



HAL
open science

Study, fabrication and characterization in view of 3D Sequential Integration

Camila Toledo de Carvalho Cavalcante

► **To cite this version:**

Camila Toledo de Carvalho Cavalcante. Study, fabrication and characterization in view of 3D Sequential Integration. Micro and nanotechnologies/Microelectronics. Université Grenoble Alpes [2020-..], 2021. English. NNT: 2021GRALT050 . tel-04884409

HAL Id: tel-04884409

<https://theses.hal.science/tel-04884409v1>

Submitted on 13 Jan 2025

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 L'UNIVERSITE GRENOBLE ALPES

Spécialité : NANO ELECTRONIQUE ET NANO TECHNOLOGIES

Arrêté ministériel : 25 mai 2016

Présentée par

**Camila TOLEDO DE CARVALHO
CAVALCANTE**

Thèse dirigée par **Gerard GHIBAUDO**, Directeur de Recherche,
Université Grenoble Alpes

et codirigée par **Claire FENOUILLET-BERANGER**
et **Perrine BATUDE**

préparée au sein du **Laboratoire CEA/LETI**
dans **l'École Doctorale Electronique, Electrotechnique,
Automatique, Traitement du Signal (EEATS)**

**Étude, fabrication et caractérisation
pour
l'intégration 3D séquentielle**

**Study, fabrication and
characterization in
view of 3D Sequential Integration**

Thèse soutenue publiquement le **26 mai 2021**,
devant le jury composé de :

Monsieur Emmanuel DUBOIS

DIRECTEUR DE RECHERCHE, CNRS DELEGATION HAUTS-DE-
FRANCE (Rapporteur)

Monsieur Brice GAUTIER

PROFESSEUR DES UNIVERSITES, INSTITUT NATIONAL SC.
APPLIQUEES DE LYON, (Rapporteur)

Monsieur Emmanuel JOSSE

INGENIEUR DOCTEUR, ST MICROELECTRONICS, (Examineur)

Monsieur Francis BALESTRA

DIRECTEUR DE RECHERCHE, CNRS DELEGATION ALPES,
(Président)

Monsieur Gerard GHIBAUDO

DIRECTEUR DE RECHERCHE, CNRS DELEGATION ALPES,
(Directeur de thèse)



Acknowledgements

I would like to thank the members of my jury, Dr. Dubois, Dr. Gautier, Dr. Josse and Dr. Balestra, for reading my manuscript, attending the thesis defense and for your valuable feedback. It was an honor having you in my jury.

Now moving to my supervising team: many thanks to my thesis director Gerard. You taught me so much, always available for meeting and discussing. I always felt welcomed at imep, and you made me feel comfortable to ask even the most basic questions. You are so nice, it was a pleasure working with you. My other thesis director Claire, my reference for anything related to thermal stability. Many thanks for everything you taught me about integration and especially for reading my manuscript (some chapters even more than once). Your feedback was critical for improving and making it clearer. Also, many thanks of course to my closest supervisor, Perrine. You always pushed me, and because of that you made me improve a lot. I really appreciated working together, your explanations with drawings, answering the same thing at least 3 times because I forgot...you know how much you helped me and I am very happy to have done my thesis with you. In addition, I also had the luck to have unofficial supervisors to help me too. Xavier, I am sure we had no idea I would work so much on the gate stack, and this was only possible thanks to you, who took the time to show and teach me, thank you for everything. Joris, the TCAD and equations specialist. Thank you for all your help, either debugging the deck, Vth calculation..I will miss stopping by in your office on my way to charac to chitchat. Overall, this goes for you all, I was very lucky to work with you, I am grateful that I had such a good and supporting supervising team. I hope we will meet again soon!

People from CEA: (LICL - Jean-Michel, Bernard, Christophe, Valerie, Zdeneck, Cyrile, Nils, Manu, Laurent Brunet, Laurent Brevard, Maud, François, Louis, Alex, Hervé with his lot maudit), (simulation - Benoit, Olivier), (charac - Jacques, Alain, Fabienne, Giovanni, Rabah), (BEOL - Stephane, Chloe), (Leti - Remi, Sebastien Kerdilles, Pablo, Anne-Sophie), (phd's and interns - Rony, Lucid, Tarcisius, Yoonji, Mariia, Theo, Sylvain, Noemie) - Thank you for everything, for your help with characterizing my lot, answering my questions, for our talks...it was nice meeting you all and I wish you all the best (sorry if I forgot someone).

Still on people from CEA: Mathieu (my first supervisor), I came to CEA thanks to the internship with you. I hope we will meet soon and play some board games together. Jean-Pierre, I cannot even; thank you for everything, I hope that you are enjoying your retirement and seeing hammerhead sharks in your dives.

My bureau peeps, you made my days in the office: Giulia you were the first to leave, but I am happy we got to meet and that we keep in touch. Jessy, we had good happy hours together discussing about phd life. Claaaaaude, thanks for your help with the SIAM, the nespresso capsules, for explaining the taxes..I hope you are enjoying the retirement. Lina we had so many lunches, shopping, starbucks in moments of crisis, baking, voice messages...Daphne, we started together, I have great memories of our talks, lunches, bars, games...Mickael, I hope that my photo is still hanging besides your computer! Also great times together, the office would not have been the same without you. I miss you and want to see you all soon!

My other peeps: (Amanda my buddy from Brazil, it was great meeting you, courage for your defense), (Larry, Samer, Johana, Alberto and Nassim, from NSE to PhD, thanks for the good times together).

Lastly, my friends and family from Brazil, thank you for your moral support, love you all. To my best buddy / fiancé Alex: thank you for being there for me, love you.

Oh man, can't believe it is over! It was not always easy but it was worth it, I am writing this full of good memories and gratitude.

Ce qui est est.

TABLE OF CONTENTS

Acknowledgements	2
Table of contents	2
Glossary	8
Introduction	13
1.1 Moore's Law	16
1.1.1 <i>Physical limits to scaling</i>	16
1.2 Scaling Alternatives	17
1.2.1 <i>Silicon germanium channel</i>	17
1.2.2 <i>High-k materials</i>	18
1.2.3 <i>New MOSFET architectures</i>	18
1.3 Introduction to 3D Sequential Integration	20
1.4 More Moore by device stacking.....	22
1.4.1 <i>Technological implications: bottom tier stability</i>	24
1.4.1.1 Impact on device performance (FEOL)	25
1.4.1.2 Impact on the interconnects (BEOL)	26
1.4.2 <i>Low temperature process flow</i>	27
1.5 More than Moore by device stacking.....	28
1.5.1.1 Computing and data storage.....	30
1.5.1.2 Image Sensors	32
1.5.1.3 This work: Smart sensor interfaces.....	33
1.6 High voltage Devices	34
1.6.1 <i>Requirements</i>	34
1.6.2 <i>FOM for analog circuits</i>	35
1.6.3 <i>Proposed process flow</i>	37
Bottom tier thermal stability	44
2.1 Bottom tier Stability: What has been done	45
2.1.1 <i>FEOL stability under standard anneal</i>	46
2.1.1.1 Motivations for this work.....	47

2.1.2	<i>Intermediate BEOL stability under standard anneal</i>	48
2.1.2.1	Motivations for this work.....	53
2.1.3	<i>FEOL and iBEOL stability under ultra-short anneal</i>	54
2.1.3.1	Motivations for this work.....	56
2.2	FEOL stability under standard anneal for More than Moore applications	57
2.2.1	<i>Gate stack stability</i>	58
2.2.2	<i>Silicide Stability</i>	65
2.2.2.1	Suggested improvements	70
2.2.3	<i>Global performance stability</i>	70
2.3	FEOL stability under standard anneal for More Moore applications	72
2.3.1	<i>Static FOM</i>	73
2.3.2	<i>Dynamic FOM</i>	75
2.3.3	<i>Yield</i>	78
2.4	iBEOL stability under standard anneal for More Moore applications	80
2.4.1	<i>Robustness Characterization</i>	80
2.4.2	<i>Yield</i>	82
2.5	FEOL and BEOL stability under ultra-short anneals for More Moore applications	83
2.6	Conclusion	86
	Low temperature gate stack	91
3.1	Challenges of Low Temperature Gate Stack	92
3.1.1	<i>Challenges for analog stack</i>	95
3.2	Overview of microscopic oxide defects	97
3.2.1	<i>Interface traps</i>	98
3.2.2	<i>Border traps</i>	101
3.2.3	<i>Deep traps</i>	101
3.3	Methods to Improve the Oxide Quality	101
3.4	Oxide Splits Definition	103
3.4.1	<i>Low temperature oxides</i>	103
3.4.2	<i>OX from BOX</i>	105
3.4.3	<i>Choice of TiN</i>	107

3.5	Electrical Results	107
3.5.1	<i>NMOS Evaluation</i>	113
3.5.2	<i>PMOS Evaluation</i>	114
3.5.3	<i>Trapping Mechanisms in LT Oxides</i>	115
3.5.3.1	NBTI Discussion	116
3.6	Electrical Results of OX from BOX	117
3.6.1	<i>NMOS Discussion</i>	117
3.6.2	<i>PMOS Discussion</i>	118
3.6.3	<i>Trapping Mechanisms on OX from BOX</i>	119
3.6.4	<i>Fourier-transform Infrared Spectroscopy</i>	119
3.7	Impact of LT Gate Stack on Analog FOM	122
3.8	Gate Stack Optimizations Guidelines	124
3.8.1	<i>Thermal Treatments</i>	124
3.8.2	<i>Atomic Hydrogen Treatment</i>	125
3.8.3	<i>SiGe channel</i>	126
3.8.4	<i>Junction-less transistors</i>	128
3.8.4.1	Gate Stack Poly	132
3.9	Hot Carrier Injection	140
3.9.1	<i>HCI Results</i>	140
3.10	Conclusions of the Chapter	142
	Low temperature junction	147
4.1	Dopant implantation	148
4.2	Dopant activation	148
4.3	End of Range defects	151
4.4	Dissolution of End of Range defects	153
4.5	Low temperature junction for high voltage devices ($V_{DD}=2.5V$)	154
4.5.1	<i>Process flow description</i>	155
4.5.2	<i>LDD implants</i>	157
4.5.3	<i>HDD implants</i>	159
4.5.4	<i>SPER activation temperature</i>	161

4.6	Electrical results @500°C 30min	161
4.6.1	<i>Understanding the I_{ON} results</i>	162
4.6.2	<i>Access resistance definition</i>	167
4.6.2.1	Access resistance extraction.....	168
4.6.2.2	Discussion on junction activation mechanisms	170
4.6.2.3	Discussion on chemical and electrical profile	172
4.6.2.4	Access resistance evaluation	178
4.6.3	<i>Understanding I_{OFF} Results</i>	179
4.7	Optimization Guidelines.....	181
4.8	Conclusion of this Chapter	183
	Conclusion and Perspectives	186
	List of Publications	190
	Abstract	191
	Résumé	192

GLOSSARY

ϕ_f : fermi potential

ϕ_M : metal workfunction

ϕ_S : semiconductor workfunction

μ_0 : low field mobility

1/f : flicker noise

3DSI : 3D sequential integration

a/c : amorphous/crystalline

ALD : Atomic Layer Deposition

AR : aspect ratio

ATR : attenuated total reflectance

BBT : Band to Band Tunneling

BEOL : back end of line

BOX : buried oxide

BSI : Back-side Illumination

CD : critical dimension

C_d : depletion capacitance

CDE : cyclic deposition/etch

CET : Capture and Emission Time

CIS : CMOS image sensors

CLM : channel length modulation

CMF : correlated mobility fluctuation

CMP : chemical mechanical polishing

CNF : carrier number fluctuation

CNTFETs : carbon nanotube field effect transistors

C_{OX} : gate oxide capacitance

C_{ox} oxide capacitance

C_{SOL} : solid solubility limit

Cu : Copper

CVS : Constant Voltage Stress

DIBL : drain induced barrier lowering

D_{it} : density of interface states

DSA : Dynamic Surface Anneal

E_a : activation energy

EDX : Energy dispersive x-ray spectroscopy

EOR : End of Range defects

EOT : effective oxide thickness

FEOL : front end of line

FGA : forming gas anneal

FOM : figure of merit

FSI : Front-side Illumination

FTIR : fourrier transform infrared spectroscopy

GAA : gate-all-around

g_D : output drain conductance

g_m : transconductance

GP : ground plane

HCI : hot carrier injection

HDD : high doped drain

HfO_2 : hafnium oxide

HTO : high temperature oxide

HV : high voltage

i.e. : that is

iBEOL : intermediate back end of line

IC : integrated circuit

IMD : intermetal dielectric

I_{STAT} : static current

JL : junction-less

KMC : Kinectic Monte Carlo

kT : thermal energy

LASSE : Laser Systems & Solutions of Europe

LDD : lower doped drain

LFN : low frequency noise

LO : optical longitudinal mode

M1 : first level metal

M4 : forth level metal

MOSFET : metal oxide semiconductor field effect transistor

ms : milisecond

MtM : More than Moore

n : body factor

N_a : substrate doping

$N_{a,poly}$: acceptor density of p-type polysilicon

NBTI : Negative Bias Temperature Instability

$N_{d,poly}$: donnor density of n-type polysilicon

NiPt : nickel platinum

NiSi : nickel silicide

N_t : oxide trap volumetric density

PDYN : dynamic power

PSTAT : static power

PAI : Pre-amorphization implantation

PBTI : Positive Bias Temperature Instability

PECVD : plasma-enhanced chemical vapor deposition

PMD : pre metal deposition

POR : Process of Reference

Pt : Platinum

PVT : Process-Voltage-Temperature

Q_{dep} : depletion layer charge

QW : quantum well

r_o : output resistance

R_{ch} : channel resistance

R_{con} : contact resistance

R_{HDD} : HDD resistance

R_{LDD} : LDD resistance

RMG : replacement metal gate

R_s : sheet resistance

R_{SD} : access resistance

R_{sil} : silicided region resistance

$R_{sil/dop}$: interface resistance between silicided region and doped region

R_{SPA} : spacer region resistance

R_{TOT} : total resistance

S/D : source and drain

SCE : short channel effect

SEM : scanning electron microscope

SiGe : silicon-germanium

SiN : silicon nitride

SiO₂ : silicon dioxide

SoC : System-on-Chip

SOI : silicon-on-insulator

SOTA : State-of-the-Art

SPER : Solid-Phase Epitaxial Regrowth

SRAM : static random access memory

SRH : Shockley Read Hall

SS : subthreshold slope

SSI : smart sensor interface

S_{vfb} : flatband voltage power spectral density

Ta : Tantalum

TaN : Tantalum nitride

TAT : Trap-Assisted Tunneling

TB : thermal budget

TCAD : Technology Computer-Aided Design

TDDB : Time-Dependent Dielectric Breakdown

TED : transient enhanced diffusion

TEM : Transmission Electron Microscopy

ToF-SIMS : Time of Flight Secondary Ion Mass Spectrometry

T_{ox} : oxide thickness

T_{OX} : oxide thickness

TRR : Time Resolved Reflectometry

T_{si} : silicon thickness

TSVs : Through Silicon Vias

TTF : Time-To-Failure

ULK : ultra low-k

UV-NLA : Ultraviolet Nanosecond Laser Annealing

V_{EA} : Early voltage

V_{FB} : flatband voltage

V_{TH} : threshold voltage

W : Tungsten

w.r.t. : with respect to

X^{1st} : extension first

X_{last} : extension last

ϵ_{si} : silicon permittivity

ϵ_{OX} : dielectric constant

ϵ_{SiO_2} : permittivity of silicon dioxide

τ_p : propagation delay

$CEFF$: effective capacitance

$IDYN$: dynamic current

Introduction

An alternative approach to conventional planar integration for future nodes is 3D sequential integration (3DSI), also called 3D Monolithic. This kind of integration consists in stacking active device layers on top of each other, in a sequential manner. It offers a unique opportunity to interconnect devices at the transistor level owing to the feasibility of an ultra-small 3D contact pitch ($<100\text{nm}$). The challenge of 3D sequential integration relies in processing high performance devices at low thermal budget (TB) in order to prevent degradation of the bottom layers, including both device and interconnects. For this reason, the top tier TB process window must be addressed, enabling one to safely process the top tier without damaging what is below. Previous studies have set this processing window at 500°C 2h, through the analysis of the device under DC operation. One of the goals of this work is to confirm this TB processing window but this time, with industry relevant Figure of Merit (FOM), such as the yield and the device's performance under AC operation. Further reliability tests on the interconnects shall also be addressed, with different test conditions and structures.

3DSI is deeply studied in LETI and demonstrations on 300mm wafers have already been realized, with a top transistor level integrating for the first time all the features compatible with high performance logic applications (defect free Si monocrystalline channel, high-k/metal gate stack, raised source & drain, etc) with medium cold temperature (600°C). However, even though the initial driver of 3DSI consisted in furthering Moore's Law for high performance applications, 3DSI has been identified as a key technology for More than Moore. This type of integration is thought to be particularly adapted for miniaturized smart sensors arrays. For example, partitioning analog and digital devices on separate levels offers the opportunity of independent tuning of each layer. In this case, analog high voltage ($V_{\text{DD}}=2.5\text{V}$) devices are needed. However, these devices are significantly different from digital devices ($V_{\text{DD}}\leq 1\text{V}$). For example, analog devices do not scale the same way as digital devices, so the short channel effect (SCE) control is not primordial as the channel length is relaxed ($L_{\text{G}}=300\text{nm}$). For the same reason, the access resistance is less critical, but more attention has to be taken in spite of the leakage current and hot carrier degradation. Additionally, the gate stack quality is crucial in order to attain reliable low

noise devices. The aforementioned features also directly impact analog FOM, such as the device voltage gain.

In the integration scheme where these devices are used on the bottom tier, its thermal stability must also be addressed. This is because the thermal stability process window can vary depending on the technological node and the device's architecture. Contrarily, there is also the integration scheme where these devices are used on the top tier. Despite the previous efforts to develop the technological bricks to meet the needs of More Moore devices, further technological development is needed for the More than Moore case. Therefore, another goal of this work is to successfully fabricate a high voltage device ($V_{DD}=2.5V$), at a limited thermal budget ($500^{\circ}C$), that can meet some analog specifications.

Manuscript organization:

The **first chapter** is dedicated to the state of the art of 3DSI, aiming at highlighting the benefits of this integration for More than Moore applications. Different topics shall be addressed, such as the motivation as to why stacking became an alternative to scaling, its thermal budget limitations, and an overview of the technological bricks developed targeting digital applications. Then, the interest of 3DSI for More than Moore, as well as the needs of an analog high voltage device shall be clarified. Lastly, a process flow for the proposed device architecture shall be described.

As it was previously stated, in a 3DSI, the devices must be processed at low TB, as to prevent degradation of what is on the bottom. Consequently, the **second chapter** is dedicated to the thermal stability of the underlying tiers, including the front-end and the back-end of line. Additionally, the study shall be performed on different device architectures, allowing one to draw a complete TB process window.

The **third chapter** is dedicated to the fabrication of a high-quality low temperature gate stack. The adapted oxide splits shall be discussed, where different fabrication methods and post-deposition annealing treatments are used. Through in-depth characterization, the gate stack quality is discussed, and hints about the traps mechanisms in low temperature oxides shall be presented. Based on the results and on the literature, optimization guidelines shall also be proposed.

The **fourth chapter** focus on the low temperature junction. The problematic shall be presented and compared with respect to the one used for digital applications, addressing both device performance and junction leakage. Lastly, optimizations shall be proposed, considering the impact of LT spacers and activation temperature.

Lastly, the conclusions and perspectives of this work shall be presented.

1.1 MOORE'S LAW

In 1965, Gordon E. Moore, co-founder of Intel, upon noticing the emerging trends in chip manufacturing, made a prediction that became one of the driving principles of the semiconductor's industry. Moore's paper titled "Cramming more components onto integrated circuits [1]" stated that the number of transistors on a chip would roughly double every two years, and that this rate of growth would continue for at least another decade (Fig. 1 (a)). This historical trend has become known as "Moore's Law", and even though empirical and based on only six years' data, it has held remarkably well. Moore also argues about the integrated circuit (IC) cost reduction proportioned by miniaturization, as highlighted on Fig. 1 (b). Overall, increasing the number of components in a die improves circuit performance whilst reducing its cost.

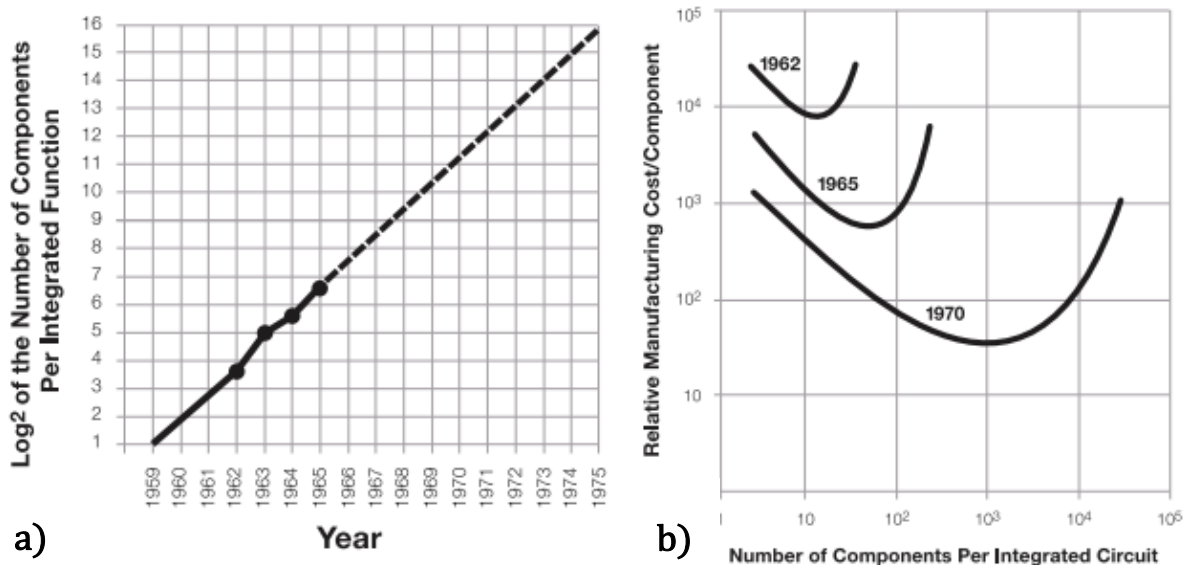


Figure 1.1 a) Moore's extrapolation showing the exponential growth of the number of transistors per integrated circuit over the years and the b) cost decrease for increasing number of components per IC [1].

Together with the scaling rules set by Dennard in 1974 [2], the industry's ability to follow Moore's law has thus relied on technological progress by continually shrinking the metal-oxide-semiconductor field-effect transistor (MOSFET), driven by the cost reduction.

1.1.1 Physical limits to scaling

As a result of extreme transistor scaling, performance gain could no longer occur without any trade-off: undesired SCE has risen, affecting device performance, and increasing leakage current for small gate lengths. Due to SCE, the potential in the channel is no longer controlled solely by the gate but also by the source and the drain (S/D): as the gate length

is reduced, the source and drain depletion regions and its corresponding electric field penetrate the channel, and this effect is amplified as the drain voltage is increased. Drain-induced barrier lowering (DIBL) and a degradation in the subthreshold slope (SS) are undesired effects resulting from SCE (Fig. 1.2). As a result, the leakage current of the transistor is increased, limiting further scaling of MOSFETs.

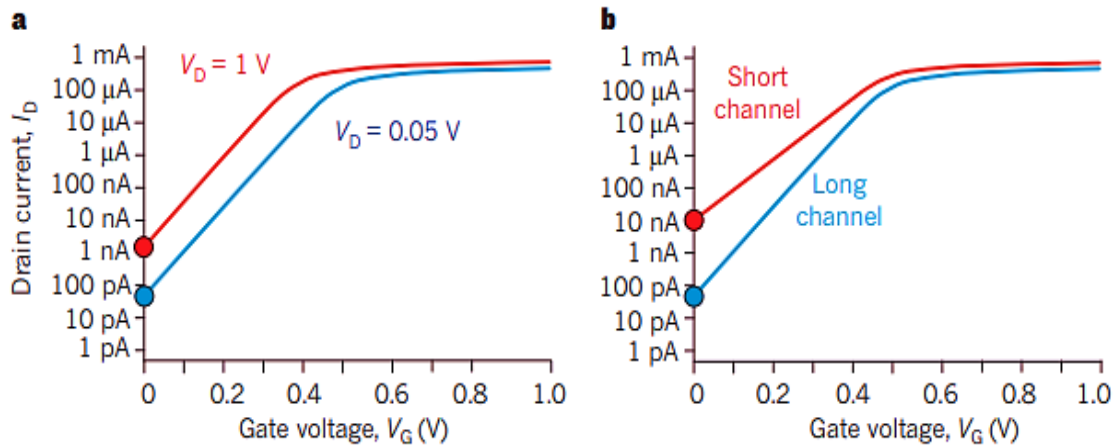


Figure 1.2 MOSFET's drain current (I_D) versus gate voltage (V_G) characteristics for low drain voltage in blue ($V_D=0.05V$) and for high drain voltage ($V_D=V_{DD}=1V$) in red. The increase in leakage current is depicted by the red circles a) due to DIBL at V_{DD} , which causes the threshold voltage (V_{TH}) to decrease, and b) due to SS degradation when the channel length is reduced [3].

1.2 SCALING ALTERNATIVES

In order to continue with Moore's Law, a simple parameter scale was no longer possible due to the previously described SCE. Instead, new device solutions were proposed to deliver the expected circuit performance. In order to compensate for the drive loss while supply voltage is scaled, additional knobs to boost the performance were used, briefly discussed in the next sections.

1.2.1 Silicon germanium channel

Mechanical strain engineering was first proposed by Intel as a performance booster to compensate the current loss due to the SS degradation. In the case of NMOS devices, a tensile stress is created in the channel due to a nitride-capping layer. As for PMOS devices, a compressive stress is induced in the channel thanks to the silicon-germanium (SiGe) used in the S/D, improving the hole mobility. More details on this integration scheme can be found in [4], [5].

1.2.2 High-k materials

A better control of SCE can be achieved by scaling down the gate oxide, hence increasing the gate capacitance and improving the capacitive coupling between the gate and the channel, according to Eq. (1.1):

$$C_{OX} = \frac{\epsilon_{OX}}{T_{OX}} \quad (\text{Equation 1.1})$$

where C_{OX} is the gate oxide capacitance, ϵ_{OX} the dielectric constant and T_{OX} the oxide thickness. Silicon dioxide (SiO_2) continued to be scaled down until it reached its limits ($\sim 1.1\text{nm}$ [6]). Significant gate leakage current occurs as a result of further scaling, a quantum mechanical effect caused by carriers tunneling through the insulator potential barrier. As a result, the semiconductor industry started to look for materials with a higher dielectric constant (high-k) in order to replace the SiO_2 gate oxide. The concept of effective oxide thickness (EOT) from Eq. (1.2) can be used to see how to achieve the same coupling effect with thicker oxide thus limiting tunneling:

$$EOT = \frac{\epsilon_{\text{SiO}_2}}{\epsilon_{OX}} \cdot T_{OX} \quad (\text{Equation 1.2})$$

with ϵ_{SiO_2} the relative permittivity of SiO_2 equal to 3.9 and ϵ_{OX} the permittivity of the new dielectric. Thanks to its high permittivity ($K=25$), hafnium oxide (HfO_2) has been widely used as gate dielectric as a substitute or in combination with SiO_2 (for reducing interface defect density) [7].

1.2.3 New MOSFET architectures

Despite some efforts, it continued to be difficult to counteract SCE. Consequently, new device architectures such as multi-gate MOSFETs were proposed [3]. In these structures, the gate electrode is wrapped around several sides of the channel region, improving the gate's electrostatic control. Some of these architectures, such as FinFets and Nanowires, have replaced planar transistors for sub-30nm technologies, where more details can be found elsewhere [8], [9].

Within the scope of this thesis, thin film devices based on Silicon-on-Insulator (SOI) [10] architectures shall be highlighted. These devices are possible thanks to the fabrication process of SOI substrate called SmartCut™ [11], developed by Soitec in collaboration with CEA-Leti. A simplified schematic is provided on Fig. 1.3. The process starts with a thermal oxidation of the silicon wafer (A), which will be the buried oxide (BOX) of the final wafer.

After the hydrogen implantation is performed, wafers (A) and (B) are bonded, where (B) represents the substrate of the SOI. An annealing step then induces the layer splitting. Lastly, Chemical Mechanical polishing (CMP) is performed to uniformly flatten the silicon thickness. The split wafer (A) is recycled.

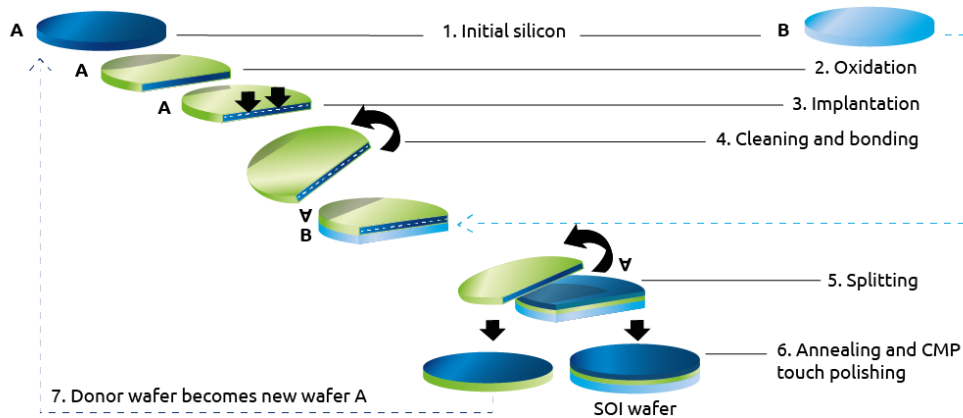


Figure 1.3 SmartCut™ process.

The BOX and silicon film thickness of the final wafer can be adapted by means of controlling the oxidation steps and by tuning the hydrogen implantation energy, respectively. As a result, SOI devices are produced.

The addition of the BOX provides a much better electrostatic control, isolating the channel from the bulk substrate, preventing leakage currents and parasitic capacitances between the S/D and the bulk. If the silicon film is thin enough, the device is Fully Depleted-SOI (FD-SOI) as opposed to Partially Depleted (PD-SOI), and the electrostatic control is improved [12]. Lastly, SOI devices also benefits from a strong back-bias polarization, where the V_{TH} can be adjusted to achieve the best trade-off between performance and power consumption [13], [14].



Figure 1.4 Schematic of Bulk (on the left) and FD-SOI technology (on the right).

1.3 INTRODUCTION TO 3D SEQUENTIAL INTEGRATION

There has been an increasing difficulty to continue Moore's Law. Other than the transistor's physical limitations, its increasing number on chip also leads to interconnects delay due to the long length of metal wires to connect them. Despite the technological efforts on the back end of line (BEOL), the delay was still limiting IC's performance. From an economical point of view, the development costs have become more expensive, due to the necessity of high-precision equipment and design tools. As an alternative to scaling, a 3D integration scheme can be used, which consists in stacking the transistor levels rather than reducing its dimensions.

3D Sequential Integration, also referred as 3D monolithic integration or CoolCube™, consists in stacking active device layers on top of each other, in a sequential manner [15]. It differs from 3D Parallel Integration (also known as 3D Packaging), where different chips are processed independently, then stacked vertically and connected using for example Through Silicon Vias (TSVs). The 3D Sequential flow, depicted on Fig. 1.5, starts with the fabrication of the bottom layer (also called tier 1), followed by the realization of the top active layer. The latter needs to be thin because the first lithography step of the top tier uses the alignment marks to be able to align with the lower level. An interesting method to obtain the top active is via SOI direct bonding, enabling to obtain perfect pure monocrystalline channel quality with a controlled thickness. The other approaches for top active fabrication can be found elsewhere [16]. Afterwards, the top layer (also called tier 2) is processed on this active layer, finishing by the 3D contact processing. The 3D contact has dimensions similar to the contacts of the sources and drains of the transistor, and they are located between the upper active areas to be isolated. It is stressed that whilst the tier 1 has no processing constraints, the tier 2 must be fabricated with a low thermal budget, in order to prevent degradation of the underlying tiers.

3DSI offers unique 3D connectivity opportunities if compared with 3D Parallel Integration thanks to the high alignment of the technique. On a 3D Parallel scheme, the alignment is performed during the bonding step, where the wafers have already been patterned: this implies that the interconnects density will be limited by the bonding alignment accuracy ($\mu \pm 3\sigma = 200\text{nm}$ [17]). Differently, on a 3DSI, the alignment is done during a photolithography step. Hence, alignment accuracy and feature size of stacked tiers are only limited by lithography stepper resolution ($\mu \pm 3\sigma < 20\text{nm}$ [18], [19]). This is due to

the fact that the transferred top active consists of a thin blanket wafer with no patterns on it, so the alignment marks of the bottom layer are used to align the first patterning.

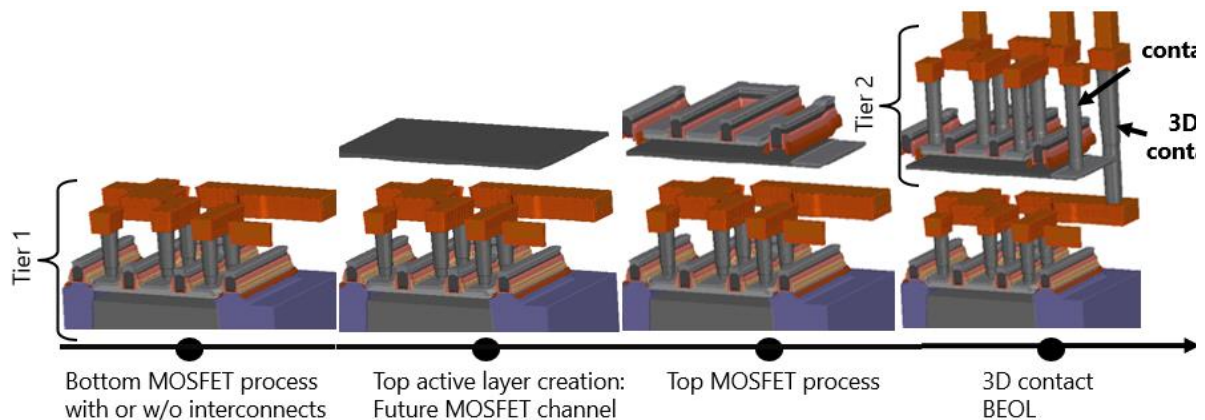


Figure 1.5 3D Sequential Integration flow.

Another advantage of 3DSI is the small 3D contact dimension, where the contacts are less deep and less large than the TSVs. Overall, a 3D contact has the same process as a normal contact, only with a slightly higher aspect ratio (AR). The 3D contact AR depends on its height and width ($AR = W_{\text{contact}}/H_{\text{contact}}$): The width scales with the used technology, whereas the height is slightly larger due to the additional etching of the underlying layers (thus, it depends on the top active thickness and inter-layer dielectric). As a result, it can be said that the height of a 3D contact equals the sum of the top channel thickness ($T_{\text{channel,top}} \sim 10\text{nm}$) and the thickness of the dielectric ($T_{\text{ILD}} \sim 100\text{nm}$).

Overall, both the high alignment and the contact dimension contributes to the superior 3D contact density and via pitch achieved with 3DSI, as shown on Fig. 1.6. While 3DSI offers a maximum density demonstrated of 2×10^7 via/ mm^2 (with 65nm design rules), 3D Parallel is limited to only 10^6 vias/ mm^2 .

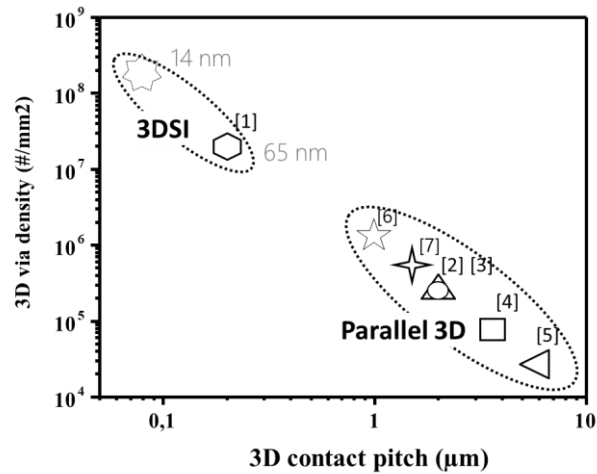


Figure 1.6 3D via density as a function of the 3D contact pitch comparing 3D Sequential and Parallel integration [20].

1.4 MORE MOORE BY DEVICE STACKING

Furthering Moore's law for high-performance applications was the original drive for 3D sequential integration. On a 3DSI scheme, different partitioning between tiers is possible, allowing for a certain degree of freedom as to which part of the circuit will be on the top level and which will be on the bottom one. This is represented on the granularity scale shown on Fig. 1.7.

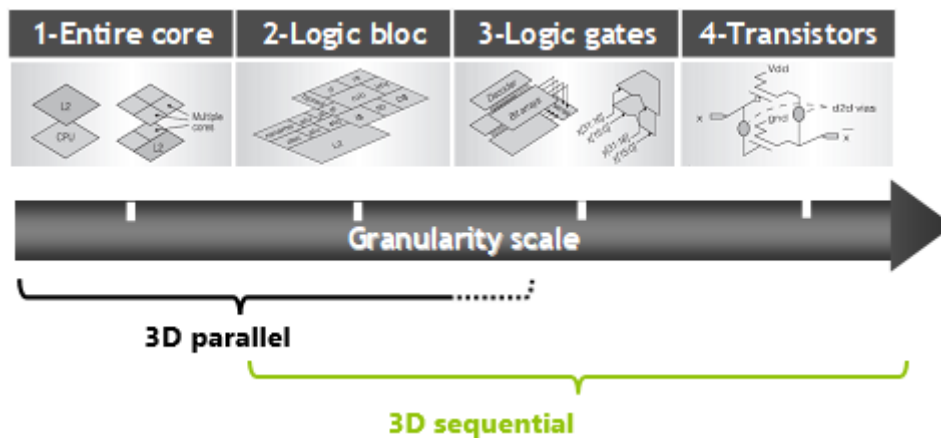


Figure 1.7 Granularity scale, highlighting that only with a 3DSI transistors partitioning can be attained [20].

From the granularity scale, it can be seen that whilst 3D Parallel is suitable to connect entire core or logic blocks, the highest granularities are only attained through a 3D Sequential scheme, such as connecting logic gates (CMOS over CMOS) and transistor over transistor. This is only possible thanks to the high contact density previously described.

With that in mind, a first strategy for More Moore consisted in stacking NMOS over PMOS (N/P), or contrarily (P/N) [21]. The main interest of this integration is to independently optimize the NMOS and PMOS transistors with its most suitable performance booster, such as channel material, substrate orientation, and so on. For example, CEA Leti has successfully demonstrated the integration of a PMOS transistor with an upper active area in Germanium (thus boosting the holes mobility) above an NMOS transistor in Silicon [22]. However, later on, it became evident that moving towards the most advanced nodes reduced the transistor delay, but the trend was different concerning the interconnects. The reduced width and pitch of the lines leads to an increased resistance and parasitic capacitance, respectively (Fig. 1.8). Additionally, it was highlighted that to benefit from the gate delay improvement of the advanced nodes, the metal lines had to become shorter.

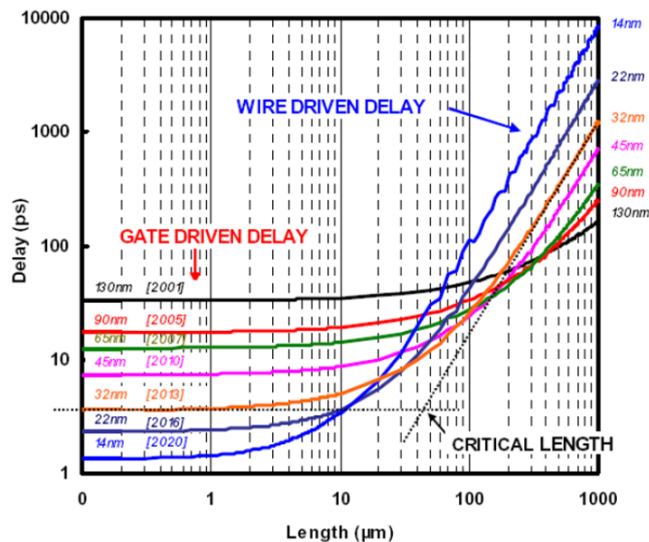


Figure 1.8 Delay as a function of wire length for several ITRS predicted nodes [23].

At that stage, partitioning at the logic bloc or logic gate level (CMOS over CMOS) became an interesting option to overcome this issue. CMOS over CMOS was possible by adding an intermediate back end of line (iBEOL) in between the stacked FETs, shortening the lines length.

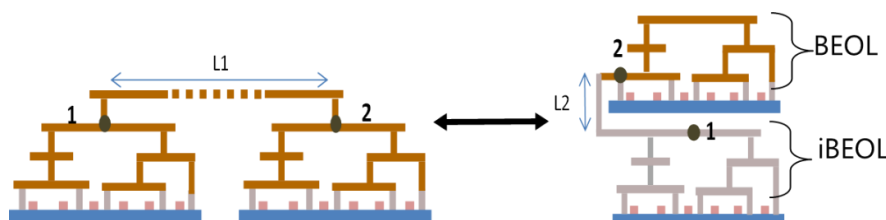


Figure 1.9 iBEOL introduction.

Leti demonstrated a full 3D CMOS over CMOS integration on 300mm wafers, where the top level was successfully fabricated through a low temperature process flow below 650°C [18]. Other groups have also successfully demonstrated the viability of 3DSI, shortly summarized here. IMEC [19] demonstrated the compatibility of the 3D sequential approach for aggressive device density stacking at advanced nodes by stacking FinFETs at 45nm fin pitch and 110nm gate pitch. In their integration, the top is composed of junction-less devices fabricated at low temperature ($T=525^{\circ}\text{C}$) in a top Si layer transferred wafer-to-wafer bonding. Samsung has developed memories using 3D sequential, where SRAM on three levels have been demonstrated [24]. Another example comes from Stanford, who has integrated carbon nanotube field effect transistors (CNTFETs) above a silicon CMOS level. TSMC also successfully integrated advanced Si logic circuits with low-cost and high-mobility CNT transistors in the BEOL fabricated at low, BEOL-compatible temperatures (250°C) in high volume manufacturing.

IBM have demonstrated full circuit functionality with a dense functional 3D 6T-static random access memory (SRAM) circuit [25]. The latter was proposed with an "N-over-CMOS" 3-DM integration of a replacement metal gate (RMG) InGaAs n-FinFET layer on fully-depleted silicon-on-insulator (FDSOI) CMOS with TiN/W interlayer contacts. AIST have also demonstrated an ultimate CMOS inverters composed of high mobility wire channel InGaAs-OI nMOSFETs and SGOI pMOSFETs by means of 3D sequential integration [26]. 3D integration of Ge gate-all-around (GAA) PMOS on top of the Si FinFET NMOS to form functional stacked CMOS inverters has also been successfully demonstrated by Intel [27] and proven a viable option for low power high performance CMOS logic applications and for continuing Moore's law in future technology nodes.

1.4.1 Technological implications: bottom tier stability

It has been previously stated that the main challenge of 3DSI consists in preventing degradation of the underlying tiers. Many studies have been performed in order to define the thermal budget process window of the top tier and will be extensively discussed on chapter 2. In summary, these studies consist of applying an additional TB (either standard furnace anneals, μ wave, laser...), which mimics the top transistor fabrication, either before or after BEOL processing. This allows discriminating the stability of device (front end of line - FEOL) and interconnects. Both cases are briefly summarized in the following sections.

1.4.1.1 Impact on device performance (FEOL)

Upon applying the TB, the device is further characterized, and its performance is analyzed with parameters such as I_{ON}/I_{OFF} , V_{TH} and access resistance (R_{SD}), for example. As a result, one can draw a temperature as a function of anneal duration plot, where the thermal stability process window can be defined. As an example, the bottom thermal stability of 14nm FDSOI technology has been validated (in a simplified way) to $\sim 500^{\circ}\text{C}$ 5h (Fig. 1.10).

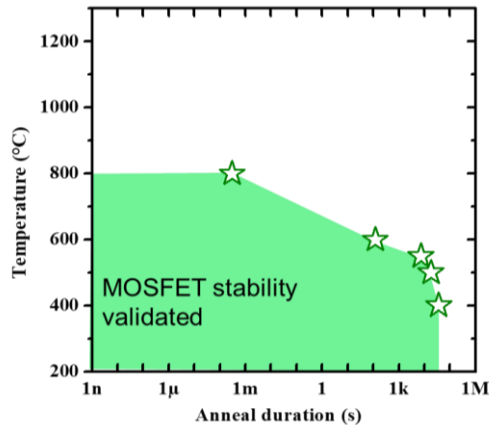


Figure 1.10 14nm FDSOI thermal stability without interconnects, with the green area representing the allowable zone [28].

In these studies, the transistor's stability has been always defined through the analysis under DC operation. One interesting way to complement these studies is to analyze the device performance with industry relevant FOM, such as its stability under dynamic operation and with the yield of real circuits, such as SRAM and logic gate arrays.

Additionally, the given example depicts the case of 14nm FDSOI, but if a 28nm or a BULK technology is used instead, the processing window of the top tier needs to be re-evaluated accordingly. As such, it can be said that the thermal stability varies depending on technological node and device architecture. Consequently, taking these considerations into account, a more accurate TB process window can be drawn, and will be presented on chapter 2.

Overall, there are different parameters that can be affected by an additional TB, but it has been shown that one of the most critical on the device stability is the silicide, where some changes in the silicide morphology can occur (such as agglomeration and abnormal silicide diffusion) [29]–[31]. This leads to undesired silicide resistance increase and additional leakage, respectively. These phenomena are however strictly dependent on the TB and on the silicide used. This means that if a more stable silicide is employed, or if a thicker silicide is needed for a specific application, its stability may change.

Another important observation from the example shown on Fig. 1.10 is that a TB corresponds to both a temperature *AND* duration, meaning that some TB with a very high temperature and short time duration could potentially be used. One example consists of a laser anneal on the millisecond range (ms), achieved with a Dynamic Surface Anneal (DSA), where a TB of 800°C 0.5ms can be reached without degrading the performance of the components [28]. Another interesting option is to use a Ultraviolet Nanosecond Laser Annealing (UV-NLA $\lambda=308\text{nm}$ with a pulse of 160ns) [32] developed by LASSE (*Laser Systems & Solutions of Europe*), which provides local heating while keeping the underlying tiers at a low temperature [33]. Overall, the short annealing durations limits the diffusion of thermal energy from the annealed surface to the lower layers, potentially making it possible to reach temperatures above 1000 °C for the upper layer whilst the bottom tiers integrity are retained. The later can be used for example to activate the S/D dopants [34] or to recrystallize the Polysilicon of the gate stack [35]. However, the heat diffusion depends on the material on which the laser is being applied and on its thickness. Also, what is below (i.e. the patterning marks) can have an impact on the uniformity of the annealing.

1.4.1.2 Impact on the interconnects (BEOL)

As previously stated, iBEOL have been incorporated to benefit from the full 3D opportunities and avoid global routing congestion. This integration brings additional challenges, seeing that the iBEOL also needs to withstand the top tier fabrication TB. In the case of a State-of-the-Art (SOTA) 28nm interconnects with the pair Copper (Cu) and an ultra low-k material (ULK), such as porous SiOCH with $k\sim 2.7$), its stability has been evaluated and set at 500°C through the RC delay and reliability measurements at 25°C. This was defined to prevent Cu diffusion into the ULK, which can take place despite the Ta-TaN barrier. It must also be noted that even though these studies were essential to have an indication of what happens in the interconnects, additional tests are still lacking. For that, the dielectric breakdown failure needs to be performed at other test temperatures and voltages, allowing one to have a more accurate time to failure and additional information concerning the breakdown mechanisms. As such, another goal of chapter 2 is to complement the BEOL thermal stability studies.

1.4.2 Low temperature process flow

Two strategies are therefore possible to overcome top tier TB limitations, where the first consists of improving the thermal stability of the bottom tiers. Within the scope of this thesis, more attention shall be given on the second strategy, which consists of reducing the TB of some crucial top tier fabrication steps. The TB of a process of reference (POR) silicon MOSFET is relatively high, hence not tolerated on a 3DSI scheme, where the TB should be kept in the 500°C 2h range (depending on the node) [29], [30]. Consequently, many advances have been done to reduce the TB of the most critical fabrication steps (Fig 1.11) and will be briefly described.

The dopant activation of the (S/D) is one of the most challenging bricks, usually performed at $T > 1000^\circ\text{C}$. UV-NLA and Solid-Phase Epitaxial Regrowth (SPER) are key technological solutions in order to activate the junction in a 3DSI TB friendly [36]. The latter is currently pursued as the preferred method used in Leti, where high performance devices have been demonstrated even at 500°C. Extensions last (Xlast) implantations are performed after raised source and drain epitaxy. It must be noted that in the most advanced nodes, due to the absence of dopant diffusion with the LT SPER process, R_{SD} penalty has been observed. To overcome this issue, an extension-first process (X^{1st}) has been proposed, where the implant is performed at the channel ends before raised S/D are formed. This scheme shows promising results in terms of R_{ON} -DIBL trade-off, compensating the insufficient dopant concentration under the spacer region .

Another challenge consists in fabricating reliable gate stack with a 500°C TB. In a standard process without TB constraints, the gate stack is formed at a relatively high temperature ($\geq 800^\circ\text{C}$). Additionally, for a standard flow, the high TB of the junction formation also helps to improve the gate quality. As a low temperature alternative, a chemical oxide followed by a high-pressure deuterium passivation could be used, thus improving the oxide quality and reducing interface states [37]. Despite, gate resistance (poly silicon recrystallization at low temp) & Negative Bias Temperature Instability (NBTI) remains the challenges for low TB schemes [38], and shall be discussed in more details on chapter 3.

Concerning the raised S/D fabrication, silicon epitaxy at 500°C has been recently demonstrated, by using a combination of wet and dry surface cleans followed by a Cyclic Deposition/Etch (CDE) strategy [39], exhibiting full selectivity over the dielectrics and

good crystallinity. Lastly, in a standard process, the offset spacer isolating the gate stack from raised S/D consists of silicon nitride (SiN_x) deposited above 600°C by Atomic Layer Deposition (ALD). There exists however a low temperature alternative, with a SiCO deposited at 400° [38]. This spacer not only meet integration requirements (such as CMOS FEOL compatibility, conformity and uniformity and stability in temperature), but it is also advantageous because of its resistance to the high HF cleaning needed for raised S/D epitaxy, while it also offers the advantage to have a low-k value.

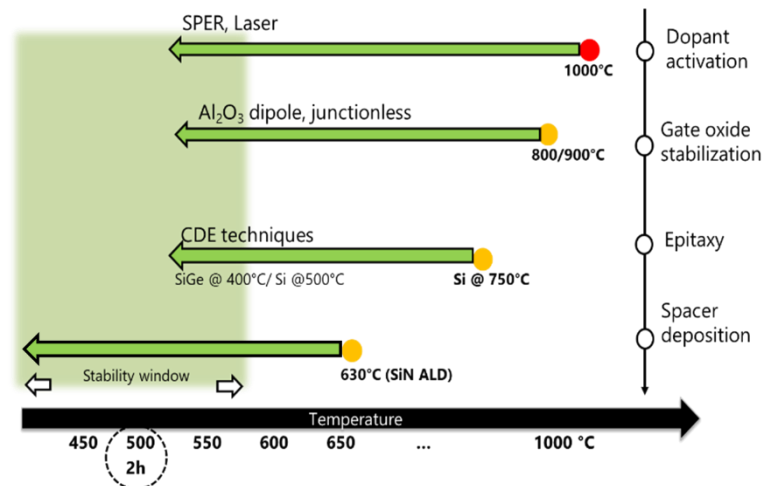


Figure 1.11 Summary of Si MOSFET thermal budget and its low temperature alternatives [28].

