Ti layer between the ferroelectric layer and TiN top electrode each had the beneficial effect of improving $2.P_R$ at scaled BEoL-integrated FeCAP level, and thus MW(0σ) at 16kbit FeRAM array level. The mechanisms behind these performance gains were proposed to be stress release, and optimization of the oxygen vacancies content through O-scavenging by Ti, respectively. In both cases however, the gain in $2.P_R$ was obtained at the expense of a degraded endurance, and favored dielectric breakdown. Indeed, the quality of the etching step, and the further increase in oxygen vacancies content during cycling were respectively pinpointed to favor the formation of conductive paths between the electrodes. While the first one could be improved through process optimization, the second one is more challenging to correct.

In addition to MW(0σ), bitcell-to-bitcell variability is carefully studied, all the more so that it reduces MW(16k) and therefore the read margin at FeRAM array level. To identify the sources of intrinsic variability at array level, a first part evaluated the contribution of process-induced variability. From material-related and dimensional metrics measurements at array and wafer scale, it was shown that process variability did not contribute significantly to the distribution slopes of both logic states. To evaluate design-related contributions, it was shown that the contribution of transistor variability along BLs and read gradient across BLs amounted each for around $10mV/\sigma$, so nearly half of the total distribution slopes each. Fixing the IR drop occurring during the read operation would enable to considerably decrease the gradient-related variability and thus improve the read margin at array level.

Finally, according to recommendations issued in Chapter 2, some technological advancements were demonstrated. Regarding the high capacity requirement, the effect of C_{BL}-limited MW in larger arrays was demonstrated on 16kbit, 64kbit and 256kbit arrays. To overcome this issue, bank partitioning was emulated and an alternative charge-based reading system was evaluated. Both solutions countered the MW decrease with increasing BL size. Regarding the high density requirement, 2D+3D FeCAP structures were fabricated on ShL wafers and confirmed the performance gain at constant footprint provided by 3D capacitors. Finally, crystallization of a 4nm thick ferroelectric film with Nanosecond Laser Anneal was proposed to address the voltage limitation in FeRAM arrays at advanced nodes. Those advancements constitute intermediary demonstrators that need to be further developed for future implementation in FeRAM arrays at advanced nodes.

4 Reliability Study of HfO_2-based Memories, from Ferroelectric Capacitor to FeRAM Array at the 130nm Node

Introduction

Chapter 3 focused on the Memory Window (MW) at FeRAM array level (*i.e.* MW(16k)) through its two main contributors: MW(0σ), and bitcell-to-bitcell variability, the latter being visualized by the distribution slopes. It also aimed at showcasing different technological realizations achieved towards FeRAM array integration at advanced nodes, while taking into account critical constraints and limitations. The MW performance was evaluated for various technological variants in given time and temperature conditions, for example at room temperature after wake-up. Chapter 4 aims at providing an extensive reliability picture of BEoL-integrated scaled FeCAPs and FeRAM arrays at the 130nm node. It will focus on three key reliability aspects, namely retention, endurance and resistance to radiations. Indeed, imprint and fatigue failure mechanisms constitute a major concern for ferroelectric memories, in particular for the HfO_2-based FeRAM technology. Therefore in the first part, the retention of a stored information will be assessed at BEoL-integrated scaled FeCAP and FeRAM array level with temperature acceleration, and quantitatively compared between both structures ; the imprint mechanism will be thoroughly studied based on electrical characterizations and demonstrated to be the main contributor to data retention loss ; finally, a possible recovery scheme from imprint will be demonstrated at FeRAM array level. Another reliability concern is the ferroelectric fatigue, which affects the endurance of FeCAP and FeRAM cycled at low electric field ; recent work [124] having demonstrated dedicated electrical operations aiming at recovering from fatigue, the second part of this chapter will study in depth different recovery strategies and their efficiency through PUND and First-Order Reversal Curve (FORC) electrical measurements ; finally, recovery demonstration at FeRAM array level will be provided. For aerospace applications, resistance to different kinds of radiations is needed ; in the third part of this chapter, we will study heavy ion (Au^{26+}) radiation effect on FeCAP and FeRAM arrays and evidence benefits of low doses on FeCAP ferroelectric performances (confirmed by consistent electrical and morphological measurements) and degradation of FeRAM arrays at higher fluences.

Summary

4.1 Data Retention . . . 107
4.1.1 Data Retention at Room Temperature . . . 107
4.1.2 Temperature-Accelerated Data Loss . . . 109
4.1.2.1 Experimental Strategy . . . 109
4.1.2.2 Limited Role of Thermal Depolarization . . . 111
4.1.2.3 Major Influence of Imprint . . . 112
4.1.2.4 Influence of Design in Data Retention at FeRAM Array Level . . . 116
4.1.2.5 Recovery Capability . . . 119
4.2 Endurance . . . 121
4.2.1 Performance . . . 122
4.2.1.1 Dielectric Breakdown and Ferroelectric Fatigue . . . 122
4.2.1.2 Endurance Exploration towards Fatigue and Potential Recovery . . . 124
4.2.2 Recovery Capability . . . 126
4.2.2.1 Efficiency of Different Recovery Strategies on FeCAP . . . 126
4.2.2.2 Towards Recovery Mechanisms Understanding . . . 128
4.2.2.3 Multiple Recovery Demonstration at FeRAM Array Level . . . 132
4.3 Radiation Hardness . . . 133
4.3.1 Effect of Heavy Ion Irradiation on BEoL-Integrated FeCAPs . . . 133
4.3.1.1 Electrical Characterization . . . 133
4.3.1.2 Morphological Characterization . . . 135
4.3.2 Effect of Heavy Ion Irradiation on BEoL-Integrated FeRAM arrays . . . 137

4.1 Data Retention

Data retention describes the ability of a memory cell to retain the stored information (or data) for a definite period of time for a specific mission profile (*i.e.* at a specific average temperature): the information has to be readable (*i.e.* a programmed '0'/'1' should be read as a '0'/'1' after retention). In ferroelectrics, data retention can be degraded by various factors. A first one, called retention loss, results in $2.P_R$ decrease over time at FeCAP level (and MW closing at FeRAM array level). Retention loss has two main root phenomena: internal depolarizing fields can cause domains to backswitch ; thermal depolarization can lead to transition from ferroelectric to paraelectric phase when the data is stored at temperatures approaching the material's Curie temperature [125]. A second key factor, called imprint, results in a shift of the coercive voltage that contributes to stabilizing the stored state [40]. From a strict retention point of view, this case is very favorable. However, imprint has the combined effect of destabilizing the complementary (or opposite) state, due to higher effective coercive voltage and apparent polarization loss [125]. The double-sided effect has to be considered as a whole to evaluate retention of the stored information, as well as ability to re-program the cell after retention. In the following, we will evidence the impact of these factors and give a quantitative comparison of retention at FeCAP and FeRAM array level.

4.1.1 Data Retention at Room Temperature

In order to evaluate the data retention of FeRAM arrays, a preliminary study was conducted at room temperature on a TiN/HSO (10nm)/TiN 16kbit FeRAM array with 0.36µm² FeCAPs. The main goals targeted were to assess the integrity of the programmed bits, as well as the reliability of the array after long-term storage. To this end, a checkerboard pattern was programmed across the array, which was then stored in a closed dark box that remained untouched during 348 days. The measurement protocol is described in Figure 4.1.a., with voltages set to 4.8V for programming and reading operations and periphery alimentation. Figure 4.1.b. shows the post-retention bitmap obtained after a digital read operation with a unique V_{ref}(read) = 0.77V, chosen exactly in the middle of MW(16kbit) from pre-retention '0' and '1' states distributions. It can be observed that the checkerboard pattern remains intact, and therefore inferred that no bitfails occurred due to long-term storage (348 days) at room temperature. This promising result is further strengthened by post-retention '0' and '1' states distributions evaluation shown in Figure 4.1.c. Indeed, no extrinsic bitfails are observed, and the distributions are almost unchanged compared to pre-retention. From there, two points can be stressed:

- Firstly, as no bitfails were measured, 10-year retention could not be extrapolated. Further retention assessments could include a longer storage time at room temperature, which is not the preferred option as there is no guarantee that the read operation would reveal bitfails, even after two or three years of storage. Another option is to accelerate the mechanisms that cause bits to fail, usually in conditions of accelered aging (*e.g.* at higher temperature).

- Secondly, although pre- and post-retention distributions are nearly superimposed, it can be seen that post-retention '0' state does not follow a normal distribution perfectly. This will be further analyzed, as it could indicate a degradation mechanism that could lead to failure.

As can be detected in Figure 4.1.d., the post-retention read bitmap obtained through a V_{ref} scan operation for the '0' state reveals the usual BL voltage gradient, but also an unusual pattern seemingly linked to the checkerboard programmed before long-term storage at room temperature. Bitmap for the '1' state exhibits only BL voltage gradient, but no additional pattern. To understand the effect of the pattern stored before retention on the post-retention distributions, differentiated '0' and '1' states distributions are plotted in Figure 4.1.e. with a different color code depending on the state stored. Slight non-normality of the post-retention '0' state distribution is due to two split 8kbit populations that bear the memory of the state stored for close to a year and have shifted in opposite directions. This could be a feature of imprint, and could lead to read failure after longer storage times, or in arrays with a smaller MW(16k), either from reduced MW(0σ) or larger bitcell-to-bitcell variability.

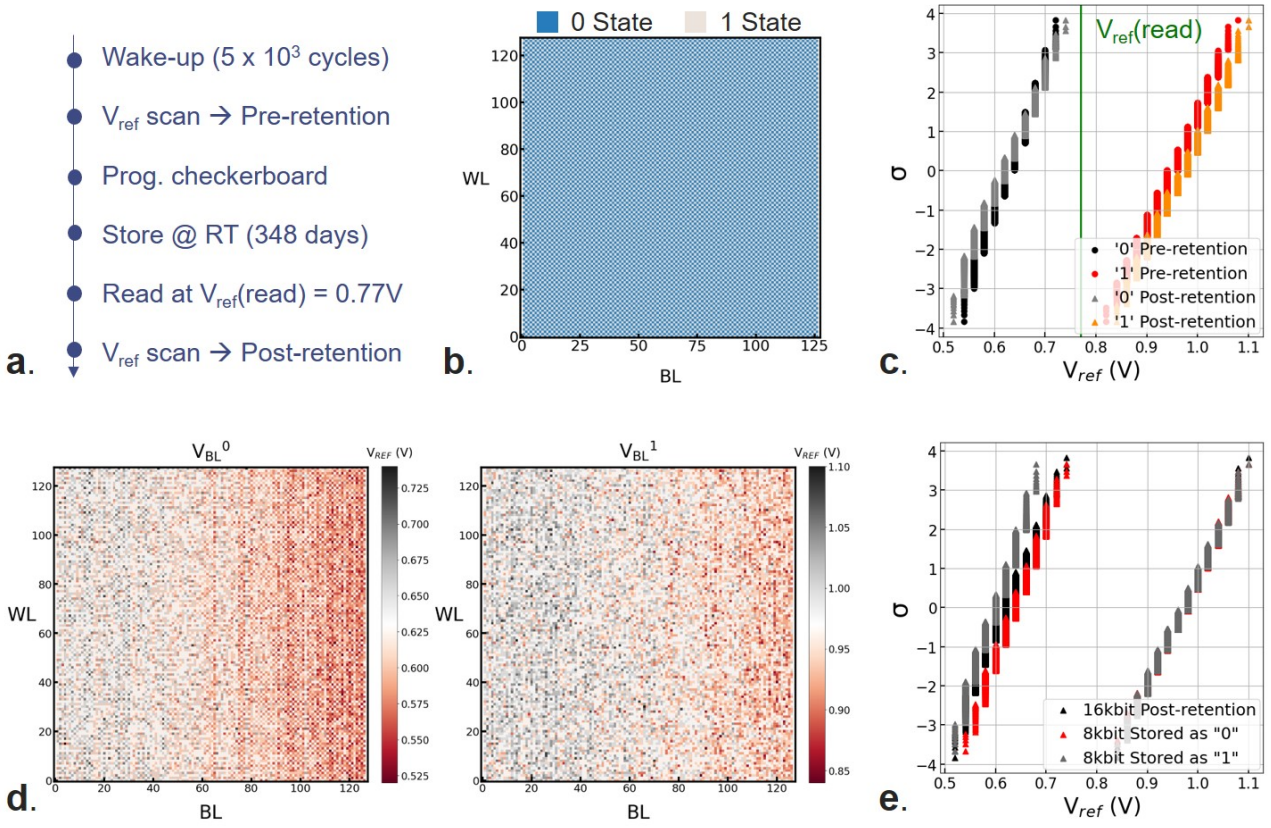


Figure 4.1: a. Measurement protocol to evaluate long-time retention at Room Temperature (RT) of 1T-1C 16kbit array ; b. 16kbit FeRAM array with 0.36µm² FeCAPs bitmap of a checkerboard pattern programmed 348 days prior ; c. Pre- and post-retention '0' and '1' states distributions featuring V_{ref}(read) ; d. $V_{BL}{}^0$ and $V_{BL}{}^1$ bitmaps resulting from post-retention V_{ref} scan ; e. Post-retention '0' and '1' states distributions and their differentiated distributions depending on the state stored. In all measurements, programming and reading operations rely on 4.8V - 2µs square pulses.

To accelerate bit rate failure due to imprint and other potential degradation mechanisms, storage at higher temperatures is adopted. To decorrelate the effects of thermal depolarization and imprint, a specific characterization pattern needs to be defined.

Recap

The long-term (348 days) retention at room temperature of a 16kbit HSO-based FeRAM array shows no bitfail and slight history effect of state stored on post-retention distributions. Though they are promising, they do not enable 10-year retention estimation, therefore thorough retention assessment of FeCAP and FeRAM arrays needs to be further conducted in accelerated conditions.

4.1.2 Temperature-Accelerated Data Loss

4.1.2.1 Experimental Strategy

Due to the destructive operation in FeCAPs and FeRAM arrays, a more complex data retention evaluation protocol is needed than for most memory technologies. Data retention at the BEoL-integrated scaled FeCAP level can be assessed by a specific pulse sequence as introduced in [110, 40], which aims at analyzing the impact of a thermal bake on each logic state programmed before bake, as well as each complementary state programmed and read at room temperature after the bake. Figure 4.2.a. presents the voltage pattern applied to two nominally identical FeCAPs to evaluate three different retention states that specifically target different degradation mechanisms. The Same-State (SS) corresponds to the state baked and the first pulse (blue) aims at evaluating the integrity of the stored information. It can be impacted by short-term retention loss and thermal depolarization [126, 110, 125]. The New Same-State (NSS) is then programmed at room temperature in the same polarity as SS, and read immediately afterwards (red). It should not be impacted by short-term retention loss, and the effects of thermal depolarization should be reversed [126, 110, 114]. Finally, the Opposite State (OS) is programmed at room temperature in the opposite polarity as SS, and read immediately afterwards (blue). It is expected to be most strongly impacted by the imprint phenomenon [40, 110]. SS (state baked), NSS (state baked restored after bake) and OS (state complementary to state baked) retention is therefore assessed after a series of bakes at 85°C, 105°C and 125°C for 10^3s, 10^4s and 10^5s, as shown in Figure 4.2.a. Different storage times and temperatures were investigated in order to be able to extract acceleration factors that will help provide accurate retention extrapolation. Write and read fields have been set equal and were varied from 3MV/cm to 5MV/cm. Nominally identical devices from different dies on TiN/HZO (10nm etched)/TiN wafer were used to investigate the effect of different write and read fields, and bake temperature conditions. Data 0 (respectively Data 1) describes the result of a read operation performed after bake on a bit programmed in a logic '0' (respectively '1') state by a a positive (respectively negative) pulse on FeCAP top electrode. For each state, 2.P_R Margin is defined as the remanent polarization integrated from the ferroelectric switching current of Data 1 - Data 0. On the contrary, P_{SW} is integrated directly from either Data 0 or Data 1 thus taking into account the dielectric contribution.

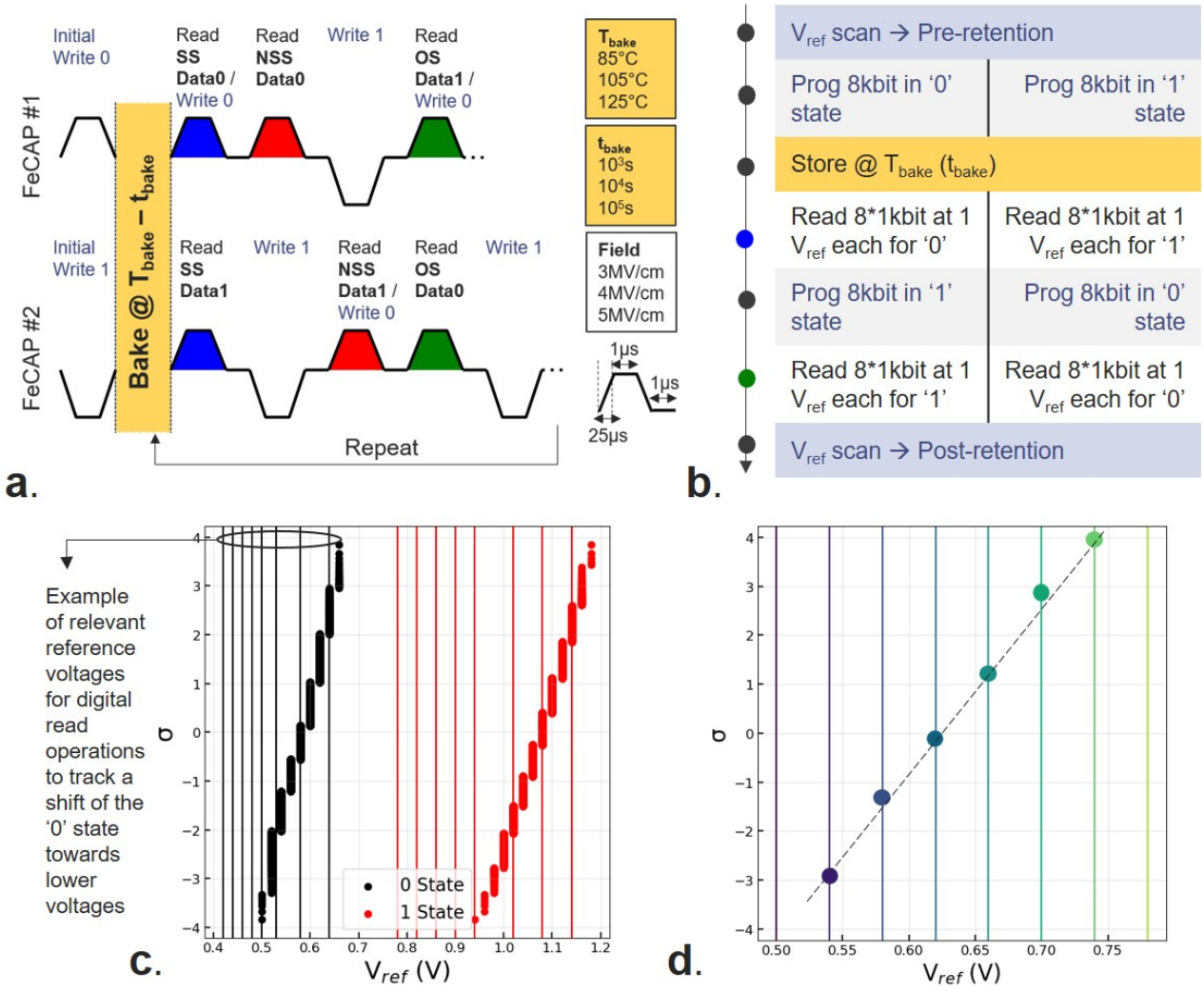


Figure 4.2: a. Voltage-time pattern applied to Top Electrode (TE) of 2 nominally identical FeCAPs to evaluate Same State (SS), New Same State (NSS) and Opposite State (OS) retention. The effect on retention of various bake times, temperatures and write and read fields is investigated. ; b. Measurement protocol to evaluate SS and OS retention of 16kbit array after bakes at different temperatures (T_{bake}) and times (t_{bake}). ; c. Representation of different V_{ref} voltages used for digital operation on 1kbit subarrays to track distribution shifts. ; d. Example of a successfully reconstructed distribution for a 16kbit array with 6 points (over 8V $_{ref}$ voltages).

The retention response to this pattern with opposite polarities, *i.e.* with a negative read voltage applied to TE, was also investigated to assess stack symmetry. Figure 4.2.b. presents an equivalent experimental strategy developed to assess SS and OS data retention at 16kbit FeRAM array level. The goal for data retention evaluation at FeRAM array level is to study the role of design as in [126] and therefore determine whether data retention studies at single device level are sufficient to extrapolate retention behavior in circuits. Similarly to the previous section reporting results from long-term

retention at room temperature, pre- and post-retention distributions were measured with a V_{ref} scan operation, and the retention of the stored information was assessed directly after bake through digital read at a fixed V_{ref}. To assess SS (OS, respectively) of both logic states simultaneously and thus spare some bake time and FeRAM arrays, it was decided to split the 16kbit in 8kbit subarrays, programmed in '0' and '1' states ('1' and '0' states, respectively) with proper scrambling in order not to be impacted by any design-induced issue. In order to assess the integrity of the information stored at array level, and to evaluate the effect of thermal bakes on both logic states distributions, each 1kbit of the 8kbit was read at a different V_{ref} after the bake to overcome destructive read limitations and still reconstruct distributions. Prior to data retention assessment at FeRAM array level, it was successfully verified that '0' and '1' states distributions on 1kbit subarrays could be used as equivalents to 16kbit, with a similar median and standard deviation. Figure 4.2.c. presents an example of V_{ref} set chosen to anticipate a slight shift of $V_{BL}{}^{0}$ towards lower voltages, and a larger shift of $V_{BL}{}^{1}$ in the same direction. Figure 4.2.d. shows that a well-chosen V_{ref} set enables the reconstruction of a distribution, even with a limited number of points. '0' and '1' states distributions at full 16kbit array level can therefore be reconstructed from digital measurements at subarray level, which enables the tracking of thermal bake induced effects on the device. This technique, already used in [68] for SS retention evaluation, was completed here to include OS retention. Note that stack asymmetry could not be assessed at FeRAM level as the reading operation does not allow for negative pulses on TE or positive pulses on BE, and that the operating field was kept constant at 4MV/cm.

Recap

At FeCAP level, specific electrical patterns are introduced and enable the characterization of thermal depolarization on both logic states (SS), and the analysis of the imprint effect on both complementary states (OS). At FeRAM array level, digital read on subarrays at different V_{ref} enables the reconstruction of '0' and '1' states distributions after bake despite destructive reading, and hence the study of SS and OS retention to evidence potential design-related effects.

4.1.2.2 Limited Role of Thermal Depolarization

Figure 4.3.a. presents SS 2.P_R Margin of FeCAPs programmed and read at 4MV/cm as a function of bake time for bake temperatures of 85°C, 105°C and 125°C. For each temperature, a clear symmetry between 2.P_R Margin positive and negative polarities is visible, indicating that top and bottom electrode interfaces are not fundamentally different from a data retention viewpoint. This result is in line with the data presented by [110, 127] and, in the following, only the data read with a positive pulse applied to TE will be analyzed. SS state is found to be stable over bake time for all temperatures, pointing towards limited thermal depolarization. To confirm that this 2.P_R Margin stability does not result from a compensation of equivalent polarization loss in Data 0 and Data 1, Figure 4.3.b. shows SS and NSS switched polarizations as a function of bake time for the intermediate bake temperature of 105°C. As can be seen on the graph, SS Data 0 and Data 1 do not significantly degrade with bake time. Moreover, it should be noted that NSS is programmed at room temperature and read

shortly after, therefore expected not to be affected by potential thermal depolarization: should SS be degraded, NSS would exhibit a clear 2.P_R Margin gain, which is not the case after bake. Both results indicate that the influence of thermal depolarization, expected to lower post-bake remanent polarization [40], is very limited. The most probable explanation is that the bake temperatures investigated, based on consumer or industrial applications requirements, are too low compared to the material's Curie temperature (estimated in the range of 600K to 1000K by [128]) to degrade SS retention. Moreover, Mueller et al. [110] evidenced that the thermal stability of HfO_2-based FeCAPs was really high on a wide range of bake temperatures.

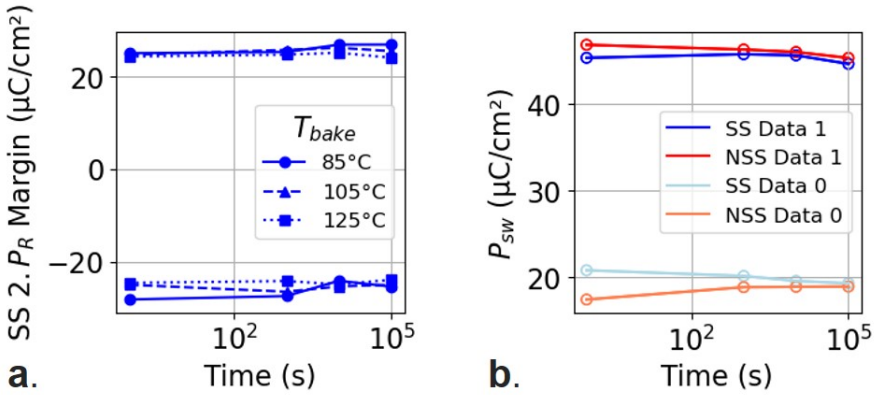


Figure 4.3: From FeCAPs programmed and read at 4MV/cm: a. Same State 2.P_R Margin as a function of bake time (at 85°C, 105°C, 125°C) ; b. SS and NSS P_{sw} for Data0 and Data1 as a function of bake time (at 105°C)

Recap

The storage temperatures investigated do not lead to thermal depolarization at FeCAP level. While short-term retention loss can be measured on ferroelectric capacitors directly after their programmation in either one of the logic states, its effect will not be accentuated with time and therefore short-term retention loss will be a second-order effect in 10-year data retention. It is thus crucial to rather evaluate the effect of imprint on OS retention, which could be a significant retention-limiting factor.

4.1.2.3 Major Influence of Imprint

To understand the electrical signature of OS retention degradation, Figure 4.4 delves deeper into the OS retention of FeCAPs programmed at 4MV/cm and subjected to a series of thermal bakes at 105°C. The individual I-V (Figure 4.4.a.) and P-V (Figure 4.4.b., left), as well as P-t_{bake} (Figure 4.4.b., right) characteristics of OS Data 0 and OS Data 1 are looked into. From the I-V curves, it can be

seen that the ferroelectric switching current peak of OS Data 1 shifts towards lower voltages with increasing bake time, consistently with results from [110, 129, 115]. Because of this shift in the coercive voltage, or imprint, the current peak is increasingly truncated with bake time, leading to a reduced amount of switchable domains, and thus to a reduced polarization, as can be seen on the P-V curve. Hence, because imprint prevents polarization switching in an increasing number of domains with bake time, OS Data 1 P_{SW} decreases, as can be seen in P-t_{bake}. This coercive voltage shift is accompanied by a remanent polarization loss due to imprint and matches data presented in [115]. Simultaneously, the I-V of OS Data 0 evidence a non-zero current peak which increases with bake time. This means that after baking a FeCAP in the '1' state, the subsequent attempt at programming a '0' (OS Data 0) will not lead to a pure dielectric state but rather to a dielectric state with a ferroelectric contribution that increases with bake time. This phenomenon gives rise to an increasing polarization and thus to an increasing OS Data 0 P_{SW} with bake time. Both trends contribute equally to OS 2.P_R Margin degradation.

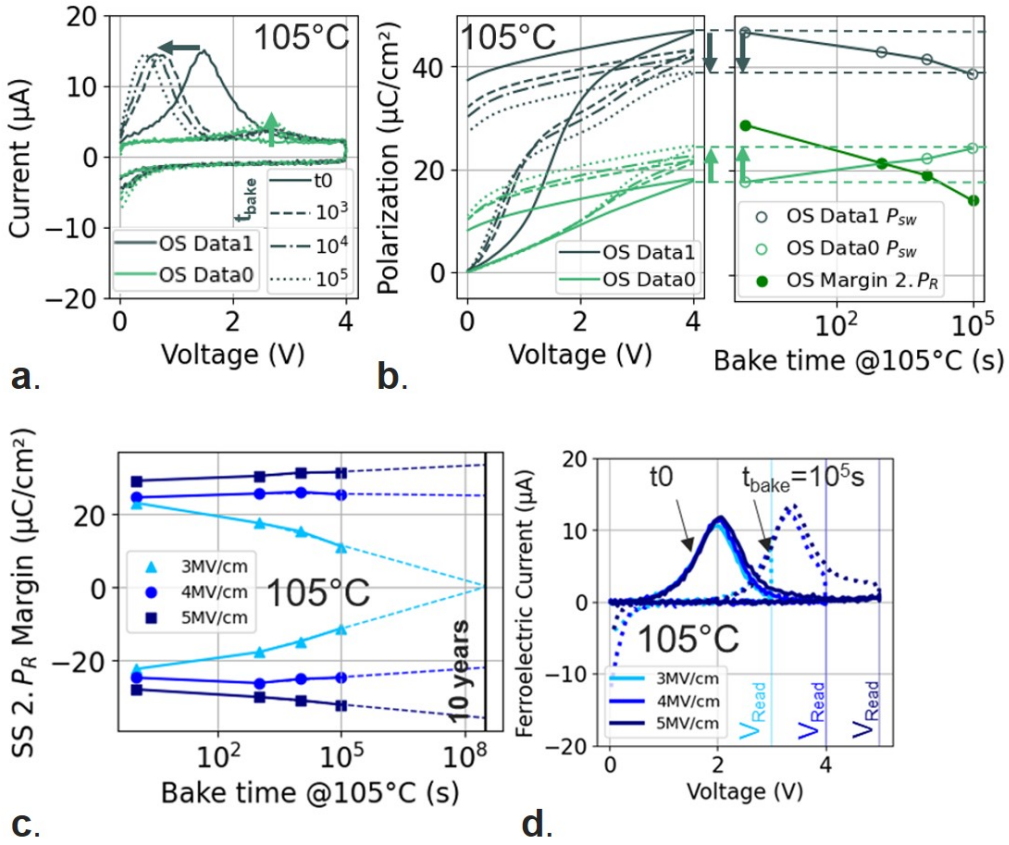


Figure 4.4: a. I-V and b. (left) P-V and (right) P-t_{bake} from OS retention test on FeCAPs operated at 4MV/cm, with T_{bake} = 105°C. ; c. SS 2.P_R Margin-t_{bake} and d. I-V from SS retention test (SS Data 1 - SS Data 0) on FeCAP operated at different write/read fields, at t_0, and after bake at T_{bake} = 105°C and t_{bake} = 10^5s.

