



Electrical and physicochemical characterization of metal gate processes for work function modulation and reduction of local V_{TH} variability in 14FDSOI technologies

Carlos Augusto Suarez Segovia

► To cite this version:

Carlos Augusto Suarez Segovia. Electrical and physicochemical characterization of metal gate processes for work function modulation and reduction of local V_{TH} variability in 14FDSOI technologies. Micro and nanotechnologies/Microelectronics. Université Grenoble Alpes, 2016. English. NNT : 2016GREAT011 . tel-01288566

HAL Id: tel-01288566

<https://theses.hal.science/tel-01288566>

Submitted on 15 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Nano Electronique et Nano Technologies**

Arrêté ministériel : 7 août 2006

Présentée par

Carlos Augusto SUAREZ SEGOVIA

Thèse dirigée par **Gerard GHIBAUDO**

et codirigée par **Charles LEROUX** et **Florian DOMENGIE**

préparée au sein de la société **STMicroelectronics**, du laboratoire **CEA-Leti (Grenoble)** et de l'**IMEP-LAHC**
et de l'**Ecole Doctorale EEATS**

Electrical and physicochemical characterization of metal gate processes for work function modulation and reduction of local V_{TH} variability in 14FDSOI technologies

Thèse soutenue publiquement le **4 février 2016**,
devant le jury composé de :

Mme. Catherine DUBOURDIEU

DR, CNRS Ecole Centrale de Lyon, Président

Mme. Nathalie MALBERT

PR, Université de Bordeaux, Rapporteur

M. Brice GAUTIER

PR, INSA Lyon, Rapporteur

M. Florian DOMENGIE

ING, STMicroelectronics, Examineur

M. Charles LEROUX

ING, CEA-LETI, Examineur

M. Gerard GHIBAUDO

DR, CNRS Alpes, Directeur de thèse



Acknowledgments

While this thesis is written in English, I feel nice to write most of these acknowledgements in French and Spanish, as it will enforce my thanks.

J'aimerais tout d'abord remercier sincèrement mon encadrant de thèse, Charles Leroux, de m'avoir toujours guidé et poussé pendant ces 3 années de thèse. J'ai particulièrement apprécié ta rigueur scientifique et toutes les discussions que nous avons pu avoir. Charles, grâce à toi j'ai beaucoup appris sur les techniques d'analyse capacitive et de photoémission interne et sur les techniques d'extraction des paramètres électriques. Merci pour la confiance que tu m'as accordé et pour le regard critique et objectif que tu as porté sur mes résultats. Cela m'a aidé à m'améliorer, à me remettre en question quand cela était nécessaire et à mieux choisir mes expériences.

Je tiens à remercier Pierre Caubet et Florian Domengie, qui a pris le relais de l'encadrement industriel chez STMicroelectronics Crolles, suite au changement de poste de Pierre. Merci à vous deux de m'avoir formé sur les équipements de dépôt et caractérisation de la grille métallique en salle blanche et de m'avoir appris plein d'astuces pour l'organisation de mes expériences et pour la gestion de mes lots. Merci pour vos conseils, bonnes idées et coups de main essentiels pour la réussite de cette thèse.

Merci également à Gérard Ghibaudo de l'IMEP-LAHC d'avoir bien voulu diriger cette thèse. Merci pour ta disponibilité dans la gestion administrative de mon travail, ton accueil et ton soutien notamment dans la rédaction de mon manuscrit et dans les différentes conférences internationales.

Je tiens à saluer la qualité de votre encadrement, votre patience et votre disponibilité, ce fut très agréable de travailler avec vous au quotidien pendant ces 3 années de thèse.

Je tiens à remercier également les membres du jury, Mme. Catherine Dubourdiou, Mme. Nathalie Malbert, ainsi que M. Brice Gautier d'avoir bien voulu évaluer mes travaux. Merci de m'avoir fait l'honneur de votre présence à ma soutenance de thèse et d'avoir bien voulu faire le déplacement jusqu'à Grenoble.

Je remercie les responsables hiérarchiques des différentes structures dans lesquelles j'ai travaillé. D'une part, j'ai été accueilli au sein de l'unité Process Development chez STMicroelectronics Crolles menée par Jean Galvier, et plus particulièrement l'équipe métal R&D, dirigée par Maxime Mellier, à qui je remercie d'avoir participé à la valorisation de mon travail au sein de l'entreprise

et ses conseils précieux pour la présentation de mes résultats. Merci aussi à Olivier Hinsinger de m'avoir donné les ressources nécessaires à mon travail. Je n'ai jamais été en manque de silicium !

D'autre part, je tiens à remercier Gilles Reimbold de m'avoir accueilli dans son équipe de caractérisation et tests électriques du CEA-LETI-Minatec. Merci à Jacques Cluzel et Denis Blachier, maîtres des SIAM et du banc IPE, respectivement, de m'avoir formé sur ces bancs de test électrique et de mesures de photoémission interne. Merci à Giovanni Romano, Alain Toffoli, Rabah Kies et Fabienne Allain, qui méritent d'être cités pour l'aide technique qu'ils m'ont apporté pour la mise en place de mes mesures.

Mes expériences n'auraient pu être menées à bien sans la participation de plusieurs experts process R&D de ST Crolles. Je remercie tout particulièrement à l'ensemble des membres du module "Gate Stack" mené par Stéphane Zoll. Merci donc à Vincent Joseph, en charge de graver par solution chimique chaque empilement de grille sacrificielle de cette thèse. Merci également à Olivier Gourhant, en charge de la réalisation des différents oxydes et recuits thermiques pour mes études de diffusion à différentes températures. Merci aussi à Virginie Beugin pour les dépôts d'oxyde high-k et les échanges instructifs. J'ai aussi bénéficié de l'expertise device de l'éminent Olivier Weber et de l'aide précieuse de Nicolas Degors et Stéphane Martin pour l'embarquement de mes structures de test. Merci aussi à Jean Philippe Manceau et Mustapha Rafik de l'équipe de caractérisation électrique de ST d'avoir partagé avec moi leurs premiers échantillons. Merci aussi à Nelly Guillot pour son aide vitale pour débloquer mes lots et faire avancer mes manips exotiques.

Une pensée particulière pour Karen Dabertrand de l'équipe de métrologie et pour Marc Juhel de l'équipe de caractérisation physique de ST pour les échanges instructifs sur les avantages et limitations des équipements XRF et SIMS, respectivement. Je n'oublie pas la collaboration avec la plateforme de nanocaractérisation du CEA-Leti Minatec pour la caractérisation physicochimique de mes échantillons, en particulier avec Patrick Gergaud et Frédéric Fillot en XRD et avec Eugénie Martinez en XPS.

Je voudrais également remercier tous les membres de l'équipe de caractérisation électrique du LETI pour les échanges intéressants que nous avons eus et pour votre bonne humeur : Xavier Garros, Mikaël Cassé, William van den Daele, Carlo Cogli et Jean Coignus. Je remercie également Stéphane Becu, Alain Lopez et Matthieu Nongaillard. Enfin, je remercie également tous les doctorants et stagiaires du labo d'avoir mis une ambiance très positive, en particulier mes collègues de bureau Blend Mohammad, à qui je remercie pour les discussions très productives et tous les bons moments passés ensemble, et

Alexandre Vernet, pour ses interminables explications sur l'évolution de la grammaire française. Merci aussi à Assawer, Jonathan, Guillaume, Johan, Antoine, Remy, Alexandre S., Giulio2, Cheikh, Illias, Clément, Niccolò, Philippe, Genaro, Gauthier, Julien pour les repas et sorties de labo, les matchs de foot, votre sympathie et votre disponibilité.

Merci également à tous les gens de l'équipe métal à ST pour leur convivialité au quotidien, leur bonne humeur contagieuse, leur dynamisme et leurs coups de mains. Je remercie profondément à Magali Grégoire, Sebastien Petitdidier, Katia Haxaire, Yannick Le Friec, Julie Poulet, Nicolas C, Jérôme, Etienne, Thierry, Eric, GG et bien d'autres.

Une pensée particulière aux doctorants de l'IMEP-LAHC qui m'ont chaleureusement accueilli et m'ont permis de rédiger ma thèse dans le calme et la bonne humeur. Et aux doctorants de ST pour la bonne ambiance pendant les repas à midi: Giulio1, Boris, Darayus, Nayera, Hassan, Antoine, etc. Merci aussi à mon pote Omar qui est venu voir ma soutenance à Grenoble.

Merci aussi à mes potes thésards ou pas, qui m'ont accompagné au quotidien au cours de ces dernières années sur Grenoble : Onintza, Bastien, Romu, Mokrane, Lucia, Giulio1, Mathilde et Bruna. Merci pour les sorties ski, les soirées raclettes, les balades en montagne, les soirées salsa et les beaux voyages que nous avons partagés !

Por último, quisiera agradecer a mi familia. A mis padres Carlos Suarez Carrillo y Silvia Segovia, por haberme dado la vida, por todo su apoyo y por haber sido una constante fuente de fortaleza y amor a lo largo de mi vida. Y a mis hermanos Rodrigo y Diego. Sin su apoyo, no estoy seguro de haber llegado hasta aquí.

Contents

Acknowledgments	i
General introduction	1
1 Gate stack technology for 14 nm FDSOI MOSFET devices	5
1.1 Impact of Metal-Oxide-Semiconductor gate stack properties on MOSFET device performance	6
1.1.1 MOSFET operation	6
1.1.2 Metal-Oxide-Semiconductor gate stack properties	9
1.1.3 MOSFET performance parameters influenced by the gate stack	18
1.2 Gate stack fabrication process	23
1.2.1 Oxidation methods and Gate dielectrics	23
1.2.2 Metal gate electrode and deposition techniques	27
1.2.2.1 Thin metal gate films deposition techniques	28
1.2.2.2 Deposition of ultra-thin Titanium nitride films and Aluminum and Lanthanum monolayers by RF-PVD magnetron sputtering	34
1.3 Metal gate integration in 14 nm Fully-Depleted SOI devices	36
1.3.1 FDSOI architecture	36
1.3.1.1 Multi- V_{TH} offer	37
1.3.1.2 Effective work function requirements for Fully-Depleted SOI devices	38
1.3.2 Sacrificial metal gate-first process integration	39
1.3.3 Random local V_{TH} fluctuations	41
1.4 Conclusion	44
2 Process flow and test methodology for electrical and physicochemical characterization of gate stack	45
2.1 Techniques of electrical characterization	46
2.1.1 Test structures	46
2.1.1.1 Hybrid devices	48
2.1.1.2 Short channel test structures	49
2.1.1.3 Matching test structures	54
2.1.2 Process flow simplification for nominal devices and specific research wafers	56
2.1.2.1 Nominal devices	56
2.1.2.2 Devices with beveled oxide	58
2.1.3 Measurements and electrical parameters extraction	62
2.1.3.1 Experimental set up for C-V measurements	62

2.1.3.2	Extraction of electrical parameters from C-V characteristics	64
2.2	Physicochemical characterization	67
2.2.1	Four-point probe resistance measurement	67
2.2.2	X-Ray Fluorescence	68
2.2.2.1	Advantages and limitations	69
2.2.2.2	Methodology for the characterization of the diffusion of gate additives	70
2.2.3	Secondary Ion Mass Spectrometry	71
2.2.4	X-Ray Diffraction	72
2.3	Conclusion	80
3	Effective work function modulation by accurate control of diffusion of <i>sacrificial</i> lanthanum and aluminum into gate stack of high-κ based NFET devices	81
3.1	State of the art of lanthanum in gate stack	82
3.1.1	Lanthanum oxide as alternate gate dielectrics	82
3.1.2	Lanthanum incorporation into Hf-based dielectrics	83
3.1.3	Lanthanum incorporation into TiN	84
3.2	Effective work function modulation by accurate control of <i>sacrificial</i> lanthanum diffusion	86
3.2.1	Device fabrication and electrical characterization	86
3.2.2	Diffusion characterization by X-Ray Fluorescence	91
3.2.3	Influence of high- κ dielectrics on the effective work function shift induced by lanthanum incorporation	97
3.3	State of the art of aluminum in gate stack	102
3.3.1	Aluminum oxide as alternative gate dielectrics	103
3.3.2	Aluminum incorporation into Hf-based gate dielectrics	103
3.3.3	Aluminum addition into TiN	105
3.4	Effective work function modulation by accurate control of <i>sacrificial</i> aluminum diffusion	107
3.4.1	Device fabrication and electrical characterization	107
3.4.2	Diffusion characterization by X-Ray Spectroscopy	113
3.4.3	Influence of high- κ dielectrics on the effective work function shift induced by aluminum incorporation	121
3.5	Sacrificial vs standard final approach	124
3.6	Conclusion	127
4	Role of TiN on effective work function and matching of 14 nm FDSOI devices	129
4.1	State of the art of the modification of the effective work function induced by the TiN	130

4.1.1	Modulation of the effective work function by tuning TiN thickness	130
4.1.2	Modulation of the effective work function by tuning N composition in TiN gate	132
4.1.3	Effect of the oxygen incorporation in TiN gate on the effective work function	132
4.1.4	Effect of the reduction of oxygen vacancy defects on effective work function	133
4.2	Characterization of the impact of TiN on the effective work function of 14 nm FDSOI devices	134
4.2.1	Impact of nitrogen content into TiN on the effective work function	134
4.2.2	Influence of high- κ dielectrics on the work function shift induced by TiN thickness	140
4.2.3	Titanium and nitrogen diffusion into HfSiON/SiON and HfON/SiON stacks	145
4.2.4	Influence of Arsenic doping into Poly-Si layer on the effective work function	147
4.3	Impact of deposition process on TiN microstructure and local V_{TH} variability	149
4.3.1	Impact of pressure conditions on microstructure of TiN deposited by RF-PVD	152
4.3.2	Impact of RF power conditions on microstructure of TiN deposited by RF-PVD	156
4.3.3	Electrical impact and local V_{TH} variability improvement with TiN microstructure engineering	160
4.3.4	TiN microstructure versus thickness	164
4.4	Conclusion	166
Conclusions and Perspectives		169
Bibliography		175

General introduction

Context

Digital electronics are omnipresent across the globe today, enriching people's lives and making communication and sharing easier than ever. This revolution is the consequence of the miniaturization of electronic systems over the past few decades. At the heart of each of these digital devices are semiconductor chips that provide the intelligence and power to drive the device. Each chip is made up with billions of transistors. The simultaneous manufacturing of these billions of transistors on a single wafer made possible the creation of integrated circuits performing complex logic operations. To build better digital devices and enhance the user experience, the size of the transistors must be reduced while increasing performance and reducing power consumption. This trend was first predicted by Intel co-founder Gordon Moore, who observed that the number of components in integrated circuits had doubled every two years from the invention of the integrated circuit in 1958, by Jack Kilby and Robert Noyce, until 1965 [1]. From the beginning of mass production, the number of transistors that the industry was able to place on a chip did double every two years. In every new generation, the down-scaling of the horizontal along with the vertical dimensions of devices allows the increase of productivity and performances [2]. This prediction, now known as the Moore's law, is responsible for the evolution of today's complementary metal-oxide-semiconductor (CMOS) technology. This law is now used in the semiconductor industry to guide long-term planning and to set targets for research and development. In the 1960's the integrated circuits market was broadly based on Bipolar Junction Transistors (BJT) due to their high switching speed and low power consumption at smaller sizes. But since 1975, integration of Field-Effect-Transistor (FET), invented by Dawon Kahng and Martin Atalla in 1959, dominated even though they were found to be slower switching devices than the bipolar transistors [3]. This was caused by the failure of BJTs to demonstrate rapid decrease in the power per circuit compared to FETs. As linear dimensions reached the half-micron level in the early 1990s, the performance advantage of bipolar transistors was overcome by FETs, then used in the CMOS circuit production until today.

For the fabrication of logic gates, most advanced integrated circuits use CMOS technology. CMOS uses both n-channel and p-channel MOSFET devices to achieve a high speed with low power dissipation. The aggressive scaling of CMOS has marked milestones in the evolution of the gate stack at STMicroelectronics. First, with the introduction of the couple high- κ metal gate (HKMG) stack that has been integrated in the conventional SiO_2 /Poly-Si gate stack to continue the down-scaling of CMOS transistors while keeping overall performances. The introduction of a high- κ dielectric is indeed key in order to

keep low gate leakage current and enhance the gate stack capacitance density, thereby increasing the transistor ON-current (I_{ON}), which leads to an increase of the circuit speed. In addition, the introduction of metal gate electrodes are required to be included in gate stack in order to lower gate resistance and eliminate poly silicon depletion and boron penetration effects, for improved performance and reliability [4], but also because of thermal instability of high- κ dielectrics in contact with Poly-Si above temperatures which can exceed 1000 °C [5]. Second, with the integration of an ultra-thin buried oxide (BOX) for V_{TH} variability improvement, back-bias efficiency and better short channel immunity compared to bulk technologies. Besides, this buried oxide layer eliminates the need to add dopants to the channel, thus making it fully depleted. This technology is called Fully-Depleted Silicon-On-Insulator (FDSOI) technology.

In this context, the V_{TH} for CMOS logic applications is mainly tuned through the work function of the metal gate, and not anymore by Poly-Si or channel doping. Metal gate electrodes with the correct work functions are therefore required for high performance CMOS logic applications. These requirements are not similar in FDSOI and bulk devices, as will be detailed in this thesis work. Furthermore, one technical challenge in metal electrode research is that the metal work function is process-dependent and depends on the electrical properties of the underlying high- κ film [6], leading to an effective work function. The gate-first integration has been chosen for the integration of the metal gate stack in FDSOI devices manufactured at STMicroelectronics. In a gate-first approach, the metal electrode is deposited before the source drain dopant activation annealing at more than 1000 °C. In a gate-last approach, the metal electrode is thus deposited after the source drain dopant activation annealing. The choice of a gate-first integration limits the choice of metal electrode materials and the tuning of V_{TH} becomes more complex. On the other hand, a gate-last approach significantly increases the process complexity, the manufacturing cost, and the design rules are more restricted. On the other hand, although the main source of local V_{TH} variability related to the Random Dopant Fluctuations (RDF) in the channel has been suppressed in undoped channel FDSOI devices, some simulations predict that the Metal Gate Granularity (MGG) becomes more important in scaled devices and must be investigated.

Objectives of this Ph. D. thesis

At the beginning of this Ph.D., the 14 nm FDSOI product was in progress of development and was scheduled to be released by end 2015. The couple TiN/HfON had been retained in 14 nm FDSOI manufacturing process for their thermal stability and reduced gate leakage currents. Based on the state of the

art, it was known that the presence of lanthanides in the dielectric shifts the effective work function towards the N+ values and the presence of alumina towards the P+ values. Nevertheless, a clear understanding of the diffusion of such additives and a methodology to provide a relationship between this diffusion and their impact on the effective work function was lacking. On the other hand, the optimal metal gate deposition conditions to face the local V_{TH} variability induced by the metal granularity were unknown.

The main objectives of this thesis are, on one hand, the understanding of the modulation of the effective work function through the incorporation of additives such as aluminum, lanthanum and nitrogen following a *sacrificial* metal gate-first approach, and on the other hand, the study of the impact of the TiN microstructure on the V_{TH} variability of adjacent MOSFET pairs placed the one from the other at minimal spacing authorized by the design rules. The goal is to validate the sacrificial gate-first integration to successfully achieve the threshold voltage requirements of the 14FDSOI technology and to propose metal deposition conditions to further reduce the local V_{TH} variability in advanced FDSOI nodes.

Outline

The first chapter introduces the gate stack technology as well as the current process challenges in gate stack integration with the aggressive scaling of Fully-Depleted Silicon on Insulator (FDSOI) MOSFET devices. The gate stack properties and their correlation with the main MOSFET performance parameters are reviewed in this chapter. The methods for dielectric growth as well as the metal deposition techniques are also described in this chapter. In particular, the deposition by Radio Frequency Physical Vapor Deposition of titanium nitride, lanthanum and aluminum layers is detailed. These metals are currently used in the development of the 14 nm FDSOI devices. We also describe in this chapter the Fully-Depleted Silicon on Insulator (FDSOI) architecture and its capability to offer multi threshold voltage (V_{TH}) solutions. We define the requirements of the effective work function of the metal for such devices and the interest of the sacrificial metal gate-first approach as a process scheme able to achieve this target by precisely controlling the amount of additives incorporated into the gate stack. Finally, the need to investigate the microstructure of metal gates is introduced.

The second chapter introduces the techniques of electrical and physicochemical characterization used in this thesis. The implementation of specially designed test structures and a simplified process flow for metal gates investigation. The Capacitance-Voltage (C-V) analysis techniques employed to characterize the modulation of the effective work function with the additives

incorporation into gate stack and to identify its physical origin are described in this chapter. Furthermore, a new methodology based on X-Ray Fluorescence is presented and validated in order to precisely characterize the percentage of effective diffusion of aluminum and lanthanum into high- κ /bottom oxide stack. Finally, the two X-Ray diffraction configurations used to fully study the microstructure of TiN layers are detailed.

The third chapter focuses on the impact of the incorporation of lanthanum (La) and aluminum (Al) in a *sacrificial* gate-first approach on the effective work function of 14 nm FDSOI devices. This chapter covers the study of the process variations used to incorporate La in a *sacrificial* gate approach in NFET devices, the electrical measurements and the diffusion characterization of lanthanum into the HfON/SiON stack by the methodology based on X-Ray Fluorescence introduced in the second chapter. The influence of high- κ dielectrics on the effective work function shift induced by lanthanum incorporation is also studied. Similarly, the study of the process variations used to incorporate aluminum following a *sacrificial* gate-first approach in NFET devices as well as the electrical characterization are described in this chapter. Likewise, the diffusion characterization of aluminum into the HfON/SiON stack and the influence of high- κ dielectrics on the effective work function shift induced by aluminum incorporation are also presented and discussed in this chapter. Finally, the benefit of the sacrificial gate-first approach compared to a standard gate-first approach is discussed, as well as the perspectives for future research.

The final chapter focuses on the impact of nitrogen, adjusted by the thickness and composition of TiN layers, on the effective work function of FDSOI devices. Afterwards, the incorporation of nitrogen into gate stack is characterized by the physicochemical measurements. Finally, an innovative process for metal deposition by RF-PVD at room temperature, allowing the modification of the microstructure of TiN, is proposed in order to further reduce the local V_{TH} variability in FDSOI devices.

This CIFRE Ph.D. thesis work has been carried out in collaboration with STMicroelectronics in Crolles, the CEA-LETI and the IMEP-LAHC laboratories in Grenoble. All the devices and blanket wafers were manufactured at STMicroelectronics in an industrial environment to better propose solutions to STMicroelectronics challenges in terms of V_{TH} modulation as well as of the reduction of V_{TH} variabilities for advanced FDSOI technologies. Electrical characterizations were performed in the Laboratoire de Caractérisations et Test Electrique (LCTE) of the CEA-LETI.

Gate stack technology for 14 nm FDSOI MOSFET devices

Contents

1.1	Impact of Metal-Oxide-Semiconductor gate stack properties on MOSFET device performance	6
1.2	Gate stack fabrication process	23
1.3	Metal gate integration in 14 nm Fully-Depleted SOI devices	36
1.4	Conclusion	44

This chapter introduces the gate stack technology as well as the current process challenges in gate stack integration with the aggressive scaling of Fully-Depleted Silicon on Insulator (FDSOI) MOSFET devices.

The gate stack properties and their correlation with the main MOSFET performance parameters are reviewed in the section 1.1 of this chapter. The section 1.2 extensively describes the methods for dielectric growth as well as the metal deposition techniques. In particular, the deposition by Radio Frequency Physical Vapor Deposition of titanium nitride, lanthanum and aluminum layers is detailed. These metals are currently used in the development of the 14 nm FDSOI devices.

The section 1.3 of this chapter presents the Fully-Depleted Silicon on Insulator (FDSOI) architecture and its capability to offer multi threshold voltage (V_{TH}) solutions. In this context, the requirements for the effective work function of the metal are defined. The sacrificial metal gate-first approach is described as a process scheme able to achieve this target by precisely controlling the amount of additives incorporated into the gate stack.

1.1 Impact of Metal-Oxide-Semiconductor gate stack properties on MOSFET device performance

1.1.1 MOSFET operation

The Metal-Oxide-Semiconductor Field Effect Transistor (MOSFET) device is based on the field effect mechanism [7]. In physics, the field effect refers to the modulation of the electrical conductivity of a material by the application of an external electric field. In MOSFETs, this field effect is then related to the control of the semiconductor conductivity near its surface.

A MOSFET is a four-terminal device, but it can be seen as a combination of two orthogonal two-terminal devices, as shown in Fig. 1.1. The objective is to control the current flow between two ohmic contacts, the *source* and the *drain*, by the modulation of the charges in the semiconductor channel. Charge carriers in the semiconductor channel are controlled by two other terminals, the *gate* and the *body* or *substrate*, by capacitive coupling¹. The threshold voltage (V_{TH}) is the minimum gate-to-body voltage difference that is needed to create a conductive channel between the source and drain terminals. It is the conductive channel that allows the carriers to flow from the source to the drain.

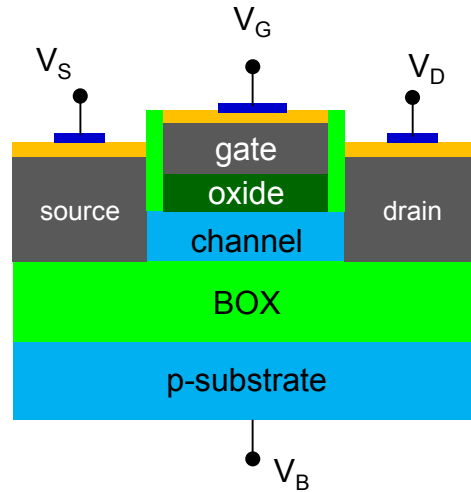


Figure 1.1: Schematic diagram of a four terminal n-channel MOSFET device.

The simplest MOS structure is the MOS capacitor and it is composed of a p-type or n-type doped silicon, a gate oxide and a layer of metal or polycrystalline

¹In electronics, coupling is the transfer of energy from a medium to another medium.

silicon. By applying Kirchhoff's voltage law, the potential difference between the gate and the body terminals (V_G) can therefore be decomposed as follows :

$$V_G = \Phi_m - \Phi_{sc} + V_{ox} + V_{sc} \quad (1.1)$$

where $q\Phi_m$ corresponds to the difference between the metal Fermi level and its vacuum level, $q\Phi_{sc}$ corresponds to the difference between the semiconductor Fermi level and its vacuum level, V_{ox} is the drop voltage across the oxide and V_{sc} is the potential drop in the semiconductor.

The relationship between the charges in the semiconductor Q_{sc} and the potential at the semiconductor surface (V_{sc}) can be described by the Poisson equation for electrostatics, according to classical electrodynamics [7]. However, for ultra-thin oxide thicknesses and low field strengths, such interactions are better described by quantum Poisson-Schrödinger simulations (section 2.1.3). The charges induced in the semiconductor may be of three types : majority carriers, minority carriers and depletion charge. The type of charges can be controlled by the voltage applied to the gate, specifically by the V_{sc} . Indeed, the ability to induce and modulate a conducting sheet of minority carriers at the semiconductor/oxide interface is the basis of the operation of the MOSFET.

Let's take the example of p-type semiconductor substrate. For an n-type semiconductor, the sign of V_{sc} has to be switched, as shown in Fig. 1.2.

- $V_{sc} < 0$: *Accumulation* regime. Majority carriers are attracted to the surface of the semiconductor.
- $0 < V_{sc} < \Phi_{fi}$: *Depletion* regime. Majority carriers are pushed towards the bulk silicon, and so, their density decreases at the semiconductor/oxide interface. The only remaining charges are the thermally ionized dopant atoms at room temperature.
- $\Phi_{fi} < V_{sc} < 2\Phi_{fi}$: *Weak inversion* or *subthreshold* regime. Minority carriers density starts to increase, but charges are essentially the thermally ionized dopant atoms at room temperature. Transistor is in the "off" state or weak inversion under threshold voltage.
- $2\Phi_{fi} < V_{sc}$: *Strong inversion* regime. The transistor is in the "on" state. Minority carriers density becomes much more important than the doping concentration. The channel is formed between the source and the drain.

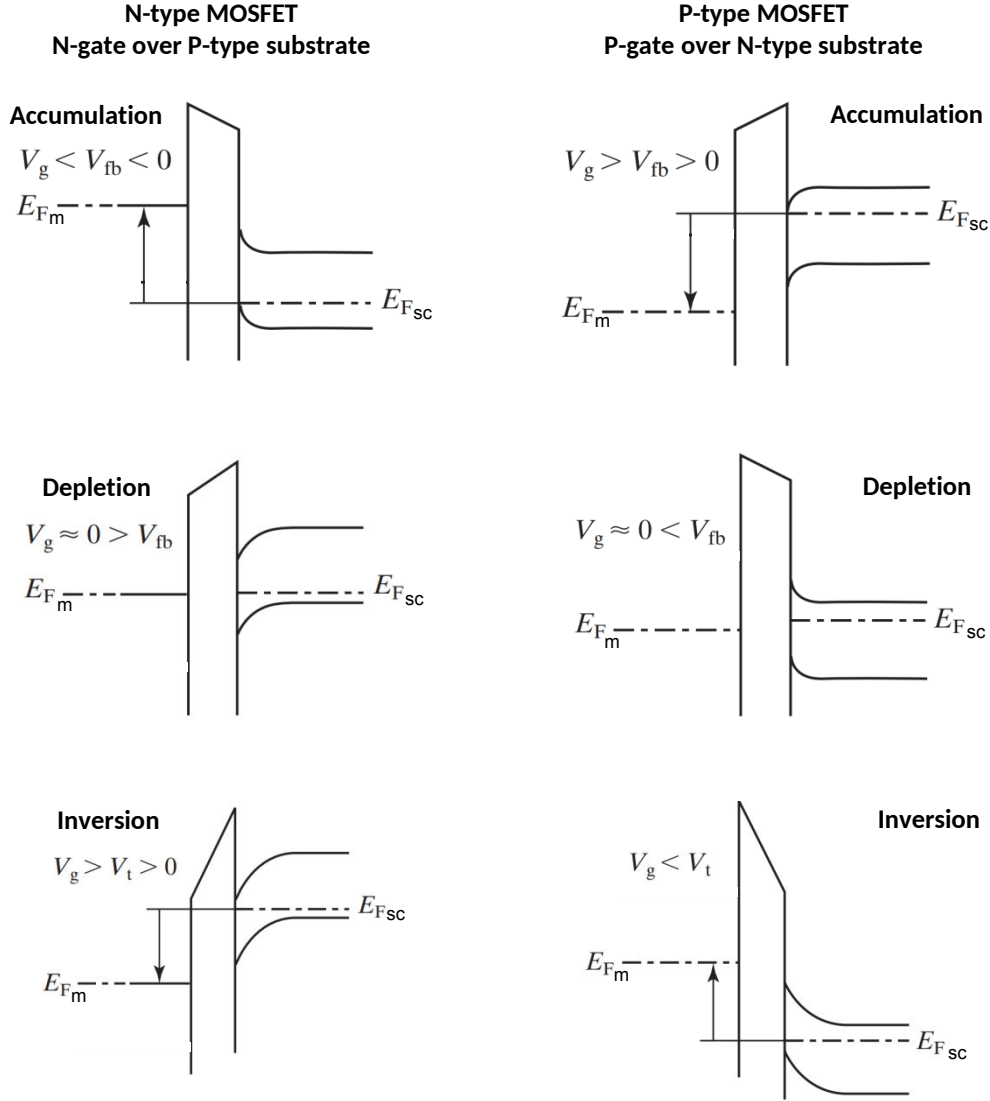


Figure 1.2: Energy band diagrams of the two dominant types of MOS capacitors.

The heavily doped (n+ for NFETs and p+ for PFETs) source/drain regions, are used to make an ohmic contact with the conductive channel for $|V_G| > |V_{TH}|$ so that a voltage difference between the source and the drain (V_{DS}) will result in a current flow (I_{DS}) of minority carriers (electrons for NFETs and holes for PFETs) from the positive voltage at the drain terminal to the negative voltage at the source. This current flow of minority carriers is also known as drive current and is one of the main MOSFET device performance parameters. In the "off" state, the drive current is very small (ideally zero) and in the "on" state, it is a function of both V_G and V_{DS} .

1.1.2 Metal-Oxide-Semiconductor gate stack properties

Gate dielectric capacitance and equivalent oxide thickness

The capacitance is the ability of a body to store an electrical charge. The most common form of energy storage device is the parallel plate capacitor, which consists of two electrical conductors (plates) separated by a dielectric (i.e. an electrical insulator that can be polarized by the application of an electric field). When a potential difference V is applied to the conductors, an electric field develops across the dielectric, causing positive charge ($+Q$) on one plate and negative charge ($-Q$) on the other plate. The capacitance C is defined as the ratio of charge on each conductor to the voltage V between them ($C = Q/V$).

In a Metal-Oxide-Semiconductor capacitor, one of the plates is the metal gate and the other, is the silicon substrate. However, in the MOS capacitor the applied voltage may be used to control the type of interface charge induced in the channel (majority carriers, minority carriers, and depletion charge). The dielectric consists of the gate oxide with a relative permittivity ϵ_{ox} (commonly known as the dielectric constant) and thickness T_{ox} . Consequently, the gate dielectric capacitance per unit area (C_{ox}) can be expressed as follows :

$$C_{ox} = \frac{\epsilon_0 \epsilon_{ox}}{T_{ox}} \quad (1.2)$$

where $\epsilon_0 = 8.854... \times 10^{-12}$ F/m is the vacuum permittivity.

In order to quickly compare different dielectric materials to the industry standard silicon oxide (SiO_2) dielectric, a new definition has been proposed :

$$C_{ox} = \frac{\epsilon_0 \epsilon_{\text{SiO}_2}}{EOT} \quad (1.3)$$

where EOT (Equivalent Oxide Thickness) is defined as :

$$EOT = T_{ox} \frac{\epsilon_{\text{SiO}_2}}{\epsilon_{ox}} \quad (1.4)$$

Both EOT and C_{ox} are important gate stack properties that influence the performance parameters of MOSFET devices such as the drive current (I_{DS}) and the threshold voltage (V_{TH}), as will be detailed in section 1.1.3.

Flat band voltage

The gate bias (V_G) always corresponds to the potential difference between the metal and semiconductor Fermi levels, $E_{F,m}$ and $E_{F,sc}$, respectively. The energy band diagram of the MOS structure for $V_G = 0$ is quite complex and shows band bending due to the semiconductor and metal work function difference. The flat band voltage (V_{fb}) is a special bias condition which corresponds to the gate bias with no charge in the substrate. In other words, V_{fb} is the bias condition leading

to no band bending at the substrate interface. It corresponds to $Q_{sc} = 0$, and for constant doping level in semiconductor, it usually leads to $V_{sc} = 0$. However, at flat band condition, the term V_{ox} is not necessarily zero [7] [8]. Indeed, it is influenced by oxide stack charges density and interfacial voltage drops [9]. The flat band expression can thus be deduced from Eq. 1.1

$$V_{fb} = \Phi_m - \Phi_{sc} + V_{ox} \quad (1.5)$$

Considering a p-type substrate and an bilayer high- κ /SiO₂ oxide stack with no oxide charges at all, the energy band diagram of the MOS structure at flat band condition is described in Fig. 1.3a. E_0 , E_C , and E_V are the energy of vacuum level, the energy of the bottom of the conduction band, and the energy of the top of the valence band of silicon, respectively.

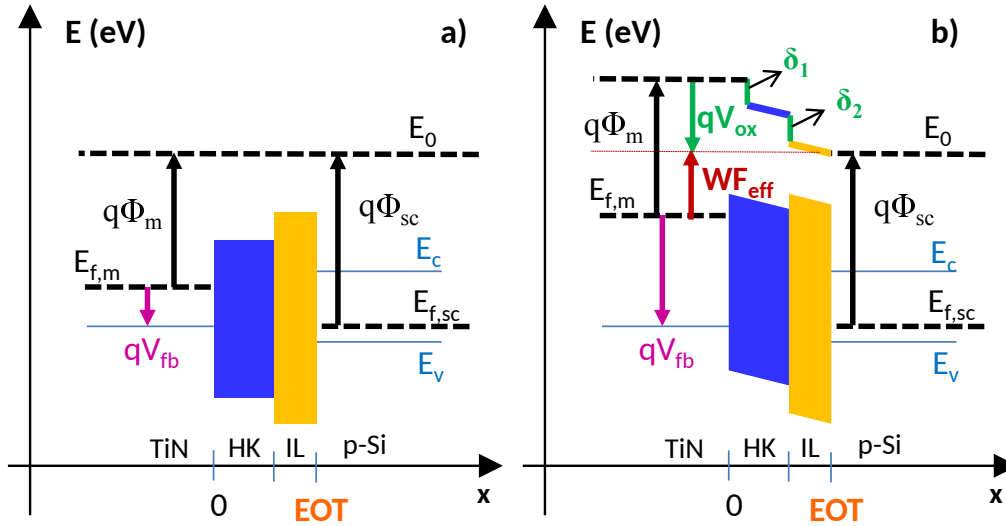


Figure 1.3: Energy band diagram of the MOS structure at the flat band condition a) without any oxide charge and b) with charges in the oxide and dipoles at high- κ interfaces.

However, V_{fb} for real devices is a process-dependent parameter and its control is currently one of the most serious problems, especially in bilayer high- κ /SiO_x oxide structures, currently used in 14 nm FDSOI technology and other modern MOSFET technologies. Gate stack charges density and interfacial voltage drops due to the formation of an electrostatic dipole (δ) at both high- κ interfaces have been reported as the origins of the V_{fb} shift. An energy band diagram for such non-ideal MOS structure at flat band condition is shown in Fig. 1.3b [9]. Note that the vacuum level is influenced by the presence of charges and interfacial drops. As a result, processes variations lead to an effective work function (WF_{eff}) different from Φ_m . The effective work function is

of first interest, since it drives V_{TH} . In the next sections, we first describe the gate stack charges and the interface electrostatic dipole and their impact on V_{fb} . Next, the expression of WF_{eff} in the case of a bilayer high- κ /SiO_x oxide structure will be defined, as well as the considerations assumed and validated to simplify the WF_{eff} expression in this thesis work.

Gate stack charges density and its impact on flat band voltage

In real MOSFET devices, four general types of charges can be identified and associated to the Si/SiO₂ system [10] [11] [12]: bulk or interfacial fixed charges (Q_f), mobile ionic charges (Q_m), interface trapped charges (Q_{it}), and oxide trapped charges (Q_{ot}). The origin of these charges has been related to some impurities or defects incorporated into the oxide during oxide growth or subsequent fabrication process steps.

The mobile ionic charges (Q_m) are primarily due to positive alkali ions in the oxide such as Na⁺, K⁺ and Li⁺ [10], incorporated during device processing steps. These impurities cause reliability problems under high temperature and high voltage operations as they can migrate from an interface to another.

The oxide trapped charges (Q_{ot}) may be positive or negative due to holes or electrons trapped in the bulk of the oxide. Trapping may result from ionizing radiation, avalanche injection, or other similar processes. Unlike fixed charge, oxide trapped charge is generally annealed out by low temperature (< 500 °C) treatment, although neutral traps may remain [11].

The fixed oxide charges (Q_f) are positive or negative charges located in the dielectric or at Si/SiO_x interface. In a bilayer high- κ /SiO_x oxide structure, fixed charges can be located at: a) bulk high- κ layer, b) high- κ /SiO_x interface, c) bulk SiO_x layer, d) SiO_x/Si interface, and e) metal/high- κ layer [13]. We name Q_{Si/SiO_x} the sheet of charges at SiO_x/Si interface and $Q_{SiO_x/HK}$ the sheet of charges at high- κ /SiO₂ interface. Fixed oxide charges do not move and exchange charge with the underlying silicon. They also do not change with the applied voltage [11].

The interface trapped charges (Q_{it}) are positive or negative charges located at the Si/SiO_x interface. They are due to structural, oxidation-induced defects, radiation-induced defects or other dangling bonds at this interface. Unlike fixed charge or trapped charge, interface trapped charge is in electrical communication with the underlying silicon and can thus be charged or discharged, depending on the surface potential V_{sc} . Most of the interface trapped charges can be neutralized by low temperature (450 °C) hydrogen annealing or forming gas annealing [11].

The location of these charges is illustrated in Fig. 1.4.

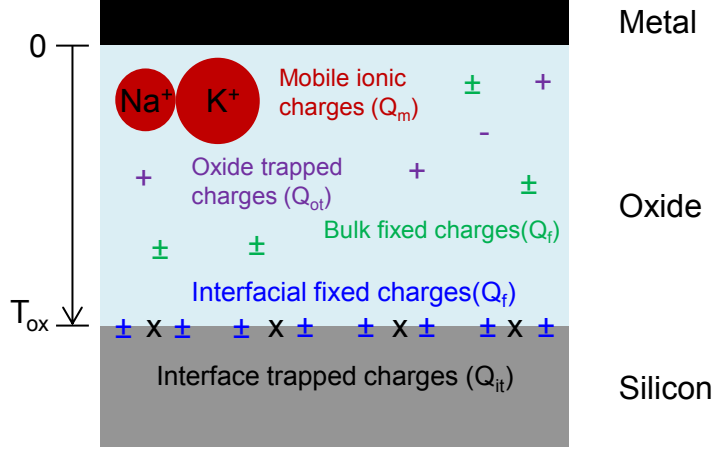


Figure 1.4: Location of charges in thermally oxidized silicon according to [10].

The effect of each charge on the V_{fb} condition depends on its distance from the oxide/semiconductor interface and can be calculated from Gauss Law. This is why it is convenient to separate the charges between bulk charges and interface charges. From Gauss Law application, we can deduce the V_{fb} shift induced by each type of oxide charge: ΔV_{fb}^{bulk} , ΔV_{fb}^f , and ΔV_{fb}^{it} for bulk charges, interface fixed charges, and interface trapped charges, respectively. Bulk charges include the mobile ionic charges, the oxide trapped charges and the bulk fixed charges and their impact on flat band voltage is given by:

$$\Delta V_{fb}^{bulk} = -\frac{1}{\epsilon_0} \int_0^{T_{ox}} \rho(z) \int_0^z \frac{du}{\epsilon_{ox}(u)} dz \quad (1.6)$$

where $\rho(z)$ is the gate dielectric charge distribution per unit volume. For a constant bulk charge density in a dielectric layer, the voltage drop will then vary with the square of its thickness. Interface fixed charges for a bilayer high- κ /SiO_x oxide structure can be expressed as follows:

$$\Delta V_{fb}^f = -\frac{Q_{Si/SiO_x}}{\epsilon_0 \epsilon_{SiO_2}} EOT - \frac{Q_{SiO_x/HK}}{\epsilon_0 \epsilon_{SiO_2}} EOT_{HK} \quad (1.7)$$

A sheet of charges at the high- κ /SiO_x interface or at SiO_x/Si interface induces a voltage drop which increases linearly with the distance from the gate. Finally, interface trapped charges located at the Si/SiO_x interface influences the V_{fb} as follows:

$$\Delta V_{fb}^{it} = -\frac{Q_{it}(V_{sc})}{\epsilon_0 \epsilon_{SiO_2}} EOT \quad (1.8)$$

It is therefore evident that the presence of charges in the oxide changes V_{fb} . As the threshold voltage is related to the V_{fb} , charges also directly influence the threshold voltage and most of the performance parameters of MOSFET devices, which will be described in section 1.1.3. Even worst, charges affect not only threshold voltages but also effective mobilities, and device stability in MOS devices, and also junction leakage, noise, and breakdown voltage in discrete transistors and digital integrated circuits. In addition, defects are undesirable for four reasons. Firstly, charges trapped in defects shift the gate threshold voltage of the transistor. Secondly, the trapped charge changes with time so V_{TH} shifts with time, leading to instability of operating characteristics. Thirdly, trapped charge scatters carriers in the channel and lowers the carrier mobility. Fourthly, defects are the starting point for electrical failure and oxide breakdown [14], and then they cause unreliability.

Interface electrostatic dipole

Discrepancies between WF_{eff} obtained through electrical measurements and Φ_m in high- κ metal gate stacks have also been related to other electrostatic effects. Indeed, although differences in fixed charges could explain some of the differences in the relative shifts, only fixed charges was not sufficiently appropriate to explain a part of the potential drop, which was independent of the thickness of the oxides. Several models have therefore been proposed to explain such discrepancies at various locations in the gate stack, in particular at metal/high- κ interface and at high- κ /SiO_x interface. We present here a synthesis of the most important hypothesis, the interpretations of experimental results, and the latest progress on the understanding of this electrical behaviour.

At first, the presence of metal-induced gap states (MIGS) was proposed at the metal/high- κ interface [15]. Oxygen vacancies in high- κ oxides, leading to charge transfer from oxygen vacancies to the gate electrode [16], was also mentioned as a possible cause of the V_{fb} shift. Indeed, reduction/oxidation ambient annealing experiments were performed to introduce/annihilate the oxygen vacancies of HfO₂ and to investigate the origin of V_{fb} shifts [17]. Positive and negative V_{fb} shifts were observed by switching from annealing in a oxidizing to a reducing atmosphere, respectively. It was stated that the V_{fb} modulation was caused by a change of the oxygen vacancy concentration in the HfO₂ near the metal/HfO₂ contact. The origins of this V_{fb} shift was also related to the electrical effect of a dipole layer at metal/high- κ interface due to the group electronegativity difference between metal gate and Hf in dielectrics [18]. According to this model, the modulation of the capping dielectric material which it is in contact with metal can also modulate the V_{fb} shift. The polarity of these shifts was also consistent with the difference in mean electronegativity for each dielectric relative to that of HfO₂ [19]. The hypothesis of the formation

of an electric dipole at metal/high- κ interface was compatible with the V_{fb} shifts successfully achieved by capping the top surface of the high- κ layer with additives such as aluminum [20] or lanthanum [21], in a conventional gate-first flow with the highest temperature in the range between 985 °C and 1050 °C. At the time, this was considered as the experimental demonstration that the metal/high- κ interface is the key reason for the observed WF_{eff} variations.

Nevertheless, other experiments demonstrated that V_{fb} shift is modulated predominantly by the high- κ /SiO₂ interface, at least in the case of low annealing temperature (500 °C) for which no interdiffusion of Hf and metallic atoms occurs during device fabrication [22], [23].

Once again, the oxygen vacancy formation [24] and a material electronegativity differences [25] at the bottom interface were considered to be the origin of the dipole creation. Some authors also concluded that generated oxygen vacancy forms an additional dipole different from the initially existing dipole at high- κ /SiO₂ interface and therefore the V_{fb} shift is determined by the contribution of two types of dipoles at high- κ /SiO₂ interface [26]. An interesting point is that the interface dipole direction of La₂O₃/SiO₂ or Y₂O₃/SiO₂ is opposite to that of Al₂O₃/SiO₂ or HfO₂/SiO₂, as shown in Fig. 1.5 on the left. In order to explain the direction and magnitude of the V_{fb} shift induced by the dipole formed at high- κ /SiO₂ interface, K. Kita and A. Toriumi proposed a model which takes into account the oxygen behaviour at the interface, since oxygen is the common atom for both materials high- κ and SiO₂ and not the metal cations in high- κ oxides [27]. In this model, the number of oxygen atoms per unit area (σ) for various high- κ oxides and SiO₂ was calculated from the weight and the density of each oxide, extracted from their crystallographic data. The results were normalized and ordered in relation to the ratio of σ/σ_{SiO_2} , as shown in Fig. 1.5 on the right.

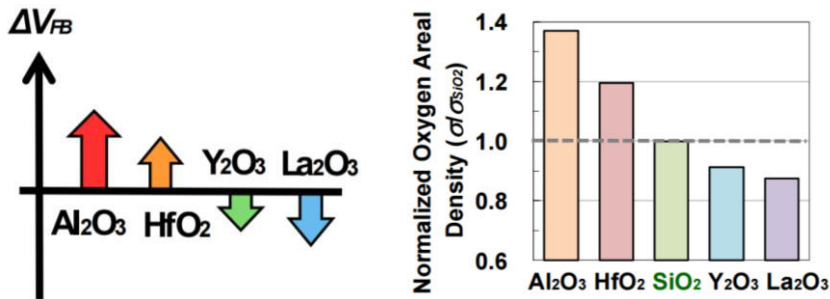


Figure 1.5: On the left, experimentally observed V_{fb} shift due to dipole formed at high- κ /SiO₂ interface. On the right, relative areal oxygen density of various oxides at interfaces normalized by that of SiO₂ calculated using their molar volume [27].

According to this model, the bonds at the interface are strained because of the deformation due to σ difference, leading to an increase of free energy of the interface. The free energy of the interface should be therefore minimized by a movement of oxygen from larger σ side to the smaller σ side. For example, the oxygen is transferred from SiO_2 to high- κ in the cases of Y_2O_3 and La_2O_3 , while it is transferred from high- κ to SiO_2 in the cases of Al_2O_3 and HfO_2 [27]. As a result, a charge imbalance at interface is induced by the movement of oxygen since oxygen is negatively charged. This model explains the formation of an equal amount of charges of opposite sign at high- κ /SiO₂ interface.

On the other hand, ab initio simulations performed recently on interface between HfO_2 and SiO_2 revealed that a substitution of Hf by La and Mg increases the valence band offset (VBO) at this interface and a substitution of Hf by Al decreases the valence band offset (VBO) [28], compared to reference interface, as shown in Fig. 1.6. These results are in agreement with WF_{eff} shifts direction, since VBO shift is equal to the opposite effect of an electrostatic dipole ($-\delta$) [25]. In other words, ab initio simulations demonstrated that interfacial Hf substitution by La and Mg shifts WF_{eff} towards N+ level, whereas Al shifts WF_{eff} towards P+.

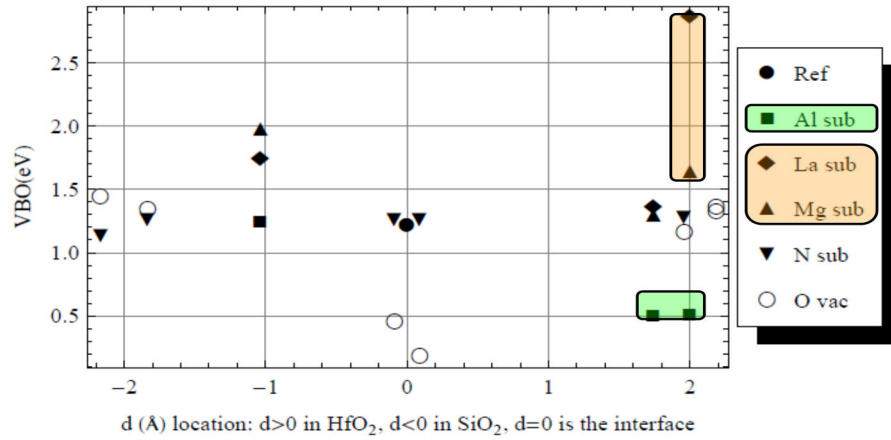


Figure 1.6: Valence band offsets for different additives at the $\text{HfO}_2(001)/\text{SiO}_2(101)$ interface [25].

Finally, it should be noted that the V_{fb} shifts successfully achieved by capping the top surface of the high- κ layer were performed in a conventional gate-first flow with temperatures in the range between 985 °C and 1050 °C. At these temperatures, diffusion of species must be taken into account. These experiments were misunderstood at the time, but the latest models described above clearly show that an energetically favorable dipole layer is formed at high- κ /SiO_x interface. For this reason, a new methodology based on X-Ray

Fluorescence will be proposed and validated in section 2.2.2.2 for accurate in-line characterization of the effective diffusion of species into the oxide gate stack after annealing.

Semiconductor doping density

For an intrinsic semiconductor, the Fermi level is at the midgap. However, for an n-doped semiconductor the Fermi level is above the intrinsic level, and for a p-doped semiconductor it is below the intrinsic level. $q\Phi_{\text{fi}}$ is the difference between the intrinsic Fermi level of silicon ($E_i = 4.61$ eV) and the resulting Fermi level energy after doping ($E_{\text{F,sc}}$) and is given by the Eq. 1.9.

$$E_i - E_{\text{F,sc}} = q\Phi_{\text{fi}} = k_b T \ln\left(\frac{N_{\text{sc}}}{n_i}\right) \quad (1.9)$$

where k_b is the Boltzmann constant, T is temperature in Kelvin, N_{sc} is the substrate doping concentration in cm^{-3} , q is the elementary charge in C and n_i is the intrinsic carrier concentration of silicon in cm^{-3} . Thus, the silicon work function can be expressed as follows:

$$q\Phi_{\text{sc}} = 4.61\text{eV} + q\Phi_{\text{fi}} \quad (1.10)$$

In practice, silicon substrate is doped with boron (for P-type Si) or phosphorus (for N-type Si) impurities by a series of ion implantation processes, followed by an annealing. As a result, the donors (or acceptors) impurities are not evenly distributed into the semiconductor and therefore the doping is not constant through the semiconductor. In Fig. 1.7, the dopant concentration profile below the buried oxide (BOX) of the 14 nm FDSOI technology after PWELL implantations and annealing at 1050 °C for 5 s has been simulated with Synopsis TCAD tool. As Φ_{sc} is a parameter related to the doping substrate level, it will be determined through C-V analysis.

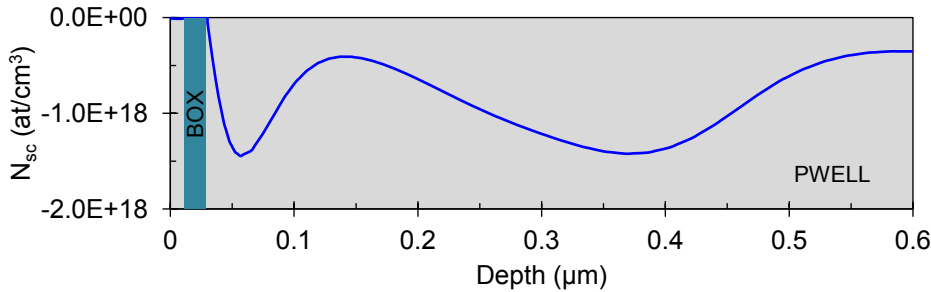


Figure 1.7: Dopant concentration profile below the buried oxide (BOX) of the 14 nm FDSOI technology after PWELL implantations and annealing at 1050 °C for 5 s simulated with Synopsis TCAD tool. *Courtesy of B. Mohamad.*

Metal effective work function

The work function of the metal is defined as the minimum energy needed to remove an electron from the metal to a point in the vacuum immediately outside the metal surface. In a MOS system, however, the metal work function seems to be dependent on the processes, leading to an effective work function. WF_{eff} takes into account not only $q\Phi_m$ but also the potential drop in the dielectric (V_{ox}) at flat band voltage in real MOS devices, as shown in Fig. 1.3b. Indeed, including the vacuum level in the equation of the flat band voltage, WF_{eff} expression is given by the work function of the metal ($q\Phi_m$) plus the impact of charges density and interfacial dipoles (δ) on the potential drop across the oxide (qV_{ox}) (see first part of Eq. 1.11), and it can be obtained by adding $q\Phi_{\text{sc}}$ to V_{fb} (see second part of Eq. 1.11). Indeed, in chapter 2, we will see that V_{fb} and $q\Phi_{\text{sc}}$ can be automatically and accurately determined by the measurement of the capacitive response (CV) of a MOS device, and then WF_{eff} directly obtained by adding $q\Phi_{\text{sc}}$ to V_{fb} .

$$WF_{\text{eff}} = q\Phi_m + qV_{\text{ox}} = qV_{\text{fb}} + q\Phi_{\text{sc}} \quad (1.11)$$

Considering constant values of permittivity along the thickness of the oxide and assuming a uniformly distributed charge per unit volume, it can be proved that V_{fb} shift induced by bulk charges (Eq. 1.6) results in a quadratic modification of the V_{fb} vs EOT plot for a given oxide [13]. On the contrary, V_{fb} shift induced by interfacial fixed charges (Eq. 1.7) results in a linear modification of the V_{fb} vs EOT plot for a given oxide. Now then, it has been experimentally demonstrated that V_{fb} vs EOT plot is described by a straight line for SiO_x and HfO_x oxides [13] [9]. Therefore, by neglecting the impact of bulk charges in the dielectric layers and only considering the presence of interfacial fixed charges and potential drops induced by dipoles at the interfaces (δ), the WF_{eff} should be rewritten as follows [9]:

$$\frac{WF_{\text{eff}}}{q} = \Phi_m - Q_{\text{Si/SiO}_x} \left(\frac{\text{EOT}}{\varepsilon_{\text{SiO}_2}} \right) - Q_{\text{SiO}_x/\text{HK}} \left(\frac{\text{EOT}_{\text{HK}}}{\varepsilon_{\text{SiO}_2}} \right) + \delta \quad (1.12)$$

Finally, recent studies demonstrated that by subtracting the potential drop due to SiO_x/Si interfacial fixed charges from the WF_{eff} , gate stacks with various HfO_2 thickness all lead to the same value. This independence of WF_{eff} shift with HfO_2 thickness implies the absence of any isolated fixed charge at $\text{HfO}_2/\text{SiO}_2$ interface or in HfO_2 bulk. Indeed, any charge at $\text{HfO}_2/\text{SiO}_2$ interface or in HfO_2 bulk would lead to increasing WF_{eff} shift when HfO_2 thickness increases [25]. Based on these results, the WF_{eff} expression may be rewritten in an even simpler way:

$$\frac{WF_{\text{eff}}}{q} = \Phi_m - Q_{\text{Si/SiO}_x} \left(\frac{\text{EOT}}{\varepsilon_{\text{SiO}_2}} \right) + \delta \quad (1.13)$$

Eq. 1.13 is the WF_{eff} expression used in this work. $Q_{\text{Si/SiO}_x}$ will be extracted from a WF_{eff} vs EOT for each manufactured device in this work. In summary, to design and accurately tune the threshold voltage (V_{TH}) in CMOS devices, comprehensive and quantitative understanding of WF_{eff} shift is strongly required. In section 2.1.2.2, a practical technique implemented on MOSFET devices completely manufactured at STMicroelectronics's 300 mm wafer fab is described in order to distinguish between an effect of charges density into gate stack or a potential offset induced by dipoles (δ). The methodology consists of varying the SiO_x layer along the wafer to decouple the impact of charges from the interface electrostatic dipole. The influence of fixed charges on V_{fb} should depend on the thickness of SiO_x layers, whereas the influence of dipoles at high- κ/SiO_x interface does not depend on the insulator thicknesses.

1.1.3 MOSFET performance parameters influenced by the gate stack

The MOSFET is the predominant semiconductor device in very large scale integrated circuits. Besides the ability to pack more and more transistors within the same given silicon area, another motivation to further reduce the size of transistors is to improve the performance of the resulting integrated circuits in terms of operating speed and power dissipation. Drive current, transconductance, channel conductance, threshold voltage, reliability and direct tunneling current density or gate leakage density are the main device performance parameters that are somewhat influenced by the MOSFET gate stack properties.

Threshold voltage

The threshold voltage equals the sum of the flat band voltage, plus the semiconductor surface potential ($V_{\text{sc}}=2\Phi_{\text{fi}}$ at threshold condition [8]), plus the voltage across the oxide due to the depletion charges (sign + for NFET devices and sign - for PFET devices).

$$V_{\text{TH}} = V_{\text{fb}} + 2\Phi_{\text{fi}} \pm \frac{\sqrt{4\varepsilon_{\text{sc}}qN_{\text{sc}}\Phi_{\text{fi}}}}{C_{\text{ox}}} \quad (1.14)$$

where V_{fb} is given by the sum of the effective work function defined in Eq. 1.12 and the silicon workfunction determined from Eq. 1.10, Φ_{fi} is the potential difference between the intrinsic Fermi level of silicon and the resulting Fermi level after doping (defined in Eq. 1.9), ε_{sc} is the semiconductor permittivity and N_{sc} is the substrate doping concentration (cm^{-3}). It is convenient to reformulate the V_{TH} in order to explicitly show its dependence with WF_{eff} and EOT. The expression

is obtained by substitution of V_{fb} by Eq. 1.11 and C_{ox} by Eq. 1.3:

$$V_{TH} = \frac{WF_{eff}}{q} - \Phi_{sc} + 2\Phi_{fi} \pm \frac{EOT\sqrt{4\varepsilon_{sc}qN_{sc}\Phi_{fi}}}{\varepsilon_0\varepsilon_{SiO_2}} \quad (1.15)$$

Drive current

For small drain-to-source voltages, MOSFET transistor acts as a linear resistor whose resistance is modulated by the gate-to-source voltage. In this regime, the MOSFET can be used as a switch for analog and digital signals or as an analog multiplier. The drive current (I_{DS}), defined as the current flow of minority carriers from the positive voltage at the drain terminal to the negative voltage at the source can be written as follows [7]:

- For $|V_{DS}| \ll (V_G - V_{TH})$

$$I_{DS} = (W/L)C_{ox}\mu(V_G - V_{TH})V_{DS} \quad (1.16)$$

- For $V_{DS} > (V_G - V_{TH})$

$$I_{DS} = (W/L)C_{ox}\mu\frac{(V_G - V_{TH})^2}{2} \quad (1.17)$$

where W is the width of the transistor channel, L is the channel length, μ is the channel carrier mobility, V_G is the voltage applied to the gate terminal, V_{DS} is the drain to source voltage and V_{TH} is the threshold voltage. An increased I_{DS} thus requires a reduction in the channel length or an increase in the gate dielectric capacitance.

Gate leakage current

Fig. 1.8 shows the 2004 International Technology Roadmap for Semiconductor (ITRS) trends for effective channel length (L_{eff}) and equivalent oxide thickness (EOT) scaling for Low Standby Power (LSTP), Low Operating Power (LOP) and High-Performance logic (HP) technologies. However, this reduction in the oxide thickness causes an important current flow through the gate, which depends exponentially on the thickness of the oxide [29]. This gate current, also known as gate leakage current, represents a limiting factor in device down-scaling and affects the performance of MOSFET devices.

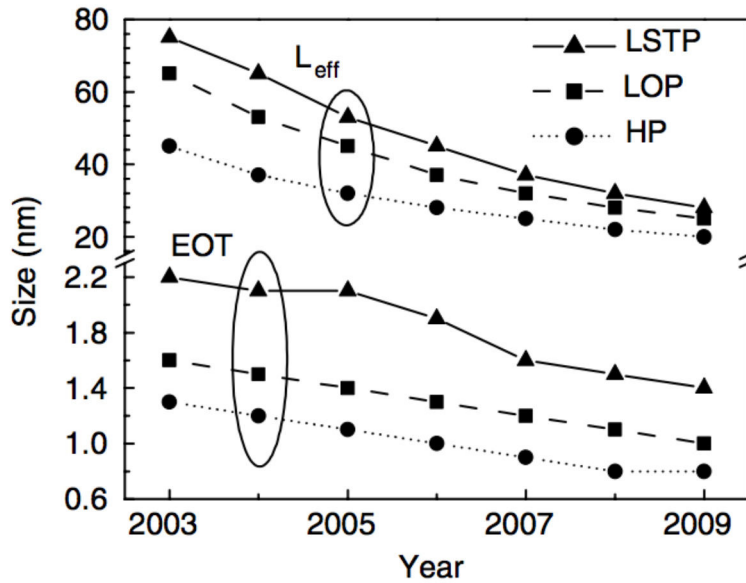


Figure 1.8: Short-term ITRS 2004 projections of effective channel length (L_{eff}) and equivalent oxide thickness (EOT) for Low Standby Power (LSTP), low operating power (LOP), and high performance logic (HP) technologies.

Moreover, for oxides with thicknesses in the order of a few nanometers, direct tunneling even at lower applied voltages is no longer negligible. In general, the direct tunneling current J_G through the energy gap of an insulator or semiconductor is given by [30]:

$$J_G = \frac{2q}{(2\pi)^3 \hbar} \int_0^\infty (f_1 - f_2) \left[\int \int P dk_y dk_z \right] dE \quad (1.18)$$

where P is the tunneling probability, E is the total electron energy, k_y and k_z are the wave vectors in the plane of the barrier (perpendicular to the tunneling direction), and f_1 and f_2 are the probabilities of occupation of the states on each side of the barrier, given by the Fermi-Dirac distribution functions.

Solving such integral is not a trivial problem. It implies the knowledge of the probability of tunneling through the barrier and requires solving Schrödinger equation for the electron wave function. Some hypothesis or other considerations are therefore needed in order to find approximate solutions. The physical mechanisms of tunneling in modern MOS transistors include effects such as carrier quantization and valence band tunneling. Various compact models taking into account such effects have already been proposed [31]. Here, it will be only cited the quasi-empirical model derived by Lee and Hu [32], which takes into account the availability of carriers in the material from where

the carriers are tunneling, and the availability of states in the receiving material. Neglecting high order terms of this complex model, the direct tunneling current J_G can be rewritten as follows [33]:

$$J_G = b \exp \{-afEOT\} \quad (1.19)$$

where a is a constant, b is a pre-exponential factor, and f is entirely dependent on the essential material properties of the gate dielectric. Simplifying to the first order f can be expressed by [33]:

$$f = \kappa(m_{\text{eff}}\phi_b)^{1/2} \quad (1.20)$$

where κ is the permittivity of the gate dielectric, m_{eff} is the tunneling effective mass of carrier in the dielectric, and ϕ_b is the tunneling barrier height. It should be noted from Eqs. 1.19 and 1.20 that direct tunneling increases exponentially with the reduction of the equivalent oxide thickness (EOT) and the attenuation of the leakage current is determined by f . The insertion of a high- κ gate dielectric has therefore successfully allowed the reduction of the high levels of gate leakage current. However, it is worth noting that the direct tunneling for a given EOT is not exclusively determined by the value of κ , but also by m_{eff} and ϕ_b , which strongly affect the tunneling characteristics. For Poly-Si/SiO₂ devices, $J_G = 1\text{A}/\text{cm}^2$ at $V_{\text{ox}} = 1\text{V}$ and a difference in EOT of just 2.1\AA can lead to a decade of magnitude change in the gate leakage current density [33].