These technological bricks have been developed to meet the needs of More Moore devices. As such, the first goal is to see if these bricks can be used if other applications are targeted, or if there is the need of further technological development.

1.5 MORE THAN MOORE BY DEVICE STACKING

The miniaturization and its associated benefits in terms of performance, the direction for further progress with the More Moore approach, as well as the benefits brought by 3DSI were already briefly described. In this section, another development axis of the Moore's Law market, called More than Moore (MtM), shall be presented.

MtM does not necessarily aim to scale the device and should not be seen as an alternative or competitor to the digital trend as described by Moore's Law. Instead, the goal is to add functionality in an integrated system, by complementing digital signal and data processing (Fig. 1.12).

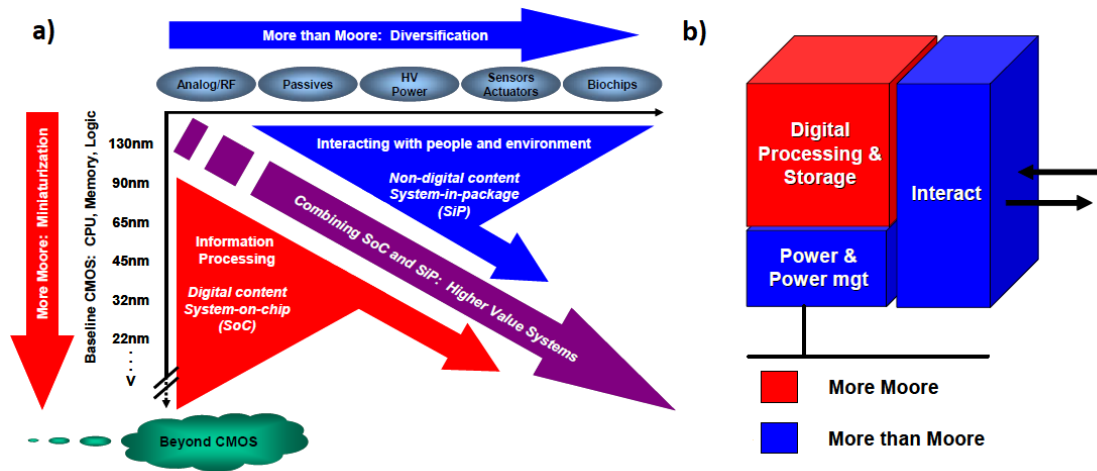


Figure 1.12: a) The combined need for digital and non-digital functionalities in an IC trend, b) where MtM devices complements the digital processing and storage elements by allowing the interaction with the outside world and in powering the system [40].

In reality, most CMOS chips are not purely digital systems and could be considered mixed-signal systems, with analog building blocks in addition to the large digital core. This heterogeneous integration of digital and non-digital functionalities is the key driver for a variety of applications. This is because overall, MtM applications require devices that interacts with the outside world. Among those, it can be mentioned the analog circuits, sensors and actuators, biochips and so on. However, one must keep in mind that as these interactions become more complex, more sensors are required to be integrated in the same product, which in turn increases the processing capability to handle all of these sensors.

Even though System-on-Chip (SoC) integrates digital, analog, and mixed-signal processing functions, a mixed-signal SoC is relatively costly per area (requiring more mask sets to offer multiple process options) and tends to offer non-optimal performance in the different signal domains (i.e. analog, digital).

It must also be said that analog and mixed signal ICs do not scale like digital circuits, mainly due to their higher sensitivity to “Process-Voltage-Temperature” (PVT) variation, noise, and the high design constraints that analog designers have when the supply voltage is low. This makes that analog circuits are actually more expensive if the latest technology is used, being the reason why older technology are usually employed for analog, simply

because it is cheaper. As a consequence, while the price per digital device is going down with more advanced CMOS technology, the price per analog device is actually going up.

3D integration scheme easily accomplishes the association of heterogeneous technologies and is a viable alternative to SoCs for highly integrated mixed-signal systems, where each tier can offer specialized options for specific signal domains (Fig. 1.13). For example, one can use advanced CMOS nodes to achieve high performance for digital processors and memories, whilst analog devices can be designed on less advanced and less expensive technologies. In addition, given that the analog performance does not necessarily improve with newer CMOS generations, the analog part does not need to be realized on the same die as the digital core, hence it does not need to cope with the CMOS evolution dictated by the digital circuit.

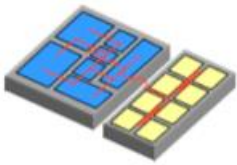
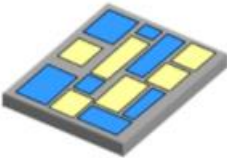
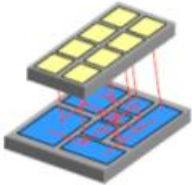
2D Interconnect	SoC (System on Chip)	3D integration
		
<ul style="list-style-type: none"> • Large form factor • Long lines / shared bus 	<ul style="list-style-type: none"> • Reduced system size • Increased performance • Increased device cost 	<ul style="list-style-type: none"> • Reduced system size • Short interconnects • Reduced packaging cost

Figure 1.13 Driving forces for 3D Integration [41].

Therefore, it becomes clear that even though the initial drive of 3DSI consisted in furthering Moore's Law, 3DSI is also a key technology for More than Moore, by partitioning analog and digital devices on separate levels. The next section shall discuss a few examples in view of this application and highlight the focus of this work.

1.5.1.1 Computing and data storage

Heterogeneous partitioning can be also interesting in order to overcome the communication bottleneck [42]. This is because as the applications evolve, massive volumes of data need to be treated, which is not an easy task to do when there are two different chips for computing and data storage with limited connections between them. Instead, transformative nanosystems, which use new nanotechnologies to simultaneously realize improved devices and new integrated circuit architectures, are required. By

inserting ultra-dense wires between these layers, 3D architectures promise to address this communication issue, providing dense and fine-grained integration for energy-efficient digital logic circuits and dense data storage, all fabricated on vertically stacked layers in a single chip.

Owing to the fine-grained and dense connectivity between vertical layers of computing, data storage and sensing, it can capture and directly store massive amounts of data from the outside world, perform in-situ processing and produce highly processed results, all on the same chip (Fig. 1.14). Due to the layering of sensing, data storage, and computing, the chip is able to measure each of the sensors in parallel, and then write directly into its memory, generating huge data bandwidth.

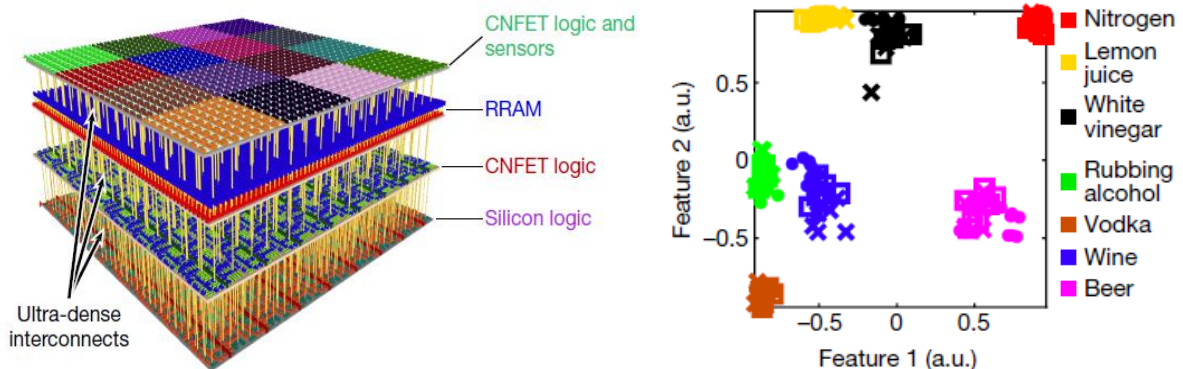


Figure 1.14 On the left, a schematic of the proposed 3D nanosystem is presented, integrating four monolithically integrated vertical layers, connected through dense vertical inter-connects. The carbon nanotube field-effect transistor sensors analysis and correctly classifies nitrogen and six other vapors [42].

The aforementioned communication issue can also be seen on data intensive applications, where real-time, low-latency object classification from images or video is critical for applications ranging from autonomous driving, to robotics, to augmented reality. In the end, the performance of such systems is being dominated not by capturing the image itself but by the rate at which data can be read from the pixel array, stored in memory, read from the memory by a processor, and then classified. Improved processing latencies are attained by integrating imagers, memory, and logic through chip stacking, such that monolithic 3D imagers can capture and compute on all of the pixels, transforming the imager output from raw pixel values to highly processed information [43].

1.5.1.2 Image Sensors

CMOS image sensors (CIS) is one of the main driving forces for MtM applications. The trend for higher pixel density, for example to increase image resolution, leads to smaller and smaller pixel size. However, extreme pixel scaling can have a negative impact on their optical performance, where more details on the impact of miniaturization on the pixel parameters can be found elsewhere [44], [45].

Technological solutions were proposed to overcome these issues, where the most effective alternative to maintain pixel size scaling trend without sacrificing sensor performance consisted in building image sensors with a Back Side Illumination (BSI) technology instead of Front Side Illumination (FSI) pixel [46]. The main difference between both is where the circuitry is positioned with respect to the photosensitive layer. For CMOS image sensors, the incident light first passed through the on-chip lens, via the color filter, then past the metal wiring, reaching the silicon, where it was converted to electrons. In an FSI, light goes through a stack of interconnects before being captured by the photodiode, reducing the light available for conversion to electrons in the silicon. By using the BSI approach, light can reach the silicon more efficiently, since it does not go through the BEOL.

However, such solutions will remain insufficient in the perspective of sub-micron pixels, hence the interest of 3DSI, required for pixel partitioning with pitch in the $1\mu\text{m}$ range [47].

Through a 3D stack process integration, the CMOS processing part can be positioned on the top of the pixel, where more advanced technology nodes for the different circuit functions can be used. Also, thanks to the combination of BSI and 3D stack, the ultimate pixel size/performance trade-off can be achieved. On this idea, it has been proposed that the image sensor can be partitioned on 2 layers (Fig. 1.15): The back-side illuminated pinned photodiodes constructed on a first layer while the three readout transistors are moved to a second SOI layer. This integration scheme allows for a 35% increase of the photosensitive area in the case of the $1.4\mu\text{m}$ pixels leading to strong Full Well enhancement. As said before, 3DSI is suitable for miniaturized pixels because via diameters of 90nm are required to ensure the performances of the shared $1.4\mu\text{m}$ pixels.

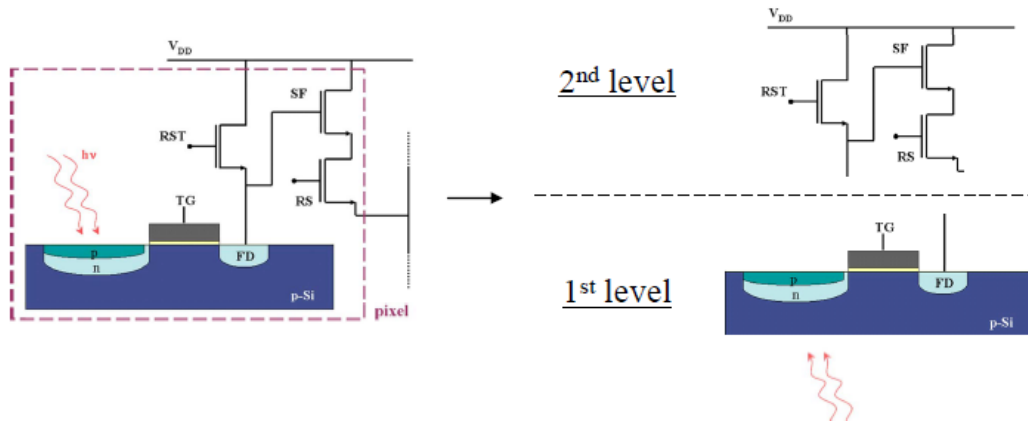


Figure 1.15 Schematic representation of the proposed architecture to go from a 2D to a 3D scheme: Back-illuminated Pinned photodiode/Transfer Gate are located on the bottom layer while three pixel transistors are moved to a second layer [47].

1.5.1.3 This work: Smart sensor interfaces

As it has been previously highlighted, this type of integration is thought to be particularly adapted for miniaturized smart sensing systems, by partitioning analog and digital devices on separate levels, with configurations depicted on Fig. 1.16.

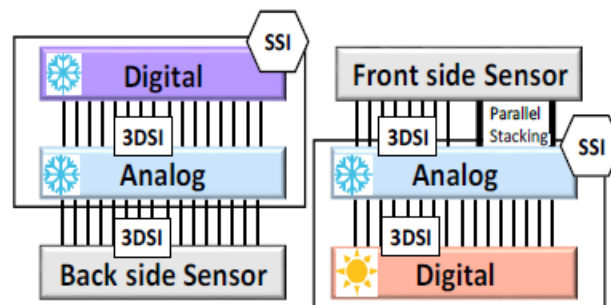


Figure 1.16: Schematics of digital/analog bilayer sensors interface module for sensing applications for a) Back side sensors and b) for above IC sensors. In both configurations, analog devices have to be fabricated at LT to preserve the underlying tier. *Smart sensor interface SSI

The sensing system will have a different structure depending on which sensor system it is designed for (for back side sensors or for above IC sensors). It is highlighted that despite the configuration, devices compatible with analog requirements are needed. In addition, in view of a 3DSI, these devices will have to be fabricated at a low thermal budget.

Given that the previous technological bricks were developed targeting digital applications, engineering needs to be done to comply with the analog needs. It must also be noted that depending on the sensor technology, their position on the stack will differ. Consequently, to offer the maximum flexibility in the partitioning options, the analog device stability shall also be evaluated.

1.6 HIGH VOLTAGE DEVICES

This section aims at highlighting the main features of the analog transistor fabricated in this thesis. Even though some applications do require analog transistors to operate at low V_{DD} , this thesis aims to design analog transistors which are required to be operated at a higher voltage (HV), typically $V_{DD}=2.5-5V$ in order to retain the integrity of the signal at a reasonable power consumption level.

1.6.1 Requirements

One needs to remember that analog circuits operate with electrical variables and not simply with numbers, therefore their dynamic range is limited by noise, offset and distortions. As a result, noise ultimately sets a lower limit on signals that can be detected and processed. **Noise** in electronic circuits is defined as random fluctuations in signal current or voltage, and it is the main limitation on the accuracy of a measuring device, as such it is highly desired to be reduced. There are several noise sources that affect the MOSFET; among those, the flicker noise ($1/f$) is attributed to the random trapping and de-trapping process of charges in the oxide traps near the Si/SiO₂ interface. So in theory, if this interface is of high quality, the noise component should be decreased. Whereas the introduction of novel high-k gate dielectrics for digital applications can pose new challenges for noise reduction, the latter is not the case for the analog gate stack, which can be composed of SiO₂ only, as the oxide thickness is relaxed.

Another parameter highly related to gate stack quality is the **reliability**. Indeed, this is not necessarily an analog requirement, but it is overall desirable to ensure the lifetime of a transistor. Bias temperature instability corresponds to a shift in threshold voltage with applied stress. When the shift exceeds some specified value, the device is considered to have failed. In the case of NMOS, the applied stress is positive (PBTI), causing a negative oxide charge that leads to a positive V_{TH} shift. The opposite applies to PMOS (NBTI): the stress is negative which induces a positive oxide trapping, resulting on a negative V_{TH} shift. The reliability can be immensely degraded if there are a lot of defects in the gate oxide, hence the importance of a high quality stack, which can be very challenging with TB limitations [38].

The mismatch is also crucial because it limits the precision of analog circuits. Ideally, identically designed components that are processed and biased under the same conditions should have the same electrical properties, where the mismatch it the

difference between their electrical parameters. Even though very important, the mismatch shall not be the focus of this study, and more details can be found elsewhere [48].

When dealing with high voltages, the **hot carrier injection** (HCI) can become a serious concern [49]. This is because the high electric field created by the large V_{DD} accelerates the particles, giving it enough kinetic energy to be injected, for example, in the gate dielectric. This can lead to significant reduction on the device threshold voltage, transconductance, and drain current. In addition to the gate stack quality, the junction profile can be engineered to decrease the high electric field intensity near the drain.

1.6.2 FOM for analog circuits

The DC gain of a transistor is represented on Eq. 1.3 as:

$$\frac{\Delta V_{out}}{\Delta V_{in}} = \frac{\Delta I_D}{g_D} \frac{1}{\Delta V_{in}} = \frac{\Delta V_{in} g_m}{g_D} \frac{1}{\Delta V_{in}} = \frac{g_m}{g_D} = \frac{g_m}{I_D} V_{EA} \quad (\text{Equation 1.3})$$

where g_D is the output drain conductance and V_{EA} is the Early voltage, with g_D being approximated by Eq. 1.4:

$$g_D = \left. \frac{\partial I_D}{\partial V_D} \right|_{V_G} = \frac{1}{r_0} \quad (\text{Equation 1.4})$$

Fig. 1.17 depicts as an example the $I_D V_D$ characteristic of a MOSFET. For a given gate voltage, the slope is described at any point of the output by the intercept of the extended $I_D V_D$ characteristic with the V_D axis, with this intercept is the V_{EA} . However, this intercept depends on g_D , which does not have the same behavior on long and short channel devices. This is because in saturation, the drain current should be independent of drain voltage and the conductance g_D should ideally be zero. There is, however, a finite increase in drain current due to channel length modulation (CLM). This increase results in a non-zero conductance and hence finite output resistance, r_0 .

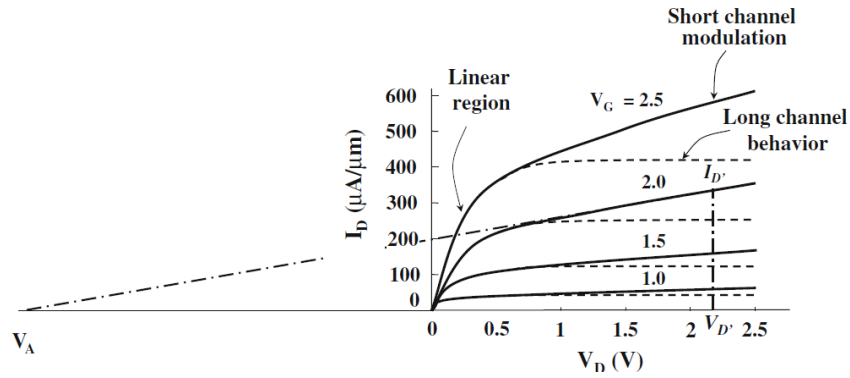


Figure 1.17 Comparison of channel length modulation in long- and short-channel NMOS. V_{EA} is the Early voltage [50].

It has been mentioned that analog transistors do not necessarily need to be scaled down [51], [52], and thus can benefit from a long channel (the relaxed L_G also comes from the higher V_{DD}). As such, in analog design, a relaxed L_G is beneficial for yielding higher gain (g_m/g_D), which should be kept as high as possible for linearity purposes.

Still regarding Eq. 1.16, another important analog FOM is the g_m/I_D , which represents the voltage gain of a MOSFET (the transconductance represents the amplification delivered by the device, and the drain current represents the power dissipated to obtain that amplification). As previously mentioned, the devices used on this work are fabricated on FD-SOI technology. The lower body factor (n) of FDSOI w.r.t. BULK devices positively impact the performance of analog circuits. The high value of FD-SOI devices should allow one to realize near-optimal micropower analog designs. Overall, the higher the ratio, the more efficient the device. In FD-SOI MOSFET's values of $35^{-1}V$ are obtained, whereas BULK devices are limited to values of $25^{-1}V$, as presented in Fig. 1.18.

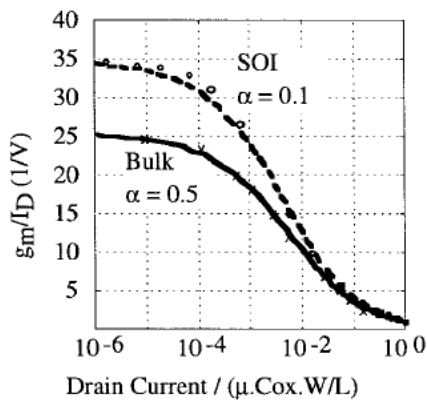


Figure 1.18 g_m/I_D ratios in saturation ($V_D=2.5V$). Typical values of n range between 1.3 and 1.5 in a bulk transistor and between 1.05 and 1.1 in a FD-SOI MOSFET [53].

There are other analog FOM, which shall not be the focus of this work, but that more details on it can be found elsewhere [54], [55].

1.6.3 Proposed process flow

Given all the differences between digital and analog design, it has become clear that for a high voltage analog transistor ($V_{DD}=2.5V$), the device does not need to be scaled down. Therefore, the aforementioned proposed low temperature process flow needs to be properly adapted. For this reason, 2.5V devices have been fabricated with a limited TB (in the 500°C range) based on the 28nm FD-SOI technology with a 25nm BOX. Thanks to the relaxed gate length of analog transistors (nominal $L_G=0.3\mu m$), the devices have a thick 30nm silicon undoped channel (<100> channel orientation). If needed, the channel can be further reduced to 22nm for a better SCE control. Note that, thanks to the relatively thick channel, raised S/D epitaxy is not needed as compared to digital FD-SOI device, thus enabling to suppress a relatively complex process with a consequent TB and to drastically reduce device cost.

To withstand the higher V_{DD} (2.5V as opposed to 1V for digital), the oxide basically needs to be thicker, where an $SiO_2=6nm$ was envisaged (The full gate stack being composed of 6nm SiO_2 / TiN / Polysilicon). For this oxide thickness, there is no need to use a high-k, which was introduced solely to comply with the aggressive scaling of digital devices. Owing to the thicker SiO_2 , different oxide deposition splits are used, as well as post deposition treatments to improve the oxide quality. More details on the gate stack conception can be found on chapter 3.

For this architecture, TCAD electrical simulations have been performed in order to verify if the device remained FD (Fig. 1.19), which is the case despite the silicon thickness ($TSI=30nm$). In addition, the strong back bias coupling (250mV/V) of this undoped device offers advantages regarding V_{TH} tuning and mismatch.

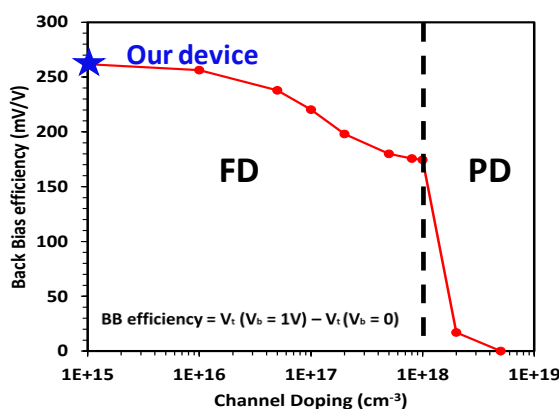
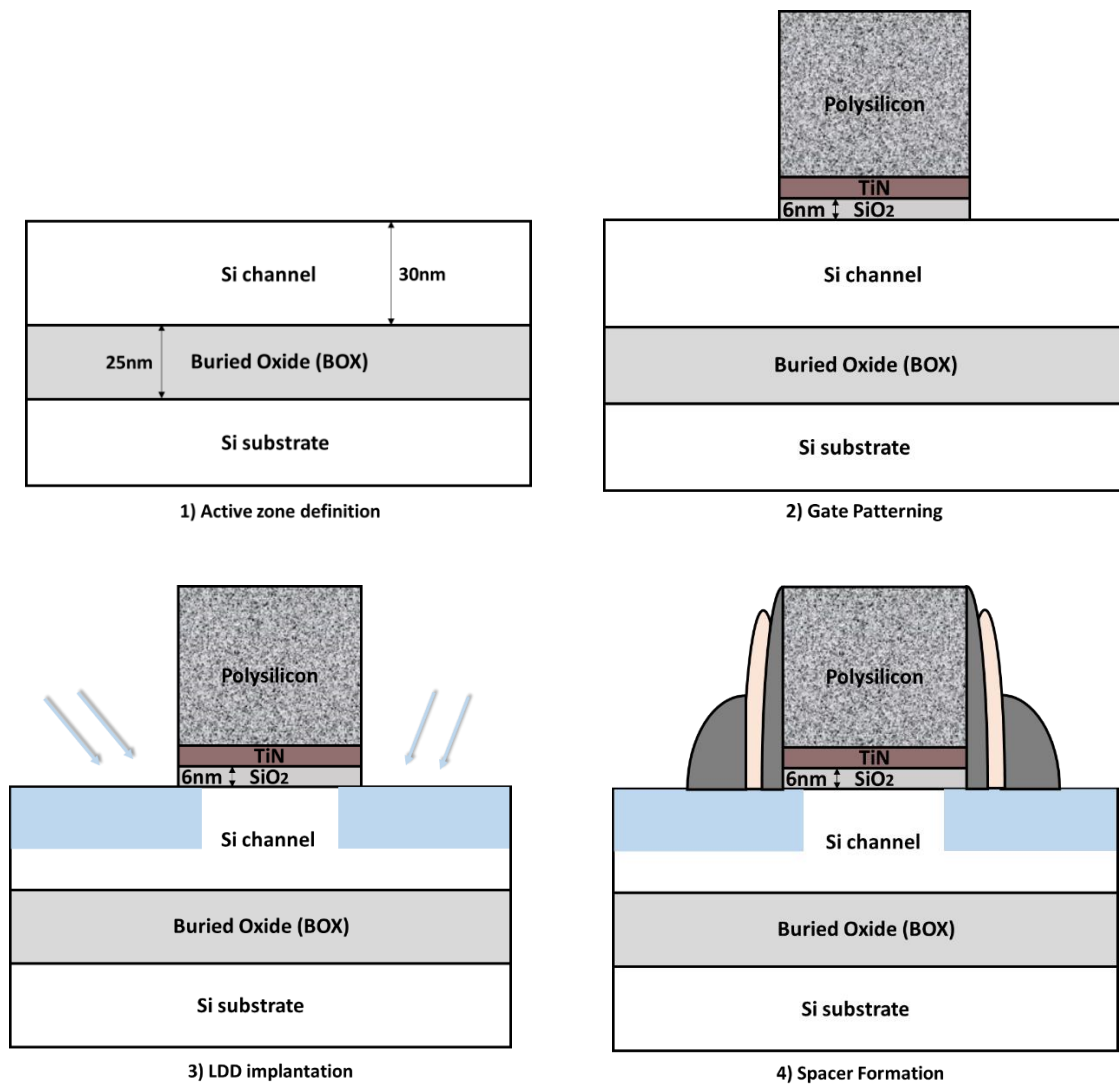


Figure 1.19 Back-bias efficiency obtained by TCAD electrical simulation showing that the proposed device is fully depleted.

The junction formation is composed of two implantation steps, a lower doped drain (LDD) and a high doped drain (HDD). The LDD is firstly implanted after the gate patterning, with dopants concentrations in the range of $5 \cdot 10^{18} \text{cm}^{-3}$. Due to its lower concentration value, different splits have been tested comprising a thermal activation at 500°C or a SPER activation. Also, tiled implants were used, in order to achieve a smooth non abrupt junction profile. The high-voltage spacers, separating the LDD from the HDD profile, are formed of a combination of ALD SiN 10nm / TeOS 10nm / Nitride 24nm. After that, the HDD implants are performed and activated by SPER. Lastly, the silicidation process is performed as for the 28 nm FDSOI technology with nickel platinum (NiPt) alloy, finishing by contact formation. The process flow is schematized on Fig. 1.20, and a transmission electron microscopy (TEM) cross section of the finalized device is provided.



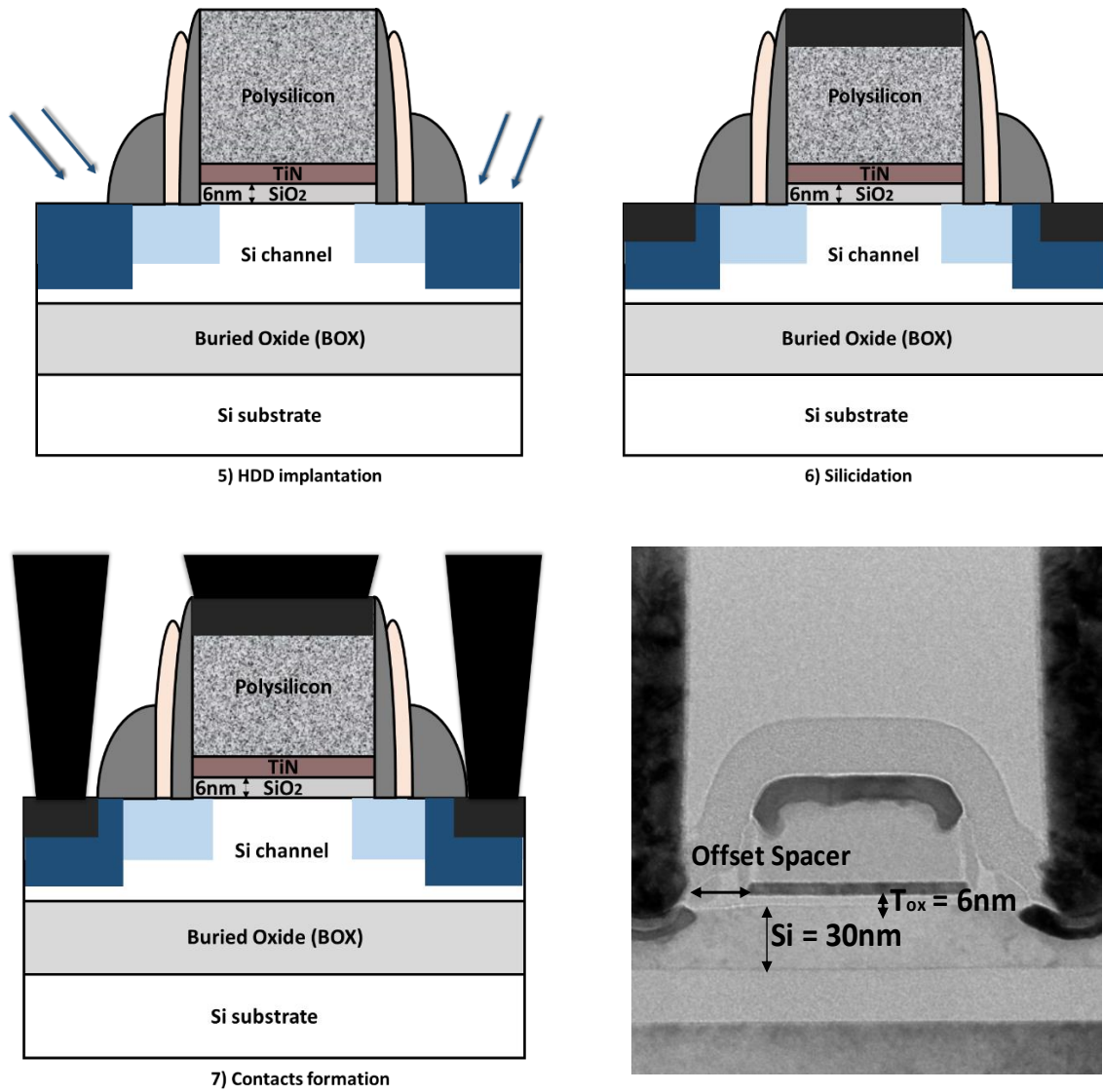


Figure 1.20 Schematized process flow of the proposed architecture for high voltage analog devices, and TEM cross section.

- [1] G. E. Moore, « Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp. 114 ff. », *IEEE Solid-State Circuits Soc. Newsl.*, vol. 20, n° 3, p. 33-35, 2006.
- [2] R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous, et A. R. Leblanc, « Design of ion-implanted MOSFET's with very small physical dimensions », *Proc. IEEE*, vol. 87, n° 4, p. 668-678, 1999.
- [3] I. Ferain, C. A. Colinge, et J.-P. Colinge, « Multigate transistors as the future of classical metal–oxide–semiconductor field-effect transistors », *Nature*, vol. 479, n° 7373, p. 310-316, nov. 2011, doi: 10.1038/nature10676.
- [4] S. E. Thompson *et al.*, « A 90-nm Logic Technology Featuring Strained-Silicon », *IEEE Trans. Electron Devices*, vol. 51, n° 11, p. 1790-1797, nov. 2004, doi: 10.1109/TED.2004.836648.
- [5] T. Skotnicki et F. Boeuf, « How can high mobility channel materials boost or degrade performance in advanced CMOS », in *2010 Symposium on VLSI Technology*, 2010, p. 153-154.
- [6] J. Robertson et R. M. Wallace, « High-K materials and metal gates for CMOS applications », *Mater. Sci. Eng. R Rep.*, vol. 88, p. 1-41, févr. 2015, doi: 10.1016/j.mser.2014.11.001.
- [7] K. Mistry *et al.*, « 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 *2007 IEEE International Electron Devices Meeting*, 2007, p. 247-250.
- [8] H.-J. Cho *et al.*, « Si FinFET based 10nm technology with multi Vt gate stack for low power and high performance applications », in *2016 IEEE Symposium on VLSI Technology*, 2016, p. 1-2.
- [9] S. Barraud *et al.*, « Performance of Omega-Shaped-Gate Silicon Nanowire MOSFET With Diameter Down to 8 nm », *IEEE Electron Device Lett.*, vol. 33, n° 11, p. 1526-1528, nov. 2012, doi: 10.1109/LED.2012.2212691.
- [10] M. Bruel, « Silicon on insulator material technology », *Electron. Lett.*, vol. 31, n° 14, p. 1201-1202, 1995.
- [11] Schwarzenbach et X. Cauchy, F. Boedt, O. Bonnin, E. Butaud, C. Girard, B.-Y. Nguyen, C. Mazure & C. Maleville, « Excellent Silicon Thickness Uniformity on Ultra-Thin SOI for controlling Vt variation of FDSOI », *IEEE Int. Conf. IC Des. Technol.*, 2011.
- [12] N. Planes, O.Weber*, V.Barral*, S.Haendler, D.Noblet, D.Croain, M.Bocat, P.-O.Sassoulas, X.Federspiel, A.Cros, A.Bajolet, E.Richard, B.Dumont, P.Perreau*,D.Petit, D.Golanski, C.Fenouillet-Béranger*, N.Guillot, M.Rafik, V.Huard, S.Puget, X.Montagner, M.-A.Jaud*, O.Rozeau*, O.Saxod, F.Wacquand, F.Monsieur, D.Barge, L.Pinzelli, M.Mellier, F.Boeuf, F.Arnaud and M.Haond, et Planes, « 28nm FDSOI Technology Platform for High-Speed Low-Voltage Digital Applications », *Symp. VLSI Technol.*, p. 133-134, 2012.
- [13] C. Gallon *et al.*, « Ultra-thin fully depleted SOI devices with thin BOX, ground plane and strained liner booster », in *2006 IEEE international SOI Conference Proceedings*, 2006, p. 17-18.
- [14] C. Fenouillet-Beranger *et al.*, « FDSOI devices with thin BOX and ground plane integration for 32nm node and below », *Solid-State Electron.*, vol. 53, n° 7, p. 730-734, 2009.
- [15] P. Batude, « Intégration à trois dimensions séquentielle: Etude, fabrication et caractérisation », Institut National Polytechnique de Grenoble-INPG, 2009.
- [16] P. Batude *et al.*, « Advances, challenges and opportunities in 3D CMOS sequential integration », in *Electron Devices Meeting (IEDM), 2011 IEEE International*, 2011, p. 7-3.
- [17] J. Jourdon *et al.*, « Hybrid bonding for 3D stacked image sensors: impact of pitch shrinkage on interconnect robustness », in *2018 IEEE International Electron Devices Meeting (IEDM)*, 2018, p. 7-3.

- [18] L. Brunet, P. Batude¹, C. Fenouillet-Beranger¹, P. Besombes¹, L. Hortemel¹, F. Ponthenier¹, B. Previtali¹, C. Tabone¹, A. Royer¹, C., Agrafeil¹, C. Euvrard-Colnat², A. Seignard¹, C. Morales¹, F. Fournel¹, L. Benaissa¹, T. Signamarcheix¹, P. Besson², M. Jourdan¹, R., Kachtouli¹, V. Benevent¹, J.-M. Hartmann¹, C. Comboroure², N. Allouti¹, N. Posseme¹, C. Vizioz¹, C. Arvet², S. Barnola¹, S. Kerdiles¹, L., et Baud¹, L. Pasini^{1,2}, C.-M. V. Lu^{1,2}, F. Deprat¹, A. Toffoli¹, G. Romano², C. Guedj¹, V. Delaye¹, F. Boeuf², O. Faynot¹ and M. Vinet¹, « First demonstration of a CMOS over CMOS 3D VLSI CoolCube™ integration on 300mm wafers », *IEEE Symp. VLSI Technol.*, 2016, doi: 10.1109/VLSIT.2016.7573428.
- [19] A. Vandooren, J. Franco, Z. Wu¹, B. Parvais², W. Li, L. Witters, A. Walke, L. Peng, V. Deshpande, N. Rassoul, G. Hellings, G. Jamieson, F. Inoue, K. Devriendt, L. Teugels, N. Heylen, E. Vecchio, T. Zheng, E. Rosseel, W. Vanherle, A. Hikavy, G., Mannaert, B. T. Chan, R. Ritzenhaler, J. Mitard, L. Ragnarsson, N. Waldron, V. De Heyn, S. Demuynck, J. Boemmels, D., et Mocuta, J. Ryckaert and N. Collaert., « First Demonstration of 3D stacked Finfets at a 45nm fin pitch and 110nm gate pitch technology on 300mm wafers », *IEEE Int. Electron Devices Meet. IEDM*, 2018, doi: 10.1109/IEDM.2018.8614654.
- [20] P. Batude, « Sequential Integration », présenté à 2019 IEDM Tutorials, 2019.
- [21] P. Batude *et al.*, « Advances in 3D CMOS sequential integration », in *Electron Devices Meeting (IEDM), 2009 IEEE International*, 2009, p. 1-4.
- [22] T. Ghani *et al.*, « A 90nm high volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors », in *IEEE International Electron Devices Meeting 2003*, Washington, DC, USA, 2003, p. 11.6.1-11.6.3, doi: 10.1109/IEDM.2003.1269442.
- [23] M. Sellier *et al.*, « Predictive Delay Evaluation on Emerging CMOS Technologies: A Simulation Framework », in *9th International Symposium on Quality Electronic Design (isqed 2008)*, San Jose, CA, USA, mars 2008, p. 492-497, doi: 10.1109/ISQED.2008.4479784.
- [24] Y.-H. Son *et al.*, « Laser-induced epitaxial growth (LEG) technology for high density 3-D stacked memory with high productivity », in *2007 IEEE Symposium on VLSI Technology*, 2007, p. 80-81.
- [25] V. Deshpande *et al.*, « Demonstration of 3-D SRAM Cell by 3-D Monolithic Integration of InGaAs n-FinFETs on FDSOI CMOS With Interlayer Contacts », *IEEE Trans. Electron Devices*, vol. 64, n° 11, p. 4503-4509, nov. 2017, doi: 10.1109/TED.2017.2755662.
- [26] T. Irisawa *et al.*, « Demonstration of ultimate CMOS based on 3D stacked InGaAs-OI/SGOI wire channel MOSFETs with independent back gate », in *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*, 2014, p. 1-2.
- [27] W. Rachmady *et al.*, « 300mm heterogeneous 3D integration of record performance layer transfer germanium PMOS with silicon NMOS for low power high performance logic applications », in *2019 IEEE International Electron Devices Meeting (IEDM)*, 2019, p. 29-7.
- [28] « 3D Sequential Integration: Application-driven technological achievements and guidelines ».
- [29] C. Fenouillet-Beranger *et al.*, « New insights on bottom layer thermal stability and laser annealing promises for high performance 3D VLSI », in *Electron Devices Meeting (IEDM), 2014 IEEE International*, 2014, p. 27-5.
- [30] C. Fenouillet-Beranger *et al.*, « FDSOI bottom MOSFETs stability versus top transistor thermal budget featuring 3D monolithic integration », *Solid-State Electron.*, vol. 113, p. 2-8, 2015.
- [31] L. J. Chen et Institution of Electrical Engineers, Éd., *Silicide technology for integrated circuits*. London: Institution of Electrical Engineers, 2004.
- [32] S. Kerdilès *et al.*, « (Invited) Sequential 3D Process Integration: Opportunities for Low Temperature Processing », *ECS Trans.*, vol. 80, n° 4, p. 215-225, août 2017, doi: 10.1149/08004.0215ecst.

- [33] B. Mathieu, C. Fenouillet-Beranger, S. Kerdiles, et J.-C. Barbe, « Thermal simulation of nanosecond laser annealing of 3D sequential VLSI », in *2015 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Washington DC, USA, sept. 2015, p. 104-107, doi: 10.1109/SISPAD.2015.7292269.
- [34] P. A. Alba *et al.*, « Nanosecond Laser Annealing for Phosphorous Activation in Ultra-Thin Implanted Silicon-On-Insulator Substrates », in *2016 21st International Conference on Ion Implantation Technology (IIT)*, 2016, p. 1-4.
- [35] X. Garros *et al.*, « RF Performance of a Fully Integrated 3D Sequential Technology », in *2019 IEEE International Electron Devices Meeting (IEDM)*, 2019, p. 25-1.
- [36] L. Pasini *et al.*, « High performance CMOS FDSOI devices activated at low temperature », in *2016 IEEE Symposium on VLSI Technology*, Honolulu, HI, USA, juin 2016, p. 1-2, doi: 10.1109/VLSIT.2016.7573407.
- [37] A. Tsiara *et al.*, « Performance and reliability of a fully integrated 3D sequential technology », in *2018 IEEE Symposium on VLSI Technology*, 2018, p. 75-76.
- [38] C.-M. Lu *et al.*, « Key process steps for high performance and reliable 3D Sequential Integration », in *VLSI Technology, 2017 Symposium on*, 2017, p. T226-T227.
- [39] V. Mazzocchi, J.-M. Hartmann, M. Veillerot, J.-B. Pin, et M. Bauer, « Benchmark of Disilane and Liquid Si for the Low Temperature Epitaxial Growth of Si, SiGe and SiGeB », *ECS Trans.*, vol. 86, n° 7, p. 177-188, juill. 2018, doi: 10.1149/08607.0177ecst.
- [40] W. Arden, M. Brillouët, P. Cogez, M. Graef, B. Huizing, et R. Mahnkopf, « More-than-Moore white paper », *Version*, vol. 2, p. 14, 2010.
- [41] « International Technology Roadmap for Semiconductors 2.0 », edition 2015.
- [42] M. M. Shulaker *et al.*, « Three-dimensional integration of nanotechnologies for computing and data storage on a single chip », *Nature*, vol. 547, n° 7661, p. 74-78, juill. 2017, doi: 10.1038/nature22994.
- [43] L. Millet *et al.*, « A 5500FPS 85GOPS/W 3D stacked BSI vision chip based on parallel in-focal-plane acquisition and processing », in *2018 IEEE Symposium on VLSI Circuits*, 2018, p. 245-246.
- [44] A. Theuwissen, « CMOS image sensors: State-of-the-art and future perspectives », in *Solid State Device Research Conference, 2007. ESSDERC 2007. 37th European*, 2007, p. 21-27.
- [45] H.-S. P. Wong, « CMOS image sensors-recent advances and device scaling considerations », in *International Electron Devices Meeting. IEDM Technical Digest*, 1997, p. 201-204.
- [46] T. Suzuki, « Challenges of image-sensor development », in *2010 IEEE International Solid-State Circuits Conference-(ISSCC)*, 2010, p. 27-30.
- [47] P. Coudrain *et al.*, « Setting up 3D sequential integration for back-illuminated CMOS image sensors with highly miniaturized pixels with low temperature fully depleted SOI transistors », in *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, 2008, p. 1-4.
- [48] P. R. Kinget, « Device mismatch and tradeoffs in the design of analog circuits », *IEEE J. Solid-State Circuits*, vol. 40, n° 6, p. 1212-1224, juin 2005, doi: 10.1109/JSSC.2005.848021.
- [49] R. Thewes, M. Brox, K. F. Goser, et W. Weber, « Hot-carrier degradation of p-MOSFET's under analog operation », *IEEE Trans. Electron Devices*, vol. 44, n° 4, p. 607-617, 1997.
- [50] B. El-Kareh et L. N. Hutter, *Silicon Analog Components*. New York, NY: Springer New York, 2015.
- [51] P. Jespers, *The gm/ID Methodology, A Sizing Tool for Low-voltage Analog CMOS Circuits*. Boston, MA: Springer US, 2010.

-
- [52] S. Wong et C. A. T. Salama, « Impact of scaling on MOS analog performance », *IEEE J. Solid-State Circuits*, vol. 18, n° 1, p. 106-114, 1983.
- [53] J.-P. Colinge, « Fully-depleted SOI CMOS for analog applications », *IEEE Trans. Electron Devices*, vol. 45, n° 5, p. 1010-1016, 1998.
- [54] D. Flandre *et al.*, « Fully-depleted SOI CMOS technology for low-voltage low-power mixed digital/analog/microwave circuits », *Analog Integr. Circuits Signal Process.*, vol. 21, n° 3, p. 213-228, 1999.
- [55] V. Passi et J.-P. Raskin, « Review on analog/radio frequency performance of advanced silicon MOSFETs », *Semicond. Sci. Technol.*, vol. 32, n° 12, p. 123004, déc. 2017, doi: 10.1088/1361-6641/aa9145.

Bottom tier thermal stability

3D Sequential Integration consists in stacking device layers in a sequential manner. This integration offers the highest 3D contact density but comes at the cost of thermal budget constraints for the top layer processing, in order to preserve the integrity of the underlying tiers. However, in order to fully benefit from the 3D opportunities, inter-tier metal layers need to be incorporated in the technology. The latter raises new technological challenges, because the introduction of an intermediate back end of line implies that the thermal budget for the top device fabrication also needs to consider the interconnects layer.

Concerning the More Moore applications, the thermal stability of 28nm FDSOI transistors has been set up to 500°C. The latter has been defined by analyzing the static performance Figure of Merit, while the Cu/ULK 28nm BEOL stability limit was set at 500°C through the RC delay and reliability measurements at 25°C. In this work, for the first time, the thermal stability of 28nm devices with industry relevant FOM, such as device performance under dynamic operation and the yield of SRAM cuts and logic gate array (5Mbit dense SRAM and 1 Million Flip-Flops). The stability of 28nm Cu/ULK BEOL process is also evaluated with yield and reliability measurements (performed at 25°C and 125°C). Additionally, the immunity of bottom device and interconnects under ultra-short anneals using Ultra-Violet Nanosecond Laser Annealing is addressed.

3D sequential integration is also a great candidate for More than Moore applications. In this case, high voltage analog devices ($V_{DD}=2.5-5V$) are required. Seeing that there are many discrepancies in the architecture of this high voltage devices w.r.t. the digital one previously studied, it is then desired to analyze its thermal stability. Therefore, this chapter shall also evaluate the maximum thermal budget that can be sustained by high voltage MOSFETs to be used as bottom transistors in a 3DSI.

2.1 BOTTOM TIER STABILITY: WHAT HAS BEEN DONE

One of the main challenges on a 3D sequential integration is that the thermal budget of the upper-level transistor must be limited in order to preserve the integrity of the underlying tiers, hence both device and interconnects.

In order to define the safe thermal budget window for top tier processing, different thermal budget splits are applied with the goal to simulate the top tier fabrication. At that point, one needs to discriminate between its impact on the front-end-of-line or back-end-of-line, which can be done by applying this additional thermal budget at a certain step of the process flow. For example, if the additional thermal budget is applied after CMP PMD (chemical mechanical polishing and pre-metal dielectric), thus impacting device only, or after the metal lines. In the latter, both device and interconnects are concerned, where a schematic is provided on Fig. 2.1. These studies are important because it allows the identification of what are the most critical parameters impacted by the top tier fabrication, allowing one to either (1) improve the thermal stability of the bottom tier in order to safely withstand this TB or (2) to develop/adapt suitable technological processes for the top tier.

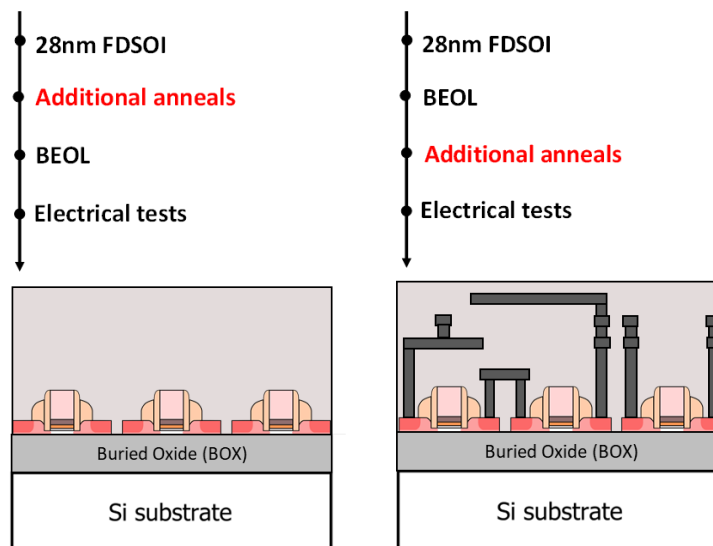


Figure 2.1 Schematic representation highlighting the difference between studies aiming to address device stability (on the left) or both device and interconnects (on the right). It must be said that even though a 28nm FDSOI is used as an example, other technological nodes can and should be employed, and shall be described later in this chapter.

Thus, this section aims at giving an overview of the previous studies on the subject and to identify which are the next steps to be taken in this regard.

2.1.1 FEOL stability under standard anneal

Concerning the thermal stability of the device only, the bottom level stability of 28nm FDSOI transistors ($V_{DD}=0.9V-1V$) has been previously defined [1] using standard furnace annealing techniques. In that work, temperatures ranging from 450° up to 650°C were applied at different times, in the minute and hour range. The additional thermal anneals have been performed after the CMP PMD step just before contacts processing, hence once more, impacting the device characteristics only. Thanks to electrical characterization, the transistor's performance has been analyzed through its static figure of merit.

It has been highlighted that the device performance integrity of both N&PMOS has been maintained for thermal budgets not exceeding the 500°C range. On the other hand, for thermal budgets higher than that, (in the 550°C range and up), some I_{ON} performance losses have been observed. Furthermore, this degradation has been attributed to an access resistance degradation. This degradation has then been ascribed to two different contributions: 1) a deactivation of dopants (for NMOS devices where the access is doped with Phosphorus and Arsenic) and the 2) degradation of the silicide ($Ni_{0.9\%}Pt_{0.1\%}$).

In order to understand this silicide degradation, the latter was analyzed through square resistance measurements of the silicided and non-silicided zones, together with TEM images as shown on Fig. 2.2-a.b, which allowed the identification of the elements responsible for these thermal instabilities. As a conclusion, the silicide morphology can be impacted by an additional thermal budget and can most likely be attributed to silicide agglomeration or phase changing. More details on the silicide thermal stability shall be discussed further on.

Overall, it has been stressed that the silicide is a limiting factor for the thermal stability of the bottom device. In this regard, for the 28 nm node, the maximum thermal budget for the upper-level processing is then set at 500°C in the hour range. More specifically, this thermal budget can slightly change depending on the type of silicide used.