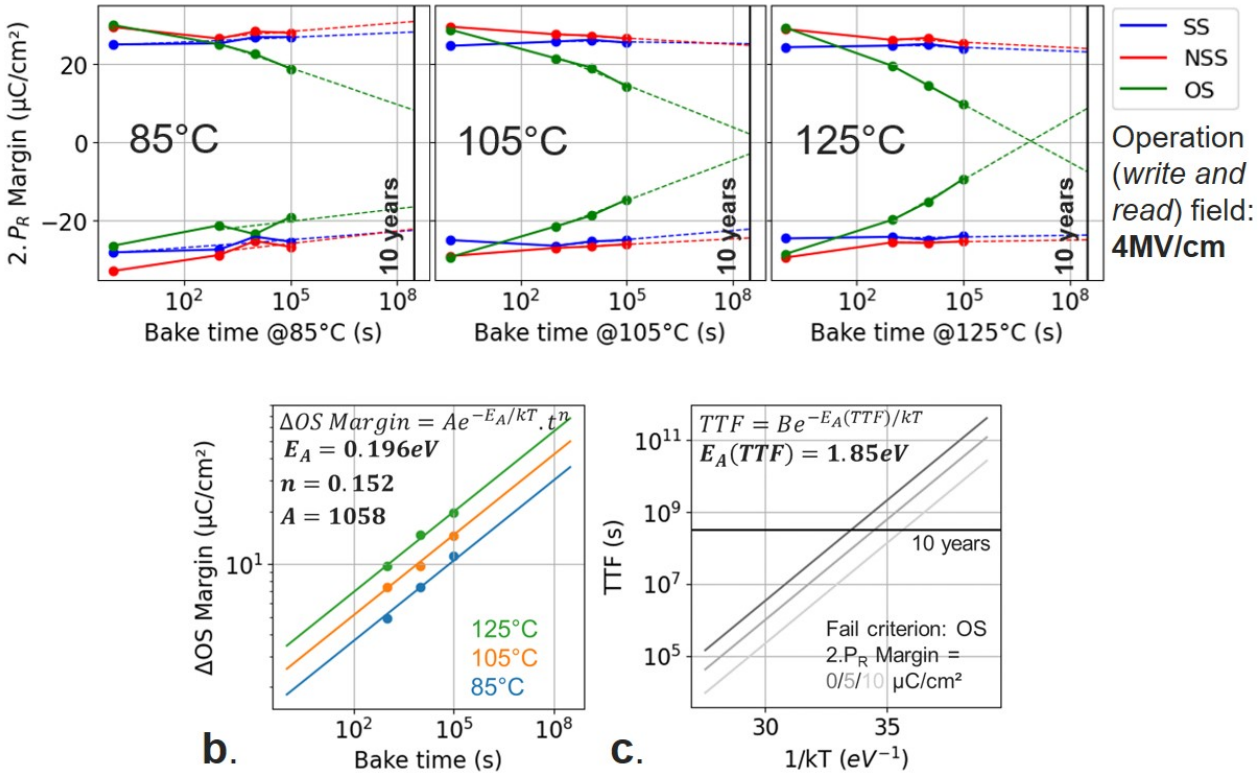


Figure 4.5: a. 2.P_R Margin for SS, NSS and OS along bake time for bake temperatures of 85°C, 105°C and 125°C. Positive margin values correspond to pattern presented in Figure 4.2.a. at 4MV/cm (read by positive pulse on TE), while negative margin values account for opposite voltages (read by negative pulse on TE). ; b. OS 2.P_R Margin shift from initial value (at t0) due to thermal bake fitted by a power law along bake time and temperature ; c. Time-To-Fail modelled by an Arrhenius plot.

Imprint occurring during bake also impacts SS retention to some extent. From Figure 4.3.a., the results presented at 4MV/cm showed no degradation of SS retention. However, Figure 4.4.c. shows that SS 2.P_R Margin can degrade rapidly over bake time when data is written and read at low field. The trends at low and high field are consistent with results from [115]. Looking at the ferroelectric current (Figure 4.4.d., from Data 1 - Data 0), writing field is found to have no influence on the post-bake peak position, *i.e.* imprint is not field-dependent, consistently with [129]. Yet, reading field strongly impacts the as-measured polarization as only some domains have their coercive field lying below E_{read}. Thus, imprint is responsible for SS (and NSS) 2.P_R Margin poor performances at low field but results from an unadapted measurement window rather than a physical polarization loss. Consistent observations were made in [130, 131, 114] and therefore, 2.P_R Margin performances could be largely improved in an array by decorrelating write and read fields. Moreover, reducing HZO thickness should also favorable to improved 2.P_R Margin performances as coercive voltages are decreased and breakdown field is increased [90]. It is worth noting that this study investigated programming fields that enabled saturated switching: no conclusions can be

drawn on field-dependency of imprint beyond these conditions.

Now that the electrical signature of imprint has been studied, it is important from an application standpoint to extrapolate 10-year retention. Extensive measurements of FeCAP OS retention loss due to thermally-accelerated imprint provide a perfect opportunity to model degradation trend and extrapolate retention.

Figure 4.5.a. gathers $2.P_R$ Margin results of SS, NSS and OS retention states in positive and negative polarities at 4MV/cm, along bake time and for bake temperatures of 85°C, 105°C and 125°C respectively. A semilog linear fit is used as a visual guide to assess 10 year-data retention. Again, positive $2.P_R$ Margin corresponds to the retention scheme presented in Figure 4.2.a., while negative $2.P_R$ Margin accounts for similar scheme with a read operation performed by negatively pulsing the TE. OS retention degrades much more quickly with bake time than SS and NSS because of imprint, as evidenced in Figure 4.4.a. and Figure 4.4.b. This degradation is found to be highly dependent on bake temperature, as already shown on PZT [40] and HZO [132], thus making OS the limiting state of data retention in ferroelectric memories, and imprint the main factor for data retention loss at FeCAP level.

For this reason, OS retention degradation is modelled in Figure 4.5.b. The metric used represents the decrease of OS $2.P_R$ Margin compared to the initial value (at t0, *i.e.* before bake) and is labelled ΔOS Margin. The experimental points of ΔOS Margin are found to be well-described by a unique thermally-activated power law

$$\Delta OS\,Margin(T,t) = Ae^{(E_A/kT)}.t^n \tag{4.1}$$

where fitting parameters are adjusted as follows: E_A, the activation energy for OS degradation, is 0.196eV, the pre-factor A is 1058 and the n-factor is 0.152. The power law thus fitted can then be used to model Time-To-Failure (TTF) as

$$TTF(T) = Be^{(E_A(TTF)/kT)} \tag{4.2}$$

in an Arrhenius plot. From (4.1), it follows that $B = \Delta OS(\text{fail})/A$ depending on the fail criterion, and $E_A(TTF) = E_A/n = 1.3eV$ regardless of the chosen fail criterion. To get an estimate of the maximal storage temperature for 10-year retention, the fail criteria used in Figure 4.5.c. account for an OS $2.P_R$ Margin (see Figure 4.5.a.) reaching 0, 5 or $10\mu C/cm^2$, respectively. Those low remanent polarization values are chosen with regard to the SA sensitivity and forebode strongly degraded performances at FeRAM level. The relationship between E_A and $E_A(TTF)$ denotes a strong link between a physically-based activation energy (E_A) and a lifetime-related effective activation energy ($E_A(TTF)$). With an initial remanent polarization of $30\mu C/cm^2$ and based on $0\mu C/cm^2$ to $10\mu C/cm^2$ fail criteria, extrapolated 10-year retentions at 74°C to 54°C are projected, respectively. Considering the microscopic imprint model proposed and supported by DFT simulations from Yuan et al. [115], $E_A = 0.196eV$ could be attributed to the position of oxygen vacancy related defects at the TiN/HZO interface. A change in the charge state of oxygen vacancies during retention bakes [40] is thought to activate shallow de-trapping sites that could affect the internal field [115]. A similar E_A of 0.2eV is reported in [133] and linked to oxygen vacancies in hafnium-based stacks.

Recap

Imprint occuring during high-temperature bake causes a shift of the coercive fields at FeCAP level. For SS, reading fields therefore need to be carefully chosen and might need to be increased to overcome imprint. Because of imprint-induced effective polarization loss, OS degrades rapidly with bake time and temperature and is therefore the limiting retention state. Modelling its time- and temperature-dependent degradation at FeCAP level requires the adjustment of fitting paramaters, whose values can then be linked to oxygen vacancies trap levels at TiN/HZO interface.

4.1.2.4 Influence of Design in Data Retention at FeRAM Array Level

The evaluation of the data retention performance at FeRAM array level is essential from an application perspective as it brings a statistical outlook on results obtained at the FeCAP level and can evidence additional array-related factors that impact FeRAM data retention compared to FeCAP. To evaluate the impact of thermal bakes on the distributions of both logic states, and because of the destructive read operation, a specific procedure (described in Figure 4.2.b.) enables the reconstruction of '0' and '1' states distributions based on digital read operations at 1kbit subarray level.

The distributions reconstructed in Figure 4.6.a. evidence a moderate shift of SS Data 0 towards lower reference voltages with bake time, and a stronger shift of SS Data 1 in the same direction. As can be seen in Figure 4.6.b. OS Data 0 shifts towards higher V_{ref} values, while shift of OS Data 1 is stronger and in the opposite direction. The direction and relative amplitude of the SS and OS distribution shifts for both logic states are consistent with FeCAP data to be presented in Figure 4.7. It is noticeable that '0' and '1' states slopes do not vary along bake time (and temperature, not shown here), demonstrating homogeneous data loss over the array with bake time and temperature. Given that OS Data 0 and OS Data 1 shift towards a central V_{ref} with bake time, hence closing the Memory Window, it can be inferred that OS still constitutes the limiting factor for data retention at FeRAM array level.

Similarly to Figure 4.5.b., ΔOS Margin in 16kbit array has been modelled and represents the loss of median MW for OS (Data 1 - Data 0) compared to pre-bake median MW. Figure 4.6.c. represents ΔOS Margin in a FeRAM array along bake time for different bake temperatures. ΔOS Margin can be fitted by a unique power law, as in (4.1), with $E'_A = 0.19eV$, $A' = 23.9$ and $n' = 0.154$. Table 4.1 summarizes the fitting parameters for OS degradation at FeCAP and FeRAM array levels.

Table 4.1: Fitting parameters of ΔOS Margin and Time-to-Fail

	$E_A\,(eV)$	$n\,(s^{-1})$	A	$E_A(TTF)\,(eV)$
FeCAP	0.196	0.152	1058	1.285
FeRAM	0.190	0.154	23.9	1.232

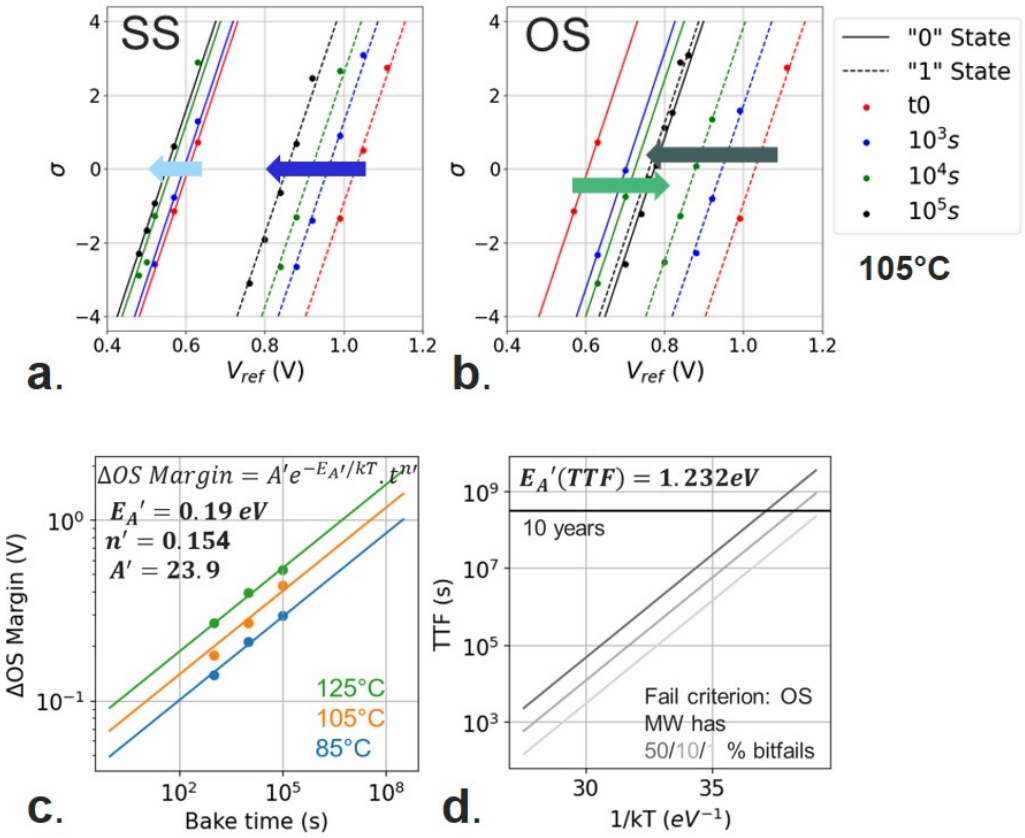


Figure 4.6: a. SS and b. OS distributions reconstructed for both logic states post-bake at 105°C. Individual capacitors of the 16kbit 1T-1C FeRAM array were programmed and read at 4MV/cm-2µs, and the experimental points were measured on subarrays read at a specific V_{ref} each. ; c. OS median Memory Window shift from initial value (at t0) due to thermal bake fitted by a power law along bake time and temperature ; d. Time-To-Fail modelled by an Arrhenius plot.

The similarity in activation energy and n-factor between FeCAP and FeRAM models highlights equivalent data retention dynamics in single devices and arrays, and confirms that data retention studies at FeCAP level are relevant to predict array behavior. A and A' are pre-factors of ΔOS Margin in FeCAP (see (4.1)) and FeRAM, respectively, and can be related with the MW formula (see Equation 2.16)

$$A' e^{(-E_A'/kT)} . t^{n'} = A e^{(-E_A/kT)} . t^{n} \times \frac{S}{C_d + C_{BL}} \tag{4.3}$$

where S is the FeCAP area, C_d its dielectric capacitance and C_{BL} the circuit Bit Line capacitance [86]. Given that the activation energy and n-factor of FeCAP and FeRAM have been found to be very close, it follows that $A'/A = S/(C_d + C_{BL})$, confirmed by numerical application with comparable dimensional units. At FeRAM level, TTF along $1/kT$ was also modelled by Arrhenius' law as in (4.2), with fail criteria of 1%, 10% and 50% bitfails (and assuming no extrinsic bitfails), taking advantage of

the constant standard deviation, and thus distribution slopes, along bake time. Hence, TTF bitfails account for intrinsic variability in the array and bake-induced distributions shifts at array level. E_A'(TTF) was found to be 1.2eV, equal to E'_A/n' and close to that of FeCAPs (Figure 4.5.c.). This strongly indicates that the imprint phenomenon governing OS data retention loss at FeCAP level occurs in the same way at FeRAM level. Based on 1% to 50% bitfails criteria at memory array level, extrapolated 10-year retention from 36°C to 51°C is projected at FeRAM level, respectively. Those values may differ from the ones obtained at FeCAP level because the fail criteria are not equivalent, or from additional array-related factors to be detailed.

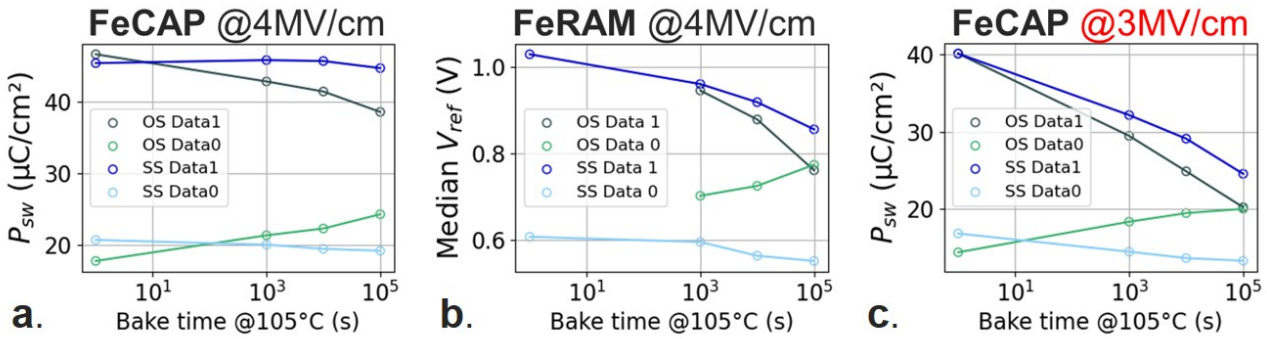


Figure 4.7: P_{sw} (a. and c.) and median V_{ref} (b.) as a function of bake time at 105°C for FeCAP (programmed and read at 4MV/cm and 3MV/cm, respectively) and 16kbit 1T-1C FeRAM array programmed and read at 4MV/cm, respectively.

From data shown in Figure 4.5 and Figure 4.6 however, SS and OS retention degradation with bake time appears stronger at FeRAM level. This is confirmed by Figure 4.7, which presents the evolution of FeCAP P_{SW} and FeRAM median V_{ref} along bake time at 105°C, for devices operated at 4MV/cm (a. and b.) or 3MV/cm (c.). All four retention states (OS Data 1, OS Data 0, SS Data 1 and SS Data 0) are investigated, and P_{SW} and V_{ref} can be related by analytical formulae presented in [68]. Global trends of shift direction and amplitude are found similar, but FeRAM time-dependent retention appears degraded compared to that of its FeCAP counterpart at 4MV/cm. Indeed, while imprint has been demonstrated to affect ΔOS Margin in the same way in FeCAP and FeRAM, pre-bake OS Margin values may differ between FeCAP and FeRAM. In this case, it is thought that the lower OS Margin value in FeRAM comes from a voltage drop (also called IR drop) in the array that leads to an overall reduced effective write and read voltage during operation at array level. As can be seen in Figure 4.7, the SS and OS trends of FeRAM data at 4MV/cm are consistent with FeCAP measurements at 3MV/cm. A preferred hypothesis would be that the voltage across the capacitor reduces as the floating node charges up during the read operation, leading to a decreased effective reading voltage across the FeCAP in a FeRAM array. To conclude, the results presented point toward an identical phenomenon governing data retention at FeCAP and FeRAM level, as well as additional design-related effects further degrading SS and OS retention in FeRAM arrays.

Recap

The imprint model provided is consistent from FeCAP to FeRAM array level. At FeRAM array level, the imprint-induced shift in OS distributions can be fitted equivalently to that of FeCAPs, indicating similar mechanisms for data retention loss in FeCAPs and FeRAM arrays. An additional voltage drop at array level reduces initial Bit Line voltage levels, thus further degrading retention at FeRAM array level.

4.1.2.5 Recovery Capability

In the previous parts, a specific electrical sequence was applied to assess SS retention and OS re-programming capability. To limit the number of FeRAM arrays under test and still reconstruct distributions, 1kbit subarrays made of 10 WLs were programmed in '0' and '1' states before the retention bake, and read at different V_{ref} voltages afterwards. Figure 4.8.a. presents a bitmap of the bitcells baked in the '0' (red) and '1' (grey) states, respectively. Previous results enabled the conclusion that SS retention remains unaltered with bake time and temperature when read at sufficient field. However, the complementary state is very difficult to re-program because of imprint, and this translates in a very limited OS retention.

To evaluate the effect of imprint beyond retention, and similarly to performance assessment after long-term storage at room temperature (Figure 4.1), a V_{ref} scan after a bake at 105°C during 10^4s was performed. Pre- and post-retention '0' and '1' states distributions shown in Figure 4.8.b. present an S-shaped distortion of the '0' state and a shift of the '1' state towards lower voltages. To understand the shape and median of these distributions, the bits were differentiated according to their state during the retention bake, and their attempted programmed state during the considered V_{ref} scan operation. Figure 4.8.c. presents post-retention distributions measured with V_{ref} scan, as well as their differentiated distributions that can be linked back Data 0 and Data 1 of SS and OS retention states. It can be seen that the S-shaped '0' state distribution in Figure 4.8.b. comes from bitcells stored in the '0' state and stabilized by imprint (SS Data 0) that have shifted towards lower V_{ref} values, and from bitcells stored in the '1' state and destabilized by imprint (OS Data 0) that have shifted towards higher V_{ref} values. The shifted '1' state distribution in Figure 4.8.b. can be attributed to a shift in SS and OS Data 1 in the same direction, as presented in Figure 4.7 (a. and b.). This is further confirmed by Figure 4.8.d. showing that the pattern observed arises from the pre-bake programming scheme of '0' and '1' states along WLs as presented in Figure 4.2.b. From an application standpoint (and not restricted to data retention), imprint (if not recoverable) could therefore permanently stabilize bits in a certain state and array operation could be compromised.

At single FeCAP level, it has been shown that imprint could be overcome by an electrical cycling sequence at field equivalent to or higher than the operating field [129, 115, 132]. On a FeRAM array presenting distorted or shifted distributions of SS and OS for both logic states following a bake at 105°C for 10^4s, bipolar cycling at operating field is demonstrated in Figure 4.9.a. to reverse the imprint effect and therefore recover most of the pre-bake memory window. This recovery, as first evidenced on FeCAPs, appears as a promising operating strategy for getting rid of retention limitation in FeRAM arrays cause by imprint of OS. However, as shown on '0' and '1' states distributions from V_{ref} scan (after a bake at 105°C for 10^5s) in Figure 4.9.b., a specific number

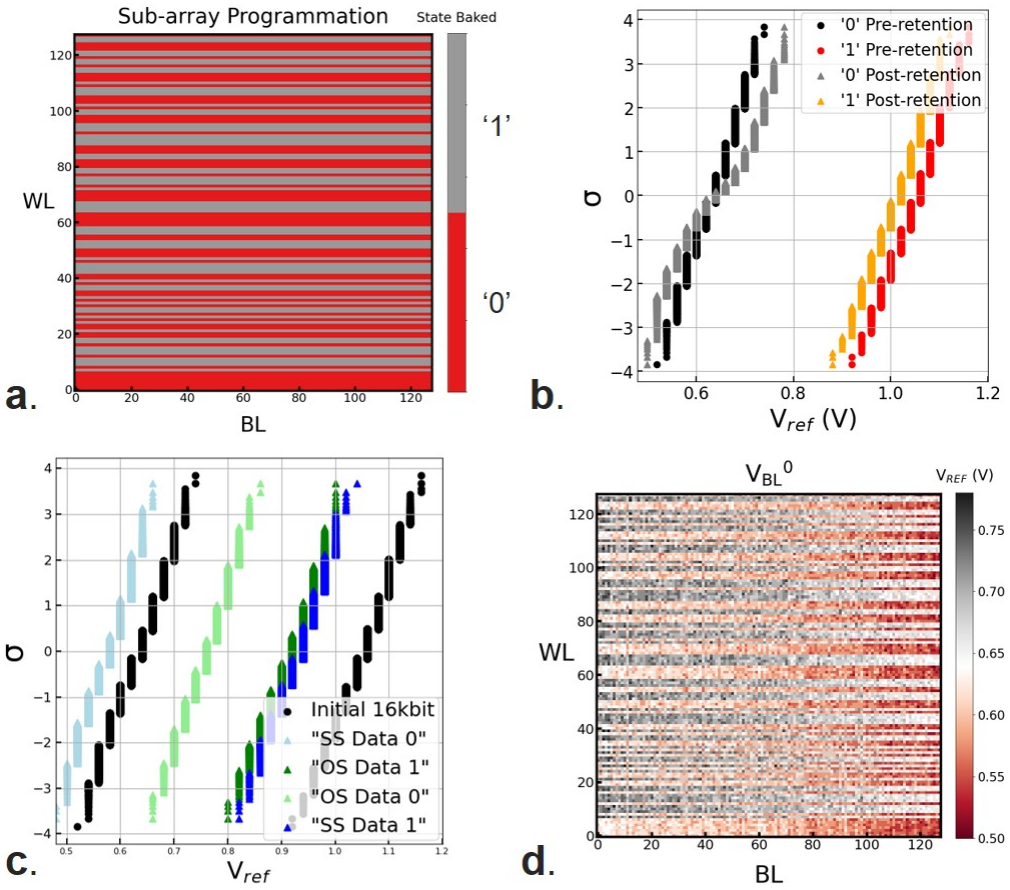


Figure 4.8: a. Bitmap summarizing the state of bitcells during retention bake. ; b. '0' and '1' states distributions from V_{ref} scan pre- and post-retention at 105°C for 10^4s ; c. Post-retention '0' and '1' states distributions and their differentiated distributions depending on the state of bitcells during retention bake. ; d. $V_{BL}{}^0$ bitmap resulting from post-retention V_{ref} scan.

of cycles needs to be applied for the optimal recovery of imprint without further degradation of FeRAM array reliability. The median reference voltage from differentiated distributions depending on the state baked is extracted from Figure 4.9.b. and represented at different recovery stages in Figure 4.9.c. It can be evidenced that differentiated distributions tend to merge with recovery cycling (to the point that no history effect can be detected after 10^6 cycles) but increasingly shift away from initial median values with cycling. An elevated number of cycles could cause ferroelectric fatigue, and although effectively recovering the history effect of imprint, could lead to other failure modes, namely fatigue-limited endurance, or breakdown-limited endurance. Imprint is therefore hard to recover as it impacts the device itself rather than the stored information. A dedicated electrical sequence is consequently needed for effective reversal to pre-bake characteristics.

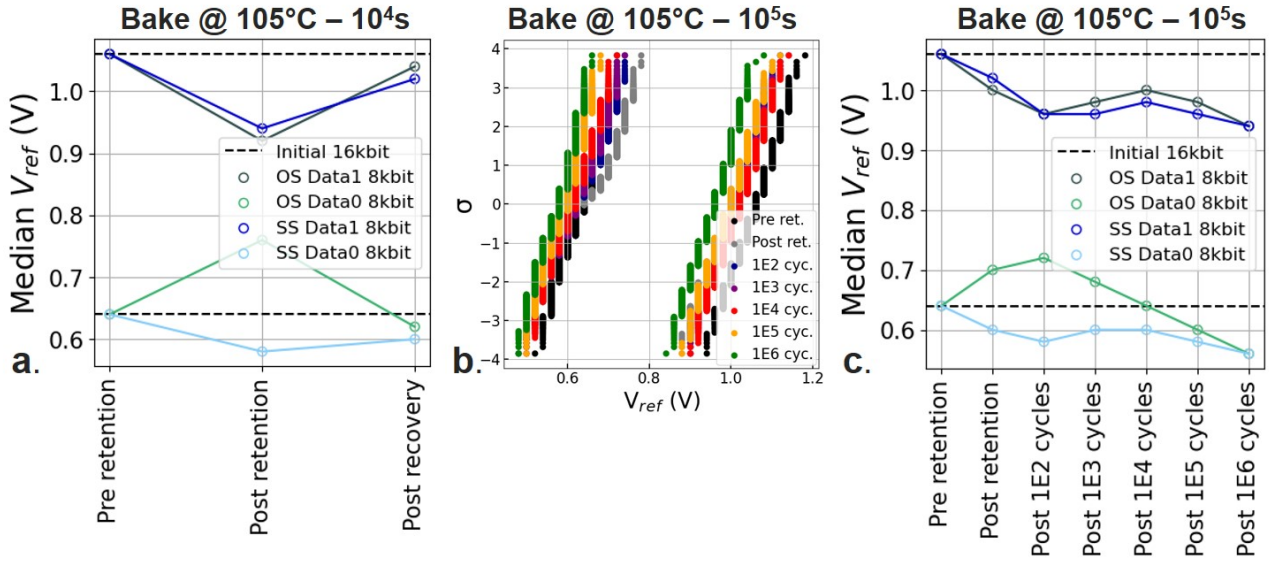


Figure 4.9: a. Median V_{ref} for all SS and OS retention states before bake, after bake at 105°C for 10^4s, and after 10^4 cycles at 4MV/cm. ; b. '0' and '1' states distributions from V_{ref} scan before retention, after retention bake at 105°C for 10^5s and after 10^2 to 10^6 recovery cycles ; c. Median V_{ref} for all SS and OS retention states before and after retention along cycling.

Recap

The effect of imprint persists after a reading operation, leading to distorted or shifted distributions that can degrade the MW. A cycling scheme at operating voltage overcomes this history effect, effectively recovering from imprint, but has to be controlled to avoid failure due to other mechanisms.

4.2 Endurance

In recent years, it has been well documented that two different phenomena limit the endurance of ferroelectric-based devices: as first mentioned in Chapter 3, dielectric breakdown can occur if oxygen vacancies (especially in high concentrations) form a conductive filament between top and bottom electrode, allowing current to flow within the oxide. Dielectric breakdown is particularly likely at high field, in large area capacitors, and in etched devices or devices with an oxygen scavenging layer (see Chapter 3). In FeCAPs, the uncontrolled formation of a filament permanently damages the oxide (ferroelectric) layer, and thus cannot be recovered. Another failure mode occurs at lower voltage, and is called the ferroelectric fatigue. It is evidenced by a decrease in the remanent polarization (or MW at FeRAM array level) with cycling and its root cause is still under discussion. Main phenomena under scrutiny are pinning of domains by charges trapped at the interfaces [124, 134], and inhomogeneous lateral distribution of oxygen vacancies [135]. It has been demonstrated that ferroelectric fatigue can be recovered, mainly through punctual application of pulses at higher

field than the operating field [124, 90, 136]. In this part, we will evidence effective recovery in FeCAP and FeRAM arrays, assess the efficiency of different strategies and try to shed light on mechanisms occurring during recovery.

4.2.1 Performance

4.2.1.1 Dielectric Breakdown and Ferroelectric Fatigue

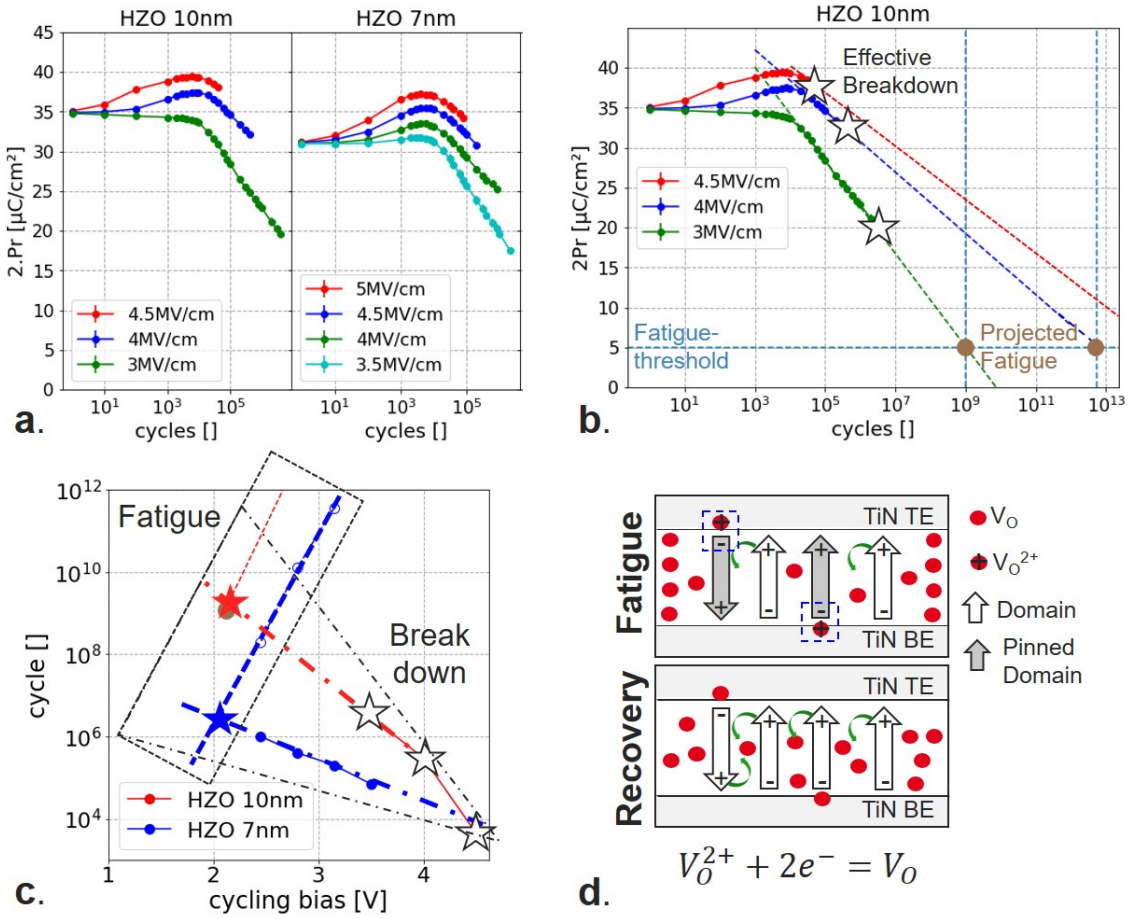


Figure 4.10: a. 2.P_R-Cycles measured with 4MV/cm 10kHz PUND along bipolar field cycling with t_{rise} = t_{fall} = t_{pulse} = 1μs with various electric fields. 300mm wafers with TiN/HZO (10nm)/TiN and TiN/HZO (7nm)/TiN stacks were studied, and for each condition, 10 100µm diameter devices were measured. ; b. Representative extraction of cycles-to-breakdown and projection of cycles-to-fatigue. ; c. Maximum number of cycles before dielectric breakdown or ferroelectric fatigue (criterion defined at 5$\mu C/cm^2$) for each HZO thickness along applied voltage. ; d. Schematics of MFM ferroelectric stack illustrating discussed mechanisms for fatigue and recovery.