Nevertheless, although the successful introduction of the high-k oxide has permitted the aggressive scaling of MOSFET devices, the ultimate down-scaling limit will be set by other factors such as noise, reliability, drain current reduction, direct tunneling between the source and drain, on-chip interconnections, power dissipation, or even the atomic dimensions [31] [14].

Transconductance

The transconductance, in MOSFET transistors, is the change in the drain current divided by the small change in the gate/source voltage with a constant drain/source voltage. It is given by the following equations [7]:

- For $|V_{\text{DS}}| \ll (V_G - V_{\text{TH}})$

$$g_{\text{m, lin}} = (W/L)C_{\text{ox}}\mu V_{\text{DS}} \quad (1.21)$$

- For $V_{\text{DS}} > (V_G - V_{\text{TH}})$

$$g_{\text{m, sat}} = (W/L)C_{\text{ox}}\mu(V_G - V_{\text{TH}}) \quad (1.22)$$

Channel conductance

The channel conductance quantifies the drain current variation with a drain-source voltage variation while keeping the gate-source voltage constant, and it can be calculated from the equations below[7]:

- For $|V_{DS}| \ll (V_G - V_{TH})$

$$g_{d, \text{lin}} = (W/L)C_{\text{ox}}\mu(V_G - V_{TH}) \quad (1.23)$$

- For $|V_{DS}| > (V_G - V_{TH})$

$$g_{d, \text{sat}} = 0 \quad (1.24)$$

As can be seen from these expressions, high values of drive current, channel conductance and transconductance are obtained for large carrier mobilities, large gate insulator capacitances (i.e., thin gate insulator layers), and large gate width to gate length ratios.

In summary, most of the performance parameters of MOSFET devices cited above are strongly influenced by the effective work function and by the gate dielectric capacitance. In this thesis, we will study the impact of the metal process conditions and materials on both the effective work function (WF_{eff}) and the equivalent oxide thickness (EOT). In particular, the impact of the incorporation of lanthanum and aluminum in a *sacrificial* metal gate-first approach on both electrical parameters will be investigated.

1.2 Gate stack fabrication process

1.2.1 Oxidation methods and Gate dielectrics

The gate dielectrics allows to insulate the gate electrode from the silicon in MOSFET devices. Although various methods, such as vapor phase reaction [34], plasma anodization² [35, 36], wet and dry thermal oxidations [37] and wet anodization [38] have been developed for the growth of oxide silicon films, the dominant technology is the thermal oxidation due to its ability to produce a minimum, stable density of oxide fixed charges and interface traps at the oxide/semiconductor interface.

A typical thermal oxidation process consists of a pre-oxidation cleaning step followed by the loading of cleaned wafers into the furnace's cassette storage area. The wafers are then put into the reaction chamber, usually in a non-oxidized ambient, at low enough temperatures so that no native oxide is grown during the insertion. Once the wafer is loaded into the reaction chamber, the furnace temperature is ramped up to process temperature and an oxidizing ambient (generally O_2 or a mixture of H_2 and O_2 reacted to produce steam) is introduced into the reaction chamber through quartz injectors. Oxidation of films thicker than 100 \AA is well described by the Deal and Groove model [39], which states that the oxidation proceeds by three consecutive reactions: i) the transfer of the oxidizing species into the oxide already formed, (ii) the diffusion of the species through the oxide, and (iii) the reaction of oxygen with silicon at the oxide-silicon interface region to form SiO_2 . The proposed kinetics of silicon oxidation results in a linear-parabolic general oxidation relationship between oxide thickness and oxidation time. Nevertheless, experimental data show that films much thinner than 100 \AA grow much faster in O_2 than the model predicts [40]. Indeed, the impossibility to predict the initial stage of the oxide growth is one of the weakness of the Deal and Groove model. Today in microelectronics industry, scaling of MOSFETs to increase performances requires the SiO_2 insulating films to be reduced from hundreds of nanometres to less than 2 nm in order to maintain high drive current and adequate gate capacitance, as it can be deduced from the drive current equation (Eq. 1.16). For this reason, other models have recently been proposed by other researchers [41] [42] [43] in order to handle this thin film oxidation resulting in better experimental fitting. The main concern is that high quality ultra-thin oxides films are not easily fabricated with conventional thermal oxidation methods. Therefore, enhanced techniques to grow ultra-thin SiO_2 films with superior quality and performance have been designed and implemented in recent years. Low pressure combined or not with low temperature oxidation, rapid thermal oxidation (RTO), laser and

²Anodization is an electrolytic passivation process used to increase the thickness of the natural oxide layer on the surface of metal parts.

plasma oxidation, VUV³ photo-oxidation and in-situ steam generation (ISSG) oxidation are examples of such improved methods.

Rapid thermal oxidation is currently one of the primary thermal oxidation techniques for growth of ultra-thin SiO₂ films. In this process, the Si wafers are heated electrically by a current passing through them or heated by lamp radiation. This technique requires low thermal budget and the Si wafer can be heated and cooled quickly. The temperature process can vary from 500 °C to 1100 °C and the oxidizing ambient may include O₂, O₃, N₂/O₂, N₂O or NO.

In-situ steam generation (ISSG) is characterized by the way the steam is generated. On one hand, batch furnaces typically use an external pyrogenic torch to produce steam. In the external torch configuration, H₂ and O₂ are combusted at atmospheric pressure in proximity to a hot element or in a hot-wall chamber that ignites the reaction, producing gaseous H₂O. The resulting steam is then introduced into the hot-wall furnace tube to oxidize the wafers. On the other hand, ISSG is a low-pressure process (typically below 20 torr) wherein pre-mixed H₂ and O₂ are introduced to the process chamber directly, without pre-combustion, where they flow across a rotating wafer heated by tungsten-halogen lamps separated by a thin quartz window [44]. The reaction between H₂ and O₂ occurs close to the wafer surface because the hot wafer acts as the ignition source.

Interlayer dielectric growth

As mentioned in section 1.1.2, the electrical properties of the MOSFET devices are strongly correlated with structure and defects near and at the Si/SiO₂ interface. Channel mobility and other reliability parameters such as leakage current, time-dependent dielectric breakdown (TDDB), and hot electron induced effects are also influenced by this Si/SiO₂ interface. As the thickness of gate oxide scales down, the transition region between crystalline silicon and amorphous SiO₂ becomes a significant part of the dielectric and much more attention has therefore to be paid to the quality and integrity of this interface. Furthermore, with aggressive scaling, direct tunneling and dopant penetration through ultrathin SiO₂ layers become critical issues.

In the search for a new material to replace SiO₂ and to overcome these problems, the incorporation of nitrogen atoms into SiO₂ to form oxynitrides (SiO_xN_y) was at the beginning an appropriate solution to increase the performance of MOSFET devices [45]. Indeed, with the incorporation of nitrogen, the diffusion barrier properties of oxynitrides are enhanced. As a result, it lowers the diffusion rates for oxygen and other dopants such as boron,

³Vacuum Ultraviolet is the ultraviolet radiation below 200 nm.

decreasing the growth rate of any further oxydation or nitridation [46]. In other words, nitrided oxides are more resistant to further oxidation and diffusion of dopants from the gate to the channel and viceversa compared to non-nitrided films. In addition, nitrogen incorporation at the Si/SiO₂ interface was found to reduce defect generation and traps in the oxide [47]. Moreover, the dielectric constant of the oxynitride increases linearly with the percentage of nitrogen from $\varepsilon_{SiO_2} = 3.9$ to $\varepsilon_{Si_3N_4} = 7.8$.

For the integration of these oxynitride dielectrics, several techniques such as RTCVD (Rapid Thermal Chemical Vapor Deposition) or RPECVD (Remote Plasma Enhanced Chemical Vapor Deposition) have been introduced [48]. Nitridation by DPN (Decoupled Plasma Nitridation) or RPN (Remote Plasma Nitridation) of thermally grown oxides fabricated by RTO or ISSG have also been explored [49]. A simple way to introduce relatively high concentrations of nitrogen (≈ 10 -15 at.%) into the SiO₂ is nitridation in ammonia (NH₃). Furthermore, nitridation of SiO₂ in NH₃ allows the simultaneous nitridation of the Si/SiO₂ interface and the SiO₂ interface. However, one limitation of this approach is the high concentration of hydrogen, which can act as traps, introduced into the oxynitride film. Actually Si-H bonds are convenient to passivate dangling silicon bonds but ironically they can break at elevated temperatures and high electric fields due to their lower binding energy and reactivate the interface defects [50]. This disadvantage can be overcome by performing a thermal reoxidation of the oxynitride film in dry O₂, which completely removes the hydrogen from the film [51].

The ultra-thin oxynitride (SiON) film used at STMicroelectronics as interlayer dielectric is fabricated by performing first an enhanced ISSG oxidation of silicon at 800 °C, followed by NH₃ nitridation and RTP anneal at 700 °C, resulting in a SiON layer with thickness between 8 Å and 12 Å.

High- κ deposition

As predicted by ITRS in 2004, the oxynitrides by themselves are not able to satisfy the gate leakage requirements at the 45 nm node and beyond. The gate leakage increases exponentially as the thickness of the dielectric is reduced, but is also influenced by the metal/dielectric barrier height (conduction band offset) and the semiconductor/dielectric hole barrier height (valence band offset). Maintaining the same channel charge induced into the semiconductor with the same gate voltage (i.e. the same capacitance) is one of the device constraints when considering different possible gate dielectric materials. The dielectric thickness needed to keep the same capacitance actually vary directly with the dielectric constant (κ), according to the capacitance equation (Eq. 1.2). It means that for a material with a κ -value N times higher than that of SiO₂ (3.9),

the oxide film can be deposited until N times thicker than that made of SiO_2 . However, as seen in Fig. 1.9, there is a strong tendency for materials with higher κ -values to also have smaller bandgaps, which must be translated into lower barrier heights.

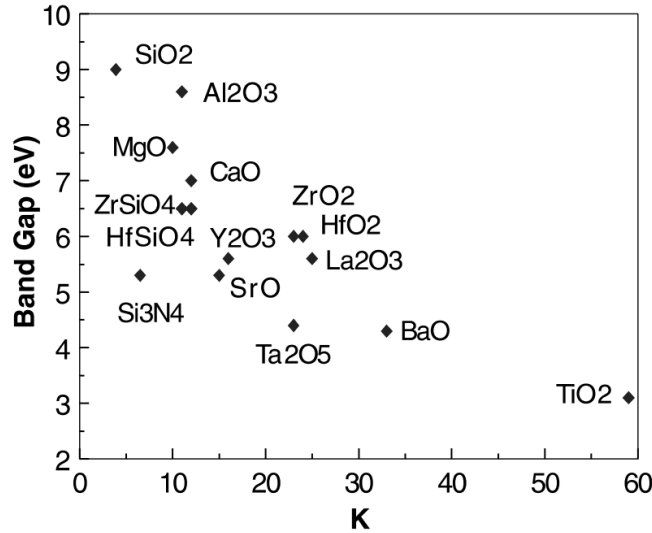


Figure 1.9: Energy bandgap vs. dielectric constant for different gate oxides [52].

Some of the requirements to be met for a high- κ material to be considered as a candidate to replace SiO_2 are therefore : a) a sufficiently high- κ dielectric constant, b) large bandgap, c) large band offsets relative to the conduction and valence bands of silicon (both over 1 eV), d) thermodynamic stability in contact with silicon, and e) retention of the amorphous phase at high temperatures. Indeed, a polycrystalline material could result in grain boundaries acting as current leakage paths [53]. In the search for the ideal high- κ dielectric, the most promising materials in the 14 nm node result to be HfO_2 and Hafnium silicates, with and without some form of nitrogen, despite of their relative poor thermodynamic stability in contact with silicon compared to SiO_2 and their lower crystallization temperature, leading to a degradation of both the surface mobility and the dielectric breakdown, and larger shifts in the flat band voltage due to the high fixed charge density. This is why the ultra-thin oxynitride (SiON) film is grown on silicon before the Hf-based oxide deposition.

On the other hand, hafnium oxide (HfO_2) has a dielectric constant ranging from about $\kappa \approx 18$ -25 [52], and HfSiO has been reported to have a dielectric constant about $\kappa \approx 15$ [54]. In addition, Hf-based gate dielectric compounds also have high band offsets, $\Delta E_c \geq 1.5$ eV and $\Delta E_v \geq 3$ eV [52]. Leakage current density of at least two orders of magnitude lower than SiO_2 was reported for

HfSiON-based devices. The presence of nitrogen in the film enhances the thermal stability relative to HfSiO, and no crystallization was observed for anneals up to 1100 °C [55]. Indeed, annealing the Hf-based film in nitrogen ambient improves its thermal stability making it suitable for subsequent source and drain high temperature anneals. The first Hafnium-based gate dielectrics were deposited by Physical Vapor Deposition (PVD) [55] [56]. Today, these layers are now being deposited by Chemical Vapor Deposition (CVD) or Atomic Layer Deposition (ALD) to meet the requirements of compositional control and conformality. On one hand, the CVD processes uses metal organic precursors such as a combination of tetrakis ethylamido hafnium (TDEAHf) and tetrakis methylamido silicon (TDMAS), and of hafnium tert-butoxide (HTB) and tetraethoxysilane (TEOS) or disilane (Si_2H_6). On the other hand, in ALD processes, HfO_2 uses hafnium tetrachloride (HfCl_4) as precursor while HfSiO uses metal organic precursors such as tetrakis ethylmethylamido silicon (TEMASi) and tetrakis ethylmethylamido hafnium (TEMAHf). In addition, nitridation methods such as Decoupled Plasma Nitridation are currently used to drive nitrogen at the surface of Hf-based oxides in order to further increase the dielectric constant.

The introduction of a high- κ materials into gate stack at 45 nm node marked a milestone in the history of semiconductor industry [57] because this transition required significant changes in process integration flows and appropriated tools to implement it in manufacturing fabs to be produced in high volume. However, the inclusion of a high- κ material introduces issues such as non-compatibility with poly-Si processing and significant degradation of channel mobility. To overcome this problem, the high quality interlayer dielectric is not suppressed and metal gates are introduced between Poly-Si and high- κ . In 14 nm MOSFET devices fabricated at STMicroelectronics, HfO_2 is deposited over the interlayer dielectric (SiON) by Atomic Layer Deposition (ALD) in order to keep a high mobility interface, followed by nitridation using DPN and Post-Nitridation Anneal before metal gate deposition. The resulting HfON layer has a thickness between 19 Å and 21 Å.

1.2.2 Metal gate electrode and deposition techniques

Metal gate electrodes are required to be included in gate stack together with high- κ in order to eliminate boron penetration effects and poly-Si depletion, thus reducing EOT penalty and increasing the ON current (I_{DS}) of the transistors. Furthermore, surface phonon scattering degrades the electron mobility in the high- κ , reducing the speed at which transistors can switch. It was shown that the influence of the phonon oscillations can be screened out by significantly increasing the density of electrons in the gate electrode [58]. Therefore, the higher density of electrons in a metal screen out the vibrations and let current to flow more smoothly [58], compared to Poly-Si gate with lower free electron concentration.

In addition, the defect formation at poly-Si and high- κ interface, which was most likely the cause of the Fermi level pinning in first poly gate studies, is reduced with the deposition of a metal gate. Last but not least, metals show less resistance compared to Poly-Si.

1.2.2.1 Thin metal gate films deposition techniques

Chemical Vapor Deposition

Chemical Vapor Deposition and derived improved techniques are one of the most employed techniques to deposit thin films in microelectronics. Nevertheless, for metal gate applications, high temperatures of film deposition might degrade the devices performance.

Atomic Layer Deposition

Atomic Layer Deposition (ALD) is a thin film deposition technique that is based on the sequential use of a gas phase chemical process. Although ALD is similar in chemistry to CVD, the main difference between them is that in ALD the precursor materials are kept separated during the reaction. In other words, the surface of the substrate is exposed to each precursor one at a time in a succession of sequential pulses. Precursors are never present simultaneously in the reactor. Each single exposure of the precursor is called a cycle. Separation of the precursors is possible by pulsing a purge gas (typically nitrogen or argon) after each precursor pulse. ALD makes atomic scale deposition control possible. This is because the amount of deposited material is completely controlled by the precursor-surface reaction and it finishes once all the reactive sites on the surface are consumed. As a consequence, the film thickness depends only on the number of reaction cycles, and extremely conformal and uniform thickness, and low impurity level of the film can be obtained by using ALD.

ALD has successfully been used to deposit metallic films such as TiN from TiCl_4 and NH_3 precursors [59] [60]. TDMAT and NH_3 have been also used as precursors to deposit TiN, but the films shows high porosity, which leads to easy oxidation of the TiN films and lowers the film resistivities [61]. Nevertheless, despite the clear advantage of ALD compared to CVD to deposit thin films with high degree of thickness control at lower temperatures, ALD has not been chosen for TiN metal gate deposition at STMicroelectronics. The major limitation of the ALD for industrial applications is the slowness of the deposition. Besides the low throughput, another concern is the high cost of ownership, limiting the proliferation of ALD in the semiconductor industry. However, the slowness of ALD is not such an important issue on extremely scaled devices and ALD has recently emerged as the preferred way of metal deposition in gate-last integration due to the low global thickness non-uniformity, low micro-loading and conformal step coverage. Nevertheless, unlike gate-last processes, such conformal step coverage is not essential in the

gate-first process used to fabricate 14 nm FDSOI devices, thanks to both the integration and the planar architecture of this technology. For this reason, the deposition technique called Radio-Frequency PVD allowing deposition at room temperature at higher deposition rates have been chosen for metal gate applications at STMicroelectronics. The Radio Frequency PVD is introduced in the next section, after a reminder of sputtering systems.

Physical Vapor Deposition and Sputtering systems

Sputtering is the process of removing surface atoms or molecules from a solid target by the bombardment of ions and sputter deposition involves ejecting those atoms or molecules from a source (target) onto a substrate (silicon wafer). Sputtering, unlike evaporation, does not require melting the metal to be deposited. Therefore, refractory metals such as titanium (Ti) and tungsten (W) that are very difficult to melt can be used. Sputtering also preserves the original composition of the target material. In addition, sputtering allows the deposition of many different materials by combining materials from solid and gaseous sources, which are introduced into the vacuum chamber either before deposition or during deposition. This form of sputter deposition is called reactive sputter deposition. Physical vapor deposition (PVD) processes include a variety of thin film deposition methods used to deposit thin films *atomically* by means of fluxes of individual neutral or ionized species. Cathodic arc deposition, electron beam physical deposition, evaporative deposition, pulsed laser deposition and sputter deposition are some examples of physical vapor deposition methods. Here, we use sometimes the term PVD to refer to the sputtering process by misuse of the language.

From simple DC to reactive Radio-Frequency Magnetron sputtering systems

The sputtering process is carried out in a vacuum chamber in order to avoid contaminants to interfere with the deposition process and to establish the pressure required by the sputtering plasma. The high vacuum pumping (in the 10^{-9} Torr range) is done by cryopumps. In a simple DC sputtering system, the target plays the role of the cathode and the substrate, that of the anode. When a DC voltage equal to a voltage known as breakdown voltage (V_b) is applied between both electrodes, a stream of electrons is created. The electrons collide with the argon and create positively charged argon ions (Ar^+), which are strongly attracted to the negatively charged cathode (the target). The argon ions collide with the target surface and some of them cause a surface atom of the target to be ejected. The sputtered atoms travel across to the substrate where they are deposited as a film. For a given gas, the minimum voltage V_b necessary to create a stream of electron between the electrodes is a function only of the product of the pressure and the electrodes gap distance (parameter pd in Fig. 1.10). The curve voltage versus the gas pressure-electrodes gap

distance is called Paschen's curve. Examples of Paschen's curves obtained for different gases are shown in Fig. 1.10.

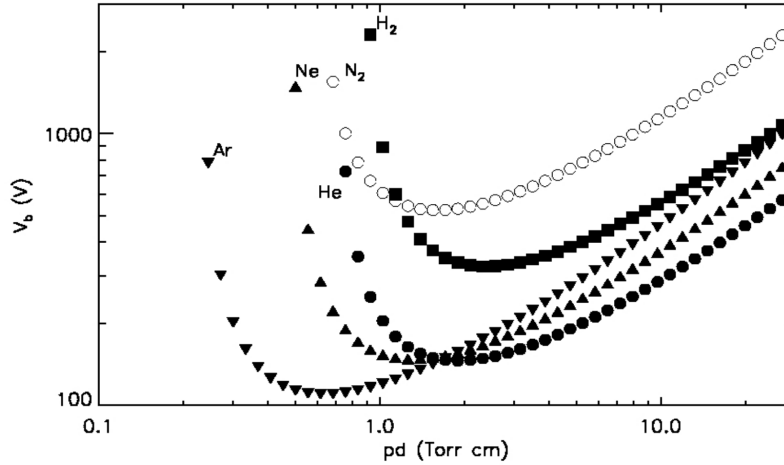


Figure 1.10: Paschen curves obtained for Helium, Neon, Argon, Hydrogen and Nitrogen [62].

Improved sputtering effects can be achieved by adding magnets behind the cathode to a simple DC sputtering system. In this case, the sputtering rate is improved because the ionization efficiency is enhanced. Electrons are confined because they tend to turn around the magnetic field lines, ionizing more argon atoms in their path. In this way, both electron and Ar^+ ions density are increased in a DC magnetron system. On one hand, with the increased electron density, the required applied voltage to sustain the plasma can be decreased from 5 000 V - 10 000 V in a simple DC sputtering system to only 400 V - 800 V in a DC magnetron system. On the other hand, because the density of Ar^+ ions also is increased, it is possible to lower the sputtering chamber pressure. At lower pressures, the sputtered atoms have fewer collisions on their path to the substrate. Fewer atom collisions result in an increased deposition rate. Actually, the pressure has to be high enough to keep a high density of ions in the plasma in order to sputter the target but low enough to reduce the collisions of the sputtered atoms. Too low or too high pressures have to be compensated by higher breakdown voltages, as described in Fig. 1.10. Finally, the electron confinement in a magnetic field near the target also reduces electron bombardment of the substrate. This results in much less heating of the substrate. Nevertheless, the cathode in DC discharge must be an electrical conductor, since an insulating surface will develop a surface charge that will prevent ion bombardment of the surface.

In addition, a Radio Frequency (RF) generator operating at a frequency of 13.56 MHz (standard in industry) can be coupled to a DC power. The main interests of RF-PVD deposition are the reduction of the breakdown voltages and the improvement of deposition uniformity. Indeed, the voltage needed to ignite the plasma is reduced because the oscillating electrons can ionize more the argon gas. A direct consequence of the reduction of the breakdown voltage is that the metal atoms are ejected with less energy, limiting the damage that could be caused on the high- κ layer by bombardment on the wafer during the plasma sputtering. This is a critical point because any significant damage can be detrimental for gate stack integrity and transistor electrical properties. In addition, less high energetic electrons are provided in a RF sputtering plasma compared to DC plasma for a same density of electrons, as shown in Fig. 1.11.

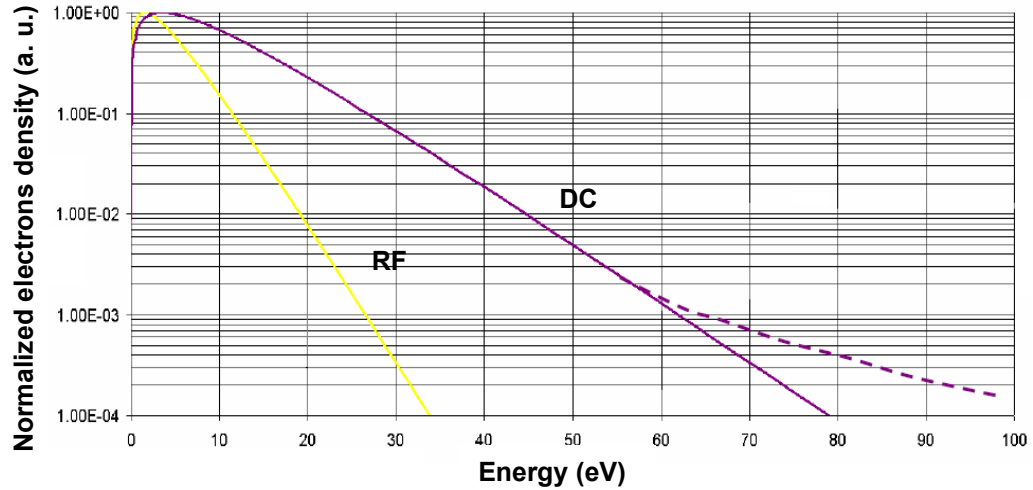


Figure 1.11: Energetic distribution of electrons in RF (yellow) and DC (purple) plasma [63].

The breakdown voltage reduction also allows to increase significantly the sputtering angle of the ejected metal atoms, as shown in Fig. 1.12. Consequently, the thickness uniformity (σ) can be improved by $\approx 1\%$ for thicknesses $< 100 \text{ \AA}$.

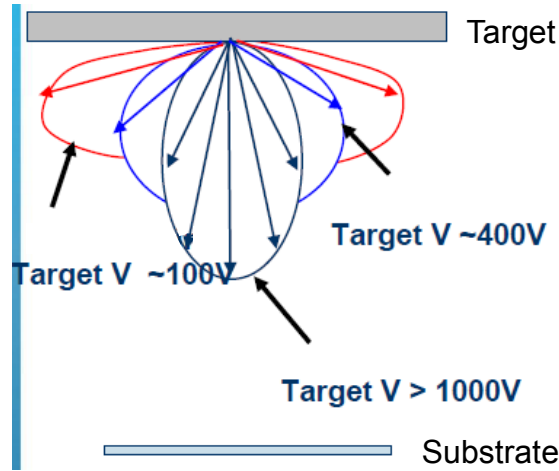


Figure 1.12: Representation of sputtering angle vs. breakdown voltage [63].

The Endura platform supplied by Applied Materials has been used in this work for metal gate deposition in RF-PVD process chambers. These chambers are connected by a main transfer chamber pumped to an ultra-high vacuum (in the 10^{-9} Torr range), as shown in Fig. 1.13. The design of the Endura RF-PVD platform allows the system to precisely and quickly move the wafers from a chamber to another without breaking the high-vacuum. Each RF-PVD process chamber contains a target material for thin films deposition. The target materials used for metal gate applications are titanium (Ti), aluminum (Al) and lanthanum (La). In general, all RF-PVD chambers are composed of a RF generator, a DC power supply, a cryo pump system for ultra high-vacuum and an Ar mass-flow controller to regulate the argon flow into the chamber. Reactive RF sputter chambers used to deposit TiN are provided, additionally, of a nitrogen mass-flow controller, as shown in Fig. 1.14. The stoichiometry of the deposited thin films can be controlled by adjusting the flow of these gases. The volumetric flow rate is measured in standard cubic centimeters per minute (sccm) units.

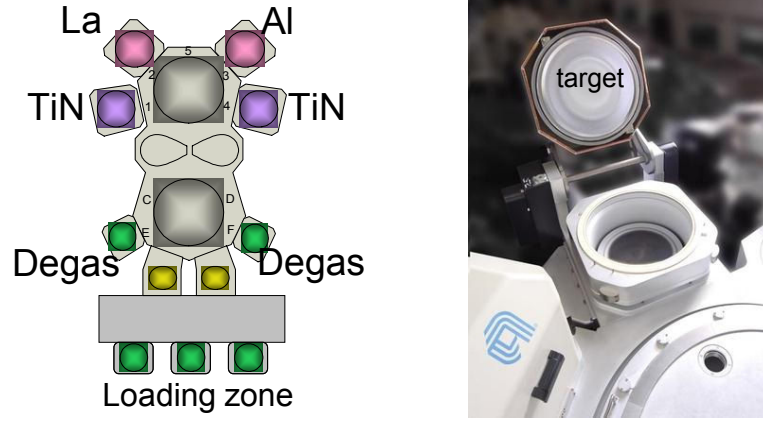


Figure 1.13: On the left, Endura platform with 4 RF-PVD chambers for TiN, lanthanum and aluminum deposition. On the right, photo of a deposition chamber.

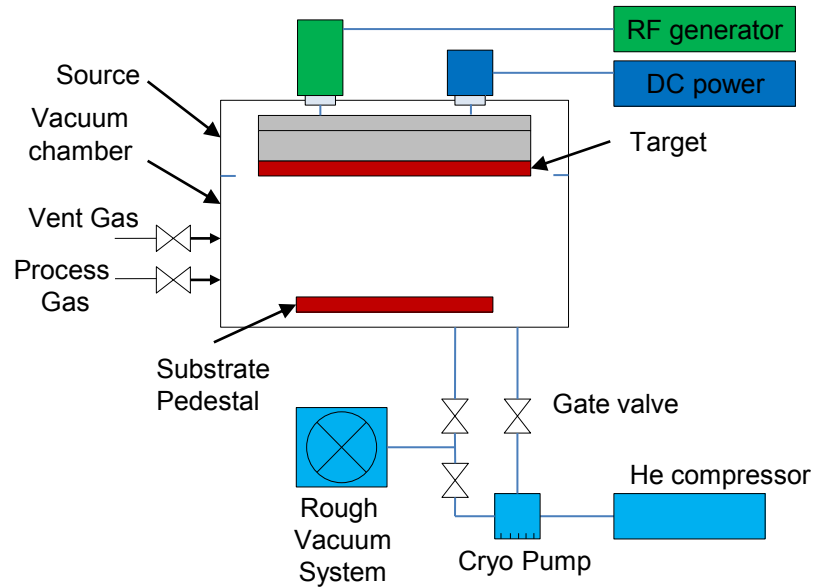


Figure 1.14: Schema of a RF-PVD chamber used to deposit TiN metal gate layers [63].

Once wafers are introduced into the Endura RF-PVD system, immediately after high- κ deposition and respective nitridation and post-nitridation

annealing, they are subjected to a high temperature bake in an Ar ambient called degas, in order to desorb aqueous and volatile species from the wafer. Degas is performed in specific chambers (Fig. 1.13 on the left) and is limited at 200 °C for 30 s to avoid the desorption of the nitrogen.

All the films in the Endura RF-PVD chambers are deposited at room temperature (20 °C). One argon back side flow enables the heat transfer from the electrostatic chuck to the wafers electrostatically clamped. The target and the chamber walls are also thermalized ⁴ at 16 °C through a common deionized water circuit. In order to limit the heating of the chamber walls, the maximal deposition time is restricted to 100 s. For longer times, deposition is carried out in several deposition cycles separated by a cooling step of 10 s. Actually, during deposition, certain particles may condensate on chamber walls. Heating might cause particles to be detached from the chamber walls, increasing the risk of defectivity or non-uniformity. Deposition rate is in the order of 1 Å/s.

1.2.2.2 Deposition of ultra-thin Titanium nitride films and Aluminum and Lanthanum monolayers by RF-PVD magnetron sputtering

Target conditioning (Warm-up)

Prior to the deposition of metal thin films, the RF-PVD chamber is warmed up. The warm up sequence consists here of the introduction of the gases, the ignition of the plasma and the target sputtering and film deposition on a shutter inside the chamber. The aim of the warm up is to simulate the deposition conditions in order to avoid undesirable "first wafer" effects. The target is thus conditioned with the same process parameters than those used later for film deposition on the wafer, except in the special case of incorporation of nitrogen into the film with a gradient profile [64], detailed in the next paragraph.

TiN films deposition

Titanium nitride (TiN) is formed by reactive sputtering of a pure Ti target in a nitrogen-containing ambient, typically Ar/N₂, in a RF-PVD chamber with combined DC and RF generators. Unlike, TiN films deposited by ALD or CVD, the nitrogen and titanium do not combine in the gas phase, but on the surface of the target, chamber walls, and on the wafer. The TiN thickness is in the range between 10 Å and 120 Å.

Unless noted otherwise, the TiN layer used in this work as metal gate has a gradient profile of nitrogen and its deposition proceeds as follows: first, during the warm-up, the Ti target is completely denitrided by pure Ar sputtering.

⁴In physics, thermalization is the process of reaching thermal equilibrium by mutual interaction.

Next, during film deposition, the chamber is backfilled with an Ar/N₂ mixture and plasma is initiated with a high volumetric flow rate of nitrogen ($N_2/(Ar + N_2) = 0.7$) in order to progressively cover the surface of the target with nitrogen. As a result, only titanium is sputtered from the surface of the target for the first monolayers deposited over HfON. As nitrogen covers the target surface, TiN begins to sputter and the nitrogen composition gradually increases into the TiN film being deposited. At this point, a noticeable drop of the sputter rate occurs. This decrease is due to both the difference in the sputtering yield achieved with only Ar and a mixture of N₂ and Ar ions, and to the decrease in sputtering yield caused by the nitridation of Ti target surface. Accordingly, the deposition rate decreases from 1.6 Å/s to 0.4 Å/s during the TiN deposition with a total pressure of about 3 mTorr and a DC and RF power of 700W and 600W, respectively. Intentional gradient composition of nitrogen into TiN results in better gate leakage (reduction of $\approx 20\%$ on NFET) and better reliability [64].

Deposition of Aluminum and Lanthanum monolayers

Both aluminum (Al) and lanthanum (La) monolayers can be incorporated as additives into the TiN metal gate. While TiN is deposited with combined DC and RF powers, aluminum and lanthanum are deposited with only RF sputtering at smaller deposition rates because desired thicknesses are much thinner (from 2 Å to 6 Å) than for TiN.

RF power and Ar flow conditions are chosen in order to optimize the deposition rate and the uniformity of the Al monolayers. Increasing the RF power and reducing the Ar flow results in a higher deposition rate, but an increment of the Ar flow improves the uniformity, as a result of the reduction of deposition rate. Table 1.1 indicates the retained deposition conditions of aluminum and lanthanum in terms of Ar flow in sccm and RF power. In these conditions, the deposition rate is extremely low, in the range between 0.1 Å/s and 0.3 Å/s. Obviously, the processing time t is the variable that is used to control Al and La thicknesses.

	Aluminum	Lanthanum
time (s)	t	t
RF power (W)	200	75
Ar flow (sccm)	30	5

Table 1.1: RF power and Ar flow into RF-PVD chamber for deposition of Al and La monolayers.

1.3 Metal gate integration in 14 nm Fully-Depleted SOI devices

1.3.1 FDSOI architecture

As CMOS technology scales down, two approaches are pursued by the industry to overcome the fundamental limits of traditional planar bulk transistors. One is the introduction of a Tri-gate or FinFET transistor at the 22 nm and 16 nm nodes [65] [66]. The other is the Fully Depleted Silicon On Insulator (FDSOI) architecture, schematized in Fig. 1.15, which provides a simplified planar manufacturing process compared to 3-D FinFET technology.

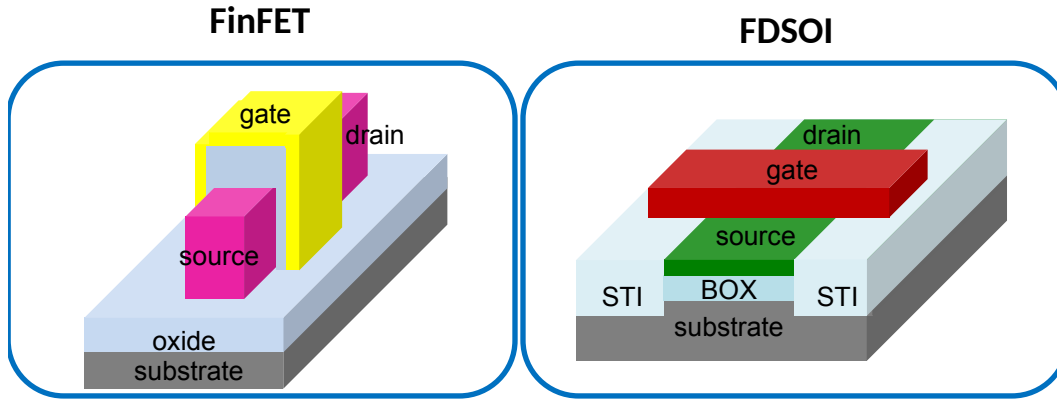


Figure 1.15: Scheme of the FinFET architecture on the left, and the Fully-Depleted Silicon-on-Insulator architecture on the right.

Although FinFET architectures offer impressive drive currents per footprint at low supply voltages because of the 3-D conduction channel and excellent electrostatic control, they have high gate and parasitic capacitances, proportional to the 3-D effective width increase, which negatively impacts both the speed and active power consumption [67]. In addition, complexity of FinFET technology has created new challenges for many design teams because their current tools and techniques may not enable them to design their IP blocks optimally for FinFET processes, delaying time to market. FinFET requires a new generation of design experience, expertise, equipments and tools in order to get the most from the technology.

In contrast, the 14 nm FDSOI offer is derived from 20 nm bulk design rules and from 28 nm FDSOI process technology [68] and the design transfer is thus less complicated. Only few analog parts need to be adopted to the FDSOI.

FDSOI is a planar technology that reuses ninety percent of the process steps used in the bulk counterpart and the overall manufacturing process in FDSOI is 15% less complex, leading to lower cycle time and reduced manufacturing costs. The manufacturing tools for FDSOI are identical to the last generation of bulk processes. FDSOI transistors are manufactured on a ultra-thin buried oxide (BOX) layer of 20 nm. Silicon channel thickness is about 6 nm. Shallow Trench Isolation (STI) feature authorizes the lateral insulation between adjacent semiconductor device components.

1.3.1.1 Multi- V_{TH} offer

Another feature of the FDSOI architecture is the possibility to use the BOX layer as second gate oxide. Even in the absence of a back bias, the use of a ground plane below an ultra thin BOX (< 30 nm), formed by doping implantation of either Indium or Arsenic, allows to achieve several threshold voltages with a single metal gate [69]. Fig. 1.16 demonstrates that multi- V_{TH} solutions exist in undoped channel FDSOI by a gate-first approach and the integration of 2 different metal gates together with 2 different WELL types.

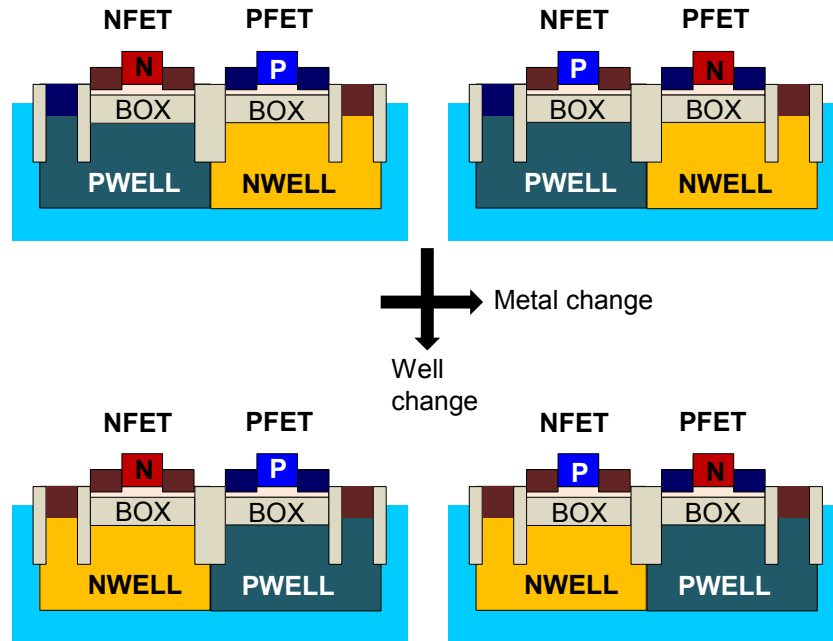


Figure 1.16: Schematics CMOS configuration to achieve 4 V_{TH} s with two metal W_{eff} (N & P types) and two WELL types (NWELL vs PWELL). *Adapted from [69].*

In addition, one of the main advantages of thin BOX in FDSOI architecture is

the ability to apply a back bias below the BOX. The performance of the transistors can thus be adapted according to the applied back bias (V_{BB}). This technique is called the back biasing [70]. For different thicknesses of BOX, in the presence of a ground plane or not, the threshold voltage of the transistors is changed when voltage is applied below the BOX. A positive voltage in the NFET decreases the threshold voltage and a negative voltage induces a decrease of $|V_{TH}|$ on the PFET. This type of bias is called Forward Back Bias (FBB) and increases the I_{on} current flow. By inverting the sign of V_{BB} , it is also possible to increase the threshold voltages, which can be used to decrease the current I_{off} . This type of bias is called Reverse Back Bias (RBB).

1.3.1.2 Effective work function requirements for Fully-Depleted SOI devices

Metal gate electrodes with the correct effective work functions are required in order to keep a low threshold voltage for both NFETs and PFETs. Fig. 1.17 [71] shows that the effective work function target depends on the channel doping density, as predicted by the term N_{sc} in Eq. 1.14. According to this figure, for high doping concentration, metals with work functions close to the silicon conduction band (CB) or the silicon valence band (VB) are required, for NFETs and PFETs respectively [72]. In contrast, for low doping concentration, the target is almost midgap work functions, at only ± 100 meV from either side of the intrinsic silicon Fermi level. Therefore, in FDSOI devices with undoped Si channel almost midgap metals will be required for low V_{TH} , in contrast to bulk technologies [73] [69]. But actually, according to Eq. 1.14, it is the entire voltage drop across the oxide due to depletion charges ($Q_{dep} = qN_{sc}x_{dep}$) at threshold voltage condition that sets the requirements for effective work function of the metal electrodes. Therefore, it is worth noting that even in highly doped channel FDSOI devices, depletion charges are too low ($\approx 6 \times 10^{11} \text{C/cm}^2$) because the width of the depletion region into the substrate is limited to 6 nm, and therefore metals with effective work functions close to the midgap will still be required. This is a great advantage of FDSOI devices compared to bulk technologies. However, from the point of view of a process engineer, this is also a new challenge because the work function range to switch from an N-gate to a P-gate in FDSOI devices is shorter (between 100 meV and 200 meV) than in bulk devices, which makes even more necessary a very fine adjustment of the effective work function of metal electrodes.

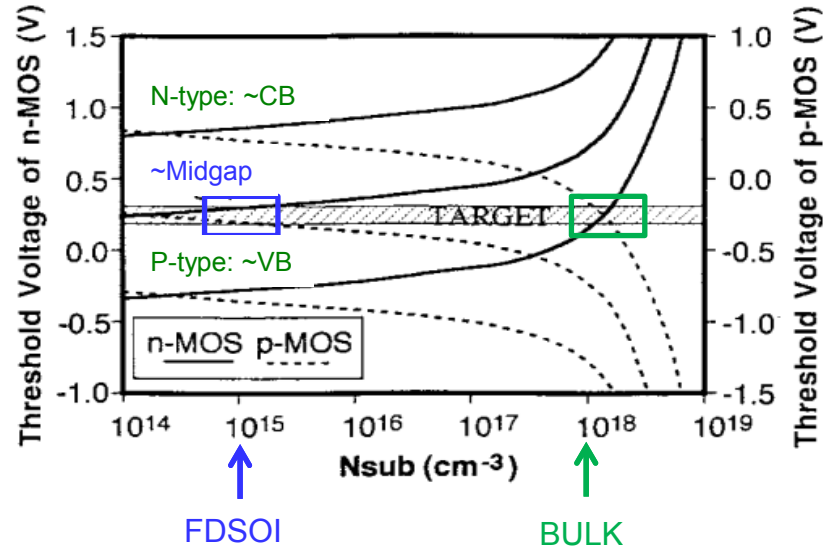


Figure 1.17: Metal work function requirements as a function of channel doping density, in order to keep low V_{TH} for both NFET and PFET devices. *Adapted from [71].*

In the next part, a new approach is proposed to carefully incorporate the lanthanum and the aluminum into TiN gate stack in order to authorize a fine modulation of work function of metal electrodes. This approach is called the *sacrificial* metal gate-first integration and their benefits compared to the standard gate-first approach are studied in chapter 3.

1.3.2 Sacrificial metal gate-first process integration

Once interlayer and high- κ dielectrics have been deposited on the silicon on insulator substrate, the metal gate is deposited following the next steps, as shown in Fig. 1.18.

1. First, the *sacrificial* multi-layer metal gate stack, composed by a TiN layer and monolayers of additives such as lanthanum or aluminum, is deposited in the Endura RF-PVD system.
2. Next, silicon is deposited as a capping layer on the top, followed by a thermal treatment under N_2 atmosphere at 900 °C for 10 s in order to activate the diffusion of additives into the HfON/SiON stack.
3. The sacrificial gate stack is then removed by a chemical wet solution.
4. Finally, a Poly-Si/TiN electrode without additives is deposited, followed by

gate patterning. Later, this final electrode is exposed to the spike annealing at 1005 °C, responsible of the Source/Drain dopant activation.

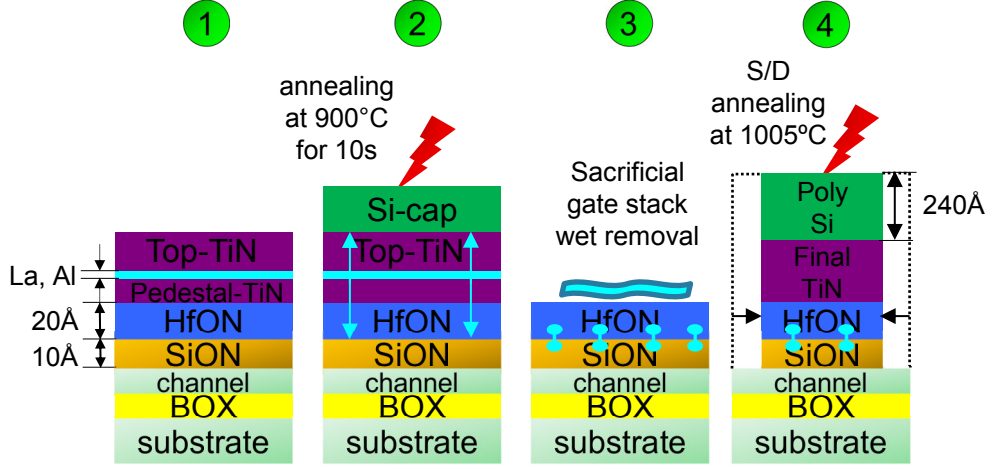


Figure 1.18: Sacrificial gate-first approach: 1) sacrificial gate deposition, 2) annealing at 900 °C for 10 s, 3) sacrificial gate stack wet removal and 4) final gate deposition, gate patterning and S/D annealing.

We have taken advantage of the sacrificial gate-first integration to decouple the main objectives of this thesis. On one hand, it is mainly with the sacrificial gate that we will seek to modulate the V_{TH} . On the other hand, it is with the final gate that we will seek to reduce the local V_{TH} variability.

1.3.3 Random local V_{TH} fluctuations

The variation of threshold voltage between devices undergoing same manufacturing process is one of the most critical challenges for the future CMOS technology nodes [74] [75]. The origins of the local V_{TH} fluctuations have been mainly attributed to the random dopant fluctuations (RDF), line edge roughness (LER), polysilicon gate granularity (PGG) [76] and metal gate granularity (MGG). With the integration of an undoped or lightly doped ultra-thin FDSOI architecture, V_{TH} fluctuations are improved, as compared to bulk MOSFETs [77] [78] [79], due to the reduction of the RDF contribution. Although the RDF component is dramatically decreased, there are still other sources of variability such as LER and MGG, which become more important on scaled devices, as shown in Fig. 1.19.

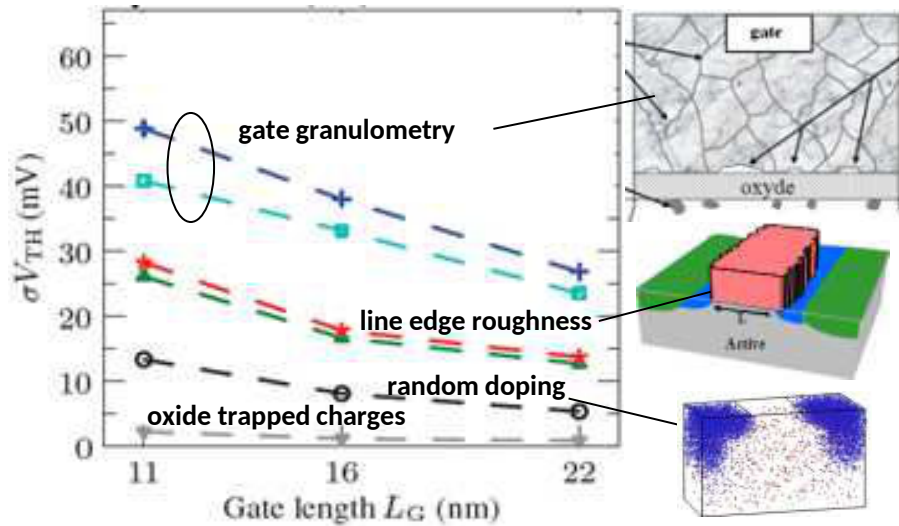


Figure 1.19: Simulation of the individual contributions of the Random Dopant Fluctuations (RDF), Oxide Thickness Fluctuations (OTF), Line Edge Roughness (LER), and Metal Gate Granularity (MGG) to σV_{TH} [80].

In this thesis, focus will be done on metal gate granularity (MGG). The work function depends on the grains orientation. Indeed, the work function is experimentally known to be an anisotropic property, i.e. to be different from different faces of a crystal [81]. Recently, the dependence of the metal work function on the orientation of its grains has been recognized as one of the main sources of random V_{TH} fluctuations [82, 83] and the main contributor on undoped FDSOI [80] and FinFET devices [84], where it is assumed that there is little or no random V_{TH} fluctuation due to Random Dopant Fluctuation (RDF).

On the other hand, the work function of a metal should be modeled as probabilistic distribution rather than a deterministic value. Since metal gates are composed of grains with a random distribution of orientations, each one with a corresponding probability P_i , the metal work function for a large number of metal gates should be written as follows [82]:

$$\Phi_m = (P_1)\Phi_1 + (P_2)\Phi_2 + \cdots + (P_k)\Phi_k \quad (1.25)$$

where P_i is the percentage of a particular grain orientation in the total population of grains. However, for each metal gate with a particular area A , the metal work function should be rewritten as [82]:

$$\Phi_m = \left(\frac{X_1}{N}\right)\Phi_1 + \left(\frac{X_2}{N}\right)\Phi_2 + \cdots + \left(\frac{X_k}{N}\right)\Phi_k \quad (1.26)$$

$$N = \sum_{i=1}^k X_i \quad (1.27)$$

where X_i are random variables that represent the number of grains with work function values of Φ_i , and X_i/N is the percentage of grains with work function values of Φ_i in the particular area A with a number of grains N . These equations were adapted from [82] to intentionally show that X_i/N is not necessarily equal to P_i but tends to P_i for a large number of metal gates or as the area of the metal gate increases. Now, suppose one does an experiment of extracting N grains of k different orientations from a metal gate, replacing the extracted grain after each trial. Grains from the same orientation are equivalent. Denote the variable which is the number of extracted grains of orientation i ($i = 1, \dots, k$) as X_i , and denote as P_i the probability that the grain will be oriented i for a given extraction. This is a typical case of a multinomial distribution and its probability mass function (i.e. the function that gives the probability that the discrete random variables are exactly equal to a combination of values [85]) is expressed by:

$$\text{Probability}(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{N!}{x_1! \cdots x_k!} P_1^{x_1} \cdots P_k^{x_k} \quad (1.28)$$

For instance, suppose we want to extract 6 titanium nitride grains of 3 different orientations in a metal gate, and each grain is extracted one by one, replacing the extracted grain after each trial. The number of extracted grains of each orientation is given by X_1 , X_2 , and X_3 and their corresponding probabilities are 0.2, 0.5, and 0.3, respectively. The Eq. 1.28 allows to calculate for example, the probability to obtain $X_1 = 1$, $X_2 = 4$, and $X_3 = 1$, or $X_1 = 2$, $X_2 = 2$, and $X_3 = 2$ as follows:

$$\text{Probability}(X_1 = 1, X_2 = 4, X_3 = 1) = \frac{6!}{1!4!1!}(0.2)^1(0.5)^4(0.3)^1 = 0.1125 = 11.25\%$$

$$\text{Probability}(X_1 = 2, X_2 = 2, X_3 = 2) = \frac{6!}{2!2!2!}(0.2)^2(0.5)^2(0.3)^2 = 0.081 = 8.1\%$$

The corresponding metal work function and the probability of all the possible combinations (i.e. the distribution of probability) can be then calculated from Eqs. 1.26 and 1.28, respectively. By repeating the same experiment but increasing the number of TiN grains, it can be corroborated that this distribution of probability of the work function is described by a Gaussian distribution [82], as predicted by the central limit theorem. Finally, the expected value and the standard deviation of the TiN work function can be estimated from this Gaussian distribution. As a result, it can be proved that the standard deviation of the work function distribution is inversely proportional to the number of grains (which in turn is proportional to the area of the metal gate).

In summary, the local V_{TH} variability due to metal gate granularity (MGG) is related to a different value of work function associated to each crystallite orientation. The number of TiN crystallites in a metal gate is inversely proportional to the standard deviation of the metal work function and therefore contributes to the local V_{TH} variability. The section 2.2.4 in chapter 2 covers the determination of the orientation and size of the TiN crystallites by performing X-Ray Diffraction measurements with specific configurations of diffractometers. Different process conditions during TiN deposition are proposed in section 4.3 in chapter 4 in order to engineer metal gate granularity and therefore to identify and limit this source of V_{TH} variability.

1.4 Conclusion

In conclusion, most of the performance parameters of MOSFET devices are strongly influenced by the equivalent oxide thickness (EOT) and the effective work function (WF_{eff}), which is influenced by the presence of dipoles at both high- κ interfaces and by the density of interfacial fixed charges into dielectrics.

FDSOI devices with undoped silicon channel need metals with effective work functions close to the midgap. Therefore, the work function range to switch from an N-gate to a P-gate in FDSOI devices is shorter than in BULK devices, which makes even more necessary a very fine adjustment of the effective work function of metal electrodes.

The precise thickness control, improved uniformity and low plasma damage achieved by RF-PVD sputtering systems is appropriate for ultra-thin TiN, aluminum and lanthanum monolayers deposition. The *sacrificial* gate-first integration is proposed to carefully incorporate such materials into gate stack in order to reach the WF_{eff} targets for both NFETs and PFETs in FDSOI devices.

Finally, although the main source of local V_{TH} variability related to the Random Dopant Fluctuations has been suppressed in undoped FDSOI devices, the Metal Gate Granularity becomes more important in scaled devices and must be investigated.

Process flow and test methodology for electrical and physicochemical characterization of gate stack

Contents

2.1	Techniques of electrical characterization	46
2.2	Physicochemical characterization	67
2.3	Conclusion	80

This chapter introduces the techniques of electrical and physicochemical characterization used in this thesis. The implementation of specially designed test structures and a simplified process flow for metal gates investigation is detailed in sections 2.1.1 and 2.1.2. The Capacitance-Voltage (C-V) analysis techniques employed to characterize the modulation of the effective work function with the additives incorporation into gate stack and to identify its physical origin are described in section 2.1.3.

Furthermore, the principle of the different spectroscopic techniques used here to study the diffusion of additives is reviewed in section 2.2. In particular, a new methodology based on X-Ray Fluorescence is presented and validated in order to precisely characterize the percentage of effective diffusion of aluminum and lanthanum into high- κ /bottom oxide stack. Finally, a full study of the microstructure of *final* TiN layers is achieved by combining two configurations of X-Ray diffraction in section 2.2.4.

2.1 Techniques of electrical characterization

One of the most important aspects of assessing new processes, materials, devices, and circuits is the development and use of fast, accurate and non-destructive electrical characterization techniques. They allow to determine not only the most important performance parameters of MOSFET devices described in section 1.1.3 but also relevant information about the Metal-Oxide-Semiconductor properties, detailed in section 1.1.2. Carrier doping density, oxide charges density, effective work function, gate dielectric capacitance, flat band voltage and equivalent oxide thickness are some of these MOS properties.

An important peculiarity of electrical characterization is the availability of appropriate test structures [86]. All the test structures used or implemented in this thesis work are described in section 2.1.1. These test structures allow to determine the effective work function, to measure the threshold voltage (V_{TH}) and the local variability of V_{TH} in 14 nm FDSOI devices fully manufactured in an industrial environment.

Many of the test structures for MOSFET device characterization may be implemented at the first level of metal (M1). Hence, the first major test stop for electrical tests occurs immediately post M1. Nevertheless, the processing time to M1 is typically more than half of the total fabrication time. Such fabrication time would delay the research and development of metal gates. In consequence, a simplified process flow for intensive metal gates investigation has been implemented and it is detailed in section 2.1.2.

2.1.1 Test structures

During the development of a new CMOS technology node, test structures placed either outside the product chip area or in the full mask area are used in both the development of the process technology and the optimization of device behaviour. The use of small test structures for analysis and debug of product functionality, reliability and yield has become more important with the increase of Very Large Scale Integration (VLSI) circuits [87]. Standard or specially designed test structures are used to determine the characteristics of CMOS devices and functional circuit blocks for logic and memory applications.

The device or the circuit block to be characterized is called DUT for Device Under Test. A DUT may be a CMOS circuit, a ring oscillator, a resistor, a capacitor, a MOSFET device, etc. A test structure comprises a DUT and a peripheral circuitry required for carrying out the measurements. Indeed, the design of test structures is related to the hardware interface between the test

equipment and silicon wafer, since electrical measurements are performed by directly connecting the microscopic test structure on a silicon wafer to the macroscopic test equipment. These connections are simplified with sharp metal needles or probe cards contacting I/O metal pads on silicon with dimensions of about $80\text{ }\mu\text{m}$ long and $80\text{ }\mu\text{m}$ wide, as shown in Fig. 2.1. These metallic pads are connected to the different terminals of the device under test and are built over a thick oxide of about 200 nm in order to neglect their parasitic capacitances. In addition, the use of metallic pads is appropriate to avoid degradation of the structures to be tested, especially for devices with thin oxides.

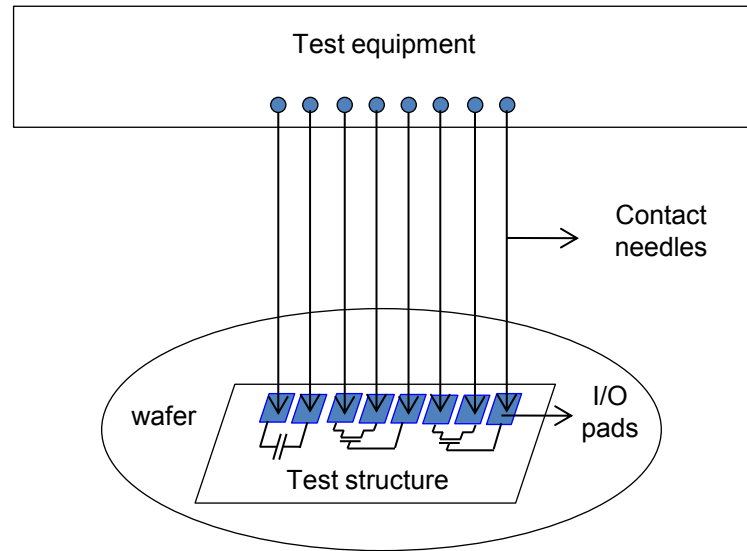


Figure 2.1: Interface between test equipment and test structures on silicon wafer. Sharp metal needles or probe cards connect test equipment to the test structures through I/O metallic pads.

The layout of a test structure is its top view representation in terms of planar geometric shapes, which correspond to the patterns of metal, oxide, or semiconductor layers that constitute the components of the test structure. The layout view typically includes the components, metal routing tracks, vias and electrical pins. All these components are placed and connected each other in order to minimise both the area required and the negative effects of layout parasitics. Finally, the layout must pass a series of checks to ensure that the circuit is both functional and manufacturable. When all verification is complete, the data is translated into an industry standard format, typically GDS (Graphic

Database System), and sent to a semiconductor mask shop¹. The data are there converted into another format to generate the photomasks used in photolithographic process of semiconductor device fabrication.

The test structures of this thesis work consist of either MOSFET devices on SOI or bulk areas (hybrid devices), or a pair of MOSFET devices, or even a matrix of MOSFET devices for V_{TH} measurement, local V_{TH} variability characterization and short-channel MOSFET studies, respectively. These electrical test structures are integrated in the full mask area of the 14 nm technology mask and are detailed in the next sections 2.1.1.1, 2.1.1.2 and 2.1.1.3.

2.1.1.1 Hybrid devices

The hybrid devices are manufactured over bulk areas. These bulk areas are formed on silicon on insulator (SOI) substrates by etching the buried oxide (BOX), and then growing epitaxial silicon before Shallow Trench Isolation (STI) patterning. These specific bulk areas are defined in the layout of the test structure by adding a CAD² mask level called NOSO (for No Silicon On Insulator) around the test structures and their pads. The hybrid devices are currently used for variable MOS capacitors, vertical bipolar transistor and ESD devices integration [88] and were used here as practical tool for electrical analysis and device characterization. The combination of the hybrid devices with a specific process (beveled oxide) authorizes an accurate evaluation of effective work function. The process and the validation of the beveled oxide will be described in section 2.1.2.2.

On the other hand, the key steps of the STI process involve the etching of silicon to form trenches, the deposition of one or more dielectric materials (such as silicon dioxide) to fill the trenches, and the removal of the excess dielectric using a technique such as chemical-mechanical planarization (CMP). As a result, quite flat SOI/bulk transitions are achieved after STI patterning, as shown in TEM pictures of Fig. 2.2.

¹A semiconductor mask shop is a factory which manufactures photomasks for use in the semiconductor industry.

²CAD = Computer-aided design. CAD standards are used in industry to increase the productivity of designer, improve the quality and interchange of designs and create a database for manufacturing.

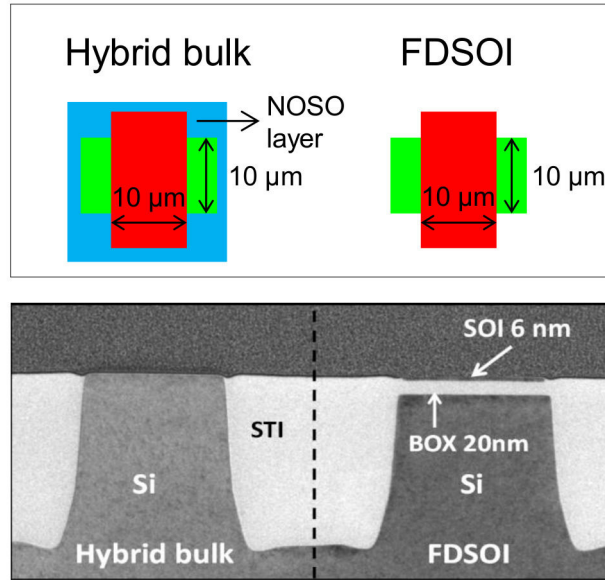


Figure 2.2: Test structure layout of Hybrid bulk devices and FDSOI devices embedded in the 14 nm technology mask. TEM pictures of the corresponding Hybrid bulk areas and FDSOI areas after STI [67].

2.1.1.2 Short channel test structures

C-V characterizations used to extract EOT and WF_{eff} are usually performed on $10 \times 10 \mu m^2$ MOSFET structures, and it would be useful to get such C-V characteristics at nominal 14 nm FDSOI dimensions.

Limitation of current test methodology

We report on Fig. 2.3, C-V characteristics for MOSFET devices with short gate lengths. The impact of parasitic capacitances (interconnect, overlap capacitances) is obvious, leading to strong discrepancies from the expected C-V characteristics. In consequence, the EOT and WF_{eff} extraction through comparison between experiment and simulation would result in inaccurate values. As the MOS capacitance is assumed to be negligible under V_{TH} bias, one strategy to deal with such parasitic capacitance would be to impose the capacitance to zero at reference bias. Figures 2.4 and 2.5 corresponds to such calculation considering this reference bias either equal to the bias at minimum capacitance (around -1.3 Volt) or to 0 Volt, respectively. Fig. 2.4 leads to large parasitic capacitance around V_{TH} and Fig. 2.5 would conclude to very large maximum capacitance discrepancies, and thus EOT variation.

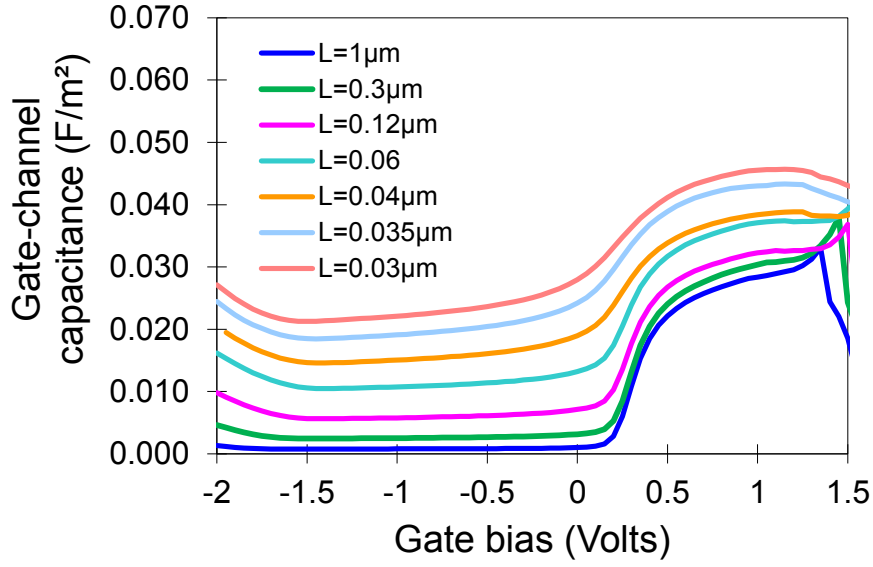
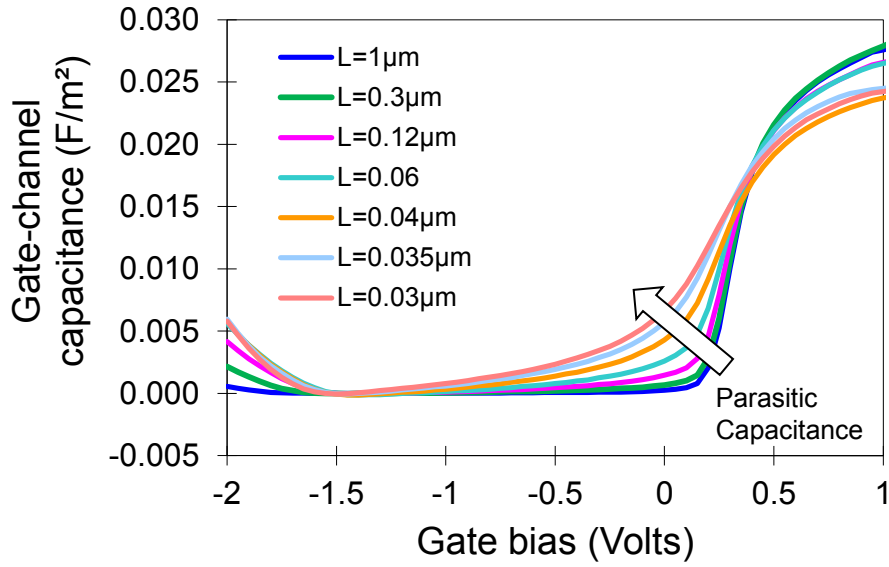


Figure 2.3: C-V characteristics for various gate lengths.