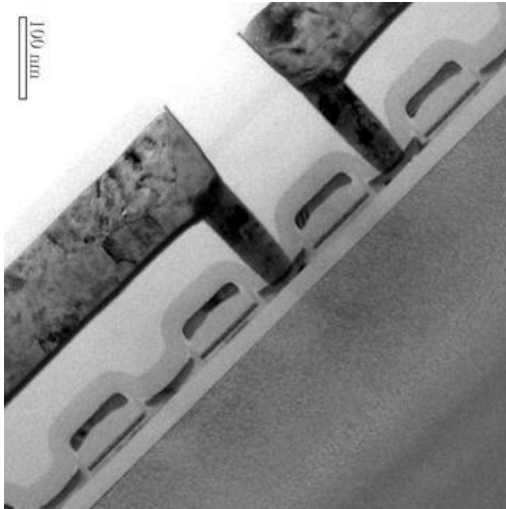


Figure 2.2-a TEM cross section of a 28nm design rules device for the reference wafer without post anneal.



Figure 2.2-b TEM cross section of a 28nm design rules device for the wafer with post anneals 600°C + 650°C 30min (on the right). Various silicide morphological degradations are observed [1].

2.1.1.1 Motivations for this work

As previously mentioned, the stability of advanced digital bottom devices ($V_{DD}=0.9-1V$) has been validated up to 500°C, a thermal budget that is sufficient to process the top high performance digital devices. Even though those studies gave a good indication of the maximum allowable thermal budget, to have a more complete processing window, additional studies are needed. For example, by analyzing the device under dynamic operation in complement to static, further insights on the transistor's stability can be drawn. Another way to enrich these studies is by analyzing the thermal stability of these 28nm devices with industry relevant FOM, such as with the yield of SRAM cuts and logic gate arrays (5Mbit dense SRAM and 1 Million Flip-Flops). This is because by taking these FOM, one comes closer to the real operation of the transistor. That said, one part of this work aims at further extending the analysis of the thermal device stability for More Moore application with 28nm digital devices.

The other part of the work shall focus on the devices used for More than Moore applications, consisting of the stacking of analog on digital devices. As said on chapter one, these devices have a higher operating voltage (which can range from 2.5V-5V), and consequently, a different architecture. So far, it has become clear that the maximum allowable thermal budget is extremely dependent on the device architecture, so if a different device is used, it is preferable to perform a thermal stability study targeting this

device. As such, the thermal stability of transistors for More than Moore applications shall also be evaluated.

To conclude, Fig. 2.3 summarizes the motivations and what shall be discussed on later sections.

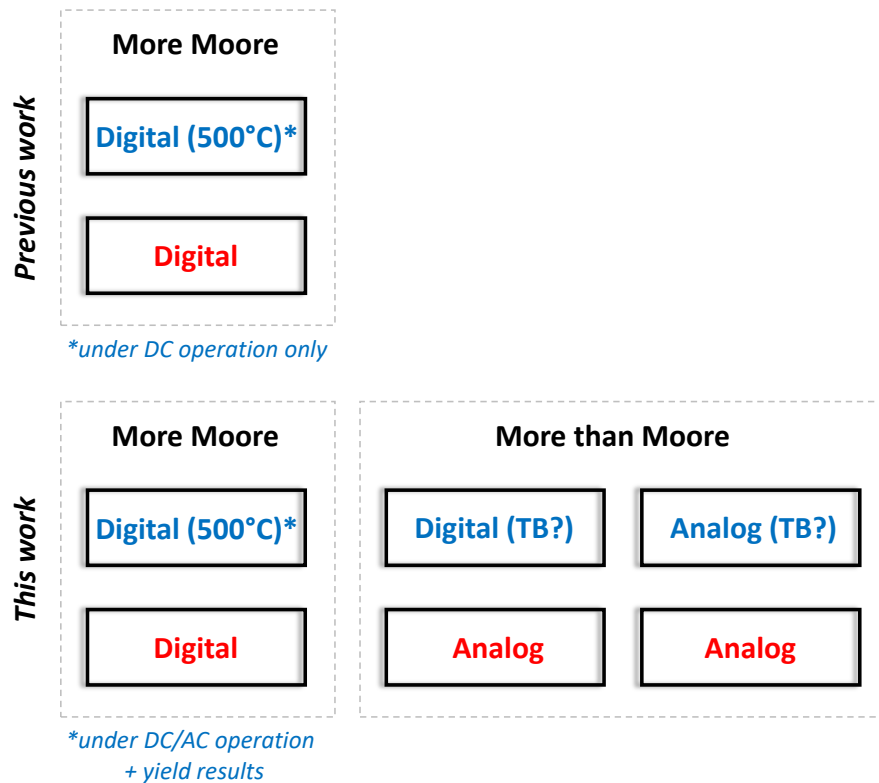


Figure 2.3 Description of this work objectives: complement the thermal budget processing window with industry relevant FOM for More Moore applications, and to quantify the maximum thermal sustainable by high voltage analog devices for More than Moore applications.

2.1.2 Intermediate BEOL stability under standard anneal

As previously stated on chapter 1, 3DSI allows partitioning at the higher granularity level, thus making CMOS over CMOS stacking an option. In addition, in order to fully benefit from the 3D opportunities and avoid global routing congestion, inter-tier metal layers need to be incorporated in the technology. However, the latter raises new technological challenges. This is because the introduction of an intermediate back-end-of-line implies that the TB for the top device fabrication also needs to consider the stability of the interconnects layer below. Additionally, although not the scope of this study, it should also be mentioned that another challenge concerning the iBEOL introduction is that it can also

lead to contamination issues, because the top FEOL must be processed after bottom BEOL, thus risking contaminating the tool. More details on that can be found elsewhere [2].

Focusing on the thermal stability only, it is now clear that in addition to the bottom device, the stability of the iBEOL must also be evaluated.

Overall, the BEOL consists of metal lines connecting the transistors to each other, isolated by layers of dielectric materials. It is built on several levels of routing and allow the integration of complex functions on a chip. An interconnection level generally consists of lines, vias and a stack of dielectrics. The interconnects can be composed of different combinations of metals and dielectrics, with the goal of minimizing the propagation delay. This delay is proportional to the R*C product, with R related to the line resistance and C to the line parasitic capacitances, as depicted on the equations below:

$$R = \frac{\rho L}{WH} \quad (\text{Equation 2.1})$$

$$C = \epsilon_0 \epsilon_r \frac{LH}{S} \quad (\text{Equation 2.2})$$

From Eq. 2.1, ρ is the resistance of the material ($\Omega \cdot m$), and L, H and W the materials dimensions, length, height, and width, respectively (m). Concerning Eq. 2.2, ϵ_0 is the vacuum permittivity ($8.85 \cdot 10^{-12} \text{Fm}^{-1}$), ϵ_r is the dielectric permittivity and S the distance between two metallic lines.

In a simplified way, the intrinsic properties of the materials directly impact the R*C product and needs to be reduced in order to overcome the signal propagation delay in the interconnects (particularly important for more advanced nodes) [3]. The delay can be improved by choosing appropriate materials, such as a less resistive metal (that at the same time is more resistant to electromigration), associated with an insulator with a lower dielectric constant. As it can be seen on Fig. 2.4, different materials combinations can be used, with the goal to achieve the best compromise of lower resistance and capacitance.

When considering Cu to be used in a 3DSI, despite its advantages concerning the lower R, the latter is prohibited in FEOL and poses the risk of contamination during the manufacture of the components, for example if the wafer breaks inside a FEOL tool. Therefore, tungsten (W), although having a higher resistivity than that of Cu, was also evaluated for the intermediate BEOL, the latter being less problematic in terms of

contamination. It has been highlighted that W interconnects combined with ULK are stable up to 550°C 5h respectively [4]. Despite its advantages and relatively high thermal stability, the best RC compromise is still obtained with the Cu when paired with insulating materials with increasingly lower values of dielectric constant (i.e. low-k).

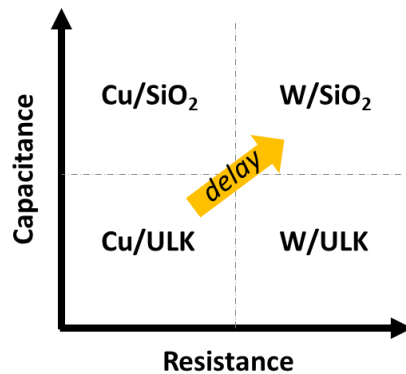


Figure 2.4 Schematic representation adapted from [5] depicts the impact on the RC delay for interconnects with different materials. Best trade-off achieved with Cu/ULK. Note that in this example, SiOCH is used as ULK.

As a result, this work shall focus on the standard 28nm design rules damascene BEOL integration [6], which is composed of copper for the metallic lines coupled with a low permittivity oxide such as a porous ULK material (SiOCH with $k \sim 2.7$). In addition to that, the copper filling is usually preceded by the deposition of a diffusion barrier, typically Ta-TaN [7]. The latter is usually employed to prevent degradation of the interconnects due to Cu diffusion.

After having described the interconnects, it is now important to understand the implications of choosing the Cu/ULK and how an additional thermal budget can have an impact on those.

Concerning the Cu/ULK line stability upon standard anneals, previous results have shown through material analysis, that the lines are stable up to 500°C 2h [8]. As observed on Fig. 2.5, even though there is no Cu resistance increase upon the additional thermal budget, the ULK integrity does not behave the same, where a capacitance increase of 5-7% has been observed. This behavior has been attributed to a shrinking of the porous SiOCH and a slight change in its permittivity, seen even for the lower thermal budget of 450°C 2h. It is also interesting to mention that SiO₂ is indeed more thermally stable and could be used instead the ULK, the drawback being its higher permittivity ($k \sim 3.9$).

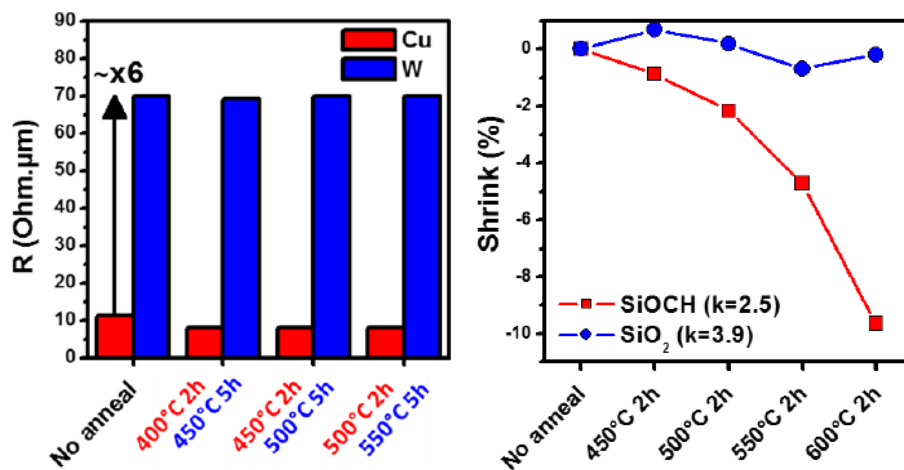


Figure 2.5 Cu line resistance depicted on the left, and ULK shrink % shown on the right. Despite the R being stable, the porous ULK can suffer from change in properties, does affecting the line capacitance [9].

Other than the RC analysis and materials characterization, a particularly important way to address the interconnects stability is to perform failure tests, such as reliability. The reliability is one of the main problems encountered in the development of Cu/ULK technology for several reasons that shall be shortly summarized.

When using a low- k material, it generally presents a weaker intrinsic breakdown strength than traditional SiO₂ dielectrics. This problem is further exacerbated by the aggressive shrinking of the interconnect pitch size due to continuous technology scaling [10]. In addition to that, it was found that the time dependent dielectric breakdown (TDDB) of the SiCOH was sensitive to all aspects of integration. As an example, due to unoptimized process steps, the latter can be exposed to moisture during certain process steps thus resulting in leakage current and k -value increase.

Other than that, failure mechanisms such as diffusion of Cu ions into the dielectric must also be considered. This is because the migration of Cu could lead to defects in the porous ULK material during the application of an electrical stress [11]. In particular, where the application of a thermal budget can promote the Cu ion migration in the dielectric [12], especially problematic for the integration of Cu in the iBEOL. The degradation of the reliability can also be linked to the integrity of the Ta-TaN barrier, which could potentially block the Cu diffusion. In this case, the premature breakdown could then be attributed to Cu already present in the dielectric [13]. Lastly, the Ta-TaN barrier can also diffuse into the dielectric, also leading to negative implications in reliability [14] [15]. Overall, the degradation of the reliability in the case of the Cu/ULK pair (with a Ta-TaN barrier) is a complex phenomenon where more information can be found elsewhere [5], [16]. It must

also be stressed that the reliability is not a concern only in the case of 3DSI, and planar technologies are also impacted due to continuous technology scaling.

A recent work has analyzed the reliability of the Cu/ULK interconnections through the time to breakdown, with measurements performed at the M1 level [16]. These measurements were carried out at a temperature of 25°C. It has been observed that the interconnections are stable up to 500°C 2h, as shown on the evolution of the Weibull distribution according to the thermal budgets applied (Fig. 2.6).

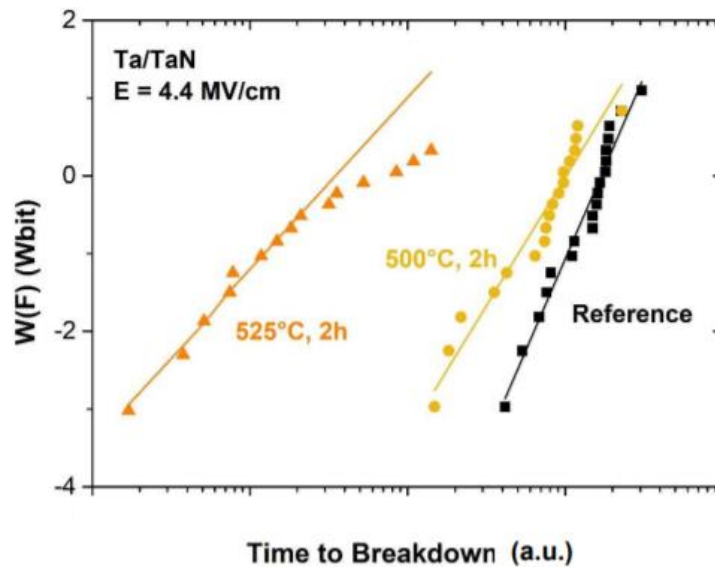


Figure 2.6 Time to Breakdown performed on Cu/ULK, stable up to 500°C 2h [16].

In order to have more insight about the observed degradation in reliability, morphology analyzes were also performed and are depicted on Fig. 2.7 for the different thermal budget splits. For the reference wafer, TEM analysis have been correlated with chemical composition analyzes by Energy-dispersive X-ray spectroscopy (EDX), showing no diffusion of Cu or Ta at the M1 or M4 levels. For the 500°C 2h split, the TEM-EDX analyzes did not clearly demonstrate the presence of defects originating from the barrier. Despite, it was observed that the diffusion barrier was slightly damaged.

For the 525°C 2h split, significant Ta and Cu diffusion into the dielectric has been highlighted due to a degradation of the diffusion barrier upon this additional thermal budget split. These results are thus consistent with the Time to Breakdown shown on Fig. 2.6.

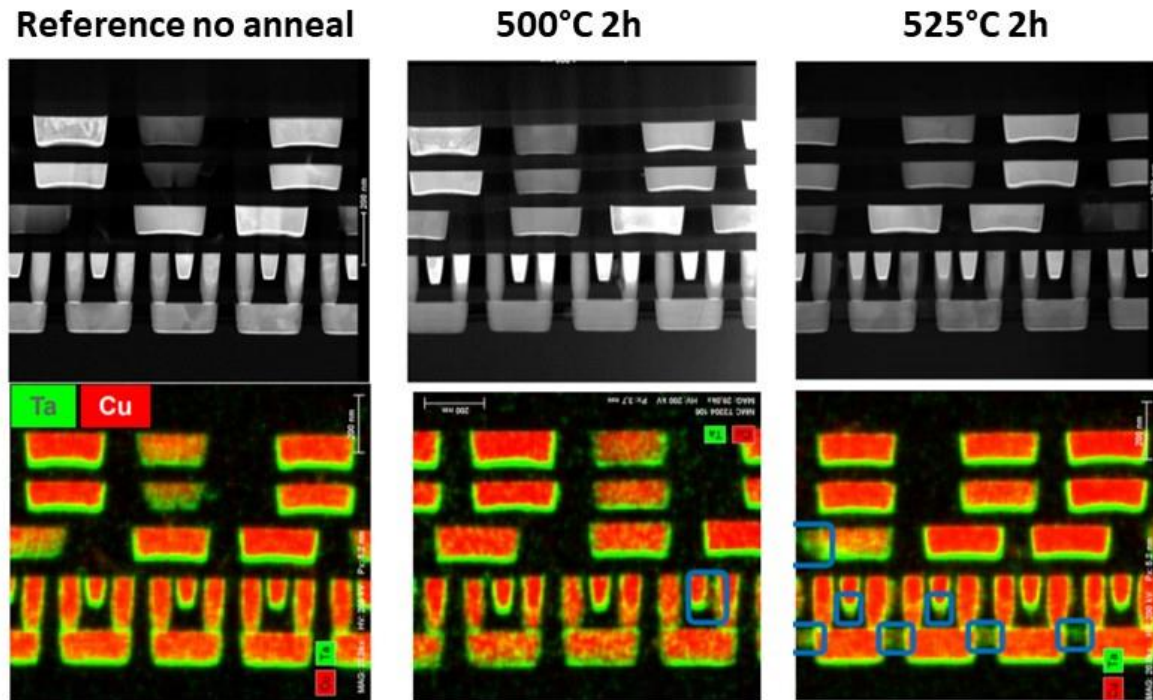


Figure 2.7 Cu/ULK interconnects morphology: the top row depicts the TEM images of 4 ULK/TaN-Ta/Cu metal levels observed before and after thermal annealing (500°C 2h and 525°C 2h). The EDX analysis shown on the bottom row highlights that diffusion in the dielectric of Ta and Cu are difficult to observe at 500 °C 2 h, but significant from 525C 2h.

2.1.2.1 Motivations for this work

As it has been previously highlighted, in view of a 3DSI, there is a degradation of the Ta-TaN barrier and the diffusion of Cu and Ta in the dielectric has been observed for the highest TB of 525°C 2h. Despite, the morphological analyzes for the split at 500°C 2h were not able to explain the dielectric breakdown degradation w.r.t. the reference wafer. Consequently, there is still the need to study the Cu/ULK pair and have a better understanding about the breakdown phenomena. In order to complement the previous reliability studies, the stability of an improved (w.r.t. the previous results [1]) 28nm Cu/ULK BEOL process shall be evaluated. In addition, the reliability measurements shall be performed at two different test temperatures and structures, described with more details in a later section. Lastly, the findings shall be supported with yield analysis.

	Previous work	This work
Reliability test temperature	25°C	25°C 125°C
Structure	M1 Line to Line	M1 Line to Line Via to M2 Line
Conclusion	stable at 500°C 2h	?

Table 1 Comparison between the test conditions exploited on [16] and the ones used in this work.

2.1.3 FEOL and iBEOL stability under ultra-short anneal

So far, the thermal stability has been mentioned only for classical non-selective annealing techniques such as furnaces, in the minute and/or hour range. In this regard, the overall thermal budget for the top transistor level should not exceed 500°C for a few hours.

However, as it has been said on chapter 1, a thermal budget corresponds to both a temperature AND duration, meaning that some annealing with a very high temperature and very short time duration could potentially be used. This is the case for laser annealing, with time durations in the millisecond and nanosecond range. This kind of thermal process is interesting because it leads to a temperature gradient between top surface and bottom layers and can eventually enable a local anneal. This means that the top tier could benefit from an extremely high temperature treatment, whereas the bottom tiers could be kept at a lower (and acceptable) thermal budget.

The principle of a laser treatment consists of the surface being exposed to a pulsed laser beam, where a high temperature anneal is achieved due to the extremely fast heating of the surface region during the pulse. In addition, this is known as a selective anneal, where only the surface is heated, not the volume of the structure [17]. This can be explained partially by its short wavelength, thus limiting the surface's absorption. Additionally, the short pulse limits in depth heat diffusion.

One example of selective heating consists of a laser anneal on the millisecond range (ms), achieved with a DSA, where a TB of 800°C 0.5ms has been applied on 14nm design rules (Fig. 2.8). The latter, applied after the silicide (NiPt) formation, showed negligible performance modification, as observed on I_{EFF}/I_{OFF} trade-off [18].

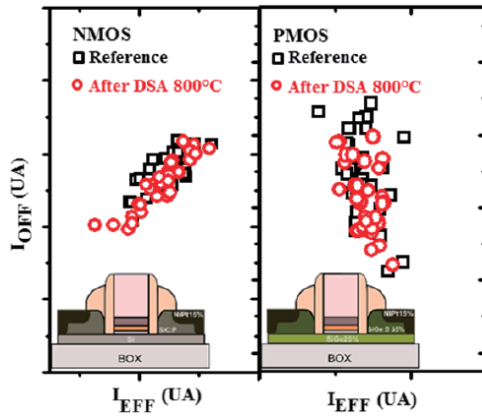


Figure 2.8 Influence of laser DSA (ms) anneal at 800°C after NiPtSi silicide formation on 14nm design rules.

Another interesting option is to use a UV Nanosecond Laser Annealing (UV-NLA $\lambda=308\text{nm}$ with a pulse of 160ns) developed by LASSE, which provides local heating while keeping the underlying tiers at a low temperature [19], [20]. Overall, the short annealing durations limits the diffusion of thermal energy from the annealed surface to the lower layers, potentially making it possible to reach temperatures above 1000 °C for the upper layer whilst the bottom tiers integrity are retained. Previous studies have suggested this kind of treatment for the S/D junction activation [21], [22]. In these works, the goal was to find a process window for the laser energy density which is large enough to reliably activate all junction types and to address its impact on the bottom tiers.

2D TCAD simulations have been performed to determine the best process structure. As a result, Fig. 2.9 depicts the best combination of oxide layer thicknesses, where the goal is that the gate temperature of the lower device to not exceed 600°C during 200ns. This can be achieved by increasing the interlevel oxide thickness from 50nm up to 120nm and reducing the bottom BOX from 145nm down to 20nm (with laser pulse tuned to heat the top transistor at 1200°C (melting point)).

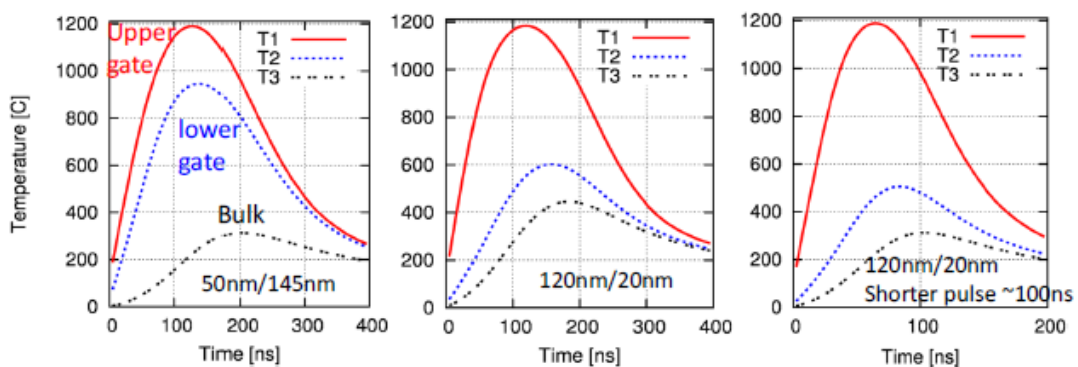


Figure 2.9 Temperature versus time for several interlevel oxide and bottom oxide layer. Best case 120nm upper/20nm lower with 308nm wavelength laser and ~200ns pulse duration. T1: T° in the gate stack of the upper layer, T2: T° in the gate stack of the lower layer, T3: T° at top of bulk silicon layer [22].

Likewise, the bottom tier thermal stability must also include the interconnections. As such, the stability of Cu/ULK inter-tiers interconnections have been addressed upon an UV-NLA treatment [23]. Those results analyzed its impact on the RC delay, with tests performed on the M1 line resistance measured on a specific multi-fingers/serpentine test structure before and after the laser annealing (Fig. 2.10).

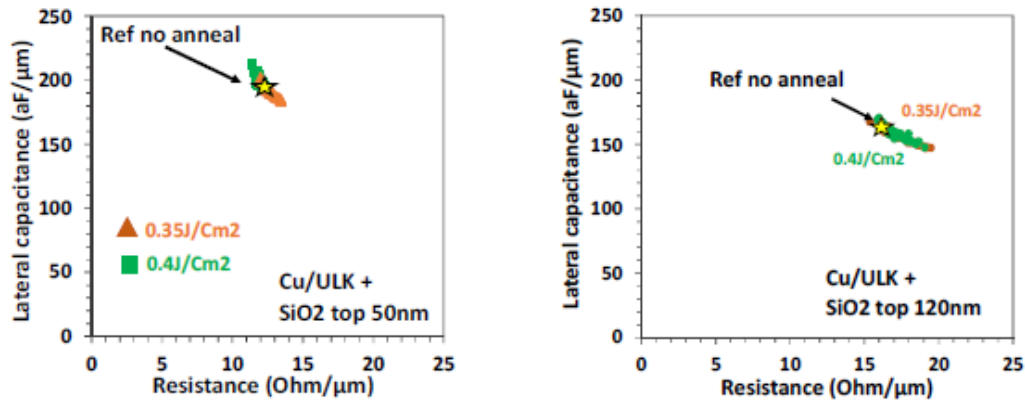


Figure 2.10 Lateral capacitance versus resistance for Cu/ULK interconnection before and after laser annealing. (left) SiO₂ interlevel is 50nm (right) 120nm [23].

No RC modification has been observed for different SiO₂ inter-layers thicknesses for a laser energy density of 0.4J/cm² (even beyond the melt threshold of the a-Si (~1150°C)). Based on these results, the laser impact on the thermal stability of Cu/ULK inter-tiers interconnections has been validated for a 28nm node back-end metal 1 design rules technology both from morphological and electrical perspectives. The interest of a UV-NLA has been previously mentioned on an earlier section, and its viability is now further confirmed concerning BEOL integrity.

2.1.3.1 Motivations for this work

In addition to dopant activation of the S/D regions, it has been reported that UV-NLA is also a promising candidate for the gate crystallization of the top tier, where very low gate resistance values are achieved [24]. As it has been previously mentioned, the heat diffusion depends on the material on which the laser is being applied and, on its thickness. Also, what is below (such as the patterning marks) can have an impact on the uniformity of the annealing.

For this reason, a process window for the laser conditions as well as the best structure design must be found for the case where the laser is being used for the top gate crystallization. As a result, the immunity of the device and interconnections under ultra-

short anneals using UV-NLA shall be evaluated through the Cu resistivity, thermal simulation analysis and the electrical characterization of the device.

2.2 FEOL STABILITY UNDER STANDARD ANNEAL FOR MORE THAN MOORE APPLICATIONS

This section aims at describing the architecture of the studied device, thus highlighting the differences between the previous studies. Furthermore, its performance stability upon an additional thermal budget shall be analyzed, with the aim to evaluate the maximum thermal budget that can be sustained by high voltage ($V_{DD}=2.5-5V$) MOSFETs to be used as bottom transistors of a 3DSI.

As previously mentioned, the devices used in this work targeting More than Moore applications present a quite different architecture w.r.t the advanced digital devices previously employed for More Moore applications. These discrepancies are summarized on Fig 2.11. It must be noted that the devices used for More than Moore shall be called “high voltage” devices, due to its higher operational voltage ($V_{DD}=5V$).

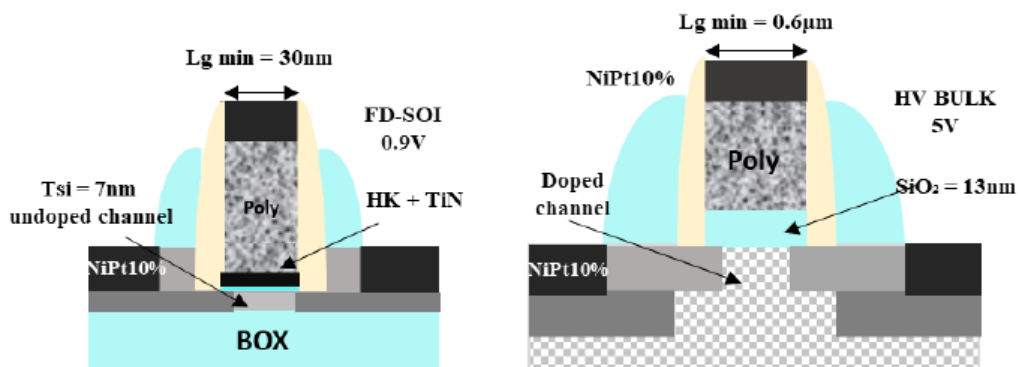


Figure 2.11 Schematic representation highlighting the architecture differences between advanced digital devices, previously analyzed on [1], and the HV device under study, adapted for More than Moore applications (such as sensing, described on chapter 1). Advanced digital devices ($V_{DD}=0.9-1V$) use FDSOI technology with $L_{G,min}=30nm$, undoped channel ($T_{Si}=7nm$) and HfSiON/TiN/Polysilicon gate stack. Meanwhile, the HV MOSFET under study presents a doped BULK channel, SiO_2 /Poly gate stack ($EOT=13nm$) and $L_{G,min}=600nm$.

In order to sustain this higher V_{DD} (as opposed to $V_{DD}=1V$ of advanced digital devices), the device’s dimensions are more relaxed. For example, its nominal gate length is $L_{G,min}=600nm$ instead of $L_{G,min}=30nm$. Likewise, the gate stack is composed of a thick $SiO_2=13nm$ + Polysilicon, instead of the high-k/TiN/Polysilicon of the device previously studied. Another discrepancy is that the high voltage are BULK devices with a doped channel, whereas the one for high performance uses the FDSOI technology with a thin

($T_{Si}=7\text{nm}$) undoped channel. All of the aforementioned differences make that the HV device can behave differently after an additional post-processing thermal budget, and these points shall be addressed in the next section.

In order to analyze the HV device thermal stability, additional thermal budgets (ranging from 400°C to 600°C 2h) has been applied to SOTA BULK 5V devices, after the CMP PMD step just before contacts processing. As a result, only the device stability shall be investigated. No further modifications have been done in the rest of the flow (Fig. 2.12).

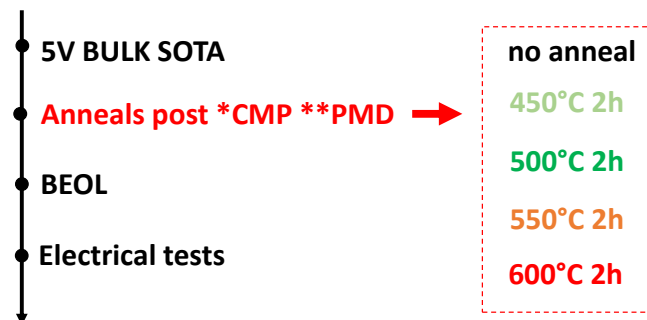


Figure 2.12 Process flow and anneal splits applied after the CMP PMD step on 5V Bulk SOTA devices. *Chemical-Mechanical Polishing **and Pre-metal Dielectric

2.2.1 Gate stack stability

One of the main motivations to study the stability of the HV devices concerns the gate stack integrity. As such, the threshold voltage for both N&PMOS after the additional thermal budget splits are shown on Fig. 2.13, for $L_G=1\mu\text{m}$. The V_{TH} was extracted by constant current method for low V_D ($V_D=100\text{mV}$), at the given current value ($I_{th}=100\text{nA}\cdot W/L$). In both cases, the V_{TH} is stable up to 550°C 2h. However, a slight V_{TH} shift of $+45\text{mV}$ and $+60\text{mV}$ starts to be observed at 600°C 2h, for N&PMOS respectively, and it needs to be understood.

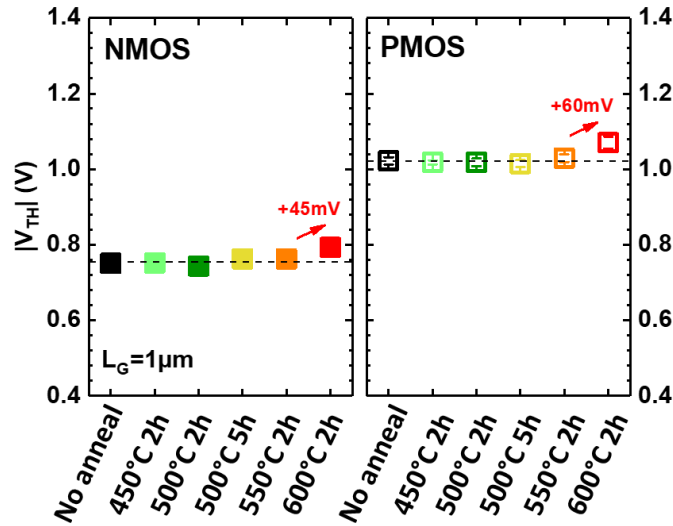


Figure 2.13 Linear threshold voltage of NMOS and PMOS extracted by constant current method [25] versus thermal budget split.

In the case of the planar bulk technology, the threshold voltage is defined by:

$$V_{TH} = V_{FB} + 2\phi_f + \left(\frac{Q_{dep}}{C_{ox}}\right) \quad (\text{Equation 2.3})$$

Q_{dep} is the depletion layer charge, C_{ox} is the oxide capacitance, ϕ_f is the fermi potential and V_{FB} is the flat-band voltage. Considering that there is no charge present in the oxide or at the oxide-semiconductor interface, the flat-band voltage simply equals the difference between the gate metal work-function, ϕ_M , and the semiconductor work-function, ϕ_S :

$$V_{FB} = \phi_{MS} = \phi_M - \phi_S \quad (\text{Equation 2.4})$$

However, in this study, one must be reminded that there is no metal (TiN) in the gate stack. Hence, for MOS structures with a highly doped polysilicon gate, the work-function of the gate based on the bulk potential of the poly-silicon is given below for a Polysilicon of p-type (for PMOS) and N-type (for NMOS), respectively:

$$\phi_{poly,S} = \phi_{poly} - \phi_S = V_t \ln\left(\frac{N_{a,poly}}{N_a}\right) \quad (\text{Equation 2.5})$$

$$\phi_{poly,S} = \phi_{poly} - \phi_S = V_t \ln\left(\frac{n_i^2}{N_{d,poly} N_a}\right) \quad (\text{Equation 2.6})$$

Where N_a corresponds to the substrate doping, $N_{a,poly}$ and $N_{d,poly}$ are the acceptor and donor density of the p-type and n-type poly-silicon gate respectively, and V_t stands for the thermal voltage at 300K = $kT/q = 0.026V$.

Similarly, since the device under analysis is on BULK technology with a doped channel, the work function of the semiconductor ϕ_s depends on the doping type and concentration. Taking all of it into account, the flat-band voltage can then be expressed as:

$$V_{FB,nmos} = \phi_{poly,N} - \chi - \frac{E_g}{2} - V_t \ln \frac{N_a}{n_i} \quad (\text{Equation 2.7})$$

$$V_{FB,pmos} = \phi_{poly,P} - \chi - \frac{E_g}{2} + V_t \ln \frac{N_d}{n_i} \quad (\text{Equation 2.8})$$

The equations above are given for a NMOS (with Poly-N and p-type substrate doping) and PMOS device (with a Poly-P and n-type substrate doping), respectively. From these equations, it is possible to conclude that the dopant concentration in both the polysilicon gate and the doped channel could potentially be sources of thermal instability upon an additional thermal budget, and its impact on the threshold voltage shall be analyzed separately.

Channel doping integrity

Once more, while the high-performance devices previously studied used the FDSOI technology, the HV device under study is fabricated on Bulk. Contrarily to FDSOI, the channel of Bulk transistor is inevitably doped.

Concerning the integrity of the doped channel of BULK devices, two different effects could lead to threshold voltage shifts. In the absence of TiN, the additional annealing steps could eventually lead to a diffusion of dopants from the heavily doped polysilicon into the channel. These diffused dopants thus result in a counter-doping of the silicon channel. This effect is more commonly observed for SiO₂/Polysilicon stacks, because for gate stacks with TiN integrated, the latter acts as a diffusion barrier for dopants. It must also be mentioned that a nitridation can be applied to also act as a diffusion barrier, but this technological option was not employed on the studied device. In addition, the counter-doping has especially been observed on PMOS devices, because boron is known to diffuse significantly faster especially in polysilicon [26], suggesting an enhanced diffusivity within or near the grain boundaries [27], [28]. However, in order to reach the channel, these dopants would also need to diffuse through the silicon dioxide, which has a lower diffusivity coefficient than the polysilicon. Lastly, it must be mentioned that this effect is more observed in the case of thin gate oxides, meaning that it would be harder to have diffusion through an oxide of 13nm.

Another key point is that in case of counter-doping, the channel is polluted by the dopants of the polysilicon gate, which have an opposite type of the dopants in the channel. For example, the counter-doping of NMOS devices would lead to an addition of N-type dopants (from the polysilicon) diffusing in the P-type channel. Its effect on the threshold voltage is opposite to what has been observed, meaning that a counter-doping lowers the V_{TH} . Given all the previous arguments, it is believed that a counter-doping did not occur in this device.

Another explanation could come from the fact that the additional thermal budget could also lead to a dopant deactivation. The latter can occur in different parts of the device architecture, but one shall focus first on the dopant deactivation in the silicon channel. However, once more, this would lead to a threshold voltage decrease, hence still not explaining the observed shift. Nevertheless, the channel doping integrity was analyzed by extracting the channel doping concentration with a back-bias sensitivity method [29]. The method consists of finding the equivalent change in depletion capacitance for the given V_{TH} shift obtained with a back-bias. The body-effect in BULK devices is given in the equation below:

$$\frac{\Delta V_{TH}}{\Delta V_{BB}} = -\frac{C_d}{C_{OX}} \quad (\text{Equation 2.9})$$

Where ΔV_{TH} stands for the extracted threshold voltage for the applied back-bias $\Delta V_{BB}=0V$ and 100mV. The depletion capacitance C_d is defined as:

$$C_d = \frac{\sqrt{q\epsilon_{Si}N_a\phi_f}}{4\phi_f} \quad (\text{Equation 2.10})$$

Where q is the elemental charge, ϵ_{Si} the permittivity of the silicon, ϕ_f the fermi potential and N_a the doping concentration. As a result, if C_{OX} is known, the channel doping concentration can be found and is shown on Fig. 2.14.

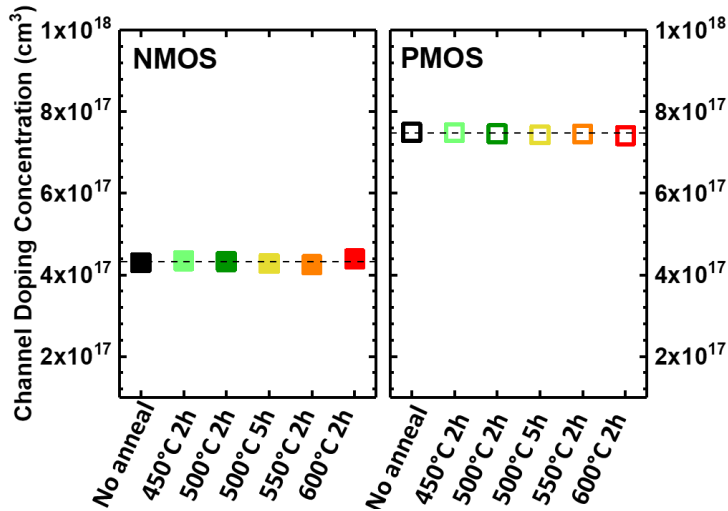


Figure 2.14 Channel doping concentration extracted with back-bias sensitivity method of NMOS and PMOS devices. The doping remains stable despite the thermal budget.

As expected, the channel doping for both N&PMOS devices is stable, meaning that *the additional thermal budget did not interfere in the channel doping integrity.*

Polysilicon gate doping integrity

As it was previously highlighted, while advanced digital devices integrate high-k (HfSiON~1nm), metal (TiN) and Polysilicon, the gate stack of the device under study is composed of only SiO₂/Polysilicon with a relaxed oxide thickness (13nm). In the absence of a metal, the device's work-function is then imposed by the phosphorus-doped or boron-doped polysilicon (for NMOS and PMOS, respectively). This is done for process simplicity, such that the polysilicon gate is doped and activated using the same ion implantation and annealing steps that form the S/D regions.

This may result in an insufficient doping at the poly-SiO₂ interface, with the dopants being predominantly present around the grain boundaries [30]. This weak activation of poly at the oxide interface may lead to polysilicon depletion, which lowers the gate capacitance and increases the effective electrical thickness of the gate insulator [31], thus weakening the action of the gate on the inversion channel. This carrier depletion increases the threshold voltage, and is specifically problematic when the depletion thickness becomes a significant fraction of the gate oxide thickness [32].

It must be said that when dealing with polysilicon gate, there is always a poly-depletion, even if small/negligible. That said, what is crucial to know is if the additional thermal budget can eventually increase this polysilicon depletion. As previously demonstrated, even though the dopant in the silicon channel is stable, the integrity of the dopants in the polysilicon gate is yet to be analyzed.

For this reason, the flatband voltage was extracted by $1/C^2$ method [29] from the C-V characteristics, and it is depicted for N&PMOS on Fig. 2.15. Indeed, the flat-band voltage remains stable up to 550°C 2h, whereas a shift is observed for the split 600°C 2h. The latter is reduced by 40mV for NMOS, and ~ 60 mV for PMOS devices. Knowing that the polysilicon is the one imposing the workfunction of the device, if there is any dopant deactivation in the interface of the polysilicon with the silicon dioxide, this would lead to an increase in the threshold voltage.

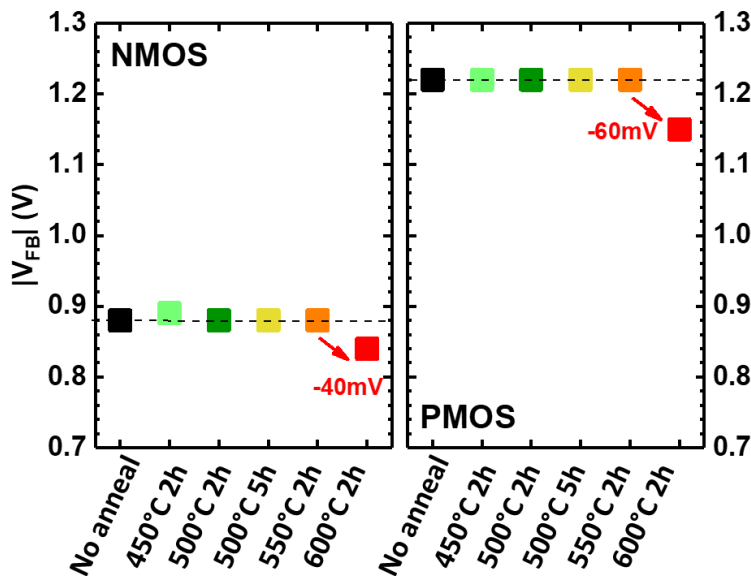


Figure 2.15 Flatband voltage extracted by $1/C^2$ method for NMOS and PMOS devices. The observed shift at 600°C 2h is well correlated with the threshold voltage instability previously highlighted.

Furthermore, thanks to C-V quantum mechanical model, one could estimate the poly doping concentration from the experimental C-V characteristics, as shown on Fig. 2.16. For the reference wafer, the poly doping concentration is estimated to 10^{20} at/cm³, while it is decreased to $5 \cdot 10^{19}$ at/cm³ for N&PMOS for the split at 600°C 2h. It must be noted that the model assumes a constant doping profile, whereas in reality, the doping profile in the polysilicon gate is not constant. This is the reason why the model does not perfectly fit the experimental data throughout the entire threshold region (roughly from 1 to $2 \mu\text{F}/\text{cm}^2$). Despite, it is evident that the doping is decreased, and can thus explain the threshold voltage shift.

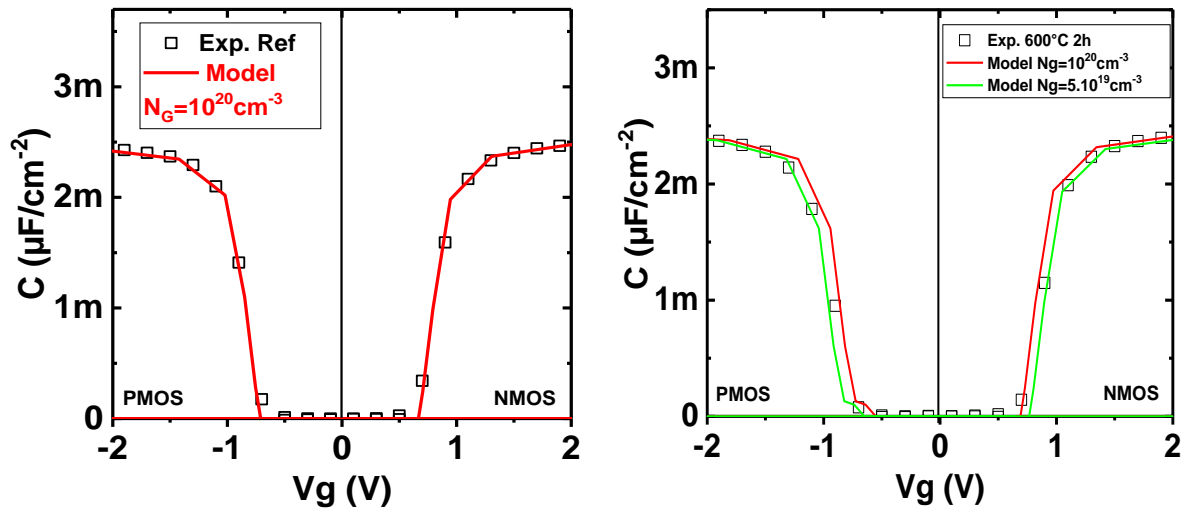


Figure 2.16 Gate doping concentration for the reference wafer and the split 600°C 2h estimated with QM model using a constant doping profile. The empty squares correspond to the experimental data for the reference wafer on the left, and for the split 600°C 2h on the right. The straight lines correspond to different values of polysilicon doping from the model used.

In addition, full sheet resistance measurements were performed on polycrystalline active areas that were implanted with its corresponding doping conditions. Fig. 2.17 shows the percentage of unsilicided resistance loss when compared to the no anneal wafer used as a reference, for the different thermal budget splits.

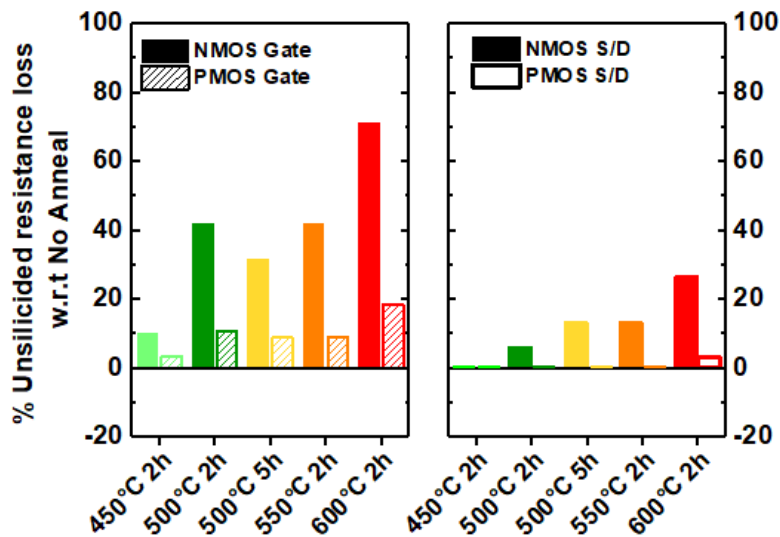


Figure 2.17 Unsilicided resistance loss in % with respect to reference for N&P type dopants for the polycrystalline (on the left) and crystalline regions (on the right). These results highlight that the dopant deactivation is not only dependent on the dopant itself, but also on the silicon template.

For the unsilicided structures, the sheet resistance is mostly increased for the split 600°C 2h and is also more attenuated on NMOS devices. Overall, the resistance increase is expected to be given by the polysilicon deactivation previously observed. In contrast, the

unsilicided resistance in the S/D region only shows a degradation of 30% at 600° 2h for the NMOS device, whereas is stable for PMOS despite the additional thermal budget. Even though the impact of this resistance increase on the S/D region shall be discussed in the following sections, it is important to highlight that the dopant deactivation is higher on polycrystalline than on crystalline regions.

As a conclusion, it has been shown that in the case of the studied high voltage device, the observed threshold voltage shift for the split 600°C 2h is well attributed to a dopant deactivation at the interface of the SiO₂ in the polysilicon gate. Despite, this issue can be solved by adjusting the channel doping concentration.

2.2.2 Silicide Stability

The silicide is a key step to form a low-resistivity electrical contact and is usually employed using a self-aligned silicide process. In this case, the term self-aligned refers to the process for which resist and lithography steps are not necessary [33], thus reducing the complexity and cost of forming the contact to the silicon areas.

The silicide formation occurs through a solid-state reaction between metal and silicon. The first step is to prepare the silicon surface, followed by the deposition of the chosen metal layer. After that, a first anneal is performed to form the silicide. After a selective wet etching to remove the metal that has not been reacted, a second anneal is carried out, this time to obtain a more suitable silicide phase, thus lowering the Si resistivity. The temperature of this anneal cannot be too high, to prevent Si diffusion. The goal of this last anneal is to achieve a complete and uniform phase transformation without agglomeration, nickel piping diffusion or subsequent phase change formation, which are a source of higher contact resistance and leakage issues.

Different materials can be used in the silicide process, as long as certain requirements are met, such as high melting point (which can be translated to generally a better thermal stability), self-alignment and etch selectivity [33]. Cobalt and Titanium silicides have initially been the material of choice. Titanium silicide presents low resistivity, relatively good thermal stability, and compatibility with silicon processes. Cobalt also presents a low-resistivity phase, but with a large silicon consumption, which limits its compatibility with ultra-shallow junctions. In addition, interface roughness caused by nucleation control growth from CoSi to CoSi₂ phase is another major issue for junction leakage.

Nickel-based silicides [34] have replaced CoSi as the contact material in advanced technologies beyond 65nm node due to its many advantages such as low temperature formation of the mono-silicide NiSi phase (350°C), low consumption of Si, compatibility with SiGe, among others. Despite its different phases from its complex phase diagram, only the NiSi phase is desired to improve the contact between the CMOS transistor and the first metallic level. This is because it is the phase with the lowest resistivity for low silicon consumption compared to previous silicides used in microelectronics.

The main drawback of NiSi silicide is its degradation at post thermal treatments, where a few degradation mechanisms shall be briefly distinguished. The first one is that the desired NiSi phase does not withstand high temperature budgets in subsequent process steps without high resistivity phase formation (NiSi_2) [35]. Another issue is that Ni can diffuse in an uncontrolled manner under the transistor gate, a phenomenon called encroachment or piping (Fig. 2.18 a) and can result in excessive leakage current. Lastly, the NiSi phase can also degrade at high temperatures thus leading to a resistance increase through a silicide agglomeration (Fig. 2.18 b).

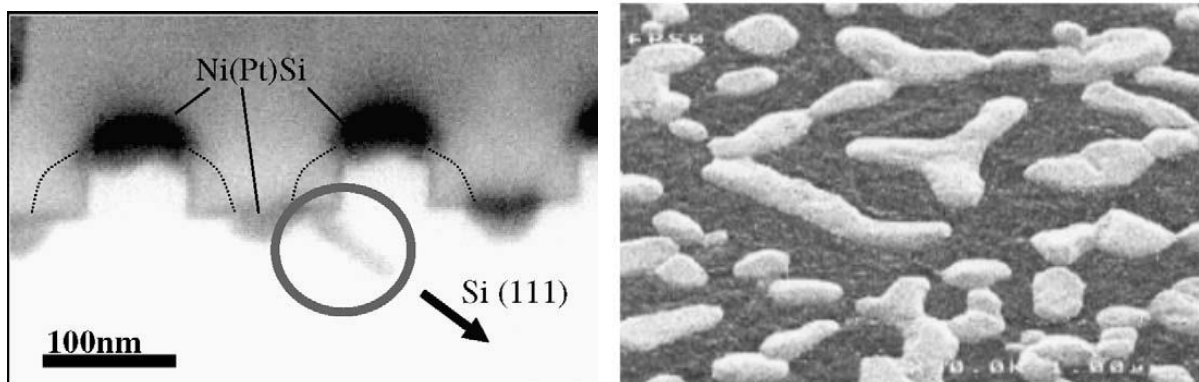


Figure 2.18 a) Scanning electron microscope (SEM) cross section realized after complete integrated circuit fabrication. The failed CMOS transistor present a nickel encroachment defect in the Si (1 1 1) substrate planes [36]. b) Agglomeration of NiSi at temperatures exceeding 550 °C [34].