As previously mentioned, the endurance of FeCAP and FeRAM arrays can be limited by two

distinct failure modes well reported in literature: dielectric breakdown and ferroelectric fatigue. The first section will thus focus on these endurance-limiting failure mechanisms through discussions based on electrical evidence gathered experimentally on Short Loop devices presenting a 100µm diameter.

Figure 4.10.a. presents the evolution of the remanent polarization during bipolar square cycling with different electric field conditions until dielectric breakdown. The effect of each condition was investigated on 10 devices of TiN/HZO (10nm)/TiN (left) and TiN/HZO (7nm)/TiN (right) stacks, and PUND sequences at 4MV/cm - 10kHz PUND were applied. Initial wake-up operation consisted of 10^3 cycles at 4MV/cm - 100kHz, regardless of the following cycling condition. This operation resulted in a unique initial remanent polarization level, and ensured that the following $2.P_R$ evolution was purely a product of the differentiated cycling field adopted. As can be seen on wafers with both thicknesses, cycling at higher electric field provides a higher maximal $2.P_R$ (attributed to prolonged wake-up due to higher field conditions), but leads to lower cycles-to-breakdown. Consistently, lower cycling fields give rise to a lower maximal $2.P_R$, but an improved breakdown-limited endurance. At the same cycling field, it can be observed that breakdown occurs after a largest number of cycles for the 7nm stack than for the 10nm stack, indicating a higher breakdown field, as reported in [122, 90]. This phenomenon is thought to arise because of lower energy barrier in thinner films that minimizes defect generation by injected electrons [90]. This result is important as it confirms a strong advantage of lower thicknesses films in terms of breakdown resiliency, as evoked in Chapter 3.

On both stacks, a progressive wear-out of $2.P_R$ can be observed at lower fields before the occurrence of dielectric breakdown. This phase is the ferroelectric fatigue and can be damaging in terms of memory operation. Indeed, $2.P_R$ loss (or MW closing at FeRAM array level) will limit the read margin, and therefore constitutes another endurance-limiting failure mode, promoted this time by lower cycling fields.

The impact of both dielectric breakdown and ferroelectric fatigue on 10nm and 7nm HZO capacitors can be assessed be measuring the cycles-to-breakdown and effectively projecting the cycles-to-fatigue based on a $5\mu C/cm^2$ threshold, as shown in Figure 4.10.b. Thus determined cycles are reported as a function of cycling bias for both breakdown and fatigue on 10nm and 7nm thick HZO-based devices in in Figure 4.10.c. Both mechanisms have opposite electric field dependence, therefore feeding the need for carefully chosen cycling field conditions to prolong endurance. The maximum number of cycles, either breakdown- or fatigue-limited, are then found equal to 2×10^6 at 2V for 7nm-thick HZO and 2×10^9 at 2.2V for 10nm-thick HZO. These conditions (represented with the red/blue star symbol) correspond to an optimal cycling voltage enabling prolonged endurance in operational conditions (*i.e.* no breakdown and sufficient $2.P_R$). Interestingly, although offering lower bias operation capability and higher breakdown field, 7nm HZO is more prone to fatigue. Note that these endurance values are limited compared to that of equivalent integrated devices and arrays because ShL devices are more prone to dielectric breakdown due to their large area, as reported in [98, 127]. Also note that the early reports of projected endurance above 10^{15} cycles based on breakdown extrapolation at lower field were missing the critical contribution of fatigue at low field [71]. This emphasizes the need to bypass fatigue in order that breakdown is the only limitation to endurance.

The mechanisms behind fatigue-limited endurance have been widely discussed in the recent years. Figure 4.10.d. (top) represents domain pinning by trapped charges at TiN/HZO interfaces,

possibly positively charged oxygen vacancies (V_O^{2+}), and laterally inhomogeneously distributed neutral oxygen vacancies (V_O). Fatigue phenomenon has been proven to be recoverable through the punctual application of higher voltage pulses, thought to effectively detrap charges and/or redistribute oxygen vacancies [134, 135].

Recap

To make the most of higher cycle-to-breakown permitted by low field cycling and higher $2.P_R$ obtained through high field cycling, optimal operating fields could be determined for 7nm and 10nm HZO stacks. Due to recent demonstration of fatigue recoverability, another operating strategy could consist in cycling at low field until a low $2.P_R$ threshold is reached, and then cycling at higher voltage to increase $2.P_R$ again.

4.2.1.2 Endurance Exploration towards Fatigue and Potential Recovery

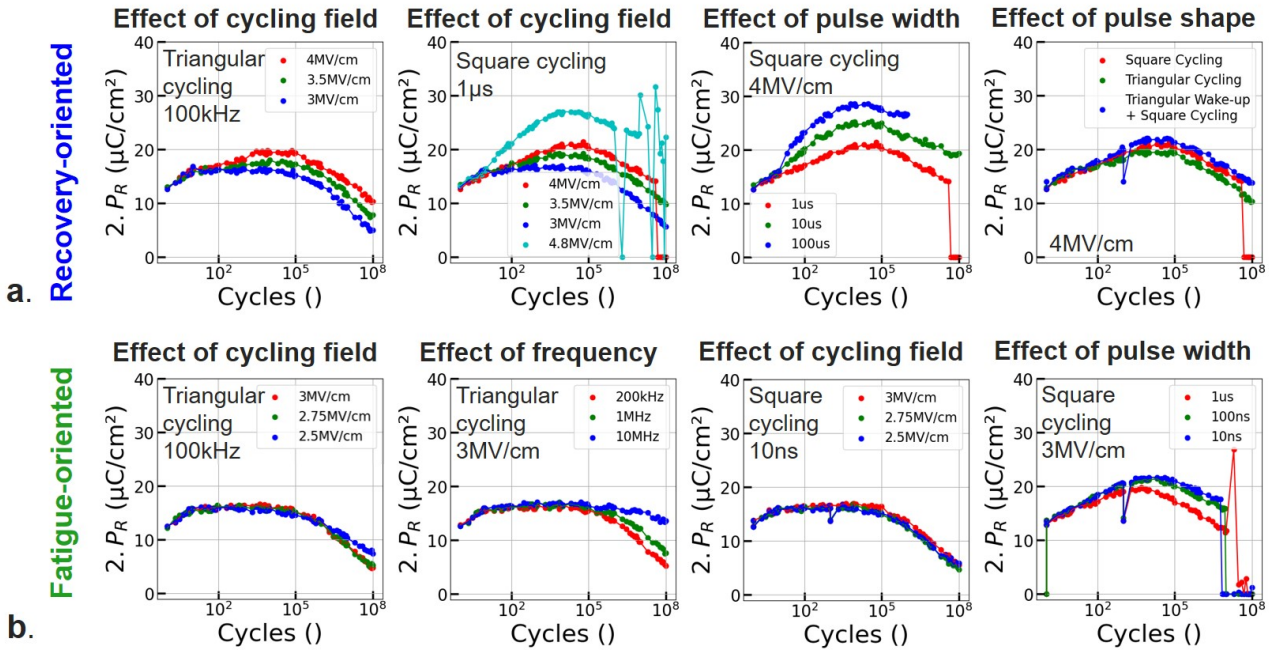


Figure 4.11: Comparison of 8 different cycling pulse sequences up to 10^8 cycles on 550nm-diameter nominally identical FeCAPs to explore a. potential recovery strategies and ; b. endurance sequences leading to fatigue.

As discussed in the previous section, ferroelectric fatigue leads to $2.P_R$ decrease during low cycling field operation, with the latter also having the downside effect of limiting the maximal $2.P_R$ achievable. From a recovery point-of-view, *i.e.* in order to increase $2.P_R$ again, a dedicated electrical

strategy at higher cycling field could induce a high performance gain while presenting a reasonably high cycles-to-breakdown, so as to avoid another failure mode. The ultimate goal of this study is to compare the efficiency of different recovery strategies applied to devices having reached a unique fatigued state. Therefore, an intermediate goal is the define a unique endurance sequence leading to fatigue in a scaled BEoL-integrated FeCAP, and several potential recovery strategies. Hence, this section aims at exploring different electrical sequences leading to ferroelectric fatigue on BEoL-integrated 550nm-diameter FeCAPs, and at determining potential recovery conditions. More specifically, the highly different endurance responses induced by variations in cycling pulses' shape, amplitude and width or frequency will be carefully studied.

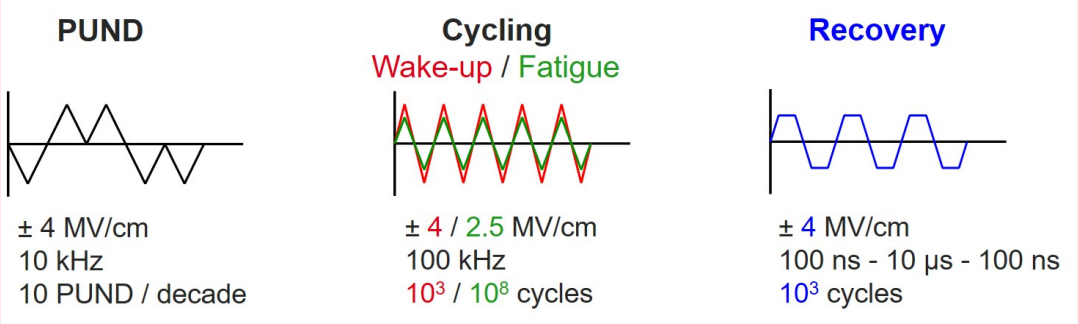
Figure 4.11 presents comparative $2.P_R$-Cycles (until 10^8 cycles) graphs measured with 4MV/cm - 10kHz PUND sequences on scaled FeCAPs from a TiN/HZO (7nm)/TiN 200mm wafer. Figure 4.11.a. targeted cycling conditions susceptible to constitute recovery strategies, while Figure 4.11.b. focused on inducing ferroelectric fatigue.

- From Figure 4.11.a., the clear advantage of 4MV/cm cycling field in terms of breakdown resiliency, and of square cycling with pulse widths of 10µs (rise and fall width at 1µs) in terms of $2.P_R$ gain can be observed. A standard recovery strategy can therefore be defined as 10^3 square cycles at 4MV/cm with 10µs pulses, and it is expected that this strategy could be repeated at least 10^5 times without causing dielectric breakdown.
- From Figure 4.11.b., the clear advantage of triangular cycling in terms of breakdown resiliency, and of 2.5MV/cm cycling field in terms of ferroelectric fatigue ($2.P_R$ loss above an arbitrarily defined $5\mu C/cm^2$ low threshold) can be observed. As such, an endurance sequence consisting in triangular cycling at 2.5MV/cm - 100kHz for 10^8 cycles can be adopted as the preferential electrical strategy to induce ferroelectric fatigue.

Note that other fatigue- and recovery-oriented sequences could have been adopted for this study, and that the choices were made to maximize chances of effective fatigue and recovery while minimizing the number of conditions under study.

Recap

To conduct a study on potential recovery effect of ferroelectric fatigue, an extensive endurance study lead to the choice of the following sequences:



Recap

Measurement scheme for recovery capability assessment is given below:



4.2.2 Recovery Capability

4.2.2.1 Efficiency of Different Recovery Strategies on FeCAP

The goal of this section is to evaluate the efficiency of the chosen recovery strategy on a BEoL-integrated FeCAP after fatigue. The endurance electrical sequence leading to fatigue and the subsequent recovery operation were repeated ten times in a row, which means that an overall 10^9 cycles were reached in the low field operation mode, and that a total of 10^4 cycles at higher operating field were applied. Judging from Figure 4.11.a., this number of cycles at 4MV/cm should not lead to dielectric breakdown. Figure 4.12.a. presents superimposed $2.P_R$-Cycles for all ten repetitions of the endurance sequence leading to fatigue at FeCAP level between which recovery sequences were interspersed. As can be seen for the first endurance iteration after fatigue (rep = 1), $2.P_R$ initially reaches 20 $\mu C/cm^2$ and then decreases until it reaches 8.5 $\mu C/cm^2$ after a cumulated 2×10^8 cycles at low field. Recovery enabled the improvement of initial $2.P_R$ compared to pre-recovery fatigue (rep = 0), probably thanks to the wake-up process of previously uncycled (*i.e.* pristine) ferroelectric domains [134]. However, similar $2.P_R$ are reached after fatigue for iterations rep = 0 and rep = 1: this interesting result is interpreted in [135] as a stable distribution of oxygen vacancies obtained after fatigue, regardless of the iteration. As can be seen from the color scale, all 10 iterations at high cycling field have led to the effective recovery of $2.P_R$, and to a unique stable state after $x \times 10^8$ cycles. The I-V characteristics at the end all fatigue iterations (10^8 cycles) are shown in Figure 4.12.c., and their very unusual shape will be studied in the next section. Although the efficiency of this standard recovery strategy decreases with incremental fatigue iteration compared to first fatigue iteration (rep = 0), the recovery ratio with regard to previous iteration (defined as $\frac{\Delta 2.P_R^{Recovery}(Cycle,rep=x)}{\Delta 2.P_R^{Fatigue}(rep=x-1)}$) is quite stable with number of iterations, as can be seen from Figure 4.12.b. All this points towards a very high number of achievable fatigue-recovery iterations, or "infinite endurance", as stated by [135]. In this case, the only limitation to FeRAM endurance comes from the breakdown contribution.

These very promising results can be broadened to a larger scale. Different recovery strategies were investigated, based on one-parameter variations to the standard recovery presented: pulse polarity, amplitude, frequency and number of cycles were varied. A comprehensive list of the studied strategies is given in Figure 4.12.d. and Figure 4.12.e. presents the efficiency of all recovery strategies as a function of applied recovery cycles. These different recovery strategies were applied to distinct nominally identical devices that experienced the same endurance sequence, thus leading to a unique fatigued state and no expected variability. The recovery metric used is defined as a ratio between $2.P_R$ gained by recovery over cycling at high field and $2.P_R$ lost by fatigue over 10^8 cycles at low field, and describes the recovery efficiency during cycling for the first iteration. It is obtained

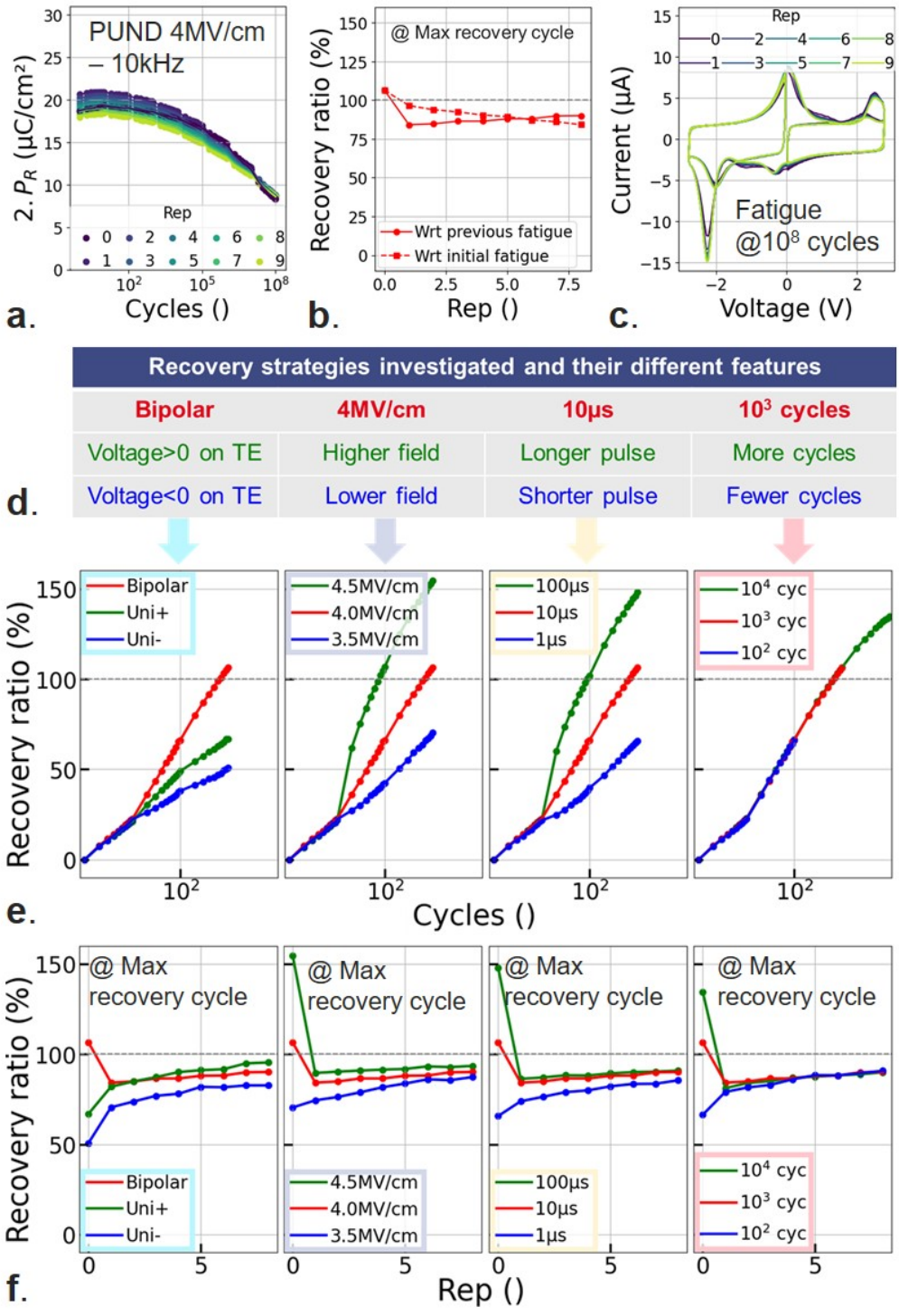


Figure 4.12: a. 2.P_R-Cycles of 550nm diameter FeCAP during fatigue for 10 repetitions (Rep) with interspersed recovery phases (Bipolar, 4MV/cm, 100ns-10µs-100ns, 10^3 cycles). ; b. Corresponding recovery ratio compared to previous or initial fatigue with regard to recovery repetition. ; c. Corresponding I-V after $x \times 10^8$ fatigue cycles for x repetitions. ; d. Table representing the strategies investigated on nominally identical FeCAPs after wake-up and fatigue phases to recover 2.P_R loss and E_c shift. ; e. Recovery ratio with regard to cycling for all recovery strategies investigated and presented in d. ; f. Corresponding Recovery ratio compared to previous fatigue with regard to recovery repetition.

through the expression $\frac{\Delta 2.P_R^{Recovery}(Cycle)}{\Delta 2.P_R^{Fatigue}}$. Overall, while all recovery strategies lead to a ratio higher than 50%, the bipolar recovery ratio is higher than that of either unipolar. It is worth noticing that higher fields, longer pulse widths or higher number of cycles all lead to the highest recovery ratios ; ratios above 100% indicate wake-up of previously uncycled domains [134]. Interestingly, an almost identical $2.P_R$ level is reached after fatigue for all devices, regardless of the recovery strategy chosen. In other words, upon re-fatigue, I-E identical to the first fatigued state is obtained, no matter the recovery strategy. As was observed for the standard recovery strategy, the recovery efficiency remains constant with the number of iterations (see Figure 4.12.f.), except for rep = 0 because of differing wake-up and recovery electrical sequences. These performances hint at a potentially much improved endurance when using repeated fatigue-recovery iterations.

Recap

The efficiency of the standard recovery strategy at high cycling field has been demonstrated after an endurance sequence of 10^8 low field cycles leading to ferroelectric fatigue. Up to 10 fatigue-recovery repetitions were demonstrated and led to a stable final state, thus pointing towards highly improved endurance thanks to a recovery scheme. The investigation of multiple recovery strategies demonstrated variable $2.P_R$ gain but a unique fatigued state and stable recovery ratio with the number of repetitions.

4.2.2.2 Towards Recovery Mechanisms Understanding

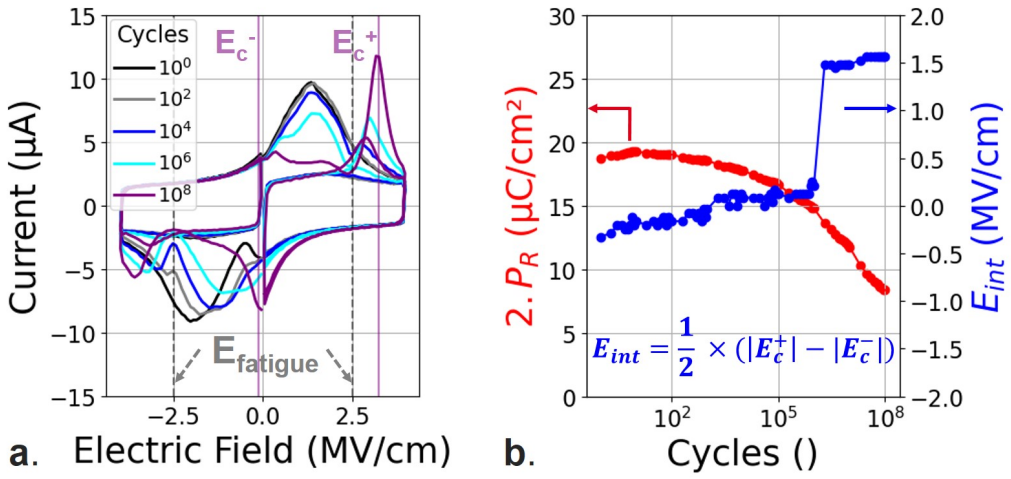


Figure 4.13: a. I-E characteristics of 550nm-diameter FeCAP measured by PUND at 4MV/cm - 10kHz during cycling at 2.5MV/cm - 100kHz. ; b. Remanent polarization and internal electric field along cycling of FeCAP during fatigue phase.

In order to shed light on recovery mechanisms with regard to phenomena presented in literature, a thorough peak analysis of I-E curves during fatigue and recovery (standard or otherwise) was conducted. Figure 4.13.a. presents post-wake up I-E characteristics measured at 4MV/cm during low field cycling at $E_{fatigue}$=±2.5MV/cm for the first fatigue iteration (rep = 0). During cycling, I-E exhibit an increasingly strong asymmetric peak splitting around $E_{fatigue}$, together with a shrink of main FE peaks, *i.e.* FE peaks within the switching region [137, 138]. Both effects result in overall remanent polarization loss and coercive fields shift with fatigue cycling, consistently with observations from [134, 136]. As can be seen in Figure 4.13.b., $2.P_R$ decreases from $19\mu C/cm^2$ to $8\mu C/cm^2$ during fatigue, as shown in Figure 4.12.a. for rep = 0. Meanwhile, internal field, defined as $E_{int} = 1/2 \times (|E_c^+| - |E_c^-|)$ increases towards peak symmetry with cycling. Discontinuity in E_{int} is caused by positive peak split outgrowing main peak with fatigue, thus abruptly shifting $E_c{}^+$ in the automated E_{int} extraction.

Dynamics of peak splitting are further studied through Figure 4.14 with the evaluation of switched polarization (P_{sw}) for each range of the field applied during fatigue (a.) and standard recovery (b.) cycling. During fatigue, main peaks of both polarities decrease with cycling, indicating loss of polarization in the cycling region. Below -$E_{fatigue}$, P_{sw} is stable with cycling, indicating that fatigue cycling causes no increase in the number of domains that switch to negative polarity state with coercive fields below -2.5MV/cm. Towards a high cycle number, P_{sw} decreases slightly: it is thought that these non-cycled domains are progressively pinned by charges trapped at the interface, and therefore start exhibiting a higher effective coercive field, as can be seen with the progressive shift of peak splitting below -$E_{fatigue}$ in Figure 4.13.a. On the other hand, P_{sw} above +$E_{fatigue}$ does increase with cycling, which could indicate domains being increasingly pinned and prevented from switching to positive polarity state with fatigue cycling field, thus shifting $E_c{}^+$ to higher values. During recovery cycling, a symmetric increasing trend of P_{sw} for main FE peaks (within ±2.5MV/cm switching region), is observed in Figure 4.14.b., while stable then decreasing P_{sw} below -$E_{fatigue}$ and decreasing P_{sw} above +$E_{fatigue}$ are shown. These P_{sw} trends during recovery are consistent with both domain depinning and oxygen vacancies redistributions mechanisms from literature [134, 135].

To further analyze the deformation of both main and split peaks, electrical results from PUND measurements are coupled with First-Order Reversal Curve, or FORC, technique. A FORC is measured by saturating a FeCAP at ±E_{sat}, decreasing or increasing the field to a reversal field E_r, and then sweeping the field back to E_{sat}. Figure 4.15 from [137] illustrates this under 'Applied Field Sweep' label (left) after repetitions of this process for different E_r values. The measured transient current and integrated polarization can be plotted as in Figure 4.15 (middle). To obtain the 'Experimental Preisach Density' (Figure 4.15 (right)) called ρ in Figure 4.16, the mixed second derivative of the polarization is calculated, and then plotted as a contour in a FORC diagram. This technique enables the study of switching and backswitching field distributions, and can give precious insights into peak splitting behavior. In the following, parameters for FORC measurement were fixed at 3V for V_{sat} and 0.06V for V_{step}.

Figure 4.16 presents PUND (top) and FORC (bottom) results post wake-up and fatigue (a. and d.), evidencing that peak splitting observed in Figure 4.13.a. comes from two distinct switching populations, labelled '1' and '2' in Figure 4.16 for more clarity. While standard bipolar recovery (Figure 4.16.b, in red) achieves almost complete re-centering and merging of both populations, and thus of FE peaks, unipolar recoveries recenter only one population each (Figure 4.16.b, in blue for uni- and c. in green for uni+).

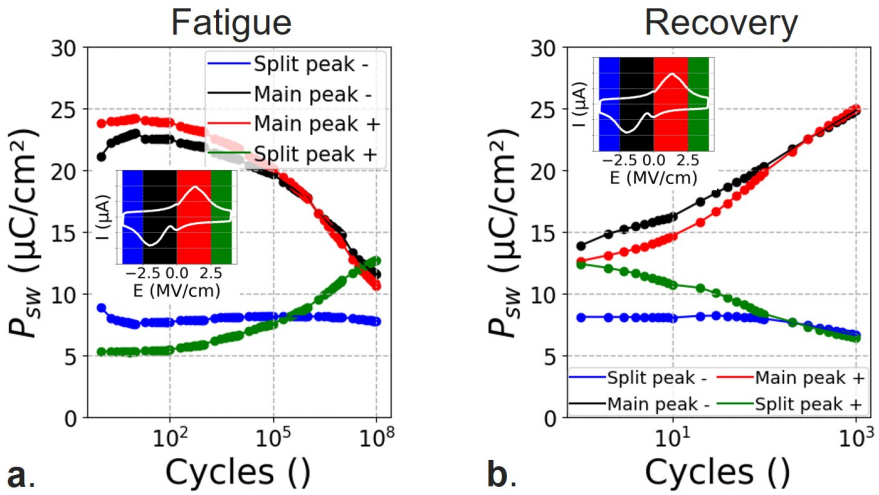


Figure 4.14: FeCAP switched polarization (P_{sw}) during fatigue (a.) and recovery (b.) cycling integrated from current measured in different field ranges shown in inset.

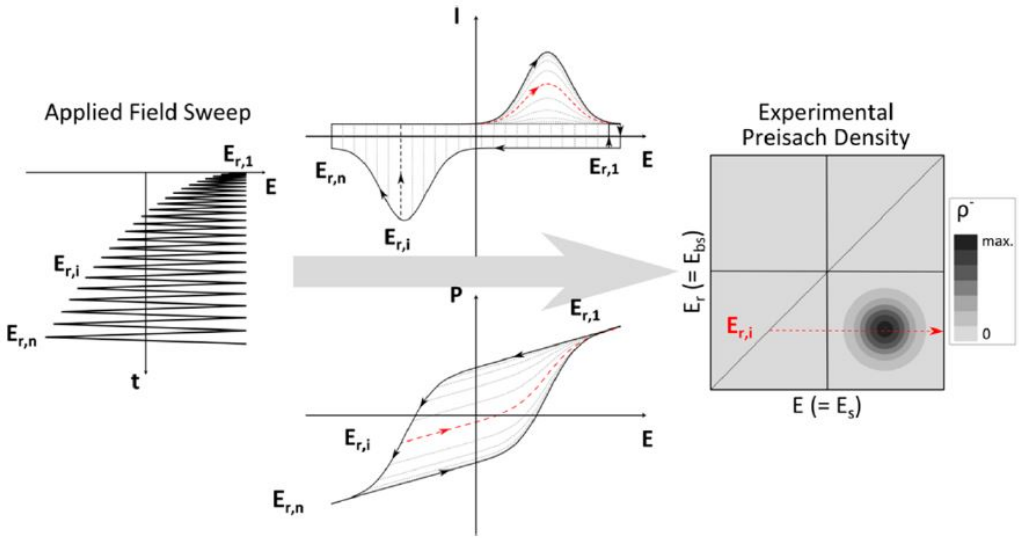


Figure 4.15: First-Order Reversal Curve from [137] presenting the applied sweeping field pattern from a positive saturation field (left) that enables to measure partial I-E and P-E (middle), which can be further derived and plotted as a switching density in a FORC diagram with E - E_r convention, where E in the electrical field applied during the sweep, and E_r is the reversal field.

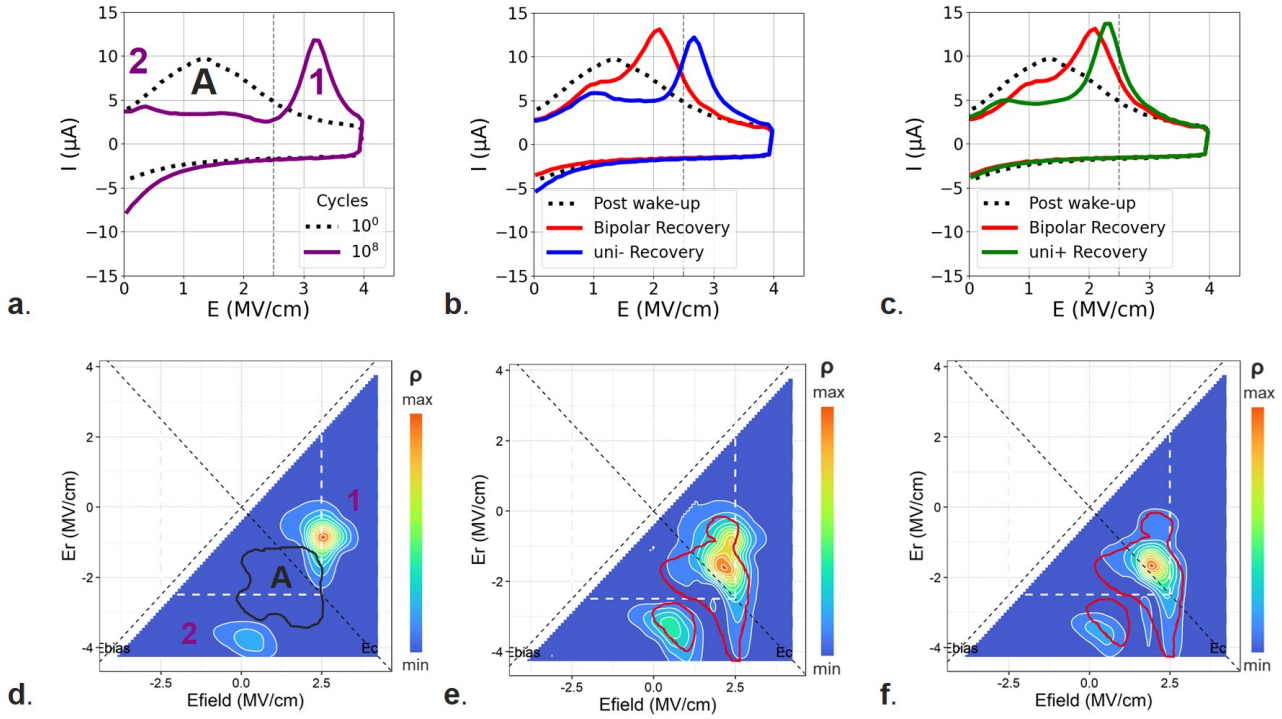


Figure 4.16: Current - Electric Field curves in positive polarity measured by PUND at 4MV/cm at the following points of the first fatigue-recovery iteration: a. After initial wake-up (10^0, in dotted black) and fatigue (10^8, in purple) ; b. After initial wake-up (dotted black), after 10^3 bipolar recovery cycles (red), after 10^3 negative unipolar recovery cycles applied to TE (blue) ; c. After initial wake-up (black), after 10^3 bipolar recovery cycles (red), after 10^3 positive unipolar recovery cycles applied to TE (green) ; First-Order Reversal Curves (FORC) of switching density populations with positive field at 4MV/cm in the E-E_r convention at the following points: d. Post initial fatigue (and contour for post wake-up population) ; e. Post uni- recovery (and contour for post bipolar recovery population) ; f. Post uni+ recovery (and contour for post bipolar recovery population).