Figure 2.4: Same set of C-V characteristics, with all C-V at 0 V for $V_g = -1.5$ V.

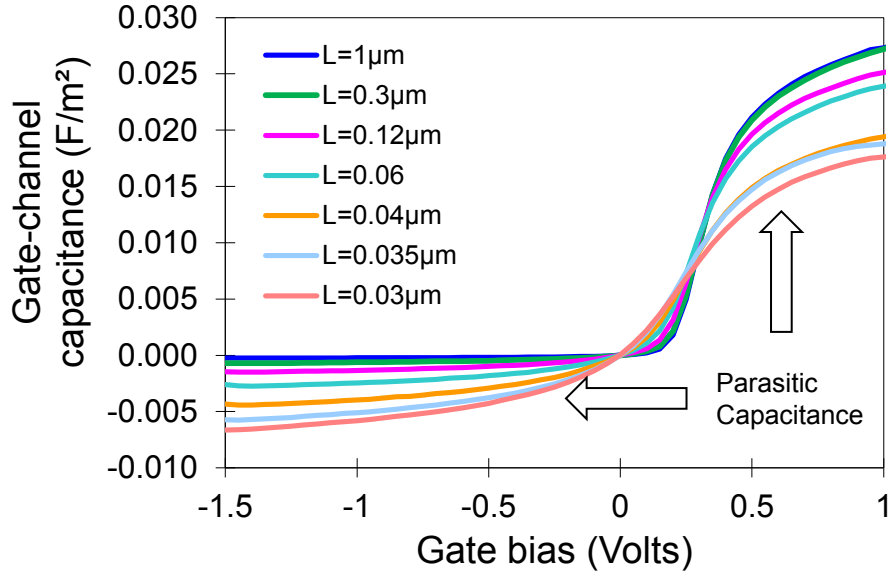


Figure 2.5: Same set of C-V characteristics, with all C-V at 0 V for $V_g = 0$ V.

New test structure layout

Dealing with parasitic capacitance is therefore a critical issue, that is the reason why a new test structure has been proposed and patented to eliminate the problem of parasitic capacitances. Considering a MOS capacitance with rather small dimensions as it is the case with nominal devices, the test structure consists of two test structures, each one having two MOS transistors in parallel, as illustrated in Fig. 2.6. The first test structure (Device 1) can be characterized by length and width of its transistors. The second test structure (Device 2) is a similar test structure with same interconnects, but having an inversion of gate lengths. As a result, the transistors have same perimeters but different areas. As we subtract the measurement of the second test structure to the measurement of the first test structure, it corresponds to the measurement of an equivalent MOS capacitance inside the first MOS of the first test structure whose length (respectively width) is the difference of the MOS lengths (respectively MOS widths). In other words, the equivalent MOS capacitance allows the study of a zone located at distances $L_2/2$ and $W_2/2$, respectively, of gate edges and active edges of the first structure (Device 1). These patented test structures have been implemented in the 14 nm technology mask at STMicroelectronics, with a large set of lengths and widths, as indicated in Table 2.1.

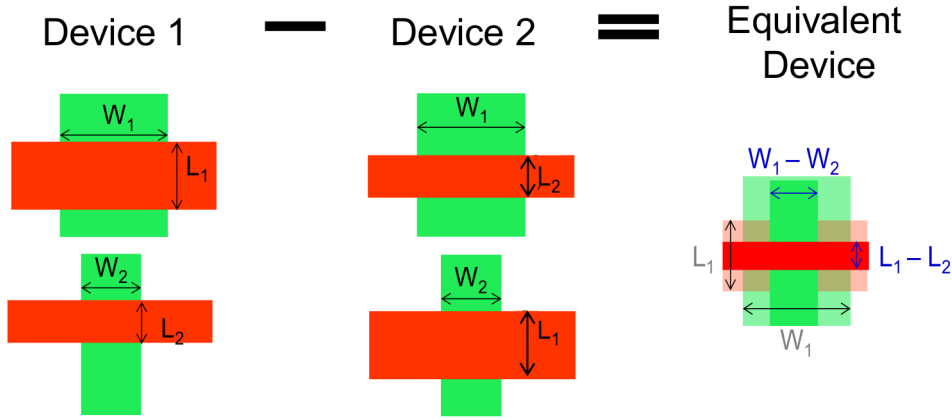


Figure 2.6: Description of the test structure for short-channel MOSFET. Each device has two MOS transistors in parallel. The second device is similar to the first one, with inversion of gate lengths.

DUT	1	2	3	4	5	6	7	8	9	10	11
W_1	12	10.08	10.08	10.08	10.08	10.08	10.08	10.08	10.08	10.08	10.08
L_1	2	1.78	1.53	1.28	0.53	0.33	0.23	0.13	0.09	0.06	0.034
W_2	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
L_2	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03

Table 2.1: Description of various set of lengths and widths embedded in the specially designed test structure on the 14 nm technology mask. All the dimensions are expressed in μm .

Validation

We report in Figs. 2.7 and 2.8 the set of CV for these different tests structures. The result for the difference of corresponding test structures is reported in Fig. 2.9, highlighting the interest of these structures. Indeed, we notice that the parasitic capacitances are almost automatically eliminated without any adjustment. We notice a remaining capacitance for the smallest device, it can be eliminated by forcing the capacitance at 0 at $V_g = 0$. For this smallest device, remaining parasitic capacitance illustrates the limit of the technique to remove parasitic capacitances, even if the two test structures are designed with similar design and interconnect, discrepancies can still exist on the parasitic capacitances. Here it corresponds to a difference of 13 fF; such low capacitances become visible as we try to measure very low capacitances, the maximum capacitance for DUT11 correspond to 0.2 pF after difference. The solution for such small dimension would be to increase the number of transistors in parallel.

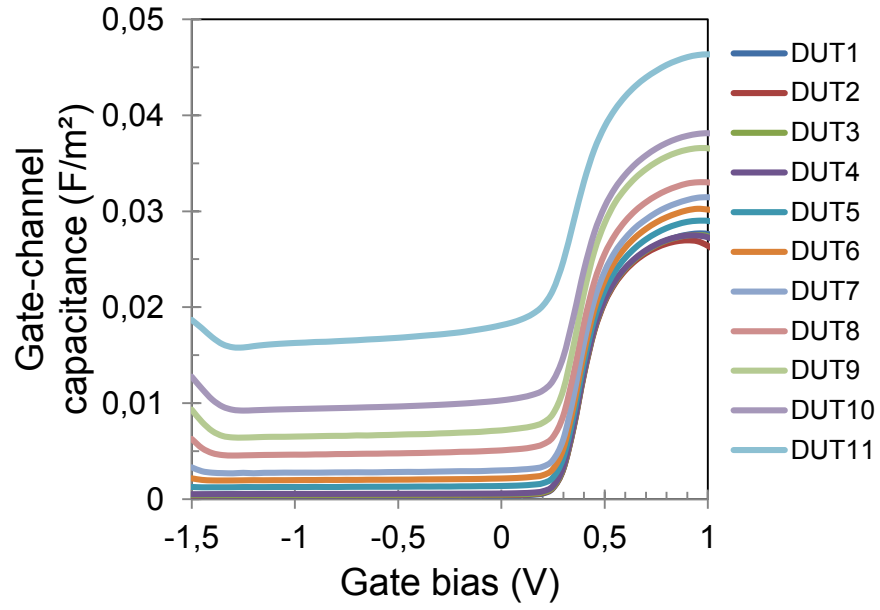


Figure 2.7: Set of C-V characteristics measured on the first test structure.

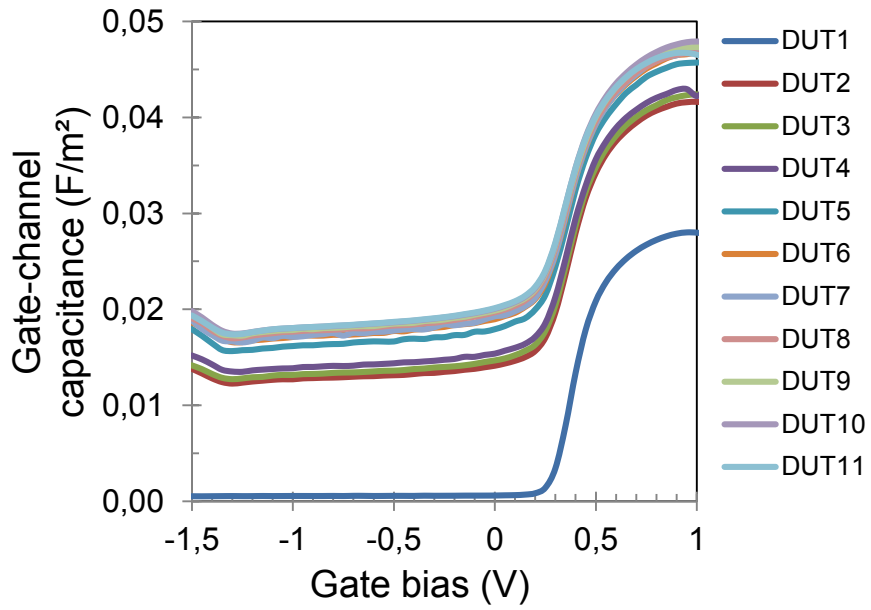


Figure 2.8: Set of C-V characteristics measured on the second test structure.

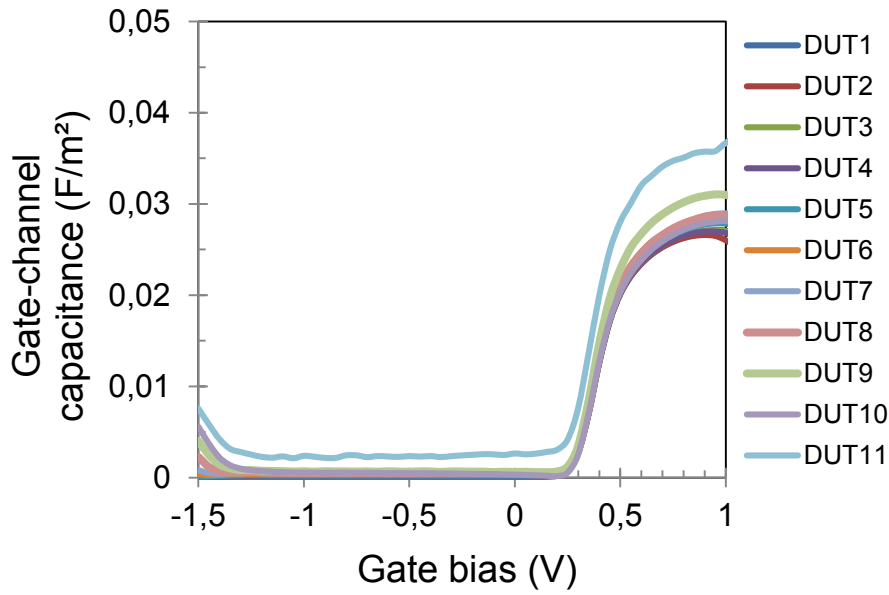


Figure 2.9: Resulting set of C-V characteristics corresponding to the difference of the two test structures.

Nevertheless for most of the capacitances, parasitic overlap capacitances are successfully eliminated. Two capacitances lead to EOTs slightly smaller. Such discrepancy on the smallest device is not so surprising; we expect to test an effective area of 10μ by 4 nm, and a gate length variation of only 1 nm would explain the discrepancy. The other device that leads to a smaller EOT is the one with gate length of 90 nm minus 30 nm, the reason in that case is still to investigate, and it could be related to discrepancies in the Optical Proximity Correction.

2.1.1.3 Matching test structures

Local fluctuations of V_{TH} were measured on matching test structures that consist of identically designed devices pairs, placed the one from the other at minimal spacing authorized by the Design Rules Checking (DRC), and individually addressable with separated drain, source, and gate, as shown in Fig. 2.10. Several identical MOSFET pairs were designed with different surface area dimensions between $0.01 \mu\text{m}^2$ and $100 \mu\text{m}^2$ in order to normalize the random fluctuation with respect to the MOSFET size.

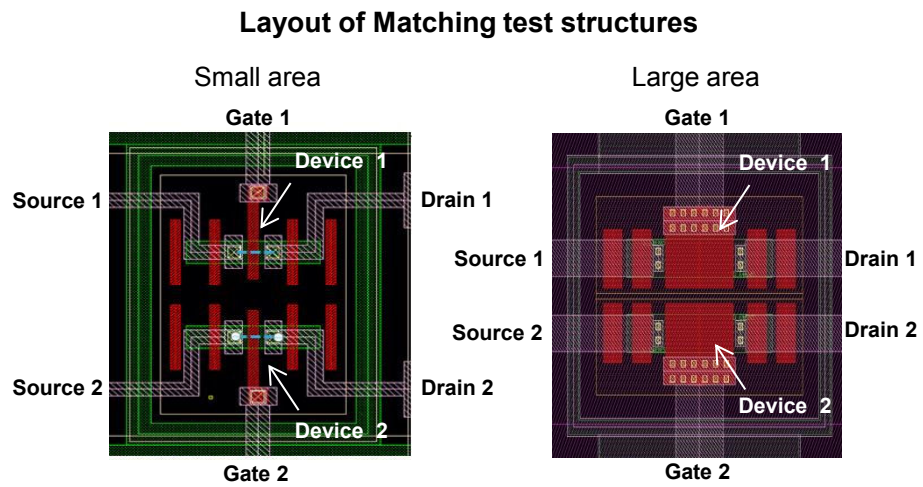


Figure 2.10: Layout of Matching test structures. Identically designed devices pairs are placed at quasi-minimal spacing the one from the other. These test structures are fabricated at different area dimensions.

2.1.2 Process flow simplification for nominal devices and specific research wafers

Most part of the integrated circuits being manufactured today are processed on 300 mm Si wafers. The fabrication process is similar regardless of the number of transistors on the chip. Integrated circuits are basically made by several sequences of thin films deposition (insulators, metals, photoresist) and their selective removal (by etching, polishing) from certain areas of the deposited thin films, defined in the layout of the test structure. Atomic species can also be introduced in the wafer by implantation or diffusion. Once photomasks have been produced in agreement with the layout rules, the information from the layout design must be transferred from the photomask to the wafer. This is done using a process called *photolithography*. The collection and ordering of these sequences for making a useful product is called a technology. The elementary MOSFET devices of the 14 nm FDSOI technology is the result of around a thousand of manufacturing steps. Elementary MOSFET transistors are usually tested after the first level of interconnections (M1), in other words, after more than half of the total fabrication time. Process flow simplification is therefore mandatory in order to intensively study the impact of the metal gate deposition processes and materials on key gate stack electrical parameters. Only the key steps, which influence the gate stack physical properties, are kept in the flow in order to reduce manufacturing steps and R&D cost. The two types of simplified devices used in this thesis for electrical characterization of metal gates are explained in the next sections. They are the nominal devices and the devices with beveled oxides.

2.1.2.1 Nominal devices

As introduced in section 1.3.1.1, the 14 nm FDSOI technology offers multi- V_{TH} solutions for logic and SRAM bitcells as a result of the smart combination of 2 metal gates with two different WF_{eff} (N vs P-type) together with 2 different Well types (N-Wells vs P-Well) and channels (Si vs SiGe), or even by applying a back bias below the BOX. However, only one technological option is enough to study the impact of the metal gate deposition processes and the incorporation of additives such as lanthanum and aluminum on the effective work function and the equivalent oxide thickness. The technological option chosen for this study is the NFET Low V_{TH} devices, which consists of NFET transistors (i.e. source and drain are doped N-type) on a P-type substrate, as illustrated in Fig. 2.11.

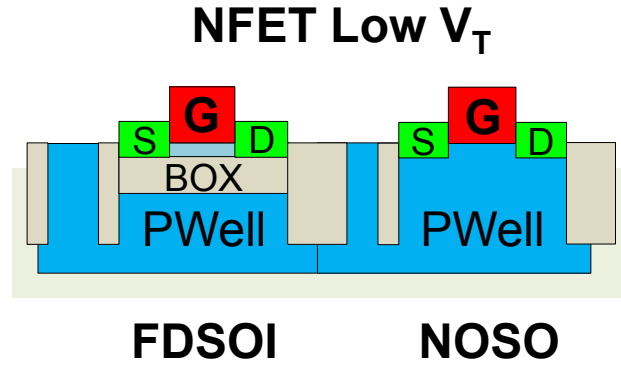


Figure 2.11: NFET Low V_{TH} option in 14 nm FDSOI technology.

Consequently, the simplified process flow consists of the standard 14 nm FDSOI process flow [67], to which some technological bricks such as strained SiGe channel formation, N-Well implants and implants required for other technological options, GO2, resistors brick, in-situ P-type Source/Drains (PEPI S/D) and Spacers for PFET, have been removed. It must be noted that the effective work function of P-type metal gates can be evaluated in NFET devices. In practice, it means that a step of photolithography for P-gate deposition can also be suppressed from the standard 14 nm FDSOI process flow. In contrast, all technological bricks required for manufacturing of NFET Low V_{TH} devices were kept in the scheme integration of the simplified process flow, as shown in Fig. 2.12.

Accordingly, Hybrid bulk definition, Silicon Trench Isolation (STI) and P-Well implants were carried out on silicon on insulator (SOI) wafers before HKMG stack deposition. The gate dielectrics consist of an interlayer dielectric (IL), followed by a 2 nm thick HfO_2 layer deposited by Atomic Layer Deposition (ALD), and decoupled plasma nitridation (DPN) during 50 seconds, as detailed in section 1.2.1. For nominal devices, the interlayer is a 1 nm thick SiON layer. Once the dielectrics are grown, the metal gate stack is deposited following a sacrificial gate-first approach, as defined in section 1.3.2. Devices were completed with Spacers for NFET, N-type S/D formation, S/D dopant activation annealing at 1005 °C, NiPt silicide, contact trench formation, back-end first level interconnects (M1) and a 400 °C forming gas annealing, in order to allow the electrical measurements.

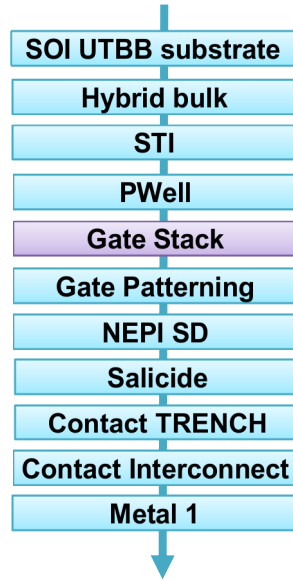


Figure 2.12: Integration scheme of the simplified process flow used to intensively study the impact of the metal gate deposition processes and materials on key gate stack electrical parameters.

2.1.2.2 Devices with beveled oxide

Beveled oxide for Gate Stack Interlayer (IL) as illustrated in Fig. 2.13 is a powerful technique if used with capacitance versus voltage measurements (C-V) to discriminate between interfacial fixed charges and effective work function shift due to dipoles or metal work function change. In this case, the interlayer dielectric consists of a beveled thermally grown SiO_x to get rid of Si/SiO_x interfacial fixed charges $Q_{\text{Si}/\text{SiO}_x}$ in WF_{eff} extraction (Eq. 1.12). However, such process must be coherent with Crolles 14 nm FDSOI facilities and it must overcome several issues: i) a bevel shape with interlayer thickness on the edges close to nominal IL of the technology, ii) significant IL variation along the wafer and iii) a good oxide quality despite the wet etching used for bevel. Indeed, etching in HF can degrade the remaining IL. The goal is to obtain a maximum Equivalent Oxide Thickness (EOT) in the center of the wafer and nominal EOT on the edges. Two processes have been compared. They correspond to a thermal oxidation of 10 nm, followed by a wet circular cleaning performed in one or two sequences. Ellipsometry measurements are reported in Fig. 2.14, and they show satisfactory shapes in both cases.

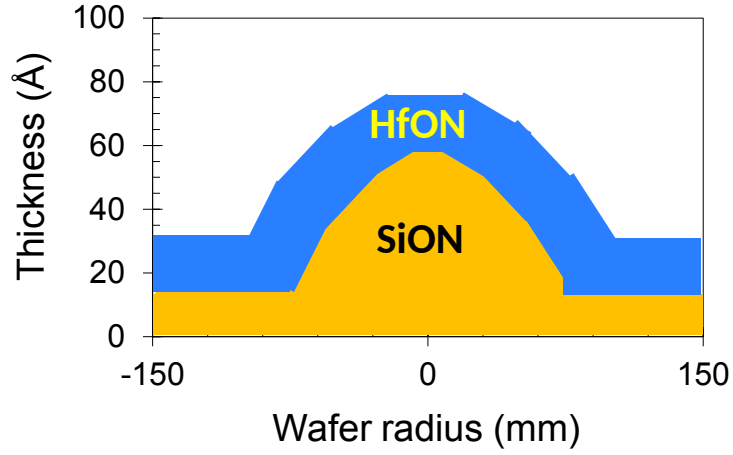


Figure 2.13: Scheme of dielectrics in beveled devices. The beveled devices of the 14 nm FDSOI technology consists of a HfON/beveled thermally grown SiO_x stack.

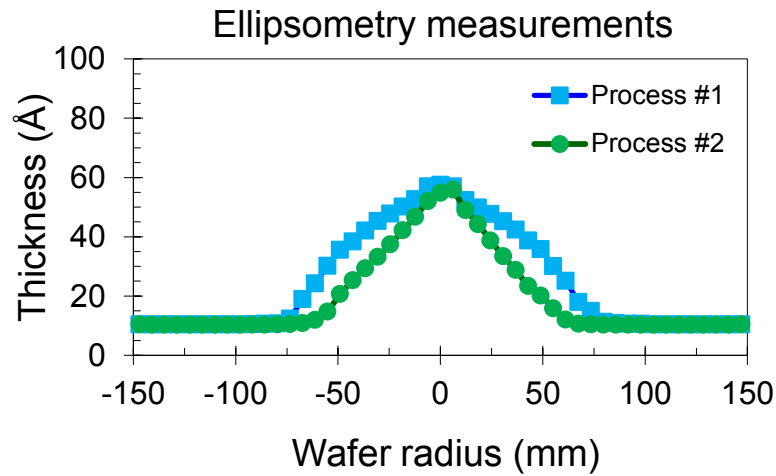


Figure 2.14: Comparison by ellipsometry of two cleaning process for bevel oxide.

This interlayer oxide has been integrated in the simplified process flow, as schematized in Fig. 2.12. After device fabrication, all CV characteristics have an expected shape without any significant evidence of interface states (Fig. 2.15). From each of these characteristics, we can extract an effective work function (Eq. 1.11) and an EOT, by fitting the C-V curves with Poisson-Schrödinger quantum simulations, as it will be described in section 2.1.3.

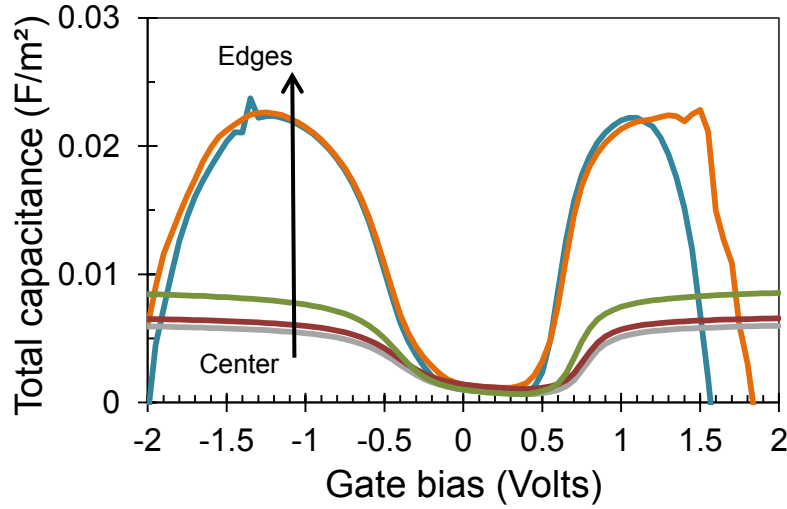


Figure 2.15: Set of C-V characteristics derived from several EOT on a same wafer.

In Fig. 2.16, the extracted Equivalent Oxide Thickness (EOT) is compared to Physical Oxide Thickness measured by ellipsometry. A nice shape agreement can be noticed with a same minimum level on wafer edge and similar EOT in the center.

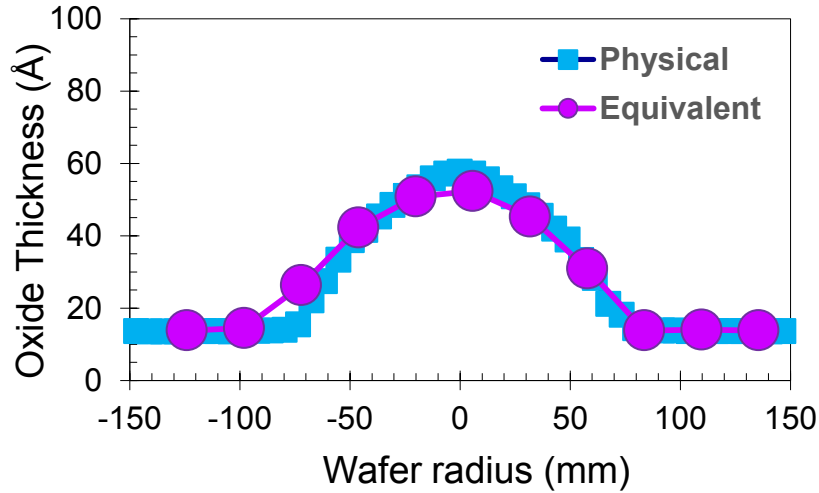


Figure 2.16: Comparison between physical thickness from ellipsometry and EOT along the wafer diameter.

As a result, a WF_{eff} vs EOT can be traced for each wafer (Fig. 2.17). As stated in section 1.1.2, note that the linear behaviour proves the absence of bulk charges in the beveled oxide and the existence of fixed interface charges around $3 \cdot 10^{11}/\text{cm}^2$

at the Si/SiO_x interface. Extrapolation at zero EOT allows an extraction of the effective work function without the impact of fixed charges leading to 4.5 eV for this gate stack (TiN/HfO₂/SiO_x/Si). In other words, the extrapolation of WF_{eff} to EOT=0 allows therefore the assessment of only metal work function (Φ_m) and drop voltage induced by interfacial dipoles (δ), without the influence of SiO_x/Si interface fixed charges [9].

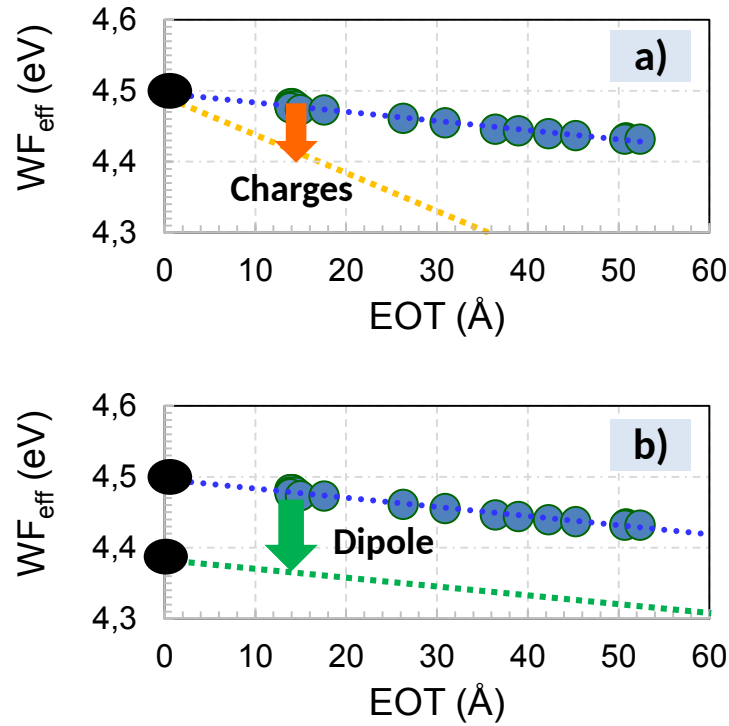


Figure 2.17: a) WF_{eff} shift is explained by fixed charges; b) WF_{eff} shift is explained by a dipole effect or metal work function change.

When two different wafers, each one with a specific process or dopant (La, Al, N) in *sacrificial* metal gate stack, are compared by combining C-V techniques and devices with beveled oxides, the origin of the resulting WF_{eff} shift can thus be identified from the plot WF_{eff} vs EOT. For instance, if the y-intercept (WF_{eff} at EOT = 0) is unaltered, but the slopes of the 2 curves are different, then the WF_{eff} shift will be explained by interfacial fixed charges induced by the process or dopant in metal gate stack (Fig. 2.17a). On the contrary, if the y-intercept is modified but the slope remains unchanged, then the WF_{eff} shift will be explained by a dipole effect or a work function change of *final* TiN induced by the process or dopant in metal gate stack, as schematized in Fig. 2.17b. Obviously, in practice the WF_{eff} shift may be a combination of both effects. If both the slope and the y-intercept are modified, the WF_{eff} shift will be explained by the sum of both

interfacial fixed charges and a dipole effect, or a work function change of *final* TiN induced by the process or dopant in metal gate stack. Assuming that the final TiN microstructure is unaltered in the particular case of evaluation of additives incorporation in a *sacrificial* gate-first approach (section 1.3.2), the WF_{eff} shift will be explained only by fixed charges or dipoles. In general, this hypothesis is acceptable, except in case of different recrystallization of TiN during S/D dopant activation annealing due to a modification of Hf-based oxides morphology.

2.1.3 Measurements and electrical parameters extraction

2.1.3.1 Experimental set up for C-V measurements

Capacitance vs Voltage (C-V) technique is one of the most common methods to evaluate new processes, materials, devices, and circuits. Here, it has been used to evaluate the impact of the incorporation of additives into metal gate stack. When a voltage is applied, the capacitance varies due to accumulation, depletion and inversion of charge carriers in the silicon substrate near the SiO_2 layer, as illustrated in Fig. 2.18.

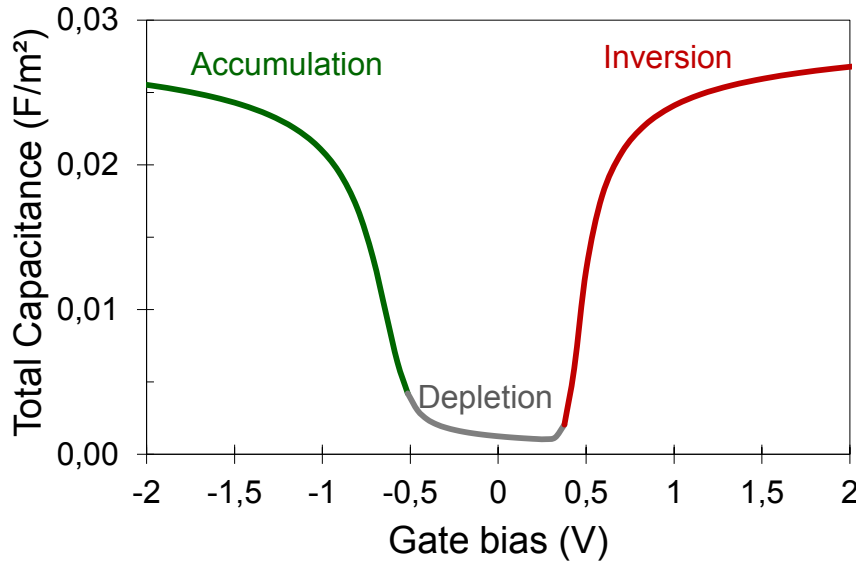


Figure 2.18: Simulation of the total capacitance of a MOSFET structure with a p-type substrate.

To measure C-V curves, the MOSFET is typically connected to a C-V analyzer, such as the Agilent 4284A or 4980A, as shown in Fig 2.19. Different test configurations are commonly employed (Fig. 2.20) to evaluate the different regimes of the MOSFET capacitances. The C-V analyzer applies a high frequency (between 10 kHz to 1 MHz) AC signal, which is superimposed on a

DC bias sweep, to the gate contact or pad by a probe needle. The signal is picked up through the body terminal or the substrate via another probe needle or the prober chuck, respectively.

The gate-channel capacitance (C_{gc}) is measured by adding a small signal voltage on the gate and measuring the small signal current at the source and drain terminals. The high terminal of an LCR-meter is connected to the gate and the low terminal to the source and the drain. The body terminal is grounded. In general, C_{gc} authorizes the measurement of only the inversion capacitance in MOSFET devices. The gate-body capacitance (C_{gb}) is measured by adding a small signal voltage on the gate and measuring the small signal current at the substrate terminal. The high terminal of an LCR-meter is connected to the gate and the low terminal to the body contact. The source and drain terminal are grounded. In general, C_{gc} authorizes the measurement of only the accumulation and depletion capacitances in MOSFET devices. The total gate capacitance (C_{tot}) is measured by adding a small signal voltage on the gate and measuring the small signal current at the source, drain, and substrate terminals. The high terminal of an LCR-meter is connected to the gate and the low terminal to the source, drain, and body contacts.



C-V analyzer Agilent 4980



Full auto wafer probe station



Keithley 630 + 300 mm full auto prober
Parametric tester

Figure 2.19: C-V analyzer Agilent 4980A connected to a 300 mm wafer probe machine for manual electrical measurements. Keithley 630 + full-auto prober for parametric tests.

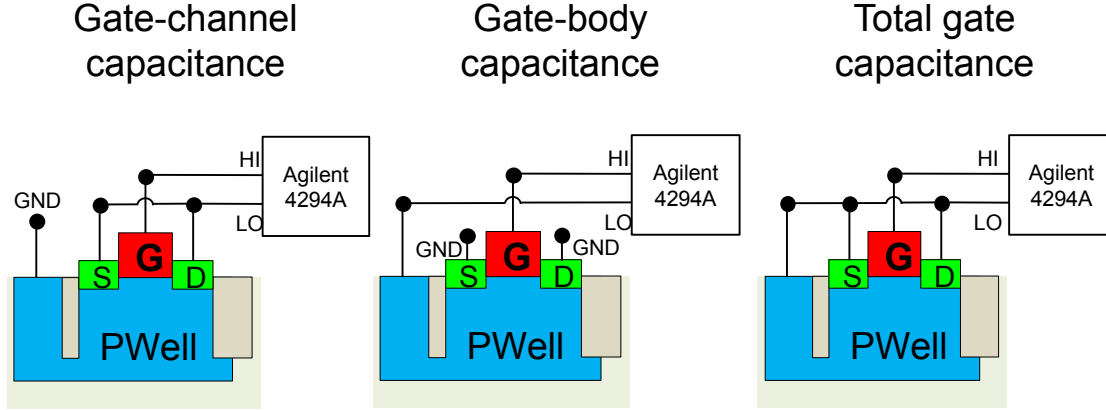


Figure 2.20: Most common C-V configurations.

2.1.3.2 Extraction of electrical parameters from C-V characteristics

In this section, we introduce the methodology for an accurate extraction of the equivalent oxide thickness (EOT) and the V_{fb} from the experimental C-V characteristics. This method allows the extraction of the EOT and V_{fb} in accumulation regime as well as in inversion regime, even if only a small part of the C-V characteristics is relevant [89]. Indeed, high leakage current or interface state densities (D_{it}) can lead respectively to a wrong evaluation of EOT or V_{fb} [90]. Furthermore, analytical solutions such as Maserjian function [91], based on the $C(V)$ derivative, are inappropriate in case of stretched C-V curves (non ideal C-V curves) [90].

The capacitance response of the gate stack is actually the sum of three capacitors connected in series: C_g , C_{ox} and C_{sc} . The use of a metal layer instead of polysilicon as gate implies $C_g = 0$, leading to the following equation:

$$\frac{1}{C_{exp}(V_g)} = \frac{1}{C_{ox}} + \frac{1}{C_{sc}} = \frac{EOT}{\epsilon_{ox}} + \frac{1}{C_{sc}(Q_{sc})} \quad (2.1)$$

The C_{sc} dependency with Q_{sc} as well as the relationship between the charges in the semiconductor and the potential at the semiconductor surface $Q_{sc}(V_{sc})$ can be described by the Poisson equation for electrostatics, according to classical electrodynamics. Nevertheless, for ultra-thin oxide thickness and low field strengths, such interactions are better described by quantum Poisson-Schrödinger simulations. Indeed, for an accurate extraction of these electrical parameters on MOSFET devices with nanoscale oxides, one should take into account the effects of quantum confinement at the dielectric/substrate

interface. These effects become significant when the thickness becomes comparable to the De Broglie wavelength of the carriers. The De Broglie wavelength for a free electron is equal to 1.2 nm at room temperature [92]. The LETI developed a one dimension Poisson-Schrödinger solver which leads to the charge (Q) and capacitance (C) dependences versus the potential at the semiconductor surface (V_{sc}) for various substrate doping levels (N_{sc}) [92]. The results are stored in a database, which is used for C-V analysis [93].

After C-V measurements of a set of MOSFET devices, from a fast visual analysis, two voltages V_{g1} and V_{g2} are defined in order to delimit the relevant part of the C-V characteristic for electrical parameters extraction. For instance, V_{g1} can be the maximum accumulation bias non-altered by high leakage currents and V_{g2} the minimum accumulation bias non influenced by interface states. At these bias conditions, Eq. 2.1 can be rewritten as follows :

$$\frac{\varepsilon_{ox}}{C_{exp}(V_{g1})} - \frac{\varepsilon_{ox}}{C_{sc}(Q_{sc1})} = EOT = \frac{\varepsilon_{ox}}{C_{exp}(V_{g2})} - \frac{\varepsilon_{ox}}{C_{sc}(Q_{sc2})} \quad (2.2)$$

$$\frac{1}{C_{exp}(V_{g1})} - \frac{1}{C_{exp}(V_{g2})} = \frac{1}{C_{sc}(Q_{sc1})} - \frac{1}{C_{sc}(Q_{sc1} - \Delta Q_{sc})} \quad (2.3)$$

The challenge is to find the substrate charge densities Q_{sc1} and Q_{sc2} that respectively correspond to the gate bias V_{g1} and V_{g2} . The difference between Q_{sc1} and Q_{sc2} (ΔQ_{sc}), can be obtained by integrating C_{exp} from V_{g1} to V_{g2} (Eq. 2.4).

$$\Delta Q_{sc} = \int_{V_{g1}}^{V_{g2}} C_{exp}(V_g) dV_g \quad (2.4)$$

By using the calculated relationships $Q_{sc}(V_{sc})$ and $C_{sc}(V_{sc})$ from the Poisson-Schrödinger database, Q_{sc1} can be determined from Eq. 2.3. Indeed, Q_{sc1} takes the value required to equalize or "fit" both sides of the equation. EOT is then obtained from Eq. 2.2.

The flat band condition (V_{fb}) is finally calculated from the Eqs. 2.5.

$$V_{g1} = q\Phi_m - q\Phi_{sc} + V_{sc1}(Q_{sc1}) - \frac{Q_{sc1}EOT}{\varepsilon_{ox}} \quad (2.5)$$

$$V_{fb} = q\Phi_m - q\Phi_{sc} \quad (2.6)$$

and its value corresponds to :

$$V_{fb} = V_{g1} - V_{sc1}(Q_{sc1}) + \frac{Q_{sc1}EOT}{\varepsilon_{ox}} \quad (2.7)$$

where the V_{sc1} value associated to V_{g1} is again taken from the

Poisson-Schrödinger database.

In summary, the methodology for EOT and V_{fb} extraction on nominal or beveled-SiO_x devices consists of: 1) the measurement of the C_{tot} and C_{gc} capacitances, 2) the suppression of parasitic capacitances, 3) the choice of a part of the C-V curve free from leakage and interface defects issues [89], 4) EOT and V_{fb} extraction by fitting experimental high-frequency C-V characteristics with Poisson-Schrödinger quantum simulations, and 5) EOT and V_{fb} extraction for different dies along the wafer either for a statistical effect or for V_{fb} dependence on EOT in the case of devices with beveled oxides.

2.2 Physicochemical characterization

Quality control and monitoring in volume production are key for successful operation of any semiconductor fabrication facility. At STMicroelectronics, statistical tools such as control charts are used to monitor an individual industrial process and feed back its performance to the operators responsible for that process. Several measurements and characterization techniques are available today for research purposes, but a relatively small number of such techniques are used for in-line monitoring in volume production. Some of the most common metrology techniques to measure the thickness of metallic thin films are the four-point probe resistance measurement and the X-Ray Fluorescence.

2.2.1 Four-point probe resistance measurement

As its name implies, this technique uses four in-line probe tips to measure sheet resistances. The measurement of the sheet resistance of the thin film is carried out by passing a current through two outer probes and measuring the voltage drop across the inner probes, as illustrated in Fig. 2.21.

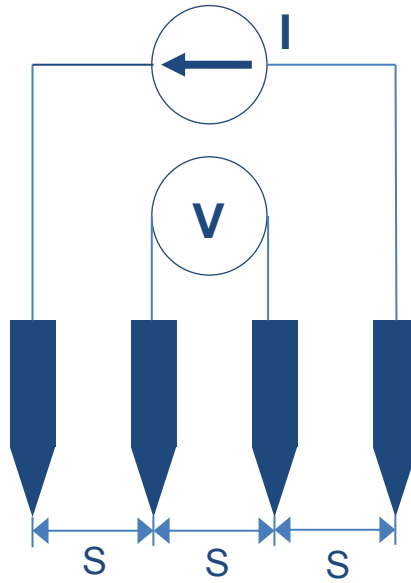


Figure 2.21: Probe configuration for a four-point probe resistance measurement.

The resistance R of a rectangular film of conducting material of bulk resistivity ρ , thickness t , length L , and width w is given by :

$$R = \frac{V}{I} = \frac{\rho L}{tw} = \frac{R_s L}{w} \quad (2.8)$$

The sheet resistance (R_s) is a special case of resistivity for a uniform sheet thickness. Assuming that the dimensions of the sample are much greater than the probe-tip spacing (S), the sheet can be considered as an infinite sheet. Therefore, the potential difference between the two inner points ($\Delta\phi$) is given by the equation below [94] :

$$\Delta\phi = V = \frac{IR_s}{\pi} \ln 2 \quad (2.9)$$

and the sheet resistance (R_s) is thus expressed as :

$$R_s = \frac{V}{I} \frac{\pi}{\ln 2} \approx 4.5324 \frac{V}{I} \quad (2.10)$$

In typical usage, the current is set to 4.5324 mA so that R_s is simply the voltage reading in mV. The units of sheet resistance are ohms per square (Ω/sq), which is dimensionally equal to an ohm, but is exclusively used for sheet resistance. Sheet resistance could be taken out of context and misinterpreted as a resistance of 1 Ω , whereas sheet resistance of 1 Ω/sq cannot thus be misinterpreted. The bulk resistivity in $\Omega\cdot\text{cm}$ is calculated by multiplying the sheet resistance by the film thickness t in cm :

$$\rho = t \frac{V}{I} \frac{\pi}{\ln 2} \approx 4.5324t \frac{V}{I} = R_s t \quad (2.11)$$

One of the limitations of this technique is that the probe tips can drill holes. For films thinner than $\approx 50 \text{ \AA}$, the contact can be realized with underlayer and in consequence, the measurements are indicative of the substrate rather than the thin film itself. For this reason, all measurements were carried out on thin films deposited on thicker SiO_2 oxides. In addition, nowadays in semiconductor industry, the deposited metal films are so thin that surface-scattering and grain morphology effects must be accounted, and it leads to a resistivity of thin PVD films significantly larger than bulk values found in references.

2.2.2 X-Ray Fluorescence

X-Ray Fluorescence (XRF) is an alternative thickness measurement technique which is non-destructive and independent of resistivity effects. Here, this technique has also been proposed and validated to accurately measure the effective diffusion of additives into the gate stack after diffusion annealing. The sample under analysis is illuminated by X-rays (or gamma rays), which results in the excitation of core level electrons to excited states or in the ejection of one or more electrons from the inner orbitals (shells) of the atom. The radiative decay of these electrons from the excited states back to their respective ground states results in the emission of *fluorescence* or secondary X-rays that are characteristic of the energy levels of each atomic species.

The emitted photon energy is equal to the energy difference of the two orbitals involved. The main transitions are given names: an $L \rightarrow K$ transition is traditionally called K_α , an $M \rightarrow K$ transition is called K_β , an $M \rightarrow L$ transition is called L_α , and so on. The principle of the X-Ray Fluorescence is illustrated in Fig. 2.22.

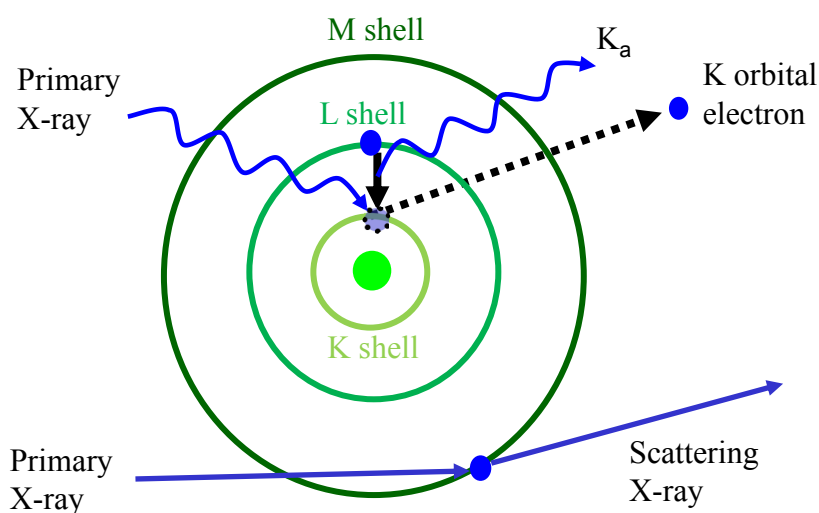


Figure 2.22: Principle of the X-Ray Fluorescence.

2.2.2.1 Advantages and limitations

The basis of this powerful technique is that the intensity of fluorescence radiation is directly related to the amount of each element in the sample. XRF signal for each atomic species correlates directly to the number of atoms present. Therefore, the XRF signal can be used as a direct measurement of the thickness of metals and alloys. The primary strength of XRF as a measurement technique lies in its ability to assess thin film thicknesses and alloy concentrations independently of any numerical modeling techniques. The main challenge in measuring very thin layers is the long acquisition time needed for the XRF detector to acquire a statistically significant number of fluorescence counts. Moreover, for rigorous quantitative results, the XRF signal must be calibrated against known thickness standards.

The depth of analysis of the XRF depends on the sample matrix, and is basically determined by two factors: the depth of penetration of the primary X-Ray beam into the sample, and the escape depth from which fluorescent X-Rays can be detected. The first one is in practice handled by the voltage of the X-Ray generator and the angle of incidence of X-Rays into the sample. The

second one directly relates to the elements being detected. Indeed, lighter elements such as Na, Mg, Si, etc. have low X-Ray energies, and thus it will be difficult to detect even at relatively small depths, whereas heavier elements such as Cu, Ag, Au have much more energetic X-Rays which will be able to pass through large distances within the sample. In contrast, heavier elements absorb strongly X-Rays and thus the chances of X-Rays escaping from deep within the sample are reduced. In other words, heavy elements (i.e. energetic fluorescent X-Rays) will be detectable relatively deep (several millimeters) within a sample matrix mainly composed of light elements (i.e. low absorption coefficient). On the contrary, light elements (i.e. low energy fluorescent X-Rays) will be detectable only a few micrometers or less within a sample composed mainly of heavy elements (i.e. strong absorption coefficient). Finally, XRF is also restricted when it comes to measure elements with low atomic numbers, typically $Z < 11$, due to their weak fluorescence.

Although XRF can explore deeply the sample with a depth of analysis of a few micrometers to the whole thickness of the wafer (≈ 700 micrometers), it cannot be used for depth profiling. It can measure the thickness of stacked films but typically cannot tell which one is on top of the other or where the films are located. For substrates and thin films composed of similar materials (typically the case of Poly-Si on silicon substrate), it is also not possible to differentiate between the spectral signal coming from the substrate and the signal from the layer on the top. For example, signal of silicon deposited as a capping layer over a TiN/HfO₂/SiO₂/Si stack overlap with the signal of interlayer and substrate. Nevertheless, this characteristic can be exploited to explore the whole gate stack and to accurately determine the effective diffusion of gate additives such as lanthanum and aluminum, as it will be detailed in the next section.

2.2.2.2 Methodology for the characterization of the diffusion of gate additives

XRF measurements were carried out on a RIGAKU WaferX300 equipment, with X-ray tube with a rhodium target, operated at 40 kV, 90 mA and a spot size of 40 mm. XRF measurements were performed on at least 9 points on blanket wafers with nominal dielectrics thicknesses a) after sacrificial gate stack deposition and b) after diffusion annealing and sacrificial gate stack removal in order to evaluate the effective dose incorporated into HfON/SiON stack as a function of the as-deposited dose and the bottom TiN thickness (d) in the sacrificial gate. It is worth noting that both measurements are necessary in order to determine the percentage of diffusion (i.e. the ratio eff/as-dep dose) for a given element. The following scheme illustrates the process steps in which the measurements were carried out.

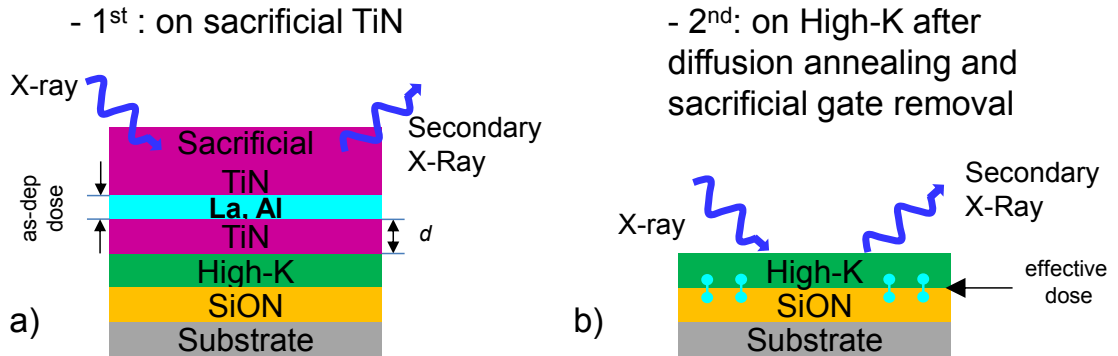


Figure 2.23: XRF measurements on blanket wafers 1) after sacrificial gate stack deposition and 2) after diffusion annealing and sacrificial gate stack removal.

The La doses in at/cm^2 were determined from the characteristic spectral line of lanthanum corresponding to a M-L transition (L_α). The Al doses in at/cm^2 were determined from the characteristic spectral line of aluminum corresponding to a L-K transition (K_α). In addition, the titanium K_α lines were also systematically analyzed to ensure that sacrificial gate had been correctly removed.

2.2.3 Secondary Ion Mass Spectrometry

The Secondary Ion Mass Spectrometry (SIMS) was used in this thesis as a complementary technique of characterization. The surface of the sample is bombarded by a focused pulsed beam of high energy ions (typically Ga^+ , Bi^+ , Cs^+ , Ar^+ with an energy of around 25-30 keV). As a result, both neutral and charged (+/-) species from the surface of the sample are ejected or sputtered. The collection and analysis of the positive charged species (also called secondary ions) by a mass spectrometer enables to determine the elemental, isotopic or molecular composition of the surface with a depth resolution < 1 nm. The SIMS technique is one of the most sensitive surface analysis techniques, with detection limits in the ppm/ppb range.

One of the most powerful mass spectrometers used in SIMS is the Time-of-Flight (ToF-SIMS) mass spectrometer. In ToF-SIMS spectrometers, the secondary ions are accelerated towards an extraction grid structure by the application of a potential difference V between the sample and the extraction grid structure, which is separated from the detector by a field free drift region (a linear flight tube or a reflectron) of length L . A secondary ion with mass m and charge q has therefore a kinetic energy given by Eq. 2.12:

$$\frac{1}{2}mv^2 = qV \quad (2.12)$$

In the field free drift region, the secondary ion has a constant speed since there is no electromagnetic fields. As a result, the time of flight in the field free drift region is given by Eq. 2.13:

$$t = L \left(\frac{m}{2qV} \right)^{1/2} \quad (2.13)$$

It can be deduced from Eq. 2.13 that the time of flight vary with the ratio m/q of the secondary ion. The analysis of the distribution of secondary ions as a function of time authorizes the acquisition of a complete mass spectrum. Compared to magnetic sector or quadrupole mass analysers, the ToF-SIMS spectrometer is capable of simultaneously analysing the complete mass range of secondary ions, which is one of its advantages.

Unlike the X-Ray fluorescence, the technique of SIMS authorizes depth profiling measurements. Indeed, a depth profiling of the sample may be obtained by progressively eroding the surface of the sample with the incident ion beam probe and recording sequential SIMS spectra. In modern ToF-SIMS spectrometers, best resolution is achieved by using a separate beam of low energy ions (typically Cs^+ with an energy of around 500 eV - 1keV) to progressively etch a crater in the surface of the sample, combined with short pulses of the primary beam of high energy ions to analyse the floor of the crater.

X-Ray Fluorescence intensities were calibrated with ToF-SIMS measurements performed on reference samples with known concentrations in order to obtain the doses in at/cm^2 .

2.2.4 X-Ray Diffraction

In this section, we first describe the basics of X-Ray diffraction theory in order to determine the position and intensity of peaks in a diffraction pattern. Next, the relative intensities of the expected peaks in the diffraction pattern of a TiN crystal structure are introduced. Finally, we detail the different configurations used to perform diffraction measurements and the determination of crystallite size and orientation from the diffraction pattern.

Theory

The technique of X-Ray diffraction enables the identification of the crystalline structure of solid films, as long as the X-ray waves scattered by the atoms of a crystal, interfere constructively. Constructive wave interference occurs when the difference of path lengths between the scattered waves is an integer multiple of the wavelength of the incident radiation and therefore

satisfies the Bragg equation (Eq. 2.14). In this equation, d_{hkl} is the spacing between lattice planes or interplanar spacing, θ_{hkl} and λ are, respectively, the angle of incidence and the wavelength of the incident radiation. The Miller indices are used to identify different planes of atoms.

$$2d_{hkl} \sin \theta_{hkl} = n\lambda \quad (2.14)$$

In general, most angles do not correspond to a planar spacing, so there are only a few sharp peaks in the diffraction pattern, each corresponding to a different interplanar spacing. X-Ray diffraction analysis can therefore give information about microstructural aspects of thin metal films such as crystal structure, size and preferred orientation of the crystallites. The crystal structure describes the atomic arrangement of a material. Crystalline materials are characterized by the long-range periodic arrangement of atoms. The unit cell is the smallest unit of volume that contains all the structural and symmetry information to build the macroscopic structure of the lattice by translation. The unit cell repeats in all dimensions to fill space and produce the macroscopic grains or crystals of the material. These crystals are usually randomly oriented in polycrystalline metals.

The position and intensity of peaks in a diffraction pattern are determined by the crystal structure. The position of the diffraction peaks (2θ) is determined by the distance between parallel planes of the atoms, according to Bragg's Law (Eq. 2.14). The intensity of the diffraction peaks (I_{hkl}) is determined by the arrangement of atoms in the entire crystal. These intensities vary widely for the different combinations of h , k , and l , depending on the crystal system. The structure factor F_{hkl} describes how atomic arrangement influences the intensity of the scattered beam. It is the resultant of the waves scattered by all the atoms in the unit cell, in the direction of the hkl reflection. Indeed, every atom in the unit cell contributes to every reflection according to its chemical nature and its relative position. Due to this shift in position relative to other atoms, the scattered photons from an atom have a phase shift ($0 \leq \Delta\theta \leq 2\pi$) relative to those from other atoms. As a result, destructive interferences occur between the reflection of different planes. For instance, consider a face-centered cubic (FCC) structure, such as in Fig. 2.24. Note that these central atoms leads to destructive interference for some reflections. In Fig. 2.24, the reflection from the 200 plane is exactly out of phase with the 100 reflection and destructive interference occurs. As a result, 100 reflections are not observed in the diffraction pattern. Furthermore, the hkl indices must be all odd or all even in order for a reflection to be observed in a face-centered cubic structure.

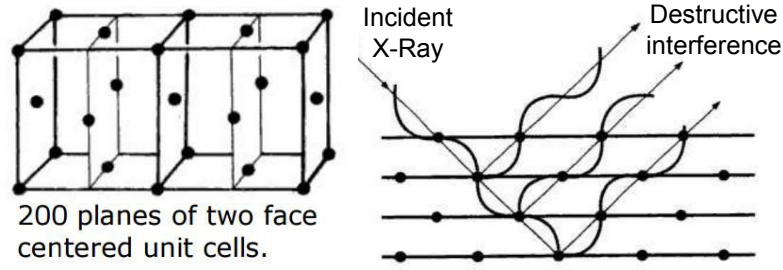


Figure 2.24: Destructive interference between 200 and 100 reflections in FCC structure.

The intensity of a peak in the diffraction pattern is proportional to the square of the structure factor (Eq. 2.15) since it represents the amplitude of the scattered waves. The structure factor F_{hkl} sums the result of scattering from all of the atoms in the unit cell to form a diffraction peak from the (hkl) planes of atoms. In 3-dimensions, the position of atoms is described by $\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$. By using Miller indices, the structure factor is rewritten as follows, where N is the total number of atoms or molecules of the unit cell:

$$I_{hkl} = |F_{hkl}|^2 \quad (2.15)$$

$$F_{hkl} = \sum_{n=1}^N f_n e^{2\pi i(hx+ky+lz)} \quad (2.16)$$

As deduced from Eq. 2.16, F_{hkl} depends not only on the position of each atom, but also on the chemical nature of the atom, characterized by the atomic form factor f_n . In other words, the amplitude of scattered waves is determined by where the atoms are on the atomic planes and what atoms are on the atomic planes. Actually, each atom is composed by electrons, or regions of charge density probability $\rho(\mathbf{r})$ of finding an electron, within a small volume δV . Each electron is located at a distance \mathbf{r} of the center of the atom. Therefore, X-ray photons not only are scattered by each atom of the unit cell, but also are scattered by the electrons of each atom. As a result, the scattered X-ray photons can also be out of phase and destructive interferences can occur due to electrons located at different distances of the center of a single atom. Accordingly, f_n quantifies the efficiency of X-ray scattering at any angle by the group of electrons in each atom.

The maximum value of the form factor is equal to the total number of electrons of the corresponding atom, which is by definition the atomic number Z of the atom in the sample. As the scattering angle 2θ increases, the destructive interference

between the X-rays scattered by different electrons also increases. f values have been tabulated against $\sin\theta/\lambda$ for most common atoms and ions and can be extracted from the International Crystallographic Tables, knowing the atomic number and the diffraction angles, or more precisely $\sin\theta/\lambda$, as shown in Fig. 2.25.

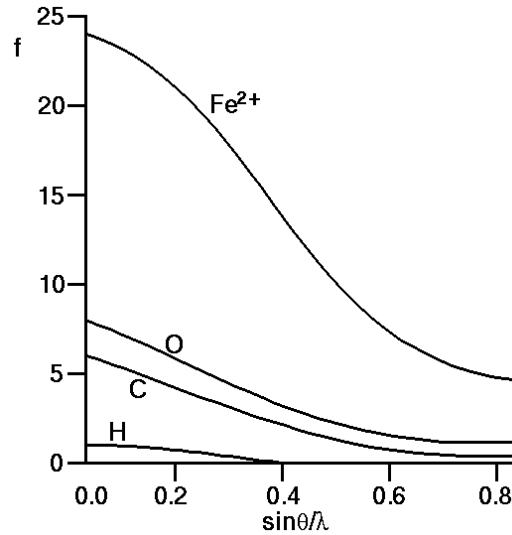


Figure 2.25: Form factor (f) against $\sin\theta/\lambda$ for different atoms [95].

Position and intensity of peaks in TiN crystal structure

The structure factor equation (Eq. 2.16) implicitly shows that intensities of the different diffraction peaks are function of the interplanar spacing between the planes. For a monochromatic X-ray source, the interplanar spacing of each (hkl) combination determines the angles for which scattered waves interfere constructively, according to Eq. 2.14. Then, the atomic form factor f can be extracted from the International Crystallographic Tables. Finally, the intensity of a peak diffraction from an ideal crystal is proportional to the square of $|F_{hkl}|$. It should be deduced that, according to Eq. 2.16, the diffraction intensities of different peaks in the diffraction pattern will not have the same values, even if the amount of crystallites in a sample corresponding to each preferred orientation is physically the same. For example, typical TiN compounds have a cubic NaCl-type crystal structure with a lattice constant (for stoichiometric TiN) of 4.242 Å [95]. The structure factor for cubic NaCl-type crystal structure enables reflections of planes with hkl Miller indices of same parity, but intensities are zero for hkl planes of mixed parity. This means that TiN (111), (200), (220), (311), (222), (400), (331), (420), (422), (511) planes are therefore expected to diffract, but others such as (100), (110), (210) are not expected in the diffraction pattern, as it will be corroborated in section 4.3.

The International Centre for Diffraction Data (ICDD) maintains a database of powder³ diffraction patterns, including the d-spacings (related to angle of diffraction) and relative intensities of observable diffraction patterns. For a monochromatic X-ray source with $\lambda = 1.540\,56\,\text{\AA}$, the most intense diffraction peaks of TiN are summarized in Table 2.2.

h	k	l	I_{hkl}	2θ	d (Å)
1	1	1	72	36.6621	2.449170
2	0	0	100	42.5959	2.120710
2	2	0	45	61.8124	1.499670
3	1	1	19	74.0680	1.278920

Table 2.2: Relative intensities (I_{hkl}) and peak position (2θ) for most intense diffraction planes of TiN [95].

In summary, unlike the X-Ray fluorescence, the signal intensity of (hkl) peaks in X-Ray diffraction pattern is not directly correlated to the amount of (hkl)-oriented crystallites in a sample.

The θ - 2θ and in-plane X-Ray Diffractometer

The essential components of a typical X-Ray diffractometer (like the ones used in laboratory in this work) are an X-ray source, a goniometer, an X-Ray detector, optics such as a monochromator, beam and detector slits, and electronics in order to synchronize the detector counts versus 2θ angle, and therefore generate the diffraction pattern.

The X-ray source is a X-ray tube with a copper anode. A filament is heated to produce electrons which are then accelerated in a vacuum by a high electric field (here 40 kV) towards the copper target. The corresponding electric current is 25 mA. Electrons from one of the inner electron shells of the Cu are then ejected, and electrons from higher atomic levels drop to the vacant level with the emission of X-ray photons characteristic of the difference between the two involved levels. The characteristic line used as source is the Cu $K\alpha$ radiation, which is actually a doublet with two wavelengths $K\alpha_1 = 1.540\,56\,\text{\AA}$ and $K\alpha_2 = 1.544\,39\,\text{\AA}$.

The goniometer provides precise mechanical movements of the detector and the sample(s) with respect to the source of monochromatic X-rays. In a Ω - 2θ goniometer, both source and detector are on the circumference of a goniometer circle, with the sample in the center. Ω is the angle between the X-Ray source and the sample and 2θ is the angle between the incident beam and the detector angle, as shown in Fig. 2.26a. The diffraction vector, perpendicular to the

³Ideally, every possible crystalline orientation is represented equally in a powdered sample

diffracting planes and parallel to the bisectrix of incident and diffracted beam, defines the direction along which structural information, such as lattice parameter or crystallite size, is obtained. It is worth noting that only the crystallites of the thin film that are oriented with lattice planes perpendicular to the bisectrix or diffraction vector contribute to the intensity of these planes in the diffraction pattern. For this reason, this study has been limited to only two configurations: the θ - 2θ and the in-plane configuration, as shown in Fig. 2.26.

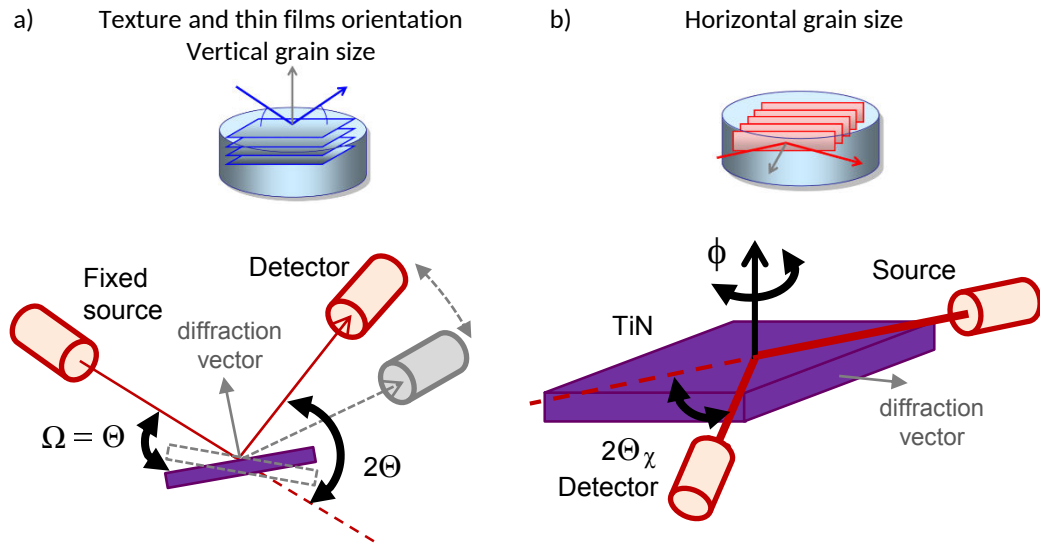


Figure 2.26: a) θ - 2θ XRD configuration (where $\Omega = \theta$) and b) In-plane XRD configuration.