It must be mentioned that even though this work is focusing in a 3DSI, even when dealing with planar devices, the aforementioned issues can also occur during subsequent interconnect manufacturing stages. As a result, the silicide stability can be a source of instabilities on all manufacturing devices, where increasing the thermal stability of NiSi has thus become an important research topic in the semiconductor industry [36].

One way to improve the stability of NiSi is by adding some alloy element to help stabilize the film, such as Platinum (Pt) [37]. On that note, the silicide that has been previously studied, as well as the one that shall be analyzed on this work is the $\text{Ni}_{0.9\%}\text{Pt}_{0.1\%}$. Despite the advantages brought by adding Pt, the thermal budget involved for the elaboration of the top-level can still lead to a silicide morphological degradation, thus deteriorating contact properties (i.e. increase of leakage or even shorts in the junctions).

Previous studies have emphasized the silicide as a main source of thermal stability upon an additional thermal budget, degraded beyond 500°C . This section shall thus investigate the silicide stability on the gate and on the S/D region. Full sheet resistance measurements were performed on polycrystalline and crystalline active areas, thus corresponding to the gate and S/D regions, respectively. Therefore, Fig. 2.19 depicts the percentage of silicided resistance loss when compared to the no anneal case used as a reference, for the different thermal budget splits.

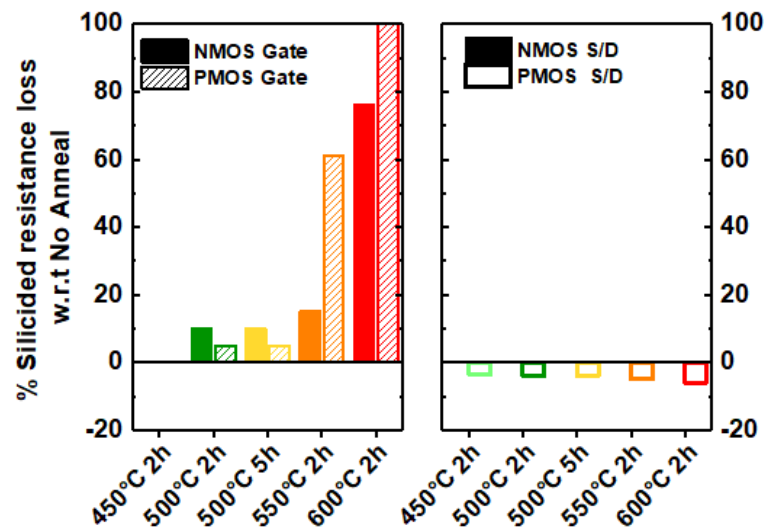


Figure 2.19 Silicided resistance loss in % with respect to reference for N&P type dopants for the polycrystalline (on the left) and crystalline regions (on the right).

Concerning the silicide in the gate region, even though small (up to 10% for NMOS and 5% for PMOS), a resistance degradation is observed starting at 500°C 2h. As the thermal budget is further increased, the silicide resistance is greatly degraded, especially for PMOS devices, that shows 60% degradation starting at 550°C 2h. This rise in sheet resistance could be an indication of the silicide morphology modification on the polysilicon regions [36]. Contrarily, the silicide resistance on the S/D region is stable for NMOS and even improved for PMOS devices. These results thus highlight that the silicide stability is not

the same on poly and crystalline regions. The resistance increase observed on the polysilicon gate, could be attributed to the silicide changing phase, which appears to be strongly dependent on the crystallinity of the substrate. All things considered, even though the silicide seems stable in the S/D area, the TB must be kept at 500°C 5h to preserve the integrity of the polysilicon gate.

Despite the fact that silicide seems stable in the S/D region, its impact on junction leakage shall also be investigated. This is because of the possible anomalous gate-edge leakage current, caused by the encroached growth of nickel silicide across the p-n junction, caused by the isotropic and rapid diffusion of Ni atoms during the silicidation annealing [38]. Although this encroached growth can be observed by cross-sectional secondary electron microscopy (X-SEM) and transmission electron microscopy (X-TEM), it is difficult to analyze it quantitatively because a large number of images are necessary for this analysis. A proposed method to evaluate this stochastic encroached growth is a simpler way is to quantitatively evaluate the leakage current on a large number of devices [39]. That said, the leakage current for N&PMOS devices was measured on 130 devices is shown on Fig. 2.20, where no degradation is observed even for the split 600°C 2h, thus further highlighting a good silicide stability in the access.

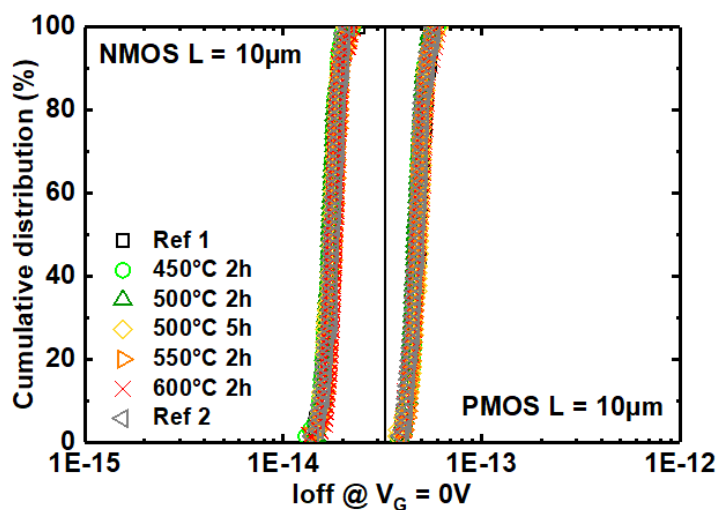


Figure 2.20 Cumulative probability of I_{off} of NMOS and PMOS for 130 devices.

At first sight, one may think these are surprising results, given the previous emphasis on the silicide degradation. However, other than the silicide itself ($\text{Ni}_{0.9\%}\text{Pt}_{0.1\%}$), there are other factors that have not been mentioned so far, and that also impacts the silicide thermal stability, such as the device dimensions. This is because with device scaling, the

silicide has also shrunk, thus becoming thinner in order to reduce the parasitic resistance and to provide drain current gain.

However, the main drawback of scaling the silicide is that its thermal stability is also reduced. This can clearly be seen on (Fig. 2.21-a), depicting the sheet resistance (R_s) for various NiPt thicknesses (T_{NiPtSi}) as a function of additional annealing temperature. An increase in R_s is observed as the annealing temperature is increased (due to the agglomeration of NiPt silicide). Additionally, the agglomeration starts to be observed sooner (for lower temperatures) for thinner silicides.

Likewise, another parameter that impacts the junction leakage is the spacing between gates, as highlighted on Fig. 2.21-b. For the case of a gate spacing of 45 and 75nm, the junction leakage current increases with thick T_{NiPtSi} due to the approach of NiPt silicide to the metallurgical junction. Furthermore, at narrow gate spacing of 45nm, the junction leakage current also increases in thin film regions.

It is then concluded that the degradation of junction leakage properties related with NiPt silicide strongly depends on T_{NiPtSi} , the gate pitch, and the size of diffusions, and that thin NiPt-silicide formation in scaled CMOS is a serious issue.

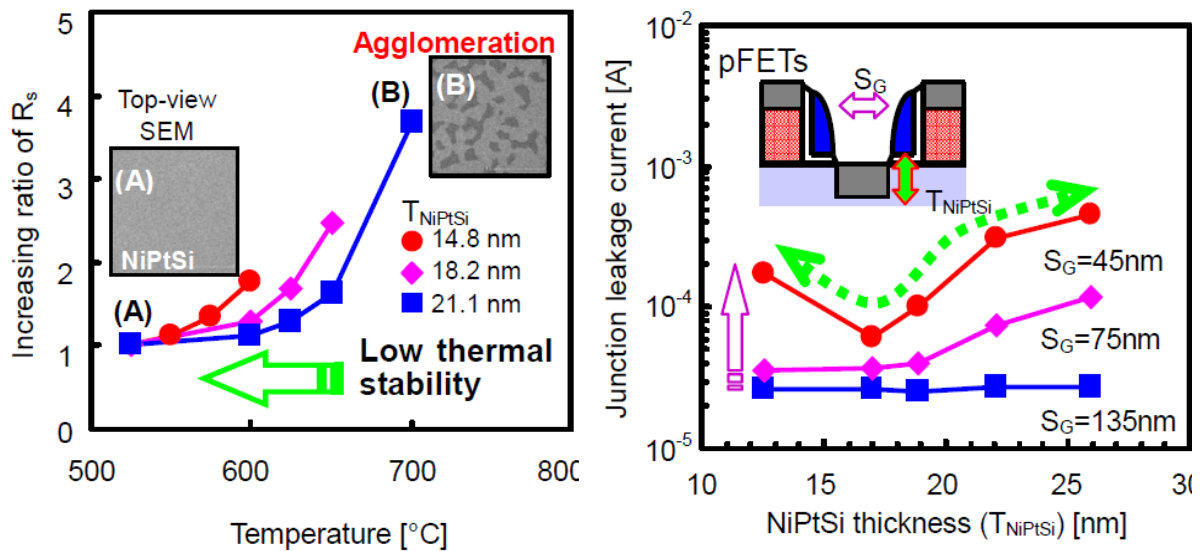


Figure 2.21 a) Increasing ratio of R_s with various T_{NiPtSi} as a function of additional annealing temperature and b) junction leakage current in pFETs with various spaces between gates (S_G) as a function of T_{NiPtSi} [39].

From these results, it is possible to try to correlate these ideas to the device studied in this work. For example, the device under analysis presents relaxed dimensions, meaning that the silicide has not been scaled as it is done for high performance devices. In addition, the

gate pitch being greater than 135nm as in the example above, makes that the silicide is far away from the metallurgical junction. Lastly, the silicide used on this studied has been optimized to avoid the encroachment phenomena, by adjusting the thermal steps during the silicide formation. All these geometries are indeed beneficial for the silicide thermal stability of the S/D regions, thus explaining the good silicide stability observed.

2.2.2.1 Suggested improvements

So far, the silicide has shown good stability in the S/D regions and appears to be more critical in the polycrystalline silicon region. At this stage, the repercussions of this silicide instability from thermal budgets above 500°C 5h are not known. Amongst these possible repercussions, it could be mentioned a potential increase in gate resistance, thus affecting the AC performance of the bottom device. For this reason, this section aims at exploring some ways of improving the silicide instability in the gate region, once more, if the latter becomes necessary.

In view of a 3DSI, some studies have already highlighted the interest of NiCo_{10%} silicide (NiCo_{10%}Si) in order to extend its thermal stability, given that the (NiCo)Si phase appears to be stable until temperatures such as 800°C [40]. Therefore, it could be interesting to verify its validity/thermal stability when used on polycrystalline silicon regions as well. A second possible optimization to improve the silicide's thermal instability would be to use a pre-amorphization implantation (PAI) with Germanium, this time in the gate region. This approach modifies the silicide roughness, thus limiting the agglomeration phenomena [5], [41].

2.2.3 Global performance stability

The overall performance of the device is lastly studied with some of the most important DC FOM. The I_{on}/I_{off} of both N&PMOS is shown on Fig. 2.22 for the different thermal budget splits and compared to the no anneal one used as a reference. In order to avoid any shift due to the threshold shift previously observed, the latter is given on $V_{GT}=3.5V$. It can be seen that no performance degradation nor additional junction leakage are observed.

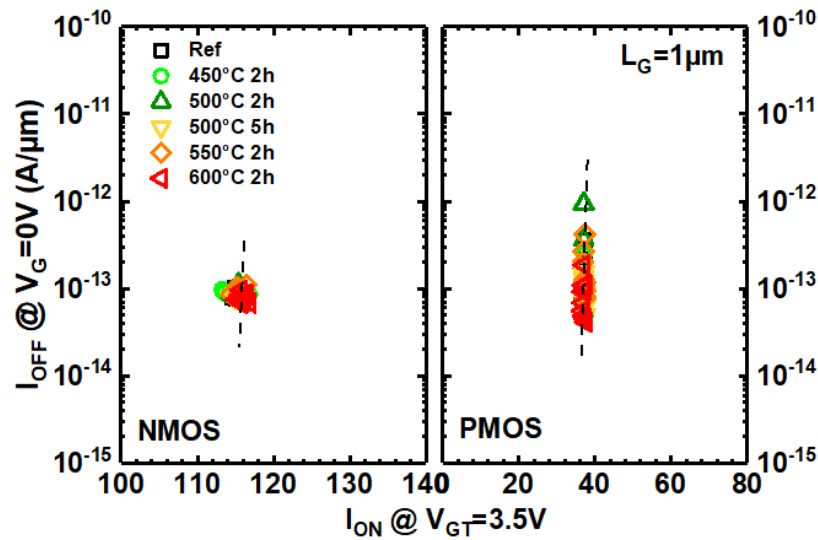


Figure 2.22 I_{ON} - I_{OFF} trade-off versus thermal anneals for NMOS and PMOS, respectively. $L_G=1\mu\text{m}$. $V_{DD}=5\text{V}$.

Similarly, the effective mobility (Fig. 2.23) measured on long channel devices is also unaffected despite the thermal budget split.

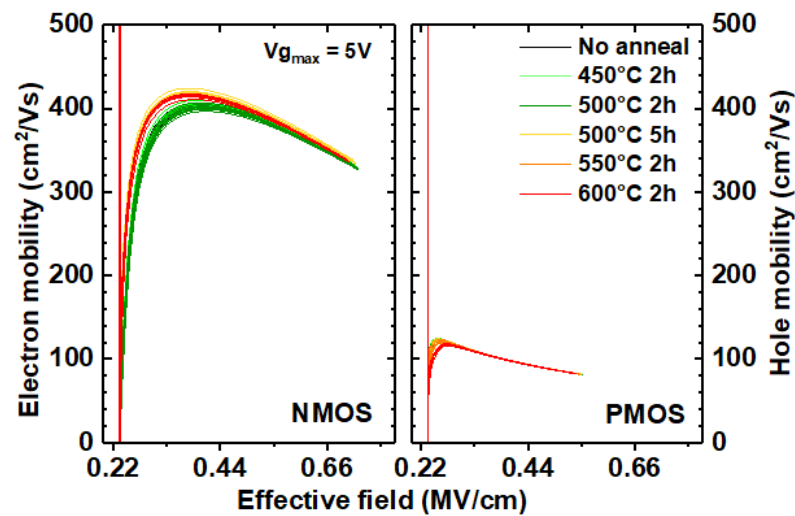


Figure 2.23 Effective mobility of NMOS and PMOS versus anneals. $L_G=10\mu\text{m}$.

Additionally, the access resistance (R_{SD}) was also extracted using the R_{TOT} method [42] and is compared with the total resistance (R_{TOT}) on Fig. 2.24. It can be seen the R_{SD} is also stable, as expected since there is no performance degradation. This can be explained by the fact that in a long channel device as the one studied here, the R_{SD} becomes negligible in comparison to the channel resistance, so even if there is a small silicide resistance increase, the latter cannot be observed on these large structures.

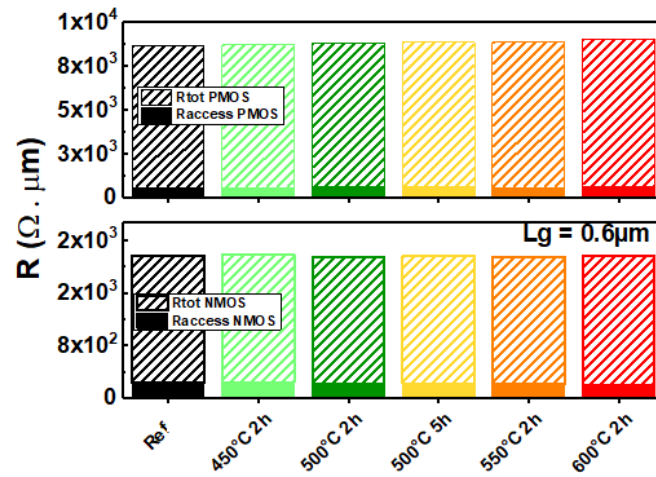


Figure 2.24 Total and access resistance of NMOS and PMOS extracted for $L_{gmin} = 0.6\mu\text{m}$.

Lastly, the voltage gain g_m/I_D is also shown (Fig 2.25), and it remains stable for all regions of operation, which is in agreement with the results shown so far.

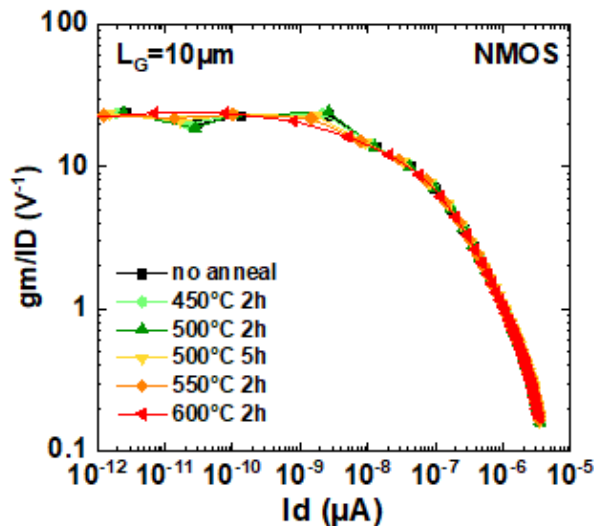


Figure 2.25 NMOS voltage gain. $L = 10\mu\text{m}$.

To conclude, the device global performance is ensured even for the higher thermal budget of 600°C 2h as seen on the DC figures of merit. However, as to maintain the gate stack integrity, the thermal budget must be limited at 500°C 5h. These results implies that BULK high voltage devices ($V_{DD}=5\text{V}$) can already be safely used as the bottom device in view of a 3D Sequential integration.

2.3 FEOL STABILITY UNDER STANDARD ANNEAL FOR MORE MOORE APPLICATIONS

The aim of this section is to now further the studies concerning the thermal stability of advanced digital devices used for More Moore applications. The methodology used here

is the same as the one mentioned in the previous section, where an additional thermal budget is applied and its impact on device performance analyzed through in-depth characterization. As such, anneals were performed after the CMP PMD at several thermal budgets (ranging from 450°C-600°C 2h), this time on standard SOTA 28nm FDSOI (Fig. 2.26). Once more, seeing the moment on which the anneal is applied, only the device is concerned.

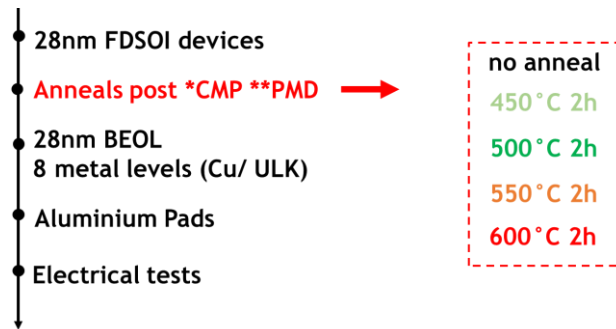


Figure 2.26 Split table of the TB applied to state-of-the-art 28nm FDSOI devices.

2.3.1 Static FOM

As it was previously mentioned, in order to have further understanding on the matter, the tests shall be performed also under AC operation. However, standard tests under DC operation must be done first, as to ease the understanding later on.

The I_{ON} - I_{OFF} trade-off is then shown for N&PMOS for the different thermal budget splits (Fig. 2.27) for two gate lengths, including the nominal one of $L_G=30\text{nm}$ and $L_G=34\text{nm}$. Concerning the O_{FF} regime, no additional leakage is observed. In the case of the PMOS, it remains stable despite the thermal budget. On the other hand, the NMOS remains stable up to 500°C 2h, whereas an important degradation of -8.5% of I_{ON} is seen for the split 600°C 2h.

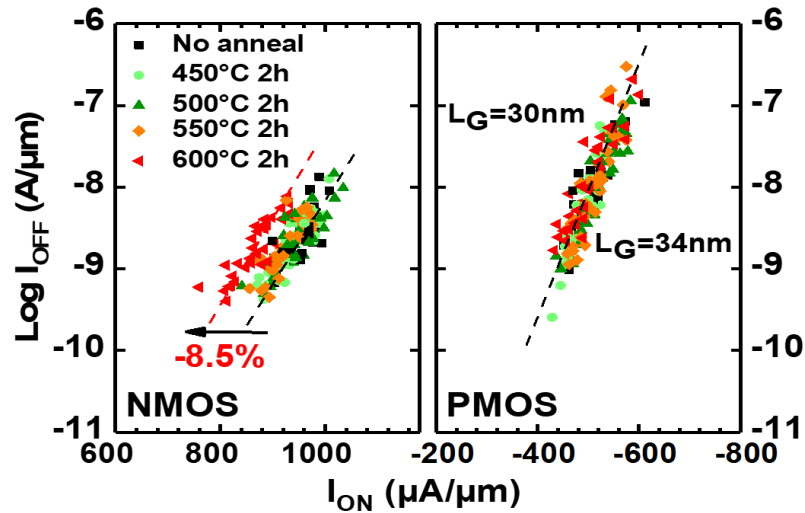


Figure 2.27 I_{ON} - I_{OFF} trade-off versus thermal anneals for N&PMOS, respectively. $L_G=30$ and 34nm and $V_{DD}=1V$.

The threshold voltage as depicted on Fig. 2.28 also seems stable, not being able to explain the I_{ON} shift observed on NMOS starting from 550°C 2h.

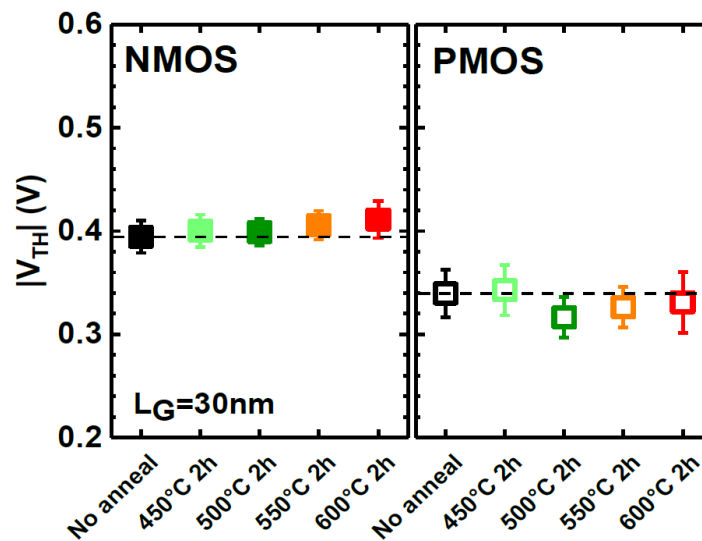


Figure 2.28 Threshold voltage versus thermal anneals for N&PMOS, respectively.

At this point, it is already understood that an additional thermal budget may lead to dopant deactivation, as mentioned on a previous section. For this reason, Fig. 2.29 depicts the I_{ON} current as a function of the unsilicided resistance. As before, this resistance value has been measured on active zones, and can be used as an indication to know what is happening in the device. It can be seen that whilst there is no trend for PMOS devices, the performance loss is well correlated to an increase in the unsilicided resistance in the case of NMOS.

As a result, the major contribution to the I_{ON} loss of the NMOS device is well correlated with R_{SD} degradation due to dopant deactivation. It must also be said that the access of these devices is composed of an implant cocktail, in order to yield the best performances. This means that even though a dopant deactivation is shown, further studies are required in order to distinguish between Arsenic or Phosphorus deactivation, which are the dopants implanted on NMOS devices.

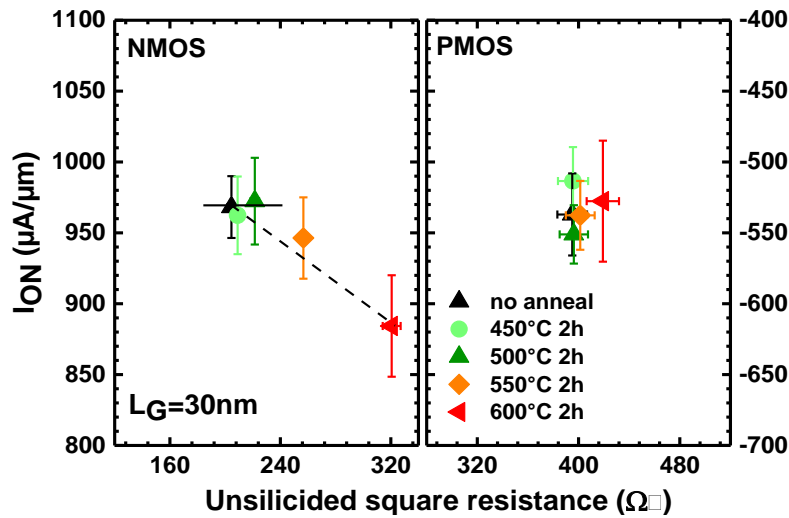


Figure 2.29 I_{ON} vs. unsilicided square resistance. I_{ON} degradation above 500°C 2h is well correlated to dopant deactivation, shown as an unsilicided resistance increase.

2.3.2 Dynamic FOM

Now that the DC performance has been evaluated, one may proceed with the AC analysis. Whilst the previous section focused on evaluating the thermal stability of single transistors, the AC analysis allows one to address the impact of an additional thermal budget in a manner closer to its real operation.

Before this analysis, a short review of the ring oscillator metrics shall be presented. Ring oscillators consists of a circular chain of odd number inverters, where the circuit speed is directly linked to the ability of the logic gates to perform their operation [43]. In a ring-oscillator, the output of the last inverter is connected to the input of the first inverter, making it oscillate between the "1" and "0" states (that is, V_{DD} and ground). The oscillation frequency is given as:

$$f = \frac{1}{2N\tau_p} \quad (\text{Equation 2.11})$$

And it depends on the number of inverters (N) and the propagation delay of each inverter (τ_p). When the inverter is not switching, the current flowing through the inverter is the leakage current I_{STAT} . This current is also referred as the stand-by or static current, which is how it shall be referred to in this work. It is measured through the supply voltage in the quiescent state. This leakage current is linked to the NMOS and PMOS leakage currents I_{OFF} .

When a stage of inverter is switching because of a change of input, a dynamic current loads the output capacitance. In a ring-oscillator, this dynamic current I_{DYN} is successively generated by each inverter stage. It flows alternatively through the NMOS and the PMOS according to the switching operation ("0" to "1" or "1" to "0").

Concerning the delay, it can be expressed as a function of the dynamic current:

$$\tau_p = \frac{V_{DD} C_{EFF}}{2 I_{DYN}} \quad (\text{Equation 2.12})$$

where C_{EFF} is the effective capacitance to be loaded. This capacitance includes the gate capacitance of the next level inverter and all parasitic capacitances. In terms of power, the dynamic and static powers are deduced from the currents. The dynamic power is given by:

$$P_{DYN} = I_{DYN} V_{DD} = \frac{V_{DD}^2 C_{EFF}}{2 \tau_p} \quad (\text{Equation 2.13})$$

and the static power by:

$$P_{STAT} = I_{STAT} V_{DD} \quad (\text{Equation 2.14})$$

The device performance under dynamic operation was evaluated through the characterization of ring oscillators with a fan-out 3 (FO3) and 28nm nominal dimensions. The static leakage is depicted on Fig. 2.30 as a function of frequency, for the different thermal budget splits. As it was observed in the leakage current of N&PMOS, it is highlighted that the same static leakage is also unchanged. Concerning the frequency, the latter is stable up to 500°C 2h, whereas an important degradation of -9% is seen for the split 600°C 2h.

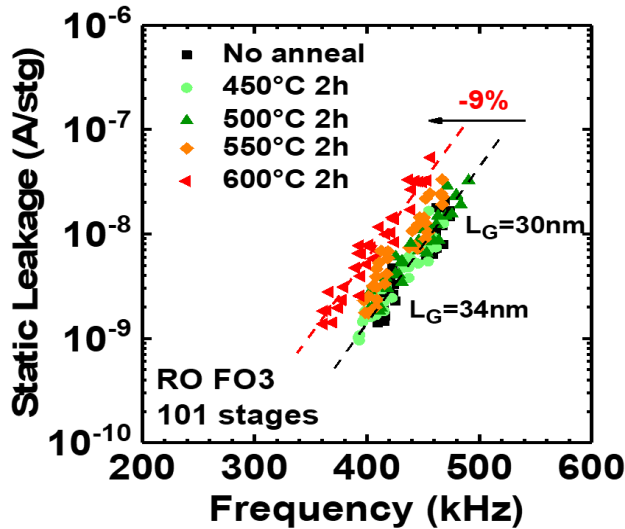


Figure 2.30 Static leakage vs. frequency for $L_G=30$ and 34nm measured on Ring Oscillator with 101 stages. $V_{DD}=1V$.

Fig. 2.31 depicts the dynamic power as a function of frequency. Despite the degradation already mentioned, the most interesting feature of this plot is that it shows a linear behavior, which is an indication that the front-end capacitances are stable. Furthermore, given that $P_{DYN}(Freq)$ is unchanged with the TB, this can already be an indication that the BEOL capacitances are also unchanged.

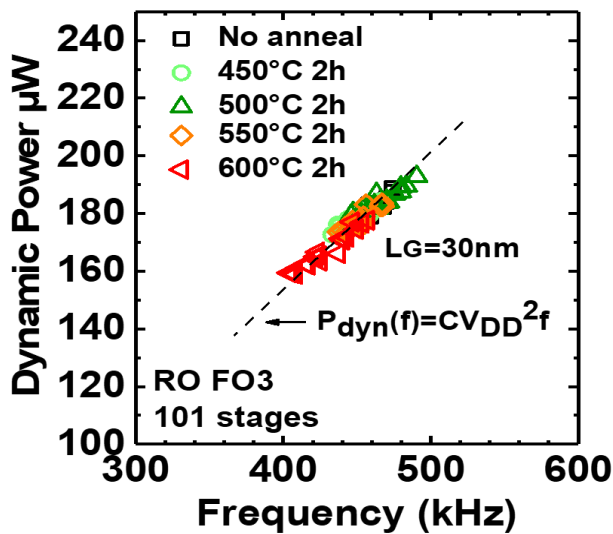


Figure 2.31 Dynamic power vs. frequency for $L_G=30\text{nm}$. Linearity shows no change in the front-end capacitances.

Lastly, it is necessary to understand the frequency degradation above 500°C 2h. One may be reminded that through the static analysis, the NMOS devices had already suffered a degradation due to dopant deactivation for the higher thermal budget splits. By plotting the frequency as a function of the I_{ON} current, it can be seen that the -9% frequency degradation for the split 600°C 2h is well correlated with the I_{ON} decrease (Fig. 2.32).

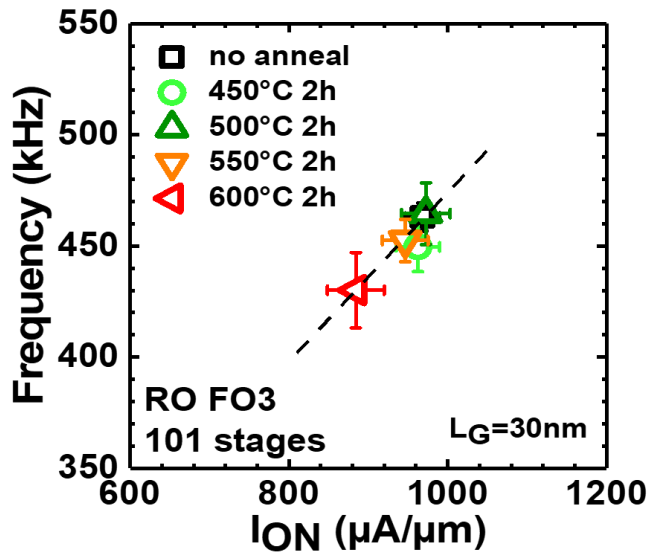


Figure 2.32 Frequency (101 stages) vs. NMOS drive current I_{ON} .

To conclude, the dynamic performance of 28nm FDSOI devices is ensured up to 500°C 2h, and the observed degradation at higher thermal budget is well correlated with the DC characteristics, such as the I_{ON} reduction due to NMOS dopant deactivation.

2.3.3 Yield

Previous studies highlighted the silicide ($\text{NiPt}_{10\%}\text{Si}$) as one of the most critical parameter upon an additional thermal budget, especially on advanced digital devices due to its scaled dimensions [1], [22]. Similarly to what has been shown on an earlier section, the silicide has been characterized on different active zones, and these results are shown on Fig. 2.33. It can be seen that on the large silicide resistance structures ($W=2\mu\text{m}$), a small resistance increase starts to be observed at 600°C. The latter could be ascribed for example to silicide agglomeration. It is interesting to notice that as the width of the active area is further decreased ($W=0.1\mu\text{m}$), this phenomenon is exacerbated. This is because silicide agglomeration is a random effect, and when it happens exactly where the contact is located it leads to an enormous resistance increase. Hence, the silicide appears to be more critical for smaller dimensions, that is why dense circuits with minimum active zone areas are necessary to properly address the impact of the thermal budget on the silicide. The same reasoning can also be applied to detect Ni encroachment. This abnormal penetration of nickel towards critical areas of the MOS transistors, is a marginal phenomenon also more easily detected on dense structures with small features, like SRAMs. Ultimately, both these phenomena lead to poor device yield.

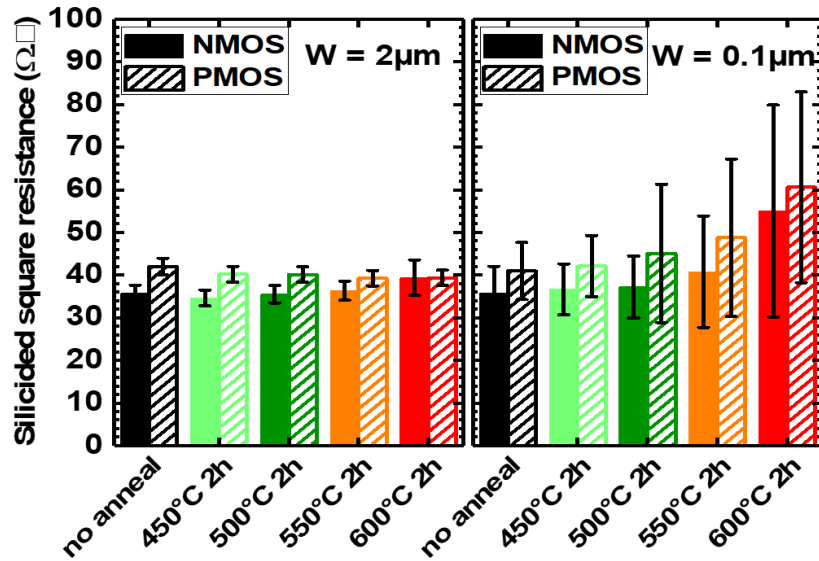


Figure 2.33 Silicided square resistance for $W=2\mu\text{m}$ and $W=0.1\mu\text{m}$. $\text{NiPt}_{10\%}$ agglomeration exacerbated as the width is decreased.

The yield of dense SRAM cells (5Mbit dense, $\text{area}=0.120\mu\text{m}^2$) and logic gate array (1 Million Flip-Flops) was evaluated. The yield stability is unaffected up to 500°C 2h (Fig. 2.34), whereas a yield of 0% is highlighted for superior thermal budgets. Similarly, the yield of the logic structures is retained up to 500°C 2h . In both cases, the yield is also ensured for a wide supply voltage range (at least $V_{\text{DD}}\pm 30\%$).

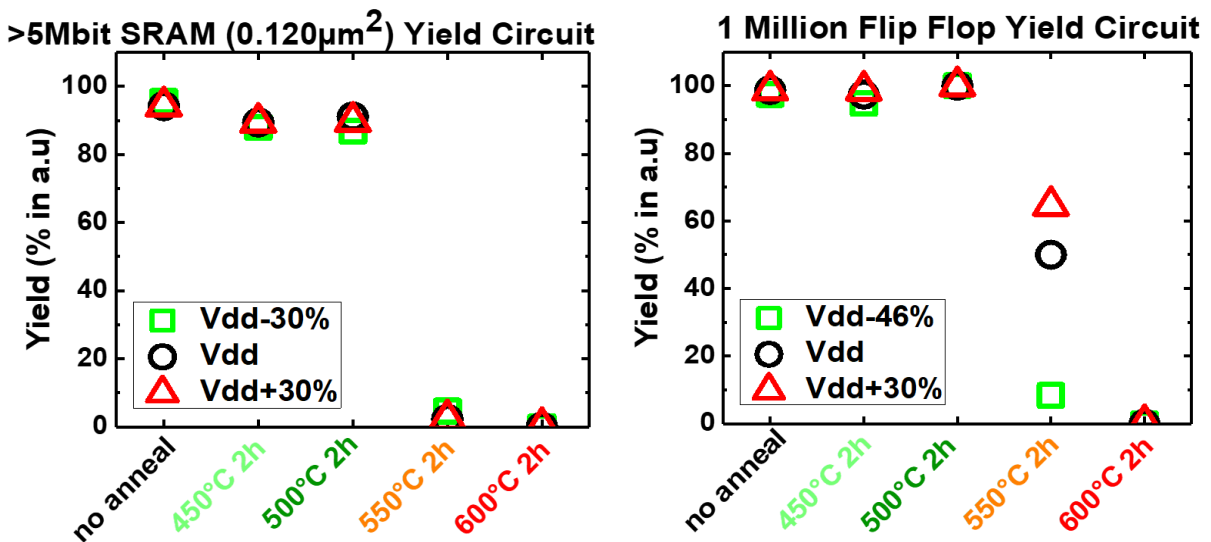


Figure 2.34 Yield evaluation with TB applied after PMD CMP on SRAM cuts (SRAM $\text{area}=0.120\mu\text{m}^2$ and 5Mb) and logic structures composed of 1M Flip Flops.

These results allow one to strongly conclude on device's thermal stability up to 500°C 2h through interesting industrial FOM.

2.4 IBEOL STABILITY UNDER STANDARD ANNEAL FOR MORE MOORE APPLICATIONS

This section now investigates the thermal stability of both BEOL Cu/ULK interconnections and FEOL devices upon an additional thermal budget split. This time, the latter is performed after M8 level (Fig. 2.35), thus affecting both devices and the interconnects. The thermal budgets splits were chosen as ranging from 450°C-525°C 2h. This range was chosen based on the previous results, and already excludes the stronger thermal budgets due to the device degradation observed. In addition, short anneal splits were also tested, in order to assess if higher TB such as 600°C could be afforded for a relatively short duration (2 min). Afterwards, the flow continued without any modification, and the characterization was performed.

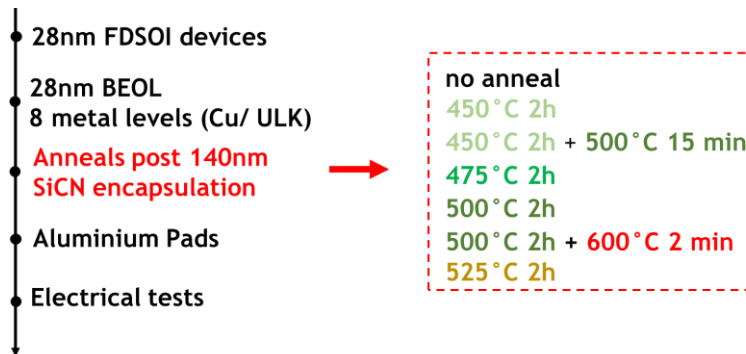


Figure 2.35 Process flow of the thermal budget splits applied to standard 28nm FDSOI devices after M8.

2.4.1 Robustness Characterization

In this work, the BEOL with 28nm design rules are used, and the intermetal dielectric (IMD) is a porous SiOCH with a permittivity of 2.7. The test structures used for reliability tests are schematized on Fig. 2.36, design at M1 and M2 levels.

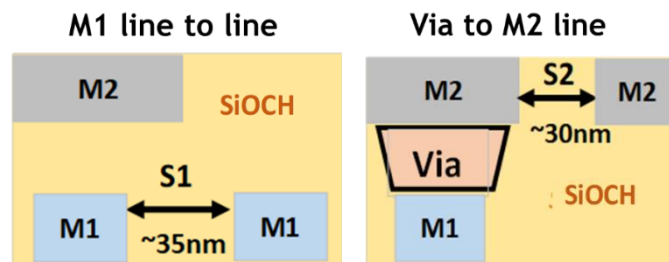


Figure 2.36 Schematic representation of reliability test structures. S =metal lines spacing, with $S2$ being more critical.

From these schematics, it is important to highlight the different metal lines spacing (here denoted S1 and S2), where the Via to M2 line test structure presents a shorter critical spacing.

The Time-Dependent Dielectric Breakdown (TDDB) of BEOL dielectrics was characterized via CVS (Constant Voltage Stress) tests [16], where a stress voltage is applied to the dielectric in order to measure the time that the breakdown of the structure is reached. From that, the Weibull distribution is plotted.

Fig. 2.37 denotes the Line to Line tests were performed at 25°C and 125°C for the no anneal and 500°C 2h split, subjected to a stress voltage of 20V. No degradation is evidenced after 500°C anneal. Additionally, same Weibull slopes and activation energy ($E_a=0.14\text{eV}$) are found in both cases, confirming that breakdown failure mechanism is unchanged.

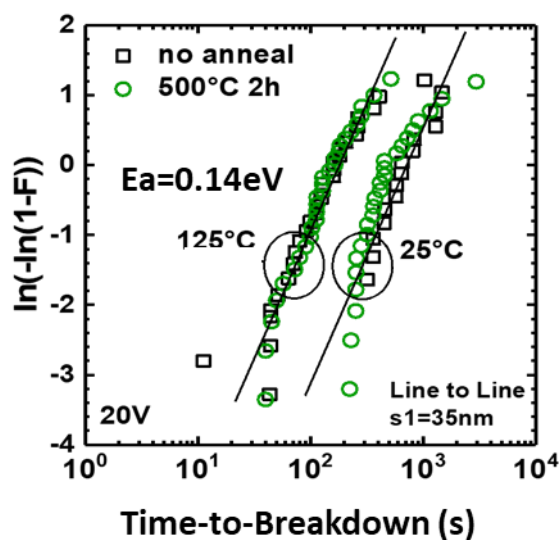


Figure 2.37 Time to breakdown in Weibull scale for Line to Line ($s1=35\text{nm}$) case with and without an anneal split (500°C 2h) after M8.

These results are further confirmed in the Via to Line configuration (Fig. 2.38), meaning that even for a reduced critical spacing, the ULK integrity is retained up to 525°C 2h.

The Cu/ULK thermal stability has never been demonstrated up to this thermal budget. Additionally, these results are shown after M8, whereas the previous studies [16] have been performed after M4. This can have an impact on the reliability, because the system may not yet be stabilized at the lower metallic lines. Still concerning the previous results, the furnace used to apply the thermal budget must also be considered. In this work, a cleaner furnace (with less chance of contamination) has been used. Also, the reliability of the interconnections is very process dependent, so the improvement shown can be attributed to a more robust and mature BEOL process.

Coming back to the Fig. 2.37, it can also be seen that as +600°C 2min thermal budget is added, TBD is already reduced.

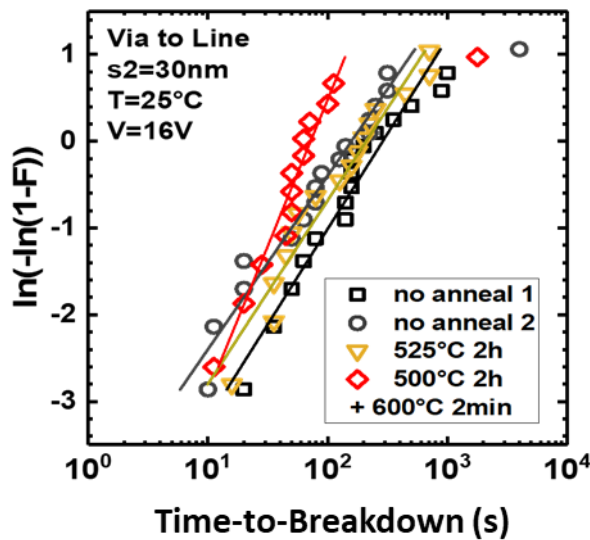


Figure 2.38 Time to breakdown in Weibull scale for Via to Line ($s1=30\text{nm}$) case for anneal splits after M8. The results are now compared with two reference wafers, named no anneal 1 and no anneal 2.

To conclude, a first demonstration of the thermal stability of BEOL up to 525°C-2h through the TDDB analysis. Additionally, this work is also in agreement with the stability demonstrated at 500°C 2h. One of the main implications is that these results evidence that top FET at 500°C 2h can be processed above BEOL with 28nm design rule without requiring the integration of alternative materials such as Co or W.

2.4.2 Yield

The yield of dense SRAM cells (5Mbit dense, area=0.120 μm^2) and logic gate array (1 million flip-flops) was once more evaluated and are depicted on Fig. 2.39. No degradation on the yield of the SRAM has been observed even at 525°C 2h. In contrast, the yield of the logic structures is stable up to 500°C 2h. It is observed that an additional 600°C 2min already leads to -20% yield reduction, hence not tolerated.

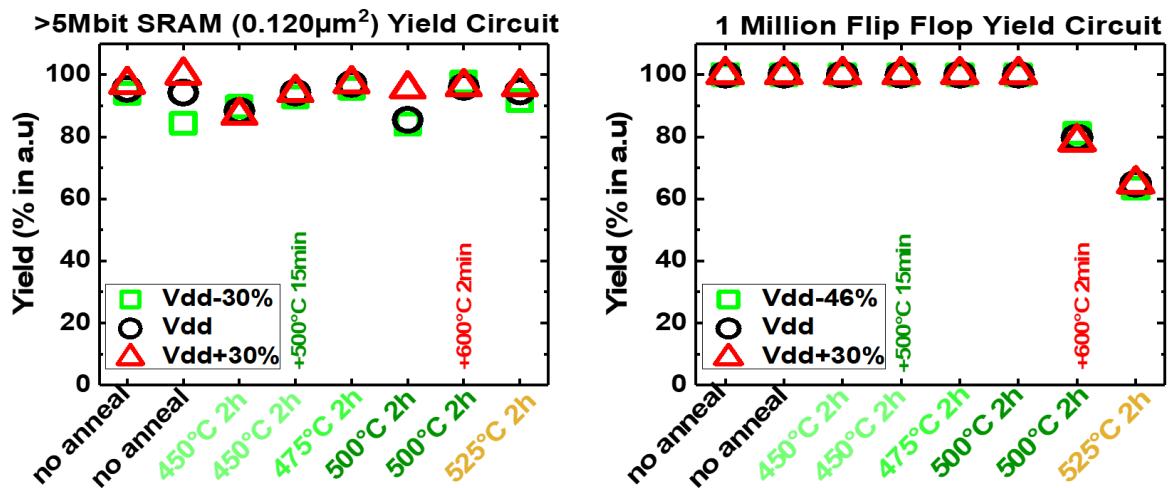


Figure 2.39 Yield on SRAM (SRAM area=0.120µm² and 5Mb) on the left and on logic structures composed of 1M Flip-Flops on the right. The thermal annealing splits were applied after M8.

Overall, these results are somehow in agreement with the reliability ones previously shown. It remains to be understood why the logic structures seems to be more affected by the additional thermal budget than the SRAM, where further studies need to be performed. However, it must be said that in a full process, all the steps can impact the final yield. Any thermal treatment or etch step performed on the BEOL in order to rend it more robust needs to be considered.

2.5 FEOL AND BEOL STABILITY UNDER ULTRA-SHORT ANNEALS FOR MORE MOORE APPLICATIONS

In this section, the thermal stability upon UV-NLA is evaluated on 28nm FDSOI technology with Cu/ULK M1 level, meaning that both device and interconnections are considered this time. In this work, the UV-NLA has a UV radiation (308nm), meaning that the absorption length is comparable to the typical top Si film thickness. In addition, its short pulse (<200ns) allows for a limited heat diffusion.

The test structure and the corresponding process flow are given on Fig. 2.40. The bottom device corresponds to a SOTA 28nm FDSOI. Then, a 20nm thick SiCN layer is deposited, followed by a 120 nm PECVD SiO₂ film used as ILD. To simulate the upper device level, a thin amorphous layer is deposited at low temperature, where two different thicknesses shall be tested ($T_{Si,top}=7nm$ and $T_{Si,top}=20nm$). Single pulse (160 ns) UV-NLA is then

performed with energy densities up to $0.5\text{J}/\text{cm}^2$. Lastly, electrical tests have been carried out on serpentine structures to detect any damage on the M1 Cu/ULK level.

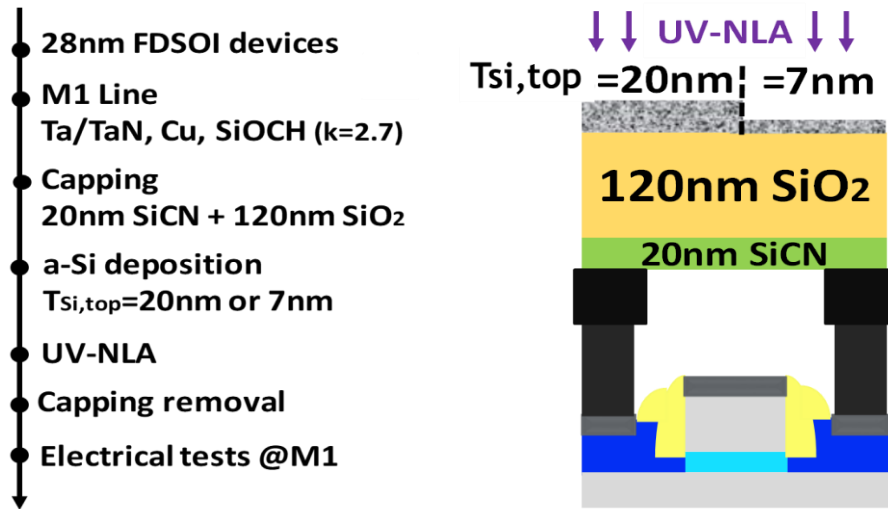


Figure 2.40 Schematic representation of the structure used for the ns laser experiment. a-Si ($T_{\text{Si,top}}=7\text{nm}$ and 20nm) simulates top device.

Cu line resistivity (Fig. 2.41) remains unchanged upon laser annealing even beyond the melt threshold of the a-Si ($\sim 1150^\circ\text{C}$), regardless $T_{\text{Si,top}}$ thickness.

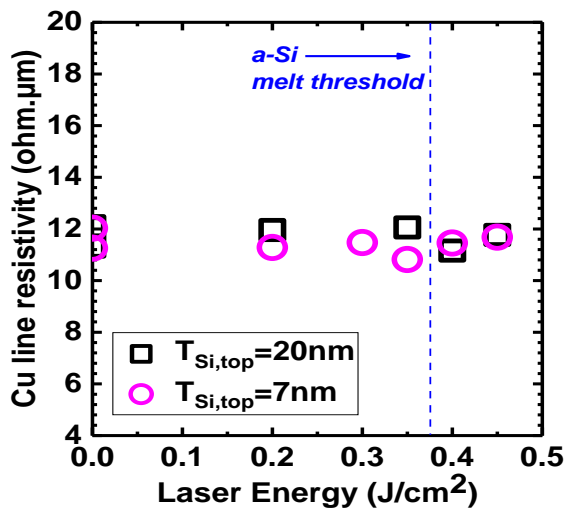


Figure 2.41 Cu line resistivity evolution as a function of the laser energy density after ns laser UV.