Uni+ recovers the largest population (region labelled '1', pinned at $E_c > +E_{fatigue}$), possibly by neutralizing $V_O{}^{2+}$ at BE, making it the most efficient unipolar strategy. Uni- helps recovering the smaller region '2' (indicating recovery of negative peak splitting, at $E_c < +E_{fatigue}$), possibly neutralizing $V_O{}^{2+}$ at TE, and only part of region '1', making it less efficient, and highlighting the asymmetric $V_O{}^{2+}$ profile in the HZO layer. FORC measurements therefore shed a light on a previously unseen population fraction that cannot be recovered with the negative uni- recovery strategy. The optimal strategy is to use bipolar pulses, as it enables to detrap charges from both interfaces at the same time.

Recap

In-depth peak splitting study on low-field fatigued and high-field recovered BEoL-integrated FeCAPs showed that most efficient bipolar cycling recovery enables domain de-pinning by recombination of positive charges (possibly V_O^{2+}) at HZO interface with both TiN electrodes. PUND and FORC measurements revealed an asymmetry in top and bottom interfaces, probably linked to the integration process.

4.2.2.3 Multiple Recovery Demonstration at FeRAM Array Level

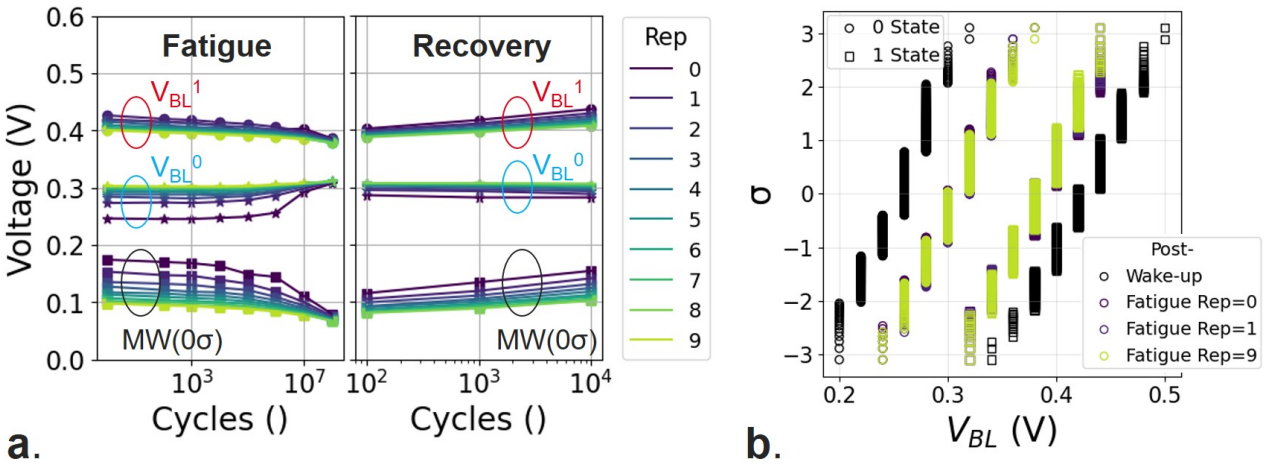


Figure 4.17: a. Bitline voltages measured on a 1kbit 1T-1C FeRAM array with 0.24µm² FeCAPs during 10 fatigue (left) - recovery (right) repetitions. ; b. Post wake-up and fatigue (10^8 cycles) distributions.

At 1T-1C 16kbit FeRAM array level, it is important to assess if previously established recovery strategies can be proven as effective. Indeed, in literature, only [124] have demonstrated that one iteration of fatigue - recovery - re-fatigue enabled a prolonged endurance without loss in read margin. Based on the knowledge gathered at FeCAP level, an equivalent measurement scheme was proposed to assess potential of recovery strategy at FeRAM array level. Similar fatigue conditions were applied (wake-up at 4MV/cm and cycling 2.5MV/cm), except for the signal shape as only square pulses are available at FeRAM array level. The efficiency of a recovery process combining the best conditions measured at FeCAP level (10^4 cycles at 4.5MV/cm) was assessed on a FeRAM array with 0.24µm² FeCAPs on the same HZO/TiN (7nm)/HZO wafer. Again, 10 repetitions of the fatigue-recovery sequence were applied. Figure 4.17.a. (left) shows '0' and '1' median Bit Line voltages (and the Memory Window) degradation during 10^8 fatigue cycles. The application of 10^4 cycles at 4.5MV/cm enable partial recovery, as can be seen from gain in dVBL for rep = 1. 10 recovery repetitions interspersed between fatigue cycles lead to an overall 10^9 endurance with no

MW degradation compared to endurance performance measured after 10^8 cycles. No additional bitfails are measured post recovery, as can be observed on Figure 4.17.b. showing '0' and '1' states distributions after wake-up, and after several iterations of fatigue. From this study, it can be inferred that recovery is a realistic option to improve endurance at FeRAM array level. Note that the recovery at FeRAM array level is however less efficient than at FeCAP level: this might be due to the the recovery sequence being identical to the wake-up sequence (and thus generating no new switchable domains), and could be overcome by applying a higher number of recovery cycles. In [124], as many as 10^{10} cycles were needed to fully recover the MW loss occurring during fatigue. Moreover, because only 10 repetitions were applied due to measurement time limitations and lead to no dielectric breakdown, the maximum endurance remains unknown: it would be an immediate follow-up axis to determine the real endurance gain achievable at FeRAM array level thanks to recovery.

Recap

To prolong endurance at FeRAM array level, recovery sequence using 10^4 cycles at 4.5MV/cm proved to be effective. In product-oriented FeRAM circuits at advanced nodes, it could be interesting to design specific features for punctual high field operation.

4.3 Radiation Hardness

In order to complete the investigation of ferroelectric memories reliability, the third part of this chapter aims at assessing the performance of FeCAPs and FeRAM arrays in radiation-challenging environments, particularly targeting their potential for integration in aerospace applications. Indeed, while memory technologies like Flash rely on charge-based storage and are therefore particularly vulnerable to ionizing radiations, PZT-based ferroelectric memories have shown a promising resilience to irradiation effects in the past [139]. To widen these studies to hafnium oxide based ferroelectric memories, FeCAPs and FeRAM arrays were exposed to heavy ion irradiation, considered one of the most severe forms of ionizing radiation due to its high energy, intense ionization, and difficulty to shield against. While devices can face reversible and non-destructive interactions with the ionizing radiation, they can also be impacted by breakdown, leading to the eventual destruction of the storage devices. In the following sections, the impact of Au^{26+} ions irradiation with a 1.635 GeV energy is carefully evaluated. The devices and arrays were irradiated at the X0-beamline of the UNILAC accelerator of the GSI Helmholtzzentrum fur Schwerionenforschung in Darmstadt.

4.3.1 Effect of Heavy Ion Irradiation on BEoL-Integrated FeCAPs

4.3.1.1 Electrical Characterization

In order to evaluate the impact of heavy ion irradiation on scaled BEoL-integrated FeCAPs, the focus was directed to a classic Metal - Ferroelectric - Metal stack integrated on 200mm wafers at the 130nm node. First, the TiN Bottom Electrode was deposited using PVD technique. Following

this, ALD enabled the deposition of a 10nm-thick hafnium oxide layer, which was subsequently silicon-doped by ion implantation, forming a Si:HfO_2 (or HSO) layer. TiN Top Electrode was then PVD-deposited and patterned to define 24 capacitors fields per die with critical dimensions ranging from 300nm to 600nm. At the end of the process, thermal annealing was performed at BEoL-compatible temperature to crystallize HSO in the orthorhombic ferroelectric phase.

Before irradiation, electrical characterization was performed on specific devices using the Positive Up Negative Down (PUND) technique at 4V - 10kHz, and in between measurement sequences, up to 10^6 triangular cycling pulses were applied at 4V - 100kHz. $2.P_R$ along cycling exhibited the expected behavior of wake-up, plateau and fatigue. Although the maximal $2.P_R$ reached $25\mu C/cm^2$, a wide dispersion of the ferroelectric properties measured on 70 dies across the wafer was observed. Indeed, after 10^4 cycles, $2.P_R$ spanned across a $16\mu C/cm^2$ range, and this variability is to take into account while analyzing post-irradiation behavior. Eight dies were diced from the original wafers and irradiated at distinct fluences ranging from 10^{10} to 7×10^{12} Au^{26+} ions/cm². On each die, six pristine devices with different areas were measured so as to provide a statistical analysis of the effect of heavy ion irradiation. Note that, due to a random permutation of samples, targeted comparison between devices before and after irradiation could unfortunately not be performed.

Figure 4.18 presents the results of electrical measurements on different FeCAPs irradiated at various fluences. Figure 4.18.a. and Figure 4.18.b. show the I-V and P-V curves of the $69\mu m^2$ area FeCAP irradiated at 8×10^{10} ions/cm² in the pristine state and after 10^5 cycles. Pristine measurement reveals antiferroelectric-like behavior with double current peaks in Figure 4.18.a. and a pinched PUND-corrected hysteresis loop in Figure 4.18.b. After 10^5 cycles, a single current peak, and hence a non-pinched hysteresis are observed, denoting classic ferroelectric behavior. Figure 4.18.c. presents $2.P_R$ evolution as a function of cycling measured on all FeCAPs irradiated at 8×10^{10} ions/cm², and confirms the expected wake-up - plateau - fatigue/breakdown behavior with a very high device-to-device homogeneity on a given diced sample, regardless of the size (450nm to 550nm) and shape (circular or square) of the devices. From this, it can be inferred that Au^{26+} irradiation at low fluence does not modify the classic ferroelectric behavior of the tested devices.

Beyond this low fluence, Figure 4.18.d. shows the evolution of $2.P_R$ measured at 10^5 cycles as a function of fluence for the $69\mu m^2$ area FeCAP. The main trend observed is an initial increase of $2.P_R$ upon irradiation, reaching a maximum at fluences of 8×10^{10} ions/cm², followed by a $2.P_R$ decrease at very high fluences. These results point to a beneficial effect of low fluences Au irradiation on ferroelectric properties. Au irradiation has been shown to generate defects in the stack, including oxygen vacancies [140]: at low fluences ($< 10^{11}$ ions/cm²), the achieved content of oxygen vacancies generated could be favorable to reach high $2.P_R$, as was the case in Chapter 3 with oxygen vacancies engineered by Ti scavenging layer. This boost in $2.P_R$ is more significant after the wake-up process, but can still be detected in the pristine case.

At intermediary and high fluences ($> 5 \times 10^{11}$ ions/cm²), a detrimental effect on $2.P_R$ is observed. It is worth noting that $2.P_R$ stabilizes just below $5\mu C/cm^2$ for the three highest fluences (3×10^{12}, 5×10^{12}, 7×10^{12} ions/cm²), which is accompanied by an improved endurance, as is shown in Figure 4.18.e. Indeed, while all lower fluences lead to breakdown before 10^7 cycles (the maximal cycle value investigated), no breakdown is observed for fluences above 10^{12} ions/cm². This gain in endurance could however have a limited interest as $2.P_R$ stabilizes at a very low value.

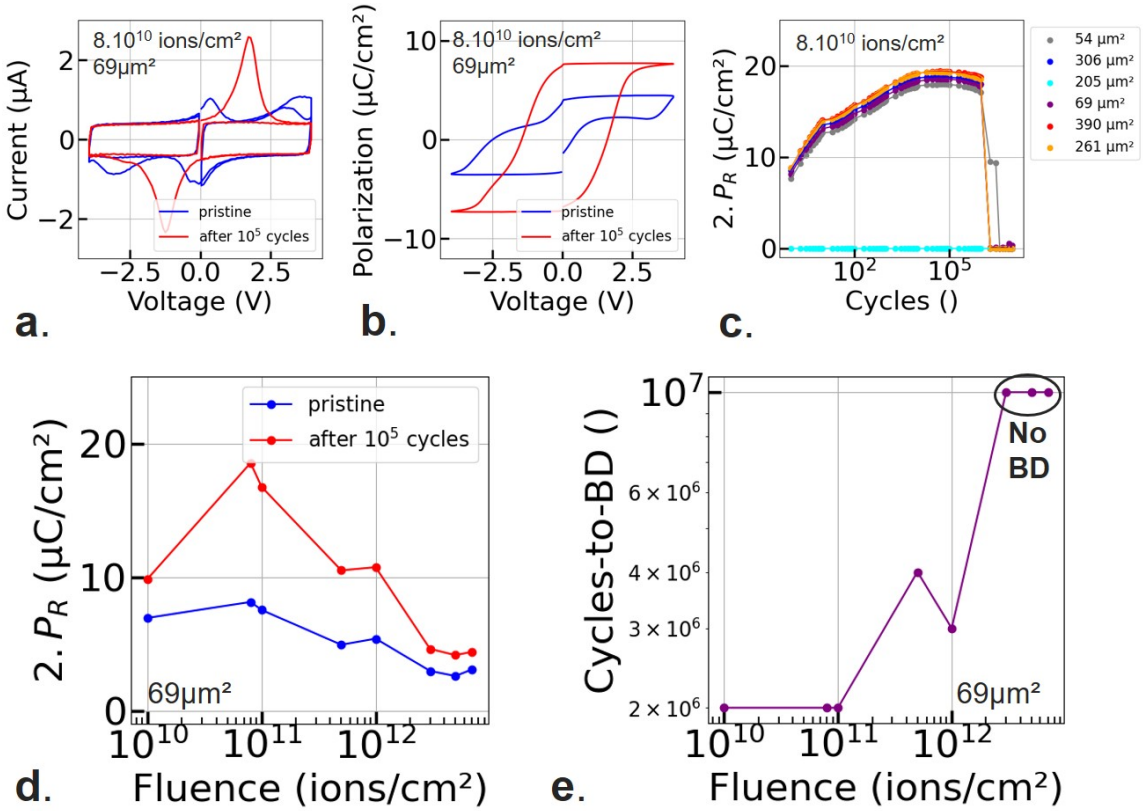


Figure 4.18: Pristine and woken-up I-V (a.) and P-V (b.) measured at 4MV/cm - 10kHz on $69\mu m^2$ area HSO-based FeCAP after irradiation at 8×10^{10} ions/cm². ; c. $2.P_R$-Cycles of devices featured on diced die irradiated at 8×10^{10} ions/cm². Triangular cycling was performed at 4MV/cm - 100kHz. ; d. Pristine and woken-up $2.P_R$-Fluence measured on $69\mu m^2$ area HSO-based FeCAP. ; e. Cycles-to-breakdown (BD)-Fluence measured on $69\mu m^2$ area HSO-based FeCAP.

4.3.1.2 Morphological Characterization

To determine if the variation of ferroelectric properties following irradiation is phase-related, XRD measurements were performed on TiN/Si:HfO_2 (10nm) /TiN full sheet stacks, irradiated during the same campaign. Figure 4.19.a. presents the spectra obtained from in-plane XRD measurements on 10-nm thick HSO full sheet samples irradiated at 5×10^9, 5×10^{10}, 5×10^{11} and 7×10^{12} or 5×10^{12} ions/cm², respectively. The orthorhombic (o) / tetragonal (t) phases are characterized by an o/t peak at 30.5°. It is often difficult to deconvoluate the contribution of both phases in this peak, especially if it is highly symmetrical. In the spectra, it can be seen that intensity of the o/t peak increases when fluence is increased from 5×10^9 to 5×10^{10} ions/cm², and then decreases at higher fluences to an almost null peak at 7×10^{12} ions/cm². This trend is consistent with what was observed electrically on scaled integrated FeCAPs, confirming that the boost of ferroelectric properties at lower fluences is related to an increased o/t peak in full sheet samples. For a more quantitative analysis, the o/t peak was fitted using a 5 peak model on the whole measured range, as shown

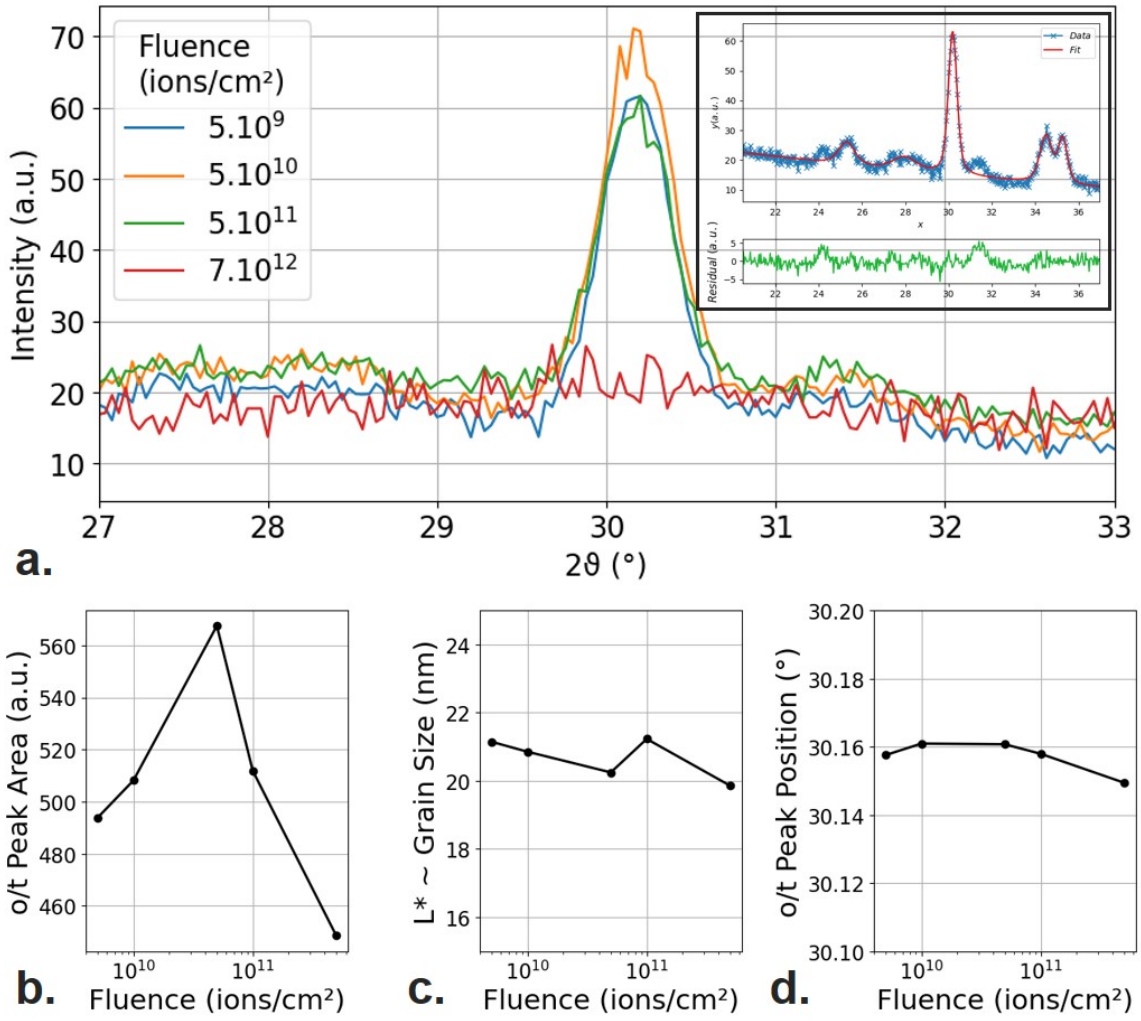


Figure 4.19: a. In-plane XRD spectra measured on TiN/HSO (10nm) /TiN full sheet sample and centered on the 30.5° reflection attributed to the orthorhombic / tetragonal (o/t) phases. Inset shows result of fitting method using a 5 peak model, and fitted o/t peak area (b.), corresponding grain size (c.) and (d.) o/t peak position are presented below as a function of fluence.

in inset. In this fit, the o/t peak is considered as a single peak and no attempt at deconvolution was made. The impact of irradiation fluence on three parameters is further analyzed: peak area (Figure 4.19.b.), L* (equivalent to grain size, Figure 4.19.c.), and peak position (Figure 4.19.d.). Peak area trend with fluence confirms indications from spectra peak maxima that low fluences boost crystallinity in o/t phases, and that high fluences lead to a reduced crystallinity. Figure 4.19.c. indicates that grain size remains constant with irradiation, therefore grain fragmentation effect can be excluded. Finally, as can be seen in Figure 4.19.d., o/t peak does not shift in one direction or the other with irradiation, indicating no obvious t-o or o-t crystalline-crystalline phase transition. For these reasons, it can be concluded that crystallinity degradation at high fluences occurs from

amorphization of HSO. It can be proposed that the amount of defects created permanently damaged the crystalline organization. Reduction in the number of orthorhombic grains explains the $2.P_R$ degradation observed electrically.

Recap

From electrical and morphological measurements, it can be inferred that HSO-based FeCAPs are resilient to heavy ions irradiation. Moreover, two major trends can be observed: boost of $2.P_R$, or crystallinity, at low Au fluences, and significant decrease in $2.P_R$, or crystallinity, at higher fluences. Generation of an optimal content of oxygen vacancies at low fluences, and material amorphization at higher fluences, respectively, could explain these trends.

4.3.2 Effect of Heavy Ion Irradiation on BEoL-Integrated FeRAM arrays

To obtain a good estimation of ferroelectric memories' resilience in harsh environments, FeRAM arrays were similarly exposed to heavy ion irradiation. Before irradiation, a 1T-1C 16kbit FeRAM array with 10nm-thick HSO and individual square FeCAPs of 0.36µm² area was woken-up with 5×10^3 cycles at 4.8V - 2µs. A scan of the internal reference voltage (V_{ref}) was then performed to determine σ versus V_{ref} distributions of '0' and '1' logic states, as presented in black and red in Figure 4.20.b. A checkerboard pattern was programmed in the 16kbit, and then Au irradiation was performed at a fluence between 5×10^9 and 5×10^{10} ions/cm². These limited fluences were used as the transistors (from the bitcell, as well as the periphery) are particularly vulnerable to ionizing irradiation due to their charge-based operation [141]. From the initial distributions, a reference voltage was chosen at 0.8V for digital reading operation to assess pattern integrity after irradiation. The reading operation on the array revealed a degraded pattern with entire word lines (WLs) failing (479 bits), as well as individual bit (1) failing, as shown in Figure 4.20.a. A subsequent scan of V_{ref} (grey and orange in Figure 4.20.b.) indicates that irradiation no only degraded the stored information, but also impacted FeRAM reliability, with bitfails in the distributions that close the memory window fully. As can be seen in Figure 4.20.c., extrinsic bitfails from '0' and '1' states distributions correspond to the WLs failing in Figure 4.20.a. This indicates a probable failure of a limited number of transistors in the peripheral CMOS circuitry. The fails in WLs and individual fail could therefore be attributed to periphery 1T and bitcell 1C degradation, respectively. To try to strengthen the first hypothesis, I_d-V_g characteristics were measured on one transistor on four different dies which were exposed to Au irradiation fluences from 0 to 10^{10} ions/cm². I_{drain}-V_{gate} behavior is shown in Figure 4.20.d. and Figure 4.20.e.: although no significant diversion from specifications were measured, a higher statistic would be needed to completely confirm (or exclude) the role of transistors' vulnerability to heavy ion irradiation in observed bitfails at FeRAM array level. Slight shift noticeable in post-irradiation distributions (Figure 4.20.b.) is attributed to imprint occurring at room temperature. Strong hints supporting the proposition are given by checkerboard pattern still discernable on '0' and '1' states bitmaps (Figure 4.20.c.), as was seen in the first part of this chapter after long-term retention at room temperature.

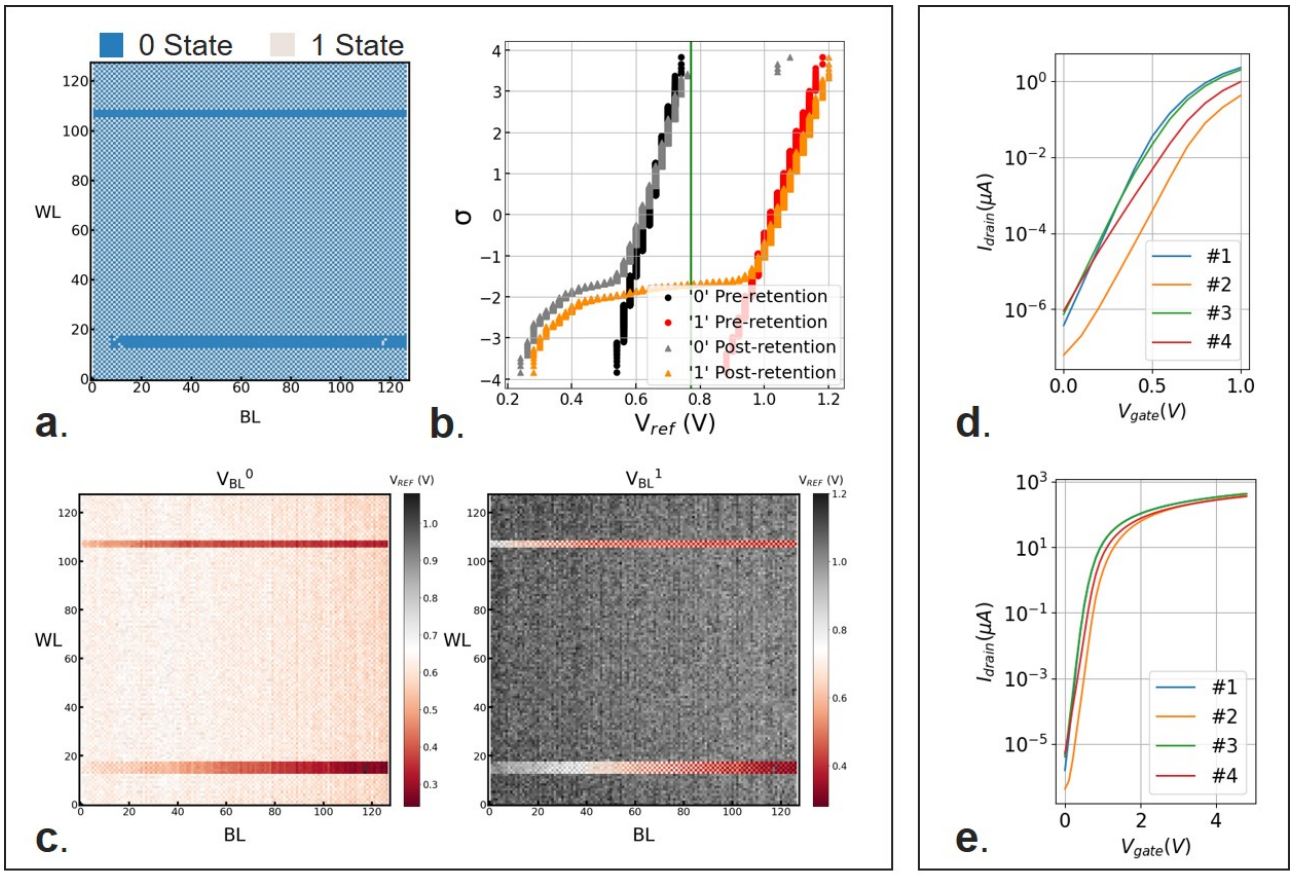


Figure 4.20: a. Checkerboard pattern read at 4.8V - 2µs with V_{ref} = 0.8V on an HSO-based 16kbit 1T-1C FeRAM array after Au irradiation. ; b. σ-V_{ref} distributions of '0' and '1' states measured pre- and post-irradiation. ; c. $V_{BL}{}^0$ (left) and $V_{BL}{}^1$ (right) bitmaps resulting from post-irradiation V_{ref} scan. ; I_{drain}-V_{gate} of 130nm thick-oxide (GO2) transistors measured on 4 dies after irradiation at different fluences in low- (d.) or high- (e.) V_{gate} range.

Conclusions

This chapter aimed at assessing the reliability performances of ferroelectric memories, and was particularly interested in bridging the gap between BEoL-integrated FeCAPs and application-oriented FeRAM arrays. To this end, three major reliability issues were evaluated, and solutions for improvement were proposed.

While evaluating data retention, the limited impact of thermal depolarization was underlined, due to the high stability of ferroelectric performances in the considered 85°C - 125°C temperature range. A key issue for retention was determined to be imprint, in line with previous work from the literature on PZT- and HfO_2-based FeRAM arrays. Indeed, while the effect of imprint at room temperature has been shown to be negligible and led to no bitfails in a FeRAM array after 348 days storage, it quickly degraded the Opposite State (OS) at elevated temperatures, thus closing the Memory Window. Equivalent measurement scheme and modelling of OS reduction in $2.P_R$ (or MW at FeRAM array level) evidenced identical imprint dynamics at FeCAP and FeRAM array levels, as well as additional array-related voltage drop resulting in apparent worsened retention in FeRAM arrays. The history effect of imprint, which leads bits programmed in one state to partially return to this state after re-programming, has been demonstrated to be recoverable through the application of a carefully-chosen number of cycles at operating field.

Recovery from fatigue has also been the subject of the second part of this chapter, which aimed at demonstrating effective recovery from ferroelectric fatigue at FeCAP and FeRAM array level. After assessing breakdown- and fatigue-limited endurance, and their opposite operating voltage dependence, extensive cycling sequence evaluation was conducted in order to determine possible recovery strategies. Ten repetitions of fatigue (at low cycling field) - standard recovery (at high cycling field) scheme demonstrated successful recovery of $2.P_R$ thus effectively improving endurance. It also revealed that a post-fatigue stable state could be achieved after each repetition, with no indication of increased leakage. The subsequent investigation of various recovery strategies enabled to determine that high fields, long pulses, and high number of cycles yield highest recovery ratios. Bipolar cycling has been demonstrated to be more effective than either unipolar scheme, which is determined to stem from domains de-pinning by recombination of positive charges trapped at HZO interface with both electrodes. At FeRAM array level, MW can also be successfully recovered, with no additional bitfail after 10 repetitions.

To assess FeCAP and FeRAM performance in radiation-challenging environments, heavy ion (Au^{26+}) irradiation was performed. On pristine and cycled FeCAPs, a beneficial effect of low fluences was evidenced, while at higher fluences, progressive amorphization evidenced by X-ray diffraction lead to degradation of the ferroelectric properties. In a FeRAM array programmed before irradiation in a checkerboard pattern, read bitfails of entire Word Lines (WLs) were demonstrated to be permanent and linked to failure of periphery transistors. Further attempt at linking WLs failure to access devices failure should be considered.