On one hand, in a θ - 2θ XRD configuration, the incident angle Ω is always $1/2$ of the detector angle ($\Omega = 2\theta/2 = \theta$). In a θ - 2θ diffractometer, the X-ray tube (source) is fixed, the sample rotates at $\theta^\circ/\text{second}$ and the detector rotates at $2\theta^\circ/\text{second}$. This set-up authorizes to get the structural information about the planes parallel to the surface of the sample (see diffraction vector in Fig. 2.26a). With this configuration, it is therefore possible to study the distribution of crystallographic orientations of the TiN crystallites and the size of the vertical grains of the TiN films, since only the crystallites of the TiN film that are oriented with lattice planes parallel to the surface of the sample contribute to the intensity of these planes in the diffraction pattern.

On the other hand, in a in-plane XRD configuration, both the incident and diffracted beams are nearly parallel to the sample surface. This set-up authorizes to get the structural information about the planes (nearly) perpendicular to the surface of the sample (see diffraction vector in Fig. 2.26b),

which are inaccessible by other configurations. This set-up was achieved with a five-circles diffractometer, as the one illustrated in Fig. 2.27. First, Ω and 2θ are fixed out of the plane of the sample, nearly parallel to the sample surface, in order to optimize the diffraction intensity. Then, the X-ray tube (source) is fixed. Finally, $2\theta_\chi$ and ϕ angles are coupled, the sample rotates at ϕ/s and the detector rotates at $2\theta_\chi/s$, so that the diffraction vector defines always the same direction, for instance from center to the notch of the wafer. In this configuration, ϕ is always 1/2 of the detector angle ($\phi = 2\theta_\chi/2 = \theta_\chi$). With this configuration, it is therefore possible to estimate the size of the horizontal grains of the TiN films.

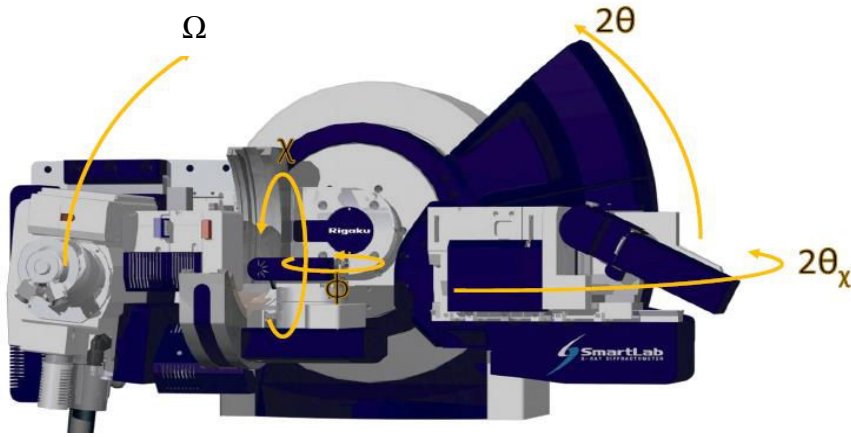


Figure 2.27: Five-circles diffractometer and definition of Ω , 2θ , $2\theta_\chi$, ϕ , and χ angles.

As stated in section 1.3.3, each crystallite orientation is associated to a different value of work function and therefore contributes to the local V_{TH} variability. In addition, it was also stated that a decrease of the horizontal crystallite size might be benefit to reduced the local V_{TH} variability. In consequence, the impact of TiN process deposition conditions on the preferred orientation of the TiN crystallites and the size of crystallite grains will be investigated in section 4.3 with the help of the θ - 2θ X-Ray diffractometer and the in-plane XRD configuration.

Determination of crystallite size from the X-Ray diffraction pattern

The most important parameters in a peak diffraction are the peak position (2θ), the peak height (I_{max}), the peak area (A) and the peak width. The most common parameters used to measure the broadness of the peak are the full width at half the maximum (FWHM) and the integral breadth β . The FWHM corresponds to the difference between the two extreme values of 2θ at which the intensity is equal to $0.5I_{max}$, as shown in Fig. 2.28. The β parameter is defined

as the ratio of the peak area over the peak maximum (A/I_{\max}). These parameters are automatically extracted by fitting the experimental spectra with a combination of Lorentzian and Gaussian functions. The diffraction intensity should not be related to the peak height but to the peak area, since the area represents the true sum of all diffracted X-ray photons that have been detected regardless of the peak shape.

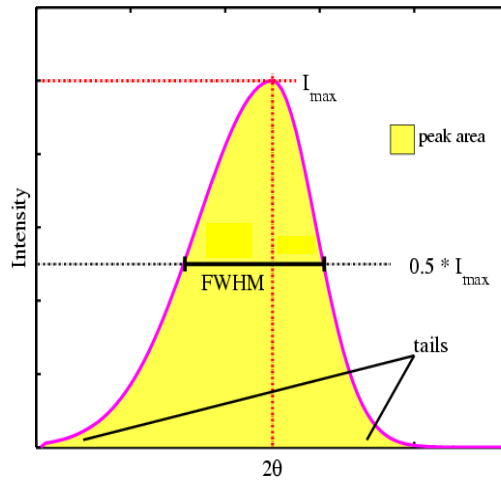


Figure 2.28: Peak in a X-Ray diffraction pattern.

The peak broadening is caused by the instrumental imperfections but also by sample factors such as crystallite size, microstrain due to dislocations, structural defects such as stacking faults, twin faults, layer mistakes, and concentration gradients in non-stoichiometric compounds. In order to get rid of the instrumental contribution, a near-perfect sample or powder was first measured to determine the diffractometer's broadening, and then this value is removed from the peak width obtained with the TiN sample.

The *mean* size of crystallites $\langle D \rangle$ was determined by using the Scherrer equation given here below, where λ is the incident wavelength, β is the integral breadth, and θ is the Bragg angle. The use of peak broadening to determine crystallite size is limited to cases where the mean crystallite size is $\leq 1 \mu\text{m}$.

$$\langle D \rangle = (0.9\lambda)/(\beta \cos \theta) \quad (2.17)$$

Note that according to Eq. 2.17, as the peak broadening increases, the crystallite size then decreases and vice versa. Nevertheless, this equation ignores other effects such as inhomogeneous strain induced by interstitials, vacancies, dislocations and other structural defects.

2.3 Conclusion

The test structures and the techniques of electrical and physicochemical characterization implemented in this thesis for characterization of metal gate stack have been extensively reviewed. In particular, specially designed test structures using LETI patent for parasitic capacitances suppression were successfully embedded in a new mask at STMicroelectronics in order to ease the extraction of WF_{eff} and EOT in short MOSFET devices.

A simplified process flow of the 14 nm FDSOI technology with only the key bricks of one technological option (NFET Low V_{TH}) was validated in order to intensively study the impact of the sacrificial metal gate deposition processes and the incorporation of additives such as lanthanum and aluminum on the effective work function (WF_{eff}) and the equivalent oxide thickness (EOT) of the gate stack. The methodology for an accurate extraction of these parameters from the experimental C-V characteristics has been introduced. Both WF_{eff} and EOT are extracted by fitting with Poisson-Schrödinger quantum simulations, a relevant part of the experimental high-frequency C-V curve, free from leakage and interfacial defects.

Moreover, the integration of a beveled oxide in this simplified process flow authorizes the identification of the origin of the modulation of the effective work function. This modulation will be explained either by interfacial charges or by dipoles or even both effects simultaneously, induced by the incorporation of additives and related processes in the sacrificial metal gate stack. On the other hand, a new methodology based on X-Ray Fluorescence was proposed to measure accurately the effective diffusion of these additives into the gate stack after annealing.

Finally, on one hand, the matching test structures used to characterize the local variability of V_{TH} between adjacent MOS transistors have been described. On the other hand, the orientation and horizontal size of the TiN crystallites as a function of the process deposition conditions can respectively be investigated with the θ - 2θ and in-plane X-Ray diffractometers.

Effective work function modulation by accurate control of diffusion of *sacrificial* lanthanum and aluminum into gate stack of high- κ based NFET devices

Contents

3.1	State of the art of lanthanum in gate stack	82
3.2	Effective work function modulation by accurate control of <i>sacrificial</i> lanthanum diffusion	86
3.3	State of the art of aluminum in gate stack	102
3.4	Effective work function modulation by accurate control of <i>sacrificial</i> aluminum diffusion	107
3.5	Sacrificial vs standard final approach	124
3.6	Conclusion	127

This chapter focuses on the impact of the incorporation of lanthanum (La) and aluminum (Al) in a *sacrificial* gate-first approach on the effective work function of 14 nm FDSOI devices. As described in section 1.3, the effective work function range to switch from a N-gate to a P-gate in FDSOI devices is shorter (between 100 meV and 200 meV) than in bulk devices, which makes even more necessary a very fine adjustment of the effective work function of metal electrodes. The aim of this chapter is therefore to assess the influence of both lanthanum and aluminum sub-nanocapping layers incorporated in a *sacrificial* metal gate-first approach on the WF_{eff} and the EOT regrowth of 14 nm FDSOI devices.

First, a review of the state of the art of lanthanum in gate stack is included in section 3.1. The section 3.2 covers the study of the process variations used to incorporate La in a *sacrificial* gate approach in NFET devices, the electrical measurements and the diffusion characterization of lanthanum into the HfON/SiON stack. The influence of high- κ dielectrics on the effective work function shift induced by lanthanum incorporation is also studied in section 3.2.

Similarly, a review of the state of the art of aluminum in gate stack is given in section 3.3. The study of the process variations used to incorporate aluminum following a *sacrificial* gate-first approach in NFET devices as well as

the electrical characterization are described in section 3.4. Likewise, the diffusion characterization of aluminum into the HfON/SiON stack and the influence of high- κ dielectrics on the effective work function shift induced by aluminum incorporation are also presented and discussed in section 3.4. Finally, the benefit of the sacrificial gate-first approach compared to a standard gate-first approach is discussed in section 3.5, as well as the perspectives for future research.

3.1 State of the art of lanthanum in gate stack

3.1.1 Lanthanum oxide as alternate gate dielectrics

As stated in section 1.2.1, besides the high- κ value, other requirements for a high- κ material to be considered as a candidate to replace SiO_2 , are a large bandgap and large band offsets relative to the conduction and valence bands of silicon. Oxides of transition metals such as Ti, Ta, Nb, Sc, Zr and Hf have generally high- κ values but low conduction band offset values, while the oxides of rare-earth metals such as Yb_2O_3 , Gd_2O_3 , Sm_2O_3 , Y_2O_3 and La_2O_3 have generally lower κ -values but higher conduction band offset values [96]. For this reason, since the beginning of the 2000s, the lanthanum oxide or lanthana (La_2O_3) [97] [98] as well as the La-silicate (LaSiO) [99] [100] have attracted much attention as alternate gate dielectrics because of their relatively high dielectric constant value ($\kappa \approx 25\text{-}30$), their energy bandgap of about 5.5 eV and conduction band and valence band offset energies higher than 2 eV. In addition, the LaSiO and the La_2O_3 films were found to exhibit an undesirable large flatband voltage shift of about 1.4 eV towards the N+ values. At the time, it was attributed to excessive positive charges in the dielectric to explain the direction of the shift [97]. These charges were attributed to positively charged oxygen vacancies (V_o^{++}) and to the formation of $\text{La}(\text{OH})_3$, by the presence of $(\text{OH})^-$ ions in place of O_2^{-2} sites.

Furthermore, LaSiO and La_2O_3 are not suitable as alternate gate dielectrics because they, as rare-earth oxides, react easily with silicon substrate to form silicate layers. The interaction of La with Si has been reported for lanthanum deposition by ALD [98], CVD [101], and PVD [100]. This is explained by the typically lower free energy of formation of rare-earth oxides compared to the one of transition metal oxides [102] [103]. Experimental work of Triyoso et al. [98] revealed a rough La-based dielectric/Si interface and also chemical interaction between La and TaSiN gates due to the formation of La-silicates. As a result, no functional devices were obtained due to short circuit of the top electrode and the semiconductor substrate. It was concluded that silicon substrate interactions may limit the utilization of La-based dielectrics in future CMOS processing.

3.1.2 Lanthanum incorporation into Hf-based dielectrics

Despite the drawbacks of integrating lanthanum oxide and its derivatives as gate dielectrics, instead of Hf-based dielectrics, Narayanan et al. proposed to insert nanoscale capping layers containing lanthanum into HfO_2/TiN stacks [21] in a conventional gate-first flow. In this way, the large V_{fb} shifts towards $N+$ induced by lanthanum can be advantageously used to control V_{TH} transistors by careful optimization of the La thickness. In this work, a large WF_{eff} shift from 4.43 (without La) to 4.05 (with La) is reported, although La thicknesses are not revealed. However, it is indicated that the La thickness is a critical optimization. It must be thick enough to shift the WF_{eff} towards $N+$, but not so much that lanthanum near or in the Si/SiO_2 interface degrades the mobility. Narayanan attributed this negative shift to two concomitant processes in the dielectric. Since the cap layer intermixed with HfO_2 , Narayanan proposed that the WF_{eff} shifts can be either due to positively charged mobile oxygen vacancies formed upon the aliovalent substitution of Hf^{4+} in HfO_2 with lower valence La^{3+} and/or a dipole at metal/high- κ interface due to differences in electronegativities between La and Hf. In contrast, Yamamoto et al. [23] experimentally evidenced that the most plausible origin for the V_{fb} shift is the formation of a dipole layer at the $\text{HfLaO}_x/\text{SiO}_2$ interface. In that work, the origin of La effect on V_{fb} was investigated experimentally by changing the La concentration in HfLaO_x films of $\text{Au}/\text{HfLaO}_x/\text{SiO}_2/\text{Si}$ capacitors, as shown in Fig. 3.1a. Electrical measurements were performed after annealing at 600 °C. Figs. 3.1b and 3.1c clearly show that V_{fb} in Cap#1 is exactly the same as that in Cap#4, while V_{fb} in Cap#2 is the same as that in Cap#3. These results clearly demonstrate that only the oxide in contact with SiO_2 plays a significant role on the V_{fb} shift. In consequence, it was experimentally proved that the metal/high- κ interface is not the origin of the modulation of V_{fb} and raised the question about the validity of the top dipole layer model, which claim the dipole layer formation at the metal/high- κ interface. Later, Kita and Toriumi [27] proposed a model based on oxygen vacancies migration to explain the direction of this WF_{eff} shift at the high- κ/SiO_2 interface. It should be noted that WF_{eff} modulation successfully achieved by capping the top surface of the high- κ in this work may seem, at first sight, to be in contradiction with the bottom layer dipole model reported by Kita and Toriumi. Nevertheless, it can be noticed that V_{fb} shift achieved by the deposition of capping layers on the top of HfO_2 in MOSFET devices reported by Narayanan [21] employed the high temperature source/drain annealing of the conventional gate-first flow. It can therefore be speculated that high thermal budget cause the diffusion of La into the high- κ /interface and result in the modification of the dipole layer at high- κ/SiO_2 interface.

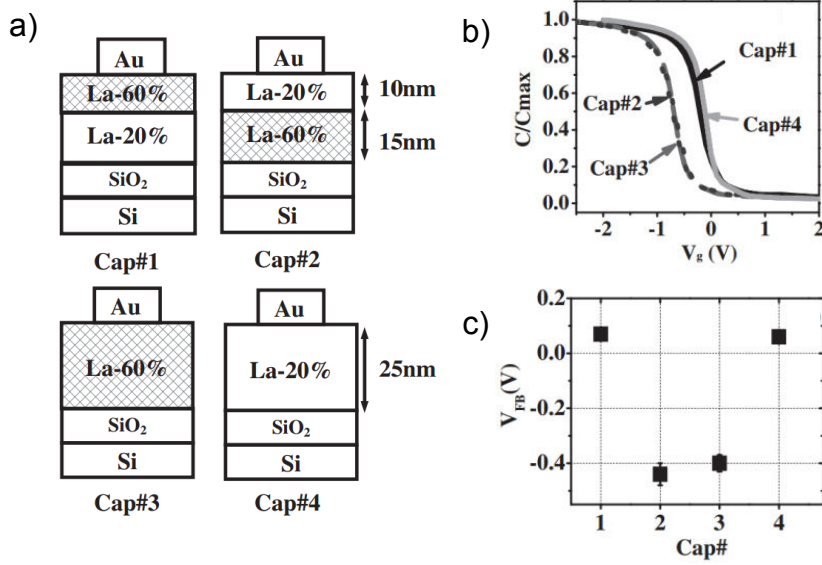


Figure 3.1: a) Schematic description of MOS capacitors with different La concentration profiles in HfLaO_x films. Cap#1: Au/La-60%/La-20%/SiO₂/Si, Cap#2: Au/La-20%/La-60%/SiO₂/Si, Cap#3: Au/La-60%/SiO₂/Si, Cap#4: Au/La-20%/SiO₂/Si; b) C-V curves and c) V_{fb} values of Cap#1-4, after [23].

Indeed, the impact of the diffusion of lanthanum on the core level energy positions in the Hf and Ti with respect to Si bulk after high temperature annealing (1065 °C under N₂ for 1.5 s) has been confirmed in a recent work [104], by performing hard X-ray photoelectron spectroscopy (HAXPES) with synchrotron radiation on TiN/LaO_x/HfSiON/SiON/Si gate stacks. After annealing, the relative differences between the binding energies of Hf4f and Ti2p core levels with respect to Si bulk were consistent with the V_{fb} shifts reported on the same stacks. In section 3.2.1, the impact of the diffusion of lanthanum on the effective work function is investigated for different gate stack variations.

3.1.3 Lanthanum incorporation into TiN

The potential of lanthanum inserted into TiN metal gate has also been evaluated in the multilayer TiN/La structure [64], TiN/La/TiN structures and in the alloyed TiLaN metal [105]. These structures were proposed due to another benefit other than WF_{eff} shift: the interlayer scavenging effect. The scavenging effect means that La ions remove oxygen from the interlayer to decrease the overall EOT of the devices. Both TiN/La/TiN structures and in the alloyed TiLaN metal suppressed interlayer growth to extremely small values (EOT around 1 Å). This was attributed to the increase of the oxygen solubility in TiN with the incorporation of lanthanum [105]. The interest of these

structures is to use the TiN barrier diffusion layer to prevent interfacial oxidation in detriment of the oxidation of TiN layers.

In this thesis work, we rather evaluate the multilayer TiN/La/TiN structure in a *sacrificial* metal gate-first approach for other purpose: the modulation of the WF_{eff} with annealing temperature and bottom TiN thickness.

3.2 Effective work function modulation by accurate control of *sacrificial* lanthanum diffusion

Here, we evaluate for the first time the impact of metallic lanthanum (La) deposited by Radio-Frequency PVD in a *sacrificial* metal gate-first approach on the effective work function of HfON-based NFET devices [106]. First, engineering of WF_{eff} towards N+ without leakage degradation will be demonstrated in section 3.2.1 by tuning both the TiN capping thickness and the as-deposited metallic La dose in the *sacrificial* gate stack. Second, in section 3.2.2, WF_{eff} shift will be correlated to the effective La dose into HfON/SiON stack after diffusion annealing, which has been accurately measured by X-Ray Fluorescence (XRF) [106]. The influence of annealing temperature on the diffusion of lanthanum into dielectrics is also evaluated in section 3.2.2.

3.2.1 Device fabrication and electrical characterization

Low- V_{TH} NFET devices were fabricated with the simplified process flow of Fig. 2.12 in order to assess the impact of the lanthanum content into the *sacrificial* Top-TiN/La/Bottom-TiN structure on the effective work function and therefore on the V_{TH} of the advanced 14 nm FDSOI devices. In addition, beveled thermally grown SiO_x was used as interlayer dielectric in concomitant research devices in order to discriminate between an effect of fixed charges from a dipole effect as the origin of the WF_{eff} modulation. Both bottom TiN and La layers thicknesses were investigated. The *sacrificial* Si/TiN/La/TiN gate stack (Si on top) with different targeted La (0 Å, 2 Å, 4 Å, 6 Å) and bottom TiN (0 Å, 10 Å and 15 Å) thicknesses was deposited on HfON. It was followed by a thermal treatment under N_2 atmosphere at 900 °C for 10 s in order to activate the diffusion of La into the HfON/SiON stack. The sacrificial gate stack was then removed by wet etching. Finally, a Poly-Si/TiN electrode was deposited, followed by gate patterning, as shown in Fig. 1.18. The Poly-Si and the final TiN layers have a thickness of 240 Å and 35 Å, respectively, for all the devices. The aim of the study is to demonstrate that WF_{eff} can be modulated by tuning the diffusion of La into the gate stack with different combinations of TiN and La thicknesses. TiN and La layers were deposited in RF-PVD chambers, as explained in section 1.2.2.2. The lanthanum XRF line intensity was used to adjust deposition times in order to reach La doses of $\approx 5 \times 10^{14}$ at/cm², $\approx 1 \times 10^{15}$ at/cm² and $\approx 1.5 \times 10^{15}$ at/cm² corresponding to 2 Å, 4 Å, 6 Å, respectively. Sample description is summarized in Table 3.1.

Sample	Interlayer oxide	La dose (at/cm ²)	Bottom thickness	Top TiN thickness
TiN45 Å	nominal	0	0	45 Å
La2 Å/TiN35 Å	nominal	5 x 10 ¹⁴	0	35 Å
La4 Å/TiN35 Å	nominal	1 x 10 ¹⁵	0	35 Å
TiN10 Å/La2 Å/TiN35 Å	nominal	5 x 10 ¹⁴	10 Å	35 Å
TiN10 Å/La4 Å/TiN35 Å	nominal	1 x 10 ¹⁵	10 Å	35 Å
TiN10 Å/La6 Å/TiN35 Å	nominal	1.5 x 10 ¹⁵	10 Å	35 Å
TiN45 Å	bevel	0	0	45 Å
La2 Å/TiN35 Å	bevel	5 x 10 ¹⁴	0	35 Å
La4 Å/TiN35 Å	bevel	1 x 10 ¹⁵	0	35 Å
TiN10 Å/La2 Å/TiN35 Å	bevel	5 x 10 ¹⁴	10 Å	35 Å
Ti rich-TiN10 Å/La2 Å/ Ti rich-TiN35 Å	bevel	5 x 10 ¹⁴	10 Å	35 Å

Table 3.1: Sample description.

EOT and V_{fb} were extracted by fitting Poisson-Schrödinger quantum simulations with experimental high-frequency gate-channel CV measurements on HfON-based and HfSiON/beveled-SiO₂ devices for different dies of the wafer [89]. WF_{eff} was then calculated from V_{fb} and the silicon work function relative to vacuum ($q\Phi_{sc}$), as presented in section 2.1.3.

WF_{eff} is decreased by up to 350 meV towards N+ with metallic La inserted in the sacrificial gate with or without 10 Å-thick TiN layer between La and HfON. However, bottom TiN thickness is efficient to moderate the La effect on WF_{eff} and on EOT, as illustrated in Fig. 3.2. Besides, as La increases, it induces a shrink of around 1 Å for the capacitance equivalent thickness in inversion (CET_{inv}). Such result is obtained without any degradation of leakage (J_g), as clearly indicated in Fig. 3.3, suggesting that a higher κ -value is achieved rather than an interlayer physical reduction. Since La is not deposited again in the final Poly-Si/TiN electrode (see step 4 in Fig. 1.18), the ΔWF_{eff} induced by tuning the La diffusion on HfON devices cannot be attributed to a metal work function ($q\Phi_m$) dependence with La dose. To investigate the origin of the ΔWF_{eff} modulation, we report in Fig. 3.4 the WF_{eff} vs EOT plot for nominal and beveled devices with different combinations of TiN and La thicknesses deposited by RF-PVD in a *sacrificial* gate-first approach and annealed at 900 °C for 10 s.

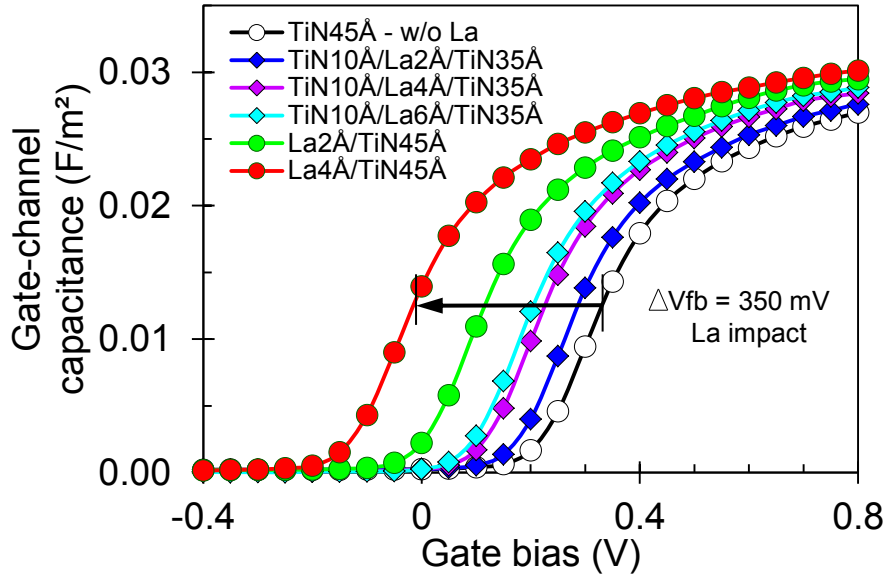


Figure 3.2: C_{gc} measurements for different La doses in sacrificial gate-first approach.

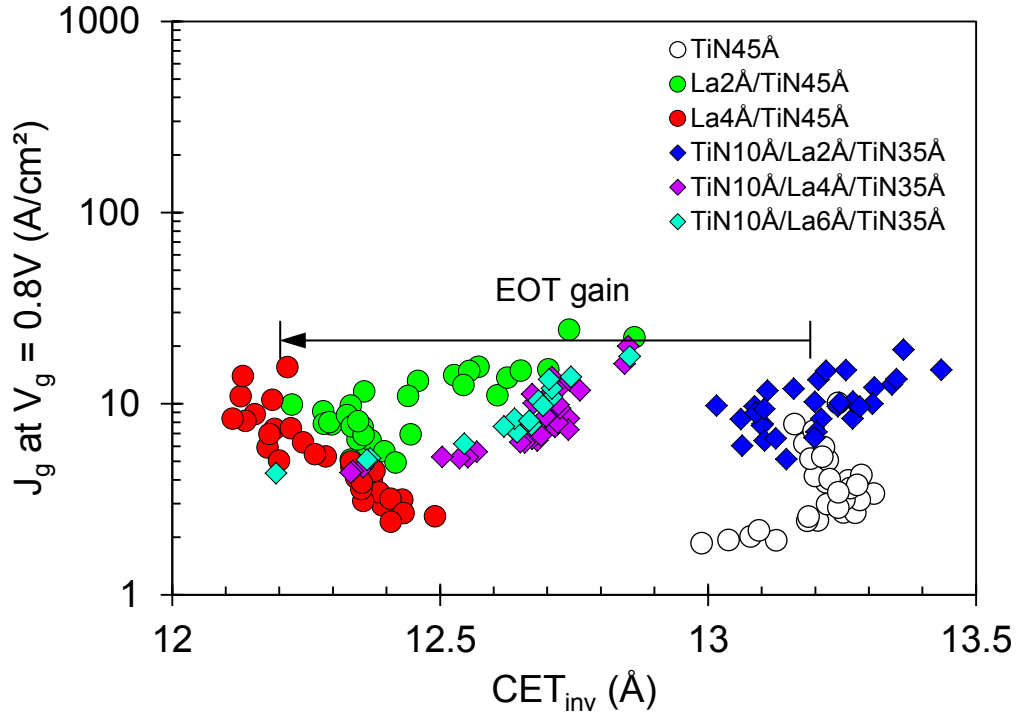


Figure 3.3: J_g/CET_{inv} trade-off at $V_g=0.8V$, as a function of La dose.

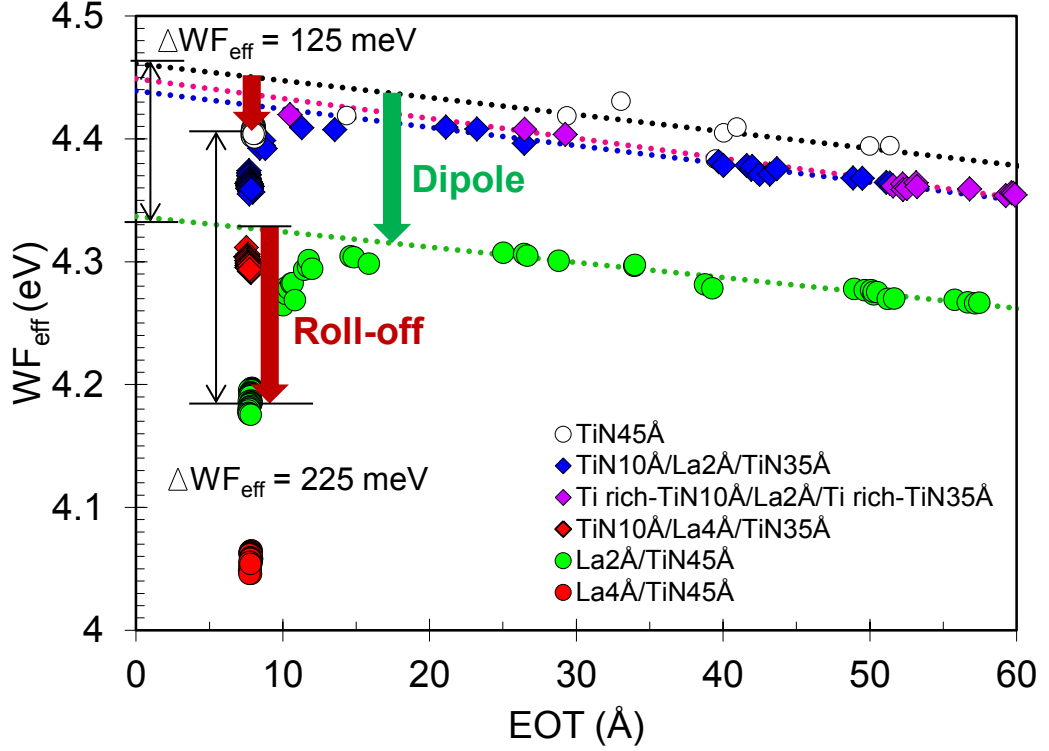


Figure 3.4: WF_{eff} vs EOT curve of HfON-based NFET devices with nominal and beveled oxides for different La doses inserted in a *sacrificial* gate-first approach. The sacrificial gate stack was annealed at 900 °C for 10s.

First, the linear dependence at largest EOT proves that interface charges exist between channel and SiO_2 and also no other charges in bulk SiO_2 (no quadratic dependence). From the slope of the WF_{eff} vs EOT plot, fixed charges density is estimated to be $7.96 \times 10^{11} / \text{cm}^2$, $8.21 \times 10^{11} / \text{cm}^2$ and $6.89 \times 10^{11} / \text{cm}^2$ for *sacrificial* stack without La (white circles), $\text{TiN}10 \text{ \AA} / \text{La}2 \text{ \AA} / \text{TiN}35 \text{ \AA}$ stack (blue diamonds), and $\text{La}2 \text{ \AA} / \text{TiN}35 \text{ \AA}$ stack (green circles), respectively. The ΔWF_{eff} extrapolated at zero EOT reveals the influence of a drop voltage induced by a dipole layer (δ) of about 125 meV (see Fig. 3.4) with the incorporation of 2 Å of La in a sacrificial gate approach (62.5 meV/Å). This drop voltage induced by a dipole layer is reduced to only 22 meV (11 meV/Å) with the incorporation of 10 Å of TiN between La and HfON. This result proves that the dipole effect is modulated with the effective dose of lanthanum incorporated into the HfON/ SiO_x stack. Additionally, it is known that only La at high- κ / SiO_x interface plays a significant role to shift the WF_{eff} [23] [25]. It can therefore be concluded that it is the effective lanthanum dose at this interface the origin of the WF_{eff} modulation [106]. However, an abrupt potential drop (called roll-off) appears when the equivalent oxide thickness (EOT) becomes thinner than 20 Å. Such roll-off is defined as the difference between the

WF_{eff} measured at a given EOT and the expected value obtained by linear fit from larger EOT (red arrows in Fig. 3.4). In consequence, the ΔWF_{eff} induced by effective La diffusion at nominal EOT of 14 nm FDSOI devices results in the sum of the dipole effect ($(\Delta WF_{\text{eff}})_{\text{EOT}=0}$) related to La diffusion and its roll-off ($\text{roll-off}_{\text{w/La}}$) minus the roll-off associated to sacrificial stack without La ($\text{roll-off}_{\text{w/o La}}$), as given in Eq. 3.1. For instance, the 225 meV of ΔWF_{eff} at nominal EOT induced by diffusion of La in a stack is the result of 125 meV (dipole) + 145 meV (La roll-off at nominal EOT) - 45 meV (TiN roll-off at nominal EOT). It should be noted that potential drop related to the dipole depends on the effective lanthanum diffusion, whereas the roll-off evolves with both the effective lanthanum diffusion and the EOT.

$$\Delta WF_{\text{eff}} = (\Delta WF_{\text{eff}})_{\text{EOT}=0} + \text{roll-off}_{\text{w/La}} - \text{roll-off}_{\text{w/o La}} \quad (3.1)$$

Several mechanisms have been proposed to explain the roll-off phenomenon. Some of them suggest a relationship between the roll-off and the generation of positively charged oxygen vacancies (V_o^{++}) in the interfacial SiO_2 next to the Si substrate [24], but in that case the effect should vary with HfO_2 thickness and the rate of V_o^{++} should be precisely controlled. Charbonnier et al. [9] demonstrated that such roll-off depends only on the SiO_2 interlayer thickness (and not of the total EOT). In that work, it has also been proved by internal photoemission (IPE) measurements that metal/ HfO_2 barrier height for devices with various SiO_2 thicknesses is unchanged, implying that roll-off phenomenon is not related to the metal/ HfO_2 interface. Furthermore, the independence of the roll-off with HfO_2 thickness proves that roll-off must be attributed to an abrupt interfacial drop (dipole) at $\text{HfO}_2/\text{SiO}_2$ interface and should not be related to fixed charges at this interface [9]. It seems therefore logical to assume that it is an evolution of the La-induced dipole at $\text{SiO}_x/\text{high-}\kappa$ interface.

In summary, the sacrificial gate-first approach offers an additional process knob to adjust the V_{TH} by the incorporation of lanthanum, whose impact can be moderated with the deposition of a thin layer of TiN prior to the deposition of lanthanum. TiN acts as an efficient barrier diffusion layer.

3.2.2 Diffusion characterization by X-Ray Fluorescence

Aiming to clarify the influence of the diffused La dose on the induced dipole at HfON/SiON interface, XRF measurements were performed on blanket wafers with nominal dielectrics thicknesses a) after sacrificial gate stack deposition and b) after diffusion annealing and sacrificial gate stack removal in order to evaluate the effective lanthanum dose incorporated into HfON/SiON stack, as explained in section 2.2.2.2. The effective La dose (La_{eff}) is determined as a function of the as-deposited lanthanum dose (La_{dep}), the bottom TiN thickness (d) and the annealing temperature after the *sacrificial* gate stack deposition.

It was found that the effective lanthanum dose can be tuned by the as-deposited lanthanum dose (Fig.3.5), the bottom TiN thickness below La layer (Fig.3.6) and the annealing temperature (Fig.3.7). As it could be expected, the effective La dose increases with the as-deposited La dose, even with a TiN layer of 10 Å as a barrier diffusion. In contrast, La_{eff} decreases from 2.36×10^{14} at/cm² to 2.34×10^{13} at/cm² with an increase of bottom TiN thickness from 0 Å to 15 Å. Furthermore, La_{eff} decreases down to 2.5×10^{12} at/cm² by decreasing the annealing temperature from 900 °C to 700 °C underwent by a TiN10 Å/La2 Å/TiN35 Å stack.

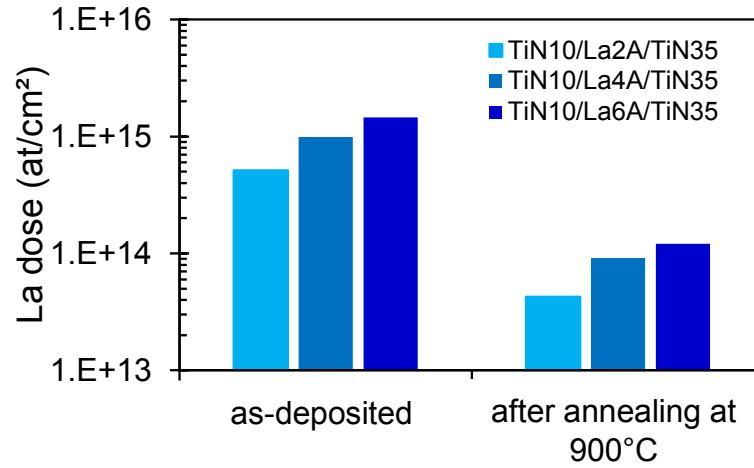


Figure 3.5: Quantification of La dose by X-Ray Fluorescence on as deposited sacrificial gate stack and after sacrificial gate removal for annealing at 900 °C for 10 seconds as a function of La dose.

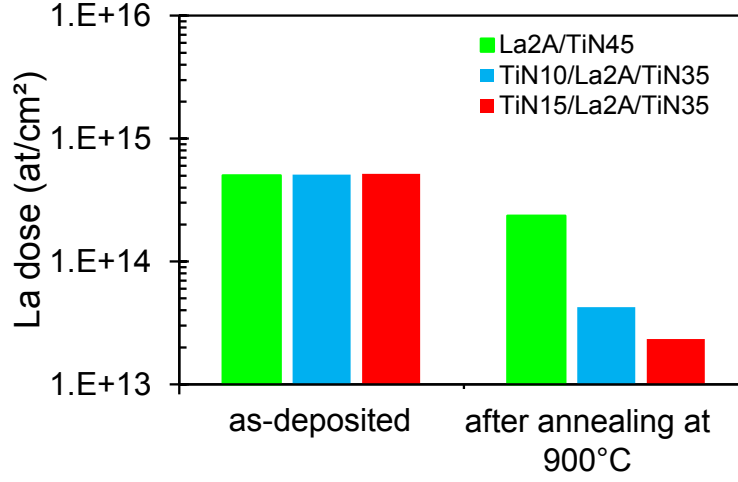


Figure 3.6: Quantification of La dose by X-Ray Fluorescence on as deposited sacrificial gate stack and after sacrificial gate removal for annealing at 900 °C for 10 seconds as a function of bottom TiN thickness.

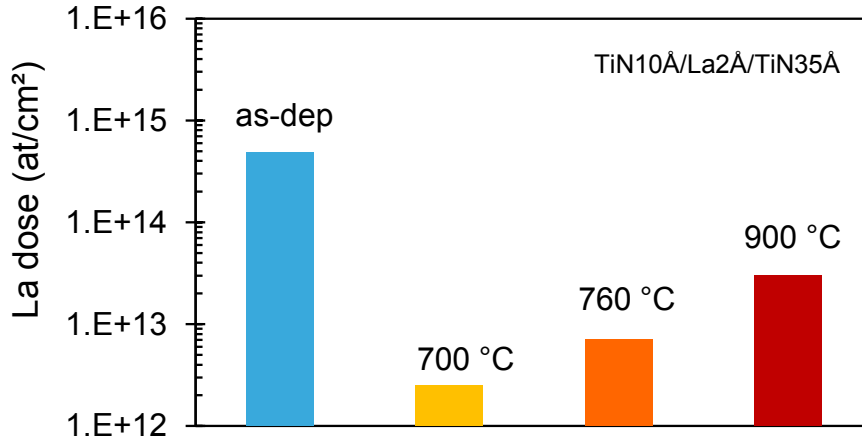


Figure 3.7: Quantification of La dose by X-Ray Fluorescence on as-deposited TiN10 Å/La2 Å/TiN35 Å stack, and after annealing and sacrificial gate removal for different annealing temperatures.

However, it should be noted that these modulations of effective lanthanum doses are not arbitrary but they follow the physical laws of diffusion. The percentage of effective lanthanum diffusion (the ratio La_{eff} / La_{dep} in %) is independent of the as-deposited La dose. In other words, La_{eff} linearly increases with La_{dep} . Furthermore, La_{eff} / La_{dep} in % is only modulated by the thickness of the bottom TiN, as shown in Fig. 3.8. A decrease of bottom TiN thickness

leads to an increase of the $\text{La}_{\text{eff}} / \text{La}_{\text{dep}}$ in %. Moreover, $\text{La}_{\text{eff}} / \text{La}_{\text{dep}}$ evidences an exponential dependence with the bottom TiN thickness, as can be observed in Fig. 3.9. In consequence, the effective lanthanum dose La_{eff} in at/cm^2 after annealing at 900°C for 10 s may be accurately controlled and predicted from the empirical equation 3.2, knowing the values of La_{dep} (in at/cm^2) and d (in \AA). The Eq. 3.2 is an exponential equation $y = A \exp(-Bx)$. The coefficient A before the exponential indicates the percentage of effective lanthanum diffusion without TiN and the coefficient B represents the efficiency of TiN to screen the diffusion of lanthanum at 900°C for 10 s.

$$\text{La}_{\text{eff}}/\text{La}_{\text{dep}} = 46.9 \exp(-0.156 d) \quad (3.2)$$

It is worth noting that the effect of La should be entirely inhibited with a 30\AA -thick bottom TiN layer below La. In addition, only $\approx 47\%$ of the as-deposited La dose diffuses into the HfON/SiON stack without bottom TiN. The rest of La, which diffuses into the top TiN layer, is totally removed by wet etching in the *sacrificial* gate-first approach.

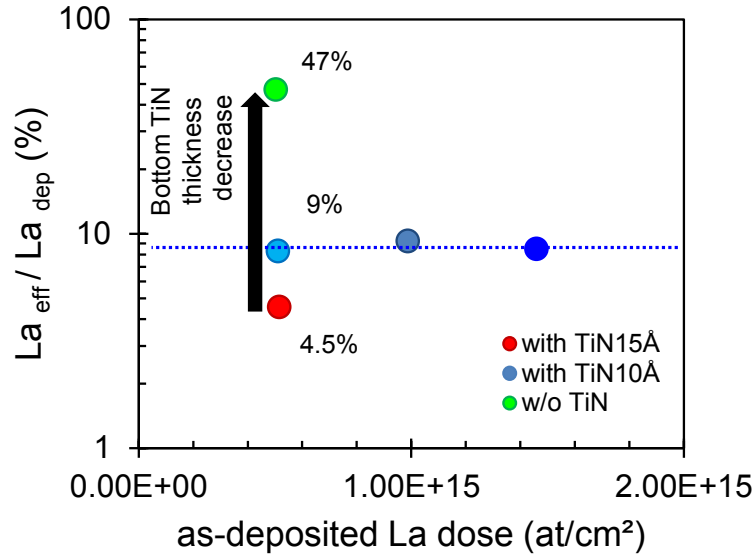


Figure 3.8: Percentage of effective lanthanum into HfON/SiON as a function of the as-deposited lanthanum dose for different bottom TiN thicknesses.

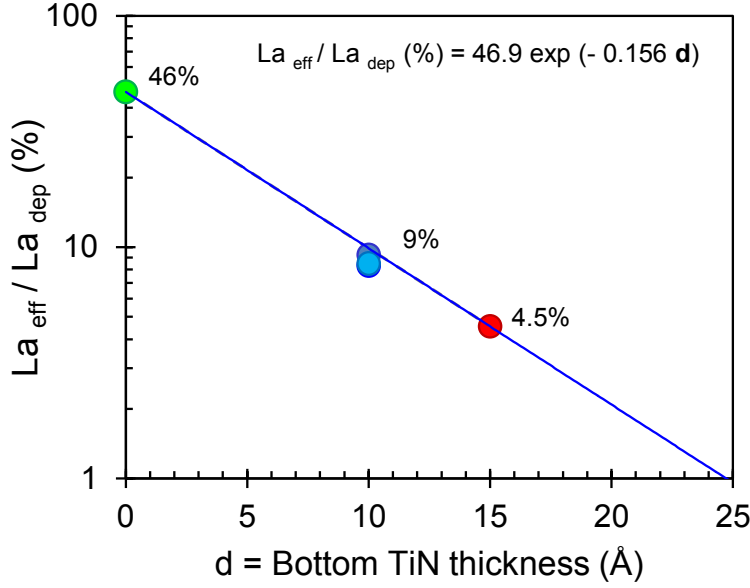


Figure 3.9: Percentage of effective lanthanum dose incorporated into HfON/SiON after annealing at 900 °C for 10 s as a function of the bottom TiN thickness for different as-deposited lanthanum doses.

Similarly, $La_{\text{eff}} / La_{\text{dep}}$ evidences an exponential dependence with the annealing temperature, as can be observed in Fig. 3.10. In consequence, for a TiN10 Å/La2 Å/TiN35 Å stack, the effective lanthanum dose La_{eff} in at/cm² may be accurately controlled and predicted from the empirical equation 3.3, knowing the values of La_{dep} in at/cm² and T in °C. Note that the effect of La in a sacrificial TiN10 Å/La2 Å/TiN35 Å stack should be entirely neglected at annealing temperatures lower than 600 °C. Here is why such La effect is not observed by capping with La monolayers the top surface of a high- κ layer in a conventional gate-last approach. And it is also for this reason that a WF_{eff} shift has been successfully achieved with such configuration in a conventional gate-first approach.

$$La_{\text{eff}}/La_{\text{dep}} = 0.00004 \exp(0.0137 T) \quad (3.3)$$

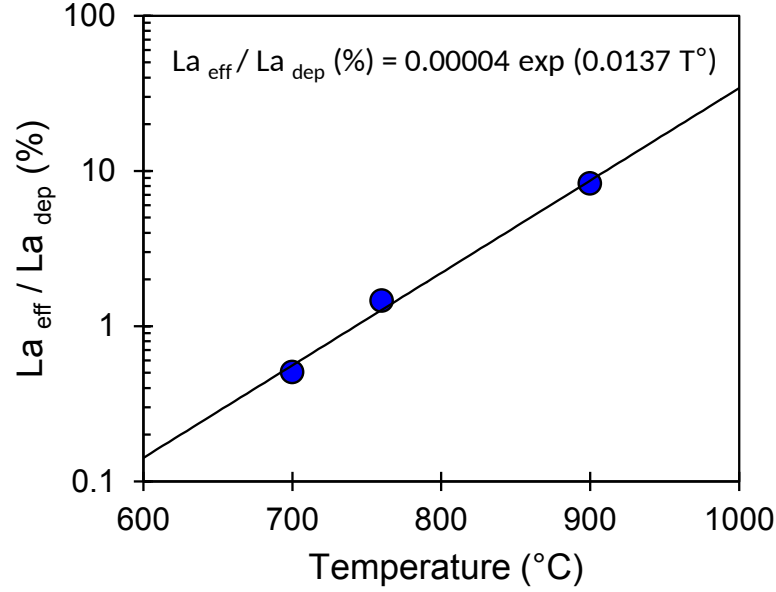


Figure 3.10: Percentage of effective lanthanum dose incorporated into HfON/SiON as a function of the annealing temperature for a sacrificial TiN10 Å/La2 Å/TiN35 Å stack.

In summary, the *sacrificial* gate-first approach allows the prediction and accurate control of diffusion of lanthanum into HfON. Furthermore, a linear relationship between WF_{eff} shift at nominal EOT and La_{eff} is evidenced for all the samples. Such WF_{eff} shift at nominal EOT takes into account the roll-off effect (see Fig. 3.4) and it corresponds to a shift of -80 mV for each La dose of 1×10^{14} at/cm² incorporated into HfON layer, as shown in Fig. 3.11 below. It should be noted that 1×10^{14} at/cm² corresponds to a thickness of about 0.4 Å of lanthanum, and thus less than an atomic monolayer. This weak effective dose can be successfully measured thanks to the detection limit of methodology based on X-Ray fluorescence and the higher fluorescence of lanthanum.

In conclusion, engineering of WF_{eff} towards N+ without leakage degradation is demonstrated by tuning both the as-deposited metallic lanthanum dose and the TiN layer thickness below lanthanum in *sacrificial* metal gate-first integration, making this approach suitable for V_{TH} requirements in FDSOI devices. The WF_{eff} shift on HfON-based 14 nm NFET devices is related to the evolution with effective La dose of a dipole at SiON/HfON interface. Its value has been linearly correlated to the effective La dose incorporated into HfON/SiON stack after diffusion annealing, which has been accurately measured through a spectroscopic method based on lanthanum X-Ray Fluorescence.

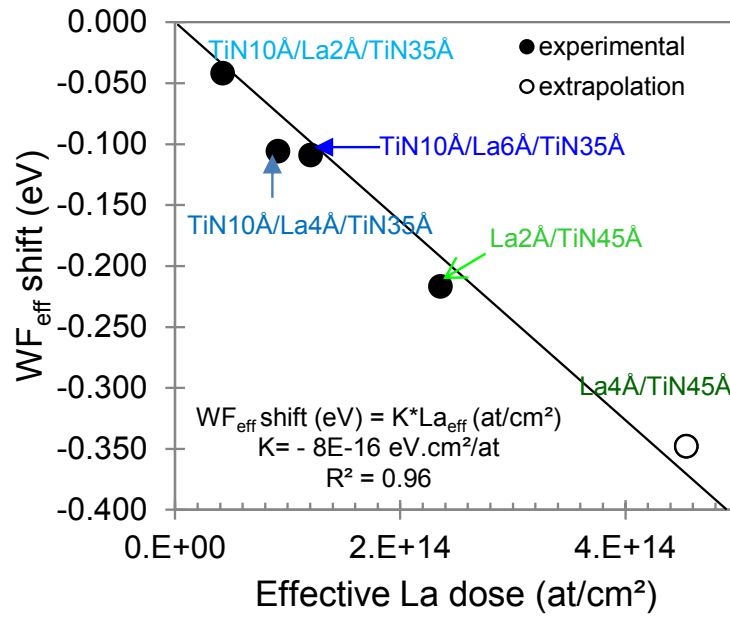


Figure 3.11: WF_{eff} shift increase towards N+ on HfON-based NFET devices vs. effective lanthanum dose incorporated into HfON/SiON stack.

3.2.3 Influence of high- κ dielectrics on the effective work function shift induced by lanthanum incorporation

In this section we will compare the impact of the effective La diffusion on the effective work function shift of both HfON- and HfSiON-based devices. HfSiON is also of particular interest since it is used as high- κ dielectric in 28 nm FDSOI technologies fabricated at STMicroelectronics. As a reminder, HfO₂ is deposited by Atomic Layer Deposition (ALD) without decoupled plasma nitridation (DPN), followed by the Post-Nitridation Anneal (PNA), as described in section 1.2.1. Equivalent HfSiON thickness is deposited by Metal Organic Chemical Vapor Deposition (MOCVD) with a concentration of 30% Si, followed by the same decoupled plasma nitridation (DPN) and the Post-Nitridation Anneal (PNA) of the 14 nm FDSOI technology. In particular, the diffusion of lanthanum into both high- κ materials is first studied by using the XRF methodology, as shown in Fig. 3.12. Then, HfSiON-based NFET devices with and without 2 Å of La in the *sacrificial* stack were fabricated with the simplified process flow of Fig. 2.12. In addition, beveled thermally grown SiO₂ was used as interlayer dielectric in these devices in order to investigate the origin of the WF_{eff} modulation, as explained in section 2.1.2.2. The purpose of the study is to assess the effective work function shift induced by lanthanum diffusion into these different high- κ materials.

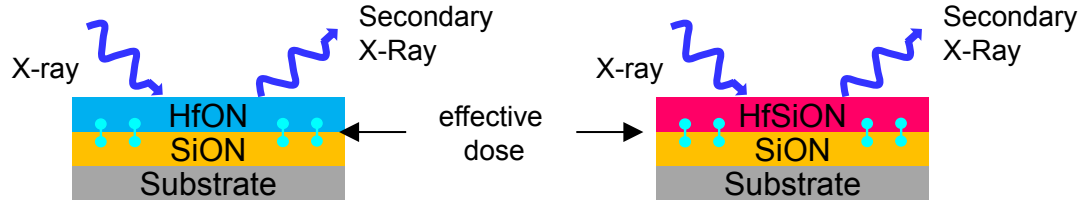


Figure 3.12: XRF scheme of the lanthanum diffusion study into HfON and HfSiON.

Based on results reported in section 3.2.2, TiN prior Lanthanum was not deposited here so that X-Ray Fluorescence signal could be high enough to study La diffusion into HfSiON after annealing at 900 °C for 10 s. La doses of $\approx 5 \times 10^{14}$ at/cm², $\approx 1 \times 10^{15}$ at/cm² and $\approx 1.5 \times 10^{15}$ at/cm² corresponding, respectively, to 2 Å, 4 Å, 6 Å were then directly deposited over 20 Å-thick HfSiON layer, following the *sacrificial* gate-first integration detailed in section 1.3.2.

The La doses in at/cm^2 resulting from XRF measurements a) after sacrificial gate stack deposition and b) after diffusion annealing and sacrificial gate stack removal, are shown in Fig. 3.13. We observe that, as in the case of HfON, the ratio $\text{La}_{\text{eff}} / \text{La}_{\text{dep}}$ is independent of the as-deposited La dose. In other words, the effective La dose is found to be constantly around 20% of the as-deposited La dose, as demonstrated in Fig. 3.14. Therefore, La_{eff} linearly increases with La_{dep} .

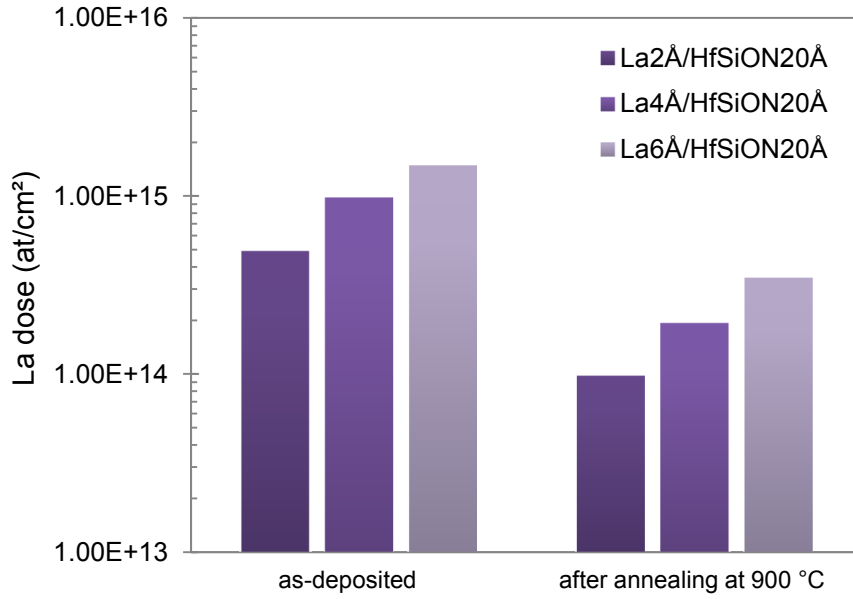


Figure 3.13: Quantification of La diffusion into HfSiON by X-Ray Fluorescence. On the left, as-deposited La dose. On the right, La into HfSiON/SiON after annealing at 900 °C and wet sacrificial gate removal for different as-deposited La doses.

Henceforth, the La diffusion into both oxides (HfON vs HfSiON) after annealing at 900 °C can be compared, as shown in Fig. 3.15. At this temperature, La clearly diffuses more easily into HfON than into HfSiON, 47% instead of 20% for the same as-deposited dose without TiN barrier diffusion layer inserted between La and both high- κ .

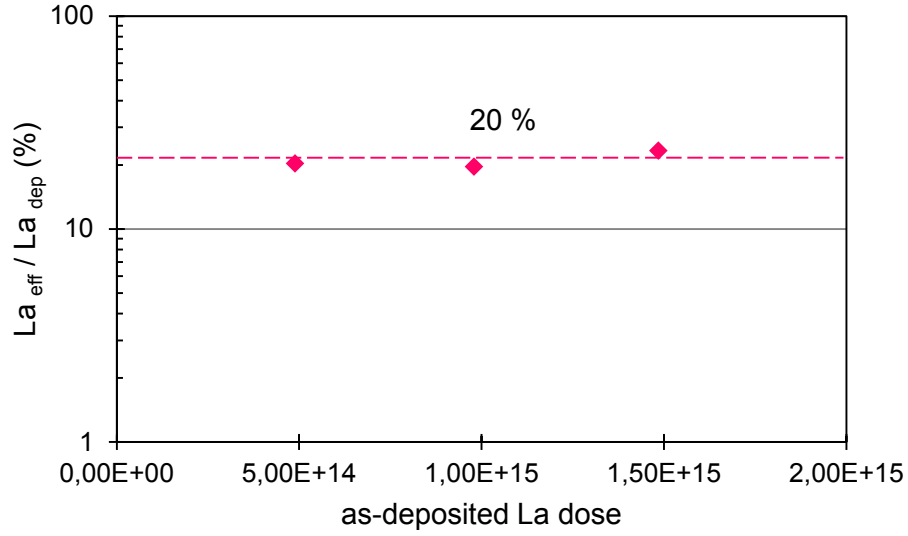


Figure 3.14: Percentage of effective lanthanum into HfSiON/SiON as a function of the as-deposited lanthanum dose.

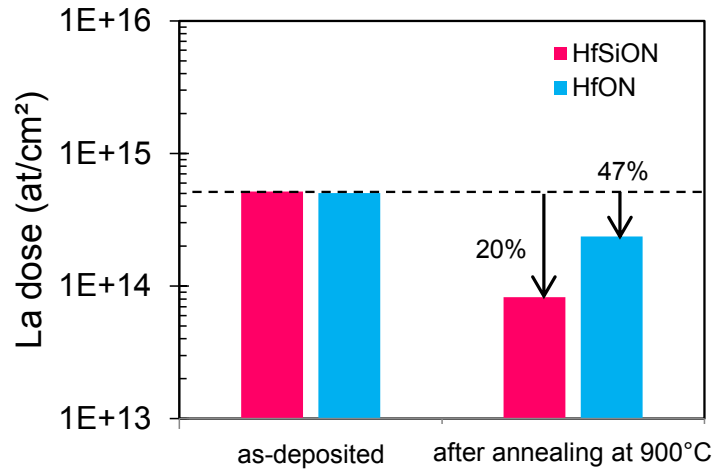


Figure 3.15: La diffusion into both oxides after annealing at 900 °C.

The WF_{eff} vs EOT plot for HfSiON/beveled-SiO₂ and HfON/beveled-SiO₂ devices is shown in Fig. 3.16. From the slope of the WF_{eff} vs EOT plot, fixed charges density is estimated to be $7.15 \times 10^{11} / \text{cm}^2$ and $5.88 \times 10^{11} / \text{cm}^2$, respectively, for HfSiON-NFET devices without and with 2 Å of La in the *sacrificial* stack. The ΔWF_{eff} extrapolated at zero EOT reveals the influence of a drop voltage induced by a dipole layer (δ) of about 45 meV (see Fig. 3.16) with the incorporation of 2 Å of La in a *sacrificial* gate approach (22.5 meV/Å).

Moreover, no significant roll-off effect is evidenced at nominal EOT. Nevertheless, it is worth noting that nominal EOT achieved with HfSiON-beveled devices ($\approx 13 \text{ \AA}$) is higher than that of HfON-based devices ($\approx 8 \text{ \AA}$). A direct comparison of ΔWF_{eff} at $EOT = \approx 8 \text{ \AA}$ is therefore not authorized. However, it must be observed that WF_{eff} shift at zero EOT induced by the diffusion of La into HfSiON devices is lower (45 meV) than that induced by the diffusion of the same as-deposited La dose into HfON devices (145 meV).

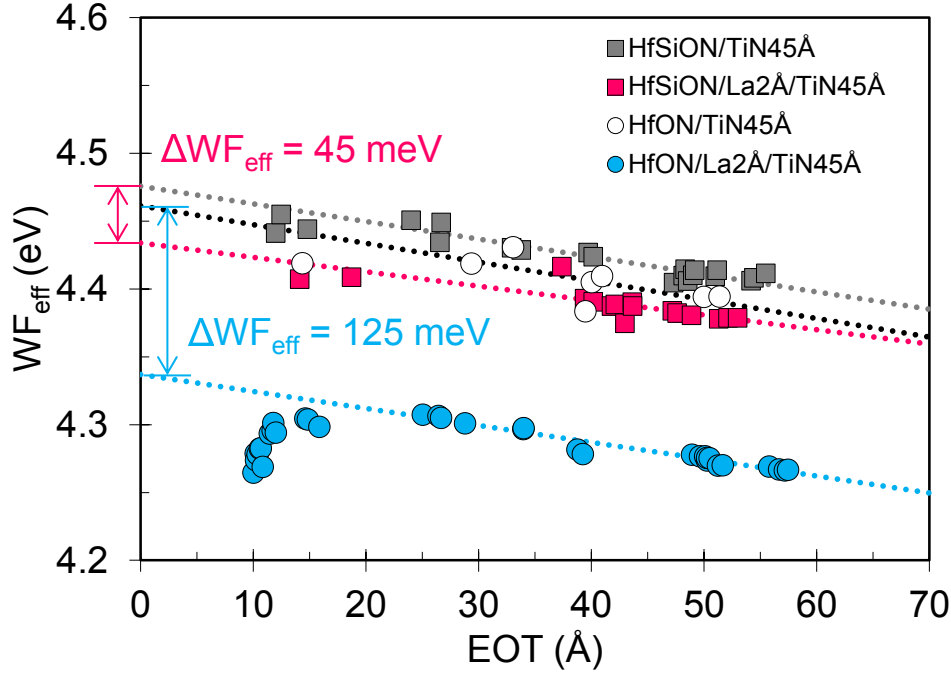


Figure 3.16: Comparison of WF_{eff} shift at $EOT = 0$ induced by a TiN45 Å stack without and with La 2 Å inserted in a *sacrificial* approach on HfSiON-based and HfON-based devices.

Discrepancies in the ΔWF_{eff} modulation at zero EOT for equivalent La doses are discussed in this part and can be attributed to two origins: 1) the differences of coefficient of diffusion between HfON and HfSiON deposited in the conditions described in the beginning of this section, and 2) the differences in strength of the potential drop induced by the substitution of an atom of Hf by La at a HfON/SiON interface compared to a HfSiON/SiON interface. By using the results of Figs. 3.15 and 3.16, we can compare the WF_{eff} shift at $EOT = 0$ for equivalent effective La doses into both high- κ devices. For HfSiON-devices, a WF_{eff} shift of around 45 meV is obtained with 20% of $\approx 2 \text{ \AA}$ of La ($\approx 8.27 \times 10^{13} \text{ at/cm}^2$) and for HfON-devices, a WF_{eff} shift of around 145 meV is obtained with 47% of $\approx 2 \text{ \AA}$ of La ($\approx 2.36 \times 10^{14} \text{ at/cm}^2$). This corresponds to a WF_{eff} shift at $EOT = 0$ of -54 meV for each La dose of $1 \times 10^{14} \text{ at/cm}^2$ incorporated into HfSiON layer and a WF_{eff} shift at $EOT = 0$ of -53

meV for each La dose of 1×10^{14} at/cm² incorporated into HfON layer. It can therefore be concluded that the ΔW_{eff} at zero EOT is proportional to the effective lanthanum dose incorporated into the high- κ /SiO_x stack. It is more significant on HfON compared to HfSiON because the effective La dose incorporated into HfON is proportionally more significant on HfON compared to HfSiON for equivalent as-deposited La doses.

Nevertheless, as explained in section 2.2.2, one of the limitations of the XRF technique is the impossibility to be used for depth profiling measurements. As a result, the effective La dose into high- κ /SiO_x stack accurately quantified by XRF does not allow to discriminate the effective La dose at high- κ /SiO_x interface. As a perspective, in order to investigate the origin and to account for the disparity of W_{eff} shift at same as-deposited La dose between HfON- and HfSiON-based devices, other physicochemical characterizations such as SIMS must be performed in order to complete the information about the diffusion of lanthanum into Hf-based dielectrics/SiON stacks. Moreover, *ab initio* simulations of the electrostatic dipole modulation due to cation substitution in HfON/SiON and HfSiON/SiON interfaces should help to improve the understanding of the W_{eff} modulation in devices with different high- κ materials.

3.3 State of the art of aluminum in gate stack

Likewise lanthanum for WF_{eff} shift towards N+, aluminum has extensively been used to shift the WF_{eff} towards P+. In order to reach the ΔWF_{eff} specification for PMOS, deposition of Al_2O_3 as high- κ dielectric [107] [108] [26], or above the high- κ layer [109] [22] and insertion of metallic aluminum by Physical Vapor Deposition in TiN gate [110] have already been proposed by many researchers. The mechanism behind such voltage shift might be explained by the formation of an electrostatic dipole layer at the high- κ /SiO_x interface in the gate stack, as mentioned in section 1.1.2.