2D numerical simulations were performed using LIAB software [44] in order to determine the temperature endured by the underlying tiers. With this sort of simulation, it is possible to have the local thermal budget seen by different parts of the structure. The Cu T° remains below 500°C for $T_{\text{Si,top}}=20\text{nm}$ (Fig. 2.42), whereas it reaches 600°C but only for 100ns when $T_{\text{Si,top}}=7\text{nm}$. This demonstrates how the UV-NLA in-depth selectivity can be improved by an appropriate choice material thickness and process conditions.

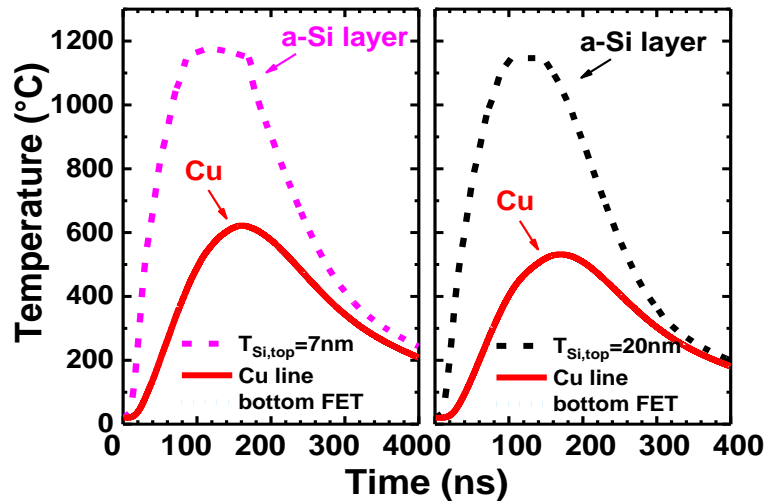


Figure 2.42 Temperature vs. time upon ns laser anneal for the two simulated structures for top a-Si layer (dashed lines), copper line (straight lines) and bottom FET (dotted lines).

Lastly, unchanged I_{ON} - I_{OFF} (Fig. 2.43) trade-off demonstrates the bottom device immunity upon UV-NLA of the top layer. Despite the thin top a-Si absorbing layer and the thin thermal barrier between the top a-Si and the embedded devices, both NMOS and PMOS transistors performance has shown no significant degradation upon UV-NLA treatment, even for conditions leading to the melt of the top a-Si layer. NiPt10% silicides, supposed to be the most temperature sensitive regions have not been affected by the laser annealing process.

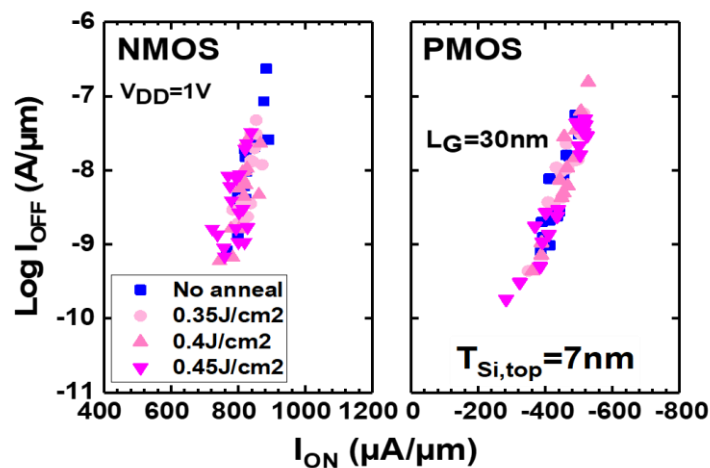


Figure 2.43 I_{ON} - I_{OFF} trade-off after UV-NLA for the thinner $T_{Si,top}=7nm$. No degradation on device performance is observed.

Therefore, laser energy conditions including melt can be successfully reached without affecting State-Of-The-Art (SOTA) Cu lines for 28nm FDSOI devices.

2.6 CONCLUSION

The thermal stability of high voltage ($V_{DD}=2.5-5V$) BULK devices is quantified. It has been shown that static figure of merit is unchanged up to 600°C 2h. A slight polysilicon deactivation is observed for the highest thermal budget (600°C 2h), affecting only the gate work-function (no additional poly depletion). However, this effect can be compensated by adjusting the channel dopant concentration. The silicide stability on the source and drain is shown to be stable, the weak point relying only in the silicide instability of the polysilicon gate starting from 550°C post anneals. Therefore, the thermal budget must be kept below 550°C 2h in order to ensure the gate resistance integrity. These results shows that high voltage SiO₂/Poly devices can be integrated on the bottom level of 3DSI with a comfortable thermal budget (at least 500°C 5h) to process the top device.

Concerning the additional studies targeting More Moore applications, for the first time, the thermal stability of a 28nm FDSOI CMOS technology is evaluated with yield measurements (5Mbit dense SRAM and 1 Million Flip-flops) and AC operation. Regarding the device stability, both AC and DC operation are stable up to 500°C 2h. For higher thermal budgets, a frequency degradation has been observed, attributed to a dopant deactivation in the NMOS device.

The SOTA CU/ULK interconnections has also been evaluated through reliability and yield measurements. Overall, the reliability is stable even up to 525°C 2h, due to a more mature BEOL process. However, as to keep the integrity of the yield, the thermal stability has been set up to 500°C 2h. These results pave the way to the introduction of BEOL between tiers in 3D sequential integration while the thermal budget allowed for the top tier is sufficient to lead to high performance device.

Lastly, the immunity of bottom device and interconnections upon an ultraviolet nanosecond laser annealing has been investigated through experimental work and numerical simulations. It has been demonstrated that even for the thin upper active Si layer (~7nm), the Cu resistivity and $I_{ON}-I_{OFF}$ remains stable. Additionally, 2D simulations helped to estimate the peak temperature seen in all the underlying stack. Therefore, it has been concluded that laser energy conditions including melt can be successfully reached without affecting State-Of-The-Art Cu lines for 28nm FDSOI devices.

- [1] C. Fenouillet-Beranger *et al.*, « FDSOI bottom MOSFETs stability versus top transistor thermal budget featuring 3D monolithic integration », *Solid-State Electron.*, vol. 113, p. 2-8, 2015.
- [2] P. Batude *et al.*, « 3D Sequential Integration: Application-driven technological achievements and guidelines », in *2017 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, déc. 2017, p. 3.1.1-3.1.4, doi: 10.1109/IEDM.2017.8268316.
- [3] M. Sellier *et al.*, « Predictive Delay Evaluation on Emerging CMOS Technologies: A Simulation Framework », in *9th International Symposium on Quality Electronic Design (ISQED 2008)*, San Jose, CA, USA, mars 2008, p. 492-497, doi: 10.1109/ISQED.2008.4479784.
- [4] C.-M. Lu *et al.*, « Key process steps for high performance and reliable 3D Sequential Integration », in *VLSI Technology, 2017 Symposium on*, 2017, p. T226-T227.
- [5] F. Deprat, « Etude et optimisation de la stabilité thermique du silicure et du beol intermédiaire pour l'intégration 3D séquentielle », PhD Thesis, 2017.
- [6] N. Planes *et al.*, « 28nm FDSOI technology platform for high-speed low-voltage digital applications », in *2012 Symposium on VLSI Technology (VLSIT)*, Honolulu, HI, USA, juin 2012, p. 133-134, doi: 10.1109/VLSIT.2012.6242497.
- [7] Qing-Tang Jiang, R. Faust, H. Lam, et J. Mucha, « Investigation of Ta, TaN and TaSiN barriers for copper interconnects », in *Proceedings of the IEEE 1999 International Interconnect Technology Conference (Cat. No.99EX247)*, San Francisco, CA, USA, 1999, p. 125-127, doi: 10.1109/IITC.1999.787098.
- [8] C. Fenouillet-Beranger *et al.*, « Guidelines for intermediate back end of line (BEOL) for 3D sequential integration », in *2017 47th European Solid-State Device Research Conference (ESSDERC)*, 2017, p. 252-255.
- [9] « Fabrication de CMOS à basse température pour l'intégration 3D séquentielle ».
- [10] F. Chen et M. Shinosky, « Addressing Cu/Low- κ Dielectric TDDB-Reliability Challenges for Advanced CMOS Technologies », *IEEE Trans. Electron Devices*, vol. 56, n° 1, p. 2-12, janv. 2009, doi: 10.1109/TED.2008.2008680.
- [11] N. L. Michael, C.-U. Kim, P. Gillespie, et R. Augur, « Mechanism of reliability failure in Cu interconnects with ultralow- κ materials », *Appl. Phys. Lett.*, vol. 83, n° 10, p. 1959-1961, sept. 2003, doi: 10.1063/1.1609242.
- [12] N. Suzumura *et al.*, « A New TDDB Degradation Model Based on Cu Ion Drift in Cu Interconnect Dielectrics », in *2006 IEEE International Reliability Physics Symposium Proceedings*, San Jose, CA, USA, 2006, p. 484-489, doi: 10.1109/RELPHY.2006.251266.
- [13] E. Chery, « Fiabilité des diélectriques low-k SiOCH poreux dans les interconnexions CMOS avancées », PhD Thesis, Université de Grenoble, 2014.

- [14] T. L. Tan, C. L. Gan, A. Y. Du, et C. K. Cheng, « Effect of Ta migration from sidewall barrier on leakage current in Cu/SiOCH low-k dielectrics », *J. Appl. Phys.*, vol. 106, n° 4, p. 043517, août 2009, doi: 10.1063/1.3202387.
- [15] D. D. Wang, W. L. Wang, M. Y. Huang, A. Lek, J. Lam, et Z. H. Mai, « Failure mechanism analysis and process improvement on time-dependent dielectric breakdown of Cu/ultra-low-k dielectric based on complementary Raman and FTIR spectroscopy study », *AIP Adv.*, vol. 4, n° 7, p. 077124, juill. 2014, doi: 10.1063/1.4890960.
- [16] S. Beaurepaire *et al.*, « 3D SEQUENTIAL TECHNOLOGY: RELIABILITY OF CU/LOW-K INTERCONNECTS AT HIGH TEMPERATURE », MRS Spring Meeting and Exhibit, 2019.
- [17] W. Skorupa et H. Schmidt, Éd., *Subsecond Annealing of Advanced Materials*, vol. 192. Cham: Springer International Publishing, 2014.
- [18] P. Batude *et al.*, « 3DVLSI with CoolCube process: An alternative path to scaling », in *2015 Symposium on VLSI Technology (VLSI Technology)*, Kyoto, Japan, juin 2015, p. T48-T49, doi: 10.1109/VLSIT.2015.7223698.
- [19] S. Kerdilès *et al.*, « (Invited) Sequential 3D Process Integration: Opportunities for Low Temperature Processing », *ECS Trans.*, vol. 80, n° 4, Art. n° 4, août 2017, doi: 10.1149/08004.0215ecst.
- [20] S. Kerdilès *et al.*, « Ultraviolet Nanosecond Laser Annealing for Low Temperature 3D-Sequential Integration Gate Stack », *ECS Trans.*, vol. 93, n° 1, p. 19-22, oct. 2019, doi: 10.1149/09301.0019ecst.
- [21] P. A. Alba *et al.*, « Nanosecond Laser Annealing for Phosphorous Activation in Ultra-Thin Implanted Silicon-On-Insulator Substrates », in *2016 21st International Conference on Ion Implantation Technology (IIT)*, 2016, p. 1-4.
- [22] C. Fenouillet-Beranger *et al.*, « New insights on bottom layer thermal stability and laser annealing promises for high performance 3D VLSI », in *Electron Devices Meeting (IEDM), 2014 IEEE International*, 2014, p. 27-5.
- [23] C. Fenouillet-Beranger *et al.*, « Ns laser annealing for junction activation preserving inter-tier interconnections stability within a 3D sequential integration », in *2016 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, Burlingame, CA, USA, oct. 2016, p. 1-2, doi: 10.1109/S3S.2016.7804375.
- [24] X. Garros *et al.*, « RF Performance of a Fully Integrated 3D Sequential Technology », in *2019 IEEE International Electron Devices Meeting (IEDM)*, 2019, p. 25-1.
- [25] A. Ortiz-Conde, F. G. Sánchez, J. J. Liou, A. Cerdeira, M. Estrada, et Y. Yue, « A review of recent MOSFET threshold voltage extraction methods », *Microelectron. Reliab.*, vol. 42, n° 4-5, Art. n° 4-5, 2002.
- [26] T. Kamins, *Polycrystalline Silicon for Integrated Circuits and Displays*. Boston, MA: Springer US, 1998.
- [27] H. P. Tuinhout, A. H. Montree, J. Schmitz, et P. A. Stolk, « Effects of gate depletion and boron penetration on matching of deep submicron CMOS transistors », in

- International Electron Devices Meeting. IEDM Technical Digest*, Washington, DC, USA, 1997, p. 631-634, doi: 10.1109/IEDM.1997.650463.
- [28] E. Josse *et al.*, « High performance 0.1 um pMOSFETs with optimized poly-Si and poly-SiGe gates », in *Solid-State Device Research Conference, 2000. Proceeding of the 30th European*, 2000, p. 548-551.
- [29] E. H. Nicollian et J. R. Brews, *MOS (metal oxide semiconductor) physics and technology*, Wiley classics library ed. Hoboken, N.J: Wiley-Interscience, 2003.
- [30] J. Schmitz, H. P. Tuinhout, A. H. Montree, Y. V. Ponomarev, P. A. Stolk, et P. H. Woerlee, « Gate polysilicon optimization for deep-submicron MOSFETs », in *Proc. ESSDERC*, 1999, p. 156-159.
- [31] F.-L. Chung et T. Lee, « EFFECT OF POLYSILICON DEPLETION ON MOSFET », *Electron. Lett.*, vol. 29, n° 13, Art. n° 13, 1993.
- [32] Y. F. Chong, H.-J. L. Gossmann, K. L. Pey, M. O. Thompson, A. T. S. Wee, et C. H. Tung, « Reduction of Polysilicon Gate Depletion Effect in NMOS Devices Using Laser Thermal Processing », *Electrochem. Solid-State Lett.*, vol. 7, n° 2, p. G25, 2004, doi: 10.1149/1.1632873.
- [33] R. Doering et Y. Nishi, *Handbook of semiconductor manufacturing technology*. CRC Press, 2007.
- [34] J. Foggiato, W. S. Yoo, M. Ouaknine, T. Murakami, et T. Fukada, « Optimizing the formation of nickel silicide », *Mater. Sci. Eng. B*, vol. 114-115, p. 56-60, déc. 2004, doi: 10.1016/j.mseb.2004.07.033.
- [35] D. Deduytsche, C. Detavernier, R. L. Van Meirhaeghe, et C. Lavoie, « High-temperature degradation of NiSi films: Agglomeration versus NiSi₂ nucleation », *J. Appl. Phys.*, vol. 98, n° 3, Art. n° 3, août 2005, doi: 10.1063/1.2005380.
- [36] B. Imbert *et al.*, « Nickel silicide encroachment formation and characterization », *Microelectron. Eng.*, vol. 87, n° 3, p. 245-248, mars 2010, doi: 10.1016/j.mee.2009.06.003.
- [37] K. Hoummada, C. Perrin-Pellegrino, et D. Mangelinck, « Effect of Pt addition on Ni silicide formation at low temperature: Growth, redistribution, and solubility », *J. Appl. Phys.*, vol. 106, n° 6, Art. n° 6, sept. 2009, doi: 10.1063/1.3204948.
- [38] T. Yamaguchi *et al.*, « Anomalous Gate-Edge Leakage Current in nMOSFETs Caused by Encroached Growth of Nickel Silicide and Its Suppression by Confinement of Silicidation Region Using Advanced Si^+ Ion-Implantation Technique », *IEEE Trans. Electron Devices*, vol. 56, n° 2, Art. n° 2, févr. 2009, doi: 10.1109/TED.2008.2010588.
- [39] T. Yamaguchi *et al.*, « Low-resistive and homogenous NiPt-silicide formation using ultra-low temperature annealing with microwave system for 22nm-node CMOS and beyond », in *2010 International Electron Devices Meeting*, San Francisco, CA, USA, déc. 2010, p. 26.1.1-26.1.4, doi: 10.1109/IEDM.2010.5703424.

-
- [40] F. Deprat *et al.*, « NiCo 10 at%: A promising silicide alternative to NiPt 15 at% for thermal stability improvement in 3DVLSI integration », 2015.
- [41] P. H. Chen *et al.*, « Investigation Pre-Amorphization Implantation on Nickel Silicide Formation », p. 4.
- [42] K. Romanjek, « Caractérisation et modélisation des transistors CMOS des technologies 50nm et en deçà », PhD Thesis, Institut National Polytechnique de Grenoble-INPG, 2004.
- [43] R. Berthelon, « Strain integration and performance optimization in sub-20nm FDSOI CMOS technology », PhD Thesis, 2018.
- [44] B. Mathieu, C. Fenouillet-Beranger, S. Kerdiles, et J.-C. Barbe, « Thermal simulation of nanosecond laser annealing of 3D sequential VLSI », in *2015 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Washington DC, USA, sept. 2015, p. 104-107, doi: 10.1109/SISPAD.2015.7292269.

Low temperature gate stack

The gate stack is one of the most important and complex bricks to optimize in a MOSFET transistor. The electrical behavior of the device strongly depends on the quality of the gate stack material, where a high-quality stack is essential to prevent degradation of the electrical parameters, such as undesired threshold voltage shifts, additional gate leakage, degraded subthreshold slope, mobility degradation via scattering phenomena, and an early degradation seen on bias temperature instability. In view of a 3D Sequential Integration, fabricating a high-quality oxide with a reduced thermal budget in the range of 500°C poses an additional challenge.

Despite the previous efforts done towards improving the quality of the gate stack for digital applications ($V_{DD}=0.8-1V$) where high-k is used, not necessarily the same techniques can be applied for an analog gate stack ($V_{DD}=2.5V$) friendly, composed of an $SiO_2=6nm$.

As a result, this section aims at giving guidelines for obtaining a low noise low-cost gate stack for More than Moore applications. An overview of the oxide defects as well as their impact on the device electrical characteristics, such as transport and reliability are also addressed. In-depth electrical characterization is provided, allowing one to have a better understanding about the nature of the charges and of the trap mechanisms observed in this work. Some analog figure of merit is also showed and compared to a high temperature process, showing the potential of the technology for some specific applications.

Lastly, based on the electrical results, guidelines are provided as to further improve the gate stack quality and reach the NBTI criteria.

3.1 CHALLENGES OF LOW TEMPERATURE GATE STACK

This section gives a short overview of the most recent solutions to achieve a high quality gate stack in view of a 3DSI for digital applications, where its main challenges can be summarized as follow [1]:

- maintaining the EOT throughout the fabrication process (specially to avoid EOT regrowth upon thermal treatments).
- reducing leakage currents (which can also be enhanced depending on thermal treatment).
- obtaining an excellent interface between the gate oxide and the channel.
- controlling of the threshold voltage (in particular for PMOS transistors).
- ensuring device reliability.

As it has been highlighted on chapter 1 and despite of the many advantages previously mentioned, SiO₂ suffers from a relatively low dielectric constant ($\kappa_{\text{SiO}_2}=3.9$), which limits the coupling effects and leads to it not being used as the gate oxide material in the integration of advanced technological nodes. In order to comply with CMOS scaling trend, a high-k material such as hafnium oxide (HfO₂) has replaced the SiO₂ layer [2]. However, this integration solution also brought new challenges, because the high-k oxides differ from SiO₂ by not being an intrinsically low defect density material. In addition, the quality of the Si/high-k interface suffer from higher interface states compared to Si/SiO₂. One simple solution to this issue is to use a thin SiO₂ layer as an interfacial layer between the Si channel and the high-k, generally formed before the high-k deposition. The drawback of this method is that this Si layer contributes to increase the EOT value of the stack, undesired for high performance applications. Despite, the later has been proved efficient for interface states reduction.

In a low thermal budget scheme, this SiO₂ layer is achieved with a chemical oxide (T~500°C), but the later still presents a higher D_{it} (4e11/cm²/eV) w.r.t. the HT (~1000°C) reference process (2e10/cm²/eV), thus requiring additional post-deposition treatments to improve it. Passivation by deuterium annealing instead of conventional forming gas has led to 1 decade D_{it} reduction [2], thus close to the HT reference process.

Other than the D_{it} , the TB reduction of 3DSI has also been investigated concerning the flatband voltage and gate leakage. Concerning the flatband voltage, it has been demonstrated that reducing the TB up to 525°C 5h has led to +100mV shift, which could be useful in reducing PMOS threshold voltage without resorting to complex gate last integration. The latter is also beneficial for the compromise between EOT and gate leakage [1].

Concerning gate stack reliability, which is extremely sensitive to process temperature, the reliability of top devices is a major concern when decreasing temperatures below 800°C [3]. Even though PBTI has not necessarily been a concern in a low TB process flow [1], NBTI remains an issue.

It must be highlighted that in a 3DSI, the TB reduction of the subsequent steps in the process flow also can negatively impact the quality of the stack. For example, in a gate-first integration of a HT process, the high-k layer is exposed to the source/drain (S/D) activation anneal (~1000°C), which reduces its defect densities and minimizes BTI. Clearly this anneal is not possible in a 3DSI, where the S/D activation anneal is in the 500°C range usually done by Solid Phase Epitaxy Regrowth (SPER). Another TB step which also have an impact is the one of the Polysilicon gate. The latter is deposited amorphous (a-Si), thus requiring an anneal for its crystallization. It has been showed that high temperature anneal ($T=1100^{\circ}\text{C}$) over very short time ($t=0.5\text{ms}$) performed by DSA after the a-Si gate deposition led to a significant gain in reliability with a minor EOT increase of 0.3\AA [4]. However, 1100°C 0.5ms DSA is still quite high in terms of temperature and time for bottom MOSFET stability in 3DSI. As a solution, a ultra-violet nanosecond laser annealing (UV-NLA) could be very promising for the crystallization of the amorphous layer with a more acceptable TB [5].

There are however other strategies when trying to overcome NBTI issues: instead of improving the stack quality or adjusting the TB of the fabrication steps, it is possible to try to overcome the high-density defects.

One alternative is to engineer dipoles at the interface between SiO_2 and HfO_2 (by depositing a thin dipole-former interlayer with ALD) to shift up or down the energy levels of the defects with respect to the Si conduction or valence band, for nMOS and pMOS reliability, respectively [6]. This method is based on reducing the density of accessible defects to a level comparable to the high temperature high-k/metal gate stack of a

commercial 28 nm technology. The dipole material choice and thickness can be properly adjusted for the type of MOSFET, where the following has been proposed for nMOS (LaSiO_x) and pMOS (Al_2O_3). In the case of nMOS, this method has also been efficient to improve channel electron mobility due to reduced carrier-defect interaction. As for the pMOS, to compensate the effective work function shift induced by the Al_2O_3 dipole, a more mid-gap work function metal such as TaN could be used instead of TiN.

More in the scope of this thesis, two approaches have also showed potential towards NBTI improvement, the use of Silicon Germanium (SiGe) channel and of Junction-less (JL) transistors [6].

It is well known that SiGe channel (also called Ge channel quantum well (QW)) can be used on pMOS devices for enhanced mobility and threshold voltage tuning, but the later has also been considered for NBTI improvement. It must be highlighted that even though the SiGe solution has initially been proposed to overcome NBTI issues resulting from aggressively scaled EOT [7] (because the thinner oxide leads to increased oxide electric field (E_{ox})), the later could potentially be used for LT stacks. The NBTI improvement is ascribe to the SiGe band gap narrowing to Si, related to a favorable alignment shift of the Fermi level in the SiGe channel w.r.t. the pre-existing oxide defect energy levels.

This effect causes misalignment of the hole injection energy with respect to pre-existing gate oxide defect energy levels, meaning that larger misalignment can cause the carriers to interact with a smaller fraction of accessible oxide traps. However, it is known that the SiGeO_x interfacial layer presents a wide distribution of defect levels due to Ge suboxide formation, which is particularly detrimental for high Ge fractions. It has been argued that this high density of negatively charged acceptor-like traps close to the SiGe E_v lowers E_{ox} and reduces NBTI degradation at fixed overdrive [8]. Nevertheless the resulting high D_{it} is not beneficial for other FOM (such as noise). A solution is to epitaxially grow a thin Si passivation layer (Si cap) on top of a pMOS SiGe channel, drastically improving NBTI reliability whilst also offering good interface quality (Fig. 3.1). Because of the E_v offset between the SiGe and the Si cap, inversion channel holes are confined in the SiGe layer, which therefore acts as a QW for holes. The efficiency of this method has been demonstrated on a standard high-k/metal gate stack consisting of a SiO_2 interfacial layer grown by oxidation of the Si cap and a HfO_2 layer [9] [10]. Integration schemes shall be proposed further down.

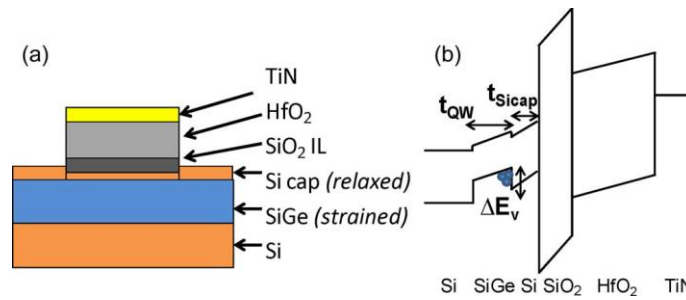


Figure 3.1 (a) Sketch of the gate stack of SiGe devices and (b) band diagram schematic in inversion. Channel holes are confined into the SiGe QW due to the valence-band offset ΔE_v between the SiGe channel and the Si cap [11].

Another interesting approach consists on overcoming the large density of defects in the low TB oxides by replacing inversion mode devices with highly doped junction-less transistors [6], achieving improved reliability due to its lower operating oxide fields. Additionally, these devices are inherently thermal-budget-friendly as they do not require S/D activation and that the channel doping is activated before bonding of the top Si level to the bottom tier passivation oxide. Further details on JL shall be discussed on a later section.

It must be highlighted that whilst these approaches have been proved successful for a high-k stack, its viability is yet to be discussed for an analog stack.

3.1.1 Challenges for analog stack

This section now aims at summarizing the additional challenges and requirements directed to an analog gate stack. It is well-established that the gate oxide quality and reliability strongly depend on the type of high- κ and interfacial oxide layer, the deposition method and the implementation of possible post-deposition annealing or passivation treatments.

In the stack used in this work, conventionally analog transistors are of G02 type, which consists of a gate oxide composed of SiO₂ only of 5 to 6.5nm thick, as to sustain the higher $V_{DD}=2.5V$ (instead of 0.9-1V for digital applications). As such, the high-k can be suppressed, which in turn can be particularly beneficial for a few analog requirements, shortly described below.

Concerning the noise component, as it has been highlighted on chapter 1, random fluctuations in signal current or voltage are a main limitation of the accuracy of a measuring device and ultimately sets a lower limit on signals that can be detected and

processed. There are several noise sources that affect the MOSFETs, such as thermal noise, shot noise, generation-recombination noise and RTN, where a complete description can be found elsewhere [12]. In the scope of this thesis, more attention shall be given to low-frequency noise, also called $1/f$, which is very dependent on the geometry of the transistor and the CMOS technology used. The LF noise is becoming a major concern for continuously scaled down devices, since the $1/f$ noise increases as the reciprocal of the device area [13].

Noise can be attributed to the random process of trapping and de-trapping of charges in the oxide traps near the Si/SiO₂ interface, thus being a parameter which is extremely sensitive to the presence of traps and charges in the gate dielectric. In theory, if this interface is of high quality, the noise component (in particular a reduction in trap density) should be decreased. Whereas the introduction of novel gate dielectrics for digital applications can pose new challenges for noise reduction, the high-k suppression of this stack benefits from the superior interface quality of SiO₂ with the Si channel.

Another issue is related to the introduction of nitrogen to reduce gate leakage and to inhibit boron penetration (from the Polysilicon) in thin oxide devices, which has also degraded BTI and mobility. Thanks to the thicker oxide of this work, this step can also be suppressed.

Slower scaling of operating voltages for digital circuits compared to more aggressive scaling of oxide thickness has gradually increased the effective field across the oxide, which in turn has enhanced NBTI [14]. Also, at a given oxide field, thin oxide devices were found to be more susceptible to NBTI than their thick oxide counterparts. So in theory, the SiO₂ 6nm only could potentially present better NBTI results.

Overall, Fig. 3.2 depicts the gate stack composition and the potential challenges for digital (previous work) and for analog applications (this work). At this point, it has become clear that the proposed stack used on this work can have promising results. Nevertheless, it is still necessary to properly fabricate a high quality SiO₂, which can be a challenge in view of a 3DSI (TB in the range of 500°C).

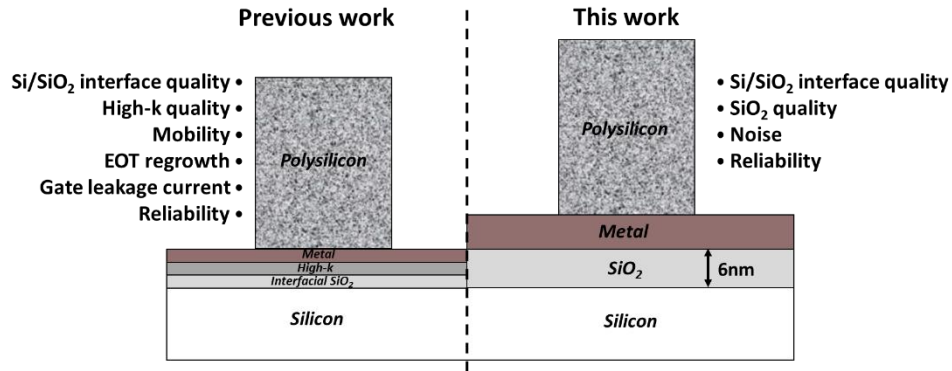


Figure 3.2 Schematic representation of the studied gate stack for digital (previous work) applications as compared to the one focused on analog applications (this work). Besides each case, the main studied requirements are highlighted.

3.2 OVERVIEW OF MICROSCOPIC OXIDE DEFECTS

Silicon dioxide (SiO₂) is a widely used material in microelectronics and can for example, be employed to isolate one device from another, to serve as a structured mask against implant of dopant atoms, and more specifically, is a great candidate to be used as the gate dielectric in MOSFET structures [15]. Amongst its advantages, one can mention its high thermal stability, ease to implement, and more importantly, the low density of defects generated at the interface with the silicon channel (especially true for a thermal oxide w.r.t. deposited oxide).

In a standard process with no thermal budget constraints, the SiO₂ employed in the gate stack is usually fabricated at high temperatures (between 800°C and 1000°C) in a thermal oxidation process, where silicon and oxide react to form SiO₂ [16]. This reaction however induces a sudden crystallographic change between crystalline Si and amorphous SiO₂, which gives origin to interface defects. Other than interface defects, there are also other defects in the bulk of the SiO₂ that can originate either during the oxidation process or afterwards throughout the various stages of the MOSFET manufacturing. All of these defects can be electrically active, thus deteriorating the short and/or long-term device performance. Despite, these defects are always less than if compared with a LT process.

Electrically active defects are defined as atomic configurations which give rise to electronic states in the oxide band gap that can trap carriers. Typically, these are sites of excess or deficit of oxygen or impurities.

Overall, there are different points for classifying defects in the gate oxide such as their chemical nature, their energy distribution or their spatial depth [17], [18]. Additionally, depending on their respective energy level in the band gap the traps can exchange charges at low fields or very high fields. In the scope of this thesis, the most common defects observed in MOSFET devices shall be divided into interface, border, and deep oxide traps, as depicted on Fig. 3.3.

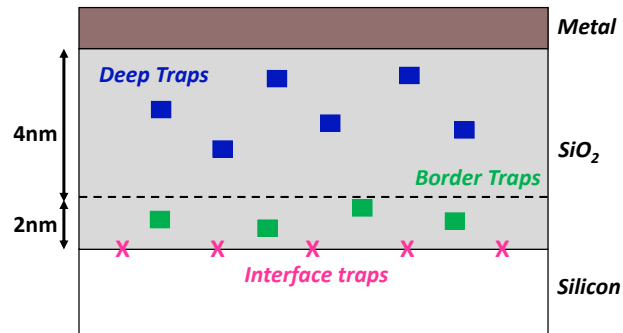


Figure 3.3 Schematic spatial depth representation of the gate oxide defects mentioned on this thesis, where the following traps are highlighted: interface traps located in the interface with the silicon channel, border traps, located in the first 1-3nm of the SiO_2 , and lastly the deep traps, located in the bulk of the oxide.

3.2.1 Interface traps

Interface states, as the name says it, exist at the Si/ SiO_2 interface and are a result of structural imperfections. The silicon atom possesses four valence electrons and therefore requires four bonds to fully saturate the valence shell. In the crystalline structure, each silicon atom establishes bonds to its four neighboring atoms, leaving no unsaturated bond behind. There are however missing atoms at the surface of the silicon crystal, where the unpaired valence electrons exist and form electrically active interface traps [19]. After oxidation, most interface states are saturated with oxygen bonds, but there are still some silicon atoms in the substrate that fail to form four covalent bonds. As such, an interface trap corresponds to an interface trivalent Si atom with an unsaturated (unpaired) valence electron, the so-called dangling bond [20]. The exact properties of the interface defects, also denoted P_b centers, depends on the atomic configuration and substrate orientation. P_b centers are found at (111) interfaces, while two distinct defects, referred to as P_{b0} and P_{b1} , are related to (100) interfaces, depicted on Fig. 3.4.

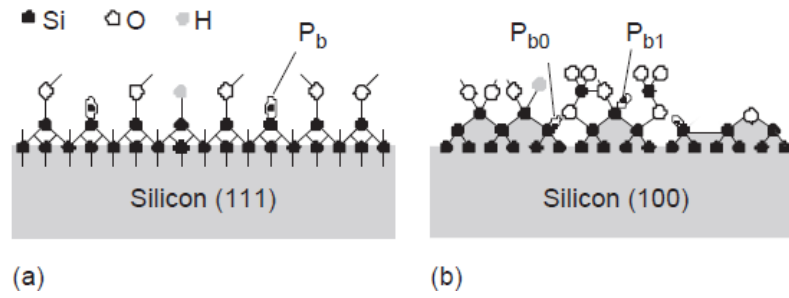


Figure 3.4 Structural representation of the (a) (111) Si surface and (b) (100) Si surface. The P_{b0} center, electrically very similar to the P_b center, is back-bonded to three Si atoms, whereas the P_{b1} center is back-bonded to two other Si atoms and an oxygen atom [17].

All three types of silicon dangling bonds mentioned (P_b , P_{b0} and P_{b1}) are of amphoteric nature, meaning that these can take two distinct states in the silicon bandgap. Interface traps at the SiO_2/Si interface are acceptor-like in the upper half and donor-like in the lower half of the Si band gap.

The traps of the donor-like energy levels can be either electrically neutral when occupied by an electron or positively charged when empty. The traps of the acceptor-like energy levels can be either electrically neutral when empty or negatively charged when occupied by an electron. In both cases, their occupancy and charge depend on the position of the fermi level. For a better understanding, Fig. 3.5 depicts the band diagrams of a p-type and n-type channel MOS device, respectively, showing the occupancy of interface traps at flatband and inversion.

In the case of the p-type channel MOS device: At flatband (Fig 3.5.a), the electrons occupying states below the Fermi energy, which means in the lower half of the band gap are neutral (occupied donors designated by “0”). Those between mid-gap and the Fermi energy are positively charged (unoccupied donors designated by “+”), and those above mid-gap are neutral (unoccupied acceptors). At inversion (Fig 3.5.b), the fermi level is now located above the mid-gap energy at the Si/SiO_2 interface. As such, the donor like trap levels in the lower half of the band gap are therefore completely filled and the acceptor like trap levels are partially filled, resulting in a negative interface charge.

Likewise, in the case of the n-type channel MOS device: At flatband (Fig 3.5.c), the electrons occupying states below the Fermi energy, which means in the lower half of the band gap are neutral (occupied donors designated by “0”). Those between mid-gap and the Fermi energy are negatively charged (occupied acceptors designated by “-”), and

those above Fermi energy are neutral (unoccupied acceptors). At inversion (Fig 3.5.d), the Fermi level is now located below the mid-gap energy at the Si/SiO₂ interface. As such, the donor-like trap levels in the lower half of the band gap are therefore partially filled and the acceptor-like trap levels are empty, resulting in a positive interface charge. For interface traps, the communication of the substrate electrons or holes with interface traps is predominantly by capture/emission.

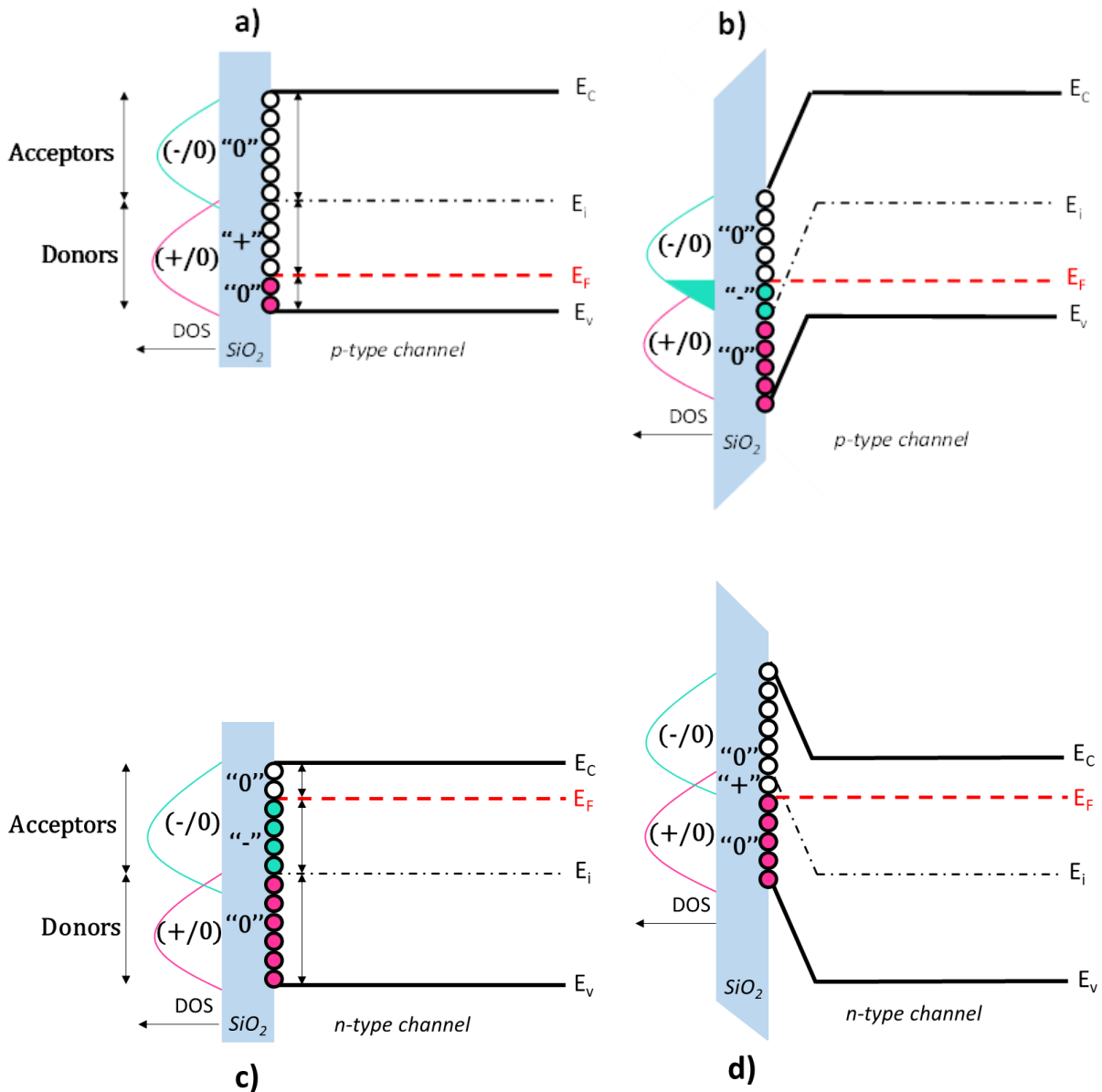


Figure 4.5 Band diagrams of the Si substrate showing the occupancy of interface traps and the various charge polarities at flat band and at inversion for a) p-Substrate and b) n-Substrate. Each of the small horizontal lines represents an interface trap. It is either occupied (solid circle) or empty (empty circle).

3.2.2 Border traps

Border traps are near-interfacial oxide traps, located within approximately 1~3nm of the Si/SiO₂ interface [21]. Similarly to interface traps, border traps can also communicate with the inversion layer in the channel by exchanging charge carriers. These are commonly considered the cause of random telegraph and 1/f noise, as well as slow charging and discharging transients causing threshold voltage shifts in the bias temperature instability. Microscopically, border traps are often associated with E' centers (trivalent silicon dangling bonds in the oxide [22]–[24]) but have also been related to hydrogen related defects (such as hydrogen bridge and hydroxyl E' center) [24].

3.2.3 Deep traps

Deep or bulk traps are considered to be located in the bulk of the SiO₂ and do not rapidly communicate with the Si. Deep traps have been observed as being one of the responsible for bias instabilities in NMOS devices [25] with high-k materials such as HfO₂. For example, it has been said that in the HfO₂, oxygen vacancies induce levels in the bandgap, where these levels are dependent on the charge state of the vacancy. That said, the process of vacancy formation and the electrical phenomena induced by them are dependent on the state of charge and these levels. An oxygen vacancy can have five charge states, such as negatively charged (V_o^{2-} , V_o^-), neutral (V_o^0), and positively charged (V_o^+ , V_o^{2+}) [26]. The importance of this example is to highlight that oxygen vacancies can present a different energy level, which is also true for the SiO₂ [27].

3.3 METHODS TO IMPROVE THE OXIDE QUALITY

Generally, in order to have a high-quality oxide, thermal treatments are often applied to reduce oxide defect densities by process control and annealing. Post oxidation thermal annealing can potentially lead to the annihilation of defects such as oxygen vacancies and to the exo-diffusion of hydrogen, which were incorporated during growth. Additionally, it can also help to reduce interface roughness and limit dangling bonds. It must be noted that other than thermal annealing, other processes such as plasma treatments can also be beneficial for the passivation of dangling bonds. The passivation of interface states tends to reduce the number of free electrons that will no longer participate in conduction once the Si-H bond is formed. Passivation does not annihilate all defects at the interface, but it is efficient enough to drop the defect below an electrically measurable level.

Interface passivation usually refers to the Si/SiO₂ interface passivation by hydrogen termination, which consists of the diffusion of hydrogen, through the oxide, to react with the P_b centers and form bonds. Hydrogen is highly effective to passivate the dangling bonds due to its high reactivity (Fig. 3.6). It migrates through the lattice and reacts with dangling bonds of defects and impurities rendering them electrically inactive. This is the basis of the widely used technique of forming gas anneal to reduce the interface trap density in MOSFET devices. The Si-H bond is a very stable entity that has no levels in the energy band gap of silicon. It must be noted that an optimum efficiency of this process and its resulting improvement on device performance depends on the treatment conditions and when it is performed during device manufacturing process.

In addition, annealing around 400°C under an N₂/H₂ atmosphere (also called FGA - forming gas anneal) is often carried out in order to passivate the dangling bonds with hydrogen. This annealing is generally carried out towards the end of the integration at the level of the formation of the contacts in order to avoid their de-passivation (i.e. breakage of the bond) by high thermal budgets. This is because even though the Si-H bonds are stable, it can eventually break if energy is provided to it. One drawback of the technique is that passivation by hydrogenated species can have various impacts on reliability because it will then be possible to release these species during the aging of the device. Additionally, there are processes that may result in the removal of H from the Si-H bond and the reappearance of dangling bonds that can act as carrier traps. One such process is the so-called de-passivation reaction.

Passivation by deuterium is an alternative process to passivate the dangling bonds, where the main advantage is that under electrical stress, bonds to deuterium are more difficult to break than Si bonds to hydrogen. Therefore, treatment with deuterium instead of hydrogen in CMOS fabrication provides a possible solution for the aforementioned reliability issues.

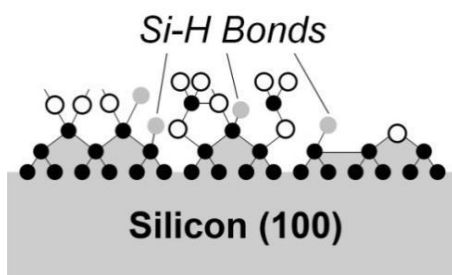
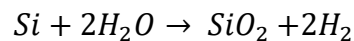
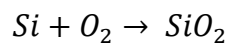


Figure 3.6 Passivation of interface states tends to reduce the number of free electrons which will no longer act as an electrically active trap once the Si-H bond has formed. Passivation does not completely eliminate defects at the interface, but it successfully reduce the defects amount below an electrically measurable level [17].

3.4 OXIDE SPLITS DEFINITION

Thermal oxidation is a complex process, where the oxidation of silicon takes place by diffusion followed by the bonding of oxygen in the network, where more details can be found elsewhere [16]. The thermal oxidation of a silicon surface is usually performed at high temperatures (800°C–1200°C), resulting in a High Temperature Oxide (HTO) layer. The chemical reaction of silicon and oxygen starts spontaneously at room temperature but stops after a very thin native oxide layer. The quality and the growth of the oxide during the manufacturing process can be controlled and are strongly influenced by the choice of oxidizing agent, the temperature and pressure, and the crystalline orientation of the substrate. The oxidation can either be "wet" by using water vapor or steam, or "dry" with an atmosphere composed of oxygen only. The reactions that take place during a wet and dry thermal oxidation are respectively described below:



A dry oxidation has as characteristics a slow growth rate and high oxide density if compared with a wet oxide (thermal, dry/wet 2.27/2.18 g/cm³). On the other hand, wet oxidation presents a superior growth rate compared to dry, which is its biggest advantage when thicker oxides are envisaged. The reason for the much higher growth rate is the oxidant solubility limit in SiO₂, which is much higher for wet (H₂O) than for dry oxidation (O₂).

The inability to use a temperature higher than 500°C makes that a thermal oxidation is not a viable option in this work, firstly because of the high temperature of the method, and also because the oxidation rates are very low for temperatures below 700°C. As such, one must explore alternative methods for obtaining the SiO₂ of the order of 6nm with satisfactory uniformity of thickness and that do not lead to interfaces with silicon rich in dangling bonds.

3.4.1 Low temperature oxides

As it has been highlighted, thermal oxidation temperature is well above the maximum allowable TB in view of a 3DSI scheme, such that alternative methods must be employed. An SiO₂ layer can also be achieved with an oxide deposition, which normally involves a much smaller TB than thermal oxidation. The main disadvantage of this alternative is that

the Si/SiO₂ interface is not as good as the one obtained with a thermal oxidation and furthermore, deposited oxides does not have the same high density as thermal grown oxides.

In this work, three types of LT oxides (<500°C) have been compared. The first one corresponds to a 6nm plasma oxidation in O₂/H₂ (named PI-O₂/H₂), obtained with a PECVD (Plasma Enhanced Chemical Vapor Deposition) μ wave remote plasma at 450°C. The benefit of a remote plasma [28] is that the substrate is remote from, i.e. not in direct contact with the plasma, so the deposited film and substrate in a remote PECVD are not subjected to ion bombardment during deposition. The plasma is generated outside of the process chamber, and so the radicals are then transported towards the chamber containing the substrate, thus limiting the damage linked to the plasma interaction. [29]

The second split (named PI-O₂+PEALD) starts with 1.6nm plasma oxidation in O₂ at 400°C. This dry oxidation has a very low deposition rate, where the target of 6nm cannot be achieved. This first oxide layer is then completed with + 4.4nm obtained with a Plasma Enhanced Atomic Layer Deposition (PEALD) at 200°C. The third split corresponds to a 6nm PEALD only.

Post Deposition Treatments (PDT) has also been tested, such as a plasma densification on O₂ at 450°C and a furnace anneal on N₂ at 500°C. Lastly, a High Temperature Oxide (HTO) at 1000°C is used as reference. A schematic of the corresponding process flow and its oxide splits nomenclatures can be found in Fig. 3.7. It must be highlighted that the entire process flow is within the 500°C temperature range, with the maximum temperature being imposed by the polysilicon deposition at 525°C, which can be further reduced and shall be discussed in another section.

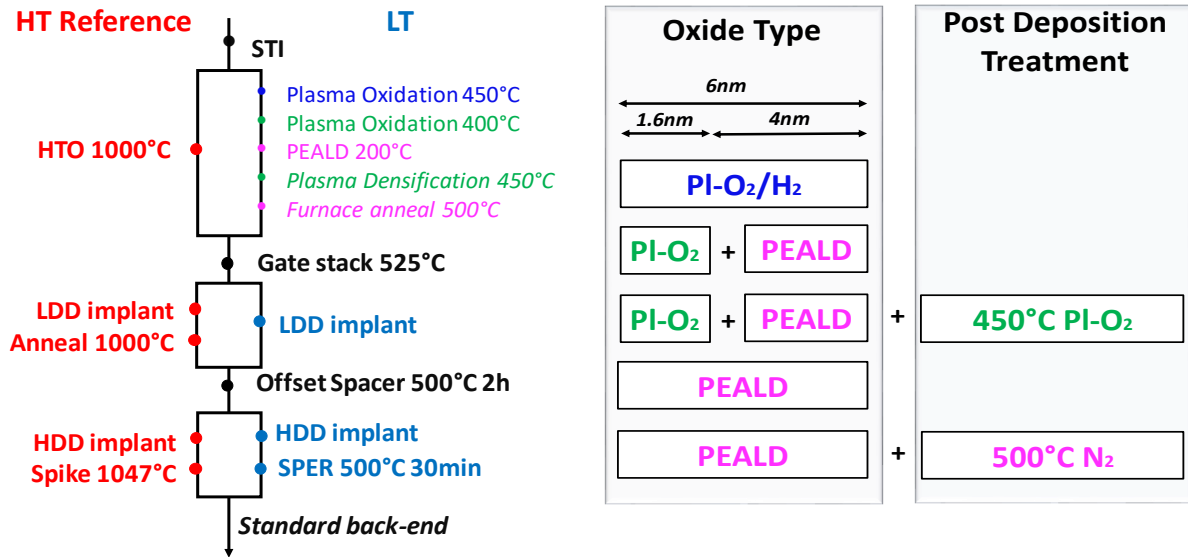


Figure 3.7 Detailed process flow of HT and LT devices and b) summary of the LT oxide with its respective split name. Owing to the thicker SiO₂, different oxide deposition are used in comparison to advanced digital devices. Post deposition treatments such as plasma densification and furnace anneal were also tested. A more detailed description of the complete process flow can be found on chapter 1.

3.4.2 OX from BOX

In view of a 3DSI, a way to obtain the silicon channel for the top level comes from a SOI reported wafer. In this process, once the bottom and top level are bonded together, the future Si channel of the tier 2 is obtained as followed: the top Si of the reported wafer is removed by grinding, followed by a chemical treatment to selectively remove the leftover Si with stopping on the buried oxide. Lastly, another chemical treatment is performed, this time to selectively remove the buried oxide with stopping on the Si. The remaining Si is then used as the top active for the fabrication of the tier 2 transistor. The main advantages of the technique comes from the excellent Si crystalline quality and its accurate thickness control [30]. Based on this method, a similar process can be done but this time to use the buried oxide as the gate oxide of the tier 2. In this thesis, the later will be called "OX from BOX". The processes are quite similar, the difference being that once the remaining Si of the reported wafer is selectively removed w.r.t. the SiO₂, the later needs to be chemically thinned to the desired oxide thickness. Both these options are schematized and compared on Fig. 3.8.