5 Scaling of HfO_2-based FeCAPs to the 28nm Node: Process Integration and Performance Assessment

Introduction

In the previous chapter, it was demonstrated that the reliability of HfO_2-based 1T-1C FeRAM arrays yielded promising performances at the 130nm node. Indeed, following the assessment of retention, endurance and immunity to radiations of FeRAM arrays, the following conclusions could be drawn: in the case of retention and endurance, different recovery strategies were demonstrated to correct the shift of the distributions caused by imprint, and the MW reduction due to ferroelectric fatigue, respectively. In FeRAM arrays, transistors were shown to be more vulnerable to heavy ion irradiation than ferroelectric capacitors. Together with results from Chapter 2 and Chapter 3, this confirms that HfO_2-based FeRAM technology is a suitable and promising candidate for scaling towards advanced nodes.

This chapter gathers initial results for FeCAP integration in the Back-End-of-Line (BEoL) of transistors at the 28nm technology node. Similarly to previously presented demonstrators at the 130nm node, the Front-End of Line (FEoL) and Middle-of-Line (MoL) process was fabricated by STMicroelectronics foundry, and the ferroelectric capacitors were integrated in the BEoL by CEA-Leti. The first section of this chapter describes the integration process implemented to account for the adoption of scaled FEoL transistors at advanced nodes. Different steps (whether observations at the microscopic scale or morphological characterization) during the process constitute a first quality check of the integration. The key method to assess the presence of ferroelectric properties in the HfO_2-based layer being the electrical characterization, scaled elementary capacitors are measured to validate the integration of 1C devices on 300mm wafers. The functionality and sanity of 28nm FEoL transistors is verified, and integrity of electrical performance of 1T-1C bitcell is demonstrated. Furthermore, the analysis of capacitor fields structures on advanced demonstrators gives insights into material and integration options for improved performances on 300mm wafers, in view of integrating FeRAM arrays at advanced technology nodes.

Summary

5.1 Process . . . 143
5.1.1 Integration Steps . . . 143
5.1.2 Morphological Characterization . . . 146
5.2 Electrical Characterization . . . 148
5.2.1 Single Devices . . . 148
5.2.1.1 1C Capacitors . . . 148
5.2.1.2 1T Transistors . . . 152
5.2.1.3 1T-1C Bitcells . . . 154
5.2.2 Capacitor Fields . . . 155

5.1 Process

As has been reported in the Introduction Chapter (Chapter 1), the development of PZT-based FeRAM technology has been limited to the 130nm technology node due to poor CMOS-compatibility and scalability [12]. On the other hand, hafnium oxide is CMOS-compatible and exhibits ferroelectric properties in thin films crystallized in the polar orthorhombic phase [57]. Chapter 2 has provided a computational study aiming at evaluating the constraints on 1C and 1T dimensioning to ensure a readable memory window in FeRAM arrays at the 28nm node. Chapter 3 has reported a few integration demonstrations at the 130nm node to prepare for optimized operation at low voltage in large and dense FeRAM arrays at advanced technology nodes, and Chapter 4 has confirmed the very promising reliability performance of FeCAPs and FeRAM arrays at the 130nm node. Therefore, the attempt at integrating FeRAM arrays at advanced nodes (28FDSOI for example) is the focus of this chapter. In this section, the process to integrate ferroelectric capacitors in the BEoL of transistors at the 28nm node is presented extensively. A description of the FEoL is given in order to understand the advantages brought by the FDSOI technology at advanced nodes, and electron microscopy and X-ray diffraction measurements enable the assessment of process quality, before a definite ferroelectric performance evaluation can be provided by the electrical characterization step.

5.1.1 Integration Steps

As transistors dimensions scale laterally with advancing planar technology nodes, undesired effects can occur such as leakage current from the drain of the transistor to the doped silicon substrate outside of the conductive channel region, as represented in Figure 5.1.a. from STMicroelectronics. The Fully-Depleted Semiconductor on Insulator (FDSOI) technology is a planar process technology that aims at reducing a transistor dimensions while maintaining either high energy-efficiency (*e.g.* through very low leakage current) or high speed. It was developed at CEA-Leti [89] and consists in a thin film of silicon in the channel region, and an isolating silicon oxide thin film (also called buried oxide layer) directly underneath, as shown in Figure 5.1.b. The electron movement is therefore confined in the undoped (fully-depleted) channel region, thus reducing leakage currents and improving the energy-efficiency. Both layers are deposited on an silicon substrate that can be biased to increase the switching speed of the FDSOI transistor, as is described in Figure 5.1.c.

In the process considered, 28nm node FDSOI transistors are manufactured on 300mm diameter silicon wafers by STMicroelectronics, and constitute the FEoL. Two different transistors flavors with varying dimensions are fabricated: GO1 transistors which have a channel length of 30nm and a gate oxide thickness of 13Å, limiting the nominal operation voltage to 0.9V ; and GO2 transistors which have a channel length of 150nm and a gate oxide thickness of 35Å, limiting the nominal operation voltage to 1.8V. For both transistors, a 10% overdrive voltage of 1.1V and 1.98V, respectively, is allowed. The wafers received from STMicroelectronics are manufactured up to Metal 5 (M5), and the hafnium-based ferroelectric capacitors are integrated at CEA-Leti between Metal 5 and Metal 6 lines. Due to the transistors scaled dimensions at the 28nm node, their gate-source capacitance (C_{GS}) is decreased compared to that of their counterparts at the 130nm node, as evoked in Chapter 2. This causes a reduction of the Bit Line capacitance (C_{BL}) at the 28nm node versus 130nm node, therefore it is expected that the '0' and '1' logic states of a 1T-1C 16kbit FeRAM can be discriminated down to 200nm diameter for a 2D planar FeCAP, as explained in Chapter 2 and reported in [86]. For

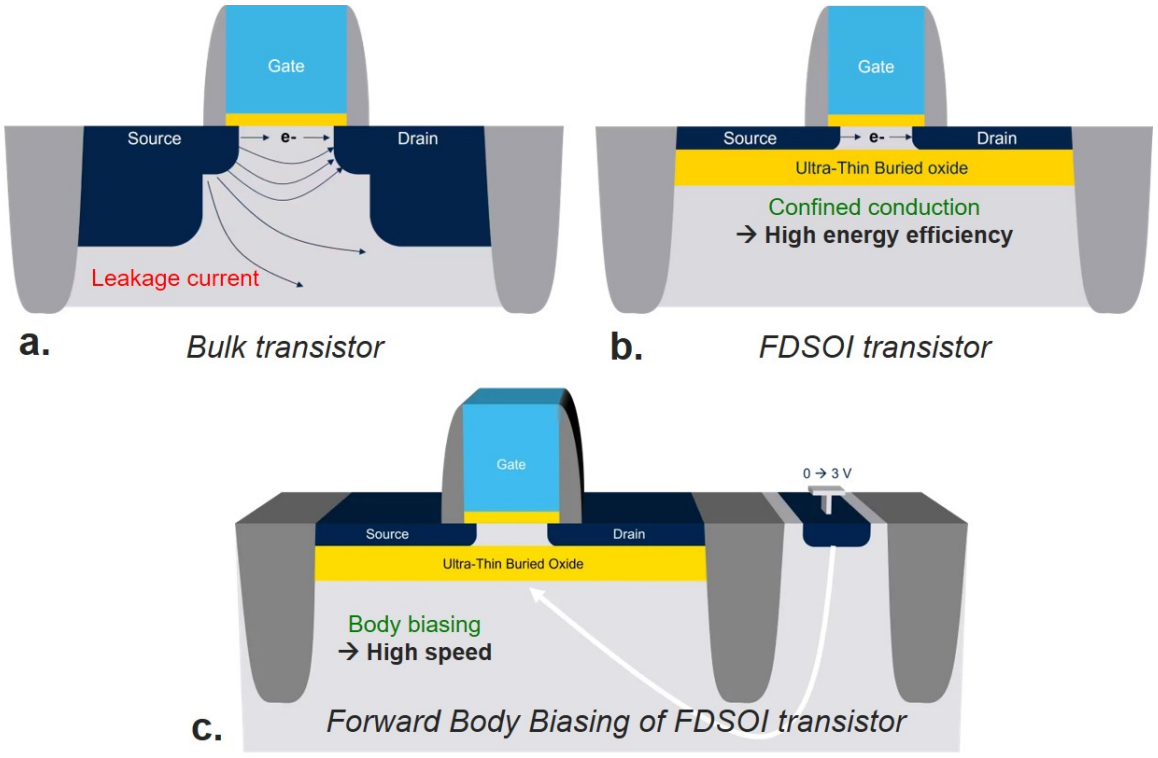


Figure 5.1: Adapted from STMicroelectronics commercial content: a. Bulk transistor with a doped silicon substrate prone to leakage current at scaled dimensions ; b. Fully-Depleted Semiconductor on Insulator (FDSOI) transistor consisting of an undoped silicon substrate, an ultra-thin silicon oxide layer and an undoped silicon thin film confining electrons in the channel ; c. Forward Body Biasing (FBB) technique in an FDSOI transistor to enable high speed switching

this reason, elementary capacitors (1C) were etched with a critical dimension of 200mm in this first demonstration. A schematic of the integration flow is presented in Figure 5.2. A lower via (Via-INF) is first integrated upon STMicroelectronics M5 by lithography and patterning, and filled with Ti-TiN on the sidewalls and W in the cavity. Following a Chemical-Mechanical Polishing (CMP) step, the ferroelectric stack (PTMEM level) is deposited. In the Process of Reference (PoR) on 300mm wafers, determined from an initial integration process at the 130nm node (on 200mm wafers) and additional material insights from 300mm Short Loop wafers, TiN top and bottom electrodes are PVD-deposited, a CMP step is performed after BE deposition, and the ferroelectric layer is a 10nm Si:HfO_2 layer deposited by ALD and doped by ion implantation. As ALD-deposited HZO shows improved ferroelectric performances with decreased wafer-level variability and decreased crystallization temperature (Chapter 3, [90]) compared to its HSO counterpart, 300mm wafers with an HZO-based ferroelectric stack will also be fabricated. In the PoR integration flow, a thermal annealing step at 500°C is applied to crystallize the hafnium-based layer. Note that variations in the temperature and position of the annealing step (post-TE deposition or at the end of the process) will also be evaluated. Then, the entire TiN/FE/TiN stack is etched after lithography using a Deep

UV (DUV) immersion scanner. Following etching of the stack, upper (Via-SUP) and then contact (Via-CONTACT) vias are defined by lithography and patterning steps, and filled with TaN-Ta on the sidewalls and Cu in the cavity. M6 is then deposited on top of the vias, followed by a final passivation and pillar lithography step to contact the top or bottom electrode of each individual FeCAP. In the following recap, the different integration steps and lithography levels are summarized.

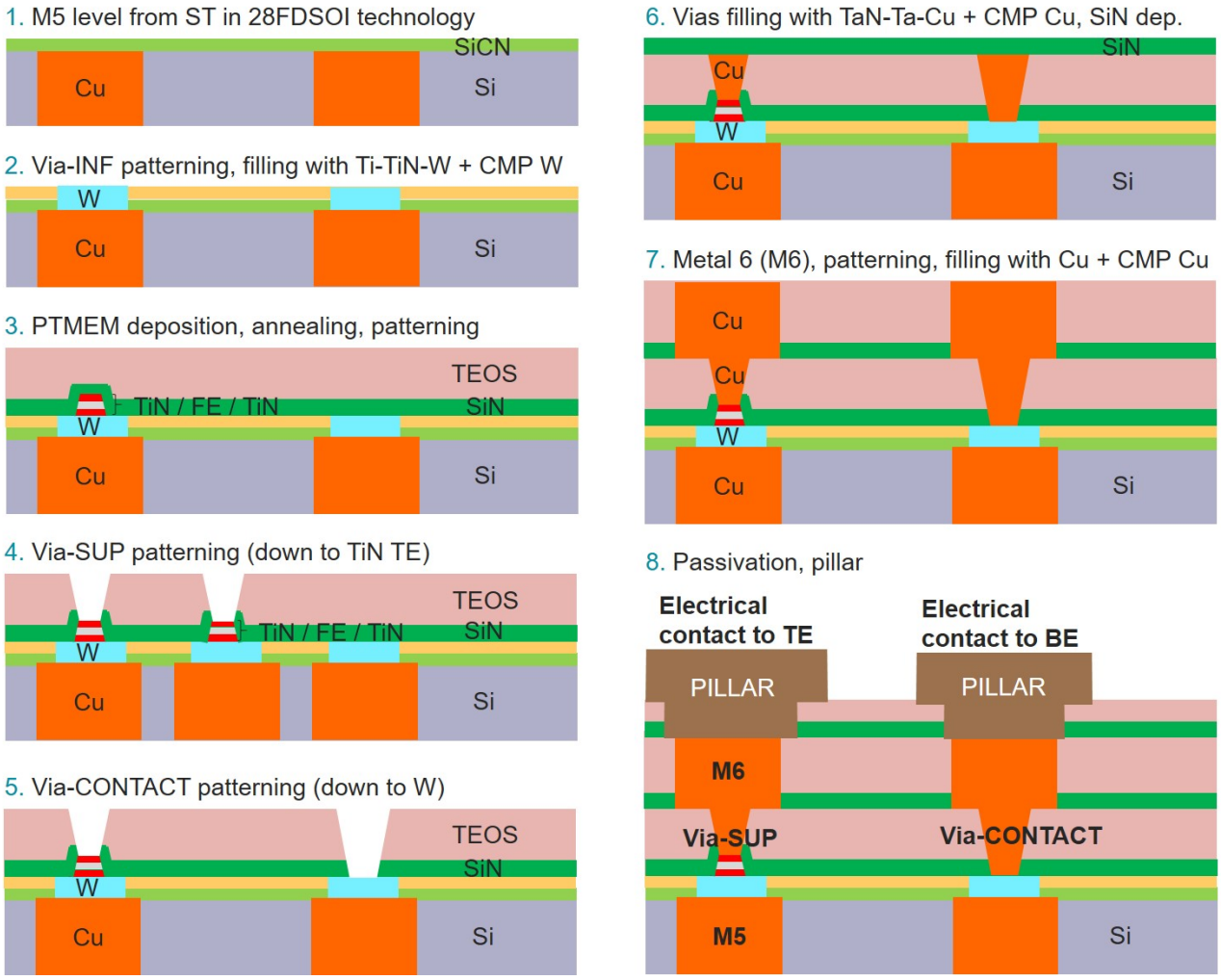
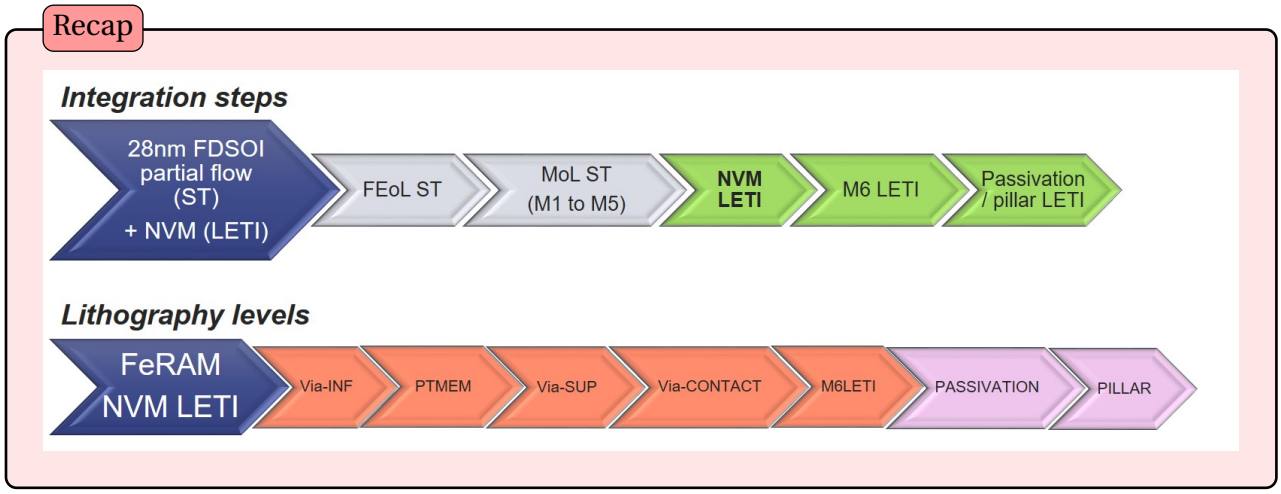


Figure 5.2: Process steps for the integration of a ferroelectric stack between M5 and M6 in the BEoL of FDSOI technology at the 28nm node from STMicroelectronics (28FDSOI).



5.1.2 Morphological Characterization

In order to assess the quality of integration of FeCAPs in the BEoL of 28FDSOI transistors, two distinct steps are adopted, and this part focuses on the obtained results. A first step is to visually assess the physical integrity of the complete stack, from transistors in the FEoL to capacitors in the BEoL. At the end of the process (*i.e.* after Metal 6 deposition and SiN encapsulation), a cross-section of the complete stack can be observed using a Scanning Transmission Electron Microscope (STEM). This cross-section was extracted from a 300mm diameter wafer using a Focused Ion Beam (FIB) technique to extract a lamella of the entire stack. As can be seen in Figure 5.3.a., the entire stack presents well-defined metal levels and interconnects and an a clear ferroelectric stack between lower and upper vias. In Figure 5.3.b., a zoom on the TiN/HSO/TiN stack evidences clean interfaces between HSO and TiN layers, indicating that the CMP step improves the quality of the interface. Indeed, while the CMP process is used only on the TiN BE and therefore efficiently decreases the TiN BE surface roughness, HSO is deposited with the ALD conformal technique, and will hence exhibit a surface roughness equivalent to that of BE. Moreover, the etching quality of the ferroelectric stack is high and no residual metal is apparent, therefore no electrical short between top and bottom electrode can be anticipated from this micrograph. To extend further the assessment of the integration quality, and more precisely evaluate the effect of crystallization anneal on the stack crystallinity, Grazing Incidence X-Ray Diffraction (GIXRD) measurements were performed during the manufacturing process at the 300mm wafer scale. Figure 5.3.c. presents the spectra measured after deposition of the top electrode (black) and after PoR crystallization anneal positioned post-TE deposition (red). The diffraction peak at 30.5° (2θ angle) suggests the presence of the orthorhombic and/or tetragonal phases, indicating that the annealing step efficiently crystallized the initially amorphous hafnium-based layer. In order to confirm the presence of the orthorhombic phase in the HfO_2-based layer and therefore evidence the presence of ferroelectric properties, electrical measurements were performed on scaled 1C devices integrated on 300mm diameter wafers in the BEoL of 28nm FDSOI transistors.

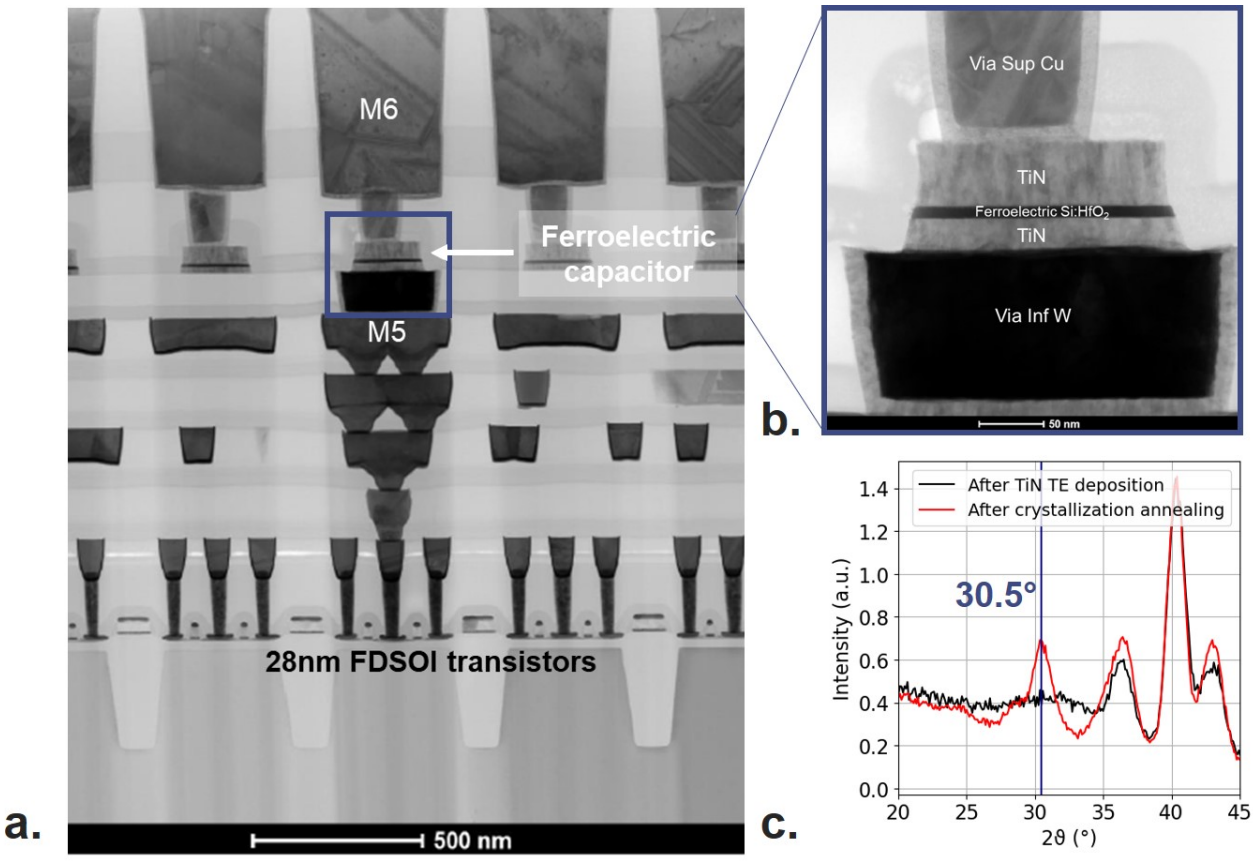


Figure 5.3: a. Bright-field STEM cross-section of HfO_2-based ferroelectric capacitors (1C) integrated in the BEoL of 28nm FDSOI transistors, between Metal 5 and Metal 6. The typical critical 1C dimension is 200nm. ; b. SEM cross-section zoom of TiN/Si:HfO_2 (HSO, 10nm)/TiN 1C between W lower via and Cu upper via. HSO layer and TiN bottom electrode are etched during the process of top electrode etching. ; c. Grazing Incidence XRD measurements during the 1C integration process: after deposition of the TE (black) and after thermal annealing (red).

Recap

Ferroelectric capacitors were integrated between M5 and M6 in the BEoL of GO1 and GO2 transistors at the 28nm node (28FDSOI from STMicroelectronics). Initial STEM observation on a cross-section of the entire stack indicates a high quality integration with no visible defects to the FEoL or the TiN/HSO/TiN stack. GIXRD measurements reveal the crystallization of HSO film in the tetragonal or orthorhombic phase after annealing. Electrical measurements should enable the determination of the transistors integrity, and confirm (or infirm) the ferroelectric properties of the 1C component.

5.2 Electrical Characterization

The first section has proposed a general integration path for ferroelectric capacitors in the BEoL of 300mm diameter wafers, with scaled transistors at the 28nm node constituting the FEoL. During the integration process, several observation steps are taken to ensure that no issue arises and to maximize the chances of electrical functionality. TEM and GIXRD results presented in Paragraph 5.1.2 are encouraging and do not reveal any blocking point against functionality at the end of the process. A preliminary PoR for FeCAP fabrication has been determined based on the fabrication process at the 130nm node, as well as additional input from 300mm Short Loop wafers. In the following parts, a variety of process variants are under investigation in order to determine the best possible integration process for optimized FeRAM performance at advanced nodes. Electrical characterization is performed on various devices featured on different 300mm demonstrators, to verify sanity, validate integration path, and determine optimal integration options. This part is voluntarily chronological as it describes the incremental procedure implemented to develop a 28nm node FeRAM technology. The 300mm demonstrators were designed and fabricated one after the other, often during the course of several months, and for cost issues, BEoL FeCAP fabrication was sometimes performed on base wafers at M5 without any transistors in the FEoL.

5.2.1 Single Devices

In this part, the term 'single device' is adopted to describe elementary transistors (1T), scaled ferroelectric capacitors (1C) and the combined 1T-1C bitcell in which the drain of a 1T is connected to the bottom electrode of a 1C. All these single devices are present in a first demonstrator, with possible variations of their dimensions, and they are the only devices available (no capacitor fields, no FeRAM arrays). This first demonstrator is replicated multiple times (different wafers in a unique lot) to assess the impact of different integration options and ferroelectric stacks on the overall ferroelectric performance. Two main kinds of wafers are investigated in this part: the first kind features scaled 1C devices integrated in the BEoL in 28nm FDSOI process ; the second kind features only scaled 1C devices (identical to 1C in the first kind) but without FEoL transistors fabricated on the silicon base wafers for cost issues.

5.2.1.1 1C Capacitors

In order to assess whether the peak reflection at 30.5° measured in-line through the GIXRD technique was indicative of a crystallization of the HfO_2 layer in the polar orthorhombic phase, it is crucial to submit the test structures to adequate electrical measurements. In this part, the single devices addressed are integrated on a wafer of the second kind (*i.e.* with no transistors in the FEoL), but are rigorously identical to 1C actually integrated in BEoL 28nm. The elementary devices measured are rectangular-shaped single 1Cs with dimensions of 320nm × 200nm, giving a total active area of $0.064\mu m^2$. Each single 1C is integrated in an 8 by 8 environment of 1Cs so as to ensure process stability in an array environment, but the Device Under Test (DUT) is not electrically connected to the surrounding 1Cs. Due to the scaled dimensions of the 1C and the sensing current limitations of the experimental setup, an elevated number of PUND sequences were applied consecutively on a single 1C to measure an averaged current response higher than the setup electrical

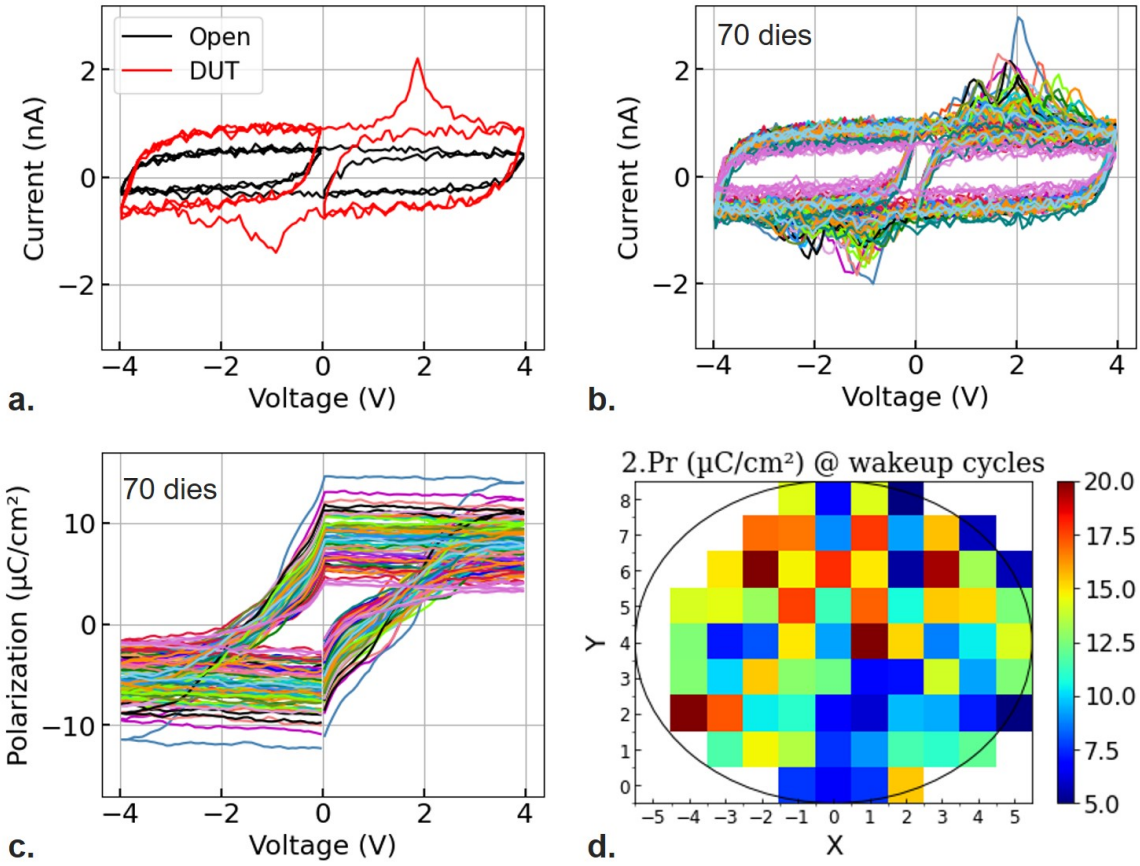


Figure 5.4: After the initial application of 9000 cycling pulses at 4MV/cm - 100kHz: a. I-V characteristics averaged over 1000 consecutive PUND sequences at 4MV/cm with probes lifted (open, black) or contacting the electrodes of a 0.064μm^2 area 1C (DUT, red) ; b. I-V characteristics averaged over 1000 consecutive PUND sequences at 4MV/cm on 70 1Cs with a TiN/HSO (10nm)/TiN ferroelectric stack on a 300mm diameter wafer ; c. Corresponding P-V hysteresis loops ; d. Corresponding wafermap of extracted remanent polarization.

noise. This method requiring stable device operation along repeated PUND,it forbids the study of wake-up in 1Cs, therefore only a picture of the ferroelectric performance after wake-up can be given. Figure 5.4 presents the electrical results measured using the PUND technique at 4MV/cm - 10kHz applied to such 1Cs integrated on a TiN/HSO (10nm)/TiN PoR wafer. The electric field at 4MV/cm is chosen to match the measurement strategy used on 200mm diameter wafers at the 130nm node, and results in 4V applied across a 10nm thick 1C. As was mentioned in Chapter 2 and Chapter 3, this voltage will degrade the performances and reliability of transistors at the 28nm node, but is adopted nonetheless on these single 1Cs to measure the expected full ferroelectric response. Figure 5.4.a., represents I-V characteristics averaged over 1000 PUND sequences after an initial 9000 wake-up cycles at 4MV/cm - 100kHz. In black, the measurement probes are lifted (open circuit) to measure the parasitic current caused by the overall setup capacitance. In red, the

current of a single 1C is measured. The current averaged over 1000 identical measurements is still very impacted by the electrical noise, in part due to the measurement caliber of 1mA chosen on the B1530 equipment needed for 10kHz measurements. However, dielectric current and ferroelectric switching peaks are clearly noticeable, indicating the functionality of the scaled 1C devices and validating the integration process on 300mm wafers. Figure 5.4.b. gather the I-V characteristics measured in the same conditions over 73 dies across the PoR wafer. While the 1Cs of three dies are defective, the other 1Cs (represented by different colors) exhibit a ferroelectric signature. Yet, a wide device-to-device variability in the shape of the ferroelectric current peak can be observed. The P-V hysteresis loops integrated from I-Vs are presented in Figure 5.4.c. While integration contributes to smoothing the signal, it remains quite noisy similarly to I-V characteristics. A large variation of $2.P_R$ can be observed on the wafermap shown in Figure 5.4.d. (from $5\mu C/cm^2$ to $25\mu C/cm^2$), with no immediate pattern noticeable. This might indicate that high variability is due to the method of averaging the current (and hence polarization) over a large number of measurements, rather than caused by a process-related device-to-device variability. Still, a median (maximal) $2.P_R$ of $12.4\mu C/cm^2$ ($24.6\mu C/cm^2$) can be achieved on this PoR wafer, indicating reasonable ferroelectric performances.