Aluminum ion implantation through TiN is also very effective in achieving PMOS low V_{TH} . However, high implantation energies can induce Equivalent Oxide Thickness penalty and gate leakage degradation from 0.01 A/cm² to 5.53 A/cm² [111]. Such effect has been attributed to the accumulation of aluminum near SiO₂/channel interface rather than near to the HfO₂/SiO₂ interface [111]. SIMS depth profiles of aluminum implantation in Fig. 3.17 shows that the concentration of aluminum at SiO₂/Si interface increases by a factor 10 for high energy implantation and high Al doses (in magenta) compared to low energy implantation and low Al doses (in green).

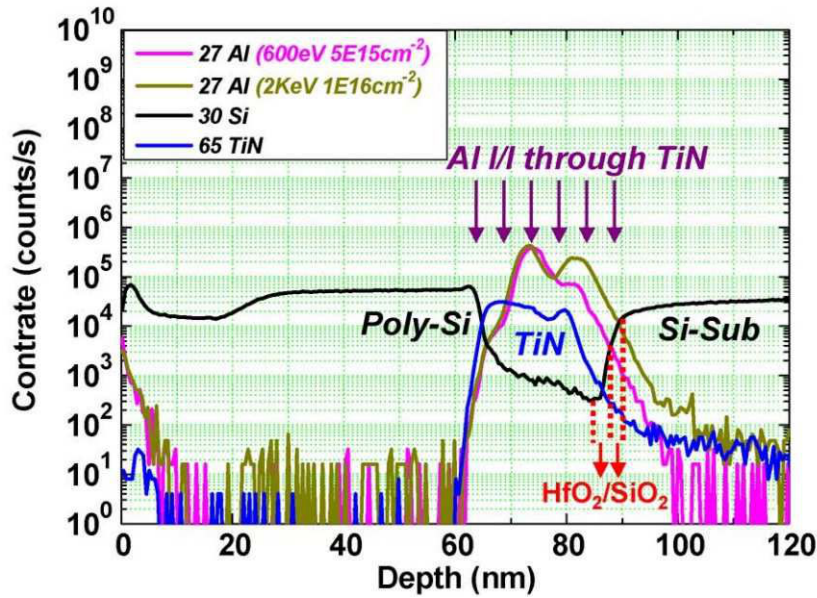


Figure 3.17: SIMS depth profiles of Al implantation through TiN (0.6 keV, $5 \times 10^{15} \text{ cm}^{-2}$ and 2keV, $5 \times 10^{15} \text{ cm}^{-2}$, according to [111].

3.3.1 Aluminum oxide as alternative gate dielectrics

Aluminum oxide (Al_2O_3) and other alumina-based compounds were originally considered to be good candidates for Flash [112], [113], [114] and CMOS gate dielectric applications [108], [115] mainly because of their dielectric constant, which is twice higher ($\kappa \approx 9$) than that of SiO_2 . Furthermore, the band gap of the Al_2O_3 is similar to that of SiO_2 , in contrast to most high- κ materials for which higher dielectric constant usually comes at the expense of narrower bandgap.

By using Al_2O_3 as high- κ oxide in $\text{SiO}_2/\text{Al}_2\text{O}_3/\text{TaCN}$ capacitors fabricated on terrace oxide wafers, Choi [108] revealed that the flatband voltage increases progressively by up to + 1.3 eV towards P+ with the alumina thickness. Such V_{fb} shift towards P+ was related to a dipole variation at the $\text{SiO}_2/\text{Al}_2\text{O}_3$ interface. Nevertheless, as the EOT values were reduced, the V_{fb} shift started to diminish. Choi attributed this roll-off phenomenon to the disruption of the dipole at the $\text{SiO}_2/\text{Al}_2\text{O}_3$ interface due to the presence of oxygen vacancies in the thin SiO_2 . The deficiency of oxygen atoms in the thin SiO_2 would result in an insufficient number of oxygen atoms to create maximum dipole moment. Choi also shown that the roll-off amount was independent of Al_2O_3 thickness and it was mainly controlled by the interfacial layer thickness. Such strong roll-off effect is one of the main limitations to integrate Al_2O_3 as alternative high- κ dielectrics of CMOS transistors. Another concern is that aluminum-based dielectrics have a lower crystallization temperature than nitrided silicates. In consequence, the mobility is degraded as a result of aluminum diffusion into the channel for higher temperatures (1000 °C and more). And unfortunately, such diffusion cannot be totally eliminated via an interfacial silicon oxynitride diffusion barrier [116].

3.3.2 Aluminum incorporation into Hf-based gate dielectrics

Nanocapping or sub-nanocapping Al_2O_3 film layers can also be incorporated either at the SiO_2/Hf -based oxide interface or above Hf-based gate dielectric. When a thin Al_2O_3 layer was deposited between the SiO_2 and HfO_2 layers, WF_{eff} increased up to +900 meV compared to SiO_2 references [9]. This shift was found to be independent of HfO_2 thickness (3-10 nm). According to Charbonnier et al. this shift cannot therefore be explained by fixed charges. In addition, the metal/ HfO_2 barrier height measured by internal photoemission measurements was insensitive to the presence of Al_2O_3 at the $\text{SiO}_2/\text{high-}\kappa$ interface. Such independence with different Al_2O_3 claims again for a dipole effect at the $\text{SiO}_2/\text{high-}\kappa$ interface.

Deposition of a thin Al_2O_3 layer on the top of HfO_2 or HfSiON also induced an increase of the WF_{eff} towards P+, but the shift was smaller (only +200 meV [109] [25]) compared to capacitors with Al_2O_3 below HfO_2 . K. Imawoto et al. demonstrated for the first time that V_{fb} shifts are mainly determined by the $\text{SiO}_2/\text{high-}\kappa$ interface [22]. For that, a bi-layer high- κ was deposited over a SiO_2 layer. By independently varying the thickness of the high- κ in contact with SiO_2 and the thickness of the top high- κ in a gate-last process, Iwamoto demonstrated that the V_{fb} shift behaviour is uniquely determined by the bottom high- κ/SiO_2 interface, as shown in Fig. 3.18. In consequence, the WF_{eff} shift observed with the insertion of Al above HfO_2 cannot be attributed to a dipole at the metal/high- κ interface. The shift towards P+ in a gate-first integration has been explained by the diffusion of Al towards the bottom HfO_2 interface [117] [109], creating a dipole layer at the $\text{HfO}_2/\text{SiO}_2$ interface as for Al_2O_3 directly deposited at this interface. The lower shift towards P+ can therefore be explained by a reduced Al diffusion through HfO_2 or HfSiO layers. Additionally, this mechanism as well as the direction of the V_{fb} shift is supported by the ab initio simulations performed on $\text{HfO}_2/\text{SiO}_2$ interfaces [28]. Indeed, the substitution of Hf by Al at this interface induces a decrease of the valence band offset and thus a WF_{eff} towards the P+ values [25].

However, despite the WF_{eff} variation range suitable for pFET on SOI devices, the introduction of alumina above the high- κ layer is not adequate for simultaneous nFET and pFET fabrication.

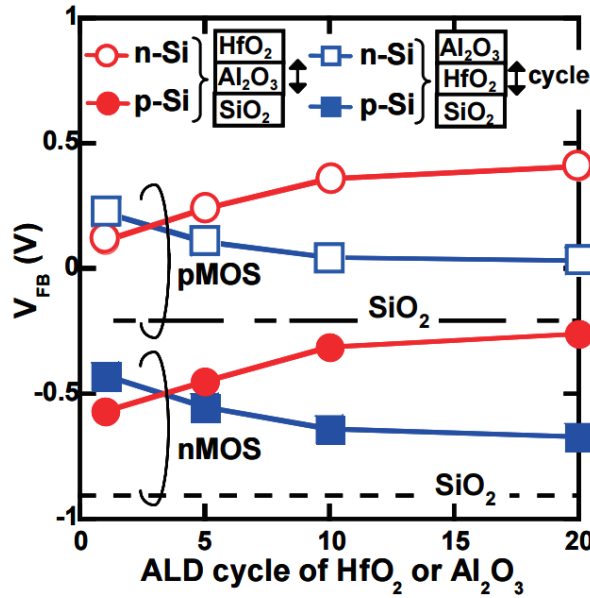


Figure 3.18: V_{fb} variations of n- and p-MOS capacitors with $\text{HfO}_2/\text{Al}_2\text{O}_3$ dielectrics [22].

3.3.3 Aluminum addition into TiN

The introduction of aluminum into TiN is another approach already reported in several articles, although opposite effects of the aluminum have been observed. For instance, increasing the Al concentration up to about 15% Al in $\text{Ta}_{1-x}\text{Al}_x\text{N}_y$ metal gates increased the WF_{eff} , but the work function decreased for higher Al concentrations. In addition, the WF_{eff} decreased with the increase of nitrogen content in the $\text{Ta}_{1-x}\text{Al}_x\text{N}_y$ metal gate [20]. At the time, the increase and subsequent drop in work function with Al content was correlated to the increase of Al-O bond formation and a subsequent reduction of the Al-O character at the metal/dielectric interface [20]. However, this explanation cannot explain the higher WF_{eff} observed in SiO_2 compared to HfO_2 observed in the same work.

The introduction of metallic aluminum deposited by PVD and inserted between 10 Å and 90 Å TiN layers (TiN/Al/TiN structure) has also been evaluated by Baudot et al. [110]. In that work, the two opposite effects of the insertion of metallic Al into TiN were clarified and are shown in Fig. 3.19: the first is the formation of dipoles at high- κ /SiO₂ interface, leading to a positive WF_{eff} shift; the second is the reduction of the TiN work function induced by the presence of Al within TiN after annealing at 1000 °C. Indeed, the work function of Al (4.2 eV) is inferior to the one of the TiN (between 4.4 and 4.6 eV). Formation of stable AlN precipitates originated from the reaction of Al with TiN for high Al doses is certainly preventing the Al diffusion into the high- κ layer [118] and should be the physical origin of the opposite effect.

Similar trends were found with TiAlN metal electrodes deposited over thin and thick HfO₂ layers, clarifying that the main mechanism is the formation of a dipole induced by aluminum diffusion. Furthermore, a reduction of the TiAlN/high- κ barrier was demonstrated by performing Internal Photoemission (IPE) measurements on 200 Å-thick HfO₂ capacitors [25].

We summarize in Fig. 3.20 the different integration schemes of the various aluminum insertion processes reported in the literature. It is worth noting that the WF_{eff} progressively decreases as the aluminum is further away from the high- κ /SiO₂ interface. Although intrinsically capable of large shift towards P+ in contact with SiO₂, the different architectures on Hf-based devices required from a process integration point of view can limit the aluminum diffusion towards the high- κ /SiO₂ interface. Besides, the opposite shift of the Al-based metal workfunction towards N+ limits the benefit of dipoles induced by Al diffusion.

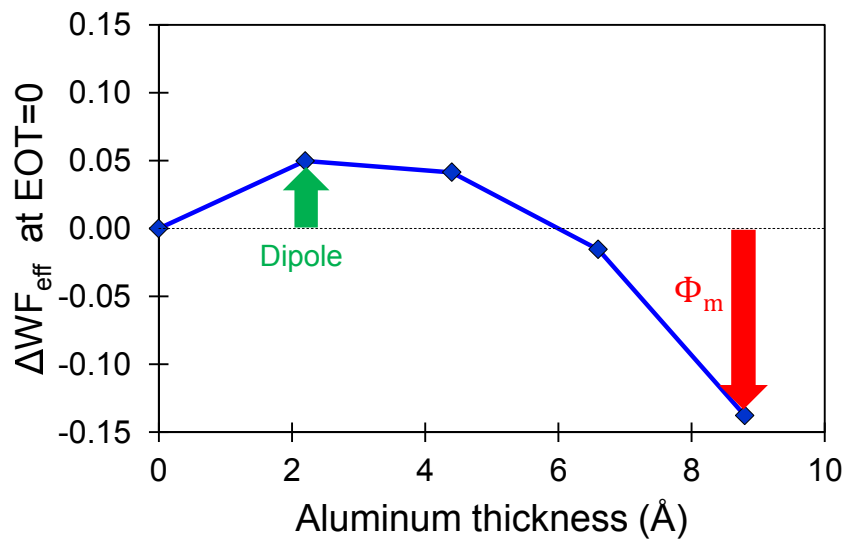


Figure 3.19: Effective work function shift at EOT=0 for different as-deposited Al thicknesses in TiN/Al/TiN/HfSiON/SiON gate stacks.

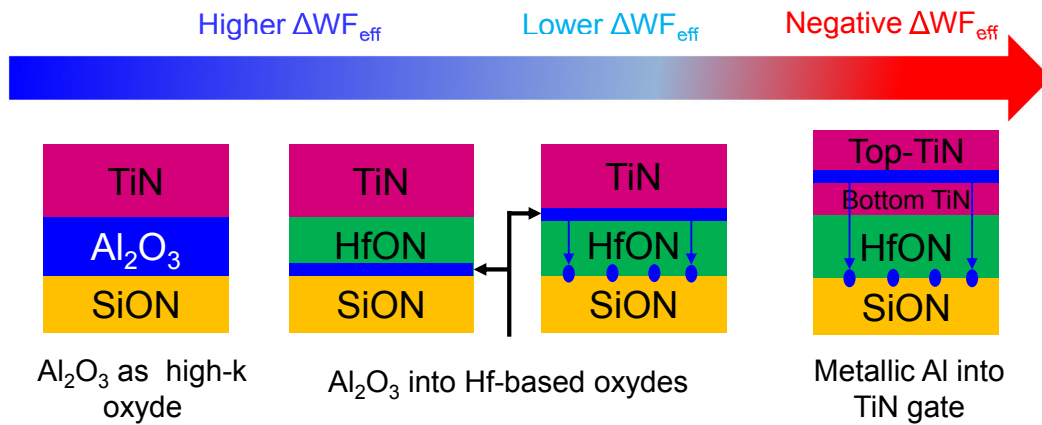


Figure 3.20: Integration scheme of the various aluminum insertion processes into gate stack in a conventional gate-first flow.

3.4 Effective work function modulation by accurate control of *sacrificial* aluminum diffusion

In this section, we evaluate for the first time the impact of metallic aluminum (Al) deposited by Radio-Frequency PVD in a *sacrificial* metal gate-first approach on the effective work function of HfON-based NFET devices [119]. Moreover, the Al diffusion is studied by X-ray Photoelectron Spectroscopy (XPS) and X-ray Fluorescence (XRF) measurements as a function of the aluminum dose, the composition of the TiN layers and the annealing temperature. In addition, the influence of high- κ on the diffusion of aluminum and its impact on the WF_{eff} is also evidenced. Finally, sacrificial gate first process efficiency to further increase effective work function (WF_{eff}) towards P+ by metallic aluminum diffusion is demonstrated in this section.

3.4.1 Device fabrication and electrical characterization

Nominal Low- V_{TH} NFET devices were fabricated with the simplified process flow of Fig. 2.12. Aluminum was incorporated following a *sacrificial* gate-first approach, as illustrated in Fig. 1.18. The impact of Al layer thickness on the WF_{eff} was investigated. For that, the *sacrificial* Si/TiN/Al/TiN gate stacks (Si on top) with different Al targeted thicknesses (0 Å, 2.2 Å, 4.4 Å, and 6.6 Å) were deposited. Bottom TiN and Top TiN were deposited with thicknesses of 10 Å and 25 Å, respectively. In addition, a supplementary layer of TiN with a thickness of 45 Å was deposited before the capping silicon layer of the *sacrificial* Si/TiN/Al/TiN structure in order to simulate the integration of P-gates in a standard 14 nm FDSOI devices. The thicknesses of TiN layers not were changed. In contrast, the influence of the composition of TiN was also investigated. For that, two other TiN compositions were deposited with a uniform profile of nitrogen into TiN. The stoichiometry of the TiN was modified by varying Ar and N₂ gas flow rates during deposition. For a N-rich TiN, $[N_2]/([N_2]+[Ar])=0.7$ and for a N-rich+ TiN, $[N_2]/([N_2]+[Ar])=1$. Once the *sacrificial* gate stack was deposited, it was followed by a thermal treatment under N₂ atmosphere at 900 °C or 1000 °C for 10 s in order to activate the diffusion of Al into the HfON/SiON stack. The *sacrificial* gate stack was then removed by wet etching. Finally, a Poly-Si/TiN electrode was deposited, followed by gate patterning and S/D spike annealing at 1005 °C, as illustrated in Fig. 1.18. The Poly-Si and the final TiN metal have thicknesses of 240 Å and 35 Å, respectively, for all the devices. All metal layers were deposited in Radio Frequency PVD chambers. The aluminum XRF line intensity was used to adjust deposition times in order to reach Al doses of $\approx 1.4 \times 10^{15}$ at/cm², $\approx 2.8 \times 10^{15}$ at/cm² and $\approx 4.2 \times 10^{15}$ at/cm² corresponding to 2.2 Å, 4.4 Å, and 6.6 Å, respectively. Samples description is summarized in Table 3.2. The aim of the

study is to demonstrate that WF_{eff} , and therefore the V_{TH} , can be modulated by tuning the diffusion of Al into the gate stack with different combinations of TiN compositions and Al thicknesses.

Sample	Al dose (at/cm ²)	Bottom TiN	Top/N gate TiN
TiN80 Å-No Al	0	10 Å	25 Å/45 Å
TiN10 Å/Al2.2 Å/TiN70 Å	1.4×10^{15}	10 Å	25 Å/45 Å
TiN10 Å/Al6.6 Å/TiN70 Å	4.2×10^{15}	10 Å	25 Å/45 Å
N-rich TiN10 Å/Al2.2 Å/N-rich TiN70 Å	1.4×10^{15}	10 Å	25 Å/45 Å
N-rich TiN10 Å/Al6.6 Å/N-rich TiN70 Å	4.2×10^{15}	10 Å	25 Å/45 Å
N-rich+ TiN10 Å/Al2.2 Å/N-rich+ TiN70 Å	1.4×10^{15}	10 Å	25 Å/45 Å
N-rich+ TiN10 Å/Al6.6 Å/N-rich+ TiN70 Å	4.2×10^{15}	10 Å	25 Å/45 Å

Table 3.2: Sample description of nominal devices.

CV curve in Fig. 3.21 shows the impact of the insertion of 0 Å, 2.2 Å, and 6.6 Å of aluminum on the V_{fb} modulation. The maximum ΔV_{fb} of around 100 mV is obtained with 2.2 Å or 1.4×10^{15} at/cm² of aluminum. However, this shift is decreased if we continue to increase the amount of Al dose, as observed in the case of 6.6 Å or 4.2×10^{15} at/cm² of as-deposited aluminum. Unlike lanthanum behaviour (Fig. 3.2), the V_{fb} shift obtained with the insertion of metallic aluminum in the *sacrificial* gate-first approach does not linearly increase with Al thickness in the *sacrificial* gate-first approach.

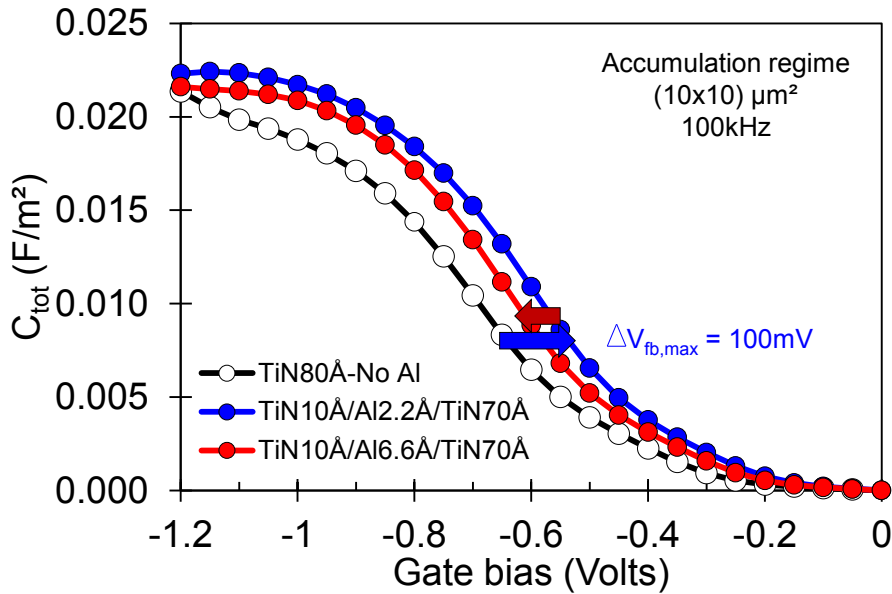


Figure 3.21: CV curve as a function of Al thickness deposited by RF-PVD at room temperature in sacrificial gate approach (annealing at 900 °C).

The reduction of the work function value (Φ_m) of the TiN with the insertion of aluminum has been confirmed by comparison of metal/oxide barrier heights for TiN and TiAlN metal layers on Internal Photoemission (IPE) measurements by Leroux et al. [25]. However, although the reduction of the Φ_m value with aluminum has been cited as the origin of a reduction of the WF_{eff} , it cannot, at first sight, explain the behaviour of the CV curves, since the *sacrificial* gate stack is removed after diffusion annealing. In other words, Al is not deposited again in the final Poly-Si/TiN electrode, so the opposite effect induced by the presence of Al within TiN should be excluded to explain this non-linear behaviour. In addition, the titanium K_α lines are also systematically analysed to ensure that sacrificial gate stack is correctly removed (see section 2.2.2.2). Therefore, possible traces of aluminum in the TiN due to issues to remove the sacrificial gate stack should also be excluded.

Beveled devices with Al thicknesses variations in the *sacrificial* gate structure were also fabricated in order to analyse the influence of fixed charges density and Al-induced dipoles at $\text{HfO}_2/\text{SiO}_2$ interface with the insertion of different aluminum thicknesses. Sample description is summarized in Table 3.3.

Sample	Al dose (at/cm ²)	Bottom TiN	Top/N gate TiN
TiN35 Å/TiN45 Å	0	35 Å	0 Å/45 Å
TiN10 Å/Al2.2 Å/TiN70 Å	1.4×10^{15}	10 Å	25 Å/45 Å
TiN10 Å/Al4.4 Å/TiN70 Å	2.8×10^{15}	10 Å	25 Å/45 Å
TiN10 Å/Al6.6 Å/TiN70 Å	4.2×10^{15}	10 Å	25 Å/45 Å

Table 3.3: Sample description of beveled devices.

EOT and V_{fb} were extracted from CV measurements for different dies of the wafer. WF_{eff} was then calculated from V_{fb} and the silicon work function relative to vacuum ($q\Phi_{\text{sc}}$) values (see section 2.1.3). Fig. 3.22 shows the WF_{eff} vs EOT plot for beveled devices with different Al thicknesses deposited by RF-PVD in a *sacrificial* gate-first approach and annealed at 900 °C for 10 s. Again, the linear dependence at largest EOT proves the existence of interface charges between channel and SiO_x and also no other charges in bulk SiO_x (no quadratic dependence). From the WF_{eff} vs EOT slope, fixed charges density is estimated to be $6.98 \times 10^{11} \text{ /cm}^2$, $1.43 \times 10^{12} \text{ /cm}^2$, $8.29 \times 10^{11} \text{ /cm}^2$, and $8.17 \times 10^{11} \text{ /cm}^2$ for 0 Å, 2.2 Å, 4.4 Å, and 6.6 Å of aluminum thickness, respectively.

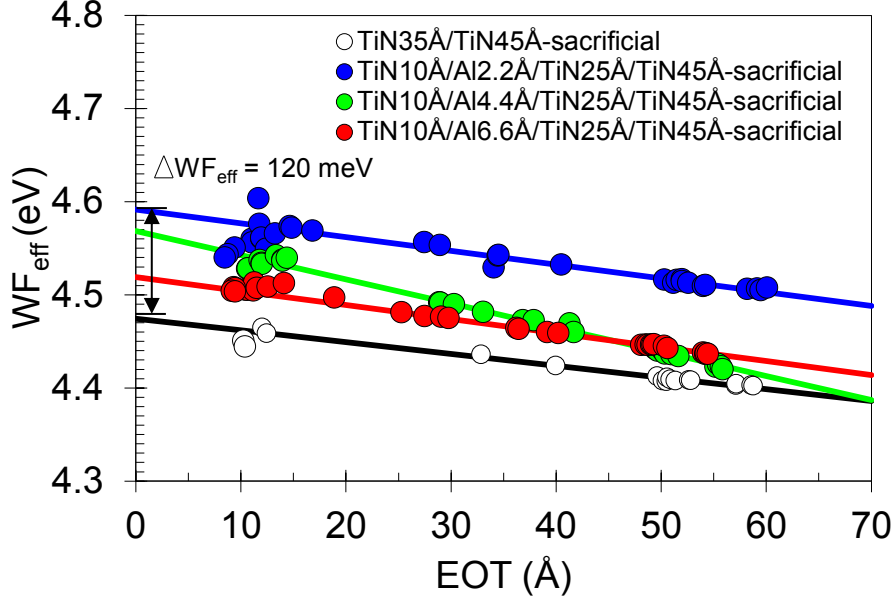


Figure 3.22: WF_{eff} vs EOT plot for beveled devices with thicknesses of 2.2 Å, 4.4 Å and 6.6 Å of aluminum deposited by RF-PVD in a *sacrificial* gate-first approach (annealing at 900 °C for 10 s).

The ΔWF_{eff} extrapolated at $EOT = 0$ reveals the influence of a maximum drop voltage induced by an Al dipole layer of about 120 meV for 2.2 Å. This maximum drop voltage cannot be attributed to a work function change (Φ_m) of TiN. Such ΔWF_{eff} decreases to 100 meV for 4.4 Å and to only 50 meV for 6.6 Å. Therefore, beveled devices also confirm the non linear behaviour of aluminum thickness after annealing at 900 °C for 10 s observed in nominal devices. On the other hand, the beveled oxide methodology is crucial to assess, once again, the abrupt potential drop (called roll-off) appears when the equivalent oxide thickness (EOT) becomes thinner than 20 Å. Such roll-off is only 20 meV for 2.2 Å of aluminum and negligible for 4.4 Å and 6.6 Å of aluminum. Again, the ΔWF_{eff} induced by effective Al diffusion at nominal EOT of 14 nm FDSOI devices can also be expressed as the sum of the dipole effect $(\Delta WF_{\text{eff}})_{EOT=0}$ related to Al diffusion and its roll-off ($\text{roll-off}_{w/\text{Al}}$) minus the roll-off associated to sacrificial stack without Al ($\text{roll-off}_{w/o \text{ Al}}$), as given in Eq. 3.4.

$$\Delta WF_{\text{eff}} = (\Delta WF_{\text{eff}})_{EOT=0} + \text{roll-off}_{w/\text{Al}} - \text{roll-off}_{w/o \text{ Al}} \quad (3.4)$$

In section 3.2.1, it was concluded in the case of lanthanum that the potential drop related to the dipole increases with the effective lanthanum diffusion. Similarly, it is therefore logical to assume that the effective aluminum dose incorporated into HfON/SiON stack reaches a maximum for 2.2 Å and decreases for 4.4 Å and 6.6 Å of as-deposited aluminum. In section 3.4.2, we will measure the effective aluminum dose incorporated into HHfON/SiON stack after

annealing at 900 °C for 10 s for different as-deposited aluminum doses. The aim of the study is to prove that the dipole effect is modulated with the effective dose of aluminum incorporated into the HfON/SiON stack.

Furthermore, no modulation of the V_{fb} with the variation of nitrogen composition from 51.6% to 54.6% of TiN layers is reported on devices with *nominal* dielectrics in Fig. 3.23. The incorporation of more nitrogen into the *sacrificial* TiN seems to have no impact on the WF_{eff} modulation, in agreement with the results obtained in [120].

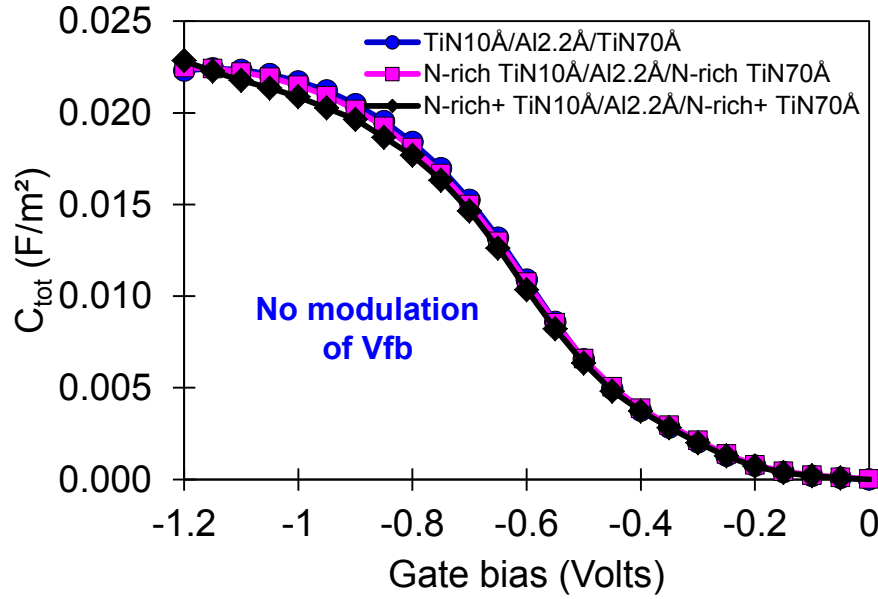


Figure 3.23: CV curve for devices with nominal dielectrics and different TiN compositions (51.6% for stoichiometric TiN, 52.6% for N-rich TiN and 54.6% for N-rich+ TiN), with the insertion of 2.2 Å of Al in a *sacrificial* gate approach (annealed at 900 °C).

The WF_{eff} vs EOT plot for nominal devices with aluminum inserted in the sacrificial gate and for different TiN compositions is shown in Fig. 3.24. Notice that a WF_{eff} vs EOT trend of around 70 meV/Å can be traced. The bigger is the ΔWF_{eff} shift, the greater is the reduction of EOT. Moreover, the EOT reduction is not accompanied by any degradation of leakage (J_g), as clearly demonstrated in Fig. 3.25. Instead, the insertion of aluminum in the *sacrificial* gate-first approach enables to improve the J_g -EOT trade-off, suggesting that a higher- κ value of HfON is achieved rather than an interlayer physical reduction.

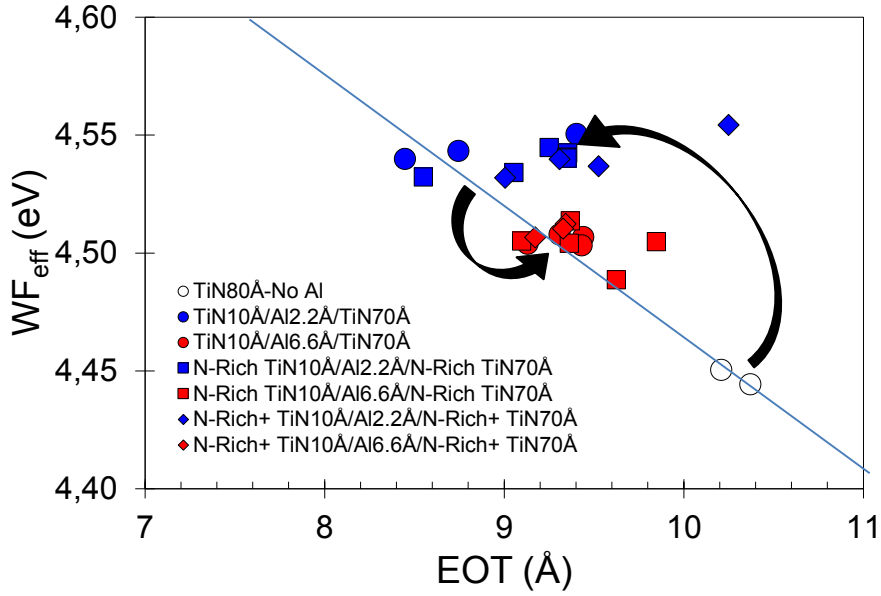


Figure 3.24: WF_{eff} vs EOT for devices with nominal HfON/SiON dielectrics and different Al thicknesses (annealed at 900 °C). Blue (red) symbols refer to 2.2 Å (6.6 Å) of aluminum.

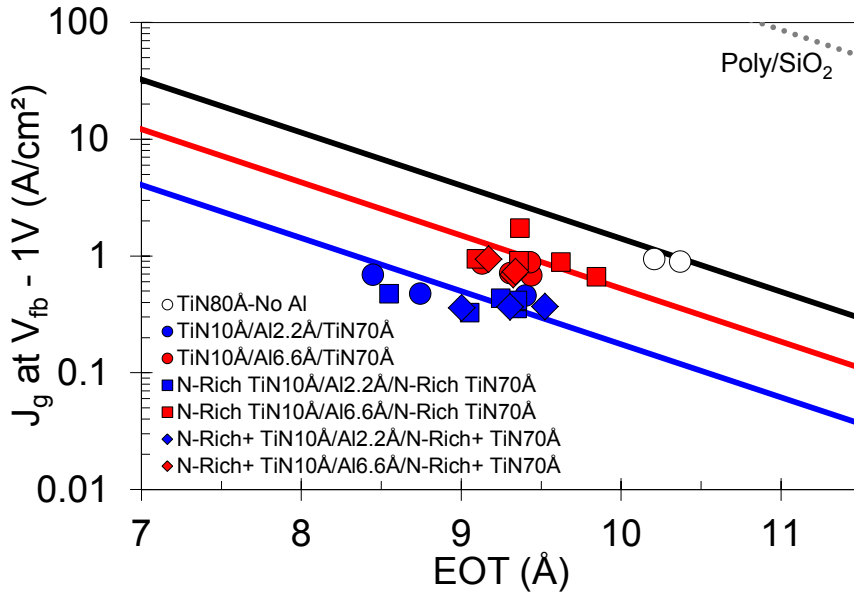


Figure 3.25: J_g vs EOT for devices with nominal HfON/SiON dielectrics and different Al thicknesses (annealed at 900 °C). Blue (red) symbols refer to 2.2 Å (6.6 Å) of aluminum.

3.4.2 Diffusion characterization by X-Ray Spectroscopy

The non-linear dependence of the WF_{eff} with the as-deposited Al thickness is clarified by X-Ray Fluorescence and X-Ray Photoemission Spectroscopy measurements performed a) after sacrificial gate stack deposition and b) after sacrificial gate stack removal [119] (see section 2.2.2.2). The aim of the study is to prove that the dipole effect is modulated with the effective dose of aluminum incorporated into the HfON/SiON stack. We report in Table 3.4 the variations of Al process (dose and annealing temperature) on blanket wafers that have been characterized by XRF and XPS.

Sample	Aluminum	Temperature	Bottom TiN	Top/N-gate TiN
1	2.2 Å	900 °C	10 Å	25 Å/45 Å
2	6.6 Å	900 °C	10 Å	25 Å/45 Å
3	11 Å	900 °C	10 Å	25 Å/45 Å
4	2.2 Å	1000 °C	10 Å	25 Å/45 Å
5	6.6 Å	1000 °C	10 Å	25 Å/45 Å
6	11 Å	1000 °C	10 Å	25 Å/45 Å

Table 3.4: Variations of Al dose and annealing temperature on blanket wafers for sacrificial aluminum diffusion study by XRF.

We notice in Fig. 3.26 a substantially lower Al diffusion at 900 °C in the case of a 6.6 Å-thick Al layer compared to the case of a 2.2 Å-thick Al layer, corroborating the electrical results shown in Figs. 3.21, 3.22, and 3.24. Indeed, XRF measurements show that Al diffusion into the HfON/SiON stack is limited even after annealing at 900 °C : only 24% for 2.2 Å of as-deposited Al compared to 6% for 6.6 Å of as-deposited Al. Accordingly, this result helps to understand why the WF_{eff} shift achieved with 6.6 Å of aluminum inserted in the sacrificial gate is inferior to that of 2.2 Å of aluminum. Indeed, even if we increase the Al dose (from 1.4×10^{15} at/cm² to 4.2×10^{15} at/cm²), less Al is incorporated into the HfON/SiON stack after annealing at 900 °C for 10 s (2.3×10^{14} at/cm² vs 4×10^{14} at/cm², respectively).

XRF results also indicates that an as-deposited Al thickness by 11 Å is required to further increase the effective Al dose. Nevertheless, X-ray Photoelectron Spectroscopy measurements in Fig. 3.27 reveal a Ti_{2p} peak abnormally intense after sacrificial gate wet removal. This means that the solution employed to remove the *sacrificial* TiN with Al inserted as additive might not be efficient for aluminum doses as high as 7×10^{15} at/cm² or 11 Å. In other words, the XRF signal intensity might come from the as-deposited aluminum encapsulated by the TiN layers of the sacrificial gate that has not been removed.

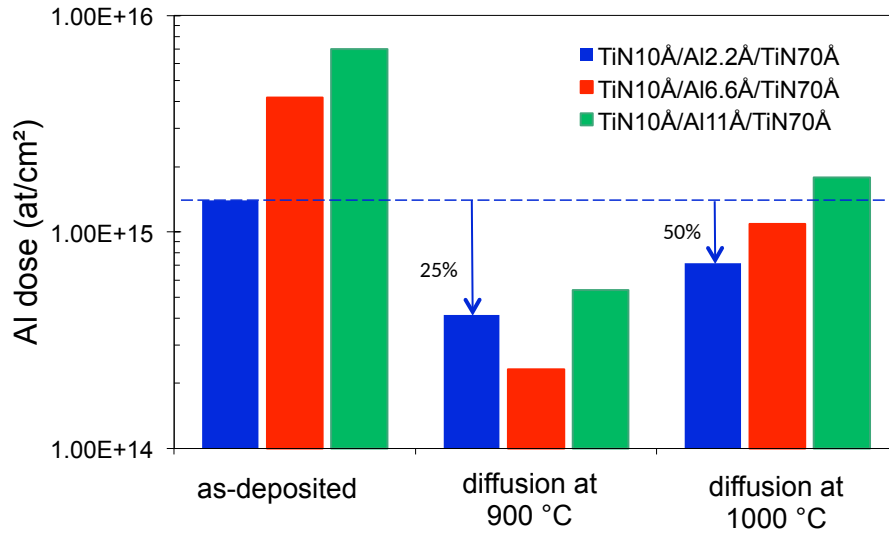


Figure 3.26: Quantification of Al dose by X-Ray Fluorescence on as-deposited sacrificial stack (left) and after sacrificial gate removal for annealing at 900 °C (middle) and at 1000 °C (right).

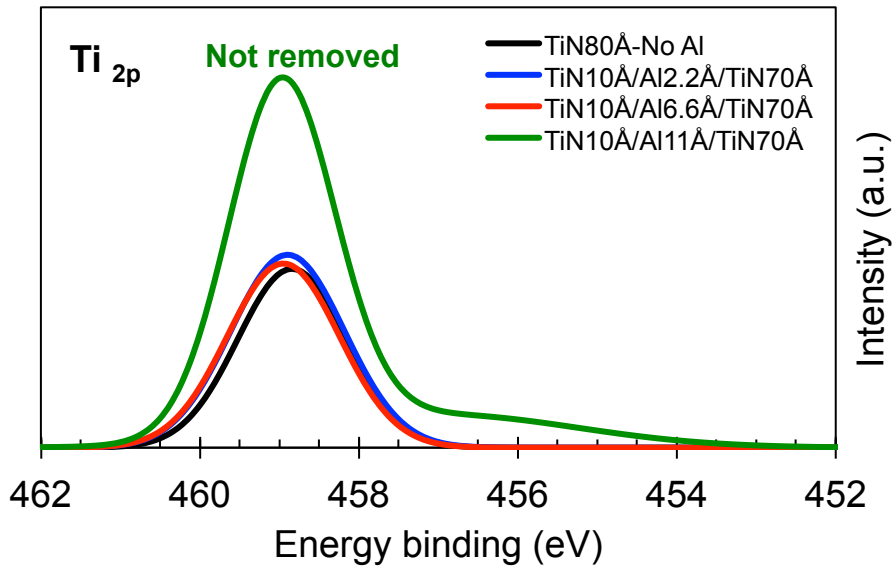


Figure 3.27: X ray Photoelectron Spectroscopy of Ti_{2p} peak after sacrificial gate removal by wet etching for different Al doses (annealing at 900 °C).

According to these XRF measurements, another way to incorporate significantly Al into the HfON/SiON stack is to perform an annealing at 1000

°C. This results in the linear increase of effective Al incorporated into dielectrics stack with the as-deposited Al thickness inserted in the *sacrificial* gate. Nevertheless, it must be clarified that Ti_{2p} was not measured for blanket wafers annealed at 1000 °C. Indeed, it is not impossible that the linear increase of effective Al dose measured by XRF is, again, due to issues to remove the *sacrificial* gate after annealing at 1000 °C. As the temperature increases from 900 °C to 1000 °C, the wet solution appears to be not efficient to remove the *sacrificial* stack, in particular the 200 Å-thick Poly-Si layer used as Si-cap for diffusion annealing.

Indeed, no obvious *sacrificial* gate stack removal at temperatures higher than 900 °C is confirmed by Scanning Electron Microscopy (SEM) measurements on nominal 14 nm FDSOI devices in Fig. 3.28. These pictures were performed after gate patterning on wafers annealed at 900 °C and 1000 °C. As expected, HfON-based devices annealed at 1000 °C were not functional probably because of a short circuit due to the impossibility to pattern the Poly-Si/TiN *final* electrode with the film residues of the *sacrificial* stack. It is possible that the the silicon capping forms a thicker oxide layer after annealing at 1000 °C.

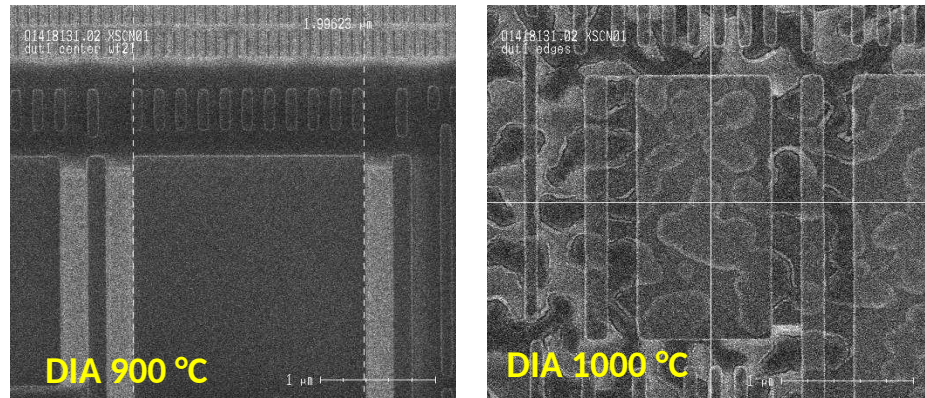


Figure 3.28: SEM pictures of 14 nm FDSOI devices after gate patterning. On the left, the *sacrificial* stack was annealed at 900 °C. On the right, the *sacrificial* stack was annealed at 1000 °C.

On the other hand, the simultaneous impact of the annealing temperature and the as-deposited aluminum dose on the Al diffusion is also observed through XPS measurements in Fig. 3.29. Note that while less Al is incorporated into the HfON/SiON stack after annealing at 900 °C, the incorporated Al is tremendously increased after annealing at 1000 °C for 6.6 Å. As a reminder, the analysis depth of XPS technique is limited to a few nanometers (\approx 8-9 nm). In consequence, the XPS intensity observed for aluminium is not coherent with SEM pictures shown in Fig. 3.28.

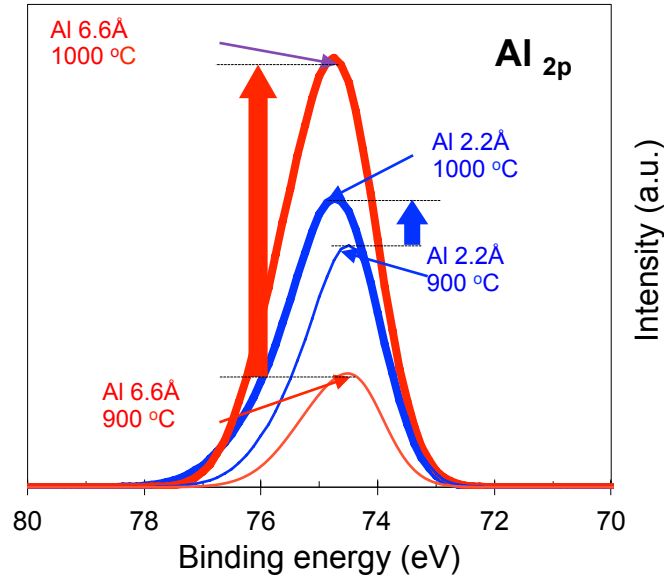


Figure 3.29: X ray Photoelectron Spectroscopy of Al_{2p} peak after sacrificial gate deposition, annealing and wet removal for different Al doses and annealing temperatures.

In order to assess the crucial role of the annealing temperature on aluminum diffusion without dealing with the wet solution removal, ToF-SIMS measurements were performed on blanket wafers directly on the Si-capping of the *sacrificial* gate stack with 2.2 Å of aluminum on 200 Å-thick HfO_2 and annealed at 800 °C, 900 °C, and 1000 °C for 10 s. The aluminum profile after diffusion annealing is shown in Fig. 3.30. It is evident that the amount of aluminum incorporated into HfO_2 increases with the annealing temperature.

Furthermore, a linear relationship between WF_{eff} shift (extracted from CV measurements in section 3.4.1) and the effective Al dose incorporated into HfON/SiON stack (Al_{eff}) is highlighted in Fig. 3.31, corresponding to a shift of +25 meV for each Al dose of 1×10^{14} at/cm² incorporated into HfON/SiON stack. It is worth noting that the electrical impact of the lanthanum is stronger than the one of the aluminum for the same effective dose incorporated into HfON/SiON stack (i.e. -80 meV and +25 meV for 1×10^{14} at/cm² of effective lanthanum and aluminum doses, respectively).

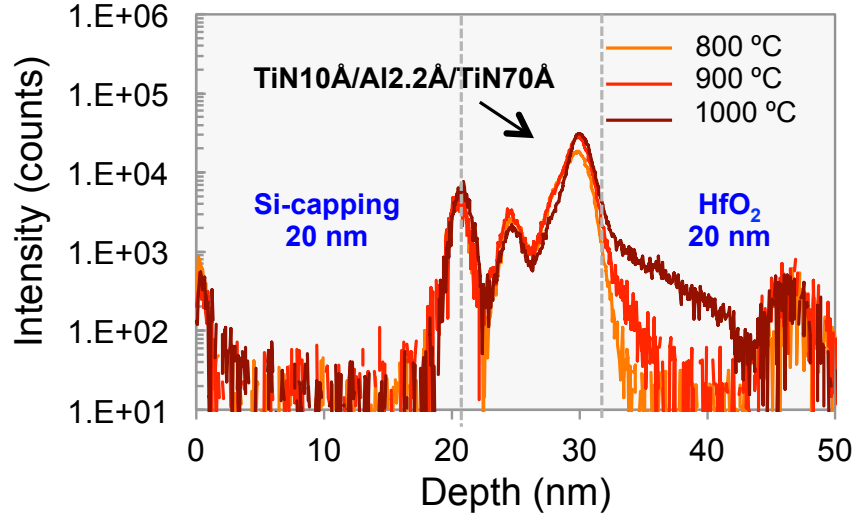


Figure 3.30: ToF-SIMS depth profiles of 2 Å-thick Al in 20 nm-thick HfO₂ as a function of annealing temperature. Si⁺, Ti⁺ and HfO⁺ ions spectra (not shown here) were used to place Si-capping/TiN and TiN/HfO₂ interfaces.

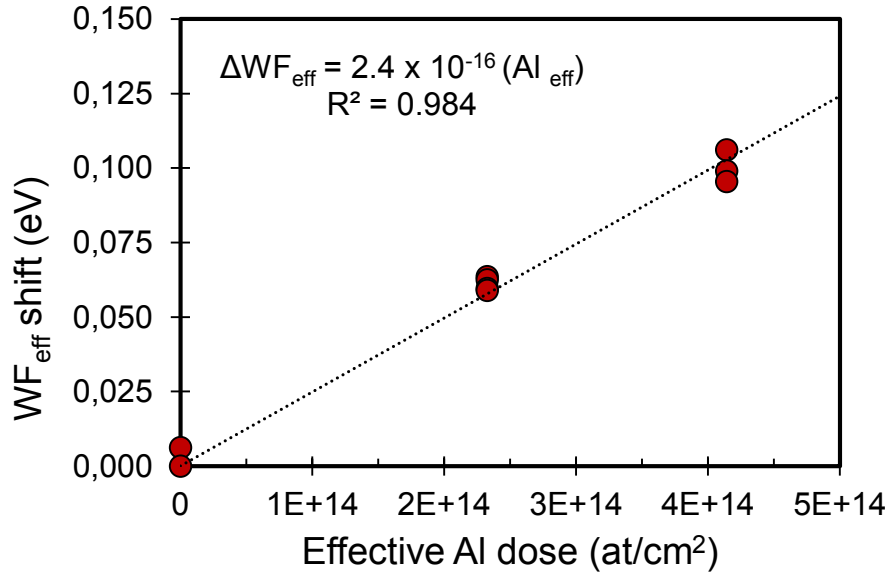


Figure 3.31: WF_{eff} shift increase towards P⁺ on HfON-based NFET devices vs effective aluminum dose incorporated into HfON/SiON stack.

In order to further increase the WF_{eff} towards P+, we have seen the role of the annealing temperature to further increment the effective aluminum dose incorporated into the HfON/SiON stack. We will examine now the impact of the thickness of the bottom TiN on the aluminum diffusion after annealing at 900 °C and 760 °C. Four different bottom TiN thicknesses (0 Å, 5 Å, 10 Å, and 25 Å) were evaluated by applying the methodology explained in section 2.2.2.2. XRF results are illustrated in Fig. 3.32. The effective dose is simultaneously modulated by the bottom TiN thickness and the annealing temperature. As expected, aluminum diffusion decreases with the annealing temperature and with the increase of the bottom TiN thickness. Again, the ratio $Al_{\text{eff}} / Al_{\text{dep}}$ in % can be reported to analyse the percentage of effective aluminum diffusion. In Fig. 3.33, the percentage of the aluminum dose incorporated into HfON/SiON is quantified as a function of the bottom TiN thickness after annealing at 900 °C and 760 °C. Just like lanthanum behaviour, the ratio $Al_{\text{eff}} / Al_{\text{dep}}$ evidences an exponential dependence with the bottom TiN thickness. Similarly, the effective aluminum dose Al_{eff} may be accurately controlled and predicted from the empirical equation 3.5, knowing La_{dep} and bottom TiN thickness (d) values. As a reminder, the pre-exponential coefficient A indicates the percentage of effective aluminum diffusion without TiN and the coefficient B represents the efficiency of TiN to screen the diffusion of aluminum at 900 °C. However, Eq. 3.5 is only valid after annealing at 900 °C and becomes Eq. 3.6 after annealing at 760 °C.

$$Al_{\text{eff}}/Al_{\text{dep}} = 47.1 \exp(-0.077 d) \quad (3.5)$$

$$Al_{\text{eff}}/Al_{\text{dep}} = 12.4 \exp(-0.138 d) \quad (3.6)$$

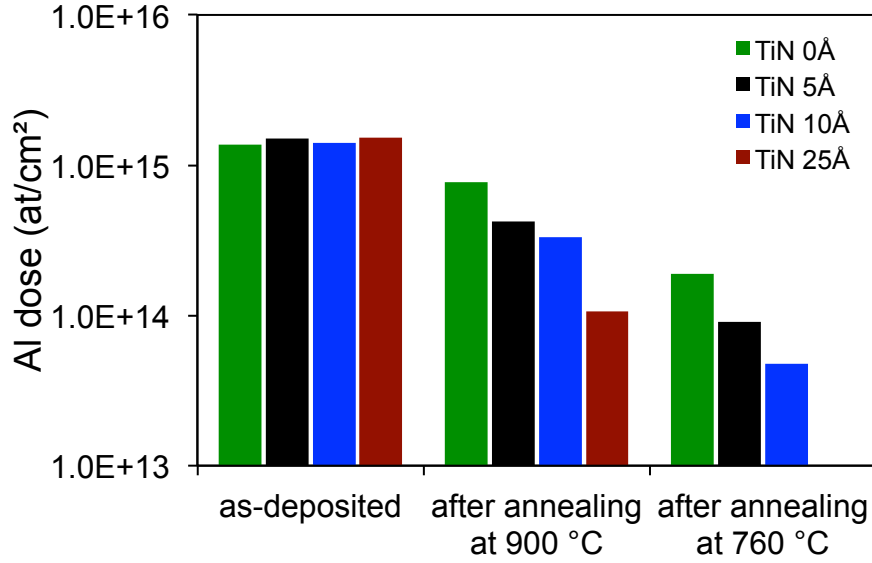


Figure 3.32: Quantification of Al dose by X-Ray Fluorescence on as-deposited wafers and after sacrificial gate removal for annealing at 900 °C and 760 °C for 10 seconds as a function of bottom TiN thickness.

By reducing the thickness of the TiN from 10 Å to 0 Å, the incorporated aluminum can be augmented until 47% of the as-deposited dose after annealing at 900 °C but only to 12.4% after annealing at 760 °C, as shown in Fig. 3.33. In addition, the efficiency of TiN to screen the diffusion of aluminum decreases as the annealing temperature increases. Henceforth, the effective diffusion of lanthanum and aluminum as a function of bottom TiN thickness after annealing at 900 °C can be compared, as indicated in Fig. 3.34. It is worth noting that Al and La diffuse similarly into HfON ($\approx 47\%$) but bottom TiN is more efficient to screen the diffusion of La than that of Al, as deduced by comparing empirical Eqs. 3.2 and 3.5. It is not a surprising result given the size of La and Al atoms.

In summary, the *sacrificial* gate-first approach also allows the prediction and accurate control of diffusion of aluminum into HfON. Furthermore, the linear relationship between WF_{eff} and Al_{eff} dose makes this approach suitable for very fine V_{TH} adjustment in 14FDSOI devices.

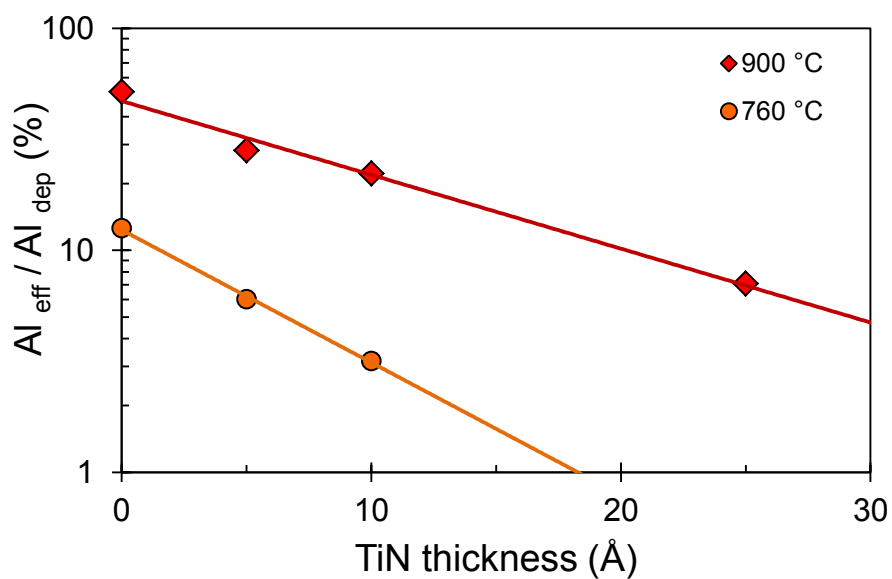


Figure 3.33: Percentage of effective aluminum dose incorporated into HfON/SiON as a function of the bottom TiN thickness after annealing at 900 °C and 760 °C for 10 s.

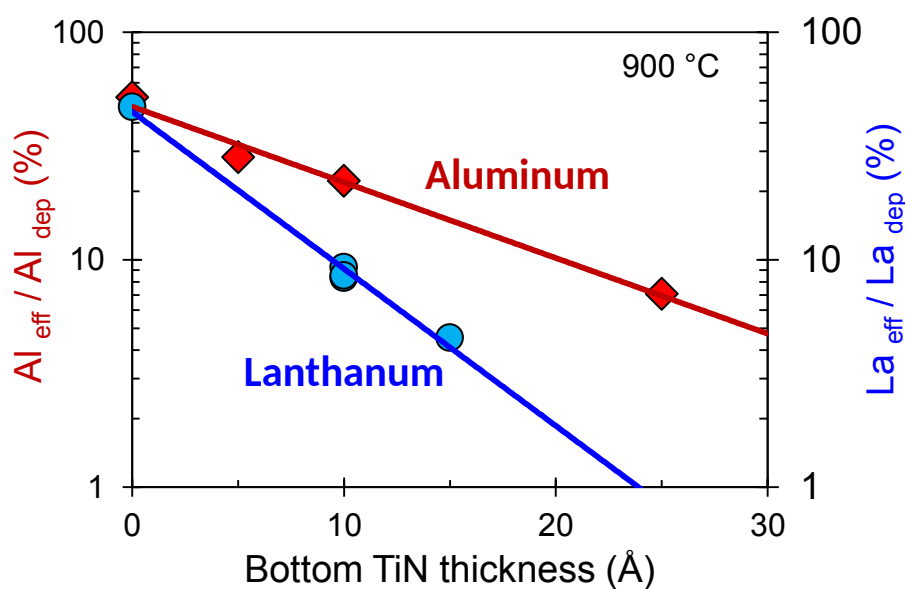


Figure 3.34: Comparison of percentage of effective aluminum and lanthanum doses incorporated into HfON/SiON as a function of the bottom TiN thickness after annealing at 900 °C for 10 s.

3.4.3 Influence of high- κ dielectrics on the effective work function shift induced by aluminum incorporation

In this section we will compare the impact of the effective Al diffusion on the effective work function shift of both HfON- and HfSiON-based devices. First, the diffusion of aluminum into both high- κ materials is studied on blanket wafers by using the XRF methodology, as shown in Fig. 3.12. Then, HfSiON-based NFET devices with and without 2 Å of Al and with 10 Å-thick bottom TiN in the *sacrificial* stack were fabricated with the simplified process flow of Fig. 2.12. In addition, beveled thermally grown SiO₂ was used as interlayer dielectric in these devices in order to investigate the origin of the WF_{eff} modulation, as explained in section 2.1.2.2. The purpose of the study is to assess the effective work function shift induced by aluminum diffusion into these different high- κ materials.

We compare the Al diffusion into HfON and HfSiON after annealing at 900 °C for 10 s. A 10 Å-thick bottom TiN was deposited between the high- κ layer and the 2 Å of aluminum. Both HfSiON and HfON were deposited with a thickness of 20 Å. We observe in Fig. 3.35 that Al clearly diffuses more easily into HfON than into HfSiON, 24% instead of 12%, at this temperature.

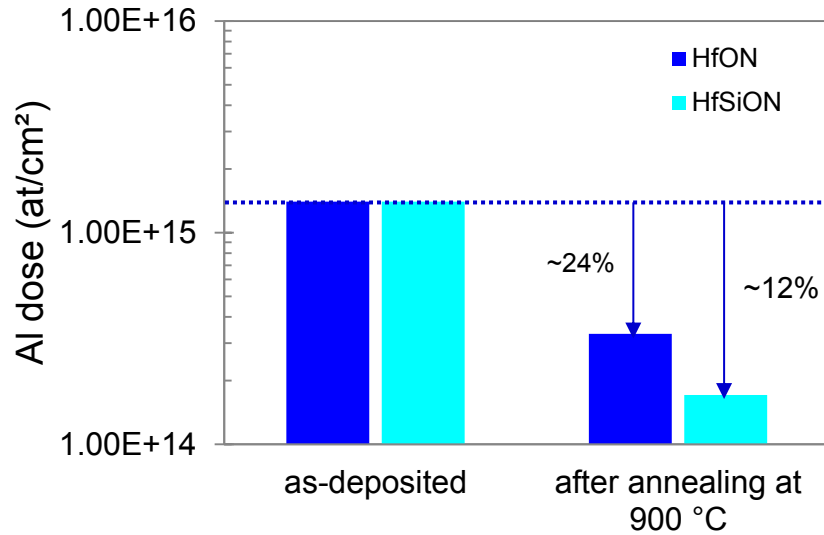


Figure 3.35: Al diffusion into both oxides after annealing at 900 °C for 10 s. At this temperature, Al clearly diffuses more easily into HfON than into HfSiON, 24% instead of 12%.

The WF_{eff} vs EOT plot for HfSiON/beveled-SiO_x devices is shown in Fig. 3.36. The ΔWF_{eff} extrapolated at zero EOT reveals the influence of a drop voltage induced by a dipole layer (δ) of about 70 meV with the incorporation of 2 Å of Al with 10 Å-thick bottom TiN inserted in a *sacrificial* gate approach (35 meV/Å). A direct comparison of ΔWF_{eff} at EOT = ≈ 8 Å is not authorized due to different nominal EOT achieved with each high- κ oxide. However, it must be observed that WF_{eff} shift at zero EOT induced by the diffusion of Al into HfSiON devices is lower (70 meV) than that induced by the diffusion of the same as-deposited La dose into HfON devices (120 meV).

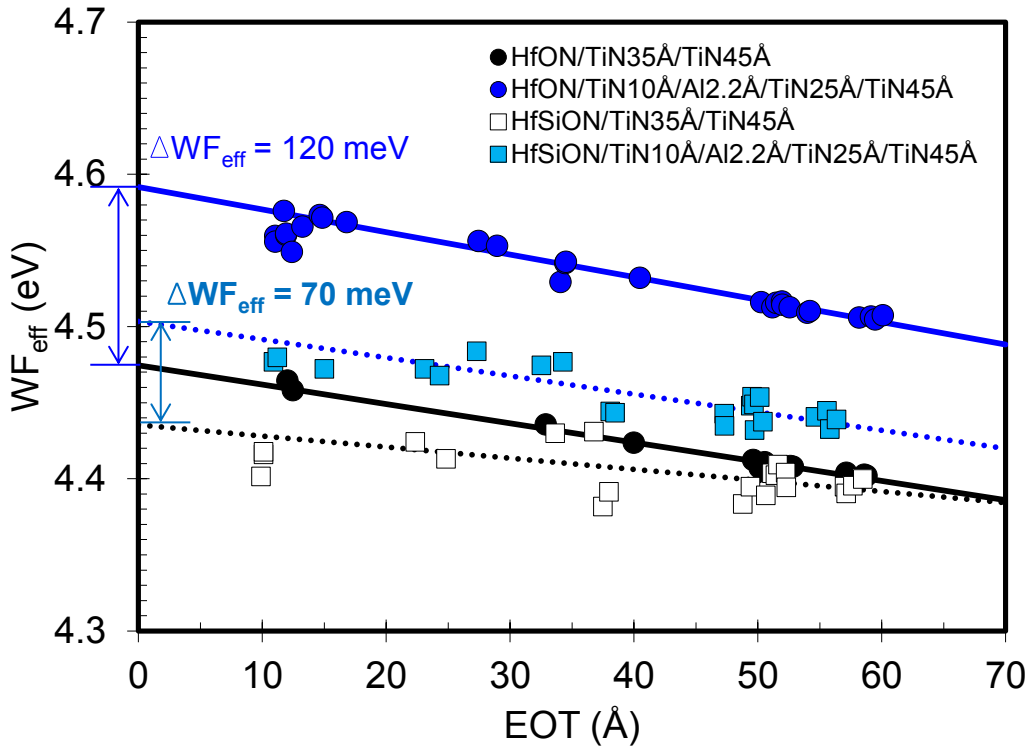


Figure 3.36: Comparison of WF_{eff} shift at EOT = 0 induced by the incorporation of 2 Å of aluminum with 10 Å-thick bottom TiN inserted in a *sacrificial* approach on HfSiON-based and HfON-based devices.

Discrepancies in the ΔWF_{eff} modulation at zero EOT for equivalent Al doses could also be attributed to two origins: 1) the differences of coefficient of diffusion between HfON and HfSiON oxides and 2) the differences in strength of the potential drop induced by the substitution of an atom of Hf by Al at a HfON/SiON interface compared to a HfSiON/SiON interface. The results of Figs. 3.35 and 3.36 allow the comparison of the WF_{eff} shift at EOT = 0 for equivalent effective Al doses into both high- κ devices. For HfON-devices, a WF_{eff} shift of around 120 meV is obtained with 24% of ≈ 2 Å of Al ($\approx 3.33 \times 10^{14}$ at/cm²) and for HfSiON-devices, a WF_{eff} shift of around 70 meV is

obtained with 12% of $\approx 2 \text{ \AA}$ of Al ($\approx 1.71 \times 10^{14} \text{ at/cm}^2$). This corresponds to a WF_{eff} shift at $\text{EOT} = 0$ of +36 meV for each Al dose of $1 \times 10^{14} \text{ at/cm}^2$ incorporated into HfSiON layer and a WF_{eff} shift at $\text{EOT} = 0$ of +41 meV for each Al dose of $1 \times 10^{14} \text{ at/cm}^2$ incorporated into HfON layer. Accounting for this weak difference, it is therefore logical to conclude that the $\Delta\text{WF}_{\text{eff}}$ at zero EOT is proportional to the effective aluminum dose incorporated into the high- κ/SiO_x stack. It is more significant on HfON compared to HfSiON because the effective Al dose incorporated into HfON is proportionally more significant on HfON compared to HfSiON for equivalent as-deposited Al doses.

Henceforth, we can also compare the electrical impact at $\text{EOT} = 0$ of both lanthanum and aluminum for the same effective dose incorporated into HfON/SiON stack. It should be noted that roll-off at nominal EOT was considered when it has been said that lanthanum shifts the WF_{eff} around -80 meV for each $1 \times 10^{14} \text{ at/cm}^2$ effective lanthanum dose whereas aluminum does around +25 meV for each $1 \times 10^{14} \text{ at/cm}^2$ of effective aluminum dose. Such roll-off effect has been taken into account for V_{TH} modulation of the 14 nm FDSOI devices in industrial applications. However, the strength of the dipole without roll-off can also be determined for effective doses of aluminum and lanthanum incorporated into HfON/SiON stack. This corresponds to a WF_{eff} shift at $\text{EOT} = 0$ of **+41 meV** for each Al dose of $1 \times 10^{14} \text{ at/cm}^2$ and a WF_{eff} shift at $\text{EOT} = 0$ of **-53 meV** for each La dose of $1 \times 10^{14} \text{ at/cm}^2$ incorporated into HfON layer. A higher relative shift induced by lanthanum is not surprising given the valence band offset (VBO) shifts induced by the substitution of Hf by La compared to the VBO shift induced by the substitution of Hf by Al [28].