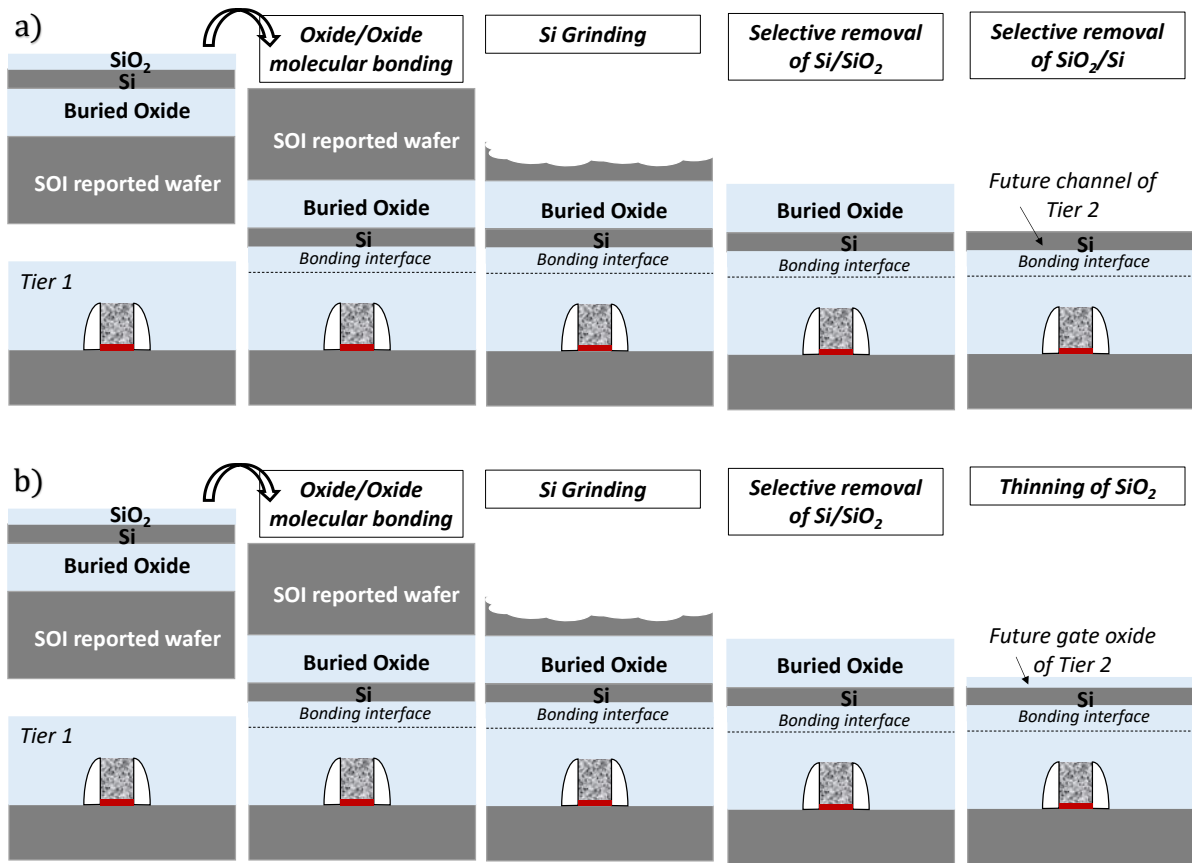


Figure 3.8 Schematic representation of bonding process to a) obtain a high quality crystalline silicon channel for top tier device [30] and b) to obtain high SiO₂ layer for the gate stack of tier 2.

In this thesis, the viability of this process was simulated starting with 8nm of thermal oxide processed at 800°C, followed by a wet etching with TEAH, finishing by an hf 0.5% treatment to thin the SiO₂ and obtain the desired thickness of 6nm. This process is schematized on Fig. 3.9. It must be highlighted that the TEAH is there to simulate the removal of the leftover Si substrate with stop on the BOX. Lastly, the plasma densification PI-O₂/H₂ was employed. The impact of this chemical treatment on the electrical characteristics of the device, if any, shall be further analyzed.

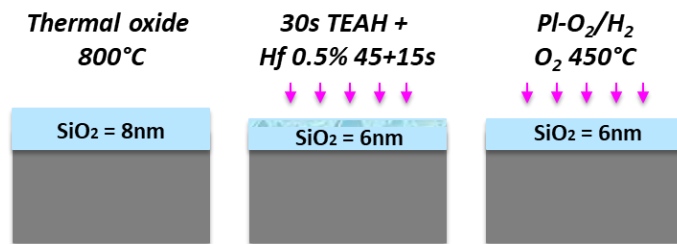


Figure 3.9 Schematic representation of the OX from BOX flow simulated on this work.

3.4.3 Choice of TiN

Titanium nitride has been widely used in gate first integration, because it is a mid-gap material, it has a high thermal stability and as previously mentioned it is necessary for the high-k integration [31]. A point to consider when studying the gate is to consider the ensemble of the stack ($\text{SiO}_2/\text{TiN}/\text{Poly}$), not only the gate oxide. There are two types of TiN that can be used, called TiN M2 or TiN M4.

With a TiN M2, the target itself is composed of TiN, meaning that the entire thickness of the layer will be composed of TiN. Contrarily, with a TiN M4, the process starts with a Ti rich target, and the nitridation is done during the deposition process. As a result, only Ti is sputtered from the surface of the target for the first monolayers deposited over the SiO_2 . As nitrogen covers the target surface, TiN begins to sputter, and nitrogen composition gradually increases in the TiN film being deposited. This means that there is a gradient profile of nitrogen in the layer. The drawback of this TiN M4 is that it could potentially lead to scavenging effects, because the first monolayers are composed of TiO instead of TiN. Scavenging effect is attributed to the diffusion of oxygen towards the TiN, due to the high solubility of oxygen in titanium nitride (TiN). This has been observed on $\text{SiO}_2/\text{HfO}_2/\text{TiN}$ stacks, where the strong affinity of TiN for oxygen makes it possible to reach EOTs of the order of 1 nm.

In this work, a TiN M4 was used. However, to ensure that there is no scavenging effects, a TiN M2 is preferable.

3.5 ELECTRICAL RESULTS

When analyzing the gate stack, it is crucial to have both the physical and electrical thickness of the gate oxide.

Post-oxidation or post-deposition, the oxide thickness can be measured by ellipsometry, an optical method that allows one to measure the thickness of the oxide (T_{ox}). The technique is based on using polarized monochromatic light to illuminate the oxidized silicon surface at an angle, which is then reflected from both the silicon substrate and the oxide surface. As a result, it outputs the angles related to the refractive index of the silicon substrate, the refractive index of the silicon oxide, and the thickness of the film. T_{ox} gives

a good indication as to knowing if the oxide thickness is within the target (in this work of 6nm).

After the lot fabrication, it is necessary to know if the thickness of the oxide remained the same, so another measurement must be done. A common method to extract the EOT relies on capacitance and current measurements is the so-called CV split method [32].

In this batch, the polysilicon was doped at the same time as the implantation of the S/D (HDD implants, the reader may relate to chapter 4 for more information) was performed. This implant however was not optimized to properly dope the polysilicon in a LT scheme (500°C). As a result, the polysilicon is not properly doped and presents a high resistance. Combined with the fact that a thicker gate oxide of 6nm was used, the resulting capacitance should be smaller than the one usually measured on G01 devices, where the relation between the oxide capacitance and thickness is given by the relation:

$$C_{OX} = \frac{\epsilon_{SiO_2}}{T_{OX}} \quad (\text{Equation 3.1})$$

Where ϵ_{SiO_2} is the dielectric constant ($3.89 \times 8.85 \cdot 10^{-12}$ F/m). To characterize this small value, one needs to lower the frequency, but very low frequencies end up being in the noise range, complicating the analysis and leading to erroneous EOT values. To overcome this issue, bigger devices must be used, but even for the larger one of $5 \times 5 \mu\text{m}$, it was not enough. Then, capacitance networks of ($2 \mu\text{m} \times 2 \mu\text{m}$) were used instead, with 200 connected fingers, thus increasing the overall device area. Lastly, upon a CV measurement, there is always a parasitic capacitance from the wiring that must be normalized. In our case, this step was not used, because the overlap capacitances of the devices were too high and this normalization would also lead to incorrect EOT results.

Fig. 3.10 displays the extracted EOT values and compares it with the T_{OX} ones measured by ellipsometry. It can be seen that even though the trend is the same, there is roughly a 2nm discrepancy between methods. In theory, for a SiO_2 gate stack, EOT and T_{OX} should be the same. which can be explained by different propositions: (1) the ellipsometry measurements were not properly calibrated or (2) oxygen scavenging from the TiN. Concerning the variation of EOT among the different oxide splits, it could be attributed to a variation of deposition rate of the techniques.

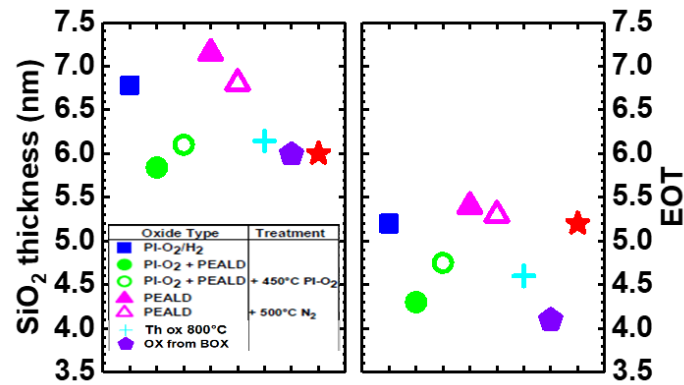


Figure 3.10 Comparison of T_{OX} measured by ellipsometry and EOT values obtained by CV split method.

The fact that there is a variation of EOT between the wafers is going to directly have an influence on other electrical parameters, highlighted here for the V_{TH} and the I_{ON} current. Even though there was not a specific V_{TH} target envisaged for this work, Fig. 3.11 displays the V_{TH} as a function of EOT for where a N-type ground plane (GP) was used in both nMOS and pMOS devices. The V_{TH} was extracted by constant current method for low V_D ($V_D=100\text{mV}$), at the given current value ($I_{th}=100\text{nA}\cdot W/L$). The red dashed line depicts the expected V_{TH} values obtained from TCAD electrical simulation. The overall observation is that the V_{TH} of some splits are either shifted upwards or downwards w.r.t. the expected TCAD line. The later can be attributed to negative or positive fixed oxide charges, thus shifting the V_{TH} accordingly. These charges do not change state with the underlying silicon channel, and their density, which depends on the oxidation temperature and ambient atmosphere, tends to decrease with increasing final oxidation temperature. In the case of the nMOS, it is also interesting to point it out that opposite fixed oxide charges are displayed on the LT splits (negative fixed oxide charges) in comparison to the OX from BOX case (positive fixed oxide charges). This is clearly an effect attributed to the chemical treatment to thin down the oxide.

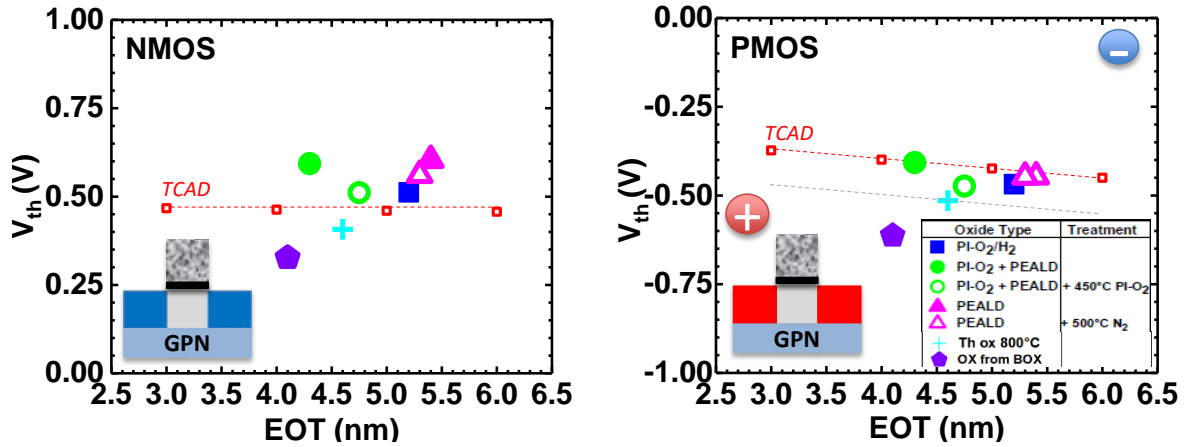


Figure 3.11 Threshold voltage of nMOS and pMOS devices. The shift in comparison to the expected TCAD simulated value is attributed to fixed oxide charges.

The saturation current I_{ON} ($V_G=V_D=V_{DD}=2.5V$) in an SOI MOSFET can be expressed as follow:

$$I_{ON} = \frac{1}{2n} \frac{W}{L} \mu \cdot C_{OX} (V_{GS} - V_{TH})^2 \quad (\text{Equation 3.2})$$

Where n stands for the body effect coefficient, defined in [33]. According to the equation above, it is expected that the V_{TH} discrepancy between splits shall shift the I_{ON} accordingly. However, simply displaying the I_{ON} on V_G-V_{TH} does not account for the impact of C_{OX} on the wafers. As such, the I_{ON} presented here is normalized by C_{OX} and given in ($\mu A/\mu m/F/cm^2$). The I_{ON}/I_{OFF} trade-off is shown (Fig. 3.12) for both N&P devices, characterized for $L=1\mu m$ and $W=2\mu m$. In both cases, a clear I_{ON} shift is observed for all oxide splits w.r.t HTO used as a reference. Given that all the LT splits have the same S/D process, the effect is attributed to the gate stack properties, analyzed independently in the following part. Concerning the I_{OFF} regime, no significant increase has been observed for the NMOS, whereas a 1 decade increase is seen for PMOS.

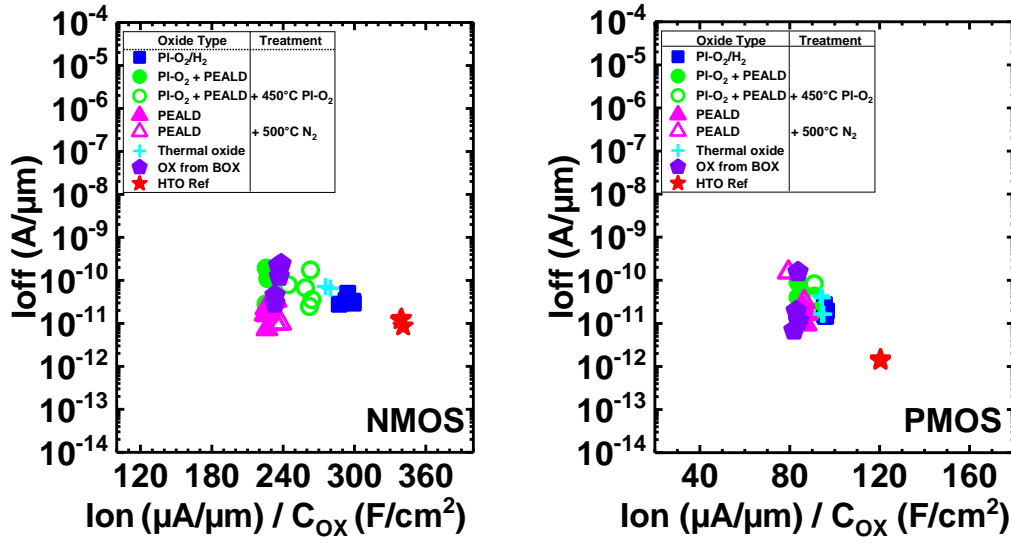


Figure 3.12 I_{ON}/I_{OFF} trade off of N&P devices for all LT oxide splits and HTO reference. $L = 1\mu\text{m}$ $V_{DD} = 2.5\text{V}$. I_{ON} shift between LT splits is attributed to the gate stack dielectric.

In order to extract the oxide features, several techniques were used, and will be shortly described. Since capacitance array were used instead of transistors, the mobility was not able to be extracted with a classical method based on CV split. Instead, the **low field mobility** (μ_0) was extracted from the Y function [34], based on the combined exploitation of $I_D(V_G)$ and $g_m(V_G)$ in the linear regime. By definition, Y function is expressed as:

$$I_{DS} = \frac{W}{L} C_{OX} \frac{\mu_0}{1+\theta_1(V_{GS}-V_{TH})} (V_{GS} - V_{TH}) V_{DS} \quad (\text{Equation 3.3})$$

With W and L the width and length of the device, μ_0 the mobility extrapolated at low-field, θ_1 the 1st order attenuation factor of mobility. The transconductance g_m is given by:

$$g_m = \frac{dI_{DS}}{dV_{GS}} = \frac{W}{L} C_{OX} \frac{\mu_0}{(1+\theta_1(V_{GS}-V_{TH}))^2} V_{DS} \quad (\text{Equation 3.4})$$

The Y function is defined by:

$$Y = \frac{I_{DS}}{\sqrt{g_m}} = \sqrt{\frac{WC_{OX}\mu_0V_D}{L}} (V_{GS}-V_{TH}) \quad (\text{Equation 3.5})$$

As such, the low-field mobility μ_0 can be determined from the slope of the Y function without R_{SD} effects. Because μ_0 is extracted with a relatively small V_G , it is mainly affected by Coulomb scattering.

The **interface trap density** (D_{it}) was calculated from the Subthreshold Slope (SS), given that in the presence of a significant D_{it} ($D_{it}=q2C_{it}$), SS can be sensitively degraded by

definition. By knowing C_{ox} and the Subthreshold Slope (SS), the D_{it} on FDSOI devices can be calculated as followed [35]:

$$SS = \frac{kT}{q} \ln(10) \left[\left(1 + \frac{C_{it1}}{C_{ox1}} + \frac{C_{si}}{C_{ox1}} \right) - \frac{\frac{C_{si}}{C_{ox2}} \frac{C_{si}}{C_{ox1}}}{1 + \frac{C_{it2}}{C_{ox2}} + \frac{C_{si}}{C_{ox2}}} \right] \quad (\text{Equation 3.6})$$

With C_{it1} $C_{it1}=qN_{it1}$, $C_{it2}=qN_{it2}$, C_{ox1} and C_{ox2} the top interface trap capacitance, the bottom interface trap capacitance, the front gate oxide capacitance, and the bottom gate oxide capacitance, respectively. C_{si} is the silicon film capacitance, and is defined by $C_{si}=\epsilon_{si}/t_{si}$, where t_{si} is the silicon film thickness.

Concerning **low frequency noise measurements** (LFN) performed at 25°C [13], the carrier number fluctuation (CNF) with correlated mobility fluctuation (CMF) is applied to the average normalized noise at $f=12\text{Hz}$ and in the strong inversion region where the noise spectrum showed $1/f$ behavior:

$$\frac{S_{I_d}}{I_d^2} = \left(\frac{g_m}{I_d} \right)^2 S_{Vfb} \left(1 + \Omega \frac{I_d}{g_m} \right)^2 \quad (\text{Equation 3.7})$$

By extracting the flat-band voltage power spectral density (S_{vfb}) from the fitting of normalized noise with the CNF/CMF model, the oxide trap volumetric density (N_t) is extracted using the eq. below:

$$S_{Vfb} = \frac{q^2 \lambda k T N_t}{W L C_{ox}^2 f \gamma} \quad (\text{Equation 3.8})$$

Where λ is the tunnel attenuation distance (~ 0.1 nm), kT is the thermal energy, C_{ox} the oxide capacitance per unit area, f the frequency, W and L the width and length of the transistor gate respectively and γ is the characteristic exponent (in this case equals 1 because the normalized noise is extracted in a frequency where the spectra is $1/f$ like).

Bias Temperature Instability (BTI) is also characterized using a fast measurement technique [36]. The characterization is usually performed at elevated temperatures ($T=125^\circ\text{C}$) and stress voltages (in this case, $V_{\text{stress}}=3.6\text{V}$, 4V , 4.4V and 4.8V) to accelerate the aging of the device. In CMOS technology, the reliability criteria can be defined as the maximum allowable V_{TH} shift of 50mV after a 5 years operation at the circuit operating voltage, in this case for a $V_{DD}=2.5\text{V}$. In order to compare the oxide splits more rigorously, the Time-To-Failure (TTF) shall be plotted as a function of the oxide field, given by the equation:

$$E_{OX} = \frac{V_{stress} - V_{TH}}{EOT} \quad (\text{Equation 3.9})$$

3.5.1 NMOS Evaluation

Fig. 3.13 a-c depicts the maximum low-field mobility (μ_0), D_{it} and N_t for NMOS devices. It is observed that higher D_{it} values are extracted for the LT oxides w.r.t. HTO. This results in an enhanced Coulomb scattering effect, thus also degrading the mobility. As a consequence, the Pl-O₂/H₂ split exhibit better mobility amongst the LT splits, explained by its lower $D_{it}=2 \times 10^{-11} \text{cm}^{-2} \text{eV}^{-1}$. The latter also presents the lower $N_t=3 \times 10^{18} \text{cm}^{-3} \text{eV}^{-1}$. The fact that the split Pl-O₂/H₂ presents overall superior results when compared to the other LT variations could be attributed from a better passivation of the dangling bonds by H⁰ incorporated during the plasma oxidation step [14]. Concerning PDT efficiency, better surface passivation is achieved with a reactive plasma O₂ applied on Pl- O₂+PEALD, while the effect is limited for the furnace anneal N₂ split. Further improvements of the Si/SiO₂ interface can be achieved with a forming gas under HPD2 anneal [15]. Finally, PBTI is investigated (Fig. 3.13-d). PBTI in LT oxides are comparable, or even better for the Pl-O₂/H₂, w.r.t HTO. It also largely meets the 5 years requirement at $V_{DDmax}=2.5V$. From these results, it becomes clear that N_t and PBTI are not correlated. This is because the traps responsible for both phenomena are not same. Noise is mainly sensitive to D_{it} while PBTI is due to trapping in a band of “deep” defects localized above the silicon conduction band.

Lastly, it should be mentioned that the HTO underwent a post-deposition nitridation, commonly performed on these types of oxide. Even though superficial, this could explain the superior PBTI results of the Pl-O₂/H₂ w.r.t. the HTO.

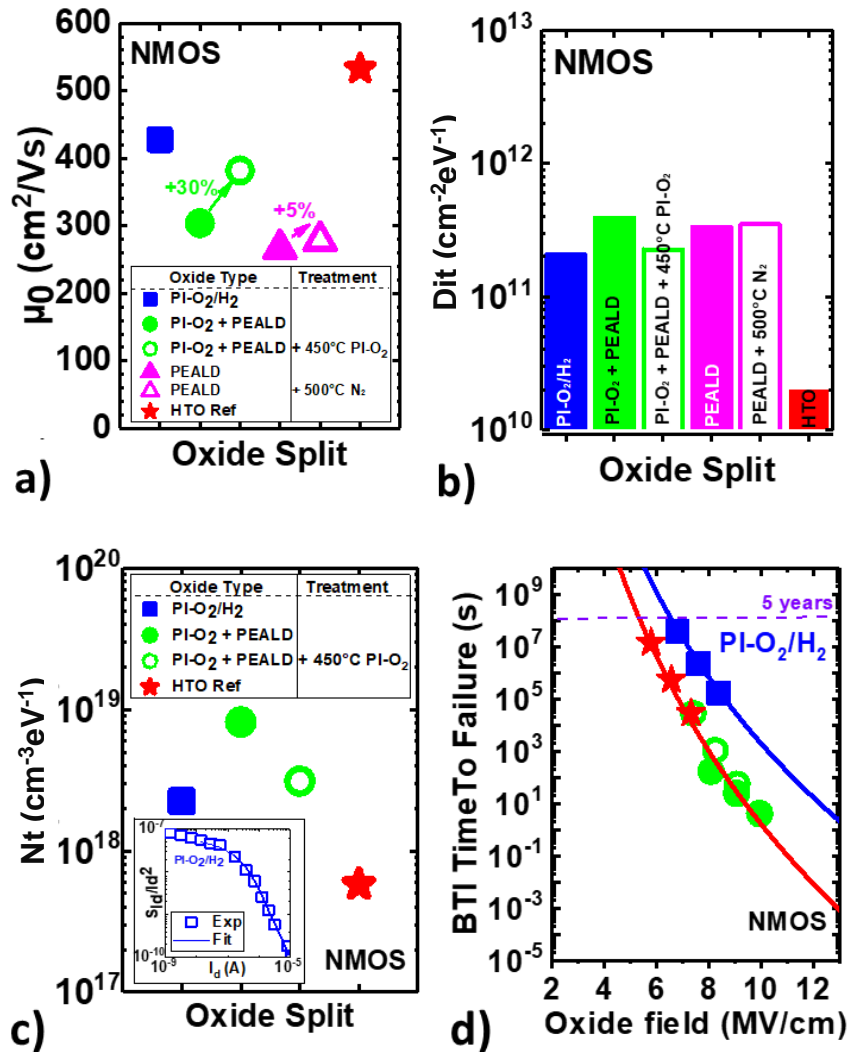


Figure 3.13 Mobility peak μ_0 . b) Calculated interface state density. c) Bulk oxide trap density deduced from LFN measurements with the CNF/CMF model. Inset: fit of the normalized drain current noise S_{i_d}/I_d^2 with drain current I_d . d) PBTI TimeToFailure vs Oxide field. All oxides fulfill the 5 years requirements (purple dashed line). ΔV_{th} of 50mV is used as criterion for BTI lifetime prediction. The grey square area represents the forbidden zone, where the criteria is not reached.

3.5.2 PMOS Evaluation

Fig. 3.14-a shows the low field field mobility, where all LT oxides are degraded w.r.t. the HTO. Fig. 3.14-b-c report D_{it} and N_t densities. It is highlighted that both are (1) much higher than the values observed for NMOS and the HTO and (2) almost independent of the LT oxide. This asymmetry between N&P implies that not only silicon dangling bonds contribute to D_{it} and N_t , but also border traps. This means that the D_{it} values observed for PMOS devices actually refers to the sum of interface and border traps, and not to pure dangling bonds, with the method to extract D_{it} from the SS not being able to separate the two contributions. Moreover, these N_t values are well correlated to NBTI results [Fig. 3.14 d)], suggesting that, this time, the defects responsible for noise and NBTI are the same.

Comparing the oxides, the TTF for the best LT oxide PI- O₂/H₂ is still degraded by ~4 decades w.r.t. HTO. Additionally, no improvement has been achieved with PDT. This confirms that NBTI remains a major issue for LT gate stacks [15-17].

The NBTI degradation has been associated to traps located in the interlayer and high-k [17]. Despite the high-k suppression of our stack, NBTI remains problematic, which could be an indication that the traps are actually mainly located in the SiO₂.

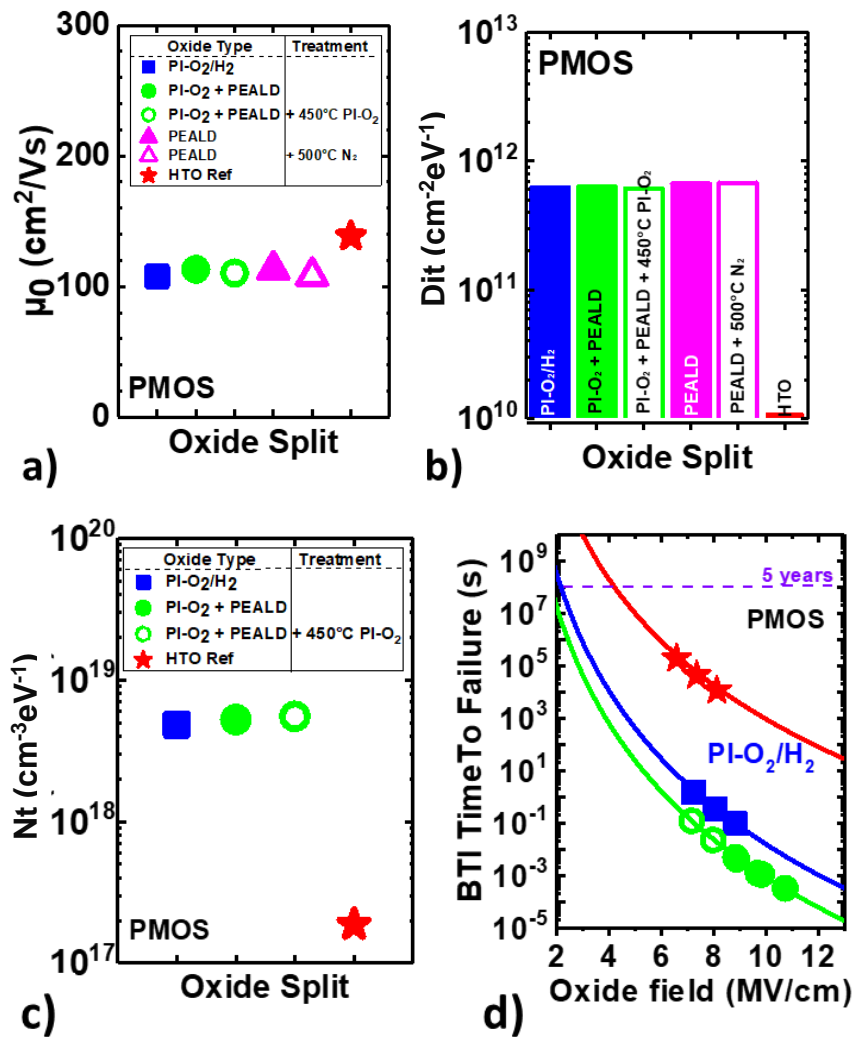


Figure 3.14 a) Mobility peak μ_0 . b) Calculated interface state density. c) Bulk oxide trap density N_t d) NBTI TimeToFailure vs oxide field. Unlike NMOS, no significant differences between LT splits are seen on the mobility. Also, all splits present a larger amount of D_{it} and N_t and no improvement after PDT is observed. The NBTI remains critical for all LT oxides - extrapolated 5y TTFs fall inside the forbidden zone.

3.5.3 Trapping Mechanisms in LT Oxides

Three types of oxide traps have been identified: “deep”, border and interface traps (Fig. 3.15). For NMOS, the majority of active traps are interface/dangling bonds located in the bandgap, which affects mobility features. There are also some high energetic “deep” traps

in the bulk of the SiO₂, responsible for electron trapping. These deep traps only become energetically available upon high field, such as the one used during PBTI stress. For PMOS, in addition to the dangling bonds, border donor traps must be considered. These are located close to the Si/SiO₂ interface (~1nm depth), and affect all the device characteristics such as SS, noise and NBTI. It can thus be said that NBTI in PMOS is controlled by the filling and creation of both Si dangling bonds and hole border traps [37].

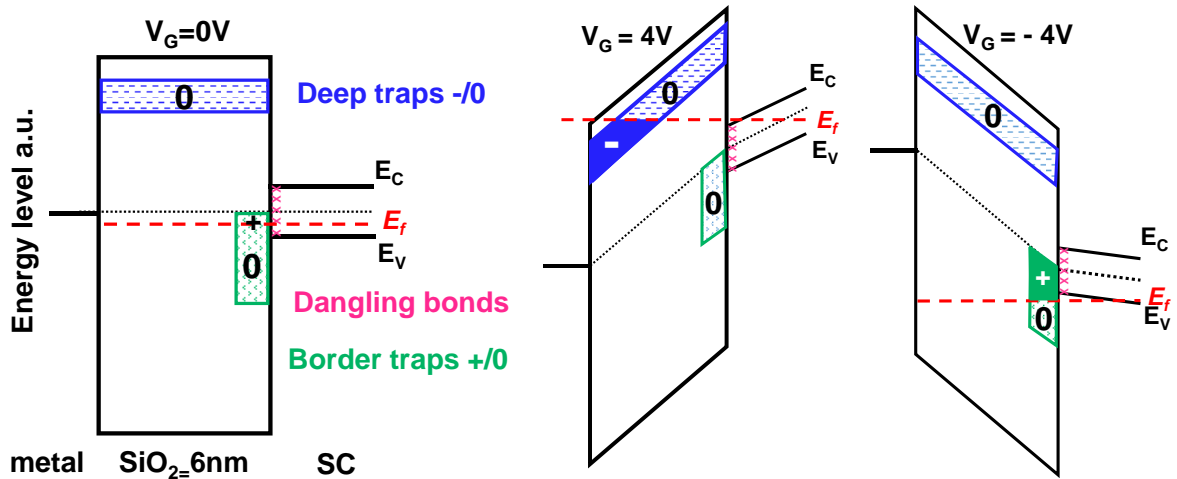


Figure 3.15 Summary of oxide traps impacting device performance and reliability. a) 3 types of traps are identified namely deep -/0, border +/0 and interface defects. b) PBTI in NMOS is driven by e^- trapping in high energetic defects c) NBTI in PMOS is controlled by the filling & creation of both Si dangling bonds and hole border traps.

3.5.3.1 NBTI Discussion

This section aims at further discussing the NBTI degradation observed on LT devices. It has been highlighted that the latter is not only due to the degradation of the Si/SiO₂ interface, but it is also strongly related to the capture of carriers by defects in the gate oxide. One possible method to further exploit BTI degradation is by building the Capture and Emission Time (CET) map [38], which represents the trap density for a given combination of capture and emission time. This map can thus be seen as a reflex of the stress and relax characteristics.

As such, Fig. 3.16 depicts the CET map of the PI/O₂+PEALD split, where two trap populations have been identified, a quasi-permanent and recoverable. It can be seen that the quasi-permanent degradation has a very high emission for any capture time and could be mainly due to interface defects. The second observed trap population is a recoverable degradation, which presents a correlation between capture and emission, and it could be

mainly due to oxide traps. These observations agree with the proposed oxide traps impacting LT devices reliability previously discussed.

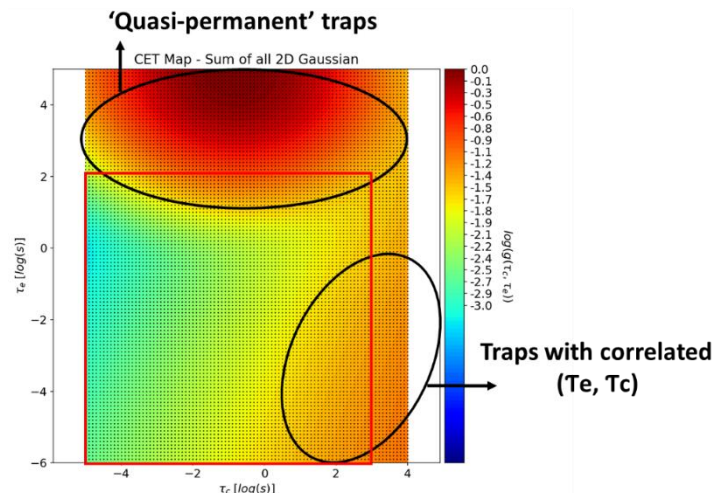


Figure 3.16 CET map revealing the two populations according to their time constants. The map was built with the RC model and performed @ 125°C for $V_{G, stress} = -4,8$ V, $T_{stress} = 1$ ks and $T_{relax} = 100$ s.

It must be stressed that this CET map is only indicative, such that for a better comprehension of the traps mechanisms, further tests should be performed for different stress times and much more dies. Additionally, the tests should be done as well on the HT reference. This would allow one to compare if the traps present in LT devices w.r.t. HTO (1) are different or (2) if they are of the same type but in a higher density.

3.6 ELECTRICAL RESULTS OF OX FROM BOX

This section will now analyze the OX from BOX results. The same methods previously described used on the LT splits are also used in this case.

3.6.1 NMOS Discussion

Fig. 3.17 a-c depicts the maximum low-field mobility (μ_0), D_{it} and N_t . Even though only a small increase on D_{it} and N_t values is observed, there is still an increase. Likewise, -20% on the μ_0 is also seen. It should also be highlighted that even the thermal oxide is degraded in all FOM w.r.t. the HTO reference. Fig. 3.17-d depicts the PBTI shift in decades w.r.t. the HTO reference. Similarly, to the LT oxides, the PBTI target is also reached for the OX from BOX split. The PBTI of the thermal oxide could not be extracted, because the measurements were in the noise level, meaning that to finally see a degradation, a higher stress voltage (>4.4V) should have been applied. Despite reaching the target, one can

assume that the chemical treatment indeed has an impact in the volume of the oxide as well, otherwise the PBTI measurements for the OX from BOX split would also be in the noise range.

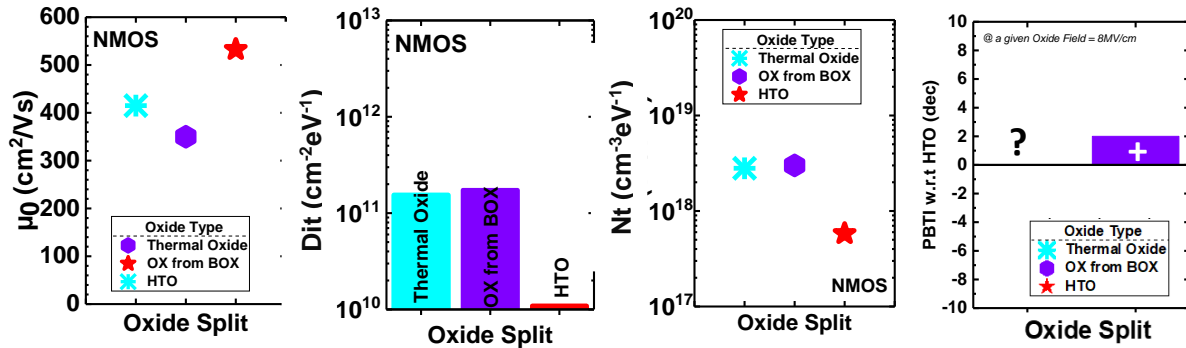


Figure 3.17 Mobility peak μ_0 . b) Calculated interface state density. c) Bulk oxide trap density deduced from LFN measurements with the CNF/CMF model. d) PBTI Time To Failure value with respect to the HTO reference, given for a fixed oxide field of 8MV/cm.

3.6.2 PMOS Discussion

Similarly, Fig. 3.18 a-c report μ_0 , D_{it} and N_t densities. Concerning the μ_0 or D_{it} , the results are pretty much similar to those observed in the LT oxides: no significant differences between the thermal oxide and the OX from BOX, whereas both are degraded w.r.t. HTO reference. Both are also much higher than for NMOS. As for the N_t , the OX from BOX split presents the highest value observed so far of $3 \times 10^{19} \text{ cm}^{-3}\text{eV}^{-1}$. Likewise, this split also presents by far the worst NBTI degradation (Fig. 3.18-d), where a degradation of -7 decades w.r.t. HTO is observed.

It must be noted that for all the FOM shown so far, for N&P, the thermal oxide is never aligned to the HTO. This could be attributed to its lower oxidation temperature w.r.t. HTO (800°C instead of 1000°C) and absence of post-oxidation annealing. For future studies, either another high-quality oxide should be used as a starting base for the OX from BOX split, or a thermal annealing should be done after the oxidation, to compensate its lower temperature of 800°C. As such, it must be stressed that the chemical treatments undergone on the OX from BOX did not possibly have the best oxide to begin with, which could partially explain its large degradation.

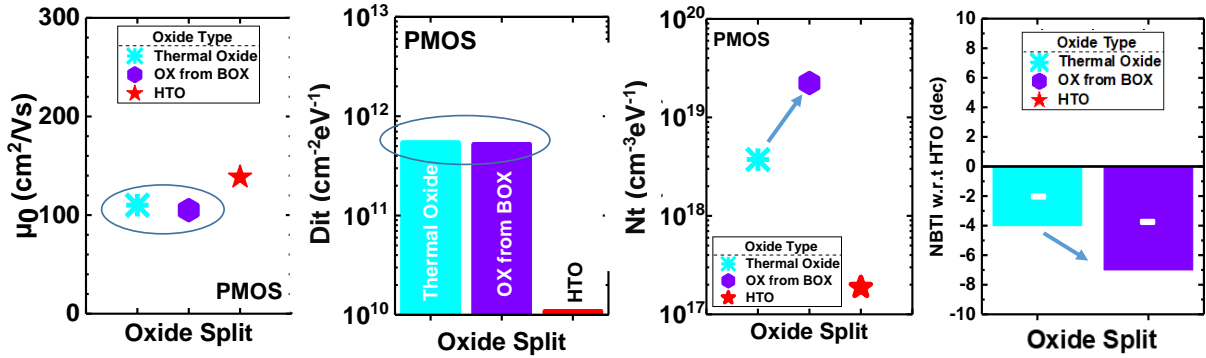


Figure 3.18 a) Mobility peak μ_0 . b) Calculated interface state density. c) Bulk oxide trap density N_t , d) NBTI Time To Failure value with respect to the HTO reference, given for a fixed oxide field of 8MV/cm.

3.6.3 Trapping Mechanisms on OX from BOX

Based on the electrical results, it could be said that the chemical treatment does have a negative impact on the performance of the device. According to N&P results, it seems that the main impact of the chemical thinning is on the BULK of the oxide. Apparently, this treatment goes deep enough in the oxide to create more border traps (as observed on the NBTI and N_t of PMOS), also energetically located near the valence band edge.

3.6.4 Fourier-transform Infrared Spectroscopy

As to have more insights about the degradation of the OX from BOX split, physical-chemical analysis were performed on full sheet wafers. The technique used is Fourier-transform infrared spectroscopy (FTIR), which allows one to identify the functional groups present in organic and inorganic compounds by measuring their absorption of infrared radiation over a range of wavelengths. As a result, one can identify the type of molecular bond present in the spectrum by their corresponding peak. For this work, the same processes as observed on the electrical results were reproduced, such as the thermal oxide 800°C and the OX from BOX. Additionally, other processes are also compared, such as a thermal oxide + a thermal annealing, the OX from BOX split without the $\text{PI-O}_2/\text{H}_2$ densification, and lastly a proposed optimization for the OX from BOX. In the optimization case, the OX from BOX is done on a better quality oxide as an initial template (supposedly) and the thinning is achieved solely with a more superficial hf 0.5%. This is done because there are suspicious of a possible contamination by TEAH that would diffuse in the SiO_2 layer. Table 3.1 summarizes the respective splits.

Split name	Oxide	Thermal Treatment	Chemical Treatment	Thermal Treatment
OX from BOX	Thermal Oxide 800°C		30s TEAH + Hf 0.5% 45+15s	
OX from BOX + PI-O ₂ 450°C	Thermal Oxide 800°C		30s TEAH + Hf 0.5% 45+15s	Plasma O ₂ 450°C
OX from BOX opt	Thermal Oxide 800°C	1050°C 30min	HF 0.5%	
Thermal Oxide	Thermal Oxide 800°C			
Thermal Oxide + anneal	Thermal Oxide 800°C	1050°C 30min		

Table 3.1 depicts the splits where the FTIR was performed.

Fig. 3.19 shows the first molecular bonding analyzed, the Si-O-Si, located at ~ 1240 - 1230cm^{-1} range. This band has several components, but in ATR (Attenuated total reflectance), the one who is excited the most is the optical longitudinal mode (LO), showed in the inset of Fig. 3.19. The position of the LO mode depends a lot on the quality/disorder of the layer (composition, stress), where usually the higher the value, the better is the layer in general. Good quality oxides can be found with a $\text{LO} \sim 1240\text{cm}^{-1}$. Surprisingly, the lowest $\text{LO} = 1228.2\text{cm}^{-1}$ is showed for the thermal oxide + anneal. This comes as a striking result, given that this should be the best oxide. It is highlighted on the Fig. 3.19, that the splits present different thicknesses, ranging from 8.3nm (thermal oxide + anneal) to 5.9nm (OX from BOX). This discrepancy has an impact on the FTIR results, where simulations shows that a maximum shift up to 5cm^{-1} can occur. This explains why the striking results at first, where the unexpected LO shift is a result of SiO_2 thickness.

That said, it does not seem that the additional anneal of 1050°C 30min led to any improvement of the oxide, and that roughly all the splits are somehow equivalent. Despite, one can compare the impact of the $\text{PI-O}_2/\text{H}_2$ on the OX from BOX split, where it seems that this plasma treatment actually degraded a bit the quality of the oxide. This could be possible and attributed to the bombarding of the plasma itself on the substrate.

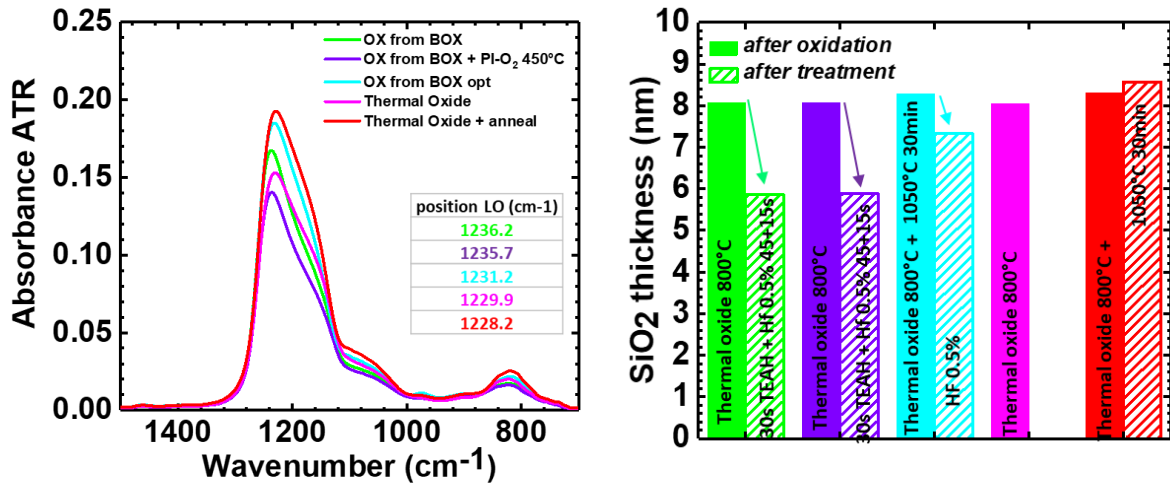


Figure 3.19 Shows on the left the absorbance for the Si-O-Si bond, where the inset highlights the position of the LO band. On the right, it is displayed the SiO₂ oxide thickness measured by ellipsometry before oxidation and after thermal treatment.

Another interesting band is the Si-OH, usually found on “wet” oxides and are associated to the creation of defects during electrical stress. Concerning the Si-OH band displayed on the left of Fig 3.20 and located ~980cm⁻¹, the flatter the peak, the less Si-OH bond the layer contains. This is the case for the thermal oxides and the OX from BOX + PL-O₂/H₂, contrarily to the splits without any thermal treatment (OX from BOX and OX from BOX opt). These results do not explain the BTI or N_t degradation, because on FTIR all the layer thickness is probed, one cannot make a distinction if these bonds are at the interface or not. Lastly, the right of Fig. 3.20 shows the Si-H band, where no conclusion can be draw because the spectra is on the noise range.

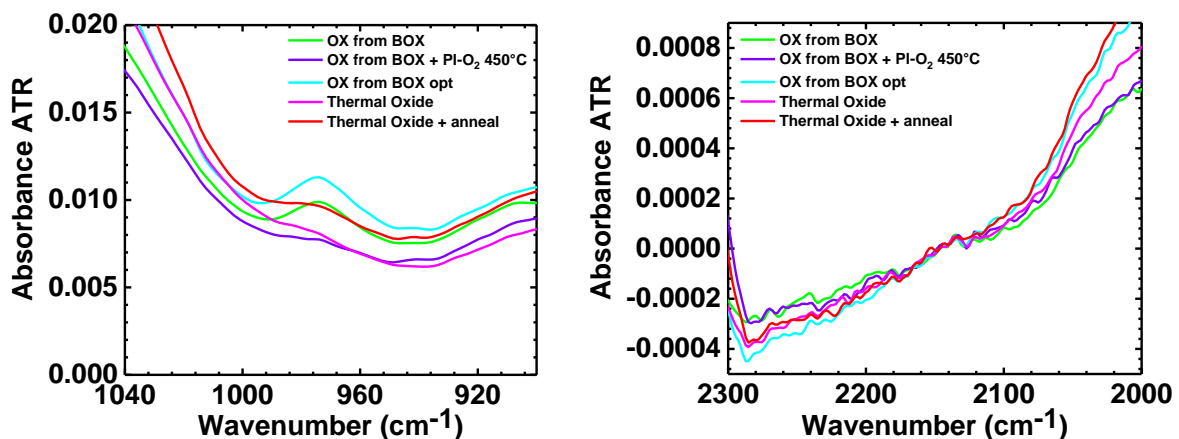


Figure 3.20 Shows on the left the absorbance for the Si-OH bond, whereas the Si-H bond is shown on the right.

Overall, even with the physical-chemical results, it is hard to address what caused the oxide degradation observed on the electrical results. Only a full integration on an electrical lot can say if the OX from BOX is truly a viable solution.

3.7 IMPACT OF LT GATE STACK ON ANALOG FOM

On this section, some analog figure of merit shall be compared for the best LT oxide split ((PI-O₂/H₂) and the HTO reference. Firstly, a very useful approach for analog performance assessment of different split is calculating gm/I_D as a function of normalized drain current. Such plot gives a complete picture of the studied device, which is valid for different applications, depicted on Fig. 3.21 for N&PFETs. The ratios of the devices are compared when the devices are operating in weak or moderate inversion (as for low-frequency base-band applications) and for devices in strong inversion (as for high-frequency applications).

For base-band applications, where high gain, high precision is needed, it is observed that in the region below threshold ($I_D=1 \times 10^{-9} \text{A}$), gm/I_D values in the range of 30V⁻¹ are found for both HTO and the LT device. Additionally, the same voltage gain is observed between long and short L_G (L_G=1μm and L_{G,nom}=0.3μm), offering a range of utilization for circuit designers. Given that the gm/I_D is a reflex of the subthreshold slope (hence, impacted by oxide features), its stability up to L_{G,nom}=0.3μm is highlighted on Fig 3.21-b, shown for PFETs.

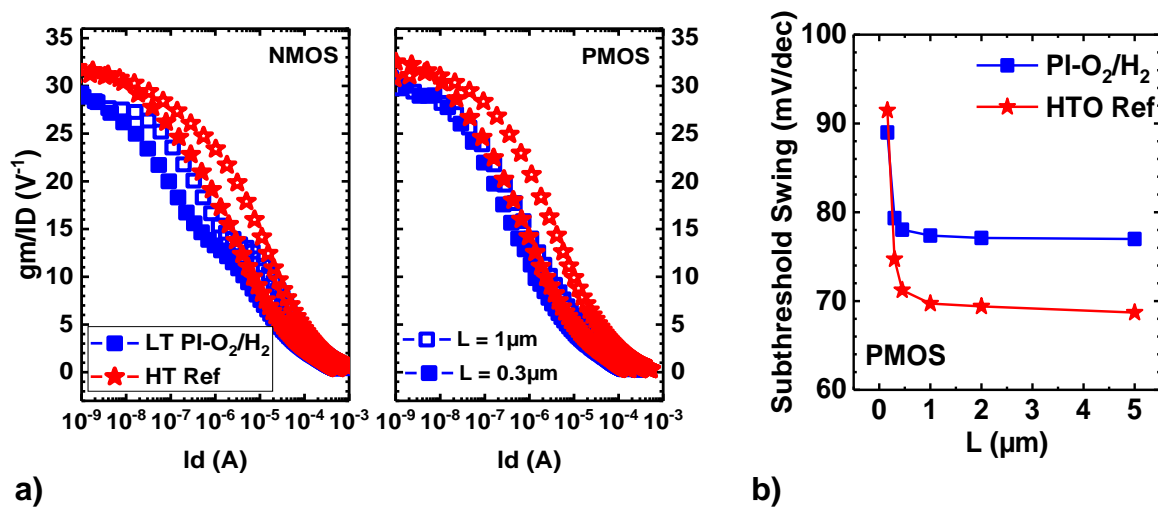


Figure 3.21 (a) N&P FETs gm/Id ($V_d=0.9\text{V}$) of the HTO and the best LT split PI-O₂/H₂. ($L_g= 1\mu\text{m}$ and $L_g = 0.3\mu\text{m}$), (b) subthreshold swing as a function of gate length shown for PMOS.