In order to determine an optimal integration process for the fabrication of FeCAPs on 300mm wafers (and therefore with mostly different clean room tools than for 200mm wafers), the performances of 1C devices on different wafers with various process variants were assessed, as can be seen in Figure 5.5. In particular, the impact of the bottom electrode, crystallization anneal and ferroelectric layer thickness were assessed. Due to the low current levels and high measurement variability, only a binary conclusion can be safely drawn regarding the functionality (or lack thereof) of the studied split.

- Regarding the bottom electrode, the PoR split consists in a TiN bottom electrode deposited on the W lower via, followed by a CMP step to planarize TiN. From the electrical results, it can be inferred that deposition of the ferroelectric layer directly on W via does not lead to ferroelectric behavior, whereas the CMP step can be omitted without cancelling ferroelectricity.

- Regarding the crystallization anneal, the PoR split consists in an annealing step at 500°C for 15 minutes in a N_2 atmosphere post-TE deposition. From comparison with different wafers, it can be concluded that the absence of post-TE deposition annealing step (or annealing step at 475°C) prevents crystallization in the ferroelectric (orthorhombic) phase, while an annealing step at 525°C guarantees ferroelectricity in the measured 1C devices. Note that this temperature might not be compatible with 28nm transistors.

- Regarding the ferroelectric layer thickness, the PoR split consists in a 10nm Si:HfO_2 (HSO) layer. A wafer with a 8nm thick HSO layer and post-TE deposition annealing step at 525°C [1] has been verified to exhibit ferroelectricity.

[1] As mentioned in Chapter 3, thinner films require a higher annealing temperature to crystallize in the polar orthorhombic phase [90].

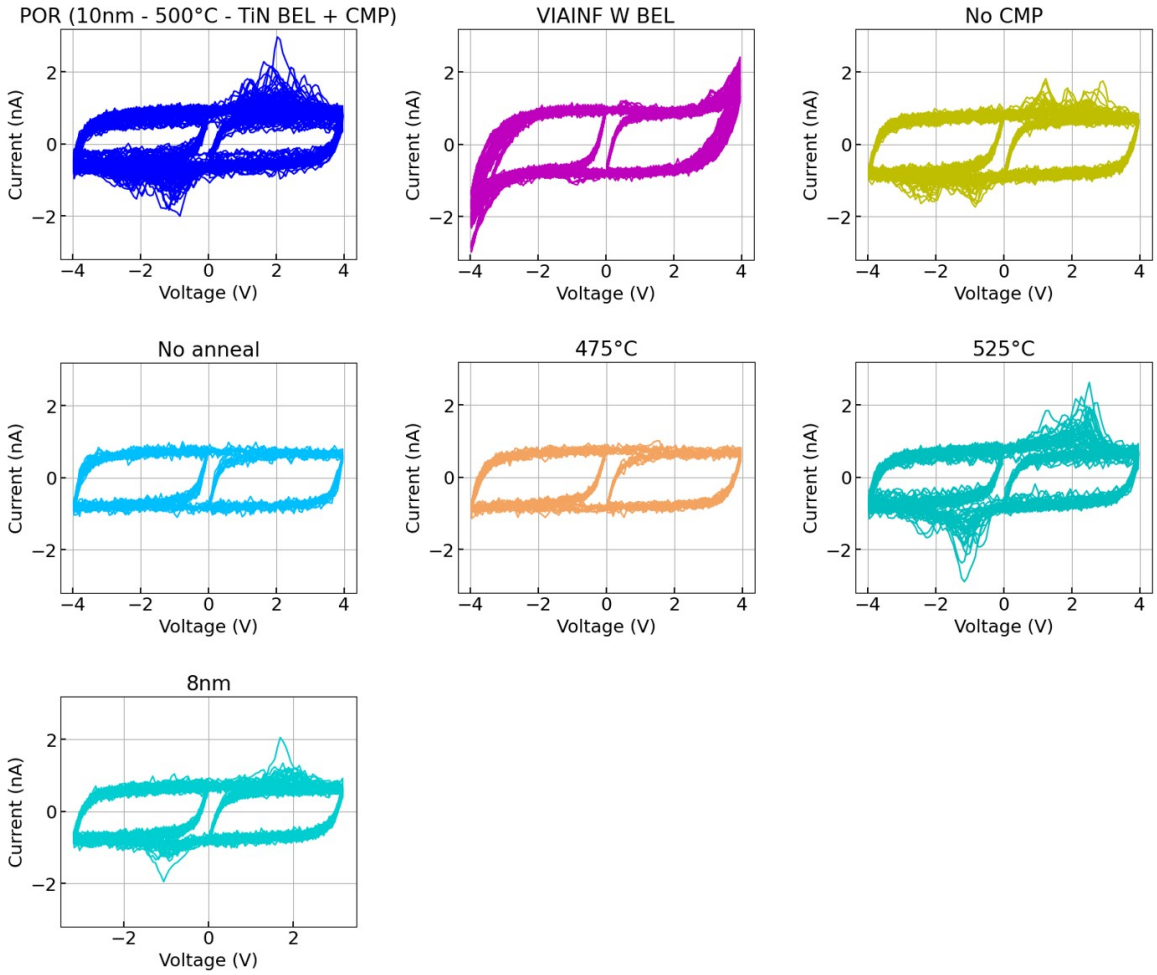


Figure 5.5: I-V characteristics averaged over 1000 consecutive PUND sequences at 4MV/cm on $0.064\mu m^2$ area 1C after the initial application of 9000 cycling pulses at 4MV/cm - 100kHz: several split variants in terms of bottom electric, anneal step and ferroelectric layer thickness are presented.

Recap

Scaled 1C devices with a PoR TiN/HSO (10nm)/TiN stack were integrated on a 300mm diameter wafer with no FEoL transistors, and demonstrated to exhibit ferroelectricity properties. Accurate evaluation of the ferroelectric performances is made difficult by the scaled dimensions of the considered devices in the limits of the sensing precision of the B1530 equipment used. Split variations around the 10nm HSO PoR give insights into optimal integration process of FeCAPs on 300mm wafers.

5.2.1.2 1T Transistors

The electrical results presented in Paragraph 5.2.1.1 were obtained on scaled 1C devices deposited on 300mm diameter base wafers with no FEoL transistors. In this part (and the following part), wafers with FEoL transistors at the 28nm node are analyzed. Specifically, dedicated single transistors (1T) are probed to assess FEoL functionality and to verify that no degradation in performances occurred during the process of integrating the ferroelectric stack in the BEoL. A sanity check of the transistors is first conducted on five wafers presenting different BEoL splits with variation of bottom electrode, crystallization anneal and ferroelectric layer thickness, as evoked in Paragraph 5.2.1.1 on transistors-free wafers. The batch of wafers is tested in a fully-automated probe station programmed for standardized measurements. On each wafer, parametric testing is used on all dies to measure two different NMOS transistors at the 28nm node: a GO2 1T with width dimension of 500nm, and a GO1 1T with width dimension of 1000nm. I_{ds}-V_{gs} characteristics are measured in the linear regime with V_{ds} fixed at 50mV for the GO1 and GO2 NMOS, and in the saturated regime with V_{ds} fixed at 0.9V (1.8V) for the GO1 (GO2) NMOS. The obtained I_{ds}-V_{gs} characteristics enable the determination of on-current (I_{on}), off-current (I_{off}) and threshold voltage (V_{th}) of the transistors, as can be seen on the left-hand side of Figure 5.6. On the right-hand side, Figure 5.6 represents the variability of the three aforementioned metrics across the 66 dies of five different wafers featuring both GO1 and GO2 transistors. Highlighted areas show the specified value ranges defined by STMicroelectronics foundry for I_{on}, I_{off} and V_{th}. The boxplots extracted for the three metrics from measurements on both GO1 and GO2 transistors lie within boundaries provided by foundries specifications. Therefore, it can be concluded that 1C integration in the BEoL and thermal budget (up to 525°C) brought by the annealing step do not degrade the transistors characteristics after process. Note that the reliability of the transistors is not assessed here, and if 1C integration and thermal annealing did lead to a degradation of the reliability of FEoL transistors, it would not be revealed by these tests.

Recap

Both GO1 and GO2 NMOS transistors at the 28nm technology node have been found to fit foundry specifications for I_{on}, I_{off} and V_{th} after the integration of scaled 1C in the BEoL and a thermal annealing step up to 525°C. From a pure specification point of view, scaled 1Cs can be safely integrated in the BEoL of 300mm wafers without compromising the integrity of FEoL transistors at advanced nodes. In addition to these standardized tests, it could be interesting to verify that the process adopted to integrate 1Cs in the BEoL does not degrade transistors reliability.

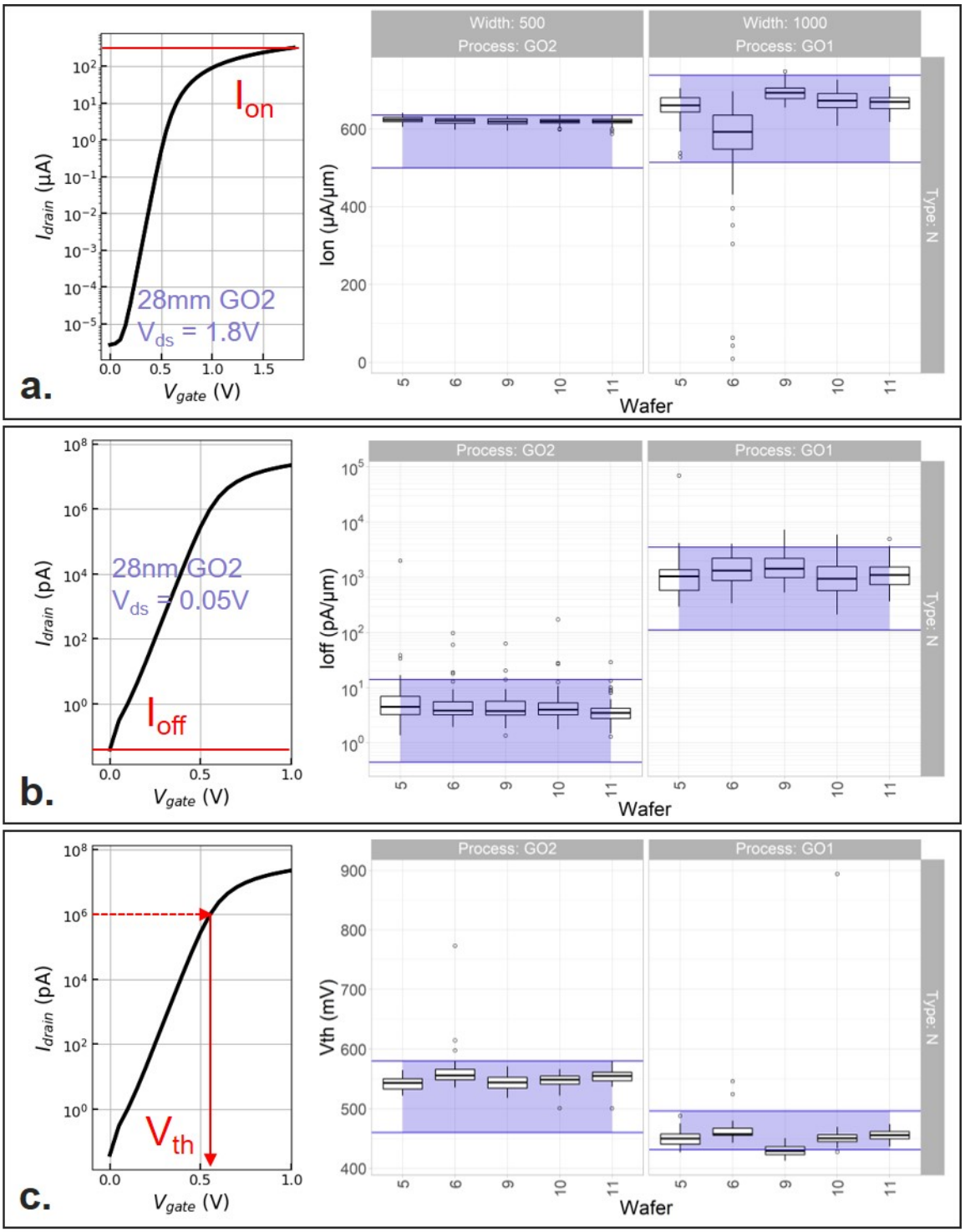


Figure 5.6: Left: Parametric test results for a GO2 NMOS transistor at the 28nm node with L=150nm and W=500nm: a. I_d-V_g at saturation with V_{ds} = 1.8V ; b. and c. I_d-V_g in the linear regime with V_{ds} = 0.05V. Right: Boxplots for I_{on} (a.), I_{off} (b.) and V_{th} (c.) for GO1 NMOS (L=30nm, W=1000nm) and GO2 NMOS (L=150nm, W=500nm) transistors at the 28nm node. The boxplots are extracted over 66 dies measured on five different wafers with various process splits. The purple area represents the foundry-specified values for I_{on} (a.), I_{off} (b.) and V_{th} (c.).

5.2.1.3 1T-1C Bitcells

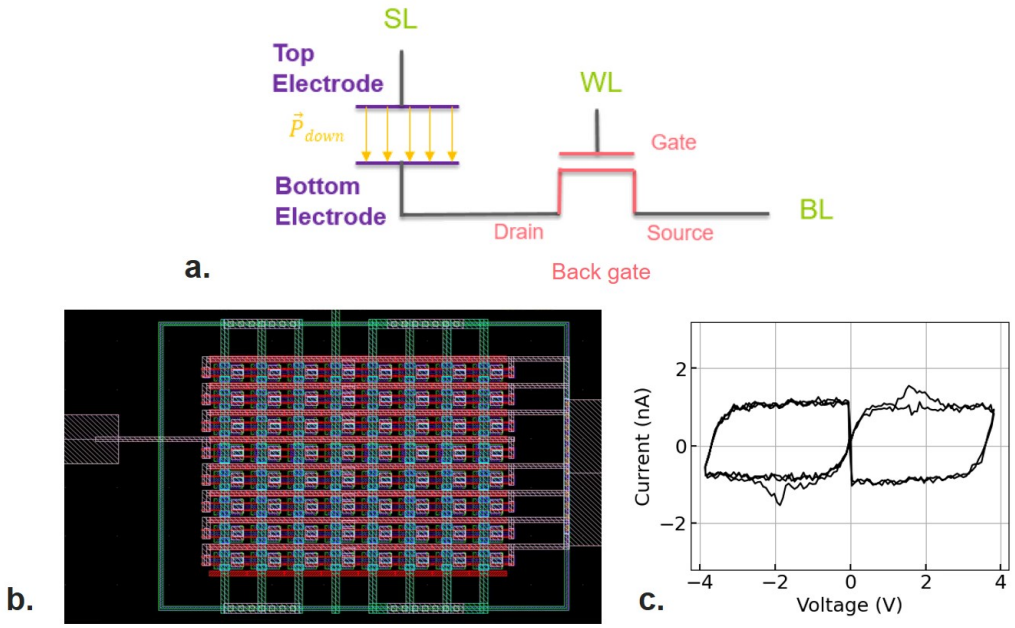


Figure 5.7: a. Schematic of a 1T-1C bitcell summarizing the different ports of the 1T and 1C components and showing its connection to SL, BL and WL. ; b. Layout view of an 8 by 8 environment for the 1T-1C bitcell. ; c. I-V characteristics measured on a 1T-1C bitcell with a GO2 28nm transistor and a FeCAP of 200nm × 320nm dimensions evidencing ferroelectric peaks. 1000 PUND sequences with V_{SL} fixed at 4V and V_{WL} fixed at 2V were applied after 9000 wake-up cycles.

On the second kind of wafers (introduced in Paragraph 5.2.1.2), scaled 1C devices are integrated in the BEoL of transistors at the 28nm node. In Paragraph 5.2.1.1, the functionality of scaled 1Cs was demonstrated on transistors-free wafers, and reasonable 2.P_R levels were obtained. In Paragraph 5.2.1.2, the integrity of transistors was assessed after 1C fabrication process in the BEoL, and key metrics were determined to fit foundry specifications. It is finally of prime importance for FeRAM array integration at advanced nodes to evaluate the performance of 1C devices actually integrated in the BEoL of transistors at the 28nm node, and more importantly of 1T-1C devices. Similarly to the 1T-1C bitcells in FeRAM arrays, these four-port structures enable to polarize and probe the top electrode of the 1C (SL), the gate of the 1T (WL), the source of the 1T (BL) and the back Gate of the 1T. A schematic of a 1T-1C bitcell is given in Figure 5.7.a. As was presented in the previous parts regarding single 1C and single 1T structures, 1T-1C structures are electrically isolated but manufactured in an 8 by 8 environment of dummy bitcells. A view of the layout is presented in Figure 5.7.b. To characterize 1T-1C bitcells, a measurement approach similar to the one applied to single 1C devices is adopted. In order to maintain a reasonable ferroelectric performance, an operating field of 4MV/cm is applied to the top electrode of the 1C (SL), resulting

in a 4V voltage across 1T-1C bitcell. Indeed, from Figure 5.4.b., it can be observed that the positive coercive voltage measured on scaled 1Cs with an identical split is 2V, therefore the operating voltage should be 4V [90]. This voltage is clearly over the nominal (and overdrive) voltage for GO1 and GO2 transistors at the 28nm node, and can consequently be expected to degrade the 1T performance and reliability. However, it is important to note that these stringent operation requirements are specifically designed to ensure a near-perfect yield of transistors in intensive use conditions: it is highly unlikely that this short performance assessment will lead to significant damage to the selection transistors (breakdown, hot-carrier injection or otherwise). Nevertheless, in order to limit the risk of dielectric breakdown of the gate dielectric, the transistor gate is biased at 2V, close to the 1.98V allowed overdrive voltage for GO2 transistors at the 28nm node. Finally, the back gate of the transistor is connected to the ground. As can be seen in Figure 5.7.c., the I-V characteristics obtained on a 1T-1C bitcell with a GO2 transistor and a scaled 1C with 200nm × 320nm dimensions exhibits ferroelectric peaks. The obtained signal is less noisy compared to Figure 5.4.b., as the measurement caliber was revised to the more adapted 1µA range. As had been demonstrated in [67] at the 130nm node, the I-V (and P-V) characteristics are expected to be very close for 1C and 1T-1C structures, although the measuring technique leads to uncertainty in the extracted remanent polarization values. This result validates the 1C integration in the BEoL of transistors at the 28nm node, and the functionality of 1T-1C bitcells. This constitutes a first-of-a-kind demonstration, and a crucial step towards the integration of FeRAM arrays at the 28nm node. In the literature, the only integration result close to this one was demonstrated by Intel in [119], which reported a functional anti-ferroelectric single capacitor integrated with advanced nodes FinFET transistors. Other than that, FeFET configurations were achieved with planar transistors at the 28nm and 22nm nodes [85, 142], but encountered reliability issues that hindered their progression towards application.

Recap

The ferroelectric behavior of a 1C in the 1T-1C bitcell configuration validates the integration process at the 28nm node. In order to preserve the transistors performance and reliability, lower voltages need to be applied, and therefore lower 1C coercive voltages have to be achieved. In this demonstrator, a 10nm thick FE layer was used. Consistently with simulations from Chapter 2 and first demonstrations from Chapter 3, it is necessary to tend towards lower FE layer thickness to enable low voltage operation at advanced nodes.

5.2.2 Capacitor Fields

In order to refine the integration process and to determine an adequate PoR for the integration of functional FeRAM arrays at advanced nodes, a valuable intermediate step is to assess the performance of FeCAP fields integrated on 300mm wafers. Therefore, to complete the analysis reported in Paragraph 5.2.1.1 and based on measurements at the single device level, an adequate demonstrator is fabricated and integrated on 300mm wafers. While there are no 28nm FDSOI transistors in the FEoL of the 300mm wafers considered, the individual 1C constituting the FeCAP fields are scaled

and dimensioned to match those of BEoL-integrated devices. In this study, the FeCAPs measured are constituted of 2728 circular 1C devices with diameters of 130nm or 200nm depending on the chosen DUT, for a total FeCAP field area of $36\mu m^2$ and $86\mu m^2$, respectively. In Paragraph 5.2.1.1, only a binary assessment of ferroelectricity could be provided because the initial demonstrator featured only single devices that were challenging to finely characterize because of their limited switching current. In this part, the FeCAP fields are much easier to characterize and precise ferroelectric performance, as well as endurance behavior, can be evaluated. Moreover, a new PoR based on the ALD-deposited HZO material is defined for improved performances. For this demonstrator, the annealing step of PoR wafers is performed at the end of the process, *i.e.* post-M6 deposition. Figure 5.8.a. presents the pristine and post wake-up I-V characteristics measured by non-averaged PUND sequences at 4MV/cm - 10kHz during the course of cycling at 4MV/cm - 100kHz. 63 devices with a $36\mu m^2$ area were measured across a 300mm wafer with a TiN/HZO (10nm)/TiN ferroelectric stack. While pristine characteristics evidence antiferroelectric-like behavior with two current peaks, they tend to merge after 10^4 cycles to form a very sharp single peak. Although the intensity of the peak varies from one DUT to the other, a high homogeneity is observed. The P-V loops presented in Figure 5.8.b. exhibit pinched hysteresis loops in the pristine state that straighten with cycling. Figure 5.8.c. presents the remanent polarization evolution as a function of cycling, and the classic behaviors of wake up, maximum and ferroelectric fatigue can be observed. Similarly to the results obtained on FeCAP fields integrated on 200mm wafers, the peak $2.P_R$ is detected after 10^4 cycles. Figure 5.8.d. presents the wafermap of $2.P_R$ after wake-up, and as hinted by the overlapping I-V and P-V characteristics, a good homogeneity is observed, confirming that previous dispersion (Paragraph 5.2.1.1) was adequately attributed to measurement variability and not device (1C) variability. Finally, Figure 5.8.e. presents positive and negative coercive fields variation with cycling. In this 10nm HZO stack, coercive fields are greater than 1MV/cm, which leads to operating voltages greater than 2V ($2V_c$, fom [90]). Those values are still high for operation at the 28nm node, especially for GO1 transistors. From Chapter 2 and Chapter 3, it has been demonstrated that thinner ferroelectric layers should enable the reduction of coercive voltages, and hence of operating voltages.

Apart from this 10nm thick HZO split, three other wafers were fabricated: the impact of the CMP step after BE deposition was studied, and the position of the thermal annealing step, as well as the ferroelectric layer thickness were varied to explore various integration options. Figure 5.8.f. gathers the cumulated $2.P_R$ distributions after wake-up for the four wafers studied: the PoR wafer (red), a second wafer without the CMP step after deposition of the bottom electrode (green), a third wafer without the CMP step and with the annealing step right after the deposition of the TiN top electrode (blue), and a third 6nm HZO wafer without the CMP step and with the annealing step right after the deposition of the TiN top electrode (cyan). In Paragraph 5.2.1.1, electrical characterization results on 1C devices with and without the CMP step after BE deposition provided the only piece of information that both options lead to ferroelectricity in the wafers tested. Figure 5.8.f. provides the conclusion that the CMP step after BE deposition improves the median remanent polarization. Moreover, performing the annealing step at the end of the process leads to an improved remanent polarization. Finally, when the annealing step is positioned at the end of the process, less leakage current is observed at high voltage on the I-V characteristics. The position of the annealing step, as well as its temperature, might explain the poor performance of the 6nm HZO wafer. Indeed, its I-V characteristics do not exhibit ferroelectric peaks, hence the null remanent polarization observed in Figure 5.8.f. In order to improve the ferroelectric performances of FeCAPs with thinner films

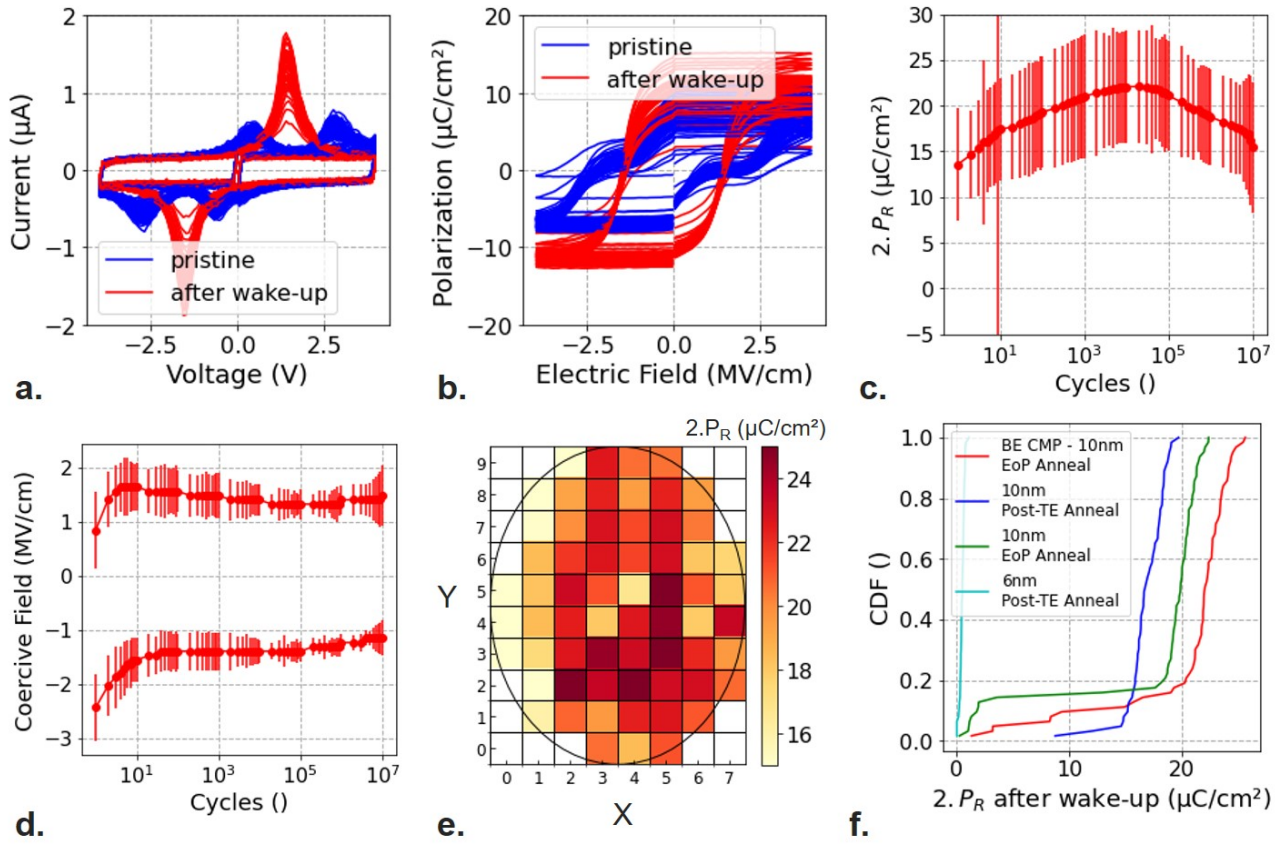


Figure 5.8: a. I-V characteristics ; b. P-V hysteresis loops ; c. Median and standard deviation of $2.P_R$ evolution versus applied cycles ; d. Median and standard deviation of coercive fields evolution versus applied cycles ; e. Wafermap of $2.P_R$ after 10^4 wake-up cycles (a. to e.) from 63 FeCAP fields of $36\mu m^2$ area with a TiN/HZO (10nm)/TiN ferroelectric stack across a 300mm diameter wafer, presenting a CMP step after TiN BE deposition, and an End-of-Process (EoP, post-M6 deposition) crystallization anneal ; f. Normalized cumulative distributions for four different splits: HZO 10nm with CMP and EoP anneal, HZO 10nm without CMP and post-TE deposition anneal, HZO 10nm without CMP and EoP anneal, HZO 6nm without CMP and post-TE deposition anneal.

integrated in the BEoL of 28nm transistors, it is necessary to optimize the crystallization annealing, as had been shown with Nanosecond Laser Anneal (NLA) in Chapter 3. It can also be useful to optimize the composition of the ferroelectric material, improve the interface quality [103], co-dope the ferroelectric layer [99], or adopt other fabrication technique, such as Atomic Layer Etching [143].

Recap

Scaled FeCAP fields were successfully integrated on 300mm diameter wafers, and exhibit good ferroelectric performances (median 2.P_R close to 20$\mu C/cm^2$) and high homogeneity. Thanks to a large developed area, wake-up and fatigue behaviors could be measured and are closely in line with results obtained on 200mm wafers. Split variation on the position of the crystallization anneal gives useful information for the integration of 1T-1C FeRAM arrays at the 28nm node. Material and process (CMP, annealing step among others) have to be further optimized to integrate very thin ferroelectric layers in the BEoL of 28nm transistors.

Conclusions

This chapter aimed at providing a first demonstration of ferroelectric capacitor integration in the BEoL of transistors at advanced nodes, as well as first electrical evidence verifying the integrity of both transistors and FeCAP devices. A secondary goal, but nonetheless important, was to obtain useful information regarding critical process options in view of determining an optimal PoR for FeRAM array integration at advanced nodes. To this end, the step-by-step development was presented and accounted for various fabrication constraints.

In this chapter, several devices were manufactured on 300mm wafer demonstrators. The base wafers on which the ferroelectric stack is integrated are fabricated until Metal 5 by STMicroelectronics, with two options: either with FDSOI transistors at the 28nm node in the FEoL, or on transistors-free wafers. The integration of the ferroelectric stack up to Metal 6 (followed by passivation and pillar levels) requires up to seven lithography steps and enables the fabrication of TiN/HSO or HZO/TiN capacitors that can be connected to the drain of GO1 or GO2 FDSOI transistors at the 28nm node. Initial observations on the entire stack reveal a very clean integration, and XRD measurements evidence the beneficial role of the thermal annealing step to crystallize the HfO_2-based material in the tetragonal or orthorhombic phase.

On a first demonstrator, only single devices (capacitors 1C and transistors 1T) were integrated. While the scaled dimensions of the capacitors ($0.064\mu m^2$ area) make the fine assessment of electrical performances challenging, ferroelectric behavior was evidenced on capacitors integrated on 300mm transistors-free wafers. Thanks to a first split evaluation on an ensemble of wafers, it can be concluded that annealing temperatures of 500°C and above enable crystallization in the polar orthorhombic phase, a TiN bottom electrode (with or without a CMP step) is required to generate ferroelectricity, and that 10nm and 8nm layers are ferroelectric in the aforementioned conditions. The integrity of GO1 and GO2 FEoL transistors at the 28nm node was then evaluated after FeCAP integration in the BEoL: the on-, off-current and threshold voltage were determined to fit the foundry specifications, thus validating the transistors performances. On this first demonstrator, the functionality of the 1T-1C bitcell was also demonstrated, with voltages higher than the overdrive voltages allowed to preserve the integrity and reliability of transistors. Thinner ferroelectric layers with lower coercive voltages are therefore required.

On a second demonstrator, scaled ferroelectric capacitor fields were integrated on 300mm wafers to enable the fine analysis of the ferroelectric behavior and the determination of an optimal PoR. It was demonstrated that the ferroelectric performances of FeCAPs were very similar to that of their counterparts on 200mm wafers, with a median remanent polarization around $20\mu C/cm^2$ obtained on a TiN/HZO (10nm)/TiN stack. Optimal process was determined to be 10nm HZO, with a CMP step after TiN BE deposition and a post-M6 crystallization anneal. The wafer presenting a 6nm HZO layer was not ferroelectric, and needs to be optimized for FeRAM array operation at advanced nodes, and therefore with a limited maximal voltage.