Finally, as it has been said for the study of the diffusion of lanthanum, one of the limitations of the XRF technique is the impossibility to be used for depth profiling measurements. As a result, the effective Al dose into high- κ/SiO_x stack accurately quantified by XRF does not allow to discriminate the effective Al dose at high- κ/SiO_x interface. As a perspective, in order to investigate the origin and to account for the disparity of WF_{eff} shift at same as-deposited Al dose between HfON- and HfSiON-based devices, other physicochemical characterizations such as SIMS must be performed in order to complete the information about the diffusion of aluminum into different Hf-based dielectrics/SiON stacks. In addition, *ab initio* simulations of the electrostatic dipole modulation due to cation substitution in HfON/SiON and HfSiON/SiON interfaces should enable to improve the understanding of the WF_{eff} modulation in devices with different high- κ materials.

3.5 Sacrificial vs standard final approach

In this part, we cover the advantages of the sacrificial gate-first integration to incorporate lanthanum and aluminum into gate stack compared to a standard integration. First, we demonstrate the sacrificial gate first process capability to further increase effective work function (WF_{eff}) towards P+ by metallic aluminum diffusion. On the other hand, we show that La incorporated into TiN in a standard integration is useless for WF_{eff} modulation.

The two opposite effects of the insertion of metallic Al into TiN on WF_{eff} at $EOT = 0$ of HfSiON devices [110] are reported again in Fig. 3.37: the formation of dipoles at high- κ /SiO₂ interface and the reduction of the TiN work function induced by the presence of Al within TiN. The origin of this effect has been related to the formation of AlN precipitates into TiN for high Al doses [118]. We also report in Fig. 3.37 the WF_{eff} at $EOT = 0$ obtained with the insertion of the same as-deposited doses of aluminum inserted in a *sacrificial* gate-first approach for HfON devices annealed at 900 °C for 10 s [119]. In both approaches, the bottom TiN was deposited with a thickness of 10 Å.

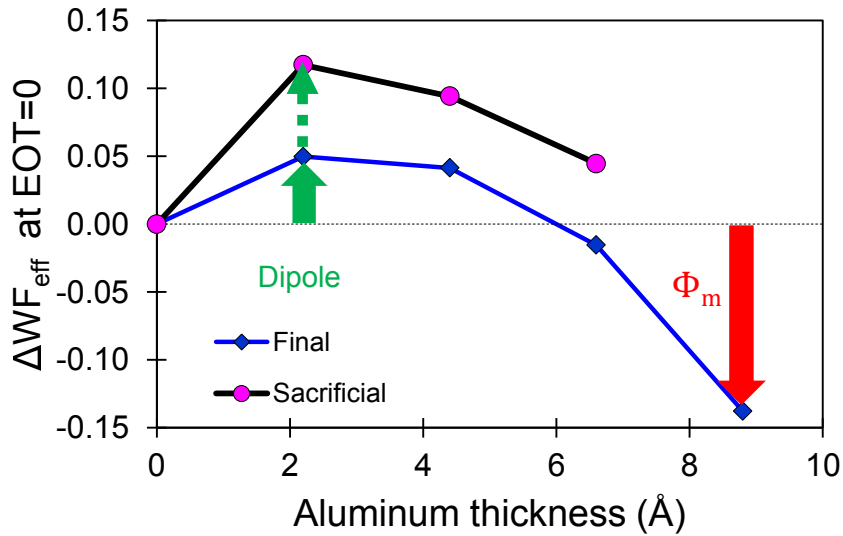


Figure 3.37: Opposite shift of the Al-based metal workfunction towards N+.

Compared to the final metal approach (Fig. 3.37), (i.e., metallic aluminum deposited in the Poly-Si/TiN final electrode and no sacrificial gate), the WF_{eff} shift induced by the insertion of 2 Å of Al is ≈ 100 meV larger with the *sacrificial* gate approach. Now, we know that this discrepancy in WF_{eff} shift can be related, on one hand, to the different effective aluminum dose incorporated into both high- κ oxides, on the other hand, to the different

annealing temperatures.

Nevertheless, since sacrificial gate stack is totally removed in a sacrificial gate-first approach, the opposite shift towards N^+ of the Al-based metal work function (see Φ_m effect in Fig. 3.37) is cancelled. Consequently the dipole benefit induced by Al diffusion is stronger with a *sacrificial* gate-first approach. On the other hand, we also notice the reduction of the WF_{eff} shift for high Al doses but the WF_{eff} is not negative. Despite the formation of AlN precipitates, the WF_{eff} remains positive. Therefore, the formation of AlN precipitates not only has a negative effect on the work function of TiN but also plays a role of diffusion barrier of aluminum into the underlying layer.

On the other hand, the lanthanum incorporated into TiN is useless for WF_{eff} modulation. Although La was inserted in the final Poly-Si/TiN electrode for HfSiON-based devices, no dependence of the TiN/HfSiON barrier height with La dose is demonstrated by performing Internal Photoemission (IPE) measurements on 20nm-thick HfSiO capacitors (Fig. 3.38). In consequence, the work function of TiN (Φ_m) is not modified with the introduction of a lanthanum layer inferior to to 6 Å. Unfortunately, the only effect of La into TiN are the undesirable gate patterning issues observed with the introduction of lanthanum [121].

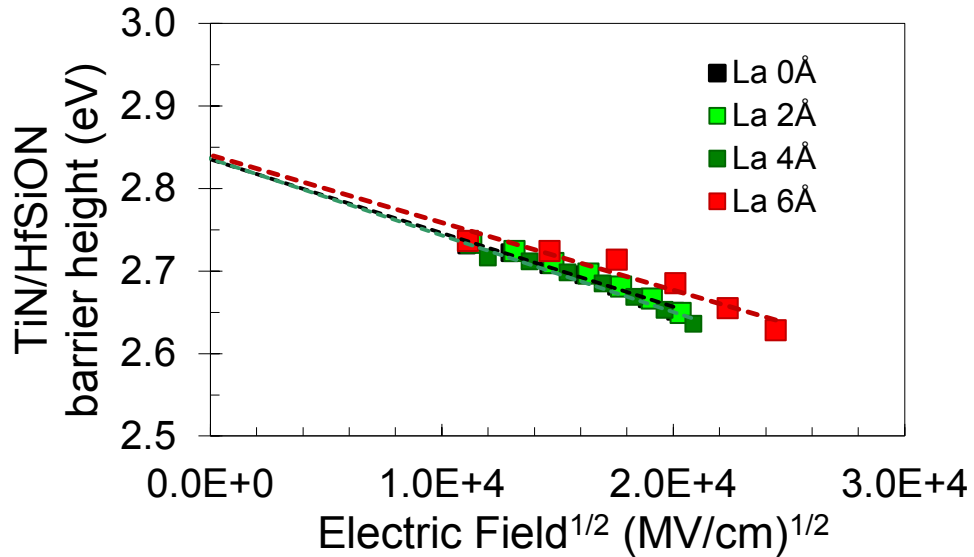


Figure 3.38: TiN/HfSiON barrier height as a function of lanthanum dose inserted between TiN and HfSiON. No modulation of the WF_m with La dose varying from 0 to 6 Å.

In summary, the *sacrificial* gate-first approach allows the prediction and

accurate control of diffusion of lanthanum and aluminum into high- κ /SiO_x stack. In addition, this approach authorizes the simultaneous NFET and PFET gate patterning in a standard 14 nm FDSOI process flow. As a reminder, *sacrificial* gate stack is entirely removed by wet etching and gate patterning is only influenced by the *final* Poly-Si/TiN electrode and high- κ stack. In this way, dry etch induced foot issues, particle-defect increase, yield loss, etc., arising from the nature and height difference of doped *final* metal electrodes, are reduced in *sacrificial* gate-first approach.

3.6 Conclusion

Both lanthanum and aluminum deposited by RF-PVD were successfully incorporated in a *sacrificial* gate-first approach in the simplified process flow validated in chapter 2. Engineering of WF_{eff} is demonstrated by tuning bottom TiN and additives thicknesses without gate leakage degradation. WF_{eff} modulation towards N+ with lanthanum and towards P+ with aluminum are efficiently moderated by the bottom TiN barrier diffusion layer, making this approach suitable for V_{TH} requirements in FDSOI devices. Furthermore, the *sacrificial* gate-first approach allows the prediction and accurate control of diffusion of additives into HfON/SiON stack. Indeed, the methodology based on X-Ray fluorescence enables to prove that the effective dose of the additives incorporated into dielectrics after diffusion annealing may be modeled as a function of the thickness of the bottom TiN in the sacrificial gate and the annealing temperature. Bottom TiN is more efficient to screen the diffusion of La than that of Al. Such capability of TiN to screen the diffusion increases as the annealing temperature decreases.

Moreover, devices with beveled oxides authorize the identification of the origin of the modulation of the effective work function at nominal EOT. This shift is attributed to a dipole layer that evolves with the effective dose of the incorporated dopant and a roll-off component which evolves with both the effective dose and the SiO_x thickness. Accordingly, empirical equations of the diffusion of additives into the gate dielectrics and their impact on the effective work function of metal gates have been proposed to precisely modulate the threshold voltage (V_{TH}) of the 14 nm FDSOI devices. By taking into account the roll-off bias drop for the nominal EOT of the 14 nm FDSOI technology, the lanthanum shifts the WF_{eff} around -80 meV for each 1×10^{14} at/cm² of effective lanthanum dose whereas aluminum does around +25 meV for each 1×10^{14} at/cm² of effective aluminum dose. For the sake of comparison, we have removed such roll-off potential drop to determine the strength of the dipole for effective doses of aluminum and lanthanum incorporated into HfON/SiON stack. This corresponds to a WF_{eff} shift at $\text{EOT} = 0$ of +41 meV for each Al dose of 1×10^{14} at/cm² and a WF_{eff} shift at $\text{EOT} = 0$ of -53 meV for each La dose of 1×10^{14} at/cm² incorporated into HfON layer.

In addition, the influence of the high- κ oxide on both the diffusion of additives and the modulation of the effective work function was investigated. Both aluminum and lanthanum clearly diffuse more easily into HfON than into HfSiON after annealing at 900 °C for 10 seconds. Such difference might explain a more significant WF_{eff} modulation achieved on HfON devices for equivalent as-deposited doses of aluminum and lanthanum. We have compared the WF_{eff} at $\text{EOT}=0$ for equivalent additive doses incorporated into both the HfON/SiON

stack and the HfSiON/SiON stack. This corresponds to a WF_{eff} shift at $EOT = 0$ of +36 meV for each Al dose of 1×10^{14} at/cm² incorporated into the HfSiON/SiON stack and a WF_{eff} shift at $EOT = 0$ of +41 meV for each Al dose of 1×10^{14} at/cm² incorporated into the HfON/SiON stack. On the other hand, we found a WF_{eff} shift at $EOT = 0$ of -54 meV for each La dose of 1×10^{14} at/cm² incorporated into the HfSiON/SiON stack and a WF_{eff} shift at $EOT = 0$ of -53 meV for each La dose of 1×10^{14} at/cm² incorporated into the HfON/SiON stack. Accounting for this weak difference, it is therefore logical to conclude that the ΔWF_{eff} at $EOT=0$ is proportional to the effective additive dose incorporated into the high- κ /SiO_x stack. It is more significant on HfON compared to HfSiON because the effective dose incorporated into HfON is proportionally more significant on HfON compared to HfSiON for equivalent as-deposited doses.

Role of TiN on effective work function and matching of 14 nm FDSOI devices

Contents

4.1	State of the art of the modification of the effective work function induced by the TiN	130
4.2	Characterization of the impact of TiN on the effective work function of 14 nm FDSOI devices	134
4.3	Impact of deposition process on TiN microstructure and local V_{TH} variability	149
4.4	Conclusion	166

This chapter focuses on the impact of nitrogen, adjusted by the thickness and composition of TiN layers, on the effective work function of FDSOI devices. The section 4.1 summarizes the state of the art of the modification of the effective work function induced by the TiN in the literature.

The impact of the nitrogen content, first into the sacrificial TiN then into the final TiN layer, on the modulation of the effective work function are studied in section 4.2. The influence of the high- κ oxide on the effective work function shift induced by nitrogen is highlighted in section 4.2.2, likewise it was evidenced for the effective work function shift induced by aluminum and lanthanum in sections 3.2.3 and 3.4.3. Afterwards, the incorporation of nitrogen into gate stack is characterized in section 4.2.3 by the physicochemical measurements and the methodology based on X-Ray Fluorescence introduced in section 2.2.2.2. Finally, an innovative process for metal deposition by RF-PVD at room temperature, allowing the modification of the microstructure of TiN, is proposed in section 4.3 in order to further improve the local V_{TH} variability in FDSOI devices.

4.1 State of the art of the modification of the effective work function induced by the TiN

Metal gate electrodes are required to be included in gate stack in order to eliminate the polysilicon (Poly-Si) depletion and boron penetration effects, thus reducing EOT penalty and increasing the ON current (I_{DS}) of the transistors. Furthermore, with the introduction of metals, mobility is improved due to the higher concentration of carriers in metal compared to Poly-Si (section 1.2.2). Among all the potential metal gate candidates, TiN has been widely used in semiconductor industry mainly for the following reasons:

- Stability in contact with high- κ dielectrics.
- Low resistivity.
- Mid-gap work function [122].
- Excellent diffusion barrier for boron [123] and gate additives such as lanthanum and aluminum [106].
- Suppression of the Poly-Si depletion.
- Easy patterning of the Poly-Si/TiN stack.

TiN thickness [124], oxygen incorporation [125], post-deposition annealing [126] and N concentration [127] are some of the process knobs which have already been investigated to modulate the effective work function of TiN in order to meet the requirements of the devices.

4.1.1 Modulation of the effective work function by tuning TiN thickness

In general, the effective work function increases as the TiN thickness increases, as experimentally demonstrated in [64] [124] [127] [128] on SiO_2 , HfO_2 and HfSiON . However, the main mechanism behind the WF_{eff} modulation remains ambiguous. Furthermore, the reported values are very different. For instance, Kai et al. reported a $\Delta\text{WF}_{\text{eff}}$ modulation by up to ≈ 150 meV as TiN increases from 3 nm to 10 nm on HfO_2 (2.14 meV/\AA) [124]. In contrast, a thicker TiN of 33 nm did not show a further increase of WF_{eff} shift. Kai attributed such behaviour to the crystallinity variation of TiN from 3 nm to 10 nm and subsequent stabilization of TiN crystallization. Widiez et al. [128] reported instead a WF_{eff} increase of 230 meV as TiN increases from 5 nm to 40 nm (0.66 meV/\AA), and Baudot found a variation of 320 meV as TiN increases from 45 \AA to 120 \AA on HfSiON (4.27 meV/\AA) [64], much more than the 1.07 meV/\AA reported by Kadoshima et al. [129]. By the way, in the latter work, it is shown that the same effect of TiN thickness on V_{fb} is not observed when TiN is

capped with electrodes poor in silicon, such as tungsten (W), as shown in Fig. 4.1. It should be noted that high WF_{eff} is obtained with W regardless of the TiN thickness. It was also demonstrated that a 2 nm-thick TiN in a Poly-Si/TiN/HfSiON structure was less oxidized than in a W/TiN/HfSiON structure. Kadoshima also observed that the reason for which a 2 nm-thick TiN was less oxidized in a Poly-Si/TiN/HfSiON structure was due to the Poly-Si oxidation. An increase of the thickness of TiN reduced the oxidation of Poly-Si. The mechanism proposed by Kadoshima et al. to explain the increase of WF_{eff} with the TiN thickness is related to the displacement of oxygen atoms at the TiN/HfSiON interface towards the Poly-Si/TiN interface. In summary, what Kadoshima observed is that an oxidized TiN leads to a high WF_{eff} shift and such oxidation can be modulated with the thickness of TiN.

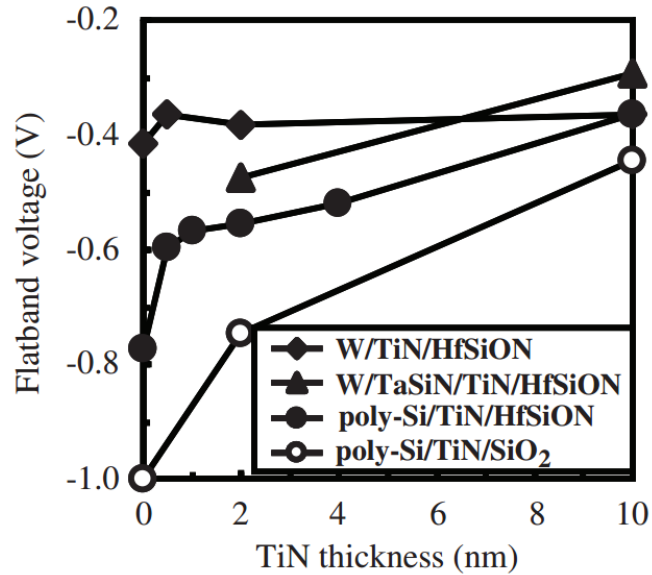


Figure 4.1: Changes in V_{fb} by changing TiN thickness in n+ poly-Si/TiN, W/TiN, W/TaSiN/TiN gate electrodes. The solid and open symbols indicate the results for HfSiON and SiO₂ dielectrics, after [129].

In addition, interlayer oxide regrowth is also observed when the TiN thickness increases on both TiN/SiO₂ and TiN/HfO₂/SiO₂ devices. Due to the EOT increase, mobility is degraded for thicker TiN films, as experimentally demonstrated in [128]. The benefit in mobility obtained with the introduction of TiN metal (section 1.2.2) is therefore limited to thin TiN films. However, this problem can be suppressed with the introduction of SiGe in the channel and the loss of mobility due to thicker TiN is overcome by the increase of mobility with SiGe.

4.1.2 Modulation of the effective work function by tuning N composition in TiN gate

For PVD processes, the concentration of N into TiN can be tuned by varying Ar and N₂ gas flow rates during TiN deposition, as described in section 1.2.2.2. The work function shift induced by the TiN composition has also been extensively studied [130] [127] [64]. S.-H. Lee reported lower V_{fb} and thinner EOT with Ti-rich TiN while a higher V_{fb} along with thicker EOT was observed with N-rich TiN on both HfO₂ and HfSiON-based MOS capacitors [127]. However, amount of V_{fb} shift and EOT reduction was found to be higher for HfO₂ than for HfSiON-based devices. The mechanism proposed by Lee to tune both V_{fb} and EOT was the generation of oxygen vacancies into high- κ gate dielectrics. For Ti-rich TiN, more Ti element in metal gate electrode scavenges more oxygen from high- κ , leading to more oxygen vacancy generation. According to Lee, the effect of oxygen vacancies into high- κ dielectrics is, in consequence, the shift of WF_{eff} towards N+. The authors attributed the differences between HfO₂ and HfSiON based MOS capacitors to the less oxygen vacancy generation due to strong Hf-Si bonds in HfSiON compared to HfO₂ [127].

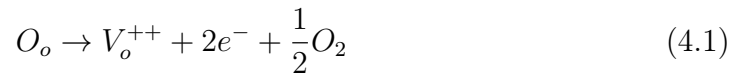
4.1.3 Effect of the oxygen incorporation in TiN gate on the effective work function

The increase of the oxygen density into TiN can modulate its effective work function towards P+ (positive V_{fb} shift), as experimentally demonstrated in [124], [125]. We can therefore identify that the electrical effect of an oxygen vacancy (WF_{eff} shift towards N+) is opposite to the electrical effect of the intentional incorporation of oxygen (WF_{eff} shift towards P+). The description of the experiments of Cartier in next section clarified the effect of oxygen and oxygen vacancies on WF_{eff} . It is also known that oxygen in metal gate can diffuse in a gate stack and react with the silicon substrate during high temperature processing, leading to the equivalent oxide thickness (EOT) increase of the gate dielectrics [131]. In the work of Kai et al. [124], TiN/HfO₂ stacks underwent annealing under N₂ and a trace of O₂ at different temperatures (from 300 °C to 550 °C) and O plasma treatments after TiN deposition to prove the impact of oxygen on C-V measurements. An increase of the annealing temperature from 300 °C to 500 °C results in a positive WF_{eff} shift of about 130 meV but also EOT increase and an interface quality degradation for temperatures higher than 500 °C. The O plasma treatment has also shown to be an effective approach for oxygen incorporation with the same electrical effect. By performing a low temperature O₂ anneal at 300 °C for 240 s after TiN deposition on HfO₂, Li et al. [125] also achieved to increase the TiN effective work function. The impact of oxygen incorporation into TiN was found to be more efficient to shift V_{fb} towards P+ with thinner TiN metal layers [125]. The mechanism proposed to explain this behaviour is that during spike anneal,

the TiN/high- κ interface becomes oxygen deficient because oxygen at this interface can diffuse and react with the poly-Si layer on top or with the silicon substrate, resulting in a low WF_{eff} . The oxygen deficient interface must be compensated with a low temperature O_2 anneal after 2 nm TiN deposition. The ability of oxygen to increase the WF_{eff} after being incorporated into metal/dielectric stack is not exclusive of TiN/high- κ layers but it also has been reported on other metal gate candidates like TaC [132], Mo [133], Ru [133] [134] [17], Re [17] and Pt [17].

4.1.4 Effect of the reduction of oxygen vacancy defects on effective work function

Unlike SiO_2 -based gate dielectrics, Hf-based dielectrics suffer from severe oxygen vacancy defects (V_o^{++}) [16]. The electrical effect of an oxygen vacancy was clarified by Cartier et al. [17], which is opposite to the electrical effect of the intentional incorporation of oxygen by annealing under oxidizing ambient or O plasma treatments. In other words, the oxygen vacancies into Hf-based dielectrics results in a decrease of the effective work function. Indeed, a WF_{eff} decrease was experimentally demonstrated by annealing in a reducing atmosphere such as forming gas anneal [17]. Higher oxygen vacancies in a high- κ gate dielectric are likely to enhance the V_{fb} shift toward a negative direction in the C-V curve because oxygen defects are positively charged, as shown in Eq. 4.1, where O_o is the oxygen atom at the oxygen site and V_o^{++} is the oxygen vacancy in high- κ dielectrics.



The high temperature thermal process has been identified to be responsible for the creation of such oxygen vacancy defects, as a result of oxygen diffusion kinetics from high- κ to metal gate during thermal budget process. For this reason, new options in thermal processing have been explored such as the dynamic surface anneal (DSA), which has been considered as a diffusionless annealing [126, 135]. In particular, in this approach the thermal budget is reduced by decreasing the annealing time from 1 s to 1 ms. Choi et al. [126] compared the C-V characteristics between DSA and conventional Rapid Thermal Annealing (RTA) of PMOS capacitors and demonstrated that a 350 mV shift of V_{fb} towards the positive direction in the C-V curve is induced by the shorter time associated to the DSA. This behaviour was explained by the authors as follows: the process time of RTA (5 s) is long enough to create oxygen vacancies, compared to the one of DSA (few ms), which is short enough to suppress oxygen vacancy generation in HfO_2 .

4.2 Characterization of the impact of TiN on the effective work function of 14 nm FDSOI devices

4.2.1 Impact of nitrogen content into TiN on the effective work function

The purpose of this study is to determine the efficiency of nitrogen compared to other additives such as aluminum or lanthanum for work function modulation. Note that the nitrogen content can be modified not only in the sacrificial TiN but also in the final TiN of the 14 nm FDSOI devices. First, we investigate the impact of the sacrificial TiN process (thickness and composition) on the WF_{eff} and EOT and then the impact of the final TiN thickness on the WF_{eff} and EOT.

Impact of nitrogen content into *sacrificial* TiN on the effective work function

Low- V_{TH} NFET devices were fabricated with the simplified process flow of Fig. 2.12 in order to assess the impact of the nitrogen content into TiN on the effective work function and thus on the V_{TH} of the advanced 14 nm FDSOI devices. Beveled thermally grown SiO_2 was used as interlayer dielectric in order to discriminate between an effect of fixed charges and a dipole effect. The impact of thickness and composition of TiN layers deposited by Radio-Frequency PVD on WF_{eff} of Hf-based devices in a *sacrificial* metal gate-first approach was evaluated for the first time. On one hand, 2 different *sacrificial* TiN thicknesses (45 Å and 80 Å) were deposited on HfON, according to the process explained in section 1.2.2.2. As a reminder, this process corresponds to a deposition where the nitrogen is incorporated into the TiN with a gradient profile. On the other hand, the impact of TiN composition on WF_{eff} was also investigated. *Sacrificial* TiN layers with different compositions ($N_2/(N_2 + \text{Ar}) = 0.5, 0.7$ and 1) were deposited with a uniform profile of nitrogen into TiN. The stoichiometry of the TiN was modified by varying Ar and N_2 gas flow rates during deposition, as shown in Table 4.1. A sample without *sacrificial* stack (i.e. only the final Poly-Si/TiN electrode deposited after HfON growth), namely TiN0 Å, was also fabricated for reference. The Poly-Si layer was deposited with a thickness of 240 Å. The *final* TiN was deposited with a nitrogen gradient profile and a thickness of 35 Å for all the manufactured NFET devices.

In order to evaluate by how much the nitrogen percentage inside TiN can be increased by tuning the process parameters, several TiN screening deposition

conditions were performed on blanket wafers made of 200 Å-thick SiO₂. The at%-N and the TiN thickness were measured by X-Ray Fluorescence (XRF) after TiN deposition on RIGAKU WaferX300, with X-ray Tube Rh (4.0 kW), 40 kV, 90 mA and a spot size of 40 mm. The results are summarized in Figs. 4.2 and 4.3. On all wafers, sheet resistance has been measured on KLA TENCOR OmniMap RS-100 Sheet Resistance Mapping System, using a 4-points probe 4.0 mil Tip - Type B, as described in section 2.2.1.

TiN process	TiN type	Deposition (sccm)	Rs (Ω/sq)	XRF (at%-N)
N-rich+	uniform	N ₂ =45 Ar=0	633.8	54.6
N-rich	uniform	N ₂ =45 Ar=20	384.1	52.6
Stoichiometric	uniform	N ₂ =20 Ar=20	252.9	51.3
Gradient	gradient	N ₂ =45 Ar=20	379.9	51.6

Table 4.1: Ar and N₂ flow, sheet resistance (Rs) and nitrogen content for various TiN processes.

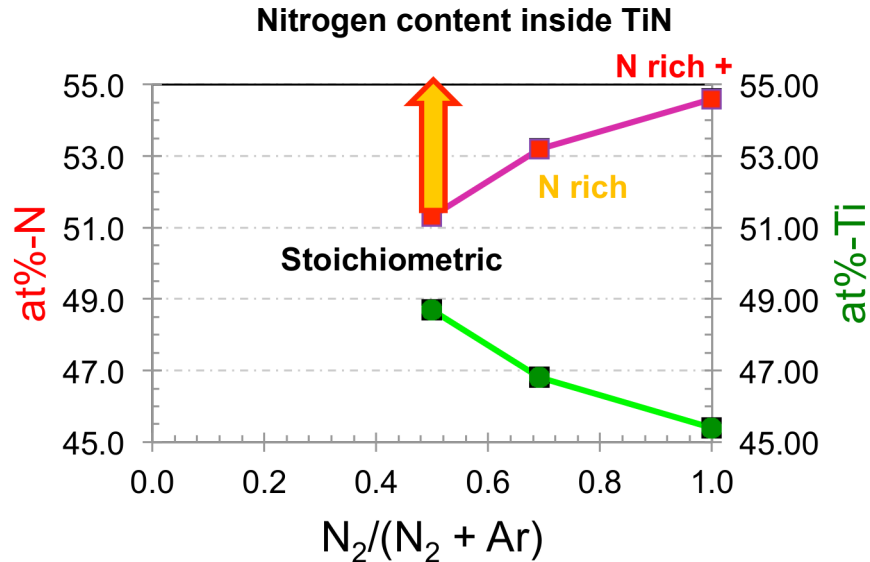


Figure 4.2: Percentage of nitrogen content inside TiN measured by XRF, expressed as a function of normalized N₂ / (Ar + N₂).

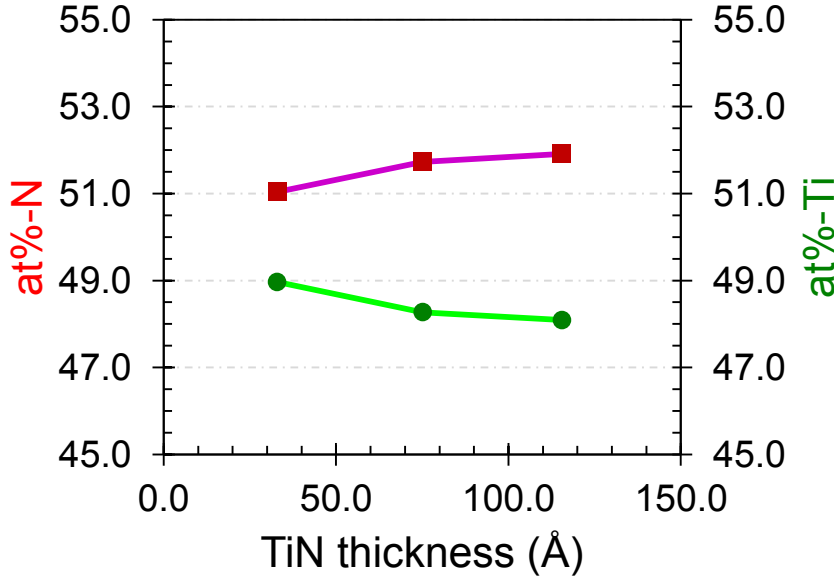


Figure 4.3: Percentage of nitrogen content inside TiN measured by XRF, as a function its thickness. In this process, nitrogen is incorporated into the TiN with a gradient profile.

EOT and V_{fb} were extracted from CV measurements for different dies along the beveled wafer and WF_{eff} was then calculated from V_{fb} and the silicon work function relative to vacuum ($q\Phi_{sc}$), as presented in section 2.1.3. Fig. 4.4 shows the WF_{eff} vs EOT plot for beveled devices with different TiN compositions deposited by RF-PVD at room temperature in a *sacrificial* gate approach (annealing at 900 °C). Fig. 4.4 clearly proves that the *sacrificial* TiN composition does not have a significant effect on WF_{eff} shift. Increasing uniformly the N content inside a 45 Å-thick *sacrificial* TiN from 51% to 55% is therefore not efficient to modulate the effective work function of 14 nm FDSOI devices. On the other hand, Fig. 4.5 shows the WF_{eff} vs EOT plot for beveled devices with 2 different TiN thicknesses with a nitrogen gradient profile deposited by RF-PVD at room temperature in a *sacrificial* gate approach (annealing at 900 °C) compared to the process of reference. From the slope of the WF_{eff} vs EOT plot in Fig. 4.5, fixed charges density is estimated to be $1.14 \times 10^{12} /cm^2$, $4.76 \times 10^{11} /cm^2$ and $6.98 \times 10^{11} /cm^2$ for 0 Å, 45 Å and 80 Å of TiN thickness, respectively. A slight decrease of interfacial fixed charges density is observed with the sacrificial gate process compared to the process of reference without sacrificial gate (TiN0 Å). The ΔWF_{eff} extrapolated at EOT = 0 also reveals the insignificant influence of a dipole effect with a variation of TiN thickness in a *sacrificial* gate-first approach, compared to the process of reference without sacrificial gate. In addition, the TiN/HfO₂ barrier height is not significantly modified, as demonstrated in Fig. 4.6 by performing Internal

Photoemission (IPE) measurements on 20 nm-thick HfO_2 capacitors.

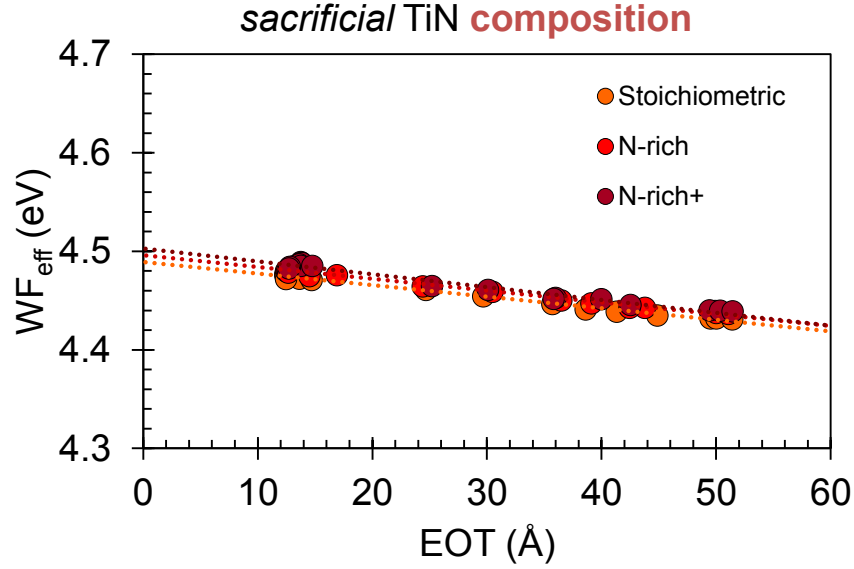


Figure 4.4: WF_{eff} vs EOT curve of HfON-based NFET devices with beveled oxides and different compositions of TiN layers inserted in a *sacrificial* approach.

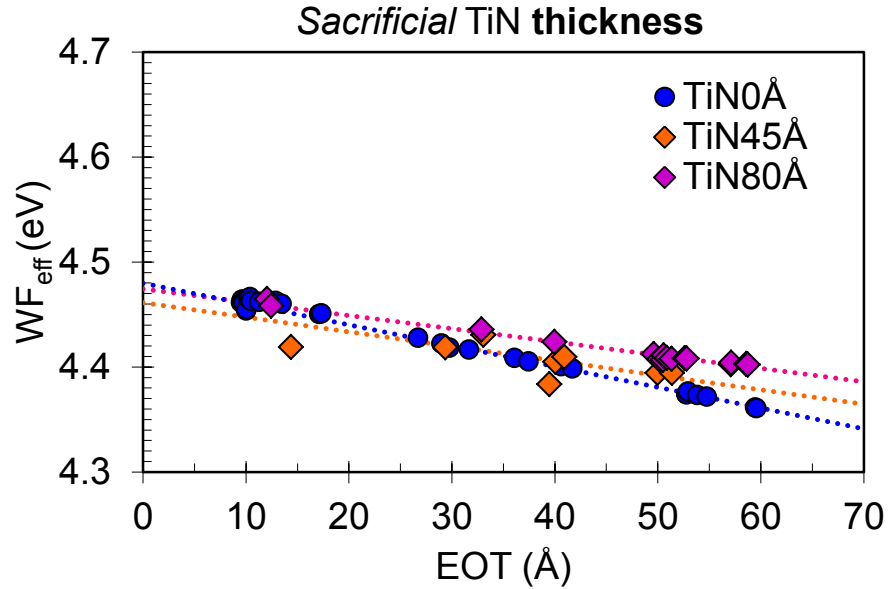


Figure 4.5: WF_{eff} vs EOT curve of HfON-based NFET devices with beveled oxides and different thicknesses of TiN layers inserted in a *sacrificial* approach.

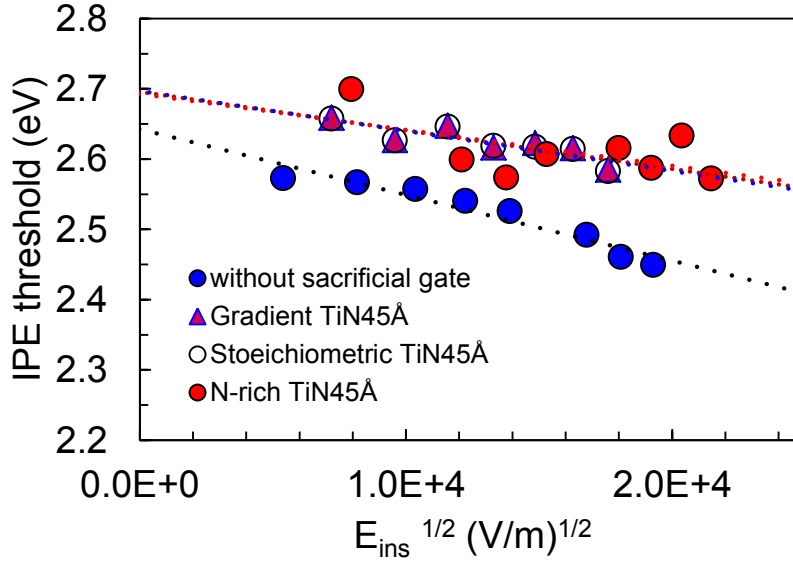


Figure 4.6: TiN/HfO₂ barrier height measurements with different nitrogen compositions.

Impact of final TiN thickness on the effective work function of 14 nm FDSOI devices

The impact of thickness of TiN layers deposited by Radio-Frequency PVD on WF_{eff} of HfON devices were also evaluated in the standard *final* metal gate-first approach (see step 4 in Fig. 1.18). Different TiN thicknesses (35 Å, 75 Å and 115 Å) were deposited in the final Poly-Si/TiN electrode of simplified Low- V_{TH} NFET devices in order to determine the capability of the TiN thickness to modulate WF_{eff} in a *final* gate approach and to compare it to a *sacrificial* gate approach. Evidently, the *sacrificial* gate stack, the annealing diffusion and its wet removal were intentionally skipped during the device fabrication for this particular study so that the analysis is simplified.

Fig. 4.7 shows the modulation of WF_{eff} with different TiN thicknesses deposited in the final Poly-Si/TiN electrode of HfON-based devices. From the slope of the WF_{eff} vs EOT plot, fixed charges density is estimated to be $1.14 \times 10^{12} \text{ /cm}^2$, $1.46 \times 10^{12} \text{ /cm}^2$, and $1.58 \times 10^{12} \text{ /cm}^2$ for 35 Å, 75 Å, and 115 Å of TiN thickness, respectively. Increasing TiN thickness from 35 Å to 115 Å leads to an increase of WF_{eff} at zero EOT of 140 meV from 4.48 eV to 4.62 eV (i.e. 1.75 meV/Å). It should be noted that this shift at EOT = 0 induced by TiN thickness with a nitrogen gradient profile is higher in a standard *final* approach than in a *sacrificial* one. This is likely because actually the main mechanism is not only related to the diffusion of the nitrogen towards the dielectrics stack but also to a change of the metal work function (Φ_m) of TiN at different thicknesses.

Nevertheless, thanks to the beveled interlayer oxide, a strong roll-off of the dipole induced by TiN is put in evidence at $EOT < 20 \text{ \AA}$. In addition, the roll-off is found to be stronger for thicker Poly-Si/TiN electrodes. As a result, the growth of the TiN thickness from 35 \AA to 115 \AA causes a WF_{eff} shift of only 50 meV (0.625 meV/\AA) towards P+ on nominal 14 nm FDSOI devices. Consequently, the efficiency of final TiN thickness to shift the WF_{eff} towards P+ at nominal EOT is limited and becomes wasteful due to this strong roll-off effect.

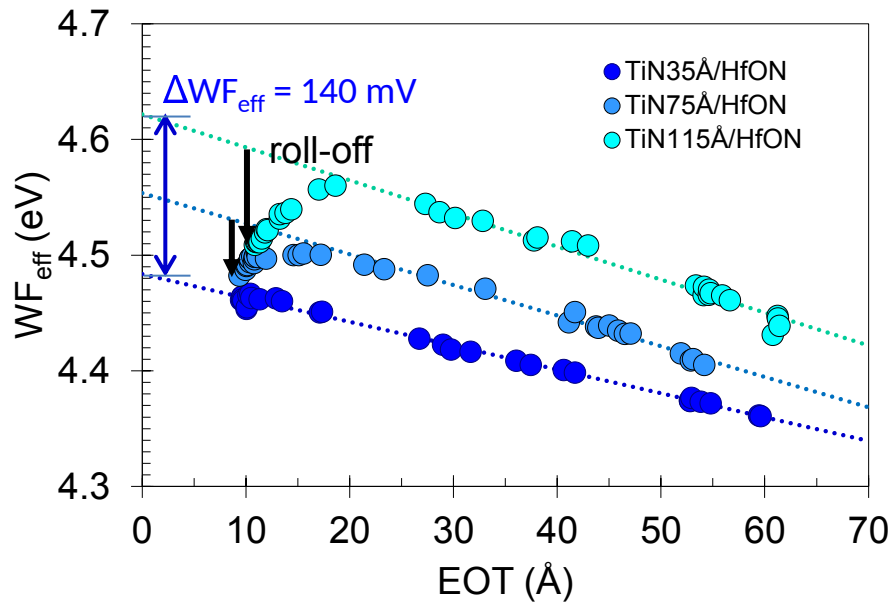


Figure 4.7: Modulation of WF_{eff} with TiN thickness in a final gate approach on HfON-based devices without *sacrificial* gate.

Comparison of capabilities of aluminum and final TiN to shift the WF_{eff} towards P+

We can now compare the electrical effect of aluminum incorporated into the HfON/SiON stack in a *sacrificial* gate-first approach after annealing at $900 \text{ }^\circ\text{C}$ for 10 s (Fig. 3.22) with the electrical effect of the increase of final TiN thickness (Fig. 4.7). Thanks to the beveled oxide methodology, we have observed a WF_{eff} shift at $EOT = 0$ of $+1.75 \text{ meV per \AA}$ of TiN and $+41 \text{ meV}$ for each Al dose of $1 \times 10^{14} \text{ at/cm}^2$. Nevertheless the WF_{eff} shift at nominal EOT is influenced by the roll-off effect. Taking into account such roll-off effect, the WF_{eff} shift at nominal EOT achieved with the final TiN thickness becomes only $+0.625 \text{ meV per \AA}$ of TiN. In the case of aluminum, it becomes only $+25 \text{ meV}$ for each Al dose of $1 \times 10^{14} \text{ at/cm}^2$.

We report in Fig. 4.8 the loss due to roll-off effect as a function of the WF_{eff}

shift at $EOT = 0$ induced by the final TiN thickness and the aluminum incorporated in a *sacrificial* gate-first approach. The incapability of the final TiN thickness to increase WF_{eff} towards P+ at nominal EOT compared to the aluminum inserted in a *sacrificial* gate-first approach after annealing at 900 °C for 10 s is demonstrated in Fig. 4.8. Indeed, the *sacrificial* metal gate process authorizes the insertion of aluminum without strong losses of dipole effect induced by the roll-off effects (only 0.08 meV/meV) compared to 0.55 meV/meV for TiN.

The origin of such roll-off differences must be investigated in a future work. Some authors consider that WF_{eff} roll-off phenomenon may be explained by a progressive oxygen vacancy generation in high- κ as interlayer oxides scale down [133], and therefore dependent on the activation anneal temperature, as observed in [9] [136]. The same phenomenon has been interpreted as an abrupt bias drop related to a dipole located at HfO_2/SiO_2 interface [9]. It is also curious to note that actually the roll-off increases with the dipole strength in the case of TiN, aluminum and lanthanum, although not with the same proportion.

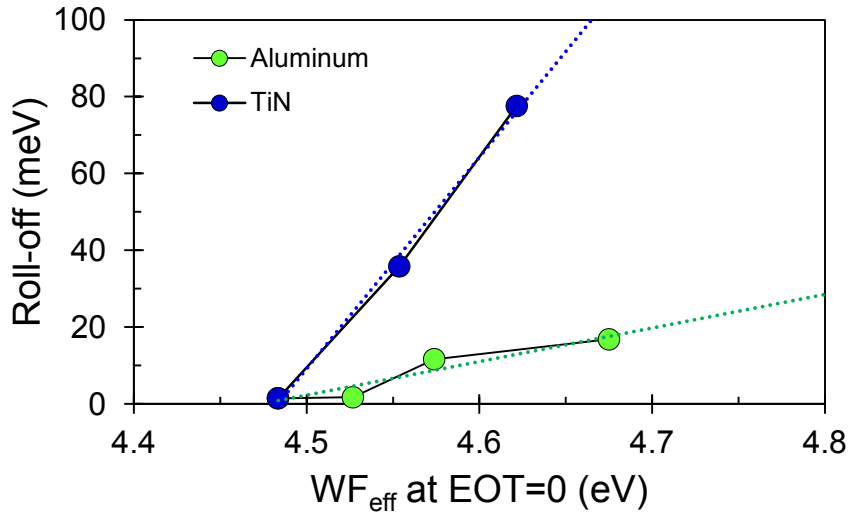


Figure 4.8: Comparison between aluminum and TiN roll-off effects and WF_{eff} gain. The *sacrificial* metal gate authorizes the insertion of aluminum without strong roll-off losses.

4.2.2 Influence of high- κ dielectrics on the work function shift induced by TiN thickness

The results obtained in the previous section are not consistent to data already reported in the literature for equivalent *final* TiN process variations on HfSiON: 340 mV shift (4.27 meV/Å) [64]). Aiming to understand this

discrepancies in WF_{eff} , we study the influence of the high- κ dielectrics on the WF_{eff} shift induced by the thickness of *final* TiN layers deposited by Radio-Frequency PVD at room temperature. Besides the standard HfON of the 14 nm FDSOI devices, two other types of high- κ oxide were investigated : i) HfO₂ deposited by Atomic Layer Deposition (ALD) without decoupled plasma nitridation (DPN), but with the Post-Nitridation Anneal (PNA) and ii) HfSiON deposited by Metal Organic Chemical Vapor Deposition (MOCVD) with a concentration of 30% Si (as used in 28 nm FDSOI technologies), followed by the same decoupled plasma nitridation (DPN) and the Post-Nitridation Anneal (PNA) of the 14 nm FDSOI technology, as shown in Fig. 4.9. Three different TiN thicknesses (35 Å, 75 Å and 115 Å) were deposited in the final Poly-Si/TiN electrode of the simplified Low- V_{TH} NFET devices. Beveled thermally grown SiO₂ was used as interlayer dielectric in order to discriminate between an effect of fixed charges and a dipole effect. Again, the *sacrificial* gate stack, the annealing diffusion and its wet removal were intentionally skipped during the device fabrication for this particular study so that the analysis is simplified.

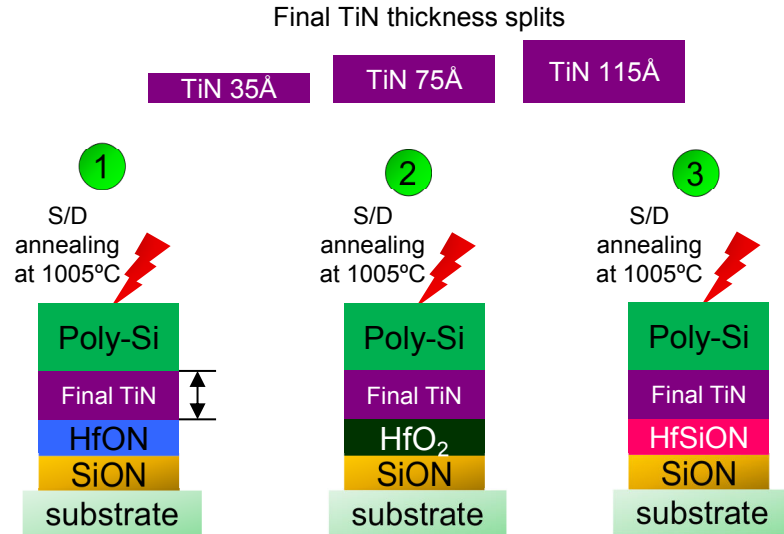


Figure 4.9: Final TiN thickness variations deposited on three high- κ oxides: HfON, and HfO₂, and HfSiON.

Fig. 4.10 shows the modulation of WF_{eff} with different TiN thicknesses deposited in the final Poly-Si/TiN electrode of HfO₂-based devices. Although the decoupled plasma nitridation is useful for decreasing the minimum value of EOT obtained with the beveled oxide technique due to reduction of the growth of the interfacial oxide after annealing, it seems however that it has no influence on the effective work function modulation induced by TiN thickness. Indeed,

the WF_{eff} is also modulated on HfO_2 -based devices by increasing TiN thickness but it is tuned exactly as when deposited on HfON gate dielectrics, from 4.48 eV to 4.61 eV with an increase of TiN thickness from 35 Å to 115 Å (see Fig. 4.7). Nevertheless, we would expect a more important shift on devices with HfO_2 oxides. Indeed a higher WF_{eff} shift should be observed for a diffusion of nitrogen from TiN to HfO_2 compared to a diffusion of nitrogen from TiN to HfON, based on the hypothesis that the diffusion of nitrogen is proportional to the gradient of concentration. Instead, not only the WF_{eff} shifts are the same, but also the WF_{eff} at $EOT = 0$ values are at the same level. Such observation suggests that nitrogen incorporated by the DPN is not efficient to modulate the WF_{eff} .

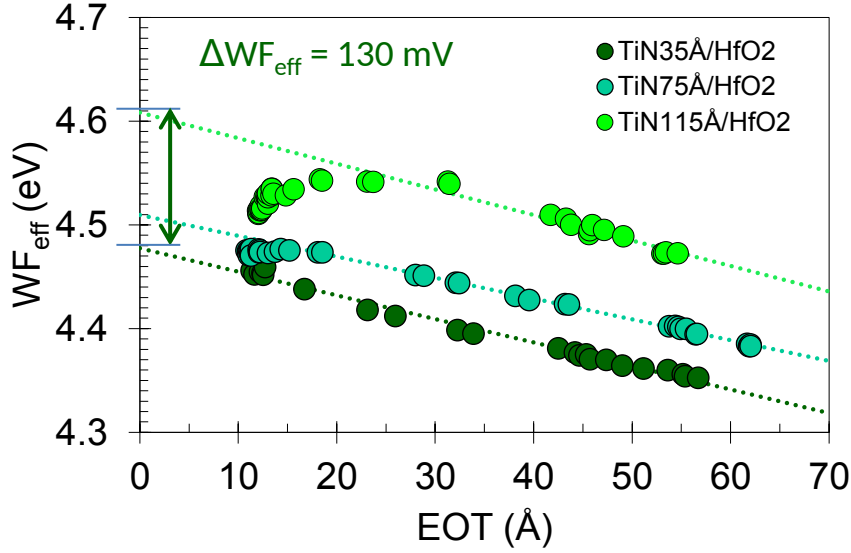


Figure 4.10: Modulation of WF_{eff} with TiN thickness in a final gate approach on HfO_2 -based devices (i.e. without plasma nitridation).

In contrast, the strength of the WF_{eff} shift induced by increasing TiN thickness seems to be reduced with $HfSiON$ dielectrics compared to HfON dielectrics, as shown in Fig. 4.11. From the slope of the WF_{eff} vs EOT plot, fixed charges density is estimated to be $1.08 \times 10^{12} / \text{cm}^2$, $1.40 \times 10^{12} / \text{cm}^2$, and $9.85 \times 10^{11} / \text{cm}^2$ for 35 Å, 75 Å, and 115 Å of TiN thickness deposited over $HfSiON$, respectively. The increase of TiN thickness from 35 Å to 115 Å leads to an increase of WF_{eff} at zero EOT of around 100 meV from 4.47 eV to 4.57 eV (i.e. 1.25 meV/Å). The WF_{eff} extrapolated at zero EOT (i.e. without the influence of interfacial Si/SiO₂ fixed charges) is summarized in Fig. 4.12 for the three high- κ dielectrics integrated in the simplified low- V_{TH} NFET devices. We can thus observe that the TiN thickness (d_{TiN} in Fig. 4.12) is more efficient to increase the WF_{eff} on HfON devices.

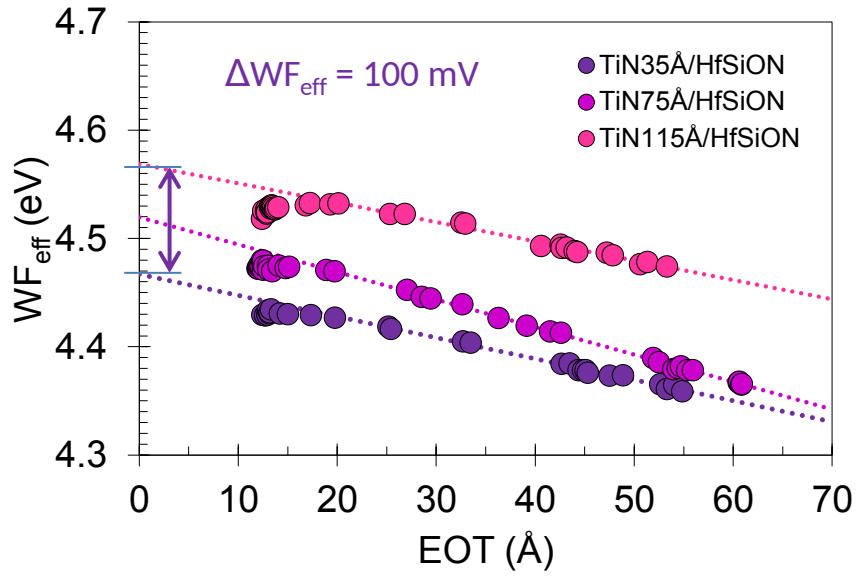


Figure 4.11: WF_{eff} vs EOT plot of HfSiON-based NFET devices with beveled oxides and different thicknesses of TiN layers in the final Poly-Si/TiN electrode.

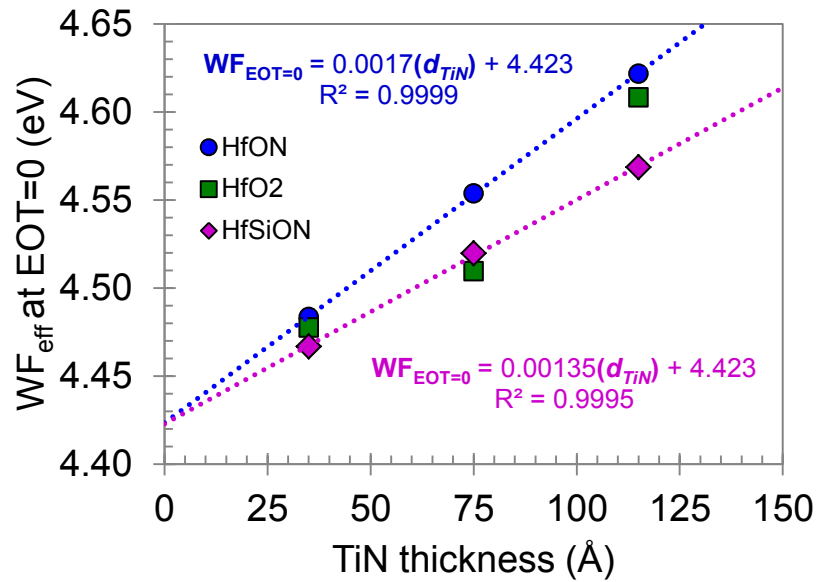


Figure 4.12: WF_{eff} extrapolated at EOT = 0 as a function of TiN thickness for the three different high- κ dielectrics studied.

On the other hand, despite the higher WF_{eff} obtained with HfON devices, the undesirable roll-off effect at nominal EOT is stronger on HfON than on

HfSiON. As a result, for nominal devices the WF_{eff} shift induced by TiN thickness is insignificant on both high- κ dielectrics. Roll-off can be defined as the difference between the value determined from the WF_{eff} vs EOT straight line and the experimental WF_{eff} , both at minimal EOT. The Fig. 4.13 compares the WF_{eff} at EOT=0 with (dotted lines) and without (solid lines) the influence of roll-off effect at nominal EOT for HfON and HfSiON dielectrics as a function of TiN thickness. Unfortunately, stronger roll-off effect neutralizes the higher shift achieved with HfON in comparison to HfSiON. Taking into account the roll-off effect, a maximum value of only 4.54 eV is obtained with 115 Å of TiN for both HfON and HfSiON gate dielectrics in nominal 14 nm FDSOI devices.

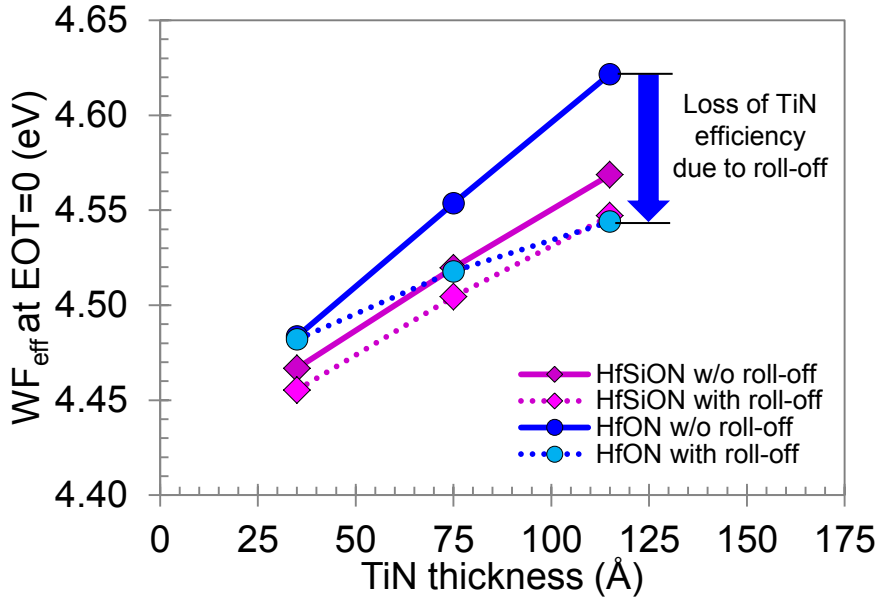


Figure 4.13: WF_{eff} extrapolated at EOT = 0 with (dotted lines) and without (solid lines) the influence of roll-off effect for HfON and HfSiON dielectrics as a function of TiN thickness.

In summary, the efficiency of TiN metal gate to shift the WF_{eff} towards P+, by adjusting the thickness, has been compared in three of the most employed high- κ dielectrics in technologies fabricated at STMicroelectronics. Although this efficiency is higher in HfON-based devices, thicker TiN layers are not suggested to be integrated in 14 nm FDSOI technology, considering the strong roll-off at nominal EOT. Instead, the use of aluminum in a *sacrificial* gate-first approach is highly recommended to conceive P-gates in order to meet the requirements of the multi- V_{TH} offer in 14 nm FDSOI devices, not only because of the precise control of the incorporated effective dose but also due to the weaker roll-off effect at nominal EOT.

4.2.3 Titanium and nitrogen diffusion into HfSiON/SiON and HfON/SiON stacks

As described in section 2.2.2.2, X-Ray Fluorescence (XRF) is a useful technique to estimate the percentage of species into the gate stack after diffusion annealing. The purpose of this section is to determine the relationship between the effective dose of N and Ti incorporated into the Hf-based dielectrics after diffusion annealing and the electrical effect on effective work function observed in section 4.2.2. In Figure 4.14, we compare N concentration at different process levels, either on HfSiON or HfON dielectrics, before and after sacrificial gate deposition and after diffusion annealing at 900 °C for 10 s and TiN wet removal.

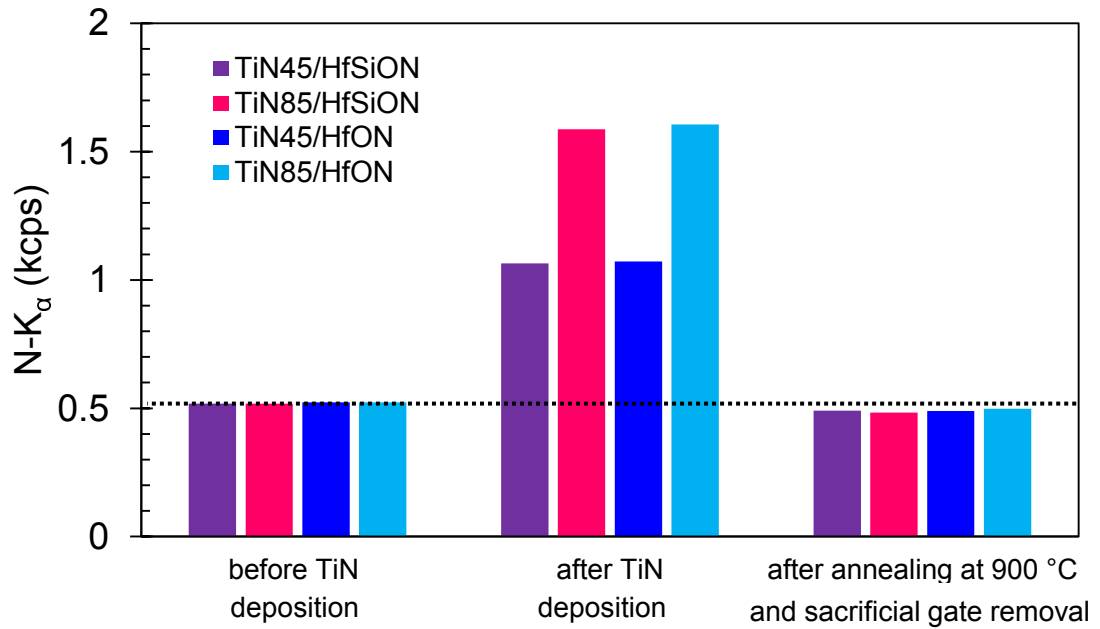


Figure 4.14: Quantification of N amount by X-Ray Fluorescence wafers at different process levels: before and after sacrificial gate, and after diffusion annealing at 900 °C for 10 s and TiN wet removal.

As expected, we notice in Fig. 4.14 a global N increase after *sacrificial* TiN deposition with an increase related to TiN gate thickness from 45 Å to 85 Å. It should also be noted that the N K_{α} intensity is not zero before TiN deposition. As a reminder, XRF technique allows to explore the whole high- κ stack. In consequence, signal below the dotted line in Fig. 4.14 is related to the nitrogen amount in HfON/SiON stack. Despite this logical trend, no obvious N increase is observed at further steps due to *sacrificial* TiN gate deposition. No effect of N enrichment into both high- κ has been induced by TiN. Instead, an insignificant decrease of N intensity is observed after the sacrificial gate process,

related perhaps to the surface condition of high- κ after removal by wet etching. Although detection of nitrogen by X-Ray Fluorescence is supposed to be difficult due to a decrease of fluorescence yield for elements with low atomic numbers, it should be noted that the results are reliable. Indeed, fluorescence intensity just after TiN deposition is proportional to the deposited thickness after suppression of N signal related to nitrogen into high- κ stack (dotted line in Fig. 4.14). In contrast, the fluorescence intensity of nitrogen after *sacrificial* TiN removal is at the same level regardless of the thickness of TiN or the type of high- κ . Therefore, the observed increase in WF_{eff} observed with thicker TiN layers in Fig. 4.5 seems to be not related to a N diffusion into high- κ stack.

Likewise, diffusion of Ti into both gate dielectrics has been investigated with the same methodology based on X-Ray Fluorescence, and no Ti diffusion was evidenced, as confirmed in Fig. 4.15. The insignificant Ti intensity observed after annealing at 900 °C and *sacrificial* gate removal and also before TiN deposition (not shown here) can only be related to a background noise of the measurement. These results allow to conclude that observed increased of WF_{eff} with TiN thickness must be explained by other reasons than Ti or N diffusion into gate dielectrics.

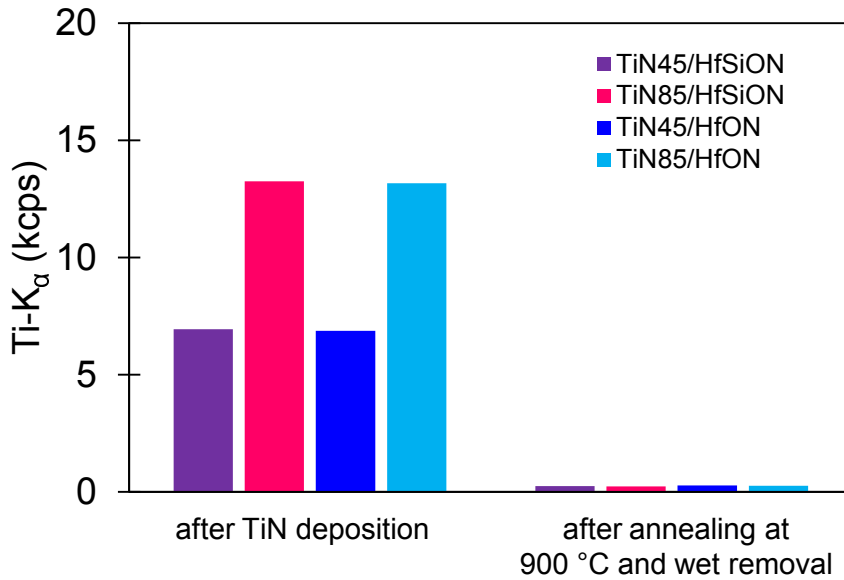


Figure 4.15: Quantification of Ti dose by X-Ray Fluorescence wafers after sacrificial gate deposition compared to after annealing at 900 °C for 10 s and wet sacrificial gate removal for thin (45 Å) and thick (85 Å) TiN layers deposited on HfON and HfSiON.

The impact of higher annealing temperatures (> 900 °C) on the diffusion of

nitrogen could not be studied because of the limitations of standard wet solutions to remove the silicon deposited as capping layer on the top of TiN. It cannot, therefore, be excluded the diffusion of nitrogen after an annealing at 1005 °C during the spike anneal to explain the WF_{eff} increase with thicker *final* TiN layers observed in Figs. 4.7, 4.10, and 4.11.