For high-frequency applications, where high drive current is requested, the devices are compared biased at a fix g_m/I_D of $10V^{-1}$ to $5V^{-1}$. In this region, the slight lower drive current for the same bias point are mostly due to differences in body factor and mobility. Similarly, the g_m/g_d plot gives a good indication at what is the analog gain of the device at a given region of operation. From Fig. 3.22, no degradation is observed for the nominal gate length ($L_{G,nom}=300nm$) between the HTO reference and the LT device.

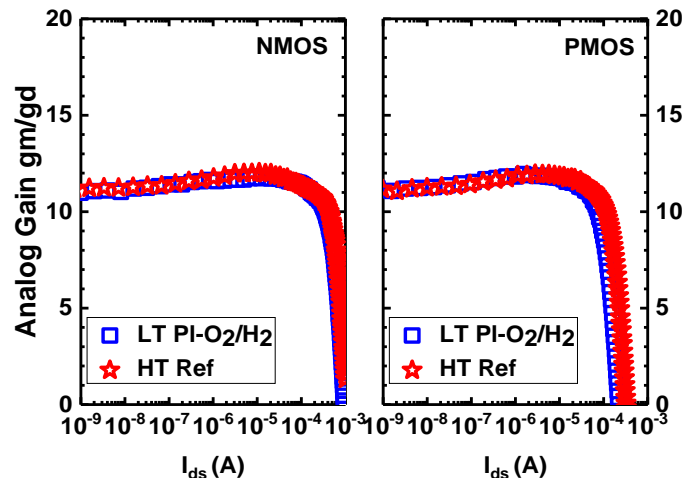


Figure 3.22 N&P FETs g_m/g_d versus L_g . The LT split $PI-O_2/H_2$ is not degraded w.r.t to the HTO reference

To the first order, the body factor of the device depend on gate oxide, film, and BOX thicknesses. That said, the architecture of the device can be further optimized to fit a specific application. For example, Fig. 3.23 depicts TCAD electrical simulations of subthreshold swing (SS) as a function of BOX thickness for both N&PFETs. The simulations were also performed for two types of ground plane (GP-N and GP-P) and silicon thickness (T_{Si}). It is observed that for lower SS values, a GP-N should be taken for PMOS, whereas a GP-P is more suitable for NMOS. If for example one takes a GP-N for both (where the NMOS presents a high $SS=85mV/dec$), the difference in SS could be attributed to the ground plane depletion (depletion of the back-side) in both cases. Therefore, the conduction does not take place in the same place which must have an impact on the slope of the device. It can also be seen that by decreasing T_{Si} , going from 30nm to 22nm, $\sim 5mV/dec$ can be improved of SS. Lastly, the BOX thickness can be further increased, knowing that this will decrease the BB efficiency (of $250mV/V$ for $BOX=25nm$).

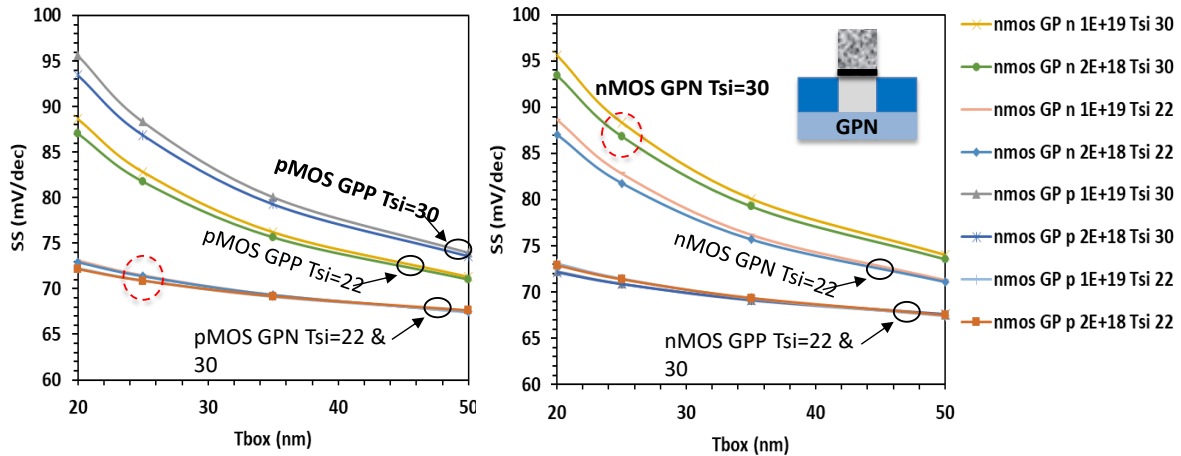


Figure 3.23 Subthreshold swing as a function of BOX thickness. The possible solutions for decreasing SS are decreasing the channel and BOX thickness and choosing the appropriate GP type for each FET.

3.8 GATE STACK OPTIMIZATIONS GUIDELINES

Based on the electrical results, some gate optimizations are suggested in this section, as to further improve the quality of the oxide.

3.8.1 Thermal Treatments

The first approach to improve the oxide quality is to optimize the thermal treatments already employed, and to applied it on the split which gave the best results (PI-O₂/H₂). On this work, no additional thermal treatment was applied on the PI-O₂/H₂. As such, despite the inefficiency of the N₂ thermal treatment when applied on the PEALD split, the later should also be tested on the best split PI-O₂/H₂. Same reasoning is valid for the plasma densification on O₂ at 450°C. Further increasing the temperature to 500°C of the plasma treatment should lead to improvements as well, but this option is not possible due to tool limitations. Concerning the split PI-O₂+PEALD, knowing that the PBTI also probes the traps in the oxide, another PEALD process should be used, at 400°C instead of 200°C.

The other approach is to employ new thermal treatments, such as a furnace annealing at 500°C on O₂, which was not tested. Even though not used on this batch, a HPD2 treatment at the end of the process should be considered. As highlighted on previous studies, this treatment should improve the interface of Si/SiO₂ [39].

Moving on other types of anneal, as it has been discussed on chapter 2, ultra-violet nanosecond laser annealing (UV-NLA) can potentially be used after the deposition of the gate oxide. The later needs to be used in the non-melt configuration, to preserve the integrity of the oxide. Additionally, it would also be interesting to address its effect by changing the number of pulses, knowing that 1000 pulses is a long process. All of the aforementioned propositions are schematically represented below on Fig. 3.24.

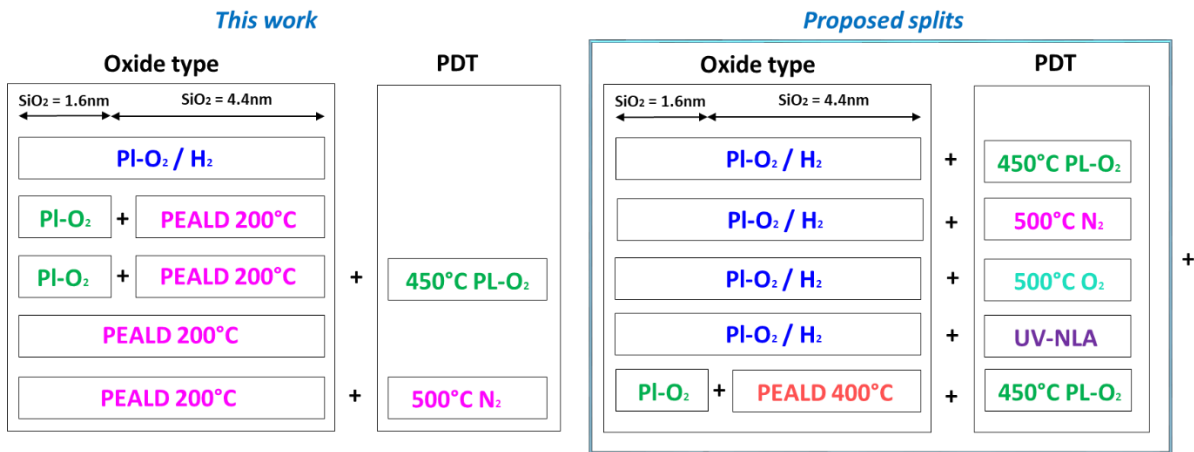


Figure 3.24 Split table comparing the oxide process used on this work as depicted on the left, and the proposed thermal treatments improvements on the right.

3.8.2 Atomic Hydrogen Treatment

It has been said that hydrogen bridges and hydroxyl E' centers could be very likely candidates responsible for the hole trapping in PMOS devices [40]. That said, in a simplified manner, the amount of oxygen vacancies and the stretching of the Si-O bonds could potentially trap or react with hydrogen, thus leading to these aforementioned defects.

In a LT process, the fact of having a reduced oxidation temperature does not allow for a complete relaxation of strain at the Si/SiO₂ interface, resulting in a larger presence of stretched Si-O bonds [41]. As reported on [42], H₀ plasmas can be very efficient to passivate defects even at T~300°C. This can be an alternative treatment to FGA (T~>400°C), interesting also if lower temperatures are needed. It can be said that H⁰ can be an interesting candidate for improving the gate oxide, but its impact on device reliability (BTI and HCI) must be further studied. Its efficiency concerning reliability is debatable, where a large concentration of H⁰ close to the Si interface may generate

additional interface defects. Additionally, for it to really be efficient, specific treatments proper for the low TB gate stack should be developed.

3.8.3 SiGe channel

As it has been highlighted, the interest of SiGe comes from the increased valence band (E_v) offset of SiGe pFETS, thus probing less traps. As such, this solution is also proposed as an optimization guideline. The SiGe channel layer can be epitaxially grown, with a thickness which can vary between 7 to 15nm. This thickness is selected based on the required channel thickness for analog and on previous results showing the improvement upon QW thickness increase. It has been said that for a very thin QW, i.e. 3nm, quantum mechanical effects increase the hole energy, and therefore artificially reduce the ΔE_v .

Concerning the choice of SiGe, it depends on the energy of defects of the stack, given the different alignments of the valence band for the different materials, as depicted on Fig. 3.25, where additional energy displacement is expected for r- and s-Ge (relaxed and strained) channels. In the example, there is a high-k material, but the idea should still hold true for the SiO_2 case.

The choice of Ge fraction ($\text{Si}_{1-x}\text{Ge}_x$) can vary from $x=0.3$ to 0.55, knowing that the valence band offset becomes larger for an increasing Ge fraction. This can be ascribed to the channel bandgap reduction, which is expected to depend on the Ge fraction in a linear relation.

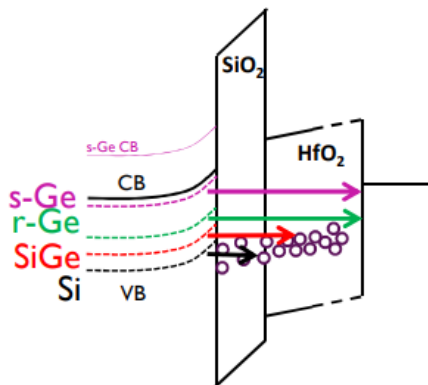


Figure 3.25 Band diagram of $\text{SiO}_2/\text{HfO}_2$ gate stacks with Si, SiGe, r-Ge, and s-Ge channel materials. Note the different alignments of the valence band (dashed lines) for the different materials [9].

Even though very good reliability devices with different Ge fraction (25% and 45%) and without a Si cap (i.e., direct oxidation of the SiGe surface), in the scope of this work, the use of a thin Si cap is a necessary integration scheme. Seeing that noise is also a device criteria, the Si Cap is crucial to obtain sufficient interface quality, especially for higher Ge fractions $> 45\%$. As such, on top of the $\text{Si}_{1-x}\text{Ge}_x$ layer, a thin undoped Si cap needs to be

grown epitaxially, in order to passivate the surface. The gate oxide is obtained afterwards with an oxidation of this cap, which can be done by knowing the ratio between Si/SiO₂ (0.45 to 0.55). This means that 1nm of SiO₂ will consume 0.45nm of Si. So in our case, for a SiO₂=6nm, the Si cap should be thicker than 3.2nm, as to have at least 0.5nm of Si left. Different Si thickness can be tested, ranging from 3.2nm to 4.7nm (leading to Si cap from 0.5 to 2nm).

The thickness of the cap has a huge impact on the NBTI performance, and it must be carefully considered. One would expect that a thicker Si cap would lead to a thicker barrier against hole tunneling, but it has been showed that a thick cap actually degraded the NBTI performance. The later has been ascribed an additional voltage drop over a thicker Si cap, which 'pushes' the channel hole energy level down and therefore more defects become energetically favorable for charging.

It must be highlighted that the Si cap thickness on NBTI degradation is still clear when correcting for the differences in E_{ox}, meaning that a comparison of degradation as a function of E_{ox} confirms that the NBTI dependence on Si cap thickness is not an artifact due to different inversion layer thicknesses (T_{INV}).

It is true that different energy distributions of the defects in the dielectric layer might exist; nevertheless, a similar beneficial effect by shifting up the Fermi level energy in the channel would hold for a very general class of defect level distributions. This means that even for a uniform energy distribution of defect levels, a fraction of defects would become unfavorable for holes at the (Si)Ge channel Fermi level.

Lastly, it is expected that the thickness of the cap also has an impact on the I_{ON}/I_{OFF}. Overall, the cap adds to the C_{ox} of the stack, so a thick one is not desired for high performance applications, whereas a very thin one is also believed to lead to reduced hole mobility values. This mobility loss was ascribed to poorer interface passivation: with a thinner Si cap more Ge from the channel segregates to the interface [30], causing a higher density of pre-existing interface states.

To summarize, NBTI should be improved by using an SiGe channel with increasing Ge fraction, increasing the QW thickness and reducing the Si cap thickness. A proposed process flow for MOS capacitors (MOSCAP) is depicted on Fig. 3.26, where the SiGe channel can be achieved by (1) a selective epitaxy @650°C or a (2) non selective epitaxy @500°C.

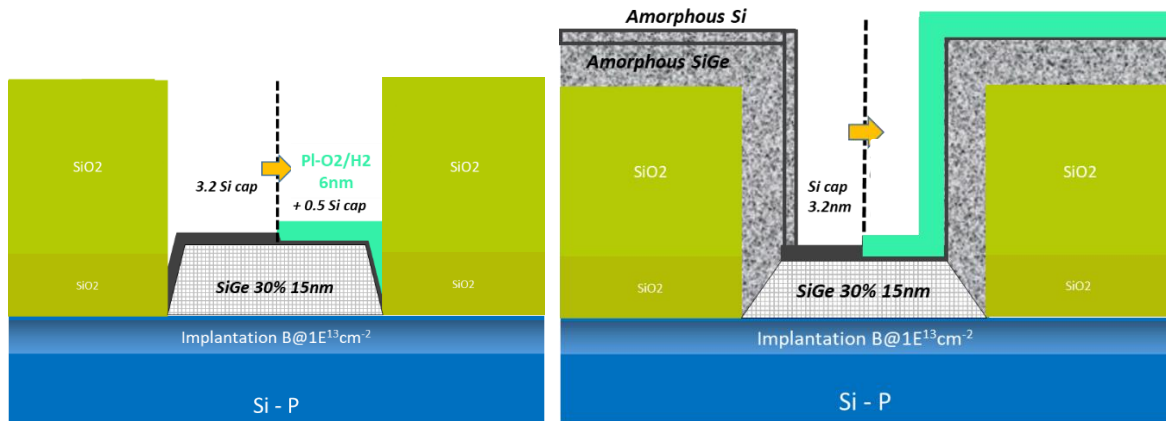


Figure 3.26 MOSCAP proposed process flow (not on scale) for SiGe channel obtained with a selective epitaxy (on the left) versus if a non-selective epitaxy is used.

In this case, the main advantage of the non-selective epitaxy is the reduced temperature. In addition, the only additional step of this process is to be sure that the CMP step also removes the additional amorphous SiGe and amorphous Si layers, as illustrated on Fig. 3.27.

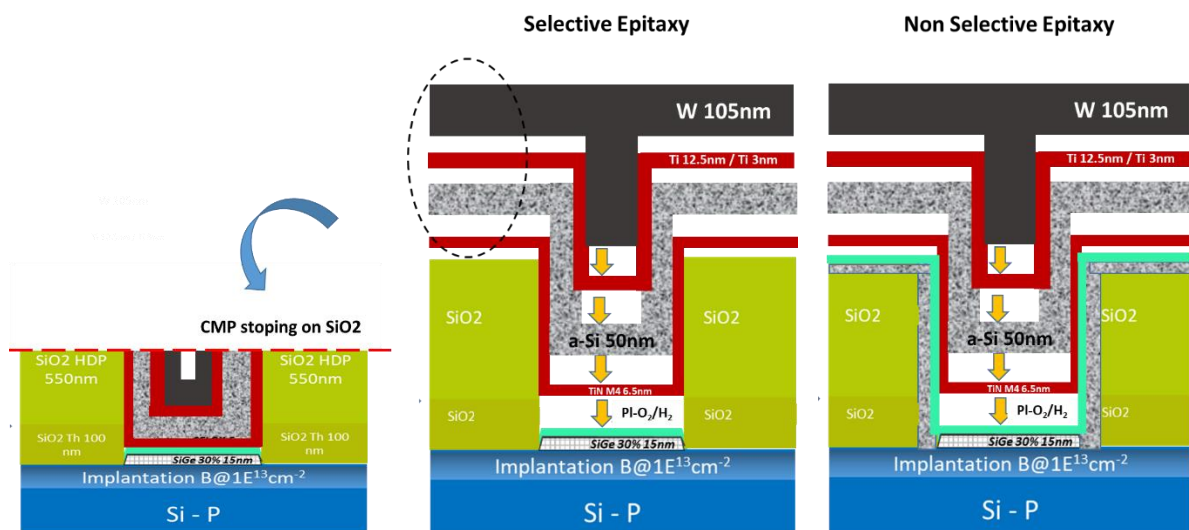


Figure 3.27 CMP step performed as a last step on a MOSCAP flow. It is performed to remove the following layers: Tungsten (W), Titanium (Ti), amorphous Silicon (a-Si) and TiN. In addition to these, in a non-selective epitaxy flow, the amorphous SiGe and amorphous Si must also be considered.

3.8.4 Junction-less transistors

The junction-less transistor (JLT) is a multigate FET with no PN nor N+N or P+P junctions, and is proposed as a solution for NBTI improvement due to its reduced oxide field (E_{ox}), as discussed on an earlier section. This is due to the fact that JL are normally

designed to operate in flat band condition at the ON-state, where the longitudinal electric field is nearly zero.

There are some obvious distinctions between JL and conventional MOSFETs. In traditional MOSFET, the inversion carriers are generated due to the gate voltage effect and confined at the thin surface channel when the device turns on. However, as for the JL transistor, the channel is often fully depleted by the high gate work function for off-state [35], [43], [44]. Then, as the gate voltage changes, the channel is turned to partially depleted and the carriers are flowing through the entire bulk of channel. The device is basically a resistor in which the mobile carrier density can be modulated by the gate. That said, there are two main criteria for successfully fabricating a JL, the first one being that the cross section of the device needs to be small enough in order to be able to turn the device off. Because of that, multi-gate architectures such as double gate (DG), tri-gate (TG) and gate-all-around nanowire (GAA-NW) configurations are often employed. In these architectures, since the channel region is under the influence of the gate electrode from more than one side, the gate-induced depletion is efficient to obtain volume depletion. The second requirement is that the semiconductor needs to be heavily doped to allow for a suitable current drive when the device is turned on.

In the scope of this thesis, the goal of this section is to propose a JL architecture on SOI structures that fits the requirements of reliability, low-leakage current and low cost. TCAD electrical simulation of p-channel silicon JL have been demonstrated, using both polysilicon and TiN gate materials. Only simulations for a p-channel shall be demonstrated, but the same reasoning can be applied for an n-channel as well.

The specifications of the technology are the device's nominal dimensions ($W=240\text{nm}$, $L=300\text{nm}$), the thick channel to avoid S/D epitaxy ($T_{\text{Si}}=22\text{nm}$), $EOT=6\text{nm}$ and $V_{\text{DD}}=2.5\text{V}$, where to ensure proper operation, the channel must be depleted at $V_{\text{G}}=0\text{V}$. As discussed in chapter 1, on this work a $T_{\text{Si}}=30\text{nm}$ was used, but it can be reduced to $T_{\text{Si}}=22\text{nm}$ for better SCE control and to ease the depletion required for JL.

The challenge of JL in a planar architecture is that the devices do not benefit from lateral depletion; the additional challenge of JL for analog coming from its bigger dimensions, unlike the case for digital devices (with reduced dimensions), as seen on Fig. 3.28 comparing JL transistor for both digital and analog.

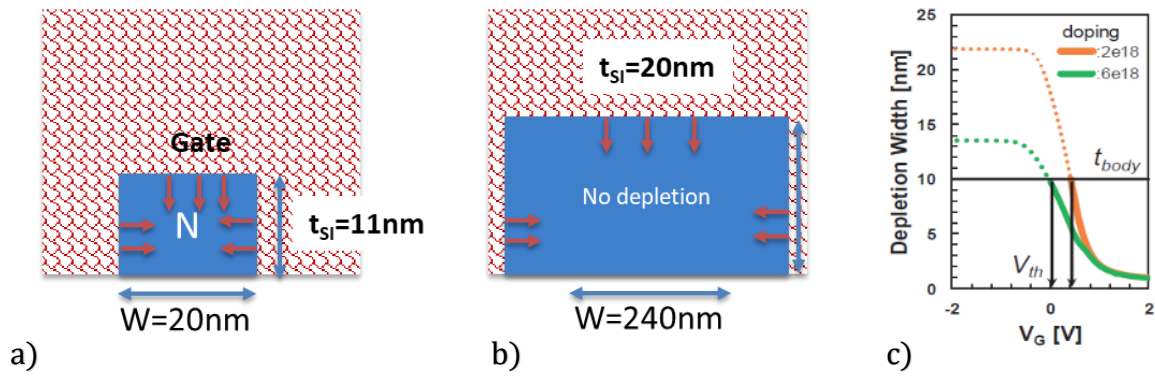


Figure 3.28 Schematic representation of the JL for digital applications (a) versus for analog (b), showing that depletion is harder to achieve on planar devices with bigger dimensions. It is shown on (c) [6] that to achieve depletion on a 22nm channel at $V_G=0\text{V}$, a channel doping of $2 \times 10^{18}\text{at/cm}^3$ is required.

It can be seen from the schematic that for this channel thickness, it is hard to achieve volume depletion, and that a $T_{\text{Si}}=22\text{nm}$ requires a low doping of $2 \times 10^{18}\text{at/cm}^3$, compromising device depletion. The proposed architecture is to sandwich two layers of N&P, thus making it that the depletion in the channel is achieved from both top (channel) and bottom, where the thickness and doping of these layers shall be defined by TCAD. For the sake of simplicity, a first schematic is proposed with a TiN gate material only. The thickness of both layers are set to 11nm, and the doping of the top layer (P-top) is set to $1 \times 10^{19}\text{at/cm}^3$. As such, the initial goal is to define the doping of the bottom layer (N-bot) for an $I_{\text{OFF}}=1 \times 10^{-10}\text{A}/\mu\text{m}$, where the $I_d(V_g)$ curves are depicted on Fig. 3.29. The simulated structure is shown as an inset on the graph.

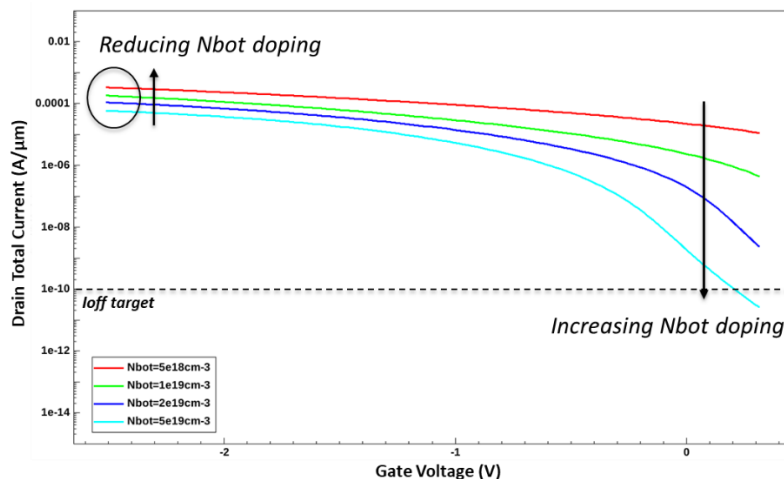


Figure 3.29 $I_d(V_g)$ curves for different doping of the N-type bottom layer for a TiN stack.

It can be seen that there is a compromise between doping and device performance: the higher the dopant concentration in the bottom layer, the lower is the leakage AND the I_{ON} current. In this case, no split fulfilled the I_{OFF} goal at $V_G=0V$.

Fig. 3.30 now compares the $I_D V_G$ curves for two gate stack materials, TiN and a n-type Polysilicon. As expected, the depletion of the top P-type layer is obtained more easily (for $N_{bot} > 2e19at/cm^{-3}$), and up to 2 decades gain on the I_{OFF} is observed for a negligible I_{ON} difference. Once the I_{OFF} was reached, the goal now is to further improve the I_{ON} .

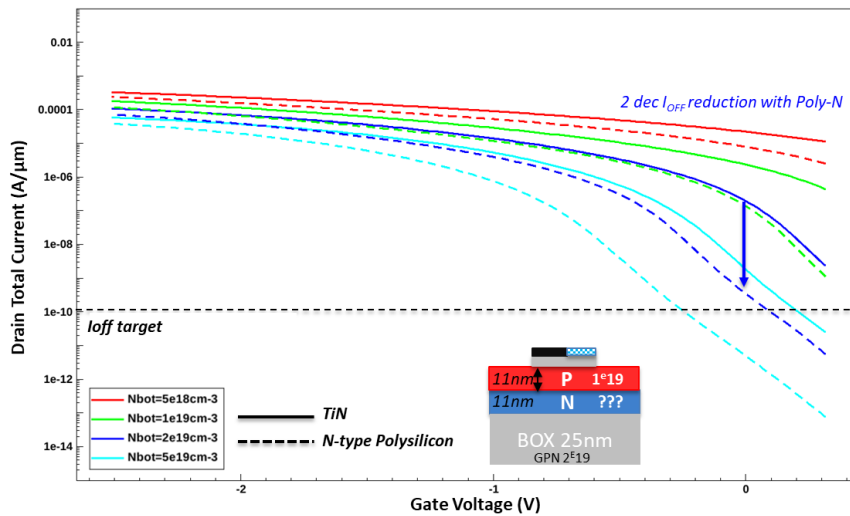


Figure 3.30 $I_d(V_g)$ curves for different doping of the N-type bottom layer for a TiN and N-type Polysilicon stack.

A next axis of improvement can be achieved by also changing the properties of the top layer (P_{top}) for a given N_{bot} ($2e19at/cm^{-3}$), as highlighted on Fig. 3.31. In this case, the doping of the bottom layer (N_{bot}) should be decreased to compensate the I_{ON} .

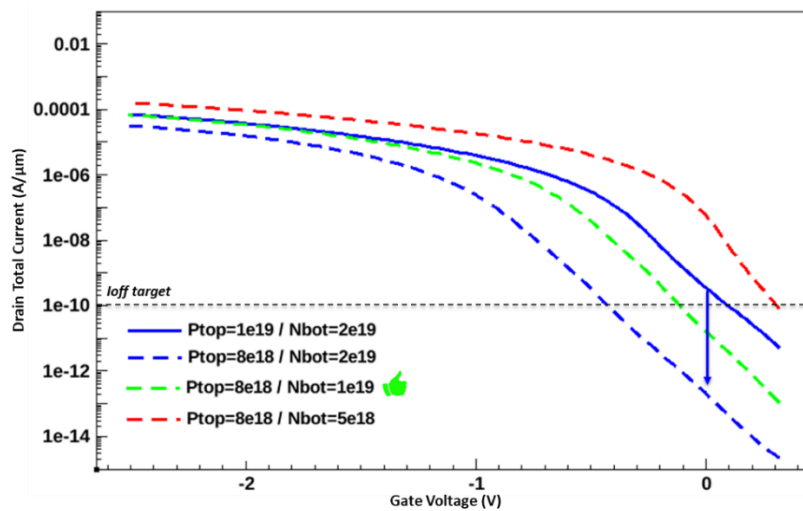


Figure 3.31 $I_d(V_g)$ curves for different doping levels of the P-type top layer.

Lastly, Fig. 3.32 depicts the impact of different top layer (P_{top}) thicknesses, varied between 10 and 12nm for a fixed $P_{top}=8e18at/cm^{-3}$ and $N_{bot}=1e19at/cm^{-3}$. Once more, if a thick layer of 12nm is chosen, thus yielding more current, a higher doping of N_{bot} is required. Contrarily for the case of a thinner layer of 10nm, the doping should be decreased. The latter option is preferred (Fig. 3.32 b), seeing that it improves the SS and lower doping levels should be beneficial for mismatch.

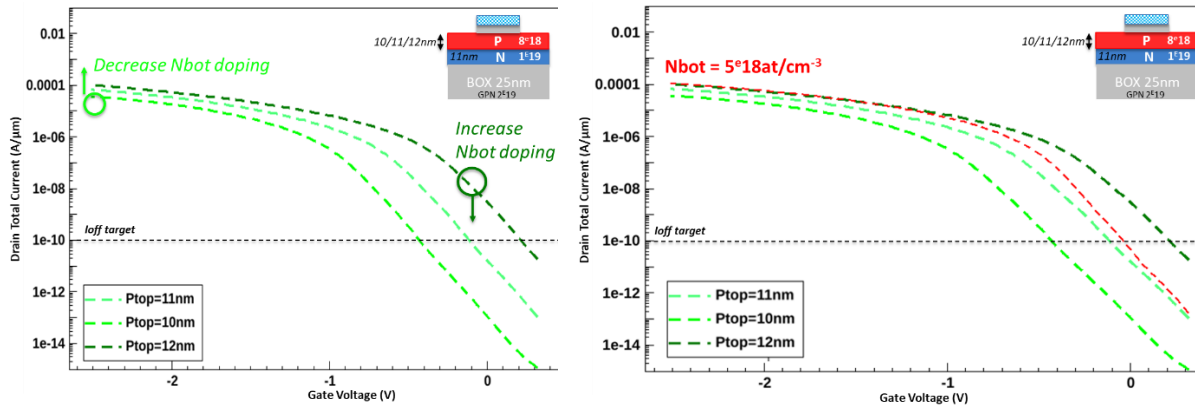


Figure 3.32 $I_d(V_g)$ (a) curves for different thickness of the P-type top layer and (b) impact of decreasing the N_{bot} doping, as depicted in the red curve.

In summary, the proposed architecture presents a $P_{top}=10nm$ and $8e18at/cm^{-3}$ with a N_{bot} of 11nm and $5e18at/cm^{-3}$. Lastly, Fig. 3.33 summarizes the ways to achieve the desired I_{ON}/I_{OFF} trade-off, which can be easily adapted according to the application.

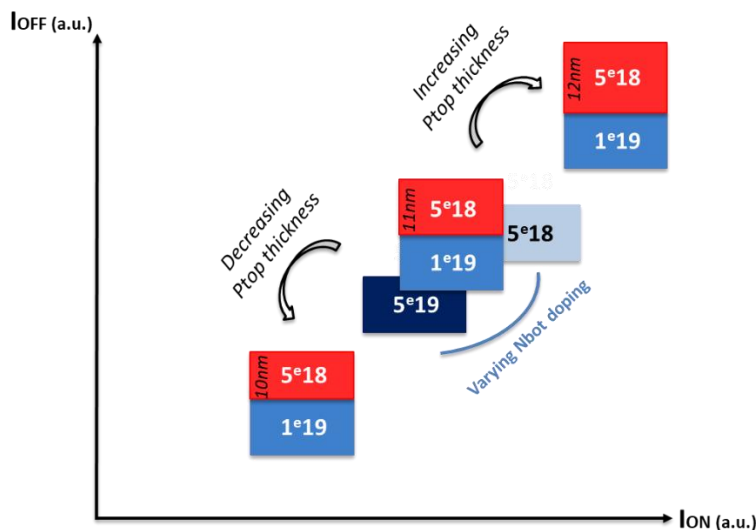


Fig. 3.33 Schematic recap of a few possible I_{ON}/I_{OFF} trade-offs when engineering a JL transistor.

3.8.4.1 Gate Stack Poly

By integrating a SiO_2 /Polysilicon only gate stack, the two main benefits come from the possible noise reduction due to the TiN suppression and the ease to implement JL

transistors. Additionally, the poly can be deposited easily, is thermally stable and does not react with silicon oxide. In comparison to digital, the TiN suppression is not so simple: if the high-k is put in direct contact with the poly, there is a phenomenon called fermi level pinning, meaning that it is not entirely the poly doping that controls the threshold voltage. This causes the modification of the poly Fermi level due to the formation of dipoles at the poly/high-k interface. This is not the case if the poly is in contact with the SiO_2 , which is the case of the proposed integration. However, when there is only a poly, one needs to attest to the possible channel counter-doping or poly depletion (the reader may refer to chapter 2 for more details on both). Summarizing, if the dopants are too energetic, upon the annealing steps, the dopants can diffuse through the gate oxide and reach the silicon channel. The latter is especially critical in the case of Boron. Of course, this is more problematic in scaled EOT, whereas a 6nm SiO_2 composes a thicker barrier for this diffusion. Still, if this happens, one may question if diffused dopants are going to affect the quality of the oxide and potentially degrade BTI. On the other hand, it is common to have poly depletion, which is a zone depleted of dopant atoms on the interface Poly/ SiO_2 , which introduces a capacitance in series and degrades the equivalent thickness. Both phenomena lead to shifts in the threshold voltage in opposite manners and are schematized on Fig. 3.34.

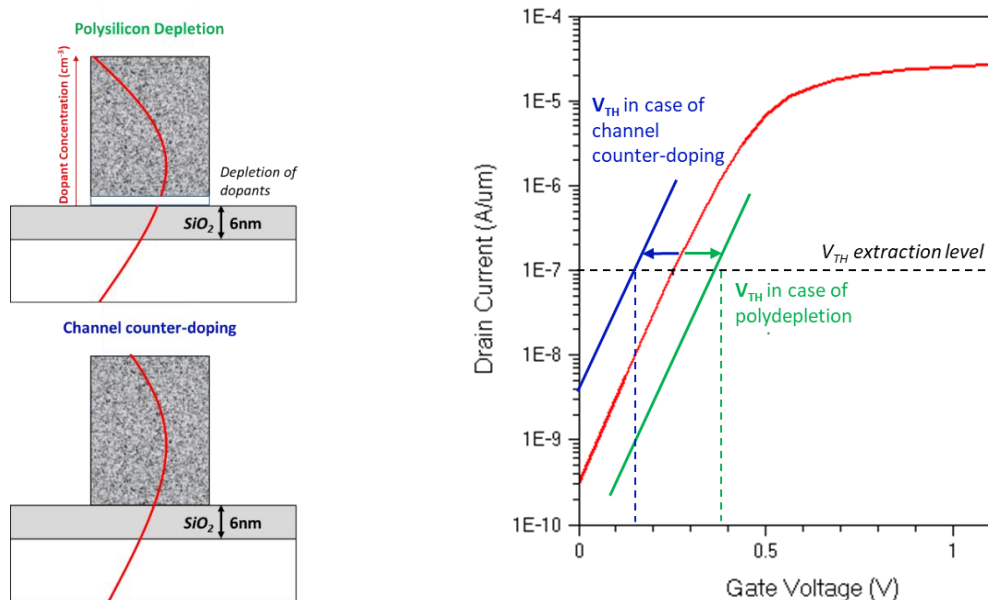


Figure 3.34 Schematic representation showing both polydepletion and channel counterdoping on the left, and its respective impact on the threshold voltage of the device on the right. Attention, this schematic sole purpose is to account to V_{TH} variation. Even though not represented here, the subthreshold slope should also be impacted.

One challenge of LT is that the diffusion is limited (even though the diffusivity coefficient is higher in polysilicon than in crystalline silicon), so being sure that the dopants will reach the interface with SiO₂ in a 500°C range is challenging. Additionally, the diffusion in poly could be worst for mismatch, due to the randomness of the polysilicon grains, thus leading to distinct peaks of dopants in different areas of the oxide.

As an alternative, thanks to its low penetration depth, a Ultra-Violet Nanosecond Laser Annealing (UV-NLA wavelength: 308nm, optimized pulse duration 160ns) is proposed for crystallizing the a-Si layer while preserving the integrity of the bottom tier (as discussed on chapter 2).

In a standard process flow, the poly is doped at the same time as the S/D implants. When using a UV-NLA, the amorphous silicon (a-Si) layer is deposited, doped in-situ, then subsequently annealed to promote the formation of a polycrystalline active layer. Despite the fact that this approach has already been demonstrated for 3DSI applications, at the moment, there is only the possibility to achieve a type-N polysilicon. It has been said however that good JL performance are actually achieved with opposite types of polysilicon (poly type-N for PMOS and poly type-P for NMOS). For this reason, the doping in-situ is then replaced by a dopant implantation step (using ion implantation). As such, the goal of this full sheet study is analyze the feasibility of using a UV-NLA and its impact on the dopant diffusion. For the sake of simplicity, bulk Si wafer was used, but for a real integration on SOI wafers, the laser energy will need to be properly adjusted (knowing that the heat diffusion depends on the underlying layers). The polysilicon thickness was chosen as 70nm because the laser step increases the rugosity of the film, so a CMP is needed afterwards. As a result, 70nm is usually deposited and then a CMP is performed to reach the desired 50nm. Even though the CMP step was not performed on this full sheet study, on a real integration flow, the later needs to be performed. The discussed steps are schematized on Fig. 3.35.

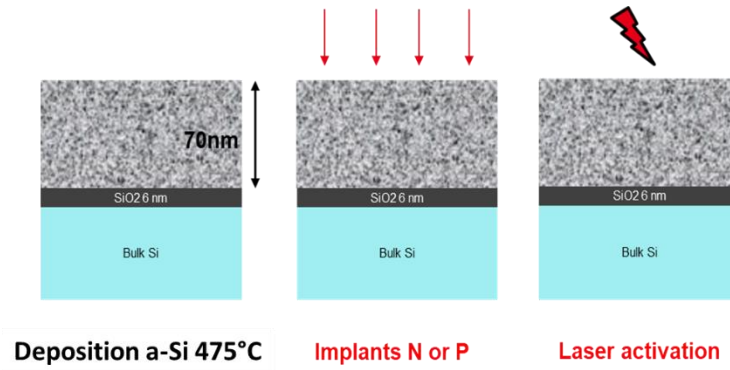


Figure 3.35 Representation of the process flow used on this full sheet study, starting with the deposition of amorphous silicon (a-Si), followed by the ion implantation and laser anneal / polysilicon recrystallization.

The oxide used was the best-case PI-O₂/H₂, to be closer to the full device integration. Also, it is more realistic to use a LT oxide, because the density of the oxide will also change the diffusivity in it. Different implantation conditions were chosen, for phosphorus and boron. Various doses and various energies were tested to see if it is possible to set the work-function.

The implant profiles were defined on crystal trim (CTRIM) 1D simulation for a 70nm a-Si. After deposition, the polysilicon was measured by ellipsometry and there was 78nm instead, so the profiles are also shown for both thickness of 70nm and 78nm. The expected concentration at the SiO₂ interface is given by integrating the dopant profile, knowing that it should be roughly between 1e20cm⁻³ or 3e20nm⁻³ for a good WF definition. For the type-P doping, BF₂ is used instead of Boron for being known to diffuse less, whereas Phosphorous was chosen for the type-N poly. The profiles are shown on Fig. 3.36.

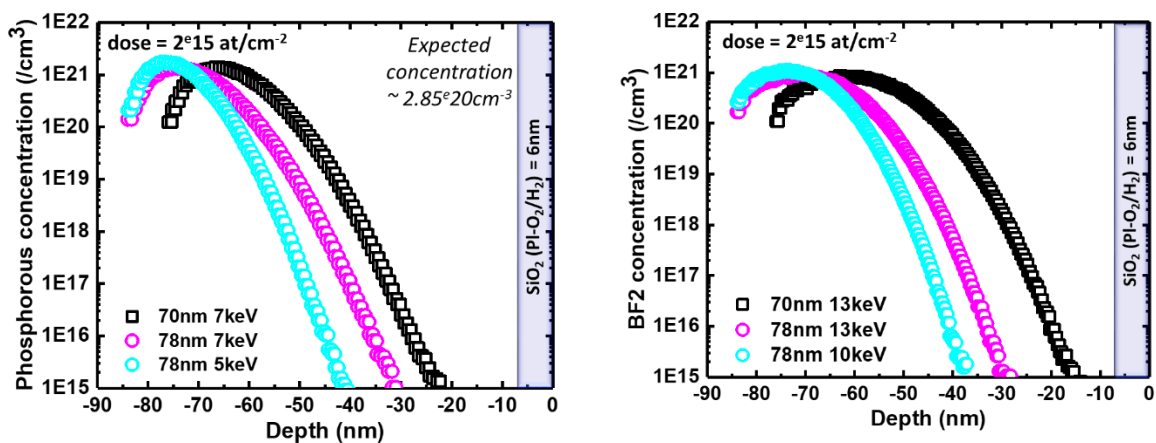


Figure 3.36 dopant concentration as a function of depth shown for Phosphorous (on the left) and BF₂ (on the right). The expected concentration of $2.85e20\text{cm}^{-3}$ [45].

The laser conditions are the following: each wafer site corresponds to 1 pulse, ranging from 0.775 to 2.200 J/cm² with 75 mJ/cm² steps. Before the laser, the first step is to perform a laser crash test, to know when the melt threshold is reached. Based on in-situ Time Resolved Reflectometry (TRR) profiles (Fig. 3.37), it is possible to detect when the film melt threshold is reached (Reflectivity decreases). As such, three regimes are identified: the amorphous Si explosive melt threshold @~0.9 J/cm², the second melt @ 1.2 J/cm² and a full a-Si layer melt reached between 2.0 & 2.1 J/cm². As such, one can draw the range of interest = 0.8 – 2.2 J/cm².

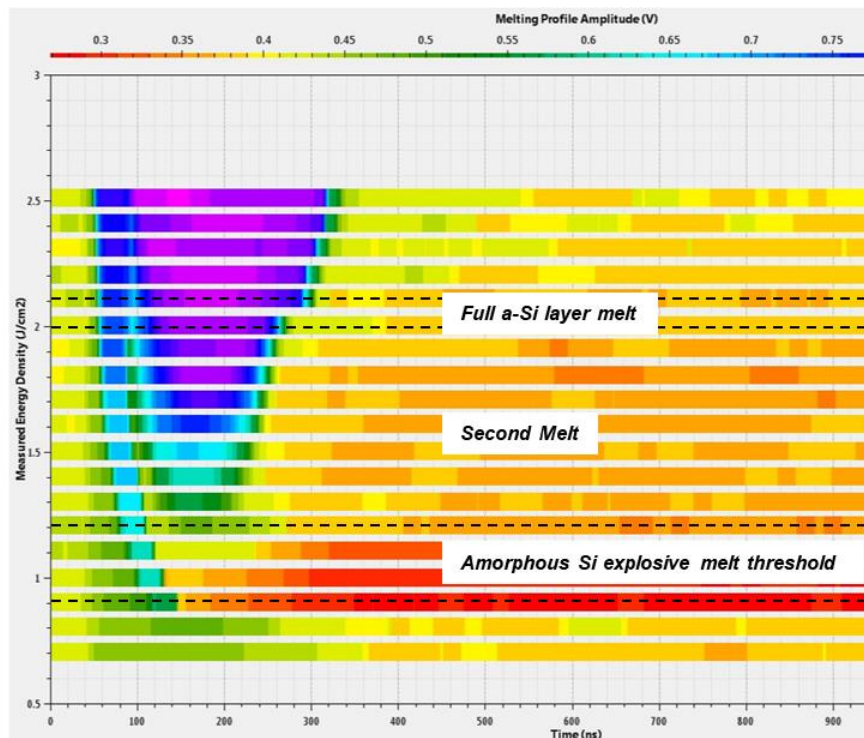


Figure 3.37 Time Resolved reflectometry: energy density screening. Three regimes are observed: explosive melt/spontaneous recrystallization, partial melt and total melt.

From the 4-point probe resistivity measurements (R_{carre}), one can have a good indication of the poly doping, knowing that the resistivity is inversely proportional to the doping and of the laser regimes. Fig 3.38 depicts the sheet resistance values for Phosphorous and BF₂ for the different implant conditions, knowing that the larger the grain size the lower the resistance. It can be seen that in the sub-melt regime, the R_s values are very high, because the poly grains are still not formed and the dopants have not yet been activated. Upon melt and recrystallization, R_s drops especially once larger grains are formed. R_s optimum when almost the whole a-Si layer melts (likely maximum grain size / maximum dopant activation). Beyond full melt, R_s increases because finer grains are

obtained. Additionally, it can be seen that there is not so much discrepancy between the resistance values for the different implant conditions.

Overall, the resistivity values obtained for BF2 are roughly in-line with [46], but it must be highlighted that on the reference paper, a classical anneal is performed instead of a laser anneal. As for the Phosphorus conditions, by comparing with [47], the data is slightly above the reference curves. The later could be explained by the fact that part of the implanted dose could be lost in the SiO₂ layer or the c-Si substrate, or lost (out-diffused), or that the Si grains in this results could be smaller than those obtained in the reference.

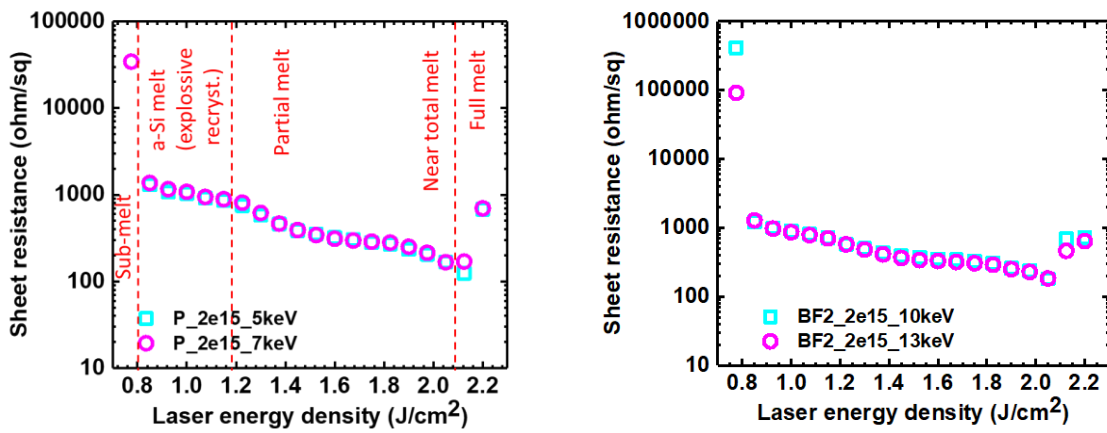


Figure 3.38 Evolution of sheet resistance as a function of the pulse laser energy density.

Seeing that the resistance depends on the polysilicon grain size, a Scanning Electron Microscopy (SEM) top view was performed (Fig. 3.39). At first order, the grain size is driven by laser energy density, and these results are well in agreement with the Rs values.

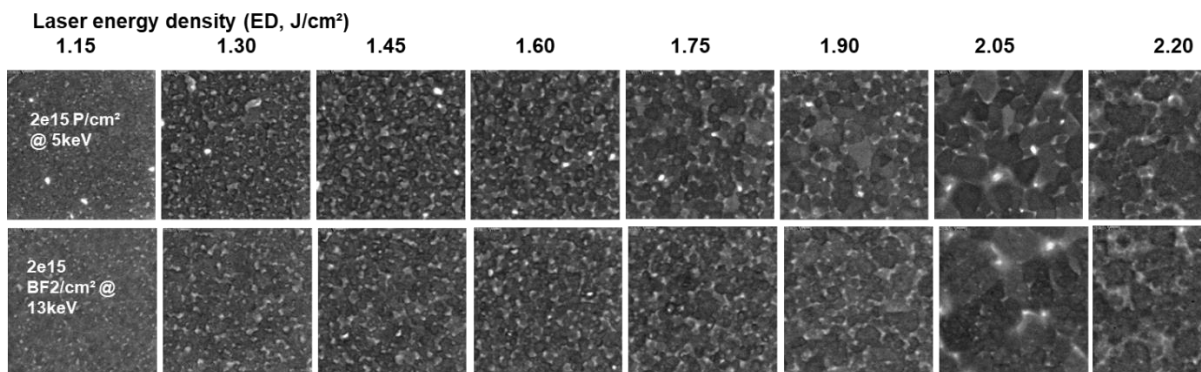


Figure 3.39 Grain size increase up to near-total melt condition, and the Larger the grains, the lower the Rs. Near-total melt (2.05) : largest grains / highest roughness, Full-melt (2.20) : grain size drop

Based on these results, it is now crucial to obtain the dopant profile upon the laser anneal. A few samples with different laser energy density were selected, based on Rs, as to not have full melt otherwise melting everything including the SiO₂ nor a value too close to full melt, given that those points are often unstable. In the case of Phosphorous, 2e15 P @ 5keV 2.05 J/cm² and 2e15 P @ 7keV 2.05 J/cm² where chosen, whereas for BF2 2e15 P @ 13keV 1.975 J/cm² was used. The question now is about dopant diffusion, so Time-of-Flight Secondary Ion Mass Spectrometry (ToF-SIMS) were performed on these samples. The quantification of boron and phosphorus was carried out by comparison with a silicon reference implanted with the two elements. Fig. 3.40 shows the Phosphorous concentration profile as a function of depth for both samples. It is highlighted that the laser successfully recrystallizes the a-Si, leading to an uniform dopant profile throughout the full layer. Additionally, the P concentration of 3e20 at/cm³ are close to the calculated value of 2.85 at/cm³.

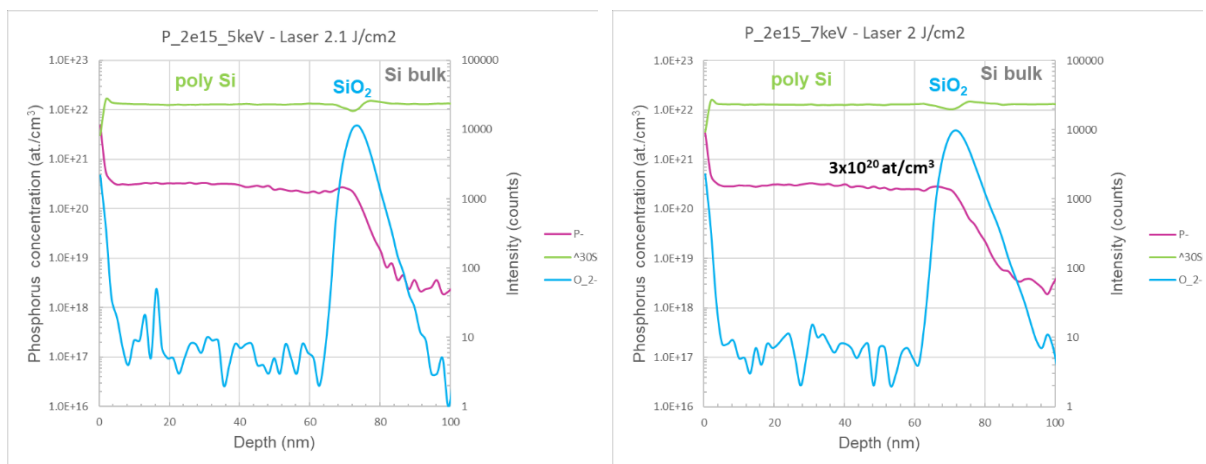


Figure 3.40 SIMS performed on Phosphorous samples. It is highlighted that in both cases, the concentration is roughly the expected value (2.85 cm^{-3}) and that it exists a flat profile.