Conclusions and perspectives

Since the discovery of ferroelectric properties in doped hafnium oxide thin films, the interest has grown for this material (otherwise extensively used in the microelectronic industry for its high-k dielectric properties) and the potentialities it offers. Indeed, while the origin of its ferroelectricity and more precisely of the stabilization of the non-centrosymmetric orthorhombic phase remains an open research subject, it soon appeared that HfO_2-based ferroelectric memories could overcome the limitations encountered by more traditional perovskite-based ferroelectrics like PZT. PZT-based FeRAM technology is commercialized today in niche applications, and seems clamped at the 130nm node. However, in the big data era currently unfolding, an exponentially growing amount of varied data is generated and needs computation and storage at increasing energy costs. Combined with the progressive scaling of transistors to advanced nodes, this creates the need for scalable embedded memories with high-performance and low power consumption. In order to evaluate the potential of HfO_2-based ferroelectric memories for fulfilling this demand, the objectives of this thesis were to optimize the performances of HfO_2-based FeRAM arrays integrated into BEoL CMOS process at the 130nm node, as well as projecting them towards integration at more aggressive nodes with levelled performances in high density and high capacity arrays. To this end, various aspects were investigated on FeRAM arrays BEoL-integrated at the 130nm node, and the results of this research was presented in four chapters aiming at giving an extensive picture of the work provided to develop high performance FeRAM arrays at the 28nm node.

In the second chapter, a theoretical approach was adopted to evaluate the constraints brought by scaled transistors at advanced nodes and accordingly propose guidelines towards transistors and capacitors dimensioning for functional FeRAM arrays at the 28nm node. A FeCAP compact model and associated array-like environment, as well as various analytical equations for the memory window or switching capability were instrumental in addressing the impact of voltage and current limitations at advanced nodes on FeCAP performances. In particular, reduced ferroelectric layer thicknesses, 3D integration of FeCAPs, and smart designing of FeRAM arrays were suggested as interesting paths to explore.

In the third chapter, therefore, technological achievements in line with recommendations from Chapter 2 were demonstrated. The performances of devices and arrays with an alternative annealing process for the crystallization of sub-5nm films, different integrations of FeCAPs for high density arrays, or an innovative reading scheme for high capacity arrays, were evaluated. Moreover, in an attempt at maximizing the read margin at FeRAM array level, stack optimization and bitcell-to-bitcell variability assessment were the focus of in-depth research. While the process- and design-induced variability study lead to the correction of some limiting effects, $2.P_R$ gains thanks to stack optimization were achieved at the expense of endurance.

It is therefore essential to take into account the reliability of FeCAPs and FeRAM arrays for a complete assessment of the potential of HfO_2-based ferroelectric memories in specific array configurations, which was especially targeted in the fourth chapter. An extensive data retention

assessment has been proposed for the first time at FeCAP and FeRAM array level ; additional studies on endurance and heavy ion irradiation hardness were also conducted. Ferroelectric fatigue, imprint, and material amorphization were shown to be critical phenomena degrading the endurance, data retention, and resistance to irradiation, respectively, at FeCAP and FeRAM array level. Those potentially blocking points were deeply studied and it emerged that carefully chosen electrical recovery strategies could enhance the reliability performances in the case of fatigue and imprint. As for the resistance to irradiations, the historical advantage of FeCAPs over charge-based transistors was confirmed.

After these promising demonstrations, an intermediate attempt at integrating scaled FeCAPs in the BEoL of transistors at the 28nm node on 300mm diameter wafers was proposed. A full process flow was presented and realized, and it achieved the fabrication of functional ferroelectric devices while maintaining the integrity of FEoL transistors. It enabled the determination of favorable process conditions in view of integrating FeRAM arrays at the 28nm node or below.

To conclude, precise integration constraints assessment and preliminary demonstrators realization prove that there is no blocking point preventing the successful integration of high performance, high density and high capacity FeRAM arrays at advanced nodes. What's more, reliability weak points can be corrected to enhance the retention and endurance of FeRAM arrays, thanks to specific electrical program / read sequences. It is therefore expected that upcoming FeRAM array integration at the 22nm will yield satisfying performances and pave the way for practical application to relevant circuits.

Looking ahead, the development of high performance, high capacity and high density FeRAM arrays will be possible only through appropriate stack optimization, as well as combined integration and design efforts. Results by Micron in [92, 144, 145] demonstrate that a wake-up free ferroelectric stack with a $2.P_R$ higher than $50\mu C/cm^2$ can be obtained through electrode material engineering and probable co-doping of the ferroelectric layer, as shown in [99]. Another promising strategy for wake-up free behavior could lie in HfO_2/ZrO_2 nanolaminates instead of HZO solid solution [146]. This intrinsic ferroelectric performance optimization highly benefits to the large read margin obtained at array level, which contributes to an enhanced reliability despite endurance and data retention limitations. Dedicated refresh sequences seem however still useful to recover from the effects of fatigue and imprint [92]. Similarly, the characterization method needs to be extended to account for the combined effect of all use cases, especially the most critical ones, in order to project towards functional applications. To account for specific voltage requirements at advanced nodes, low thickness (5.7nm) of the ferroelectric layer in the aforementioned Micron NVDRAM, as well as in Sony Mbit array (6nm, [91]) enable low-voltage operation at 1.5V and 1.8V, respectively. In both arrays, 3D capacitors were successfully integrated in DRAM structures with aspect ratios greater than 1/20 or metallic vias, respectively, to enhance the ferroelectric performance at a reduced footprint. Finally, bank partitioning was further adopted by Sony [91] for high performing large capacity arrays. These publications confirm that the objectives and achievements of this thesis pursued the right path towards FeRAM array integration at advanced nodes.

Moreover, the parallel design of dedicated circuits targeting specific applications is a useful step towards product development. Recent work by CEA-Leti ([147]) and Sony [148], and of course the Micron 32Gbit, explore the potential of HfO_2-based capacitors for NVSRAM and NVDRAM applications. Regarding AI-oriented applications, circuits by [149, 150, 151] including or relying on ferroelectric capacitors provide promising demonstrations towards in-memory computing.

Synthèse en français

Dans notre monde hyper-connecté, l'industrie des semi-conducteurs est évaluée à plus de 600 milliards de dollars dans le monde en 2024. Les circuits intégrés constituent plus de 80 % de cette valeur estimée, et incluent des secteurs tels que les mémoires stand-alone, les circuits logiques, les microcontrôleurs et les circuits analogiques. Les deux premiers présentent les tendances de croissance les plus élevées (77 % et 11 % respectivement), tandis que le secteur des microcontrôleurs, qui intègre notamment les mémoires embarquées, devrait atteindre une valeur de près de 80 milliards de dollars en 2024. Les composants considérés répondent aux besoins en puces microélectroniques dans les branches de l'électronique grand public, de l'industrie 4.0 et du secteur automobile. En plus d'être une industrie rentable, ou peut-être à cause de cela, l'industrie des semi-conducteurs est un secteur très compétitif pour les groupes industriels, et un secteur d'investissement stratégique à l'échelle nationale. En effet, la récente crise du Covid a révélé la dépendance de l'industrie européenne des semi-conducteurs vis-à-vis des fabricants de dispositifs intégrés d'Asie du Sud, qui présentent par ailleurs une vulnérabilité environnementale et géopolitique croissante. Dans ce contexte, les pays européens s'allient au travers du European Chips Act, qui vise à renforcer la souveraineté de l'écosystème européen des semi-conducteurs. Dans ce cadre, un projet de ligne pilote à grande échelle (FAMES) a été lancé par la Commission Européenne dans le but de développer des puces économes en énergie pour une électronique plus verte. Un axe particulier de ce projet est le développement de mémoires non volatiles embarquées, incluant l'intégration de mémoires ferroélectriques (FeRAM) à des nœuds technologiques avancés.

En effet, dans un contexte où la quantité de données générées augmente exponentiellement avec des exigences de calcul et de stockage plus élevées et des coûts énergétiques croissants, l'intégration de mémoires embarquées à haute densité et faible consommation d'énergie est devenue cruciale, en particulier avec la réduction progressive des dimensions des transistors vers des nœuds technologiques avancés. C'est dans cette perspective que cette thèse vise à optimiser les performances des matrices FeRAM à base d'oxyde d'hafnium (HfO_2) intégrées dans le Back-End-of-Line (BEoL) de la technologie Complementary Metal Oxide Semiconductor (CMOS) au nœud de 130 nm, et à extrapoler ces performances vers des nœuds plus agressifs tout en augmentant la densité et la capacité mémoire des matrices. L'intérêt pour ce matériau a considérablement augmenté depuis la découverte de propriétés ferroélectriques dans des films minces d'HfO_2 dopé, diélectrique par ailleurs largement utilisé dans l'industrie microélectronique pour sa haute permittivité. Bien que l'origine de la formation et la stabilisation de la phase orthorhombique, responsable de la ferroélectricité dans l'HfO_2, fasse encore l'objet de recherches approfondies, les mémoires ferroélectriques à base d'HfO_2 sont rapidement apparues comme des alternatives prometteuses capables de surmonter les limites des ferroélectriques à base de pérovskite, comme le PZT, dont l'avancée technologique est limitée au nœud de 130 nm.

Comme décrit dans les parties précédentes, la découverte des propriétés ferroélectriques dans des films minces à base de HfO_2 compatibles avec la technologie CMOS ouvre la voie à des dispositifs

de mémoire ferroélectrique ultra-basse consommation pouvant être miniaturisés en-deçà du nœud 130nm. Des travaux antérieurs au CEA-Leti ont démontré l'intégration réussie de condensateurs ferroélectriques de faibles dimensions à base d'oxyde d'hafnium dopé au silicium dans la partie BEoL de la technologie CMOS au nœud de 130 nm. Des matrices FeRAM fonctionnelles de 16 kbits de type 1T-1C ont été rapportées, et des paramètres clés tels que l'endurance, la vitesse de commutation ou la compatibilité avec le budget thermique de la soudure (solder reflow compatibility) ont été évalués. Ces résultats préliminaires constituent le point de départ de cette thèse, qui s'inscrit dans le cadre de recherche et développement sur ces matrices FeRAM 16 kbits, en vue d'optimiser leurs performances, ainsi que de les intégrer dans des matrices plus grandes et plus denses à des nœuds plus avancés. Dans cet esprit, les objectifs principaux de cette thèse sont doubles :

- Évaluer et optimiser les performances intrinsèques et la fiabilité des matrices FeRAM 16 kbits au nœud de 130 nm, afin d'évaluer le potentiel de cette technologie d'un point de vue applicatif. Cela nécessite de plonger plus en profondeur dans les mécanismes se produisant à l'échelle du dispositif unitaire et impactant les performances et la fiabilité au niveau de la matrice FeRAM.
- Évaluer le potentiel de miniaturisation des matrices FeRAM à base de HfO_2 vers des systèmes à des nœuds plus avancés. Cela inclut des études théoriques et dimensionnelles pour déterminer les contraintes imposées par les transistors à ces nœuds avancés, ainsi que le développement de démonstrateurs intermédiaires vers ces nœuds avancés.

À cette fin, cette dissertation est divisée en quatre chapitres.

Chapitre 2 : Etude théorique et simulatoire

Ce chapitre visait à proposer des recommandations argumentées pour le dimensionnement des transistors et des FeCAP en vue de l'intégration de matrices FeRAM 1T-1C à des nœuds technologiques avancés. À cette fin, plusieurs outils ont été adoptés et ont permis d'explorer quantitativement et qualitativement les conséquences de la miniaturisation des transistors 1T au nœud de 28 nm, ainsi que d'évaluer la faisabilité de l'expansion de la taille des matrices combinée aux exigences accrues en densité de celles-ci.

Le premier outil développé était un modèle compact FeCAP inspiré de Preisach. Bien que l'historique de la modélisation du comportement hystérétique remonte aux théories de Landau et Preisach des années 1930, son adoption plus large pour la modélisation des condensateurs ferroélectriques a suivi le développement progressif des FeCAPs à base de PZT et d'HfO_2 pour les mémoires ferroélectriques. De manière similaire aux modèles basés sur Preisach dans la littérature, le modèle développé visait à être efficace en termes de calcul. Contrairement à la plupart des modèles inspirés de Preisach, celui-ci visait à décrire l'évolution du courant ferroélectrique plutôt que la polarisation. Le modèle complet, calibré sur des données expérimentales, permet des simulations précises du courant pour une grande variété de signaux d'entrée. En particulier, il reproduit très bien la réduction de la polarisation rémanente avec la diminution de la tension de fonctionnement, en raison d'une commutation sous-saturée. Un circuit élémentaire a été précisément dimensionné autour de la cellule 1T-1C pour émuler le fonctionnement d'une matrice

de 16 kbits, ce qui a permis de bien reproduire la dégradation de la fenêtre mémoire (Memory Window, MW) à basse tension au nœud de 130 nm. L'adoption de transistors au nœud de 28 nm a été simulée et montre un impact limité sur la MW dans la gamme de tensions disponibles. Il est alors recommandé de réduire l'épaisseur de la couche ferroélectrique de la FeCAP à 5 nm pour maintenir une MW de 0,1 V en-dessous d'une tension de fonctionnement de 2 V. Un défi considérable soulevé par cette recommandation est de réaliser la fabrication de couches ferroélectriques très minces sans dégrader leur polarisation rémanente. En outre, des calculs simples ont permis d'évaluer que la réduction d'échelle du transistor au nœud 28 nm à des dimensions minimales ralentirait légèrement la capacité de commutation rapide du ferroélectrique, mais que la plage de nanosecondes atteignable reste acceptable pour le fonctionnement des matrices FeRAM. Selon l'application ciblée, les dimensions du transistor peuvent être ajustées pour permettre un flux de courant plus rapide.

Les transistors aux nœuds avancés (et donc de dimensions réduites et de capacité source-grille plus faible) ont un effet largement bénéfique sur le développement de matrices FeRAM plus denses et plus grandes. Cependant, les contraintes se reportent alors sur la FeCAP, et spécifiquement sur la réduction de son empreinte pour atteindre une densité plus élevée. En règle générale, des calculs analytiques supplémentaires ont indiqué que les tentatives de développement de matrices plus denses (réduction du diamètre de la FeCAP) et plus grandes (augmentation de la capacité de Bit Line, C_{BL}) conduiraient à une polarisation rémanente ($2.P_R$) irréaliste requise pour maintenir une MW de 0,1 V. Une solution envisagée pour maintenir une large fenêtre mémoire dans des matrices plus grandes tout en favorisant une haute densité consiste à développer des FeCAPs en 3D.

En conclusion, il peut être affirmé que l'optimisation de l'empilement vers des couches ferroélectriques plus fines, ainsi que les efforts conjoints de conception et d'intégration, peuvent conduire au développement prometteur de matrices FeRAM au nœud de 28 nm. Dans les chapitres suivants, les fenêtres mémoire à la fois médiane et au niveau de la matrice seront optimisées à cet égard, et la fiabilité des matrices FeRAM sera explorée avant que l'intégration à des nœuds avancés puisse être démontrée.

Recap

Le deuxième chapitre a adopté une approche théorique pour évaluer les contraintes imposées par les transistors à des nœuds avancés et proposer des directives pour le dimensionnement des transistors et des condensateurs pour des matrices FeRAM fonctionnelles au nœud de 28 nm. Un modèle compact de FeCAP et un environnement de matrice ont été utilisés pour analyser l'impact des limitations de tension et de courant aux nœuds avancés sur les performances des FeCAPs. Il a été suggéré d'explorer des couches ferroélectriques plus minces, l'intégration 3D des FeCAPs et des conceptions intelligentes des matrices FeRAM.

Chapitre 3 : Amélioration des performances et réalisations technologiques

L'objectif de ce chapitre était d'explorer les variations de l'empilement ferroélectrique, les innovations technologiques et la variabilité induite par les procédés et la conception, en vue d'améliorer les performances brutes des FeCAPs et des matrices FeRAM au nœud de 130 nm, pour une future intégration à des nœuds technologiques plus agressifs. À cette fin, les performances de la fenêtre mémoire ont été étudiées en profondeur en se concentrant à la fois sur les performances médianes de la fenêtre mémoire (MW(0σ)) et sur sa variabilité au niveau de la matrice.

Dans un premier temps, il a été démontré que le fait de graver la couche ferroélectrique ou d'insérer une couche de Ti entre la couche ferroélectrique et l'électrode supérieure en TiN avait un effet bénéfique en améliorant la $2.P_R$ au niveau des FeCAPs de dimensions réduites intégrées dans le BEoL, et ainsi la MW(0σ) au niveau des matrices FeRAM de 16 kbits. Les mécanismes responsables de ces gains en performance ont été attribués respectivement à la libération de contraintes mécaniques et à l'optimisation de la concentration en lacunes d'oxygène par le piégeage d'oxygène via le Ti. Dans les deux cas cependant, le gain en $2.P_R$ a été obtenu au détriment d'une endurance dégradée et d'une augmentation du claquage diélectrique. En effet, la qualité de l'étape de gravure et l'augmentation de la concentration en lacunes d'oxygène pendant le cyclage ont été identifiées comme favorisant la formation de chemins conducteurs entre les électrodes. Si la première peut être améliorée par l'optimisation du processus, la seconde semble en revanche plus difficile à corriger.

En plus de la fenêtre mémoire médiane, la variabilité entre cellules mémoires est soigneusement étudiée, d'autant plus qu'elle réduit la MW(16k) et donc la marge de lecture au niveau des matrices FeRAM. Pour identifier les sources de variabilité intrinsèque au niveau des matrices, une première partie a évalué la contribution de la variabilité induite par les procédés de fabrication. À partir de mesures des métriques liées au matériau et aux dimensions à l'échelle de la matrice et du wafer, il a été montré que la variabilité du procédé de fabrication ne contribuait pas de manière significative aux pentes de distribution des deux états logiques. Pour évaluer les contributions liées au design, il a été montré que la variabilité des transistors le long des Bit Lines (BL) et le gradient de lecture à travers les BL contribuaient chacun à environ $10mV/\sigma$, soit presque la moitié des pentes de distribution totales. La correction de la chute de tension pendant l'opération de lecture permettrait de réduire considérablement la variabilité liée au gradient et ainsi d'améliorer la marge de lecture au niveau de la matrice.

Enfin, conformément aux recommandations émises au chapitre 2, des avancées technologiques ont été démontrées. Concernant l'exigence de haute capacité mémoire, l'effet de la MW limitée par C_{BL} dans les matrices plus grandes a été démontré sur des matrices de 16kbit, 64kbit et 256kbit. Pour surmonter ce problème, la partition en banks a été émulée et un système de lecture alternatif basé sur la charge a été évalué. Les deux solutions ont contré la diminution de la MW avec l'augmentation de la taille des BL. Concernant l'exigence de haute densité, des structures FeCAP 2D+3D ont été fabriquées sur des wafers en boucle courte (Short Loop) et ont confirmé le gain de performance à empreinte constante fourni par les condensateurs 3D. Enfin, la cristallisation d'un film ferroélectrique de 4 nm d'épaisseur par recuit laser nanoseconde (Nanosecond Laser Anneal, NLA) a été proposée pour répondre à la limitation de tension dans les matrices FeRAM à des nœuds

avancés. Ces avancées constituent des démonstrateurs intermédiaires qui doivent être développés davantage pour une future mise en œuvre dans les matrices FeRAM à des nœuds avancés.

Recap

Le troisième chapitre a démontré des avancées technologiques conformes aux recommandations du chapitre précédent. Les performances de dispositifs unitaires basés sur des films sub-5nm cristallisés par avec une étape de recuit alternatif ont été évaluées, ainsi que des intégrations différentes de FeCAPs pour des matrices à haute densité et des schémas de lecture innovants pour des matrices à grande capacité mémoire. Cependant, bien que des gains de $2.P_R$ aient été obtenus grâce à l'optimisation de l'empilement ferroélectrique, ceux-ci ont été associés à une réduction de l'endurance, soulignant l'importance de trouver un équilibre entre performances et fiabilité.

Chapitre 4 : Fiabilité à l'échelle des composants unitaires jusqu'aux matrices

Ce chapitre visait à évaluer les performances de fiabilité des mémoires ferroélectriques, en s'intéressant particulièrement à combler l'écart entre les FeCAPs intégrés en BEoL et les matrices FeRAM orientées vers des applications. À cette fin, trois principaux problèmes de fiabilité ont été évalués, et des solutions d'amélioration ont été proposées.

Lors de l'évaluation de la rétention des données, l'impact limité de la dépolarisation thermique a été souligné, en raison de la grande stabilité des performances ferroélectriques dans la plage de température considérée de 85°C à 125°C. Un problème clé pour la rétention a été déterminé comme étant l'imprint, en ligne avec les travaux précédents de la littérature sur les matrices FeRAM à base de PZT et HfO_2. En effet, bien que l'effet de l'imprint à température ambiante ait été montré comme étant négligeable, ne conduisant à aucun bitfail dans une matrice FeRAM après 348 jours de stockage, il a rapidement dégradé l'état opposé (Opposite State, OS) à des températures élevées, fermant ainsi la fenêtre mémoire. Un schéma de mesure équivalent et une modélisation de la réduction de la $2.P_R$ de l'OS (ou MW au niveau des matrices FeRAM) ont mis en évidence des dynamiques d'imprint identiques aux niveaux FeCAP et matrice FeRAM, ainsi qu'une chute de tension supplémentaire liée au design de la matrice, ce qui a conduit à une rétention apparente dégradée dans les matrices FeRAM. L'effet d'historique de l'imprint, qui amène les bits programmés dans un état à revenir partiellement à cet état après reprogrammation, a été démontré comme étant récupérable grâce à l'application d'un nombre de cycles soigneusement choisi au champ de fonctionnement.

La récupération de la fatigue a également fait l'objet de la deuxième partie de ce chapitre, qui visait à démontrer une récupération efficace de la fatigue ferroélectrique au niveau des FeCAPs et des matrices FeRAM. Après avoir évalué l'endurance limitée par le claquage diélectrique et la fatigue ferroélectrique, et leurs dépendances opposées à la tension de fonctionnement, une évaluation extensive des séquences de cyclage a été menée pour déterminer d'éventuelles stratégies

de récupération. Dix répétitions de cycles de fatigue (à faible champ de cyclage) - récupération standard (à fort champ de cyclage) ont démontré une récupération réussie de la $2.P_R$, améliorant ainsi effectivement l'endurance d'une décade. Cela a également révélé qu'un état stable post-fatigue pouvait être atteint après chaque répétition, sans indication de courant de fuite accru. L'investigation suivante de diverses stratégies de récupération a permis de déterminer que des champs élevés, de longues impulsions et un grand nombre de cycles produisent les meilleurs taux de récupération. Il a été démontré que le cyclage bipolaire était plus efficace que tout autre schéma unipolaire, ce qui est attribué au déblocage des domaines par la recombinaison des charges positives piégées à l'interface entre la couche ferroélectrique et les deux électrodes. Au niveau des matrices FeRAM, la MW peut également être récupérée avec succès, sans bitfail supplémentaire après 10 répétitions.

Pour évaluer les performances des FeCAPs et des matrices FeRAM dans des environnements soumis aux radiations, une irradiation par ions lourds (Au^{26+}) a été effectuée. Sur les FeCAPs non cyclées et cyclées, un effet bénéfique des faibles fluences a été mis en évidence, tandis qu'à des fluences plus élevées, une amorphisation progressive révélée par diffraction des rayons X a conduit à une dégradation des propriétés ferroélectriques. Dans une matrice FeRAM programmée avant irradiation selon un motif en damier, les bitfails de Word Lines (WLs) entières se sont révélés permanents et probablement liés à la défaillance des transistors de périphérie.

Recap

Le quatrième chapitre a porté sur la fiabilité des FeCAPs et des matrices FeRAM, un aspect crucial pour leur évaluation complète. Une étude approfondie de la rétention des données a été réalisée, ainsi que des études sur l'endurance et la résistance aux irradiations d'ions lourds. La fatigue ferroélectrique, l'imprint et l'amorphisation du matériau se sont révélés être des phénomènes critiques qui dégradent l'endurance, la rétention des données et la résistance aux radiations. Cependant, il a été démontré que des stratégies de récupération électriques bien choisies pouvaient améliorer ces performances, notamment pour la fatigue et l'imprint.

Chapitre 5 : Intégration au nœud 28nm

Ce chapitre visait à fournir une première démonstration de l'intégration de condensateurs ferroélectriques dans le Back-End-of-Line (BEoL) de transistors aux nœuds technologiques avancés, ainsi qu'une première validation électrique de l'intégrité des transistors et des dispositifs FeCAP. Un objectif secondaire, mais tout aussi important, était d'obtenir des informations utiles concernant les options de fabrication critiques en vue de déterminer un Process of Reference (PoR) optimal pour l'intégration des matrices FeRAM aux nœuds avancés. À cette fin, le développement étape par étape a été présenté en tenant compte des diverses contraintes de fabrication.

Dans ce chapitre, plusieurs dispositifs ont été fabriqués sur des démonstrateurs à l'échelle de wafers de 300 mm de diamètre. Les base wafers, sur lesquels l'empilement ferroélectrique est intégré, sont fabriqués jusqu'au Metal 5 par STMicroelectronics, avec deux options : soit avec des

transistors FDSOI au nœud de 28 nm dans le Front-End-of-Line (FEoL), soit sur des wafers sans transistors. L'intégration de l'empilement ferroélectrique jusqu'au Metal 6 (suivie par la passivation et les niveaux de piliers) nécessite jusqu'à sept étapes de lithographie et permet la fabrication de condensateurs TiN/HSO ou HZO/TiN qui peuvent être connectés au drain des transistors GO1 ou GO2 FDSOI au nœud de 28 nm. Les premières observations sur l'ensemble de l'empilement révèlent une intégration très propre, et les mesures de diffraction aux rayons X (XRD) mettent en évidence le rôle bénéfique de l'étape de recuit thermique pour cristalliser le matériau à base de HfO_2 dans la phase tétragonale ou orthorhombique.

Sur un premier démonstrateur, seuls des dispositifs individuels (condensateurs 1C et transistors 1T) ont été intégrés. Bien que les dimensions réduites des condensateurs (aire de $0.064\mu m^2$) rendent difficile l'évaluation fine des performances électriques, un comportement ferroélectrique a été observé sur des condensateurs intégrés sur des wafers 300 mm sans transistors. Grâce à une première évaluation sur un ensemble de wafers, il a été conclu que des températures de recuit de 500°C et plus permettent la cristallisation dans la phase orthorhombique polaire, qu'une électrode inférieure en TiN (avec ou sans étape de polissage chimico-méchanique, ou CMP) est nécessaire pour générer de la ferroélectricité, et que des couches de 10 nm et 8 nm sont ferroélectriques dans les conditions mentionnées. L'intégrité des transistors GO1 et GO2 du FEoL au nœud de 28 nm a ensuite été évaluée après l'intégration des FeCAPs dans le BEoL : les courants on, off et la tension de seuil ont été déterminés comme conformes aux spécifications du fondeur, validant ainsi les performances des transistors. Sur ce premier démonstrateur, la fonctionnalité de la cellule 1T-1C a également été démontrée, avec des tensions supérieures aux tensions d'overdrive qui permettent de préserver l'intégrité et la fiabilité des transistors. Des couches ferroélectriques plus fines, avec des tensions coercitives plus faibles, sont donc nécessaires pour permettre une commutation saturée à basse tension.

Sur un second démonstrateur, des champs de condensateurs ferroélectriques à échelle réduite ont été intégrés sur des wafers 300 mm pour permettre une analyse fine du comportement ferroélectrique, ainsi que la détermination d'un PoR optimal. Il a été démontré que les performances ferroélectriques des FeCAPs étaient très similaires à celles de leurs homologues sur des wafers 200 mm, avec une polarisation rémanente médiane autour de $20\mu C/cm^2$ obtenue sur un empilement TiN/HZO (10nm)/TiN. Le procédé optimal a donc été déterminé comme étant une couche HZO de 10 nm, avec une étape de CMP après le dépôt du TiN BE et un recuit de cristallisation post-M6. Le wafer présentant une couche HZO de 6 nm n'était pas ferroélectrique, et doit être optimisée pour le fonctionnement des matrices FeRAM aux nœuds avancés, et donc avec une tension maximale limitée.

Recap

Le cinquième chapitre a présenté une première tentative d'intégration des FeCAPs à échelle réduite dans le BEoL des transistors au nœud de 28 nm sur des wafers 300 mm. Un flux d'intégration complet a été proposé et réalisé, permettant la fabrication de dispositifs ferroélectriques fonctionnels tout en maintenant l'intégrité des transistors FEoL. Cela a permis de déterminer les conditions de procédé favorables pour l'intégration des matrices FeRAM à des nœuds de 28 nm ou moins.

Conclusions et perspectives

Pour conclure cette thèse, une partie finale résume les résultats obtenus et explore les perspectives de ce travail en vue de l'intégration de matrices FeRAM performantes et fiables avec une capacité mémoire étendue et une empreinte réduite pour des applications de mémoire ultra-basse consommation. Les évaluations des contraintes d'intégration et les démonstrateurs réalisés prouvent qu'il n'existe aucun obstacle majeur empêchant l'intégration réussie de matrices FeRAM à haute densité et haute capacité mémoire aux nœuds technologiques avancés. Les points faibles en matière de fiabilité, tels que la fatigue et l'imprint, peuvent être corrigés par des séquences électriques spécifiques de programmation/lecture, améliorant ainsi la rétention et l'endurance des matrices. Les derniers résultats de Micron et Sony, mentionnés dans des travaux récents, confirment que l'intégration de FeRAM aux nœuds avancés est sur la bonne voie, et ces développements ouvrent la voie à des applications concrètes dans des circuits pertinents. À l'avenir, l'optimisation de l'empilement ferroélectrique, combinée à des efforts d'intégration et de design, sera essentielle pour le développement de matrices FeRAM à haute performance, haute capacité mémoire et haute densité. Les travaux dédiés à des applications spécifiques, comme ceux orientés vers des SRAM ou DRAM non-volatiles ou les circuits pour l'IA, illustrent le potentiel de ces technologies pour répondre aux défis futurs de la microélectronique.