4.2.4 Influence of Arsenic doping into Poly-Si layer on the effective work function

It is important to remark that polysilicon was kept in process flow in order to minimize process complexity and internal stress [122], which might be associated with thick metal gates. Nevertheless, Arsenic ion implantation (As I/I) has been reported to drop V_{fb} up to 600 meV for nMOSFETs [137] [138]. Both authors claim that the V_{fb} shift is due to an Arsenic dipole at metal/high- κ interface. In this section, it is demonstrated that there is no influence of incorporation of arsenic into Poly-Si on the values of WF_{eff} reported in this thesis work.

In 14 nm FDSOI devices, the polysilicon gate is implanted with As ions at 6 keV with a dose of $2 \times 10^{15} \text{ cm}^{-2}$ through a 50 Å thick liner oxide, followed by rapid thermal annealing at 1000 °C for 13 seconds. The different As ion implantation conditions shown in Table 4.2 were performed on HfON-based NFET devices in order to investigate their influence on the WF_{eff} .

Sample	Energy implantation	Dose implantation
No As	NA	NA
As 3keV 1E15	3 keV	$1 \times 10^{15} \text{ cm}^{-2}$
As 6keV 2E15	6 keV	$2 \times 10^{15} \text{ cm}^{-2}$
As 8keV 4E15	8 keV	$4 \times 10^{15} \text{ cm}^{-2}$

Table 4.2: Different As ion implantation conditions investigated in NFET devices.

Fig. 4.16 shows clearly that no modulation of WF_{eff} is achieved by the incorporation of arsenic into the Poly-Si layer. In addition, Atome Probe Tomography (APT) was used to obtain an extremely precise chemical analysis and a 3D reconstruction of the gate stack. Arsenic distribution is not homogeneous within Poly-Si. Moreover, it does not diffuse into TiN layer. Instead, segregation of arsenic was found at the TiN/Poly-Si interface. In consequence, it can be concluded that there is no dipole layer formation at TiN/Poly-Si interface. Furthermore, the validity of results and the WF_{eff} values shown in this thesis work are not influenced by arsenic ion implantation.

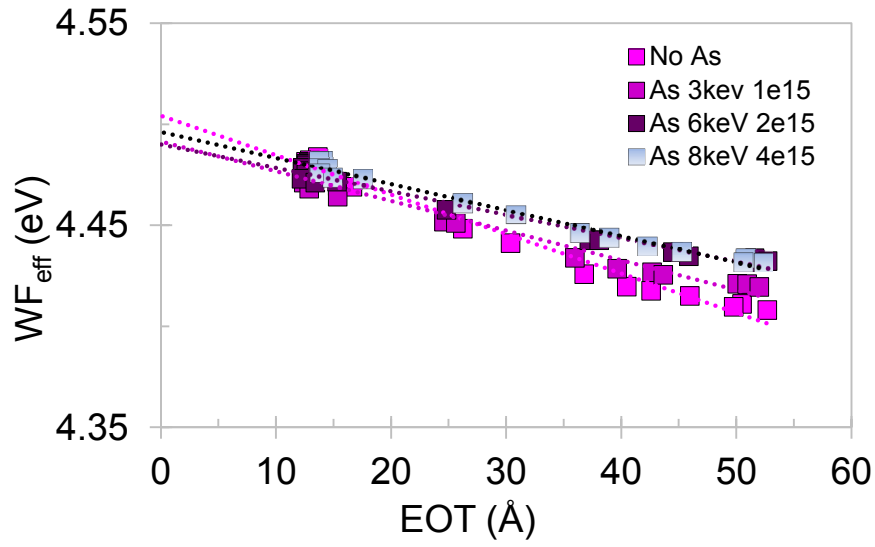


Figure 4.16: WF_{eff} vs EOT of HfON-based NFET devices with beveled oxide and different As Ion Implantation conditions.

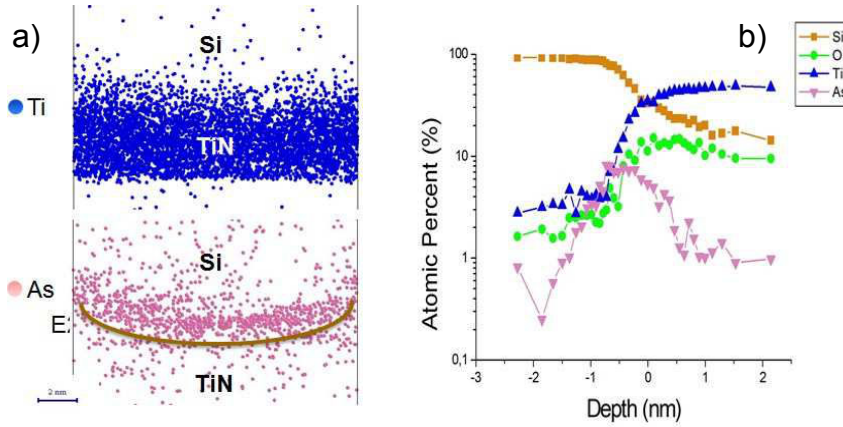


Figure 4.17: a) Titanium (top picture) and Arsenic (bottom picture) distribution at the TiN/poly-Si interface. *Courtesy of R. Estivill.*

4.3 Impact of deposition process on TiN microstructure and local V_{TH} variability

As stated in section 1.3.3, the local V_{TH} variability (also called matching) due to metal gate granularity (MGG) is related to a different value of work function associated to each crystallite orientation. In the case of TiN, (200) orientation is associated to a Φ_m value of 4.6 eV and (111) orientation is associated to a Φ_m value of 4.4 eV. For that reason, three technical solutions are proposed and listed below in order to suppress or reduce the variability of the effective work function.

1. **Same and unique metal orientation.** Indeed, if the TiN metal has only one preferred grain orientation, then the matching related to the TiN metal gate granularity is solved.
2. **Amorphous microstructure.** This approach has demonstrated to result in a lower σV_{TH} compared to polycrystalline gates. Amorphous gates such as TiCN [139], TaCN [140], TiSiN [141], TaSiN [84], and Ta₄₀W₄₀Si₁₀C₁₀ [142] have been studied and have shown promising results.
3. **Smaller horizontal grain size.** The number of TiN crystallites in a metal gate is inversely proportional to the standard deviation of the metal work function (σWF_{eff}). For a given area, smaller crystallites results in an increase of the total number of grains and σWF_{eff} is reduced by a statistical effect (see Fig. 4.18).

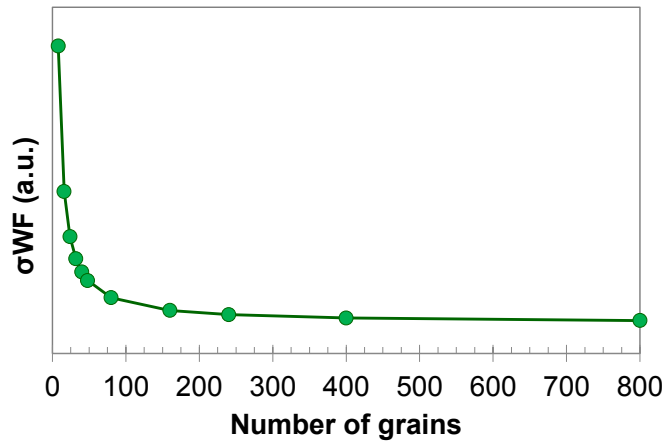


Figure 4.18: σWF_{eff} as a function of the number of grains.

The morphology of the deposited thick TiN film has been described by a zone model introduced for the first time by Mochan and Demchishin and later

extended to sputter-deposited films by Thornton [143]. Basically, there are four basic processes from the atom point of view : 1) the geometric interaction between the roughness of the growing surface (the substrate, i.e. HfON) and the angular directions of the arriving atoms, 2) the surface diffusion of the adsorbed atoms (adatoms), 3) the bulk diffusion and 4) the desorption. The first one, also known as shadowing, can be characterized by the roughness of the substrate. The other three can be quantified in terms of the activation energies for surface and bulk diffusion, and the sublimation energy, respectively [144]. Since for pure metals, these energies are related and proportional to their melting point, it can be expected that one or various of the four basic processes dominate over different ranges of T/T_m , where T is the substrate temperature and T_m is the coating material melting point.

Based on his observations, Thornton established 4 structural zones as illustrated in Fig. 4.19, which are briefly described here [143]:

- The **Zone 1** structure results when adatom diffusion is insufficient to overcome the effects of shadowing. Its persistence to elevated temperatures depends on the degree of substrate roughness. However, it is promoted, even on smooth surfaces, by increasing inert gas pressure at low T/T_m . The zone 1 is recognized by columnar crystallites defined by open boundaries that are not influenced by annealing at high T/T_m (> 0.5).
- The **Zone T** is defined as the form taken by the Zone 1 structure in the limit of zero T/T_m on an infinitely smooth substrate.
- The **Zone 2** is characterized by evolutionary growth due to adatom diffusion. It occurs for $T/T_m > 0.3$. It is generally characterized by columnar shaped crystallites which may approach a near equiaxed shape, and which are separated by dense intercrystalline boundaries. Occasionally the growth may take the form of platelets and needles.
- The **Zone 3** is characterized by bulk diffusion processes such as recrystallization and grain growth. It occurs for $T/T_m > 0.5$. Grain shape may be equiaxed or columnar, depending on structure and stress distribution formed during initial deposition.

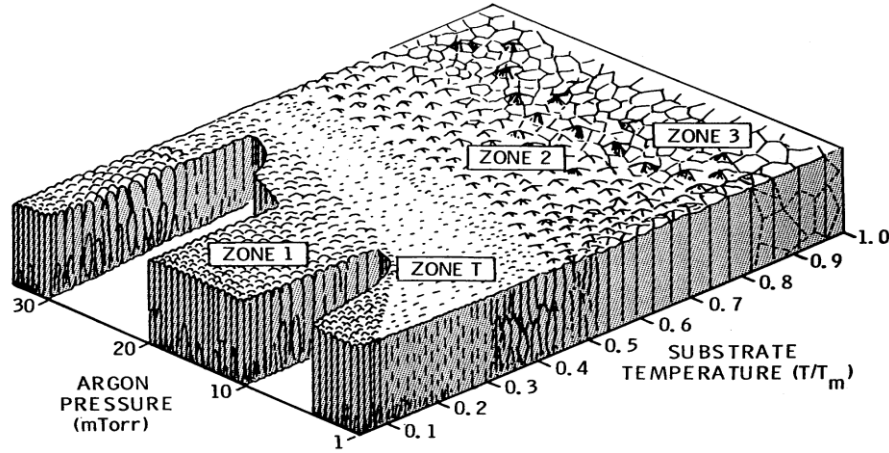


Figure 4.19: Schematic representation of dependence of thick metal films on substrate temperature and argon pressure [143].

It is worth noting that although these zone models are in agreement with experimental observations for pure metal thick films, inconsistencies might be found in complex alloys and compounds films such as the TiN. The dynamics of atomic-scale processes on compound surfaces is very complex to be modeled and requires computational methods such as density functional theory. Another source of discrepancy might be associated to thin polycrystalline films, which differ from bulk stress-free materials. Indeed, the T/T_m at which recrystallization occurs is dependent on the stored strain energy. Lastly, the specificities of RF-PVD deposition have been little investigated and Thornton zone models are not necessarily appropriated to describe this films.

In this work, the preferred orientations and the size of the crystallites of the thin TiN films, deposited by RF-PVD at room temperature, are experimentally determined by performing X-Ray Diffraction measurements in the center of the wafer, in θ - 2θ and in-plane configurations, respectively, as described in section 2.2.4. Furthermore, the impact of the main RF Physical Vapor Deposition parameters such as pressure, RF power and DC power on the preferred orientations and the size of the crystallites is investigated on TiN films deposited in blanket wafers. Although final TiN deposition is performed at room temperature, it undergoes later in the process flow the effects of the spike annealing at 1005 °C for source/drain dopant activation. In order to take into account such temperature for gate-first applications, the XRD measurements were carried out after Poly-Si/TiN deposition and spike annealing at 1005 °C, as shown in Fig. 4.20. The thicknesses of the TiN metal films were 10 nm for the thin film characterization.

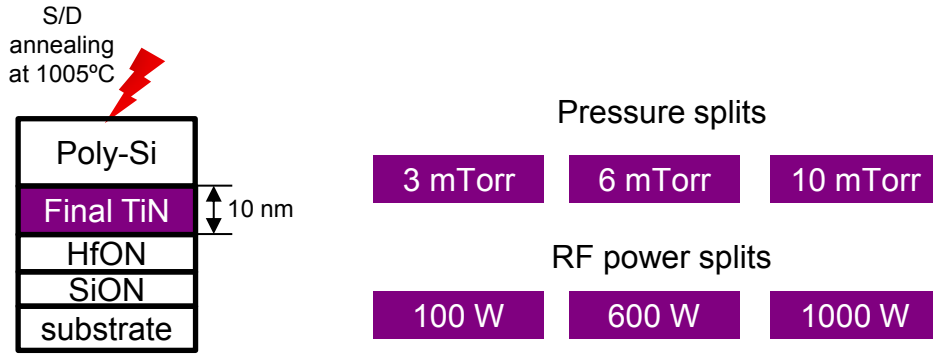


Figure 4.20: Process variations for TiN film characterization.

4.3.1 Impact of pressure conditions on microstructure of TiN deposited by RF-PVD

The impact of different pressures (3 mTorr and 10 mTorr) into RF-PVD chamber during TiN deposition on the preferred orientation is shown in Fig. 4.21. The most intense orientations of TiN crystallites are (200) and (111). From the values of the peak areas, automatically extracted by fitting the experimental X-Ray diffraction spectra with a combination of Lorentzian and Gaussian functions, it can be calculated the percentage of a particular grain orientation in the total population of grains (i.e. P_i values in Eq. 1.25). The relative intensities of the different orientations related to TiN crystallites reported in Table 2.2 must be considered to calculate these percentages. As a reminder, intensities of the possible different peaks are not supposed to be equal, even in the particular case of same proportion of each grain orientation, due to the structure factor related to each combination of Miller indices, as explained in section 2.2.4. The percentage of crystallites for each TiN orientation as a function of the pressure conditions are shown in Table 4.3. Note that only (111) et (200) orientations were considered and their sum was forced to 100 %. In other words, the (220) and (311) orientations were neglected. In fact, a first θ - 2θ long scan (not shown here) from 2θ between 25° and 75° revealed that TiN (220) and TiN (311) peaks are buried in background signal and no relevant information about these orientations can be extracted from the diffraction pattern. The range of the 2θ was reduced thereafter between 25° and 50° so that the time per step can be increased in order to improve signal. As shown in Fig. 4.21, amorphous films were not obtained. Indeed, amorphous materials like glass do not have a periodic array, so they do not produce a diffraction pattern. Monocrystalline TiN films were also not achieved by increasing pressure into the RF-PVD chamber. Instead, with an increase of pressure from 3 mTorr to 10 mTorr, the population of grains of each orientation becomes balanced, as shown in Table 4.3.

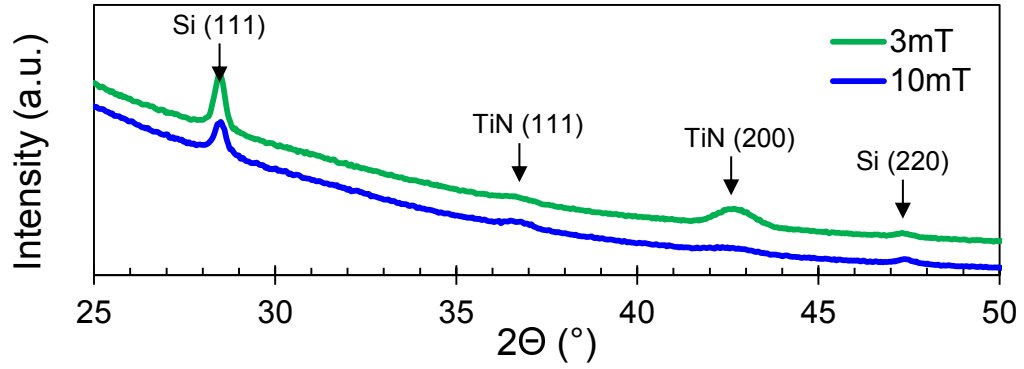


Figure 4.21: θ - 2θ X-Ray diffraction spectra for different pressures into RF-PVD chamber during TiN deposition.

Pressure	Parameter	TiN (111)	TiN (200)
3 mTorr	Normalized Area	21.1	100
	Relative intensity	72	100
	Percentage	14.7 %	85.3 %
10 mTorr	Normalized Area	75.1	100
	Relative intensity	72	100
	Percentage	52.2 %	47.8 %

Table 4.3: Determination of percentage of each grain orientation in the total population of grains from θ - 2θ X-Ray diffraction spectra for two different pressure conditions into RF-PVD chamber during TiN deposition.

In addition, we report in Table 4.4 the average grain size $\langle D \rangle$ of each crystallite orientation determined by in-plane XRD as a function of different pressure conditions (3 mTorr, 6 mTorr, and 10 mTorr). The diffraction angle position (2θ), the peak height and the integral breadth of each TiN peak orientation are also reported in Table 4.4. We highlight in Fig. 4.22 the average grain size as a function of the pressure conditions into RF-PVD chamber during TiN deposition.

TiN type	TiN peak	Position [2 θ]	Height [cts]	Integral breadth [2 θ]	<D> [nm]
3 mTorr	111	36.40	719	1.118	12.7
	200	42.38	1122	1.394	8.9
	220	61.54	413	2.021	5.8
6 mTorr	111	36.39	1084	1.510	8.8
	200	42.35	1583	1.675	7.7
	220	61.52	653	2.532	4.9
10 mTorr	111	36.41	72	1.660	7.7
	200	42.33	116	1.906	6.5
	220				

Table 4.4: Determination of lateral grain size from in-plane XRD measurements by applying Scherrer equation for different pressure conditions into RF-PVD chamber during TiN deposition.

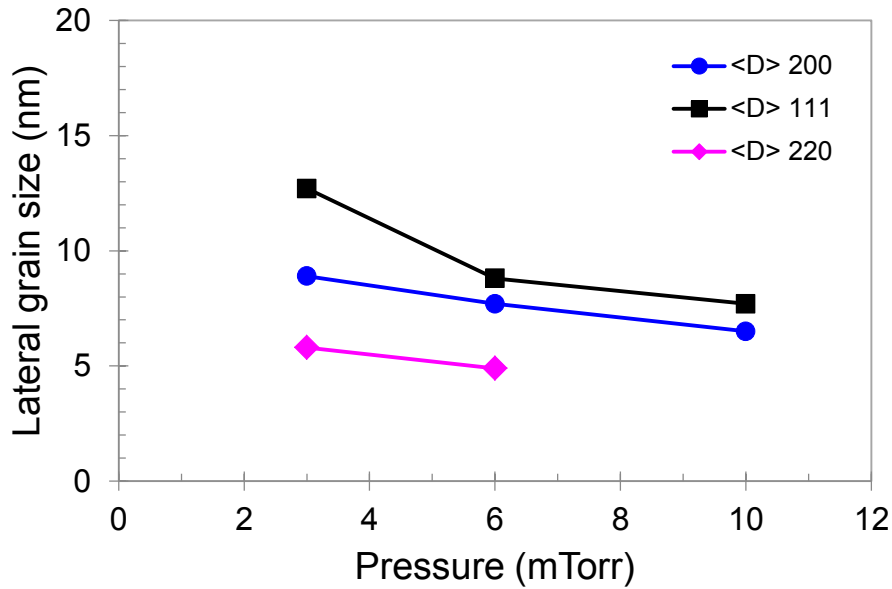


Figure 4.22: Average grain size evolution with pressure into RF-PVD chamber during TiN deposition.

Fig. 4.22 clearly shows that the average size $\langle D \rangle$ of the horizontal grains of the TiN decreases with the pressure increase into RF-PVD chamber during TiN deposition. With an increase of pressure from 3 mTorr to 10 mTorr, the average size $\langle D \rangle$ of the (200)-oriented grains roughly decreases from 90 Å to 65 Å, and the one of (111)-oriented grains decreases from 130 Å to 75 Å. It should be reminded that different grains are characterized in the Θ -2 Θ and in in-plane configurations and only lattice planes perpendicular to the diffraction vector contribute to the intensity of these planes in the diffraction pattern. Thus, it is

not surprising that the intensity of TiN (220) planes becomes significant in this configuration. The extraction of the average size of the (220)-oriented grains also shows a decrease of about 10 \AA , from 60 \AA to 50 \AA with pressure growing from 3 mTorr to 6 mTorr. Complementary Transmission Electron Microscopy (TEM) pictures corroborate the size decrease of TiN horizontal grains after gate patterning with the increase of pressure from 3 mTorr to 10 mTorr, as shown in Figs. 4.23 and 4.24, respectively. Crystallites with a columnar shape (approximately 9 grains for a gate length of around 42 nm), as described by the zone 2 in Thornton model, can be observed with a pressure of 3 mTorr (Fig. 4.23). Surprisingly, not only decrease of the average grain size (approximately 12 grains for a gate length of around 42 nm), but also recrystallization in a new grain structure with a shape similar to zone 3 in Thornton model is observed in Fig. 4.24. According to Thornton model, this recrystallization should occur for temperatures higher than the melting point of TiN, at approximately $2350 \text{ }^\circ\text{C}$.

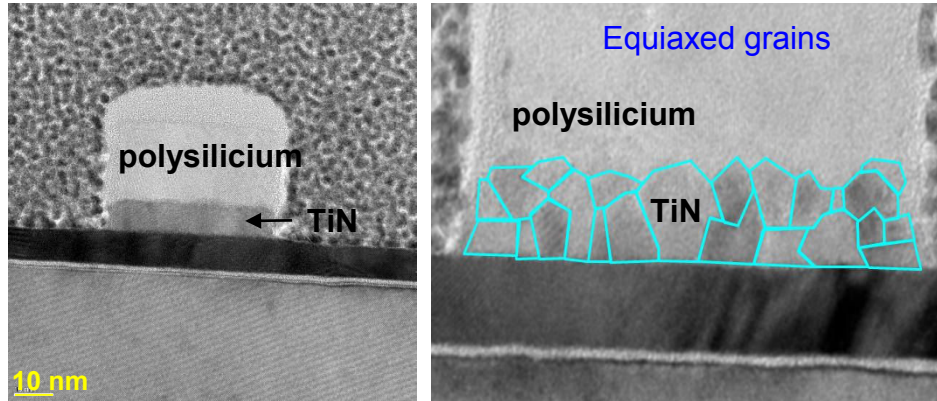


Figure 4.23: TEM image for TiN deposited at 3 mTorr and RF power of 600 Watts. *Courtesy of O. Ros.*

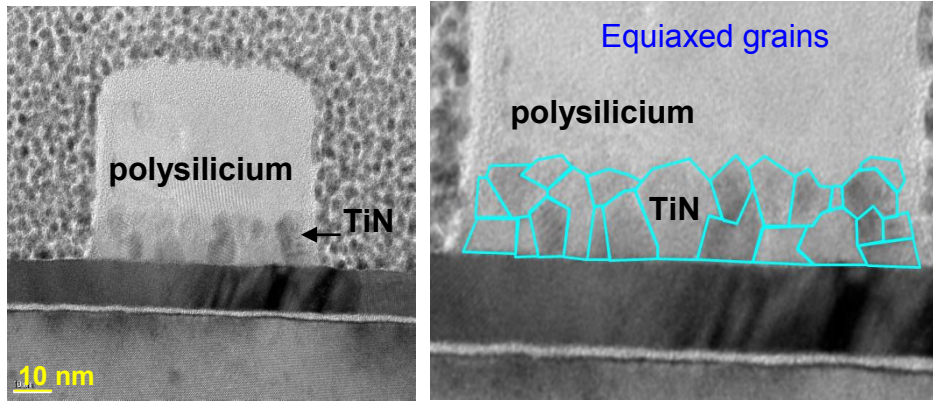


Figure 4.24: TEM image for TiN deposited at 10 mTorr and RF power of 600 Watts. *Courtesy of O. Ros.*

4.3.2 Impact of RF power conditions on microstructure of TiN deposited by RF-PVD

The impact of different Radio Frequency powers (600 Watts, 300 Watts, and 100 Watts) applied to the RF-PVD chamber during TiN deposition on the preferred orientation is shown in Fig. 4.25. The most intense orientations of TiN crystallites are (200) and (111). The percentage of crystallites for each TiN orientation as a function of the RF power conditions are shown in Table 4.5. Unlike the pressure, the increase of RF power from 100 Watts to 600 Watts does not equilibrate the population of grains of each orientation. RF power modulation in this range is also not convenient to grow an amorphous film or a monocrystalline structure.

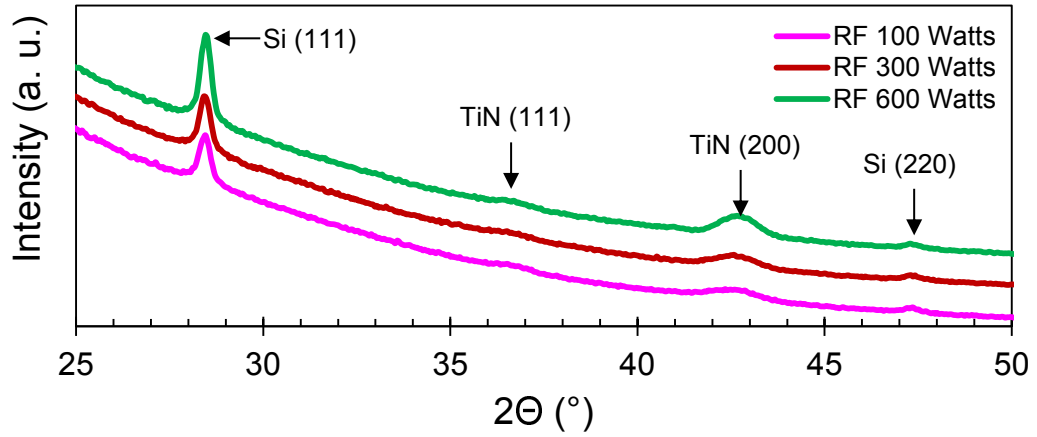


Figure 4.25: θ - 2θ X-Ray diffraction spectra for different RF powers applied to the RF-PVD chamber during TiN deposition.

RF Power	Parameter	TiN (111)	TiN (200)
100 W	Normalized Area	38	100
	Relative intensity	72	100
	Percentage	26.4 %	73.6 %
300 W	Normalized Area	43.6	100
	Relative intensity	72	100
	Percentage	30.3 %	69.7 %
600 W	Normalized Area	22.8	100
	Relative intensity	72	100
	Percentage	15.8 %	84.2 %

Table 4.5: Determination of percentage of each grain orientation in the total population of grains from θ - 2θ X-Ray diffraction spectra for different RF power conditions applied to the RF-PVD chamber during TiN deposition.

In addition, the average grain size $\langle D \rangle$ of each crystallite orientation as a function of different RF power conditions (100 Watts, 600 Watts, and 1000 Watts), determined by performing in-plane X-Ray Diffraction measurements, is reported in Table 4.6, as well as the diffraction angle position (2θ), the peak height and the integral breadth of each TiN peak orientation. The resulting lateral grain size evolution with RF power during RF-PVD TiN deposition is shown in Fig. 4.26. It can be noticed an increase of the average size of the grains when the RF-PVD power is increased, in contrast to the effect of pressure.

TiN type	TiN peak	Position [2 θ]	Height [cts]	Integral breadth [2 θ]	<D> [nm]
100 Watts	111	36.41	831	1.728	7.4
	200	42.37	1607	1.750	7.2
	220	-	-	-	-
600 Watts	111	36.40	719	1.118	12.7
	200	42.38	1122	1.394	8.9
	220	61.54	413	2.021	5.8
1000 Watts	111	36.43	1094	0.98	15.9
	200	42.37	862	1.34	9.5
	220	61.57	399	1.66	7.5

Table 4.6: Determination of lateral grain size from in-plane XRD measurements by applying Scherrer equation for different RF power conditions applied to the RF-PVD chamber during TiN deposition.

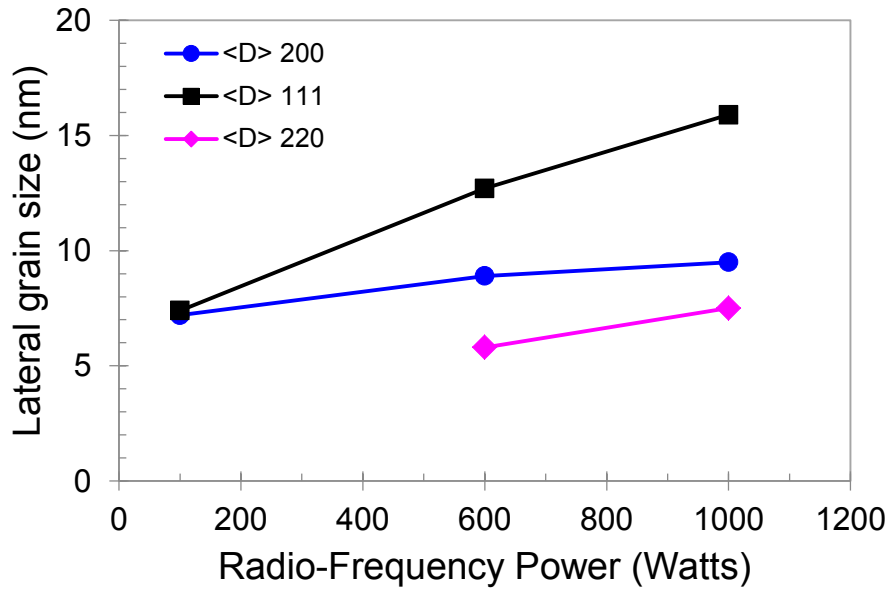


Figure 4.26: Lateral grain size evolution with Radio-Frequency Power into RF-PVD chamber during TiN deposition.

The average size decrease of TiN horizontal grains after gate patterning with the decrease of RF power from 600 Watts to 100 Watts is confirmed with the help of Transmission Electron Microscopy (TEM) pictures, as shown in Figs. 4.23 and 4.27. Once again, recrystallization in a new grain structure with a shape similar to zone 3 in Thornton model is observed in Fig. 4.27.

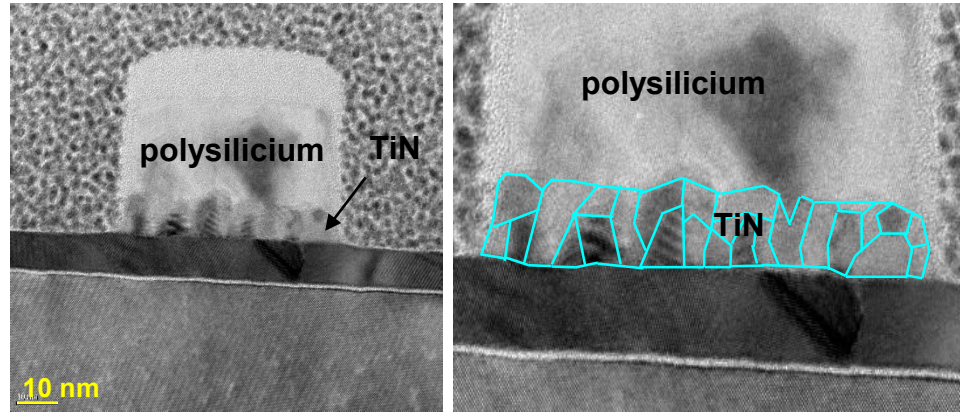


Figure 4.27: TEM image for TiN deposited at 3 mTorr and RF power of 100 Watts. *Courtesy of O. Ros.*

In summary, reduction of the average size of TiN grains has been successfully achieved by both the increase of pressure and the decrease of RF power during TiN deposition by RF-PVD at room temperature. As a result, the TiN deposition rate is reduced, as shown in Figs. 4.28 and 4.29. Consequently, the reduction of average grain size might be related to a reduction of ion bombardment during TiN deposition. On the other hand, the modulation of RF power and pressure in this range are not convenient to grow a monocrystalline structure or an amorphous film, which seems difficult to obtain without the incorporation of dopants such as carbon or silicon.

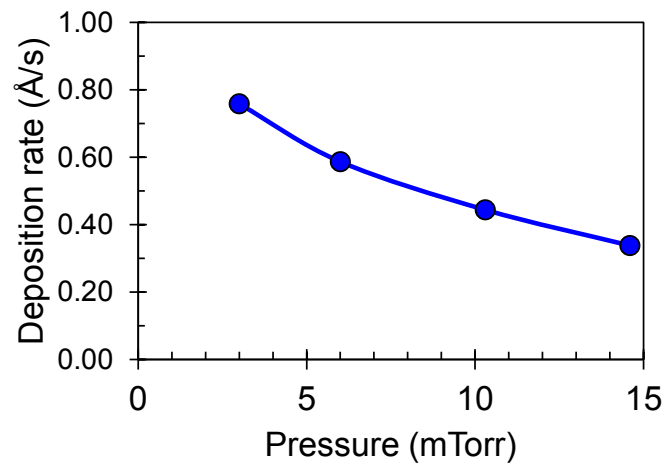


Figure 4.28: TiN thin film deposition rate as a function of pressure conditions into RF-PVD chamber during TiN deposition.

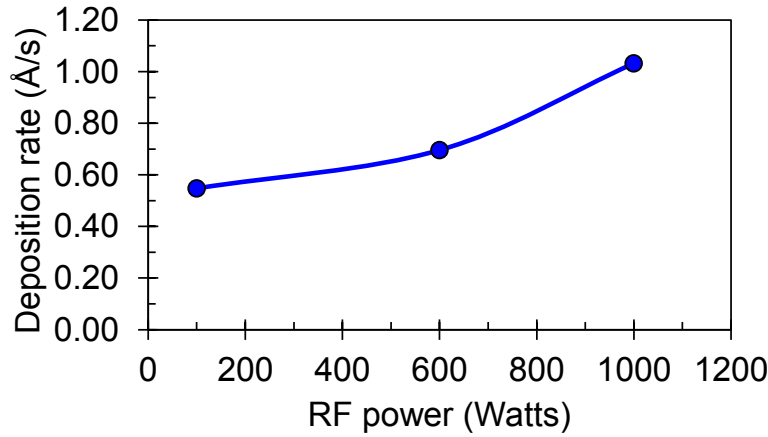


Figure 4.29: TiN thin film deposition rate as a function of RF-PVD conditions applied to the RF-PVD chamber during TiN deposition.

4.3.3 Electrical impact and local V_{TH} variability improvement with TiN microstructure engineering

In order to determine the electrical impact of the modification of TiN microstructure obtained with pressure and RF power variations, we first perform 4 point sheet resistance (R_s) measurements on 100 Å-thick TiN/200 Å-thick SiO₂ along the wafers. The results for different pressure and RF power conditions are, respectively, reported in Figs. 4.30 and 4.31. We notice that higher dispersion of R_s measurements are obtained with higher pressures and lower RF powers. At the edges of the wafer, R_s is in the 200-300 Ω/sq range regardless of the deposition conditions. In contrast, at the center of the wafer, R_s increases for higher pressures and lower RF powers.

As the in-plane XRD measurements were performed at the center of the wafer, we can compare the grain size and the sheet resistance at the center of the wafer and a nice correlation is observed in Fig. 4.32. As the grain size decreases from 9 nm to 6 nm, the sheet resistance increases from 220 Ω/sq to 860 Ω/sq. The mechanism proposed to explain this relationship is that an increase of the grain boundaries creates more obstacles for current to flow between the 4-points probe. As a result, it is translated as an increase of the sheet resistance. In consequence, it is therefore logical to deduce that the decrease of grain size with higher pressure or lower RF power is preponderant at the center of the wafer.

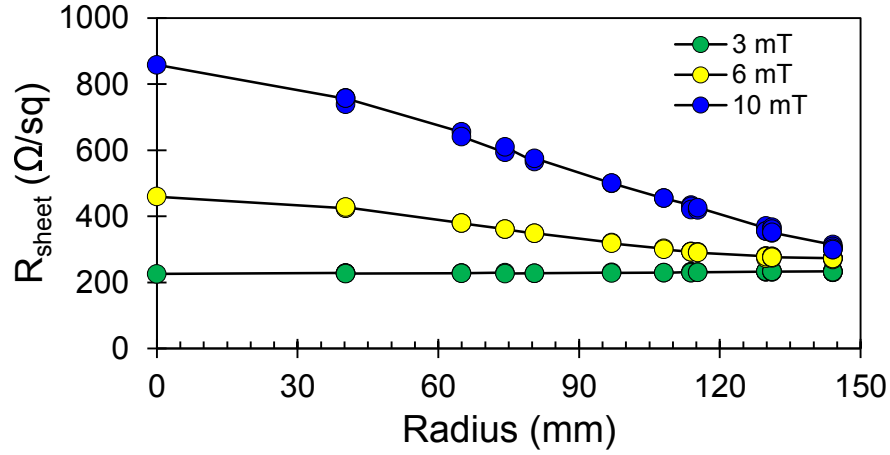


Figure 4.30: Sheet resistance measurements on 100 Å-thick TiN/200 Å-thick SiO₂ along a 300 mm wafer for different pressure conditions into RF-PVD chamber during TiN deposition.

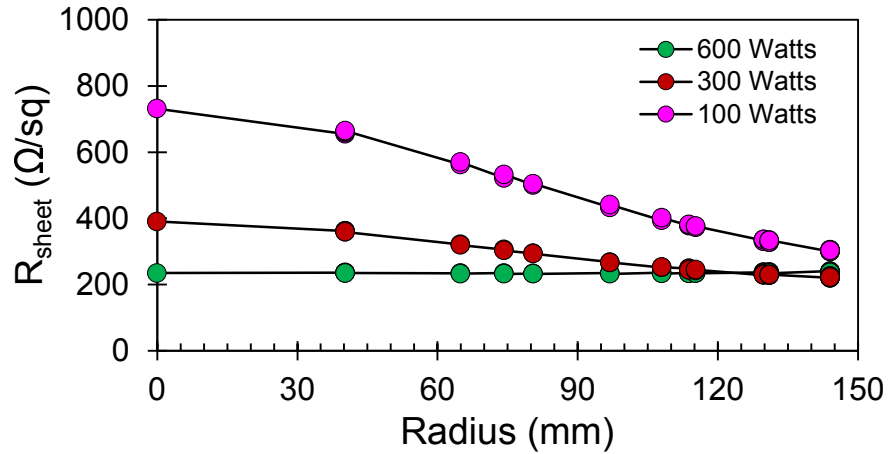


Figure 4.31: Sheet resistance measurements on 100 Å-thick TiN/200 Å-thick SiO₂ along a 300 mm wafer for different RF power conditions into RF-PVD chamber during TiN deposition.

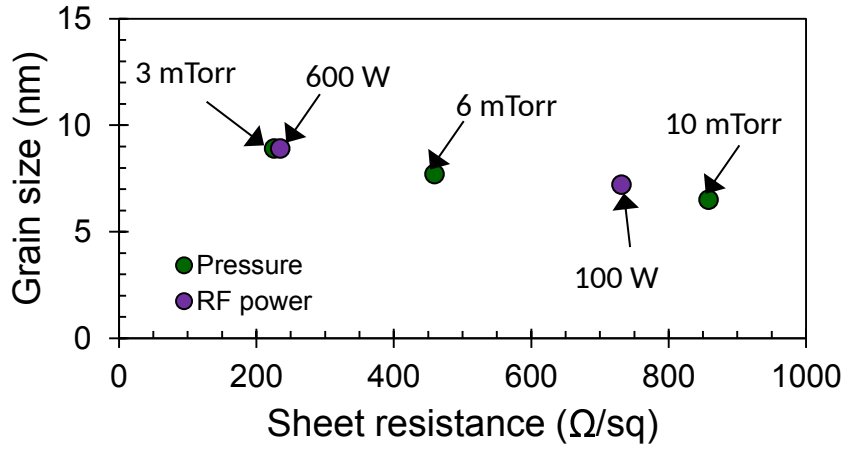


Figure 4.32: Correlation between R_s and the grain size determined by in-plane XRD at the center of the wafer.

Standard 14 nm NFET FDSOI devices were fabricated with a high pressure TiN (10 mTorr) and a low pressure TiN (3 mTorr) metal gates. The aim of the study was to investigate the influence of the decrease of the grain size at the center of the wafer on the local V_{TH} variability (ΔV_{TH}) of identical MOSFET pairs (matched devices), placed the one from the other at minimal spacing authorized by the Design Rules Checking (DRC). Such MOSFET pairs have different surface area dimensions (see section 2.1.1.3) and are located at different radius along the wafer. As the grains are smaller at the center than at the edges of the wafer for the matched MOSFET devices with a high pressure TiN metal gate, we would expect a smaller $\sigma\Delta V_{TH}$ at the center than at the edges of the wafer. In general, we would also expect a smaller $\sigma\Delta V_{TH}$ for matched devices with a high pressure TiN metal gate compared to matched devices with a low pressure TiN metal gate, especially in small areas. Indeed, in small areas, the increase of the number grains should have a more important impact on the statistical effect (see Fig. 4.18). We report in Fig. 4.33 the $\sigma\Delta V_{TH}$ of the identical MOSFET devices with a high pressure TiN metal gate as a function of the radius of the wafer for different area dimensions. However, no significant differences were observed between the center and the edges of the wafer. In contrast, the $\sigma\Delta V_{TH}$ decreases in matched devices with large area dimensions.

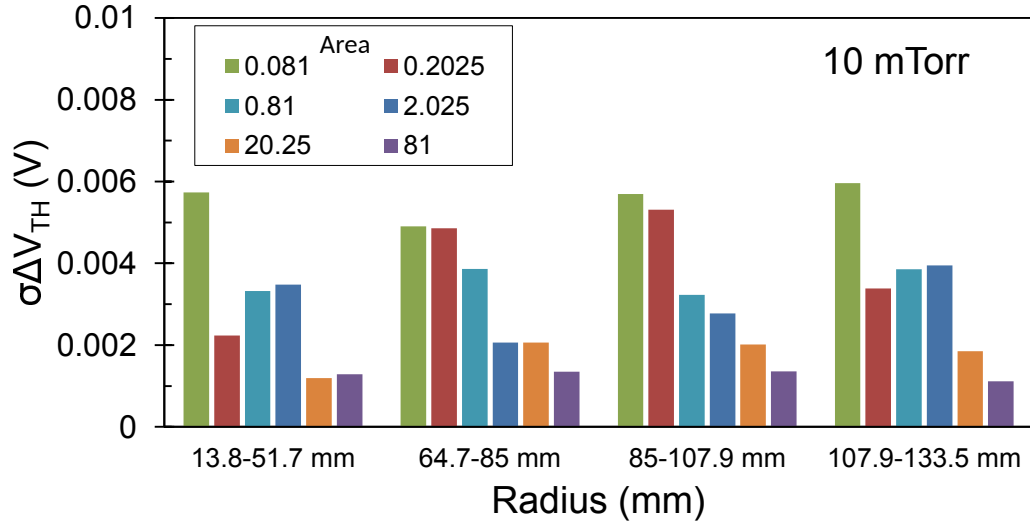


Figure 4.33: σV_{TH} as a function of the radius of the wafer for MOSFET pairs with different area dimensions.

On the other hand, the efficiency of *final* TiN deposited at higher pressure to improve the local variability of V_{TH} in FDSOI devices is demonstrated in the Pelgrom plot shown in Fig. 4.34. The slope of the Pelgrom plot (AV_{TH}) can be taken as the random fluctuation normalized with respect to the MOSFET channel size ($1/(WL)^{1/2}$). Compared to devices with low pressure TiN metal gates, AV_{TH} is reduced for TiN deposited at higher pressures. It is worth noting that AV_{TH} values are not independent on the area dimensions. The ratio $AV_{TH_{10mT}} / AV_{TH_{3mT}}$ as a function of the area of the MOSFET devices in Fig. 4.35 reveals that the $\sigma\Delta V_{TH}$ is further reduced in large area devices with higher pressure TiN metal gates. Nevertheless, it is unexpected that the greatest benefit is observed for the larger MOSFET devices. Indeed, as simulated in Fig. 4.18, the strongest decrease of the σWF_{eff} by statistical effect should be observed for the smallest areas and this benefit should saturate after a certain number of grains. In consequence, it cannot be concluded that the decrease of local variability of V_{TH} in FDSOI devices for the larger devices is due to the reduction of the average size of TiN crystallites.

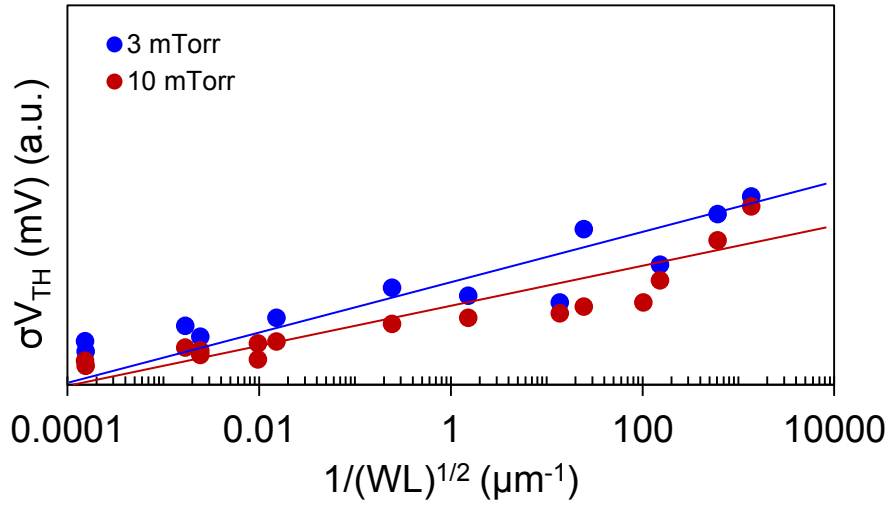


Figure 4.34: Pelgrom plot of NFET devices for low and high pressure conditions into RF-PVD chamber during TiN deposition.

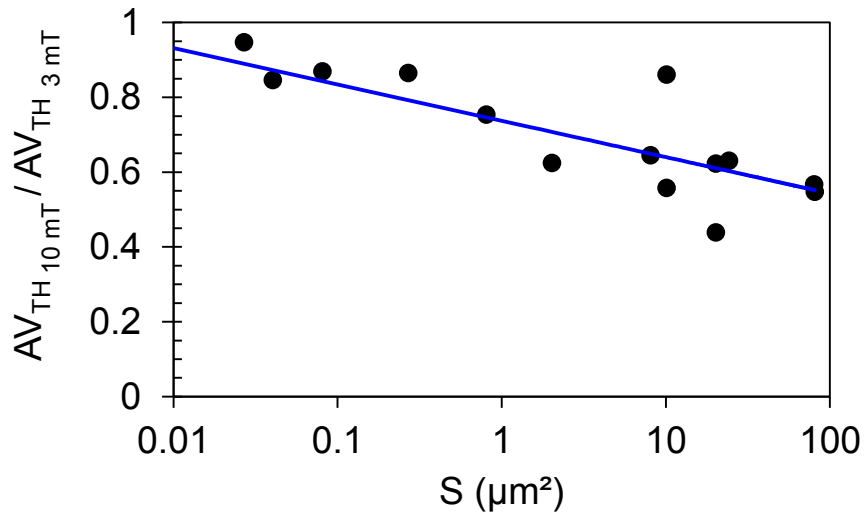


Figure 4.35: Ratio $AV_{TH_{10mT}} / AV_{TH_{3mT}}$ vs area of NFET FDSOI devices.

4.3.4 TiN microstructure versus thickness

Finally, the impact of two different thicknesses (65 Å and 100 Å) on the microstructure of TiN is shown in Fig. 4.36. The most intense orientations of TiN crystallites are still (200) and (111). Table 4.7 shows that the grains are preferentially oriented (200) with the increase of thickness from 65 Å to 100 Å. In addition, it can be noticed an increase of the average size of the grains when the thickness increases from 65 Å to 100 Å, as reported in Table 4.8. These

microstructural changes may induce a change of the average work function (Φ_m) of TiN.

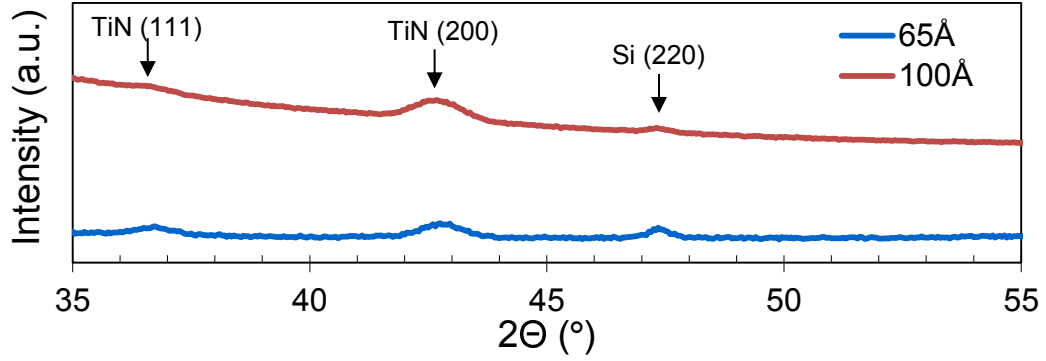


Figure 4.36: θ - 2θ X-Ray diffraction spectra for different TiN thicknesses.

TiN thickness	Parameter	TiN (111)	TiN (200)
65 Å	Normalized Area	55	100
	Relative intensity	72	100
	Percentage	36.7 %	63.3 %
100 Å	Normalized Area	21.1	100
	Relative intensity	72	100
	Percentage	14.7 %	85.3 %

Table 4.7: Determination of percentage of each grain orientation in the total population of grains from θ - 2θ X-Ray diffraction spectra for two different TiN thicknesses.

TiN type	TiN peak	Position [2 θ]	Height [cts]	Integral breadth [2 θ]	$\langle D \rangle$ [nm]
35 Å	111	36.26	462	1.333	9.6
	(200)	42.18	695	2.188	4.9
	220	61.28	386	1.896	6.3
100 Å	111	36.40	719	1.118	12.7
	(200)	42.38	1122	1.394	8.9
	220	61.54	413	2.021	5.8

Table 4.8: Determination of lateral grain size from in-plane XRD measurements by applying Scherrer equation for different TiN thicknesses.

4.4 Conclusion

The impact of nitrogen, adjusted by the thickness and composition of titanium nitride (TiN) layers on the effective work function (WF_{eff}) has been investigated in 14 nm FDSOI devices. Insignificant modulation of effective work function of HfON-based devices induced by TiN in a sacrificial gate-first approach was revealed. The fluorescence intensities showed that nitrogen is not incorporated into the gate stack after diffusion annealing at 900 °C and *sacrificial* TiN removal and is actually unchanged regardless of the thickness of TiN or the type of high- κ . Similarly, no diffusion of Ti was evidenced.

On the other hand, TiN layers proved to be more efficient to increase WF_{eff} by increasing TiN thickness in a standard final integration. The origin of this shift is uncertain and it might be related either to a variation of the TiN microstructure or to a higher oxidation of TiN. The activation of the diffusion of nitrogen after an annealing at 1005 °C during the spike anneal should not be excluded either to explain the WF_{eff} increase with thicker *final* TiN layers. Nevertheless, a strong roll-off of the dipole induced by TiN is evidenced at EOT $< 20 \text{ \AA}$, as a result of new developments in devices with beveled oxides. In addition, the roll-off is found to be stronger for thicker Poly-Si/TiN electrodes. In consequence, roll-off also makes final TiN thickness useless for V_{TH} modulation at nominal EOT in 14 nm FDSOI devices. Instead, the use of aluminum in a *sacrificial* gate-first approach is highly recommended to conceive P-gates in order to meet the requirements of the multi- V_{TH} offer in 14 nm FDSOI devices, not only because of the precise control of the incorporated effective dose but also due to the weaker roll-off effect at nominal EOT.

Finally, we proposed an innovative process of metal deposition in order to change the microstructural properties of TiN. The impact of both pressure and Radio-Frequency power conditions during TiN deposition by RF-PVD at room temperature on the preferred orientation and average size of TiN grains were investigated. Reduction of the average size of TiN grains on the center of the wafer has been successfully achieved by increasing pressure or by decreasing RF power, as experimentally demonstrated by performing X-Ray diffraction measurements in in-plane configuration. In particular, final TiN deposited at higher pressure is proposed in order to further reduce the local variability of V_{TH} in FDSOI devices. Nevertheless, it is unexpected that the greatest benefit is observed for the largest MOSFET devices. Indeed, we would expect that the strongest decrease of the σWF_{eff} by statistical effect is observed for the smallest areas and this benefit should saturate after a certain number of grains.

In consequence, the decrease of local variability of V_{TH} observed in high pressure TiN metal gates in FDSOI devices can only be partially explained by the reduction of the average size of TiN crystallites. Other sources of the local variability of V_{TH} such as line edge roughness or oxide trapped charges, should be investigate to explain, in particular, the decrease of local variability of V_{TH} in large MOSFET devices.

Conclusions and Perspectives

Summary of the context and objectives

This Ph.D. thesis work has been focused on the fabrication and the electrical and physicochemical characterization of metal gates in 14 nm high-k based FDSOI devices, manufactured at STMicroelectronics. These metal gates are made up with titanium nitride (TiN), lanthanum and aluminum layers, deposited by Radio-Frequency sputtering (RF-PVD). The main objectives of this thesis were on one hand the understanding of the modulation of the effective work function through the incorporation of additives such as aluminum, lanthanum and nitrogen following a *sacrificial* metal gate-first approach, on the other hand, the study of the impact of the TiN microstructure on the V_{TH} variability of matched MOSFET devices. The goal was to validate the sacrificial gate-first integration to successfully achieve the threshold voltage requirements of the 14FDSOI technology and to propose metal deposition conditions to further reduce the local V_{TH} variability in advanced FDSOI nodes.

In the first chapter, we have remembered the definition of the main parameters of the metal-oxide-semiconductor stack. We have seen that most of the performance parameters of MOSFET devices are strongly influenced by the equivalent oxide thickness (EOT) and the effective work function (WF_{eff}). We have identified the electrical impact of the presence of dipoles at the high- κ /SiON interface and the density of interfacial fixed charges into dielectrics on such effective work function. We have shown that the work function range to switch from an N-gate to a P-gate in FDSOI devices is shorter than in bulk devices due to the low amount of depletion charges in the thin silicon layer over the buried oxide. As a result, a very fine adjustment of the effective work function of metal electrodes is even more necessary. The *sacrificial* gate-first integration has been described as a process scheme able to reach the V_{TH} targets for both NFETs and PFETs by the careful incorporation of additives such as lanthanum and aluminum. We have taken advantage of the sacrificial gate-first integration to decouple the main objectives of this thesis. On one hand, it was mainly with the sacrificial gate that we sought to modulate the V_{TH} . On the other hand, it was with the final gate that we sought to reduce the local V_{TH} variability. Finally, we have assumed that the local V_{TH} variability due to metal gate granularity is related to a different value of work function associated to each crystallite orientation. In the case of TiN, (200) orientation is associated to a Φ_m value of 4.6 eV and (111) orientation is associated to a Φ_m value of 4.4 eV. For that reason, three technical solutions were proposed in order to suppress or reduce the variability of the effective work function: the deposition of an amorphous metal gate, the deposition of a metal gate with a same and unique metal orientation, and the deposition of a metal gate with smaller crystallites in order to reduce the variability by statistical effect.

In the second chapter, we described the test structures and the techniques of electrical and physicochemical characterization implemented in this thesis for characterization of metal gate stack. A simplified process flow of the 14 nm FDSOI technology with only the key bricks of one technological option (NFET Low V_{TH}) was validated in order to intensively study the impact of the sacrificial metal gate deposition processes and the incorporation of additives such as lanthanum and aluminum on the effective work function (WF_{eff}) and the equivalent oxide thickness (EOT) of the gate stack. We have introduced our methodology for an accurate extraction of these parameters from the experimental C-V characteristics. Both WF_{eff} and EOT were extracted by fitting with Poisson-Schrödinger quantum simulations, a relevant part of the experimental high-frequency C-V curve, free from leakage and interfacial defects. We have then introduced our approach to determine the effective work function modulation and to discriminate the influence of a dipole and the influence of fixed charges. Indeed, the integration of a beveled oxide in this simplified process flow authorizes the identification of the origin of the modulation of the effective work function. When two different wafers, each one with a specific process or dopant (La, Al, N) in *sacrificial* metal gate stack, are compared by combining C-V techniques and devices with beveled oxides, the origin of the resulting WF_{eff} shift can thus be identified from the plot WF_{eff} vs EOT. If the y-intercept of the WF_{eff} vs EOT plot (WF_{eff} at $EOT = 0$) is unaltered, but the slopes of the 2 curves are different, then the WF_{eff} shift will be explained by interfacial fixed charges induced by the process or dopant in metal gate stack. On the contrary, if the y-intercept is modified but the slope remains unchanged, then the WF_{eff} shift will be explained by a dipole effect or a work function change of *final* TiN induced by the process or dopant in metal gate stack.

Contributions

In the second chapter, a new methodology based on X-Ray Fluorescence was proposed to accurately measure the effective diffusion of these additives into the gate stack after annealing. This methodology consists in the XRF measurement on at least 9 points on blanket wafers with nominal dielectrics thicknesses after sacrificial gate stack deposition and after diffusion annealing and sacrificial gate stack removal. Both measurements are necessary in order to determine the percentage of diffusion for a given element. Finally, we described the θ -2 θ and the in-plane X-Ray configurations. The investigation of the preferential orientation and horizontal size of the TiN crystallites as a function of the process deposition conditions is allowed with such X-Ray configurations.

In third chapter, we demonstrate a successful incorporation of lanthanum and aluminum in the dielectrics stack after diffusion annealing performed at 900 °C for 10 seconds. Furthermore, engineering of WF_{eff} is demonstrated by tuning bottom TiN and additives thicknesses without gate leakage degradation. WF_{eff} modulation towards N+ with lanthanum and towards P+ with aluminum are efficiently moderated by the bottom TiN barrier diffusion layer, making this approach suitable for V_{TH} requirements in FDSOI devices. Additionally, we demonstrate that the *sacrificial* gate-first approach authorizes the prediction and accurate control of diffusion of additives into HfON/SiON stack. Indeed, the methodology based on X-Ray fluorescence was successfully employed in order to model the effective doses of additives incorporated into HfON/SiON stack as a function of the as-deposited dose, the bottom TiN thickness and the temperature annealing in the sacrificial gate. From that study, we have identified that the percentage of diffusion of both additives exponentially decreases as the bottom TiN increases. Furthermore, we have also identified that bottom TiN was more efficient to screen the diffusion of lanthanum than that of aluminum. Such capability of TiN to screen the diffusion increases as the annealing temperature decreases. In the case of lanthanum, the effective dose linearly increases with the as-deposited dose. In contrast, in the case of aluminum, the diffusion of aluminum is stopped for higher as-deposited doses. A higher temperature would therefore be required to further increase the diffusion of aluminum.

Moreover, devices with beveled oxides let us to identify the origin of the modulation of the effective work function at nominal EOT. This shift was attributed to the drop bias induced by a dipole layer. Such drop related to a dipole evolves with the effective dose of the incorporated additive, towards N+ with the incorporation of lanthanum and towards P+ with the incorporation of aluminum. In addition it is drastically reduced towards N+ values by the influence of a roll-off phenomenon at smaller EOT. Such roll-off component evolves with both the effective dose and the SiO_x thickness. Accordingly, empirical equations of the diffusion of additives into the gate dielectrics and their impact on the effective work function of metal gates have been proposed to precisely modulate the threshold voltage (V_{TH}) at nominal EOT of the 14 nm FDSOI devices. By taking into account the roll-off bias drop for the nominal EOT of the 14 nm FDSOI technology, the lanthanum shifts the WF_{eff} around -80 meV for each 1×10^{14} at/cm² of effective lanthanum dose whereas aluminum does around +25 meV for each 1×10^{14} at/cm² of effective aluminum dose. Evidently, the roll-off potential drop is translated by a maximization of the strength of the potential drop induced by the incorporation of lanthanum and by a decrease of the strength of the potential drop induced by the incorporation of aluminum. For the sake of comparison, we have removed such roll-off potential drop to determine the strength of the dipole for effective doses of

aluminum and lanthanum incorporated into HfON/SiON stack. This corresponds to a WF_{eff} shift at $EOT = 0$ of +41 meV for each Al dose of 1×10^{14} at/cm² and a WF_{eff} shift at $EOT = 0$ of -53 meV for each La dose of 1×10^{14} at/cm² incorporated into HfON layer. In addition, the influence of the high- κ oxide on both the diffusion of additives and the modulation of the effective work function was investigated. Both aluminum and lanthanum clearly diffuse more easily into HfON than into HfSiON after annealing at 900 °C for 10 seconds. Such difference might explain a more significant WF_{eff} modulation achieved on HfON devices for equivalent as-deposited doses of aluminum and lanthanum. Since the equivalent oxide thicknesses were different for devices with HfON and HfSiON, the comparison of the WF_{eff} at nominal EOT is not authorized. Instead, we have compared the WF_{eff} at $EOT=0$ for equivalent additive doses incorporated into both the HfON/SiON stack and the HfSiON/SiON stack. This corresponds to a WF_{eff} shift at $EOT = 0$ of +36 meV for each Al dose of 1×10^{14} at/cm² incorporated into the HfSiON/SiON stack and a WF_{eff} shift at $EOT = 0$ of +41 meV for each Al dose of 1×10^{14} at/cm² incorporated into the HfON/SiON stack. On the other hand, we found a WF_{eff} shift at $EOT = 0$ of -54 meV for each La dose of 1×10^{14} at/cm² incorporated into the HfSiON/SiON stack and a WF_{eff} shift at $EOT = 0$ of -53 meV for each La dose of 1×10^{14} at/cm² incorporated into the HfON/SiON stack. Accounting for this weak difference, it is therefore logical to conclude that the ΔWF_{eff} at $EOT=0$ is proportional to the effective additive dose incorporated into the high- κ /SiO_x stack. It is more significant on HfON compared to HfSiON because the effective dose incorporated into HfON is proportionally more significant on HfON compared to HfSiON for equivalent as-deposited doses.

In the fourth chapter, the impact of nitrogen, adjusted by the thickness and composition of titanium nitride (TiN) layers on the effective work function (WF_{eff}) has been investigated in 14 nm FDSOI devices. Insignificant modulation of effective work function of HfON-based devices induced by TiN in a sacrificial gate-first approach was revealed. The fluorescence intensities showed that nitrogen is not incorporated into the gate stack after diffusion annealing at 900 °C for 10 s. The level of nitrogen into both HfON and HfSiON is actually unchanged regardless of the thickness of TiN. On the other hand, TiN layers proved to be more efficient to increase WF_{eff} at $EOT = 0$ by increasing TiN thickness in a standard final integration. The origin of this shift is uncertain and it might be related either to a variation of the TiN microstructure or to a higher oxidation of TiN. The activation of the diffusion of nitrogen after an annealing at 1005 °C during the spike anneal should not be excluded either to explain the WF_{eff} increase with thicker *final* TiN layers. Nevertheless, a strong roll-off for thicker Poly-Si/TiN electrodes is evidenced at $EOT < 20 \text{ \AA}$, thanks to the beveled oxides. In consequence, roll-off also makes final TiN thickness useless for V_{TH} modulation at nominal EOT in 14 nm FDSOI devices. Instead, the use

of aluminum in a *sacrificial* gate-first approach is highly recommended to conceive P-gates in order to meet the requirements of the multi- V_{TH} offer in 14 nm FDSOI devices, not only because of the precise control of the incorporated effective dose but also due to the weaker roll-off effect at nominal EOT.

Finally, we proposed an innovative process of metal deposition in order to change the microstructural properties of TiN. The impact of both pressure and Radio-Frequency power conditions during TiN deposition by RF-PVD at room temperature on the preferred orientation and average size of TiN grains were investigated. Reduction of the average size of TiN grains on the center of the wafer has been successfully achieved by increasing pressure or by decreasing RF power, as experimentally demonstrated by performing X-Ray diffraction measurements in a in-plane configuration. In particular, final TiN deposited at higher pressure was proposed in order to further reduce the local variability of V_{TH} in FDSOI devices. The proposed new process was implemented in the TiN metal gate of FDSOI devices and shown better results for local V_{TH} variability. Nevertheless, it is unexpected that the greatest benefit is observed for the largest MOSFET devices. Indeed, we would expect that the strongest decrease of the σWF_{eff} by statistical effect is observed for the smallest areas and this benefit should saturate after a certain number of grains. In consequence, the decrease of local variability of V_{TH} observed in high pressure TiN metal gates in FDSOI devices can only be partially explained by reduction of the average size of TiN crystallites. In particular, the decrease of local variability of V_{TH} in large MOSFET devices must be investigated in a future work.

Recommendations for future research

We have paved the way for further studies in short-channel FDSOI devices at STMicroelectronics. Indeed, specially designed test structures using LETI patent for parasitic capacitances suppression were successfully embedded in a new mask at STMicroelectronics in order to ease the extraction of WF_{eff} and EOT in short-channel MOSFET devices. For a future work, it would be interesting to correlate the effective work function shift determined on such short MOSFET devices with the effective incorporated dose.

It would also be necessary to carry out a more detailed analysis of the diffusion of additives into the dielectrics stack. Although the effective incorporated doses into high- κ /SiO_x stack has been accurately quantified by XRF, the impossibility of this technique to be used for depth profiling measurements does not authorize to discriminate the effective dose at high- κ /SiO_x interface. Other physicochemical characterizations such as SIMS must be performed in order to complete the information about the diffusion of aluminum into different Hf-based dielectrics/SiON stacks. Furthermore, *ab*

initio simulations of the electrostatic dipole modulation due to cation substitution in HfON/SiON and HfSiON/SiON interfaces with different concentrations of silicon (%Si) should help to improve the understanding of the WF_{eff} modulation in devices with different high- κ materials.

Major efforts must be done in order to clarify the origin of the potential drop induced by the roll-off phenomenon. It cannot be only related to the oxygen vacancies at the high- κ /SiO_x since it has been proved that the potential drop is proportional to the effective dose of lanthanum and aluminum. Therefore, further investigation should be encouraged to provide a clear relationship between the effective incorporated dose of a given additive and the bonds formed with oxygen into dielectrics. In particular the discrepancies in the roll-off potential drop observed for lanthanum, aluminum and TiN must be investigated in a future work. Such discrepancies might be related to the resistance to oxidation of the incorporated additives.

A comparison of the results obtained on RF-PVD TiN layers with TiN films deposited using other deposition techniques, such as ALD or Plasma-Enhanced ALD, would be interesting to determine whether different metal deposition conditions may impact the preferential orientation of TiN, the average grain size or the V_{TH} variability. It would also be interesting to study the impact of the different TiN microstructures on other aspects of the V_{TH} variability such as the line edge roughness or the oxide trapped charges. Furthermore, the integration of amorphous metal gates such as TiSiN or TaSiN should be investigated.

Finally, we validate and highly recommend the sacrificial gate-first approach to incorporate additives into the gate stack of the 14 nm FDSOI devices, but also in most advanced nodes. This approach allows the prediction and accurate control of diffusion of lanthanum and aluminum into high- κ /SiO_x stack. In addition, this approach authorizes the simultaneous NFET and PFET gate patterning in a standard 14 nm FDSOI process flow. Since the gate stack containing the additives is entirely removed by wet etching, the gate patterning is only influenced by the *final* Poly-Si/TiN electrode without additives and high- κ stack. Moreover, the WF_{eff} increase towards P+ induced by Al diffusion is stronger with a *sacrificial* gate-first approach since the opposite shift towards N+ of the Al-based metal work function is cancelled after sacrificial gate removal.

Bibliography

- [1] G. E. Moore, "Cramming more components onto integrated circuits," in *Proceedings of IEEE*, vol. 86, pp. 82–85, 1989.
- [2] R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. Leo, E. Bassous, and A. R. LeBlanc, "Design of ion-implanted MOSFETs with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. SC-9, pp. 256–268, 1974.
- [3] S. K. Wiedmann, "Advancements in bipolar VLSI circuits and technologies," *IEEE Journal of Solid-State Circuits*, vol. 19, pp. 282–291, 1984.
- [4] Semiconductor Industry Association, *International Technology Roadmap for Semiconductors*, 2002.
- [5] P. Ranade, Q. Lu, I. Polishchuk, H. Takeuchi, C. Hu, and T. King, "Dual work function metal gate technology for future CMOS devices," in *3rd International Conference on Microelectronics and Interfaces, Symposium Proceedings*, pp. 154–157, American Vacuum Society, 2002.
- [6] Y.-C. Yeo, P. Ranade, T.-J. King, and C. Hu, "Effects of high- κ gate dielectric materials on metal and silicon gate workfunctions," *IEEE Electron Devices Letters*, vol. 23, pp. 342–344, 2002.
- [7] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*. John Wiley & Sons, 3rd edition, 2007.
- [8] J. Colinge and C. Colinge, *Physics of Semiconductor Devices*. Springer, 2002 edition.
- [9] M. Charbonnier, C. Leroux, V. Cosnier, P. Besson, E. Martinez, N. Benedetto, C. Licitra, N. Rochat, C. Gaumer, K. Kaja, G. Ghibaudo, F. Martin, and G. Reimbold, "Measurement of dipoles/roll-off/work functions by coupling CV and IPE and study of their dependence on fabrication process," *IEEE Trans. Electron Devices*, vol. 57, pp. 1809–1819, 2010.
- [10] B. E. Deal, "Standardized terminology for oxide charges associated with thermally oxidized silicon," *IEEE Transactions on Electron Devices*, vol. 27, pp. 606–608, 1980.
- [11] B. E. Deal, "The current understanding of charges in the thermally oxidized silicon structure," *Journal of Electrochemical Society*, vol. 121, pp. 198C–205C, 1974.
- [12] Y. C. Cheng, "Electronic states at the silicon-silicon dioxide interface," *Progress in Surface Science*, vol. 8, pp. 181–218, 1977.

- [13] R. Jha, J. Gurganos, Y. Kim, R. Choi, J. Lee, and V. Misra, "A capacitance-based methodology for work function extraction of metals on high- κ ," *IEEE Electron Devices Letters*, vol. 25, pp. 420–423, 2004.
- [14] J. Robertson and R. M. Wallace, "High-K materials and metal gates for CMOS applications," *Materials Science and Engineering R*, vol. 88, pp. 1–41, 2015.
- [15] C. C. Hobbs, L. R. C. Fonseca, A. Knizhnik, V. Dhandapani, S. B. Samavedam, W. J. Taylor, J. M. Grant, L. G. Dip, D. H. Triyoso, R. I. Hegde, D. C. Gilmer, R. Garcia, D. Roan, M. L. Lovejoy, R. S. Rai, E. A. Hebert, H.-H. Tseng, S. G. H. Anderson, B. E. White, and P. J. Tobin, "Fermi-level pinning at the Polysilicon/Metal oxide interface-Part I," vol. 51, pp. 971–977, 2004.
- [16] K. Shiraishi, K. Yamada, K. Torii, Y. Akasaka, K. Nakajima, M. Kohno, T. Chikyo, H. Kitajima, and T. Arikado, "Physics in Fermi level pinning at the polySi/Hf-based high-k oxide interface," in *2004 Symposium on VLSI Technology Digest of Technical Papers*, pp. 108–109.
- [17] E. Cartier, F. R. McFeely, V. Narayanan, P. Jamison, B. P. Linder, M. Copel, V. K. Paruchuri, V. S. Basker, R. Haight, D. Lim, R. Carruthers, T. Shaw, M. Steen, J. Sleight, J. Rubino, H. Deligianni, S. Guha, R. Jammy, and G. Shahidi, "Role of oxygen vacancies in V_{fb}/V_t stability of pFET metals on HfO_2 ," in *2005 Symposium on VLSI Technology Digest of Technical Papers*, pp. 230–231.
- [18] J. K. Schaeffer, L. R. C. Fonseca, S. B. Samavedam, Y. Liang, P. J. Tobin, and B. E. White, "Contributions to the effective work function of platinum on hafnium dioxide," *Applied Physics Letters*, vol. 85, p. 1826, 2004.
- [19] J. K. Schaeffer, D. C. Gilmer, C. Capasso, S. Kalpat, B. Taylor, M. V. Raymond, D. Triyoso, R. Hegde, S. B. Samavedam, and B. E. W. Jr, "Application of group electronegativity concepts to the effective work functions of metal gate electrodes on high- κ oxides," *Microelectronic Engineering*, vol. 84, pp. 2196–2200, 2007.
- [20] H. N. Alshareef, K. Choi, H. C. Wen, H. Luan, H. Harris, Y. Senzaki, P. Majhi, B. H. Lee, R. Jammy, S. Aguirre-Tostado, B. E. Gnade, and R. M. Wallace, "Composition dependence of the work function of $\text{Ta}_{1-x}\text{Al}_x\text{N}_y$ metal gates," *Applied Physics Letters*, vol. 88, pp. 072108–1:3, 2006.
- [21] V. Narayanan, V. K. Paruchuri, N. A. Bojarczuk, B. P. Linder, B. Doris, Y. H. Kim, S. Zafar, J. Stathis, S. Brown, J. Arnold, M. Copel, M. Steen, E. Cartier, A. Callegari, P. Jamison, J. P. Locquet, D. L. Lacey, Y. Wang, P. E. Batson, P. Ronsheim, R. Jammy, M. P. Chudzik, M. Jeong, S. Guha,

- G. Shahidi, and T. C. Chen, "Band-edge high-performance high- κ metal gate n-MOSFETs using cap layers containing group IIA and IIIB elements with gate-first processing for 45 nm and beyond," in *2006 Symposium on VLSI Technology Digest of Technical Papers*, pp. 224–225.
- [22] K. Iwamoto, H. Ito, Y. Kamimuta, Y. Watanabe, W. Mizubayashi, S. Migita, Y. Morita, M. Takahashi, H. Ota, T. Nabatame, and A. Toriumi, "Re-examination of flat band voltage shift for high-k MOS devices," in *2007 Symposium on VLSI Technology Digest of Technical Papers*, pp. 70–71.
- [23] K. K. Y. Yamamoto, K. Kita and A. Toriumi, "Study of La-induced flat band voltage shift in Metal/HfLaOx/SiO₂/Si capacitors," *Japanese Journal of Applied Physics*, vol. 46, pp. 7251–7255, 2007.
- [24] G. Bersuker, C. Park, H.-C. Wen, K. Choi, J. Price, P. Lysaght, H.-H. Tseng, O. Sharia, A. Demkov, J. T. Ryan, and P. Lenahan, "Origin of the flatband voltage roll-off phenomenon in metal/high- κ stacks," *IEEE Transactions on Electron Devices*, vol. 57, pp. 2047–2056, 2010.
- [25] C. Leroux, S. Baudot, M. Charbonnier, A. V. der Geest, P. Caubet, A. Toffoli, P. Blaise, G. Ghibaudo, F. Martin, and G. Reimbold, "Investigating doping effects on high- κ metal gate stack for effective work function engineering," *Solid State Electronics*, vol. 88, pp. 21–26, 2013.
- [26] Y. Kamimuta, K. Iwamoto, Y. Nunoshige, A. Hirano, W. Mizubayashi, Y. Watanabe, S. Migita, A. Ogawa, H. Ota, T. Nabatame, and A. Toriumi, "Comprehensive study of V_{fb} shift in high-k CMOS - Dipole formation, Fermi-level pinning and Oxygen vacancy effect," in *Electron Devices Meeting (IEDM), 2007 IEEE International*, pp. 341–344.
- [27] K. Kita and A. Toriumi, "Origin of electric dipoles formed at high- κ /SiO₂ interface," *Applied Physics Letters*, vol. 94, pp. 132902–1:3, 2009.
- [28] A. G. V. der Geest, P. Blaise, and N. Richard, "Ab initio study of electrostatic dipole modulation due to cation substitution in HfO₂/SiO₂ interfaces," *Physical Review B*, vol. 86, pp. 085320–1:13, 2012.
- [29] B. Majkusiak, "Gate tunnel current in an MOS transistors," *IEEE Transactions on Electron Devices*, vol. 37, pp. 1087–1092, 2002.
- [30] W. A. Harrison, "Tunneling from an independent-particle point of view," *Physical Review*, vol. 123, pp. 85–89, 1961.
- [31] J. C. Ranuarez, M. J. Deen, and C. H. Chen, "A review of gate tunneling current in MOS devices," *Microelectronics Reliability*, vol. 46, pp. 1939–1956, 2006.

- [32] W.-C. Lee and C. Hu, "Modeling CMOS tunneling currents through ultra thin gate oxides due to conduction and valence band electron and hole tunneling," *IEEE Transactions on Electron Devices*, vol. 48, pp. 1366–1373, 2001.
- [33] Y.-C. Yeo, T.-J. King, and C. Hu, "MOSFET gate leakage modeling and selection guide for alternative gate dielectrics based on leakage considerations," *IEEE Transactions on Electron Devices*, vol. 50, pp. 1027–1035, 2003.
- [34] E. L. Jordan, "A diffusion mask for Germanium," *Journal of Electrochemical Society*, vol. 108, pp. 478–481, 1961.
- [35] J. R. Ligenza, "Silicon oxidation in an oxygen plasma excited by microwaves," *Journal of Applied Physics*, vol. 36, p. 2703, 1965.
- [36] J. R. Ligenza and M. Kuhn, "DC arc anodic plasma oxidation: A new vacuum process for solid state device fabrication," *Solid State Technology*, vol. 13, p. 33, 1970.
- [37] R. Doering and Y. Nishi, *Handbook of Semiconductor Manufacturing and Technology*. CRC Press, Taylor & Francis Group, 2008.
- [38] P. F. Schmidt and W. Michel, "Anodic formation of oxide films on silicon," *Journal of Electrochemical Society*, vol. 104, pp. 230–236, 1957.
- [39] B. E. Deal and A. S. Grove, "General relationship for the thermal oxidation of Silicon," *Journal of Applied Physics*, vol. 36, pp. 3770–3778, 1965.
- [40] Y. J. van der Meulen, "Kinetics of thermal growth of ultra-thin layers of SiO₂ on silicon," *Journal of Electrochemical Society*, vol. 119, pp. 530–534, 1972.
- [41] H. Z. Massoud and J. D. Plummer, "Analytical relationship for the oxidation of silicon in dry oxygen in the thin-film regime," *Journal of Applied Physics*, vol. 62, pp. 3416–3423, 1987.
- [42] E. P. Gusev, H. C. Lu, T. Gustafsson, and E. Garfunkel, "Growth mechanism of thin silicon oxide films on Si (100) studied by medium-energy ion scattering," *Physical Review B*, vol. 52, pp. 1759–1775, 1995.
- [43] S. Gorantla, S. Muthuvenkatraman, and R. Venkat, "A model for thermal growth of ultrathin silicon dioxide in O₂ ambient: A rate equation approach," *IEEE Transactions on Electron Devices*, vol. 45, pp. 336–338, 1998.

- [44] S. Kuppurao, H. S. Joo, and G. Miner, "In-situ steam generation: A new rapid thermal oxidation technique," *Solid State Technology*, vol. 43, p. 233, 2000.
- [45] S. V. Hattangady, H. Niimi, and G. Lucovsky, "Controlled Nitrogen Incorporation at the gate oxide surface," *Applied Physics Letters*, vol. 66, pp. 3495–3497, 1995.
- [46] W. Ting, H. Hwang, J. Lee, and D. L. Kwong, "Growth kinetics of ultrathin SiO₂ films fabricated by Rapid Thermal Oxidation of Si substrates in N₂O," *Journal of Applied Physics*, vol. 70, pp. 1072–1074, 1991.
- [47] D. J. DiMaria and J. H. Stathis, "Trapping and trap creation studies on nitrided and reoxidized-nitrided silicon dioxide films on silicon," *Journal of Applied Physics*, vol. 70, pp. 1500–1509, 1991.
- [48] J. S. Jeon, C. Wong, J. Gray, H. S. Kim, and B. Ogle, "Preparation of sub 20a thick ultra-thin stack gate dielectrics by in-situ RTCVD processes," in *Electrochemical Society Proceedings*, pp. 183–188, 2001.
- [49] S. A. Campbell, *The Science and Engineering of Microelectronic Fabrication*. Oxford University Press, 2001.
- [50] L. Tsetseris, X. J. Zhou, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, "Physical mechanisms of negative bias temperature instability," *Applied Physics Letters*, vol. 86, pp. 142103:1–3, 2005.
- [51] I. J. Baumvol, F. C. Stedile, J. J. Ganem, I. Trimaille, and S. Rigo, "Thermal nitridation of SiO₂ films in ammonia: The role of hydrogen," *Journal of Electrochemical Society*, vol. 143, pp. 1426–1434, 1996.
- [52] J. Robertson, "Band offsets of wide-band-gap oxides and implications for future electronic devices," *Journal of Vacuum Science and Technology B*, vol. 18, pp. 1785–1791, 2000.
- [53] P. Lysaght, P. Chen, R. Bergmann, T. Messina, R. Murto, and H. Huff, "Experimental observations of the thermal stability of high- κ gate dielectric materials on silicon," *Journal of Non-Crystalline Solids*, vol. 303, pp. 54–63, 2002.
- [54] M. Koike, T. Ino, Y. Kamimuta, M. Koyama, Y. Kamata, M. Suzuki, Y. Mitani, A. Nishiyama, and Y. Tsunashima, "Effect of Hf-N bond on properties of thermally stable amorphous HfSiON and applicability of this material to sub-50 nm technology node LSIs," in *IEDM, Technical Digest*, pp. 4.7.1–4.7.4, 2003.

- [55] M. R. Visokay, J. J. Chambers, A. L. P. Rotondaro, A. Shanware, and L. Colombo, "Application of HfSiON as a gate dielectric material," *Applied Physics Letters*, vol. 80, p. 3183, 2002.
- [56] G. D. Wilk, R. M. Wallace, and J. M. Anthony, "Hafnium and zirconium silicates for advanced gate dielectrics," *Journal of Applied Physics*, vol. 87, pp. 484–492, 2000.
- [57] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C.-H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Heussner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. McIntyre, P. Moon, J. Neiryneck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, and K. Zawadzki, "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging," in *IEEE International Electron Devices Meeting (IEDM)*, pp. 247–250, 2007.
- [58] M. T. Bohr, R. Chau, T. Ghani, and K. Mistry, "The high-k solution," *IEEE Spectrum*, pp. 30–35, 2007.
- [59] K.-E. Elers, V. Saanila, P. Soininen, W.-M. Li, J. Kostamo, and S. Haukka, "Diffusion barrier deposition on a copper surface by atomic layer deposition," *Chemical Vapor Deposition*, vol. 8, pp. 149–153, 2002.
- [60] M. Ritala, M. Leskela, E. Rauhala, and P. Haussalo, "Atomic layer epitaxy growth of TiN thin films," *Journal of Electrochemical Society*, vol. 142, pp. 2731–2737, 1995.
- [61] J. W. Elam, M. Schuisky, J. D. Ferguson, and S. M. George, "Surface chemistry and film growth during TiN atomic layer deposition using TMADT and NH₃," *Thin Solid Films*, vol. 436, pp. 145–156, 2003.
- [62] M. A. Lieberman and A. J. Lichtenberg, *Principles of plasma discharges and materials processing*. John Wiley & Sons, Inc., 2005.
- [63] *Applied Endura2 RF PVD chamber manual*, 2009.
- [64] S. Baudot, *Elaboration et caracterisation des grilles metalliques pour les technologies CMOS 32/28 nm a base de dielectrique haute permittivite*. PhD thesis, Universite de Grenoble, 2012.

- [65] C.-H. Jan, U. Bhattacharya, R. Brain, S.-J. Choi, G. Curello, G. Gupta, W. Hafez, M. Jang, M. Kang, K. Komeyli, T. Leo, N. Nidhi, L. Pan, J. Park, K. Phoa, A. Rahman, C. Staus, H. Tashiro, C. Tsai, P. Vandervoorn, L. Yang, J.-Y. Yeh, and P. Bai, "A 22 nm SoC platform technology featuring 3-D tri-gate and high-k/metal gate, optimized for ultra low power, high performance and high density SoC applications," in *Electron Devices Meeting (IEDM), 2012 IEEE International*, pp. 3.1.1–3.1.4, 2012.
- [66] S.-Y. Wu, C. Y. Lin, M. C. Chiang, J. J. Liaw, J. Y. Cheng, S. H. Yang, M. Liang, T. Miyashita, C. H. Tsai, B. C. Hsu, H. Y. Chen, T. Yamamoto, S. Y. Chang, V. S. Chang, C. H. Chang, J. H. Chen, H. F. Chen, K. C. Ting, Y. K. Wu, K. H. Pan, R. F. Tsui, C. H. Yao, P. R. Chang, H. M. Lien, T. L. Lee, H. M. Lee, W. Chang, T. Chang, R. Chen, M. Yeh, C. C. Chen, Y. H. Chiu, Y. H. Chen, H. C. Huang, Y. C. Lu, C. W. Chang, M. H. Tsai, C. C. Liu, K. S. Chen, C. C. Kuo, H. T. Lin, S. M. Jang, and Y. Ku, "A 16nm FinFET CMOS technology for mobile SoC and computing applications," in *Electron Devices Meeting (IEDM), 2013 IEEE International*, pp. 9.1.1–9.1.4, 2013.
- [67] O. Weber, E. Josse, F. Andrieu, A. Cros, E. Richard, P. Perreau, E. Baylac, N. Degors, C. Gallon, E. Perrin, S. Chhun, E. Petitprez, S. Delmedico, J. Simon, G. Druais, S. Lasserre, J. Mazurier, N. Guillot, E. Bernard, R. Bianchini, L. Parmigiani, X. Gerard, C. Pribat, O. Gourhant, F. Abbate, C. Gaumer, V. Beugin, P. Gouraud, P. Maury, S. Lagrasta, D. Barge, N. Loubet, R. Beneyton, D. Benoit, S. Zoll, J.-D. Chapon, L. Babaud, M. Bidaud, M. Gregoire, C. Monget, B. Le-Gratiet, P. Brun, M. Mellier, A. Pofelski, L. Clement, R. Bingert, S. Puget, J.-F. Kruck, D. Hoguet, P. Scheer, T. Poiroux, J.-P. Manceau, M. Rafik, D. Rideau, M.-A. Jaud, J. Lacord, F. Monsieur, L. Grenouillet, M. Vinet, Q. Liu, B. Doris, M. Celik, S. Fetterolf, O. Faynot, and M. Haond, "14nm FDSOI technology for high speed and energy efficient applications," in *Symposium on VLSI Technology Digest of Technical Papers*, pp. 14–15, 2014.
- [68] N. Planes, O. Weber, V. Barral, S. Haendler, D. Noblet, D. Croain, M. Bocat, P. Sassoulas, X. Federspiel, A. Cros, A. Bajolet, E. Richard, B. Dumont, P. Perreau, D. Petit, D. Golanski, C. Fenouillet-Beranger, N. Guillot, M. Rafik, V. Huard, S. Puget, X. Montagner, M.-A. Jaud, O. Rozeau, O. Saxod, F. Wacquant, F. Monsieur, D. Barge, L. Pinzelli, M. Mellier, F. Boeuf, F. Arnaud, and M. Haond, "28nm FDSOI technology platform for High Speed Low Voltage digital applications," in *Symposium on VLSI Technology Digest of Technical Papers*, pp. 133–134, 2012.
- [69] O. Weber, F. Andrieu, J. Mazurier, M. Casse, X. Garros, C. Leroux, F. Martin, P. Perreau, C. Fenouillet-Beranger, S. Barnola, R. Gassilloud,

- C. Arvet, O. Thomas, J.-P. Noel, O. Rozeau, M.-A. Jaud, T. Poiroux, D. Lafond, A. Toffoli, F. Allain, C. Tabone, L. Tosti, L. Brevard, P. Lehnen, U. Weber, P. Baumann, O. Boissiere, W. Schwarzenbach, K. Bourdelle, B.-Y. Nguyen, F. Breuf, T. Skotnicki, and O. Faynot, "Work-function engineering in gate-first technology for multi-VT dual-gate FDSOI CMOS on UTBOX," in *Electron Devices Meeting (IEDM), 2010 IEEE International*, pp. 3.4.1–3.4.4, 2010.
- [70] C. Fenouillet-Beranger, P. Perreau, T. Benoist, C. Richier, S. Haendler, J. Pradelle, J. Bustos, P. Brun, L. Tosti, O. Weber, F. Andrieu, B. Orlando, D. Pellisier-Tanon, F. Abbate, C. Richard, R. Beneyton, M. Gregoire, J. Ducote, P. Gouraud, A. Margain, C. Borowiak, R. Biancini, N. Planes, E. Gourvest, K. Bourdelle, B. Nguyen, T. Poiroux, T. Skotnicki, O. Faynot, and F. Boeuf, "Impact of local back biasing on performance in hybrid FDSOI/bulk high-k/metal gate low power LP technology," *Solid State Electronics*, vol. 88, pp. 15–20, 2013.
- [71] H. Shimada, Y. Hirano, T. Ushiki, K. Ino, and T. Ohmi, "Tantalum-gate thin-film SOI nMOS and pMOS for low-power applications," *IEEE Trans. Electron Devices*, vol. 44, pp. 1903–1907, 1997.
- [72] Q. Lu, Y. Yeo, P. Ranade, H. Takeuchi, T. King, C. Hu, S. Song, H. Luan, and D. L. Kwong, "Dual-metal gate technology for deep-submicron cmos transistors," in *Symposium on VLSI Technology Digest of Technical Papers*, pp. 72–73, IEEE, 2000.
- [73] L. Chang, S. Tang, T.-J. King, J. Boko, and C. Hu, "Gate length scaling and threshold voltage control of double-gate mosfets," in *IEDM, Technical Digest*, pp. 719 – 722, 2000.
- [74] A. Asenov, "Simulation of statistical variability in nano MOSFETs," in *Symposium on VLSI Technology Digest of Technical Papers*, pp. 86–87, 2007.
- [75] F. Boeuf, M. Sellier, A. Farcy, and T. Skotnicky, "An evaluation of the CMOS technology roadmap from the point of view of variability, interconnects, and power dissipation," *IEEE Transactions on Electron Devices*, vol. 55, pp. 1433–1440, 2008.
- [76] A. Cathignol, B. Cheng, D. Chanemougame, A. R. Brown, K. R. ang G. Ghibaudo, and A. Asenov, "Quantitative evaluation of statistical variability sources in a 45-nm technological node LP N-MOSFET," *IEEE Transactions on Electron Devices*, vol. 29, pp. 609–611, 2008.
- [77] A. Vandooren, A. Thean, Y. Du, I. To, J. Hughes, T. Stephens, M. Huang, S. Egley, M. Zavala, K. Sphabmixay, A. Barr, T. White, S. Samavedam,

- L. Mathew, J. Schaeffer, D. Triyoso, M. Rossow, D. Roan, D. Pham, R. Rai, B.-Y. Nguyen, B. White, M. Orlowski, A. Duvallet, T. Dao, and J. Mogab, "Mixed-signal performance of sub-100nm fully-depleted SOI devices with metal gate, high K (HfO₂/sub 2/) dielectric and elevated source/drain extensions," in *Electron Devices Meeting (IEDM), 2003 IEEE International*, pp. 11.5.1–11.5.3, 2003.
- [78] C. Fenouillet-Beranger, S. Denorme, B. Icard, F. Boeuf, J. Coignus, O. Faynot, L. Brevard, C. Buj, C. Soonekindt, J. Todeschini, J. Le-Denmat, N. Loubet, C. Gallon, P. Perreau, S. Manakli, B. Mmghetti, L. Pain, V. Arnal, A. Vandooren, D. Aime, L. Tosti, C. Savardi, F. Martin, T. Salvetat, S. Lhostis, C. Laviro, N. Auriac, T. Kormann, G. Chabanne, S. Gaillard, O. Belmont, E. Laffosse, D. Barge, A. Zauner, A. Tarnowka, K. Romanjec, H. Brut, A. Lagha, S. Bonnetier, F. Joly, N. Mayet, A. Cathignol, D. Galpin, D. Pop, R. Delsol, R. Pantel, F. Pionnier, G. Thomas, D. Bensahel, S. Deleombus, and H. M. T. Skotnicki, "Fully-depleted SOI technology using high-k and single-metal gate for 32 nm node LSTP applications featuring 0.179 μm^2 6T-SRAM bitcell," in *Electron Devices Meeting (IEDM), 2007 IEEE International*, pp. 267–270, 2007.
- [79] Y. Morita, R. Tsuchiya, T. Ishigaki, N. Sugii, T. Iwamatsu, T. Ipposhi, H. Oda, Y. Inoue, K. Torii, and S. Kimura, "Smallest V_{th} variability achieved by intrinsic silicon on thin BOX (SOTB) CMOS with single metal gate," in *Symposium on VLSI Technology Digest of Technical Papers*, pp. 166–167, 2008.
- [80] S. Markov, A. S. Mohd, B. Cheng, and A. Asenov, "Statistical variability in scaled generations of n-channel UTB-FD-SOI MOSFETs under the influence of RDF, LER, OTF and MGG," in *IEEE International SOI Conference*, pp. 1–2, 2012.
- [81] R. Smoluchowski, "Anisotropy of the electronic work function of metals," *Physical Review*, vol. 60, pp. 661–674, 1941.
- [82] H. F. Dadgour, K. Endo, V. K. De, and K. Banerjee, "Grain-orientation induced work function variation in nanoscale metal gate transistors- Part I: Modeling, Analysis, and Experimental validation," *IEEE Transactions on Electron Devices*, vol. 57, pp. 2504–2514, 2010.
- [83] S. C. Song, C. S. Park, J. Price, C. Burham, R. Choi, H. C. Wen, K. Choi, H. H. Tseng, B. H. Lee, and R. Jammy, "Modeling and analysis of grain-orientation effects in emerging metal-gate devices and implications for SRAM reliability," in *Electron Devices Meeting (IEDM), 2008 IEEE International*, pp. 1–4.

- [84] T. Matsukawa, Y. Liu, W. Mizubayashi, J. Tsukada, H. Yamauchi, K. Endo, Y. Ishikawa, S. O'uchi, H. Ota, S. Migita, Y. Morita, and M. Masahara, "Supressing V_t and G_m variability of FinFETs using amorphous metal gates for 14 nm and beyond," in *Electron Devices Meeting (IEDM), 2012 IEEE International*, pp. 8.2.1–8.2.4.
- [85] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*. Wile, 4th edition.
- [86] M. J. Deen and F. Pascal, "Electrical characterization of semiconductor materials and devices - review," *Journal of Materials Science : Materials in Electronics*, vol. 17, pp. 549–575, 2006.
- [87] M. Bhushan and M. Ketchen, *Microelectronic test structures for CMOS technology*. Springer, 2011.
- [88] D. Golanski, P. Fonteneau, C.Fenouillet-Beranger, A. Cros, F.Monsieur, N.Guitard, C.-A. Legrand, A.Dray, C.Richier, H.Beckrich, P.Mora, G.Bidal, O. Weber, O.Saxod, J.-R. Manouvrier, P.Galy, N. Planes, and F.Arnaud, "First demonstration of a full 28nm high-k/metal gate circuit transfer from Bulk to UTBB FDSOI technology through hybrid integration," in *2013 Symposium on VLSI Technology Digest of Technical Papers*, pp. T124–T125, 2013.
- [89] M. Charbonnier, C. Leroux, F. Allain, A. Toffoli, G. Ghibaudo, and G. Reimbold, "Automatic full quantum analysis of CV measurements for bulk and SOI devices," *Microelectronic Engineering*, vol. 88, pp. 3404–3406, 2011.
- [90] C. Leroux, F. Allain, A. Toffoli, G. Ghibaudo, and G. Reimbold, "Automatic statistical full quantum analysis of C-V and I-V charateristics for advanced mos gate stacks," *Microelectronic Engineering*, vol. 84, pp. 2408–2411, 2007.
- [91] J. Maserjian, G. Petersson, and C. Svensson, "Saturation capacitance of thin oxide MOS structures and the effective surface density of states of silicon," *Solid-Sate Electronics*, vol. 17, pp. 335–339, 1974.
- [92] C. Leroux, *Contribution à la caractérisation électrique et à l'analyse des dispositifs utilisés en microélectronique*. Habilitation à diriger des recherches, Institut National Polytechnique de Grenoble, 2010.
- [93] C. Leroux, G. Ghibaudo, and G. Reimbold, "Accurate determination of flat band voltage in advanced MOS structure," *Microelectronics Reliability*, vol. 47, pp. 660–664, 2007.
- [94] F. M. Smits, "Measurement of sheet resistivities with the four-point probe," *The Bell System Technical Journal*, vol. 37, pp. 711–718, 1958.

- [95] W. Wong-Ng, H. McMurdie, B. Paretzkin, C. Hubbard, and A. Dragoo, *The JCPDS database (1998), data set number: 38-1420*. Powder Diffraction 2, 200, 1987.
- [96] O. Engström, B. Raeissi, S. Hall, O. Bui, M. Lemme, H. Gottlob, P. Hurley, and K. Cherkaoui, "Navigation aids in the search for future high- κ dielectrics: Physical and electrical trends," *Solid-State Electronics*, vol. 51, pp. 622–626, 2007.
- [97] S. Guha, E. Cartier, M. A. Gribelyuk, N. A. Bojarczuk, and M. Copel, "Atomic beam deposition of lanthanum- and yttrium- based oxide thin films for gate dielectrics," *Applied Physics Letters*, vol. 77, pp. 2710–2712, 2000.
- [98] D. H. Triyoso, R. I. Hegde, J. M. Grant, J. K. Schaeffer, D. Roan, B. E. W. Jr., and P. J. Tobin, "Evaluation of lanthanum based gate dielectrics deposited by atomic layer deposition," *Journal of Vacuum Science and Technology B*, vol. 23, pp. 288–297, 2005.
- [99] D. J. Lichtenwalner, J. S. Jur, A. I. Kingon, M. P. Agusti, Y. Yang, S. Stemmer, L. V. Goncharova, T. Gustafsson, and E. Garfunkel, "Lanthanum silicate gate dielectric stacks with subnanometer equivalent oxide thickness utilizing an interfacial silica consumption reaction," *Journal of Applied Physics*, vol. 98, 2005.
- [100] M. Copel, E. Cartier, and F. M. Ross, "Formation of a stratified lanthanum silicate dielectric by reaction with si(001)," *Applied Physics Letters*, vol. 78, pp. 1607–1609, 2001.
- [101] H. Yamada, T. Shimizu, A. Kurokawa, K. Ishii, and E. Suzuki, "MOCVD of high-dielectric constant lanthanum oxide thin films," *Journal of the Electrochemical Society*, vol. 150, pp. G429–G435, 2003.
- [102] S. Stemmer, "Thermodynamic considerations in the stability of binary oxides for alternative gate dielectrics in complementary metal-oxide-semiconductors," *Journal of Vacuum Science and Technology B*, vol. 22, pp. 791–800, 2004.
- [103] S. Stemmer, "Interface and electrical properties of La-silicate for direct contact of high- κ with silicon," *Solid-State Electronics*, vol. 54, pp. 715–719, 2010.
- [104] R. Boujamaa, E. Martinez, F. Pierre, O. Renault, B. Detlefs, J. Zegenhagen, S. Baudot, M. Gros-Jean, F. Bertin, and C. Dubourdieu, "Study of the La-related dipole in TiN/LaOx/HfSiON/SiON/Si gate stacks using hard X-ray photoelectron spectroscopy and backside medium energy ion scattering," *Applied Surface Science*, vol. 335, pp. 71–77, 2015.

- [105] H. K. Kim, S. Y. Lee, I.-H. Yu, T. J. Park, R. Choi, and C. S. Hwang, "Gate engineering in TiN/La/TiN and TiLaN metal layers on atomic layer deposited HfO₂/Si," *IEEE Electron Device Letters*, vol. 33, pp. 955–957, 2012.
- [106] C. Suarez-Segovia, C. Leroux, F. Domengie, K. Dabertrand, V. Joseph, G. Romano, P. Caubet, S. Zoll, O. Weber, G. Ghibaudo, G. Reimbold, and M. Haond, "Effective work function engineering by sacrificial lanthanum diffusion on HfON-based 14 nm NFET devices," in *45th European Solid State Devices Conference (ESSDERC)*, pp. 246–249.
- [107] E. P. Gusev, M. Copel, E. Cartier, I. J. R. Baumvol, C. Krug, and M. A. Gribelyuk, "High-resolution depth profiling in ultrathin Al₂O₃ films on Si," *Applied Physics Letters*, vol. 76, pp. 176–178, 2000.
- [108] K. Choi, H.-C. Wen, G. Bersuker, R. Harris, and B. H. Lee, "Mechanism of flatband voltage roll-off studied with Al₂O₃ film deposited on terraced oxide," *Applied Physics Letters*, vol. 93, pp. 133506–1:4, 2008.
- [109] M. Charbonnier, C. Leroux, G. Reimbold, F. Allain, A. Toffoli, G. Ghibaudo, F. Martin, H. Crampeix, and F. Boulanger, "Investigation of mechanisms shifting metal effective workfunction towards P+ for various Al incorporation scenarios," in *2010, Symposium on VLSI Technology Systems and Applications (VLSI-TSA)*, pp. 84–85.
- [110] S. Baudot, C. Leroux, F. Chave, R. Boujamma, E. Martinez, P. Caubet, M. Silly, F. Sirotti, G. Reimbold, and G. Ghibaudo, "Understanding reversal effects of metallic aluminum introduced in HfSiON/TiN PMOSFETs," *Microelectronic Engineering*, vol. 88, pp. 1305–1308, 2011.
- [111] Y. W. Chen, C. M. Lai, T. F. Chiang, L. W. Cheng, C. H. Yu, C. H. Chou, C. H. Hsu, W. Y. Chang, T. B. Wu, and C. T. Lin, "Effective work function modulation by Aluminum Ion Implantation on Hf-based high- κ /metal gate pMOSFET," *IEEE, Electron Device Letters*, vol. 31, pp. 1290–1292, 2010.
- [112] R. Kakoschke, L. Pescini, J. R. Power, K. van der Zanden, E.-O. Andersen, Y. Gong, and R. Allinger, "Use of Al₂O₃ as inter-poly dielectric in a production proven 130 nm embedded Flash technology," *Solid State Electronics*, vol. 52, pp. 550–556, 2008.
- [113] W. Mueller, G. Aichmayr, W. Bergner, E. Erben, T. Hecht, and C. K. et al., "Challenges for the DRAM cell scaling to 40 nm," in *Electron Devices Meeting (IEDM), Tech Dig 2005*, pp. 347–350, 2005.
- [114] P. Youngwoo, C. Jungdal, K. Changseok, L. Changhyun, S. Yuchoel, and C. B. et al., "Highly manufacturable 32 Gb multi-level NAND Flash memory

- with $0.0098 \mu\text{m}^2$ cell size using TANOS cell technology,” in *Electron Devices Meeting (IEDM), Tech Dig 2006*, pp. 29–32, 2006.
- [115] L. Manchanda, W. H. Lee, J. E. Bower, F. H. Baumann, W. L. Brown, C. J. Case, R. C. Keller, Y. O. Kim, E. J. Laskowski, M. D. Morris, R. L. Opila, P. J. Silverman, T. W. Sorsch, and G. R. Weber, “Gate quality doped high- κ films for CMOS beyond 100 nm: 3–10 nm Al_2O_3 with low leakage and low interface states,” in *Electron Devices Meeting (IEDM), Tech Dig 1998*, pp. 605–608, 1998.
- [116] S. Guha, E. P. Gusev, H. Okorn-Schmidt, M. Copel, L.-A. Ragnarsson, N. A. Bojarczuk, and P. Ronsheim, “High temperature stability of Al_2O_3 dielectrics on Si: Interfacial metal diffusion and mobility degradation,” *Applied Physics Letters*, vol. 81, pp. 2956–2958, 2002.
- [117] H.-C. Wen, S. C. Song, C. S. Park, C. Burham, G. Bersuker, K. Choi, M. A. Quevedo-Lopez, B. S. Ju, H. N. Alshareef, H. Niimi, H. B. Park, P. S. Lysaght, P. Majhi, B. H. Lee, and R. Jammy, “Gate first Metal-Aluminum-Nitride PMOS electrodes for 32nm Low Standby Power Applications,” in *2007, Symposium on VLSI Technology*, pp. 160–161.
- [118] F. Panciera, S. Baudot, K. Hoummada, M. Gregoire, M. Juhel, and D. Mangelinck, “Three-dimensional distribution of Al in high- κ metal gate: Impact on transistor voltage threshold,” *Applied Physics Letters*, vol. 100, pp. 201909–1:4, 2012.
- [119] C. Suarez-Segovia, C. Leroux, P. Caubet, F. Domengie, G. Reimbold, G. Romano, O. Gourhant, V. Joseph, and G. Ghibaudo, “Effective work function modulation by sacrificial gate aluminum diffusion on HfON-based 14 nm NMOS devices,” *Microelectronic Engineering*, vol. 147, pp. 113–116, 2015.
- [120] C. Suarez-Segovia, P. Caubet, V. Joseph, O. Gourhant, G. Romano, F. Domengie, and G. Ghibaudo, “Effective work function shift induced by TiN sacrificial metal gates as a function of their thickness and composition in 14 nm NMOS devices,” in *Extended Abstracts of Solid State Devices and Materials (SSDM), 2014*, pp. 838–839.
- [121] F. Chave, L. Vallier, P. Gouraud, S. Baudot, P. Caubet, and O. Joubert, “Etching issues with high-K metal gate stacks for CMOS 28 nm technology,” in *Plasma Etch and Strip in Microelectronics, 4th International Workshop, Mechelen, Belgium, 2011*.
- [122] J.-M. Hwang and G. Pollack, “Novel polysilicon/TiN stacked-gate structure for fully depleted SOI CMOS,” in *Electron Devices Meeting (IEDM), Tech Dig 1992*, pp. 345–348, 1992.

- [123] K. T. Kim, L. G. Tang, T. S. Park, Y. S. Shin, J. K. Park, and C. J. Lee, "Tungsten silicide/titanium nitride compound gate for submicron CMOSFET," in *1990, Symposium on VLSI Technology*, pp. 115–116.
- [124] H. Kai, M. Xueli, Y. Hong, and W. Wenwu, "Modulation of the effective work function of TiN metal gate for PMOS application," *Journal of Semiconductors*, vol. 34, pp. 086002–1:4, 2013.
- [125] Z. Li, T. Schram, T. Witters, J. Tseng, S. D. Gendt, and K. D. Meyer, "Oxygen incorporation in TiN for metal gate work function tuning with a replacement gate integration approach," *Microelectronic Engineering*, vol. 87, pp. 1805–1807, 2010.
- [126] C. Choi, K.-L. Lee, and V. Narayanan, "Impact of diffusionless anneal using dynamic surface anneal on the electrical properties of a high- κ /metal gate stack in metal-oxide-semiconductor devices," *Applied Physics Letters*, vol. 98, pp. 123506–1:3, 2011.
- [127] S.-H. Lee, R. Choi, and C. Choi, "Effects of composition and thickness of TiN metal gate on the equivalent oxide thickness and flat band voltage in metal oxide semiconductor devices," *Microelectronic Engineering*, vol. 109, pp. 160–162, 2013.
- [128] J. Widiez, M. Vinet, T. Poiroux, P. Hollinger, B. Previtali, P. Grosgeorges, M. Mouis, and S. Deleonibus, "TiN metal gate thickness influence on Fully Depleted SOI MOSFET physical and electrical properties," in *2005 IEEE International SOI Conference*, pp. 30–31.
- [129] M. Kadoshima, T. Matsuki, S. Miyazaki, K. Shiraishi, T. Chikyo, K. Yamada, T. Aoyama, Y. Nara, and Y. Ohji, "Effective work function control by varying the TiN thickness in Poly-Si/TiN gate electrodes for scaled high- κ CMOSFETs," *IEEE Electron Device Letters*, vol. 30, pp. 466–468, 2009.
- [130] J. Westlinder, G. Sjoblom, and J. Olsson, "Variable work function in MOS capacitors utilizing nitrogen-controlled TiNx gate electrodes," *Microelectronic Engineering*, vol. 75, pp. 389–396, 2004.
- [131] J.-H. Ha, H. AlShareef, J. Chambers, Y. Sun, P. Pianetta, P. McIntyre, and L. Colombo, "Oxygen transfer from metal gate to high-k gate dielectric stack: Interface structure & property changes," *ECS Transactions*, vol. 11, pp. 213–218, 2007.
- [132] W. Mizubayashi, K. Akiyama, W. Wang, M. Ikeda, K. Iwamoto, Y. Kamimuta, A. Hirano, H. Ota, T. Nabatame, and A. Toriumi, "Novel V_{th} tuning process for HfO₂ CMOS with Oxygen-doped TaC_x," in *2008 Symposium on VLSI Technology Digest of Technical Papers*, pp. 42–43.

- [133] S. C. Song, C. S. Park, J. Price, C. Burham, R. Choi, H. C. Wen, K. Choi, H. H. Tseng, B. H. Lee, and R. Jammy, "Mechanism of Vfb roll-off with high work function metal gate and low temperature oxygen incorporation to achieve PMOS band edge work function," in *Electron Devices Meeting (IEDM), 2007 IEEE International*, pp. 337–340.
- [134] Z. Li, T. Schram, L. Pantisano, T. Conard, S. V. Elshocht, W. Deweerdt, S. D. Gendt, K. D. Meyer, A. Stesmans, S. Shamuilia, V. V. Afanas'ev, A. Akheyar, D. P. Brunco, N. Yamada, and P. Lehen, "Flatband voltage shift of ruthenium gated stacks and its link with the formation of a thin ruthenium oxide layer at the ruthenium/dielectric interface," *Journal of Applied Physics*, vol. 101, pp. 034503–1:9, 2007.
- [135] D. Jennings, A. Mayur, V. Parihar, L. Haifan, R. McIntosh, B. Adams, T. Thomas, J. Ranish, A. Hunter, T. Trowbridge, R. Achutharaman, and R. Thakur, "Dynamic surface anneal: activation without diffusion," in *IEEE International Conference on Advanced Thermal Processing on Semiconductors*, pp. 47–51.
- [136] K. Akiyama, W. Wang, W. Mizubayashi, K. M. A. Salam, M. Ikeda, H. Ota, T. Nabatame, and A. Toriumi, "Vfb roll-off of metal gate/HfO₂/SiO₂/Si capacitors in thinner EOT regime," in *Proc. Int. Workshop Dielectr. Thin Films, Kawasaki, Japan 2006*, pp. 63–64.
- [137] R. Singanamalla, H. Y. Yu, T. Janssens, S. Kubicek, and K. D. Meyer, "The study of effective work function modulation by As Ion Implantation in TiN/TaN/HfO₂ stacks," *Japanese Journal of Applied Physics*, vol. 46, pp. L320–L322, 2007.
- [138] B. Kim, Y. Ji, S. Lee, B. Jeon, K. Lee, K. Hong, and S. Park, "The study of flat-band voltage shift using Arsenic Ion-Implantation with high- κ /metal inserted Poly Si gate stacks," in *Solid-State Device Research European Conference, ESSDERC 2011*, pp. 79–82.
- [139] K. Ohmori, T. Matsuki, D. Ishikawa, T. Morooka, T. Aminaka, Y. Sugita, T. Chikyow, K. Shiraishi, Y. Nara, and K. Yamada, "Impact of additional factors in threshold voltage variability of metal/high-k gate stacks and its reduction by controlling crystalline structure and grain size in the metal gates," in *Electron Devices Meeting (IEDM), 2008 IEEE International*, pp. 71–74.
- [140] C. Adelmann, J. Meersschaut, L.-A. Ragnarsson, T. Conard, A. Franquet, N. Sengoku, Y. Okuno, P. Favia, H. Bender, C. Zhao, B. J. O'Sullivan, A. Rothschild, T. Schram, J. A. Kittl, S. V. Elshocht, S. D. Gendt, P. Lehen, O. Boissiere, and C. Lohe, "Thermally stable high effective work

- function TaCN thin films for metal gate electrode applications,” *Journal of Applied Physics*, vol. 105, pp. 053516:1–8, 2009.
- [141] F. Fillot, S. Maitrejean, F. Pierre, and B. Chenevier, “Work function tuning of $\text{Ti}_x\text{Si}_y\text{N}_z$ electrodes using partial saturation of chemisorbing surface during pulsing chemical vapor deposition,” *Electrochemical and Solid State Letters*, vol. 12, pp. H272–H274, 2009.
- [142] M. E. Grubbs, X. Zhang, M. Deal, Y. Nishi, and B. M. Clemens, “Development and characterization of high temperature stable Ta-W-Si-C amorphous metal gates,” *Applied Physics Letters*, vol. 97, pp. 223505:1–3, 2010.
- [143] J. A. Thornton, “Influence of apparatus geometry and deposition conditions on the structure and topography of thick sputtered coatings,” *Journal of Vacuum Science & Technology*, vol. 11, pp. 666–670, 1974.
- [144] J. Bareno, *In situ investigation of TiN surface dynamics: The role of surface and bulk mass transport processes*. PhD thesis, Graduate College at the University of Illinois at Urbana Champaign, 2007.

Résumé: Cette thèse porte sur l'élaboration et la caractérisation électrique et physico-chimique des grilles métalliques des dispositifs FDSOI MOSFET 14 nm à base d'oxyde high- κ , fabriqués chez STMicroelectronics. Ces grilles métalliques sont composées de couches de TiN, lanthane et aluminium, déposées par pulvérisation cathodique RF. Des structures de test et un schéma d'intégration simplifié permettant l'analyse capacitive ont été mis en place pour caractériser la modulation du travail de sortie effectif des grilles métalliques en TiN avec l'incorporation d'additifs tels que le lanthane ou l'aluminium. Ces additifs ont été incorporés suivant une approche de grille sacrificielle. Par ailleurs, une méthodologie inédite basée sur la fluorescence X a été proposée et validée pour la caractérisation précise en ligne de la diffusion des additifs. Cette méthodologie permet de prouver que la dose effective de l'espèce incorporée après recuit de diffusion peut être modélisée en fonction de l'épaisseur du TiN piédestal dans la grille sacrificielle ainsi que de la température de recuit. De plus, la variation de l'épaisseur de l'oxyde interfaciel sur une seule plaquette (oxyde biseau) autorise l'identification de l'origine physique de la modulation du travail de sortie effectif, qui s'explique par un dipôle qui évolue avec la dose effective de l'espèce incorporée. En conséquence, des équations empiriques de la diffusion des dopants dans l'oxyde de grille et de leur impact sur le travail de sortie effectif des grilles métalliques ont été proposées afin de moduler avec précision la tension de seuil (V_{TH}) des dispositifs FDSOI 14 nm. En outre, l'impact de l'oxyde high- κ à la fois sur la diffusion des additifs et sur la modulation du travail de sortie effectif a été mis en évidence. Enfin, un procédé innovant de dépôt métallique, permettant la modification de la microstructure du TiN, a été développé afin de réduire davantage la variabilité locale de la tension de seuil V_{TH} des dispositifs FDSOI.

Mots-clés: grille métallique, travail de sortie, diffusion de dopants, lanthanum, aluminium, spectroscopie de rayons X, variabilité locale du V_{TH} , pulvérisation cathodique RF

Abstract: This Ph.D. thesis is focused on the fabrication and the electrical and physicochemical characterization of metal gates in 14 nm high- κ based FDSOI MOSFET devices, manufactured at STMicroelectronics. These metal gates are composed of TiN, lanthanum and aluminum layers, deposited by RF sputtering. Test structures and a simplified integration scheme allowing C-V measurements, have been implemented in order to characterize the modulation of the effective work function of TiN metal gates with the incorporation of dopants such as lanthanum or aluminum. These additives are incorporated in a sacrificial gate-first approach. Furthermore, a new methodology based on X-ray fluorescence was proposed and validated for accurate in-line characterization of the diffusion of dopants. This methodology enables to prove that the effective dose of the species incorporated into dielectrics after diffusion annealing may be modeled as a function of the thickness of the pedestal TiN in the sacrificial gate and the annealing temperature. Moreover, the variation of the thickness of the interfacial oxide along the wafer (bevel oxide) authorizes the identification of the origin of the modulation of the effective work function, which is explained by a dipole that evolves with the effective dose of the incorporated dopant. Accordingly, empirical equations of the diffusion of dopants into the gate dielectrics and their impact on the effective work function of metal gates have been proposed to precisely modulate the threshold voltage (V_{TH}) of the 14 nm FDSOI devices. In addition, the influence of the high- κ oxide on both the diffusion of dopants and the modulation of the effective work function was highlighted. Lastly, an innovative process for metal deposition, allowing the modification of the microstructure of TiN, was developed in order to further reduce the local V_{TH} variability in FDSOI devices.

Keywords: metal gate, work function, dopants diffusion, lanthanum, aluminum, X-Ray spectroscopy, local V_{TH} variability, RF-PVD sputtering.

Résumé

Cette thèse aborde la fabrication et la caractérisation électrique et physico-chimique des grilles métalliques des dispositifs FDSOI 14 nm à base d'oxyde high-k, fabriqués chez STMicroelectronics. Ces grilles métalliques sont constituées des couches de nitrure de titane (TiN), de lanthane et d'aluminium, déposées par pulvérisation cathodique assistée par plasma radiofréquence (RF-PVD). Les principaux objectifs de cette thèse étaient d'une part la compréhension de la modulation du travail de sortie effectif grâce à l'incorporation d'additifs tels que l'aluminium, le lanthane et l'azote dans une approche dite de grille sacrificielle, et d'autre part, l'étude de l'impact de la microstructure du TiN sur la variabilité locale des dispositifs MOSFET appariés. Le but était de valider l'intégration grille sacrificielle pour atteindre les exigences de tension de seuil (V_{TH}) de la technologie 14FDSOI et de proposer des conditions de dépôt métallique pour réduire davantage la variabilité locale des dispositifs MOSFET pour les applications FDSOI avancées.

Dans le premier chapitre, nous avons rappelé la définition des principales propriétés de l'empilement métal-oxyde-semiconducteur. Nous avons vu que la plupart des paramètres de performance des dispositifs MOSFET sont fortement influencés par l'épaisseur d'oxyde équivalente (EOT) et le travail de sortie effectif (WF_{eff}). Nous avons identifié l'impact électrique de la présence de dipôles à l'interface high-k/SiON et de la densité de charges fixes interfaciales dans les diélectriques sur ce travail de sortie effectif. Nous avons montré que la plage des valeurs du travail de sortie effectif pour passer d'une grille N à une grille P dans des dispositifs FDSOI est plus étroite que dans les dispositifs *bulk* en raison de la faible quantité de charges de déplétion dans la couche de silicium sur l'oxyde enterré. L'intégration grille sacrificielle a été décrite comme un schéma de procédé permettant d'atteindre les spécifications de tension de seuil à la fois sur NFETs et PFETs par l'incorporation minutieuse d'additifs tels que le lanthane et l'aluminium. Nous avons profité de l'intégration grille sacrificielle pour découpler les principaux objectifs de cette thèse. D'une part, c'est principalement avec la grille sacrificielle que nous avons cherché à moduler la tension de seuil. D'autre part, c'est principalement avec la grille finale que nous avons cherché à réduire la variabilité locale de la tension de seuil. Enfin, nous avons supposé que la variabilité locale de la tension seuil due à la granularité de la grille métallique est liée à une valeur différente du travail de sortie (Φ_m) associée à chaque orientation cristalline. Dans le cas du TiN, une orientation (200) est associée à une valeur Φ_m de 4,6 eV et une orientation (111) est associée à une valeur Φ_m de 4,4 eV. Pour cette raison, trois solutions techniques ont été proposées dans le but de supprimer ou de réduire la variabilité du travail de sortie: 1) le dépôt d'une grille métallique amorphe, 2) le dépôt d'une grille métallique monocristalline, et 3) le dépôt d'une grille métallique avec des plus petites cristallites afin de réduire la variabilité par effet statistique.

Dans le deuxième chapitre, nous avons décrit les structures et les techniques de caractérisation électrique et physico-chimiques mises en œuvre dans cette thèse pour la caractérisation de l'empilement de grille métallique. Un schéma d'intégration simplifié de la technologie FDSOI 14 nm a été validé afin d'étudier intensivement l'impact des procédés de dépôt de la grille sacrificielle et de l'incorporation d'additifs tels que le lanthane ou l'aluminium sur le travail de sortie effectif (WF_{eff}) et l'épaisseur d'oxyde équivalente (EOT) de l'empilement de grille. Nous avons introduit notre méthodologie d'extraction de ces paramètres électriques à partir des caractéristiques expérimentales C-V. Le WF_{eff} et l'EOT ont été extraites par ajustement des simulations quantiques Poisson-Schrödinger avec une partie pertinente de la courbe CV à haute fréquence, libre de fuites et de défauts d'interface. Nous avons ensuite introduit notre approche pour discriminer l'influence d'un dipôle de l'influence des charges fixes sur la valeur du décalage du travail de sortie effectif. En effet, l'intégration d'un oxyde biseauté dans ce schéma d'intégration simplifié autorise l'identification de l'origine de la modulation du travail de sortie effectif. Ainsi, lorsque deux plaquettes différentes sont comparées, chacune avec un procédé ou dopant spécifique (La, Al, N) dans l'empilement de grille sacrificielle, l'origine du décalage de travail de sortie effectif peut être identifiée à partir des graphes WF_{eff} vs EOT de chaque plaquette. Si l'ordonnée à l'origine (WF_{eff} à EOT = 0) des graphes WF_{eff} vs EOT est inchangée, mais les pentes des 2 courbes sont différentes, alors le décalage du WF_{eff} à EOT nominal sera expliqué par des charges fixes interfaciales induites par le procédé ou dopant spécifique dans l'empilement de grille sacrificielle. En revanche, si l'ordonnée à l'origine est modifiée, mais les pentes restent inchangées, le décalage du WF_{eff} à EOT nominal sera expliqué par un effet de dipôle ou par une modification du travail de sortie intrinsèque de la grille finale en TiN. Par ailleurs, une méthodologie inédite basée sur la fluorescence X a été proposée et validée pour la caractérisation précise en ligne de la diffusion des additifs dans l'oxyde high-k après recuit. Cette méthodologie consiste à mesurer sur des pleines plaques l'intensité de fluorescence des additifs après le dépôt de la grille sacrificielle et après le retrait de la grille sacrificielle. Les deux mesures sont nécessaires afin de déterminer le pourcentage de diffusion pour un élément donné. Enfin, nous avons décrit deux configurations de diffraction X dites « Θ -2 Θ » et « in-plane » afin d'étudier l'orientation préférentielle et la taille horizontale des cristallites de TiN en fonction des conditions de dépôt métallique.

Dans le troisième chapitre, nous démontrons que le lanthane et l'aluminium sont incorporés avec succès dans les diélectriques après un recuit de diffusion effectué à 900 °C pendant 10 secondes. De plus, l'ingénierie du travail de sortie effectif est démontrée avec la modulation de l'épaisseur du TiN et celle des additifs sans dégradation de fuite de grille. La modulation du travail de sortie effectif vers les valeurs N+ avec le lanthane et vers les valeurs P+ avec l'aluminium est efficacement modérée par la couche de TiN piédestal, qui joue le rôle de barrière de diffusion, ce qui rend cette approche appropriée pour répondre aux exigences de tension de seuil des dispositifs FDSOI 14 nm. En outre, il est démontré que l'approche de grille sacrificielle permet de prédire et contrôler précisément la diffusion des additifs vers l'empilement HfON/SiON. En effet, la méthodologie

basée sur la fluorescence X a été utilisée avec succès pour modéliser la dose effective d'additif incorporée dans l'empilement HfON / SiON en fonction de la dose déposée, de l'épaisseur du TiN piédestal et de la température de recuit de la grille sacrificielle. A partir de cette étude, nous avons identifié que le pourcentage de diffusion des deux additifs diminue exponentiellement lorsque l'épaisseur du TiN piédestal augmente. De plus, nous avons également identifié que le TiN piédestal est plus efficace pour écranter la diffusion du lanthane que celle de l'aluminium. La capacité du TiN pour écranter la diffusion augmente lorsque la température de recuit diminue. Dans le cas du lanthane, la dose effective augmente de manière linéaire avec la dose déposée. En revanche, dans le cas de l'aluminium, la diffusion est arrêtée pour des doses déposées plus élevées. Une température plus élevée serait donc nécessaire pour augmenter la diffusion de l'aluminium. En outre, les dispositifs avec oxyde biseauté nous permettent d'identifier l'origine du décalage du travail de sortie effectif à EOT nominal. Ce décalage a été attribué à la chute de tension provoquée par une couche dipolaire à l'interface high-k/SiO_x. Cette chute de tension liée à une couche dipolaire évolue avec la dose effective de l'additif incorporé, vers les valeurs N⁺ avec l'incorporation de lanthane et vers les valeurs P⁺ avec l'incorporation d'aluminium. De plus, la valeur de cette chute de tension est considérablement impactée par l'influence d'un phénomène de roll-off pour les plus petites EOT. Ce composant de roll-off décale le travail de sortie effectif vers les valeurs N⁺ et évolue à la fois avec la dose effective de l'additif incorporé et avec l'épaisseur de SiO_x.

En conséquence, des équations empiriques de la diffusion des additifs dans les diélectriques de grille et de leur impact sur le travail de sortie effectif des grilles métalliques ont été proposées afin de moduler précisément la tension de seuil à EOT nominal des dispositifs FDSOI 14 nm. En tenant compte de la chute de tension induite par le roll-off à EOT nominal de la technologie FDSOI 14 nm, le lanthane décale le W_{eff} d'environ -80 meV pour chaque dose effective de lanthane de 1×10^{14} at/ cm² alors que l'aluminium décale le W_{eff} d'environ +25 meV pour chaque dose effective d'aluminium de 1×10^{14} at/ cm². Evidemment, le phénomène de roll-off se traduit par une maximisation de la force de la chute de potentiel provoquée par l'incorporation de lanthane et par une diminution de la force de la chute de potentiel provoquée par l'incorporation d'aluminium. Par souci de comparaison, nous avons supprimé cette chute de potentiel induite par le roll-off pour déterminer la force du dipôle pour les doses effectives d'aluminium et de lanthane incorporées dans l'empilement HfON/SiON. Cela correspond à une modification du W_{eff} à EOT = 0 de +41 meV pour chaque dose effective d'aluminium de 1×10^{14} at/ cm² et à une modification du W_{eff} à EOT = 0 de -53 meV pour chaque dose effective de lanthane de 1×10^{14} at/ cm² incorporée dans l'empilement HfON/SiON. De plus, nous avons étudié l'influence de l'oxyde high-k à la fois sur la diffusion des additifs et sur la modulation du travail de sortie effectif. L'aluminium et le lanthane diffusent plus facilement dans le HfON que dans le HfSiON après un recuit à 900 °C pendant 10 secondes. Cette différence pourrait expliquer une modulation plus significative du travail de sortie effectif sur les dispositifs à base de HfON pour une dose déposée équivalente d'aluminium et de lanthane.

Étant donné que les épaisseurs d'oxyde équivalentes (EOT) étaient différentes pour les dispositifs avec HfON et HfSiON, la comparaison du WF_{eff} à EOT nominal n'est pas autorisée. Nous avons donc comparé le WF_{eff} à EOT = 0 pour des doses équivalentes incorporées à la fois dans l'empilement HfON/SiON et dans l'empilement HfSiON/SiON. Cela correspond à un décalage de WF_{eff} à EOT = 0 de +36 meV pour chaque dose effective d'aluminium de 1×10^{14} at/ cm² incorporée dans l'empilement HfSiON/SiON et à un décalage de WF_{eff} à EOT = 0 de +41 meV pour chaque dose effective d'aluminium de 1×10^{14} at/ cm² incorporée dans l'empilement HfON/SiON. D'un autre côté, nous avons trouvé un décalage de WF_{eff} à EOT = 0 de -54 meV pour chaque dose effective de lanthane de 1×10^{14} at/ cm² incorporée dans l'empilement HfSiON/SiON et un décalage de WF_{eff} à EOT = 0 de -53 meV pour chaque dose effective de lanthane de 1×10^{14} at/ cm² incorporée dans l'empilement HfON/SiON. Compte tenu de cette faible différence, il est donc logique de conclure que le ΔWF_{eff} à EOT = 0 est proportionnel à la dose effective incorporée dans l'empilement high-k/SiO_x. Il est plus important sur les dispositifs à base de HfON par rapport à ceux à base de HfSiON parce que la dose effective incorporée est proportionnellement plus importante sur HfON que sur HfSiON pour des doses déposées équivalentes.

Dans le quatrième chapitre, nous avons étudié l'impact de l'azote, ajusté par l'épaisseur et la composition de la couche de nitrure de titane (TiN), sur le travail de sortie effectif des dispositifs FDSOI 14 nm. Il s'est avéré que le TiN de la grille sacrificielle n'est plus efficace pour moduler le travail de sortie effectif sur les dispositifs à base de HfON. Les mesures d'intensités de fluorescence ont montré qu'en réalité l'azote n'est pas incorporé dans les diélectriques après un recuit de diffusion à 900 °C pendant 10 secondes. Le niveau d'azote dans l'empilement high-k/SiON est en fait inchangé, indépendamment de l'épaisseur du TiN et du type de diélectrique : HfON et HfSiON. Par ailleurs, les couches de TiN se sont avérées plus efficaces pour incrémenter le WF_{eff} à EOT = 0 en augmentant l'épaisseur du TiN dans une intégration finale standard. L'origine de ce décalage est incertaine et elle peut être liée soit à une variation de la microstructure du TiN soit à une oxydation importante des couches de TiN plus épaisses. L'activation de la diffusion de l'azote lors du recuit d'activation des dopants à 1005 °C ne devrait pas être exclue pour expliquer l'augmentation du WF_{eff} avec des couches de TiN final plus épaisses. Néanmoins, un fort effet de roll-off est mis en évidence pour les électrodes Poly-Si/TiN plus épaisses à EOT < 20Å grâce aux dispositifs avec oxyde biseauté. En conséquence, le phénomène de roll-off rend aussi la grille finale en TiN inutile pour la modulation de la tension de seuil à EOT nominal dans des dispositifs FDSOI 14 nm. En conclusion, nous recommandons très fortement l'utilisation de l'aluminium dans une intégration grille sacrificielle pour concevoir des grilles P+ afin de répondre aux exigences de l'offre multi- V_{TH} de la technologie FDSOI 14 nm, non seulement en raison du contrôle précis de la dose effective incorporée dans l'empilement high-k/SiON, mais aussi en raison du plus faible effet de roll-off à EOT nominal, par rapport au TiN.

Enfin, nous avons proposé un procédé innovant de dépôt métallique par RF-PVD afin de modifier la microstructure des couches de TiN. Nous avons étudié l'impact des conditions de pression et de puissance radiofréquence lors du dépôt de TiN par RF-PVD à température ambiante sur l'orientation préférentielle et sur la taille moyenne des grains de TiN. La réduction de la taille moyenne des grains de TiN au centre de la plaque a été achevée soit en augmentant la pression soit en diminuant la puissance RF, comme démontré expérimentalement par des mesures de diffraction de rayons X dans la configuration dite « in-plane ». En particulier, des grilles finales en TiN déposées à une pression plus importante ont été proposées afin de réduire davantage la variabilité locale de la tension de seuil dans les dispositifs FDSOI. Par ailleurs, l'implémentation de ces grilles métalliques « haute pression » sur des dispositifs FDSOI 28 nm ont montré une amélioration de la variabilité locale de la tension de seuil. Néanmoins, il est surprenant que le bénéfice observé soit plus important sur les dispositifs FDSOI dont la surface est plus importante. En effet, nous nous attendons à ce que la plus forte réduction de l'écart type par effet statistique soit observée pour les plus petits transistors et que ce bénéfice sature après un certain nombre de grains. En conséquence, la diminution de la variabilité locale de la tension seuil des dispositifs FDSOI observée pour les grilles TiN déposées à haute pression ne peut être que partiellement expliquée par la réduction de la taille moyenne des cristallites de TiN. En particulier, la diminution de la variabilité locale de la tension de seuil dans les dispositifs FDSOI à grande surface doit être étudiée dans un travail futur.