Bibliography

1. WSTS Semiconductor Market Forecast Spring 2024. 2024 (cit. on p. 24)
2. Bertolazzi S. Emerging Non-Volatile Memory: A 2024 Market Update. Leti Innovation Days. Grenoble, France, 2024 (cit. on pp. 24, 28, 30)
3. Léquepeys JR, Duranton M, Bonnetier S, et al. Overcoming the data deluge challenges with greener electronics. *2021 IEEE 51st European Solid-State Device Research Conference (ESSDERC)*. IEEE. 2021 :7–14 (cit. on pp. 24, 26)
4. Johnson L. Biden ends slog on semiconductor bill with signature. Politico 2022 (cit. on p. 24)
5. Taiwan's dominance of the chip industry makes it more important. The Economist 2023 (cit. on p. 24)
6. Swanson A. Global Trade Grows but Remains Vulnerable to War and Geopolitics. The New York Times 2024 (cit. on p. 24)
7. Taiwan Risk Report 2024. Fitch Solutions 2024 (cit. on p. 24)
8. Zhang K. How Water Scarcity Threatens Taiwan's Semiconductor Industry. The Diplomat 2024 (cit. on p. 24)
9. European Chips Act. European Commission (cit. on p. 24)
10. Towards energy-saving chips for digital, analog and RF. Fames Pilot Line (cit. on p. 24)
11. CEA-Leti Announces Launch of FAMES Pilot Line As Part of EU Chips Act Initiative. 2024 (cit. on p. 24)
12. Saito H, Sugimachi T, Nakamura K, et al. A Triple-Protection Structured COB FRAM with 1.2-V Operation and 10^{17}-Cycle Endurance. *2015 IEEE International Memory Workshop (IMW)*. IEEE. 2015 :1–4 (cit. on pp. 24, 143)
13. 'Tsunami of data' could consume one fifth of global electricity by 2025. Climate Home News, part of the Guardian Environment Network 2017 (cit. on p. 26)
14. Koot M and Wijnhoven F. Usage impact on data center electricity needs: A system dynamic forecasting model. Applied Energy 2021; 291:116798 (cit. on p. 26)
15. Ishimaru K. Future of non-volatile memory-from storage to computing. *2019 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2019 :1–3 (cit. on p. 26)
16. Natarajan S. The Future of Compute: Accelerating Moore's Law Innovation. Leti Innovation Days. Grenoble, France, 2024 (cit. on pp. 26, 27)
17. Reinsel D, Gantz J, and Rydning J. The Digitization of the World: From Edge to Core. IDC White Paper 2018 (cit. on p. 26)

18. Molas G and Nowak E. Advances in emerging memory technologies: From data storage to artificial intelligence. Applied Sciences 2021; 11:11254 (cit. on pp. 26–29)
19. SRC Decadal Plan Update. IEA World Energy Outlook 2022 (cit. on pp. 26, 27)
20. Microcontroller Market Monitor 2024. Yole Intelligence 2024 (cit. on p. 27)
21. Strenz R. Review and outlook on embedded NVM technologies - from evolution to revolution. *2020 IEEE International Memory Workshop (IMW)*. IEEE. 2020 :1–4 (cit. on p. 27)
22. Seidel K, Lehninger D, Müller F, et al. Hafnium oxide-based ferroelectric memories: Are we ready for application? *2023 IEEE International Memory Workshop (IMW)*. IEEE. 2023 :1–4 (cit. on pp. 27, 28)
23. Keshavarzi A, Ni K, Van Den Hoek W, et al. Ferroelectronics for edge intelligence. IEEE Micro 2020; 40:33–48 (cit. on p. 27)
24. Burr GW and Franzon P. Storage class memory. Emerging Nanoelectronic Devices 2014 :498–510 (cit. on p. 28)
25. Tsutsui A. Redefine Memory-System Interactions of SCM (Short Course). *2017 IEEE International Memory Workshop (IMW)*. IEEE. 2017 :125–61 (cit. on p. 28)
26. Backus J. Can programming be liberated from the von Neumann style? A functional style and its algebra of programs. Communications of the ACM 1978; 21:613–41 (cit. on p. 28)
27. Choe J. Memory technology 2021: Trends & challenges. *2021 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*. IEEE. 2021 :111–5 (cit. on p. 28)
28. Navarro G, Verdy A, Castellani N, et al. Innovative PCM+ OTS device with high sub-threshold non-linearity for non-switching reading operations and higher endurance performance. *2017 IEEE Symposium on VLSI Technology*. IEEE. 2017 :T94–T95 (cit. on p. 28)
29. Cheng H, Wu J, Cheek R, et al. A thermally robust phase change memory by engineering the Ge/N concentration in $(Ge, N)_xSb_yTe_z$ phase change material. *2012 IEEE International Electron Devices Meeting*. IEEE. 2012 :31–1 (cit. on p. 28)
30. Navarro G, Bourgeois G, Kluge J, et al. Phase-change memory: Performance, roles and challenges. *2018 IEEE International Memory Workshop (IMW)*. IEEE. 2018 :1–4 (cit. on p. 28)
31. Nielsen P and Bashara N. The reversible voltage-induced initial resistance in the negative resistance sandwich structure. IEEE Transactions on Electron Devices 1964; 11:243–4 (cit. on p. 28)
32. Hickmott T. Low-frequency negative resistance in thin anodic oxide films. Journal of Applied Physics 1962; 33:2669–82 (cit. on p. 28)
33. Baek I, Lee M, Seo S, et al. Highly scalable nonvolatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses. *2004 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2004 :587–90 (cit. on p. 28)
34. Tillie L, Nowak E, Sousa R, et al. Data retention extraction methodology for perpendicular STT-MRAM. *2016 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2016 :27–3 (cit. on p. 28)

35. Shih YC, Lee CF, Chang YA, et al. A reflow-capable, embedded 8Mb STT-MRAM macro with 9ns read access time in 16nm FinFET logic CMOS process. *2020 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2020 :11–4 (cit. on p. 28)
36. Edelstein D, Rizzolo M, Sil D, et al. A 14 nm embedded stt-mram cmos technology. *2020 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2020 :11–5 (cit. on p. 28)
37. Buck DA. Ferroelectrics for digital information storage and switching. PhD thesis. Massachusetts Institute of Technology, Department of Electrical Engineering, 1952 (cit. on pp. 28, 31)
38. Bondurant D. Ferroelectronic RAM memory family for critical data storage. Ferroelectrics 1990; 112:273–82 (cit. on pp. 28, 32)
39. Mikolajick T, Schroeder U, and Slesazeck S. The past, the present, and the future of ferroelectric memories. IEEE Transactions on Electron Devices 2020; 67:1434–43 (cit. on p. 28)
40. Rodriguez J, Remack K, Boku K, et al. Reliability Properties of Low-Voltage Ferroelectric Capacitors and Memory Arrays. IEEE Transactions on Device and Materials Reliability. 2004 Sep; 4:436–49 (cit. on pp. 30, 107, 109, 112, 115)
41. Tallis B. Intel Launches Optane Memory M.2 Cache SSDs for Consumer Market. ANandTech 2017 (cit. on p. 30)
42. Hady FT, Foong A, Veal B, et al. Platform storage performance with 3D XPoint technology. Proceedings of the IEEE 2017; 105:1822–33 (cit. on p. 30)
43. Micron Updates Data Center Portfolio Strategy to Address Growing Opportunity for Memory and Storage Hierarchy Innovation. Micron 2021 (cit. on p. 30)
44. Mann T. Why Intel killed its Optane memory business. The Register 2022 (cit. on p. 30)
45. Arnaud F, Zuliani P, Reynard J, et al. Truly innovative 28nm FDSOI technology for automotive micro-controller applications embedding 16MB phase change memory. *2018 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2018 :18–4 (cit. on p. 30)
46. Arnaud F, Haendler S, Clerc S, et al. 28nm FDSOI Platform with Embedded PCM for IoT, ULP, Digital, Analog, Automotive and others Applications. *2019 IEEE 45th European Solid State Circuits Conference (ESSCIRC)*. IEEE. 2019 :7–10 (cit. on p. 30)
47. Grossi A, Coppetta M, Aresu S, et al. 28nm Data Memory with Embedded RRAM Technology in Automotive Microcontrollers. *2023 IEEE International Memory Workshop (IMW)*. IEEE. 2023 :1–4 (cit. on p. 30)
48. Lee K, Bak J, Kim Y, et al. 1Gbit high density embedded STT-MRAM in 28nm FDSOI technology. *2019 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2019 :2–2 (cit. on p. 30)
49. Naik V, Lee K, Yamane K, et al. Manufacturable 22nm FD-SOI embedded MRAM technology for industrial-grade MCU and IOT applications. *2019 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2019 :2–3 (cit. on p. 30)
50. Alzate J, Arslan U, Bai P, et al. 2 MB array-level demonstration of STT-MRAM process and performance towards L4 cache applications. *2019 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2019 :2–4 (cit. on p. 30)

51. Aggarwal S, Almasi H, DeHerrera M, et al. Demonstration of a reliable 1 Gb standalone spin-transfer torque MRAM for industrial applications. *2019 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2019 :2–1 (cit. on p. 30)
52. Palomino A, Marty J, Auffret S, et al. Evaluating critical metals contained in spintronic memory with a particular focus on Pt substitution for improved sustainability. Sustainable Materials and Technologies 2021; 28:e00270 (cit. on p. 30)
53. Godefroy G. Ferroélectricité. Ed. Techniques Ingénieur, 1996 (cit. on p. 31)
54. Mikolajick T. Basics, History and conventional FeRAM (Short Course). *2018 IEEE International Memory Workshop (IMW)*. IEEE, 2017 (cit. on p. 31)
55. Valasek J. Piezo-electric and allied phenomena in Rochelle salt. Physical review 1921; 17:475 (cit. on p. 31)
56. Chandra P and Littlewood PB. A Landau Primer for Ferroelectrics. 2006 Sep 14 (cit. on pp. 31, 47–49)
57. Böscke T, Müller J, Bräuhaus D, et al. Ferroelectricity in hafnium oxide thin films. Applied Physics Letters 2011; 99 (cit. on pp. 33, 73, 75, 143)
58. Sang X, Grimley ED, Schenk T, et al. On the structural origins of ferroelectricity in HfO_2 thin films. Applied Physics Letters 2015; 106 (cit. on p. 33)
59. Wei Y, Nukala P, Salverda M, et al. A rhombohedral ferroelectric phase in epitaxially strained $Hf_{0.5}Zr_{0.5}O_2$ thin films. Nature materials 2018; 17:1095–100 (cit. on p. 33)
60. Mikolajick T, Schroeder U, and Park MH. Special topic on ferroelectricity in hafnium oxide: Materials and devices. Applied Physics Letters 2021; 118 (cit. on p. 33)
61. Batra R, Tran HD, and Ramprasad R. Stabilization of metastable phases in hafnia owing to surface energy effects. Applied Physics Letters 2016; 108 (cit. on p. 33)
62. Batra R, Huan TD, Rossetti Jr GA, et al. Dopants promoting ferroelectricity in hafnia: Insights from a comprehensive chemical space exploration. Chemistry of Materials 2017; 29:9102–9 (cit. on p. 33)
63. Materlik R, Künneth C, and Kersch A. The origin of ferroelectricity in $Hf_{1-x}Zr_xO_2$: A computational investigation and a surface energy model. Journal of Applied Physics 2015; 117 (cit. on p. 33)
64. Schröder U. HfO_2 based FeRAM and NVDRAM (Short Course). *2018 IEEE International Memory Workshop (IMW)*. IEEE, 2017 (cit. on p. 34)
65. Fina I, Fàbrega L, Langenberg E, et al. Nonferroelectric contributions to the hysteresis cycles in manganite thin films: A comparative study of measurement techniques. Journal of Applied Physics. 2011 Apr; 109:074105 (cit. on pp. 36, 54)
66. Francois T, Coignus J, Grenouillet L, et al. Ferroelectric HfO_2 for Memory Applications: Impact of Si Doping Technique and Bias Pulse Engineering on Switching Performance. *2019 IEEE 11th International Memory Workshop (IMW)*. IEEE, 2019 May :1–4 (cit. on pp. 37, 73)

67. Francois T, Grenouillet L, Coignus J, et al. Demonstration of BEOL-compatible ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ scaled FeRAM co-integrated with 130nm CMOS for embedded NVM applications. *2019 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2019 :15–7 (cit. on pp. 37, 41, 42, 155)

68. Francois T, Coignus J, Makosiej A, et al. 16kbit HfO_2: Si-based 1T-1C FeRAM arrays demonstrating high performance operation and solder reflow compatibility. *2021 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2021 :33–1 (cit. on pp. 39, 41, 42, 60, 79, 91, 97, 98, 111, 118)

69. François T. Caractérisation électrique et analyse de mémoires non-volatiles embarquées à base de matériaux ferroélectriques. PhD thesis. Aix-Marseille Université, 2022 (cit. on pp. 41, 96, 99)

70. Okuno J, Kunihiro T, Konishi K, et al. SoC Compatible 1T1C FeRAM Memory Array Based on Ferroelectric $Hf_{0.5}Zr_{0.5}O_2$. *2020 IEEE Symposium on VLSI Technology*. IEEE, 2020 Jun :1–2 (cit. on pp. 41, 42, 81)

71. Okuno J, Kunihiro T, Konishi K, et al. High-Endurance and Low-Voltage operation of 1T1C FeRAM Arrays for Nonvolatile Memory Application. *2021 IEEE International Memory Workshop (IMW)*. IEEE, 2021 May :1–3 (cit. on pp. 41, 42, 123)

72. Pešić M, Künneth C, Hoffmann M, et al. A computational study of hafnia-based ferroelectric memories: From ab initio via physical modeling to circuit models of ferroelectric device. Journal of computational Electronics 2017; 16:1236–56 (cit. on pp. 48, 49)

73. Dragosits K. Modeling and simulation of ferroelectric devices. PhD thesis. Technische Universität Wien, 2000 (cit. on pp. 47, 48)

74. Devonshire AF. Theory of ferroelectrics. Advances in Physics. 1954; 3. Publisher: Taylor & Francis:85–130 (cit. on p. 47)

75. Sentaurus™ Device User Guide (cit. on p. 49)

76. Mayergoyz I. Mathematical models of hysteresis. IEEE Transactions on Magnetics. 1986; 22:603–8 (cit. on pp. 50, 52)

77. Preisach F. Über die magnetische Nachwirkung. Zeitschrift für Physik. 1935 May 1; 94:277–302 (cit. on p. 49)

78. Krasnoselskii MA and Pokrovskii AV. Systems with hysteresis. Nauka, Moscow, 1983 (cit. on p. 50)

79. Meyer V, Sallese JM, Fazan P, et al. Modeling the polarization in ferroelectric materials: a novel analytical approach. Solid-State Electronics 2003; 47:1479–86 (cit. on p. 51)

80. Miller S, Schwank J, Nasby R, et al. Modeling ferroelectric capacitor switching with asymmetric nonperiodic input signals and arbitrary initial conditions. Journal of Applied Physics 1991; 70:2849–60 (cit. on pp. 51, 52)

81. Bo Jiang, Zurcher, Jones, et al. Computationally Efficient Ferroelectric Capacitor Model For Circuit Simulation. *1997 IEEE Symposium on VLSI Technology*. IEEE, 1997 :141–2 (cit. on pp. 51, 52, 56, 60)

82. Alessandri C, Pandey P, and Seabaugh A. Experimentally validated, predictive Monte Carlo modeling of ferroelectric dynamics and variability. *2018 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2018 :16–2 (cit. on p. 53)

83. Tung CT, Pahwa G, Salahuddin S, et al. A compact model of polycrystalline ferroelectric capacitor. IEEE Transactions on Electron Devices 2021; 68:5311–4 (cit. on p. 53)

84. Tagantsev AK, Stolichnov I, Setter N, et al. Non-Kolmogorov-Avrami switching kinetics in ferroelectric thin films. Physical Review B 2002; 66:214109 (cit. on p. 53)

85. Mulaosmanovic H, Ocker J, Müller S, et al. Switching kinetics in nanoscale hafnium oxide based ferroelectric field-effect transistors. ACS Applied Materials & Interfaces 2017; 9:3792–8 (cit. on pp. 53, 155)

86. Laguerre J, Bocquet M, Billoint O, et al. Memory Window in Si:HfO_2 FeRAM arrays: Performance Improvement and Extrapolation at Advanced Nodes. *2023 IEEE International Memory Workshop (IMW)*. IEEE. 2023 :1–4 (cit. on pp. 54, 95, 117, 143)

87. Francois T, Coignus J, Makosiej A, et al. High-performance operation and solder reflow compatibility in BEOL-integrated 16-kb HfO_2: Si-based 1T-1C FeRAM arrays. IEEE Transactions on Electron Devices 2022; 69:2108–14 (cit. on pp. 54, 79)

88. Fan Z, Xiao J, Wang J, et al. Ferroelectricity and ferroelectric resistive switching in sputtered $Hf_{0.5}Zr_{0.5}O_2$ thin films. Applied Physics Letters. 2016 Jun 6; 108:232905 (cit. on p. 55)

89. Planes N, Weber O, Barral V, et al. 28nm FDSOI technology platform for high-speed low-voltage digital applications. *2012 IEEE Symposium on VLSI Technology*. IEEE. 2012 :133–4 (cit. on pp. 63, 143)

90. Toprasertpong K, Tahara K, Hikosaka Y, et al. Low operating voltage, improved breakdown tolerance, and high endurance in $Hf_{0.5}Zr_{0.5}O_2$ ferroelectric capacitors achieved by thickness scaling down to 4 nm for embedded ferroelectric memory. ACS Applied Materials & Interfaces 2022; 14:51137–48 (cit. on pp. 65, 76, 88, 89, 114, 122, 123, 144, 150, 155, 156)

91. Okuno J, Kunihiro T, Yonai T, et al. A highly reliable 1.8 V 1 Mb $Hf_{0.5}Zr_{0.5}O_2$-based 1T1C FeRAM Array with 3-D Capacitors. *2023 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2023 :1–4 (cit. on pp. 66, 68, 81, 86, 162)

92. Ramaswamy N, Calderoni A, Zahurak J, et al. NVDRAM: A 32Gb dual layer 3D stacked non-volatile ferroelectric memory with near-DRAM performance for demanding AI workloads. *2023 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2023 :1–4 (cit. on pp. 66, 86, 162)

93. Yang J, Luo Q, Xue X, et al. A 9Mb HZO-Based Embedded FeRAM with 10^{12}-Cycle Endurance and 5/7ns Read/Write using ECC-Assisted Data Refresh and Offset-Canceled Sense Amplifier. *2023 IEEE International Solid- State Circuits Conference (ISSCC)*. IEEE, 2023 Feb 19:1–3 (cit. on p. 68)

94. Polakowski P, Riedel S, Weinreich W, et al. Ferroelectric deep trench capacitors based on Al:HfO_2 for 3D nonvolatile memory applications. *2014 IEEE 6th International Memory Workshop (IMW)*. IEEE. 2014 :1–4 (cit. on pp. 69, 85, 86)

95. Haratipour N, Chang SC, Shivaraman S, et al. Hafnia-based FeRAM: a path toward ultra-high density for next-generation high-speed embedded memory. *2022 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2022 :6–7 (cit. on p. 69)

96. Richter C, Schenk T, Park MH, et al. Si Doped Hafnium Oxide—A "Fragile" Ferroelectric System. Advanced Electronic Materials. 2017 Oct; 3:1700131 (cit. on p. 73)

97. Park MH, Schenk T, Fancher CM, et al. A comprehensive study on the structural evolution of HfO_2 thin films doped with various dopants. Journal of Materials Chemistry C. 2017; 5:4677–90 (cit. on p. 73)

98. Francois T, Grenouillet L, Coignus J, et al. Impact of area scaling on the ferroelectric properties of back-end of line compatible $Hf_{0.5}Zr_{0.5}O_2$ and Si:HfO_2-based MFM capacitors. Applied Physics Letters. 2021 Feb 8; 118:062904 (cit. on pp. 73, 123)

99. Sünbül A, Lehninger D, Hoffmann R, et al. Improved Endurance Reliability of Ferroelectric Hafnium Oxide-Based BEoL Integrated MFM Capacitors. *2024 IEEE International Reliability Physics Symposium (IRPS)*. IEEE. 2024 :1–5 (cit. on pp. 73, 157, 162)

100. Mittmann T, Materano M, Chang SC, et al. Impact of Oxygen Vacancy Content in Ferroelectric HZO films on the Device Performance. *2020 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2020 Dec 12:18.4.1–18.4.4 (cit. on pp. 73, 76)

101. Hamouda W, Mehmood F, Mikolajick T, et al. Oxygen vacancy concentration as a function of cycling and polarization state in TiN/$Hf_{0.5}Zr_{0.5}O_2$/TiN ferroelectric capacitors studied by x-ray photoemission electron microscopy. Applied Physics Letters 2022; 120 (cit. on pp. 73, 77)

102. Lederer M, Abdulazhanov S, Olivo R, et al. Electric field-induced crystallization of ferroelectric hafnium zirconium oxide. Scientific Reports - Nature 2021; 11:22266 (cit. on p. 73)

103. Segantini G, Manchon B, Cañero Infante I, et al. Interplay between Strain and Defects at the Interfaces of Ultra-Thin $Hf_{0.5}Zr_{0.5}O_2$-Based Ferroelectric Capacitors. Advanced Electronic Materials. 2023 Oct; 9:2300171 (cit. on pp. 73, 157)

104. Alcala R, Materano M, Lomenzo PD, et al. The Electrode-Ferroelectric Interface as the Primary Constraint on Endurance and Retention in HZO-Based Ferroelectric Capacitors. Advanced Functional Materials 2023; 33:2303261 (cit. on pp. 73, 77)

105. Huang F, Saini B, Wan L, et al. First Observation of Ultra-high Polarization (~ 108 $\mu C/cm^2$) in Nanometer Scaled High Performance Ferroelectric HZO Capacitors with Mo Electrodes. *2023 IEEE Symposium on VLSI Technology and Circuits*. IEEE. 2023 :1–2 (cit. on p. 73)

106. Joh H, Jung T, and Jeon S. Stress Engineering as a Strategy to Achieve High Ferroelectricity in Thick Hafnia Using Interlayer. IEEE Transactions on Electron Devices. 2021 May; 68:2538–42 (cit. on p. 73)

107. Kao YC, Peng HK, Wang YK, et al. Toward Highly Pure Ferroelectric $Hf_{1-x}Zr_xO_2$ Thin Films by Tailoring the Strain in an Unstable Thermodynamic System. ACS Applied Electronic Materials. 2022 Aug 23; 4:3897–908 (cit. on pp. 73, 75)

108. Fields SS, Cai T, Jaszewski ST, et al. Origin of Ferroelectric Phase Stabilization via the Clamping Effect in Ferroelectric Hafnium Zirconium Oxide Thin Films. Advanced Electronic Materials. 2022 Dec; 8:2200601 (cit. on pp. 73, 75)

109. Xu B, Lomenzo PD, Kersch A, et al. Strain as a Global Factor in Stabilizing the Ferroelectric Properties of ZrO $_2$. Advanced Functional Materials. 2023 Nov 10 :2311825 (cit. on pp. 73, 75)

110. Mueller S, Muller J, Schroeder U, et al. Reliability Characteristics of Ferroelectric Si:HfO_2 Thin Films for Memory Applications. IEEE Transactions on Device and Materials Reliability. 2013 Mar; 13:93–7 (cit. on pp. 73, 109, 111–113)

111. Grenouillet L, Francois T, Coignus J, et al. Nanosecond Laser Anneal (NLA) for Si-Implanted HfO_2 Ferroelectric Memories Integrated in Back-End of Line (BEOL). *2020 IEEE Symposium on VLSI Technology*. IEEE, 2020 Jun :1–2 (cit. on pp. 73, 90)

112. Lee JY, Hsiang KY, Liao CY, et al. Nanosecond Laser Annealing Based Wake-up of Ferroelectric $HfZrO_2$ Capacitors for BEOL Compatible and High Throughput FeRAM. 2022 International Conference on Solid State Devices and Materials (SSDM). The Japan Society of Applied Physics, 2022 Sep 27 (cit. on p. 73)

113. Kaiser N, Song YJ, Vogel T, et al. Crystal and electronic structure of oxygen vacancy stabilized rhombohedral hafnium oxide. ACS Applied Electronic Materials 2023; 5:754–63 (cit. on p. 77)

114. Vishnumurthy P, Alcala R, Mikolajick T, et al. Ferroelectric HfO_2-based Capacitors for FeRAM: Reliability from Field Cycling Endurance to Retention. *2024 IEEE International Reliability Physics Symposium (IRPS)*. IEEE. 2024 :1–10 (cit. on pp. 77, 109, 114)

115. Yuan P, Mao GQ, Cheng Y, et al. Microscopic mechanism of imprint in hafnium oxide-based ferroelectrics. Nano Research. 2022 Apr; 15:3667–74 (cit. on pp. 77, 113–115, 119)

116. Grenouillet L, Martin S, Laguerre J, et al. Synergetic FeRAM/OxRAM memory array operation induced by oxygen vacancy interface engineering in Si:Hf_2 ferroelectric films. *2023 International Conference on Solid State Devices and Materials (SSDM)*. 2023 :1–2 (cit. on p. 77)

117. Barrett N, Hamouda W, Lubin C, et al. Oxygen vacancy engineering in Si-doped, HfO_2 ferroelectric capacitors using Ti oxygen scavenging layers. Applied Physics Letters 2024; 125 (cit. on p. 77)

118. Billoint O, Martin S, Laguerre J, et al. Charge-based Sense Demonstration in 1T-1C HZO FeRAM Arrays to Overcome CBL-induced Bank Size Limitations. *2024 IEEE International Memory Workshop (IMW)*. IEEE. 2024 :1–4 (cit. on pp. 82, 84)

119. Chang SC, Haratipour N, Shivaraman S, et al. Anti-ferroelectric $Hf_xZr_{1-x}O_2$ capacitors for high-density 3-D embedded-DRAM. *2020 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2020 :28–1 (cit. on pp. 86, 155)

120. Sung M, Rho K, Kim J, et al. Low Voltage and High Speed 1Xnm 1T1C FE-RAM with Ultra-Thin 5nm HZO. *2021 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2021 Dec 11:33.3.1–33.3.4 (cit. on p. 86)

121. Tahara K, Toprasertpong K, Hikosaka Y, et al. Strategy toward HZO BEOL-FeRAM with low-voltage operation (< 1.2 V), low process temperature, and high endurance by thickness scaling. *2021 IEEE Symposium on VLSI Technology*. IEEE. 2021 :1–2 (cit. on p. 88)

122. Grenouillet L, Barbot J, Laguerre J, et al. Reliability assessment of hafnia-based ferroelectric devices and arrays for memory and AI applications. *2023 IEEE International Reliability Physics Symposium (IRPS)*. IEEE. 2023 :1–8 (cit. on pp. 88, 123)

123. Okuno J, Kunihiro T, Konishi K, et al. 1T1C FeRAM memory array based on ferroelectric HZO with capacitor under bitline. IEEE Journal of the Electron Devices Society. 2021 :1–1 (cit. on p. 91)

124. Okuno J, Yonai T, Kunihiro T, et al. Demonstration of fatigue and recovery phenomena in $Hf_{0.5}Zr_{0.5}O_2$-based 1T1C FeRAM memory arrays. *2022 6th IEEE Electron Devices Technology & Manufacturing Conference (EDTM)*. IEEE. 2022 :64–6 (cit. on pp. 105, 121, 122, 132, 133)

125. Mulaosmanovic H, Lomenzo PD, Schroeder U, et al. Reliability aspects of ferroelectric hafnium oxide for application in non-volatile memories. *2021 IEEE International Reliability Physics Symposium (IRPS)*. IEEE, 2021 Mar :1–6 (cit. on pp. 107, 109)

126. Rodriguez J, Remack K, Gertas J, et al. Reliability of Ferroelectric Random Access memory embedded within 130nm CMOS. *2010 IEEE International Reliability Physics Symposium (IRPS)*. IEEE, 2010 :750–8 (cit. on pp. 109, 110)

127. Alcala R, Materano M, Lomenzo PD, et al. BEOL Integrated Ferroelectric HfO_2-Based Capacitors for FeRAM: Extrapolation of Reliability Performance to Use Conditions. IEEE Journal of the Electron Devices Society. 2022; 10:907–12 (cit. on pp. 111, 123)

128. Schroeder U, Mittmann T, Materano M, et al. Temperature-Dependent Phase Transitions in $Hf_xZr_{1-x}O_2$ Mixed Oxides: Indications of a Proper Ferroelectric Material. Advanced Electronic Materials 2022; 8:2200265 (cit. on p. 112)

129. Higashi Y, Kaczer B, Verhulst A, et al. Investigation of imprint in FE-HfO_2 and its recovery. IEEE Transactions on Electron Devices 2020; 67:4911–7 (cit. on pp. 113, 114, 119)

130. Alcala R, Lomenzo P, Mittmann T, et al. The Role of Interface Dynamics on the Reliability Performance of BEOL Integrated Ferroelectric HfO_2 Capacitors. *2022 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2022 Dec 3:32.8.1–32.8.4 (cit. on p. 114)

131. Mikheev V, Kondratyuk E, and Chouprik A. Retention Model and Express Retention Test of Ferroelectric HfO_2-Based Memory. Physical Review Applied. 2022 Dec 27; 18:064084 (cit. on p. 114)

132. Sünbül A, Lehninger D, Lederer M, et al. A Study on Imprint Behavior of Ferroelectric Hafnium Oxide Caused by High-Temperature Annealing. Physica Status Solidi (a). 2023 Apr; 220:2300067 (cit. on pp. 115, 119)

133. Li Y, Zhu H, Liu X, et al. Trap characteristics of hafnium oxide-based ferroelectric field-effect transistors measured by using a current transient method. Applied Physics Letters. 2023 Mar 13; 122:112905 (cit. on p. 115)

134. Okuno J, Yonai T, Kunihiro T, et al. Investigation of Recovery Phenomena in $Hf_{0.5}Zr_{0.5}O_2$-Based 1T1C FeRAM. IEEE Journal of the Electron Devices Society. 2023; 11:43–6 (cit. on pp. 121, 124, 126, 128, 129)

135. Chen J, Qian H, Zhang H, et al. Physical Origin of Recovable Ferroelectric Fatigue and Recovery for Doped-HfO_2: Toward Endurance Immunity. *2023 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2023 Dec 9:1–4 (cit. on pp. 121, 124, 126, 129)

136. Liao P, Chang Y, Lee YH, et al. Characterization of fatigue and its recovery behavior in ferroelectric HfZrO. *2021 IEEE Symposium on VLSI Technology*. IEEE. 2021 :1–2 (cit. on pp. 122, 129)

137. Schenk T, Hoffmann M, Ocker J, et al. Complex Internal Bias Fields in Ferroelectric Hafnium Oxide. ACS Applied Materials & Interfaces. 2015 Sep 16; 7:20224–33 (cit. on pp. 129, 130)

138. Li S, Zhou D, Shi Z, et al. Involvement of unsaturated switching in the endurance cycling of Si-doped HfO_2 ferroelectric thin films. Advanced Electronic Materials 2020; 6:2000264 (cit. on p. 129)

139. Scott JF, Araujo CA, Meadows HB, et al. Radiation effects on ferroelectric thin-film memories: Retention failure mechanisms. Journal of Applied Physics. 1989 Aug 1; 66:1444–53 (cit. on p. 133)

140. Benyagoub A. Mechanism of the monoclinic-to-tetragonal phase transition induced in zirconia and hafnia by swift heavy ions. Physical Review B—Condensed Matter and Materials Physics 2005; 72:094114 (cit. on p. 134)

141. Vogel T, Kaiser N, Petzold S, et al. Defect-Induced Phase Transition in Hafnium Oxide Thin Films: Comparing Heavy Ion Irradiation and Oxygen-Engineering Effects. IEEE Transactions on Nuclear Science. 2021 Aug; 68:1542–7 (cit. on p. 137)

142. Dünkel S, Trentzsch M, Richter R, et al. A FeFET based super-low-power ultra-fast embedded NVM technology for 22nm FDSOI and beyond. *2017 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2017 :19–7 (cit. on p. 155)

143. Hoffmann M, Murdzek JA, George SM, et al. Atomic layer etching of ferroelectric hafnium zirconium oxide thin films enables giant tunneling electroresistance. Applied Physics Letters. 2022 Mar 21; 120:122901 (cit. on p. 157)

144. Ettisserry D, Visconti A, Bonanomi M, et al. Comprehensive Reliability Assessment of 32Gb (Hf,Zr)O_2 -Based Ferroelectric NVDRAM. *2024 IEEE International Reliability Physics Symposium (IRPS)*. IEEE, 2024 Apr 14:1–8 (cit. on p. 162)

145. Chavan A, Rajagopal A, Yan Y, et al. Materials Engineering for High Performance Ferroelectric Memory. *2024 IEEE International Memory Workshop (IMW)*. IEEE. 2024 :1–4 (cit. on p. 162)

146. Migita S, Ota H, Asanuma S, et al. Accelerated ferroelectric phase transformation in HfO_2/ZrO_2 nanolaminates. Applied Physics Express 2021; 14:051006 (cit. on p. 162)

147. Rhetat L, Noel JP, Giraud B, et al. A ferroelectric based non volatile SRAM optimized for critical embedded systems. *2024 International Conference on Very Large Scale Integration (VLSI-SoC)*. IFIP/IEEE. 2024 :1–6 (cit. on p. 162)

148. Shuto Y, Okuno J, Yonai T, et al. HZO-based Nonvolatile SRAM Array with 100% Bit Recall Yield and Sufficient Retention Time at 85° C. *2024 IEEE Symposium on VLSI Technology and Circuits*. IEEE. 2024 :1–2 (cit. on p. 162)

149. Martemucci M, Rummens F, Hirtzlin T, et al. Hybrid FeRAM/RRAM Synaptic Circuit Enabling On-Chip Inference and Learning at the Edge. *2023 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2023 :1–4 (cit. on p. 162)

150. Hur J, Luo YC, Lu A, et al. Nonvolatile Capacitive Crossbar Array for In-Memory Computing. Advanced Intelligent Systems 2022; 4:2100258 (cit. on p. 162)

151. Mukherjee S, Bizindavyi J, Luo YC, et al. Pulse-Based Capacitive Memory Window with High Non-Destructive Read Endurance in Fully BEOL Compatible Ferroelectric Capacitors. *2023 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2023 Dec 9:1–4 (cit. on p. 162)