It is also observed on Fig. 3.41 that there is no impact of energy split on the P profile. This means that when defining the implant conditions, if preferred, even a superficial doping of very low energy should be efficient to properly dop the poly, and mostly the dose should be carefully chosen.

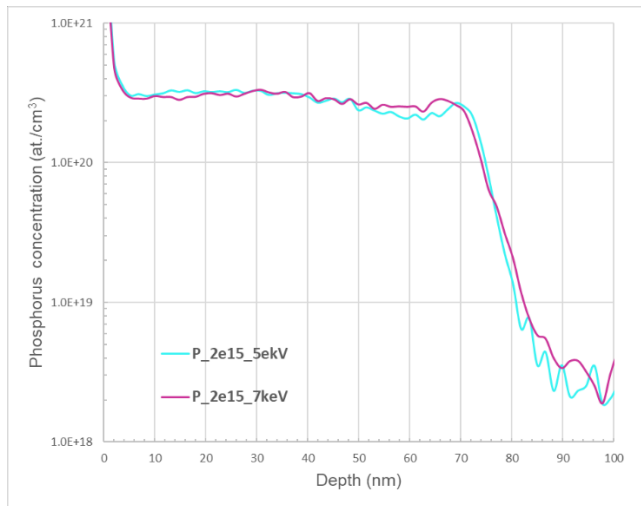


Figure 3.41 SIMS performed on Phosphorous samples for two different energy splits. No difference in dopant concentration is seen despite the energy of the implanted species.

The SIMS BF₂ profiles are now depicted on Fig. 3.42. Likewise for the P case, the SiB profile is also flat and concentrations $\sim 1e20cm^{-3}$. This lower concentration w.r.t. the P case could be explained by the different behavior of BF₂ and P; P is known to segregate at the interface (native SiO₂), whereas an exo-diffusion could have taken place in the case of B. Another interesting feature is that the SiF profile is not flat, but it is hard to conclude on the effect of laser on Fluor seeing that the SIMS were not properly calibrated for this case.

Lastly, it is observed that in both cases a high concentration is seen on the SiO₂ and the Si bulk, leading one to believe that there is a channel counter-doping. However, SIMS at the interface with SiO₂ are tricky to be properly calibrated, so the values in the oxide are indicative data, meaning that the most reliable conclusion is on the comparison between samples.

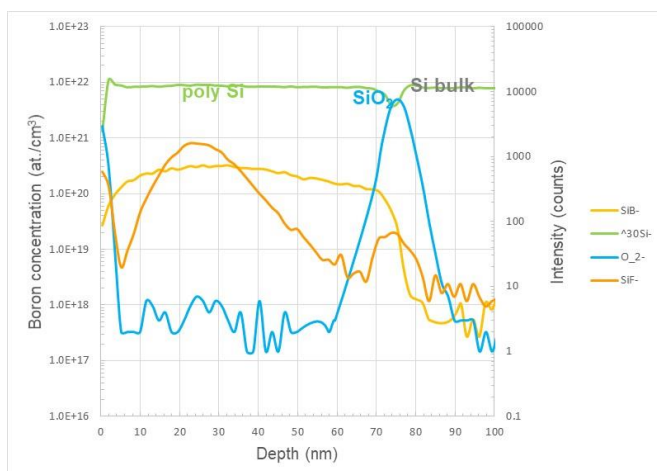


Figure 3.42 SIMS performed on BF₂ samples.

3.9 HOT CARRIER INJECTION

In addition to BTI, Hot Carrier Injection (HCI) degradation can also be evaluated (Fig. 3.43). Whilst BTI can be strictly related to both oxide and Si/SiO₂ interface quality, HCI is both a combination of the aforementioned parameters plus another component, the junction profile. This latter statement can be understood as followed: when the transistor is in saturation, the electric field near the drain is very high because of the strong potential drop between the pinch-off point in the channel and the drain. The carriers arriving in this pinch-off zone are strongly accelerated under the effect of the high lateral electric field, thus becoming very energetic. These hot carriers can then degrade the gate oxide of the channel/drain region either by direct injection into the oxide or indirectly by generating new electron/hole pairs which, in turn, can lead to the creation of interface defects.

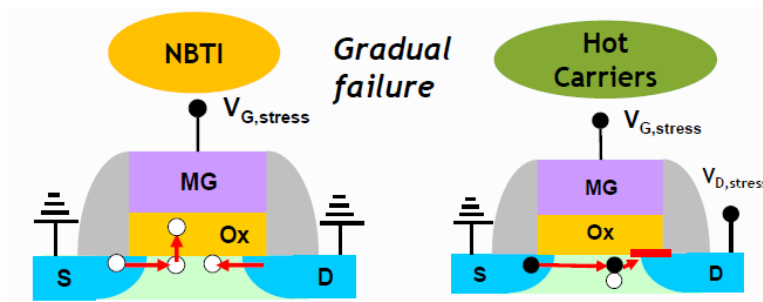


Figure 3.43 Schematics highlight the differences between NBTI and HCI degradation. For NBTI tests, no V_D is applied, whereas HCI requires current flowing through the transistor [48].

This section shall focus only on the results for NMOS devices, given that their hot carrier degradation can be more critical than for PMOS. This is because in PMOS, the channel current is carried by holes, which have a lower mobility than of the electrons. As a consequence, the holes cannot attain such a high energy and thus have a lower probability of being injected into the gate oxide.

3.9.1 HCI Results

Figure 3.44 shows the TTF extracted for different stress voltages $V_{STRESS}=3.2, 3.4$ and $3.6V$ performed under the condition $V_G=V_D/2$ at room temperature ($T=30^\circ C$) for the different oxide splits previously mentioned. The TTF is shown for the saturated current ($I_{D,sat}$), linear current ($I_{D,lin}$) and transconductance (gm), plotted as a function of V_D .

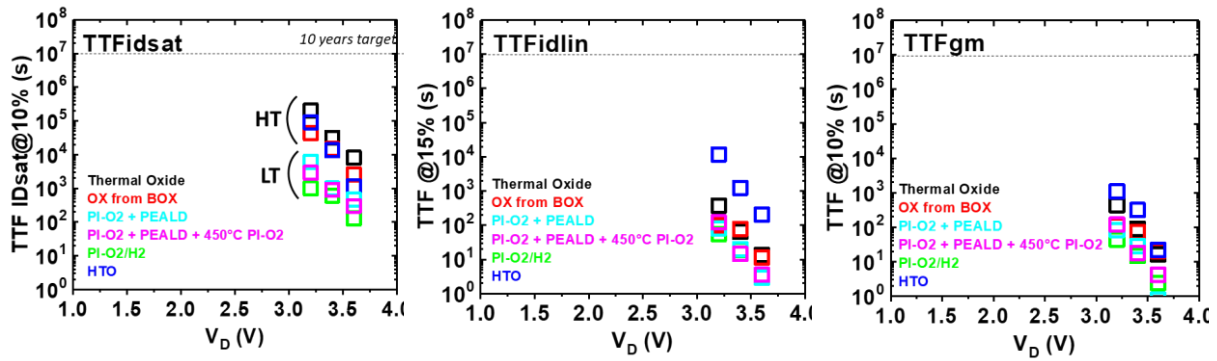


Figure 3.44 TTF for the $I_{D,sat}$, $I_{D,lin}$, and gm. TTFs are given for a 10% drift for $I_{D,sat}$ and gm, and 15% for $I_{D,lin}$ for the different oxide splits. Each square corresponds to the mean of the tested dies.

Concerning the $I_{D,sat}$ degradation, it can be highlighted that there are 2 populations, one related to the LT splits, and another for the HT, with the LT group showing inferior results w.r.t. the HTO. These results are somehow surprising, seeing that the OX from BOX split had the worst gate oxide parameters (N_t , BTI, μ_0). Likewise, the PI-O₂/H₂ split also depicts inferior hot carrier degradation. It must be mentioned that on previous sections, the superior gate stack results of the split PI-O₂/H₂ had been attributed to a better passivation of the dangling bonds by H⁰. As such, it is possible that during the hot carrier degradation, there was a breaking of the Si-H bonds, which could explain these results. This is because HCI degradation is generally ascribed to Si-H interface bond breaking in the vicinity of the drain. Even though not done on this work, one way to address this hypothesis is to perform noise measurements on the same dies tested on HCI.

It must be mentioned that $I_{D,sat}$ has been shown here, but this FOM is more interesting for digital applications. For certain analog operation conditions, it would be more relevant to look at the degradation of $I_{D,lin}$, and gm. Contrarily to $I_{D,sat}$, the split HTO used as a reference is really superior to the other ones for the $I_{D,lin}$ degradation. It can then be said that $I_{D,lin}$ is much more related to the junction quality. Lastly, gm can give an indication about the mobility of the carriers, but it is hard to conclude. The derivative was too noisy, thus requiring more dies for a reliable extraction.

As a general conclusion, it can be said that even though reliable NMOS has been already achieved, further interface engineering should be pursued as to meet hot carrier specifications. Additionally, as shall be discussed on chapter 4, this lot has a smooth junction, which should be beneficial for HCI. To this moment, the efficiency of this implantation tilt is unknown since other junction profiles were not tested.

3.10 CONCLUSIONS OF THE CHAPTER

In this chapter, the experimental procedures set up to study the properties of the proposed gate stack for a high voltage ($V_{DD}=2.5V$) analog device has been discussed, where reliable and low noise stacks have been envisaged. The gate stack, composed of a thick $SiO_2=6nm$ has been obtained through different LT oxide splits (in the $500^\circ C$ range). Thanks to the thicker gate oxide, new fabrication methods has been used, not possible before for aggressively scaled gate stacks. Additionally, post deposition treatments have also been applied, in both plasma and furnace anneal. An innovative method has also been proposed, called OX from BOX, which takes advantage of the high-quality oxide of the buried oxide.

Thanks to in-depth characterization, the gate stack features (low field mobility, interface states, noise and reliability) have been extensively studied. For the NMOS devices, there was a visible effect of the different oxide splits, with the hydrogenated plasma split $Pl-O_2/H_2$ showing the best results. Additionally, reliable NMOS devices were demonstrated, ideal for readout circuits with NFETs only. As for the PMOS devices, a less noticeable effect was observed for the oxide splits, and the post deposition treatments have not effectively improved its oxide features. All the splits were degraded w.r.t. the HTO used as a reference. Lastly, an asymmetry between N&P has been observed.

An identification of the defects present in LT SiO_2 oxides has been provided, where 3 types of traps have been identified, namely deep, border and interface defects. These traps, presenting different spatial location and energy levels, does not impact the gate oxide features the same way, thus explaining the N&PMOS results. For NMOS, noise is mainly sensitive by interface states, whereas PBTI is driven by e- trapping in high energetic defects located above the conduction band. Concerning PMOS, other than the interface states, the border traps also contribute to the trapping mechanisms, but the method to extract D_{it} from the SS not being able to separate the two contributions. Additionally, despite the high-k suppression of the studied stack, NBTI remains problematic, which could be an indication that the traps are actually mainly located in the SiO_2 .

The other high quality oxide solution, the OX from BOX, has also been degraded in comparison to the high temperature process. It is thus concluded that the chemical treatment to thin the oxide leads to detrimental results. Additionally, the post deposition plasma treatment does not seem enough to recover the oxide integrity. FTIR

measurements have been performed as to understand this aforementioned degradation. Overall, even with the physical-chemical results, it is hard to address what caused the oxide degradation observed on the electrical results. Only a full integration on an electrical lot can say if the OX from BOX is truly a viable solution.

The impact of the best LT oxide PI-O₂/H₂ on some analog FOM has also been shown. Despite the degradation of the LT split, the latter presents good analog performance, as seen on gm/Id and gm/gd.

Finally, guidelines are also proposed to improve NBTI reliability, currently the main challenge in the context of Sequential 3D integration. Other thermal treatments such as laser annealing and atomic hydrogen, SiGe channel QW and Junction-less transistors being viable options for NBTI improvement.

Overall, this chapter enable to conclude on the feasibility of stacking high voltage analog devices in a 3DSI for More than Moore applications.

- [1] C.-M. Lu *et al.*, « Key process steps for high performance and reliable 3D Sequential Integration », in *VLSI Technology, 2017 Symposium on*, 2017, p. T226–T227.
- [2] J. C. Lee *et al.*, « High-k dielectrics and MOSFET characteristics », in *IEEE International Electron Devices Meeting 2003*, 2003, p. 4–4.
- [3] A. Veloso *et al.*, « Thermal and plasma treatments for improved (sub-)1 nm equivalent oxide thickness planar and FinFET-based replacement metal gate high- k last devices and enabling a simplified scalable CMOS integration scheme », *Jpn. J. Appl. Phys.*, vol. 53, n° 4S, p. 04EA04, janv. 2014, doi: 10.7567/JJAP.53.04EA04.
- [4] « Recent advances in low temperature process In view of 3D VLSI integration ».
- [5] X. Garros *et al.*, « RF Performance of a Fully Integrated 3D Sequential Technology », in *2019 IEEE International Electron Devices Meeting (IEDM)*, 2019, p. 25–1.
- [6] J. Franco *et al.*, « BTI reliability improvement strategies in low thermal budget gate stacks for 3D sequential integration », in *2018 IEEE International Electron Devices Meeting (IEDM)*, 2018, p. 34–2.
- [7] J. Franco *et al.*, « NBTI in replacement metal gate SiGe core FinFETs: Impact of Ge concentration, fin width, fin rotation and interface passivation by high pressure anneals », in *2016 IEEE International Reliability Physics Symposium (IRPS)*, 2016, p. 4B–2.
- [8] G. Jiao *et al.*, « Acceptor-like trap effect on negative-bias temperature instability (NBTI) of SiGe pMOSFETs on SRB », in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, p. 31–2.
- [9] J. Franco *et al.*, « NBTI Reliability of SiGe and Ge Channel pMOSFETs With $\text{SiO}_2/\text{HfO}_2$ Dielectric Stack », *IEEE Trans. Device Mater. Reliab.*, vol. 13, n° 4, p. 497-506, déc. 2013, doi: 10.1109/TDMR.2013.2281731.
- [10] J. Franco, B. Kaczer, M. Cho, G. Eneman, G. Groeseneken, et T. Grasser, « Improvements of NBTI reliability in SiGe p-FETs », in *2010 IEEE International Reliability Physics Symposium*, 2010, p. 1082–1085.
- [11] J. Franco *et al.*, « SiGe Channel Technology: Superior Reliability Toward Ultrathin EOT Devices—Part I: NBTI », *IEEE Trans. Electron Devices*, vol. 60, n° 1, p. 396-404, janv. 2013, doi: 10.1109/TED.2012.2225625.
- [12] T. Bourdi et I. Kale, *LOW-FREQUENCY NOISE IN ADVANCED MOS DEVICES*. Dordrecht: Springer, 2007.
- [13] G. Ghibaudo et T. Boutchacha, « Electrical noise and RTS fluctuations in advanced CMOS devices », *Microelectron. Reliab.*, vol. 42, p. 573–582, déc. 2001.
- [14] M. A. Alam et S. Mahapatra, « A comprehensive model of PMOS NBTI degradation », *Microelectron. Reliab.*, vol. 45, n° 1, p. 71-81, janv. 2005, doi: 10.1016/j.microrel.2004.03.019.
- [15] R. Doering et Y. Nishi, *Handbook of semiconductor manufacturing technology*. CRC Press, 2007.
- [16] C. R. Helms et B. E. Deal, Éd., *The Physics and Chemistry of SiO₂ and the Si-SiO₂ Interface*. Boston, MA: Springer US, 1988.
- [17] D. M. Fleetwood, S. T. Pantelides, et R. D. Schrimpf, Éd., *Defects in microelectronic materials and devices*. Boca Raton: CRC Press, 2009.
- [18] W. Gös *et al.*, « Identification of oxide defects in semiconductor devices: A systematic approach linking DFT to rate equations and experimental evidence », *Microelectron. Reliab.*, vol. 87, p. 286–320, 2018.

- [19] W. Füssel, M. Schmidt, H. Angermann, G. Mende, et H. Flietner, « Defects at the Si/SiO₂ interface: their nature and behaviour in technological processes and stress », *Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.*, vol. 377, n° 2-3, p. 177–183, 1996.
- [20] A. Stirling, A. Pasquarello, J.-C. Charlier, et R. Car, « Dangling Bond Defects at Si- SiO₂ Interfaces: Atomic Structure of the P b 1 Center », *Phys. Rev. Lett.*, vol. 85, n° 13, p. 2773, 2000.
- [21] D. M. Fleetwood, « “Border traps” in MOS devices », *IEEE Trans. Nucl. Sci.*, vol. 39, n° 2, p. 269–271, 1992.
- [22] V. V. Afanas' ev et A. Stesmans, « Proton nature of radiation-induced positive charge in SiO₂ layers on Si », *EPL Europhys. Lett.*, vol. 53, n° 2, p. 233, 2001.
- [23] Artemisia TSIARA, « Electrical characterization & modeling of the trapping phenomena impacting the reliability of nanowire transistors for sub 10nm nodes », UNIVERSITE GRENOBLE ALPES, 2019.
- [24] T. Grasser, H. Reisinger, P.-J. Wagner, et B. Kaczer, « Time-dependent defect spectroscopy for characterization of border traps in metal-oxide-semiconductor transistors », *Phys. Rev. B*, vol. 82, n° 24, déc. 2010, doi: 10.1103/PhysRevB.82.245318.
- [25] J. L. Gavartin, D. Muñoz Ramo, A. L. Shluger, G. Bersuker, et B. H. Lee, « Negative oxygen vacancies in HfO₂ as charge traps in high-k stacks », *Appl. Phys. Lett.*, vol. 89, n° 8, p. 082908, août 2006, doi: 10.1063/1.2236466.
- [26] J. Robertson et R. M. Wallace, « High-K materials and metal gates for CMOS applications », *Mater. Sci. Eng. R Rep.*, vol. 88, p. 1-41, févr. 2015, doi: 10.1016/j.mser.2014.11.001.
- [27] C. J. Nicklaw, Z.-Y. Lu, D. M. Fleetwood, R. D. Schrimpf, et S. T. Pantelides, « The structure, properties, and dynamics of oxygen vacancies in amorphous SiO₂ », *IEEE Trans. Nucl. Sci.*, vol. 49, n° 6, p. 2667-2673, déc. 2002, doi: 10.1109/TNS.2002.805408.
- [28] S. Taylor, J. F. Zhang, et W. Eccleston, « A review of the plasma oxidation of silicon and its applications », *Semicond. Sci. Technol.*, vol. 8, n° 7, p. 1426, 1993.
- [29] G. Lucovsky, « Preparation of device-quality SiO₂ thin films by remote plasma-enhanced chemical vapour deposition (PECVD): Applications in metal-oxide-semiconductor (MOS) devices », *Adv. Mater. Opt. Electron.*, vol. 6, n° 2, p. 55–72, 1996.
- [30] P. Batude *et al.*, « Advances, challenges and opportunities in 3D CMOS sequential integration », in *Electron Devices Meeting (IEDM), 2011 IEEE International*, 2011, p. 7–3.
- [31] C. C. Hobbs *et al.*, « Fermi-Level Pinning at the Polysilicon/Metal Oxide Interface—Part I », *IEEE Trans. Electron Devices*, vol. 51, n° 6, p. 971-977, juin 2004, doi: 10.1109/TED.2004.829513.
- [32] K. Romanjek, « Caractérisation et modélisation des transistors CMOS des technologies 50nm et en deçà », PhD Thesis, Institut National Polytechnique de Grenoble-INPG, 2004.
- [33] J.-P. Colinge, *Silicon-on-Insulator Technology: Materials to VLSI*. Boston, MA: Springer US, 2004.
- [34] G. Ghibaudo, « New method for the extraction of MOSFET parameters », *Electron. Lett.*, vol. 24, n° 9, p. 543–545, 1988.
- [35] J. P. Colinge, « Silicon-on-insulator (SOI) junctionless transistors », in *Silicon-On-Insulator (SOI) Technology*, Elsevier, 2014, p. 167-194.
- [36] X. Garros *et al.*, « PBTI mechanisms in La containing Hf-based oxides assessed by very Fast IV measurements », in *2010 International Electron Devices Meeting*, 2010, p. 4–6.
- [37] T. Grasser et B. Kaczer, « Evidence That Two Tightly Coupled Mechanisms Are Responsible for Negative Bias Temperature Instability in Oxynitride MOSFETs », *IEEE Trans. Electron Devices*, vol. 56, n° 5, p. 1056-1062, mai 2009, doi: 10.1109/TED.2009.2015160.

- [38] A. Subirats *et al.*, « A new gate pattern measurement for evaluating the BTI degradation in circuit conditions », in *2014 IEEE International Reliability Physics Symposium*, Waikoloa, HI, USA, juin 2014, p. 5D.1.1-5D.1.5, doi: 10.1109/IRPS.2014.6860670.
- [39] A. Tsiara *et al.*, « Performance and reliability of a fully integrated 3D sequential technology », in *2018 IEEE Symposium on VLSI Technology*, 2018, p. 75–76.
- [40] T. Grasser *et al.*, « On the microscopic structure of hole traps in pMOSFETs », in *2014 IEEE International Electron Devices Meeting*, 2014, p. 21–1.
- [41] A.-M. El-Sayed, M. B. Watkins, T. Grasser, V. V. Afanas'ev, et A. L. Shluger, « Hydrogen-Induced Rupture of Strained Si–O Bonds in Amorphous Silicon Dioxide », *Phys. Rev. Lett.*, vol. 114, n° 11, mars 2015, doi: 10.1103/PhysRevLett.114.115503.
- [42] X. Garros *et al.*, « Process damages in HfO₂/TiN stacks: the key role of H₂/anneals », in *IEEE International Electron Devices Meeting, 2005. IEDM Technical Digest.*, 2005, p. 4–pp.
- [43] J.-P. Colinge, « Junctionless transistors », in *Future of Electron Devices, Kansai (IMFEDK), 2012 IEEE International Meeting for*, 2012, p. 1–2.
- [44] J. P. Colinge *et al.*, « Junctionless Transistors: Physics and Properties », in *Semiconductor-On-Insulator Materials for Nanoelectronics Applications*, A. Nazarov, J.-P. Colinge, F. Balestra, J.-P. Raskin, F. Gamiz, et V. S. Lysenko, Éd. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, p. 187-200.
- [45] R. Dutta et S. Kundu, « Effect of polysilicon gate doping concentration variation on MOSFET characteristics », in *Computers and Devices for Communication (CODEC), 2012 5th International Conference on*, 2012, p. 1–4.
- [46] J. Y. W. Seto, « The electrical properties of polycrystalline silicon films », *J. Appl. Phys.*, vol. 46, n° 12, p. 5247-5254, déc. 1975, doi: 10.1063/1.321593.
- [47] M. M. Mandurah, K. C. Saraswat, et T. I. Kamins, « Phosphorus Doping of Low Pressure Chemically Vapor-Deposited Silicon Films », *J. Electrochem. Soc.*, vol. 126, n° 6, p. 1019–1023, 1979.
- [48] Xavier Garros, « Reliability of 3D Architectures (Finfets, Nanowires) », 2020.

Low temperature junction

In this chapter, low temperature junction for high voltage devices is studied through electrical characterization and TCAD process and electrical simulation. The goal of this work is to reproduce the same characteristics of the high temperature junction ($>1000^{\circ}\text{C}$), but on a limited thermal budget. Different splits focusing on the pre-amorphization step are tested, in order to find the best compromise between performance and junction leakage performing dopant activation by solid phase epitaxial regrowth (SPER). Lastly, guidelines are provided in order to further decrease access resistance.

4.1 DOPANT IMPLANTATION

Ion implantation is a common method to add impurities in the intrinsic silicon, and it consists in the impact and penetration of energetic ions into the target solid. This impact damages the silicon structure, displacing the silicon atom out of its lattice position and creating defects.

Consequently, there is a lattice site with a missing atom, called a vacancy (a), and an extra atom called an interstitial (b), where both irregularities are referred to as point defects. As for the implanted species, most of them occupy interstitial positions (c), which are electrically inactive. To be active, the dopants must occupy substitutional sites (d), that is, occupy a lattice site normally occupied by silicon (Si) [1]. These configurations are represented on Fig. 4.1.

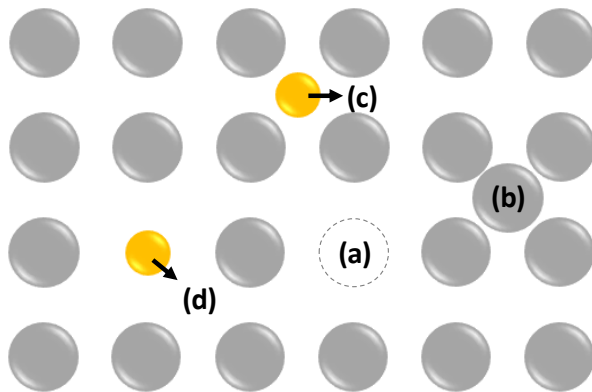


Figure 4.1 Schematic representation of crystalline lattice after ion implantation, where each letter corresponds to a lattice configuration previously described. Gray circles represent the Si atoms, whereas the implanted impurity is depicted in yellow.

As ions continually arrive and impact the target, damage builds-up progressively. If the density of defects reaches a threshold value (about $2 \times 10^{22}/\text{cm}^3$ for silicon [2]), the material changes its phase from crystalline to amorphous. Higher defects densities are expected for higher ion masses, energies, and doses.

4.2 DOPANT ACTIVATION

After an ion implantation step, a thermal anneal is usually required, in order to recover the crystalline lattice from the implantation damage and to electrically activate the dopants. The latter is usually performed at high temperatures ($>1000^\circ\text{C}$) to ensure that the damage removal is complete and at short times (second range) to limit dopant diffusion.

In order to have an estimation of the impurity activation level, one needs to look at the solid solubility limit (C_{SOL}) [1], [3]. C_{SOL} indicates the maximum number of dopants that can be incorporated into the silicon without the creation of a different phase, at thermodynamic equilibrium (for an ideal crystal without point defects). Fig. 4.2 summarizes the C_{SOL} values measured experimentally as a function of temperature for Phosphorus and Boron respectively, the most commonly used n-type and p-type dopants in silicon technologies. Since no work at 500°C has been reported, the C_{SOL} values @500°C have been obtained by a linear extrapolation.

At 1000°C, C_{SOL} values in the range of $\approx 4 \times 10^{20}$ at/cm³ and $\approx 1 \times 10^{20}$ at/cm³ are found for Phosphorus and Boron, respectively. As the temperature decreases, so does C_{SOL} , where extrapolated values @500°C are found in the range of $8 \cdot 10^{19}$ at/cm³ (for Phosphorus) and $2 \cdot 10^{18}$ at/cm³ (for Boron).

Overall, there is a strong discrepancy between the extrapolated C_{SOL} values @500°C (1 decade in the case of Phosphorus, and up to 3 decades in the case of Boron). This can be explained by the fact that solid solubility does not guarantee electrical activity, where C_{SOL} does not necessarily corresponds to the electrical solubility, here called C_{ELEC} [4]. Additionally, the ideal electrical solubility corresponds to the thermodynamic equilibrium, that is, with an infinite annealing duration and in a perfect crystal (without point defects).

These point defects can cause dopant activation anomalies [5], due to its interaction with the dopants to form immobile electrically inactive clusters. These interactions will depend on both implant and annealing conditions. If the implantation has generated a lot of point defects (or interstitials), the inactive cluster formation can increase, reducing C_{ELEC} . In the case of Boron, in the presence of a high amount of interstitials, further deactivation can occur due to the formation of Boron Interstitial Clusters (BICs) [3],[4]. Deactivation for Phosphorus can also occur, but differently from Boron, Phosphorus can interact with both interstitials and vacancies. One particular reference (S.F. Guo [7]) presents a superior C_{SOL} of $\approx 2 \times 10^{20}$ at/cm³ at 830°C, with an extrapolated value @500°C of 5×10^{19} at/cm³. The absence of point defects created by an implantation step could hence be the origin of this high C_{SOL} value. This superior Boron concentration has been obtained by Boron-Nitride

layer diffusion, which might have enabled the incorporation of boron atoms without implantation.

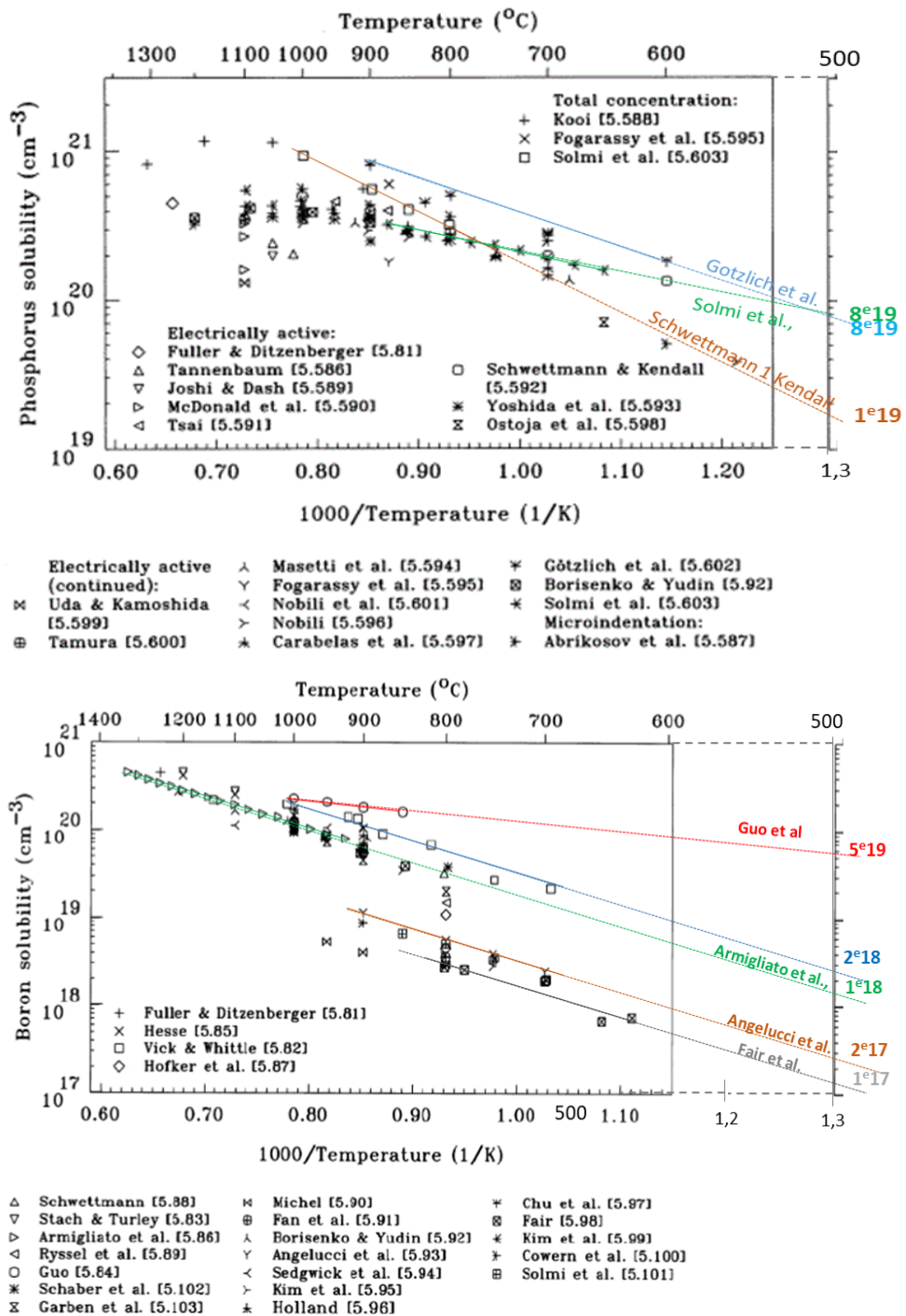


Figure 4.2 Solubility as a function of temperature of Phosphorus and Boron [1], respectively, together with extrapolated values at 500°C.

As mentioned in the introduction, high dopant concentration $>1 \times 10^{20}$ at/cm³ is required in order to achieve the low access resistance of the S/D. At 500°C, C_{sol} values are well below this target, hence another activation method must be used.

Solid Phase Epitaxial Regrowth (SPER) [8] can be used as an alternative, at temperatures (500°C) that are compatible with a 3D Sequential Integration. SPER is based on the recrystallization of an amorphous layer, as depicted on Fig. 4.3. Starting from a crystalline template, dopant implantation is performed. If the implant condition does not create an amorphous layer, a pre-amorphization implantation (PAI) with a heavy specie (for example, Germanium) can be performed. It is highlighted that a crystalline seed must be preserved. Once a small thermal anneal (for example, 500°C) is applied, the atoms in the amorphous phase reorder by local bond rearrangements at the amorphous/crystalline (a/c) interface. This process allows for a layer-by-layer epitaxial regrowth of the film, with the same crystalline orientation as the seed.

SPER active concentration (C_{SPER}) reaches a metastable solubility, significantly higher than the maximum equilibrium solubility (C_{SOL}) at 500°C. This is due to the fact that under the extremely non-equilibrium conditions just after recrystallization [7],[8], the silicon lattice can substitutionally incorporate a very high quantity of dopant atoms [11]. Above a certain high concentration level (6×10^{20} at/cm³ and 3×10^{20} at/cm³ for Phosphorus and Boron respectively, as reported in [11]), the implanted atoms start to form inactive clusters. These clusters of inactive dopants will lead to carrier mobility degradation and in turn access resistance increase.

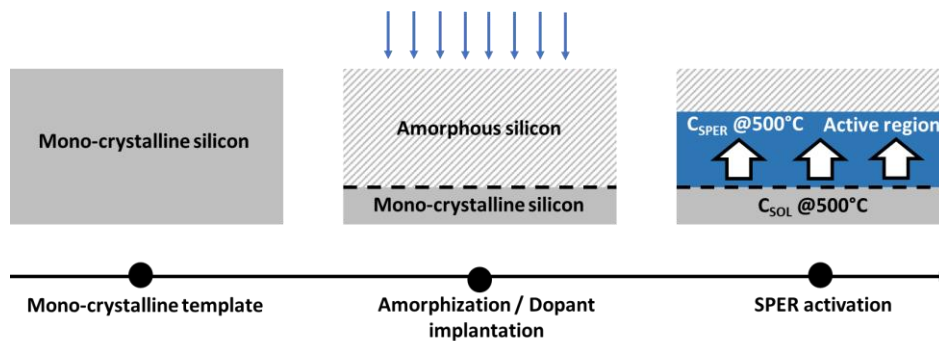


Figure 4.3 Schematic representation of SPER process.

4.3 END OF RANGE DEFECTS

After ion implantation, a high interstitial concentration is left below the a/c interface. This is because in the crystalline Si region, there are not enough Si vacancies to recombine with all the Si interstitials. During subsequent thermal anneal [12], the Si interstitials tend to

interact to form bigger and more stable defects, called End-of-Range (EOR). EOR can be generally completely removed, but for that, a long thermal budget is required.

Given the thermal budget constraint of 3DSI, after the SPER, the EOR are not eliminated and evolve with the annealing, initially forming small Si interstitial clusters that evolve into {311} defects and then dislocation loops. These {311}s and dislocation loops act as reservoirs of interstitials, which can have an impact on device performance [13].

I_{ON} current can be impacted, especially given the metastable nature of the SPER activation; this means that further thermal processing (for example during the back-end of line), could potentially lead to dopant deactivation. During annealing, there is a flow of interstitials diffusing out of the EOR region, inducing the BICs formation [14], [15]. BICs are also detrimental for the hole mobility, acting as independent scattering centers which have a strong impact on hole mobility in addition to the other scattering mechanisms such as lattice and impurity scatterings [16].

The emission of Si interstitials from EOR defects also impacts the transient enhanced diffusion (TED), observed during the initial stage of post implantation anneals [4],[15]. Overall, all dopants can suffer TED, by using point defects as a diffusion vehicle. However, TED is more important in the case of Boron because it mainly diffuses through the interstitial kick-out mechanism. This name perfectly describes the process: a Si interstitial “kicks out” the substitutional boron atom to an interstitial position where it can diffuse easily. The TED also depends on implant conditions, where more TED is observed for higher concentrations of excess Si interstitials. Diffusion also changes the junction profile, increasing short channel effects such as Drain-Induced Barrier Lowering (DIBL) [18].

Overall, the total leakage is a combination of different generation mechanisms. Among those, one can mention Shockley Read Hall (SRH) generation-recombination. The latter relies on the presence of deep levels in the depletion region, where the SRH dominated current is proportional to the width of the depletion region. Trap-Assisted Tunneling (TAT) can be considered as SRH in the presence of an electric field, or as a combination of electron capture and tunneling through the barrier. Band to Band Tunneling (BBT) occurs when the electric field across the junction is strong enough to propel a carrier from the valence band through the band gap, into the conduction band on the other side. More details on these mechanisms can easily be found elsewhere [19].

These contributions to drain leakage current are dependent on the doping profile, the field distribution around the drain edges and of course, on the presence of defects. An example of such is given on Fig. 4.4. If a defect is located at a distance larger than one minority-carrier diffusion length from the depletion boundary, its contribution to leakage is negligible. For example, the region around a contact to the N⁺-region can be highly defective but its impact on reverse leakage is minimized by keeping the contact boundary in silicon at a “safe” distance from the metallurgical junction. If the defect falls within or near the depleted region, the leakage current can increase drastically as $V_D - V_G$ increases.

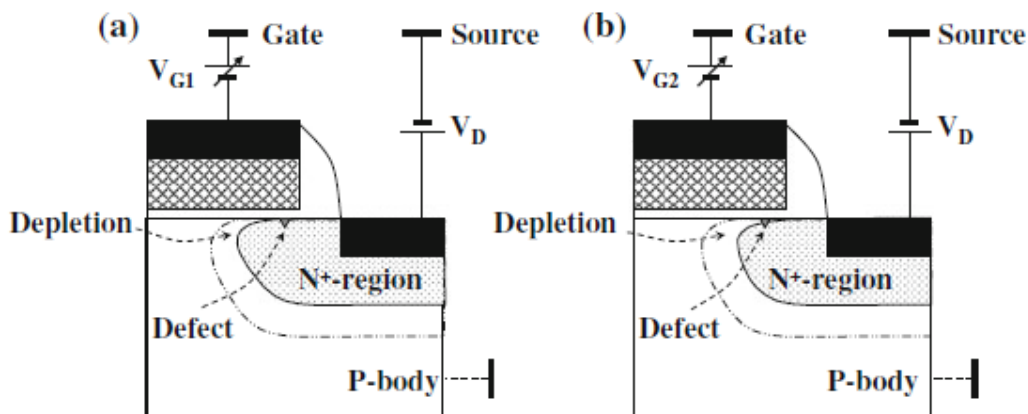


Figure 4.4 Surface defect near junction edge. a) Defect away from depletion boundary as opposed to b) defect near depletion boundary [20].

In the case of EOR, if located in or close to the space charge depletion region, it contributes to junction leakage by enhancing TAT and SRH.

4.4 DISSOLUTION OF END OF RANGE DEFECTS

As previously described, the diffusion of Si interstitials released from the EOR region towards the Si surface during annealing impacts both TED and dopant deactivation. However, as a consequence of the high interstitials value, the EOR will subsequently dissolve through the loss of free Si interstitials to nearby sinks, such as Si/SiO₂ interfaces [21], [22]. In the case of SOI devices, the BOX acts as an additional sink, competing with the top surface of the active Si layer. There will be two fluxes of Si interstitials, one towards the Si/SiO₂ of the top surface, and another towards the BOX. It must be stressed that the dissolution is stronger when the BOX is relatively close to the EOR defect band.

The BOX can also provide an additional benefit for EOR reduction due to the interstitials profile cut-off. After implantation, part of the Si interstitial defect profile get blocked into the BOX, making it that less Si interstitials are available for the formation of EOR defects [22]. A schematic comparing the EOR defect from Bulk x SOI structures is provided below (Fig. 4.5).

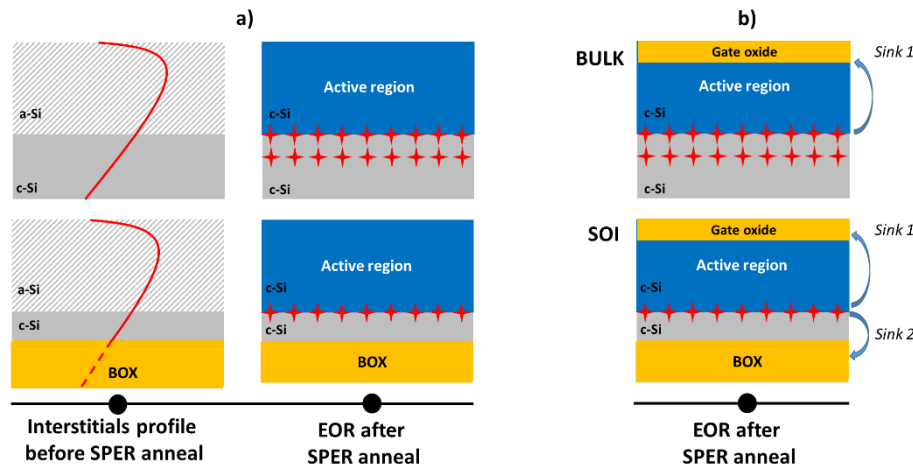


Figure 4.5 Schematic illustration of EOR cut-off (a) and dissolution (b) in BULK and SOI devices. In SOI, the BOX (as shown by the dashed line in) blocks part of the Si interstitials. This results in a smaller initial number of Si available for EOR formation, and less EOR defects are formed. The SOI structures also benefits from an additional Sink, the Buried Oxide (BOX).

4.5 LOW TEMPERATURE JUNCTION FOR HIGH VOLTAGE DEVICES ($V_{DD}=2.5V$)

As previously mentioned, SPER is a great alternative for achieving high dopant concentration of the S/D region, while having a compatible thermal budget in view of a 3DSI. Many efforts have been made in order to optimize the junction of digital devices ($V_{DD}=0.9-1V$), where good activation levels @600°C 1min have been achieved when comparing to the high temperature (HT) process of reference (POR) on 28nm FD-SOI devices. Furthermore, studies have also been performed in order to further reduce the SPER temperature to 500°C. Some examples are given on Fig. 4.6.

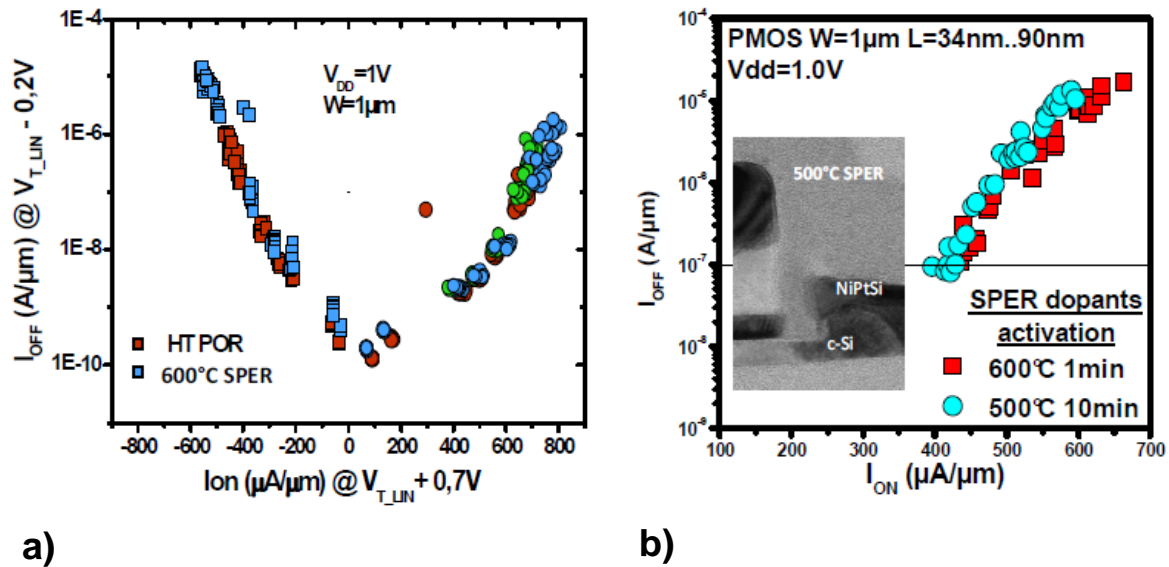


Figure 4.6 Demonstrations of SPER activated junction a) @600°C 1min compared with high temperature process of reference (POR) 28nm FDSOI as shown on I_{ON} - I_{OFF} trade-off. The SPER temperature has also been further reduced to 500°C 10min, as shown on b) for the I_{ON} - I_{OFF} trade-off of PMOS devices [23].

Given the promising results of the method, SPER shall also be used in this work, where the main objective of this chapter is to reproduce the characteristics of the high temperature devices, but in a low thermal budget (@500°C).

4.5.1 Process flow description

The high voltage devices ($V_{DD}=2.5V$) have 30nm thick undoped channel (with $\langle 100 \rangle$ channel orientation) and 25nm Buried Oxide (BOX). All splits are composed of a common gate stack of 6nm High Temperature Oxide (HTO) / TiN 6nm (PVD) / Polycrystalline silicon. The fact that all wafers have the same gate stack, allows one to discriminate between gate and junction properties. For more details about the gate stack, the reader may refer to chapter 3.

The junction is composed of two implantation steps, schematized on Fig. 4.7. A first dopant implantation is performed in the source and drain regions after gate patterning, through a residual $\text{SiO}_2=4\text{nm}$, called Low Doped Drain (LDD). This residual oxide was addressed through ellipsometry measurements performed after the gate etch (mean= 38.34 \AA with Goodness of Fit (GOF)=0.91). Phosphorus and Boron are used for NMOS and PMOS, respectively. The use of a Germanium implant for the LT case shall be described afterwards.

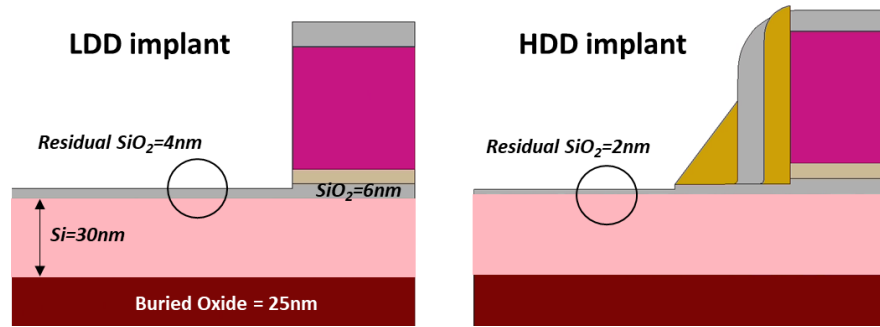


Figure 4.7 Schematic representation of the structure at the moment of LDD and HDD implants.

Concerning spacer formation, critical dimension (CD) measurements were performed after each spacer etching, as shown on Fig. 4.8. After spacer formation, a second doping step is performed through a residual $\text{SiO}_2=2\text{nm}$, called High Doped Drain (HDD).

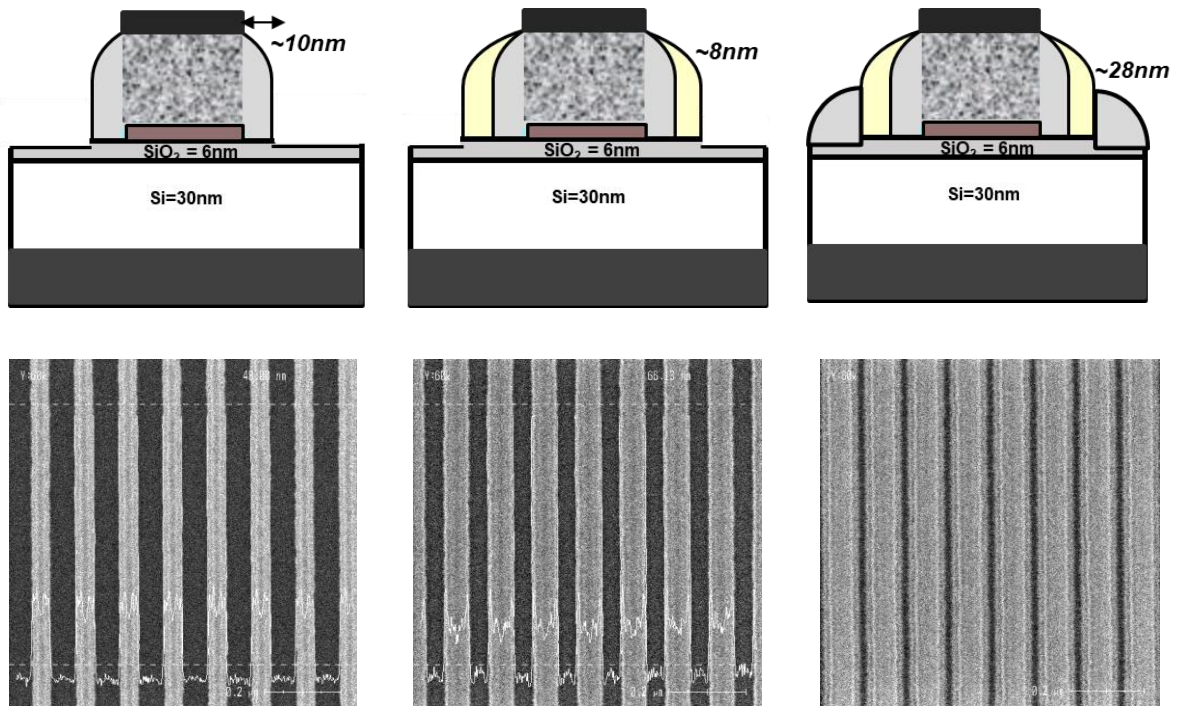


Figure 4.8 Critical dimension measurements after spacer etching.

For the low temperature process, SPER was used for junction activation. On the other hand, the high temperature device used as a reference underwent conventional high temperature LDD and S/D activation anneals. It is worth noticing that after implantations, the thermal budget to manufacture the low temperature devices is in the 500°C range. Lastly, standard silicide and BEOL was carried out. The complete described process flow is shown on Fig. 4.9.

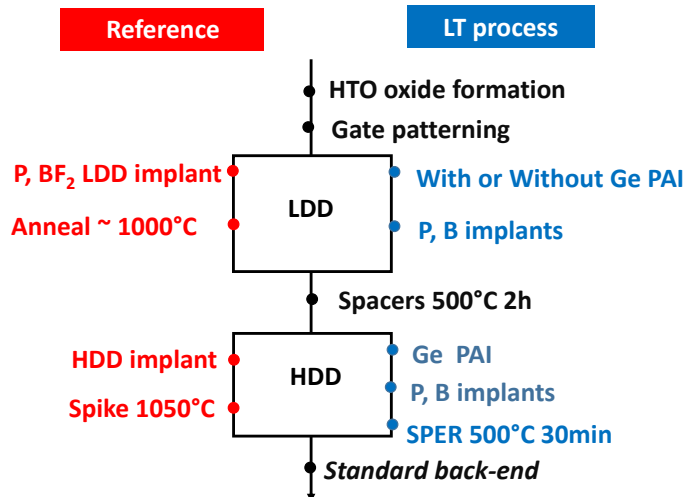


Figure 4.9 Process flow of 2.5V devices for low and high temperature process.

4.5.2 LDD implants

Kinetic Monte Carlo simulation [24] (KMC) was used to define the corresponding implant conditions of the low temperature junction. The main goal was to properly match the low temperature with the high temperature profile. Their corresponding as-implanted profile is then shown on Fig. 4.10.

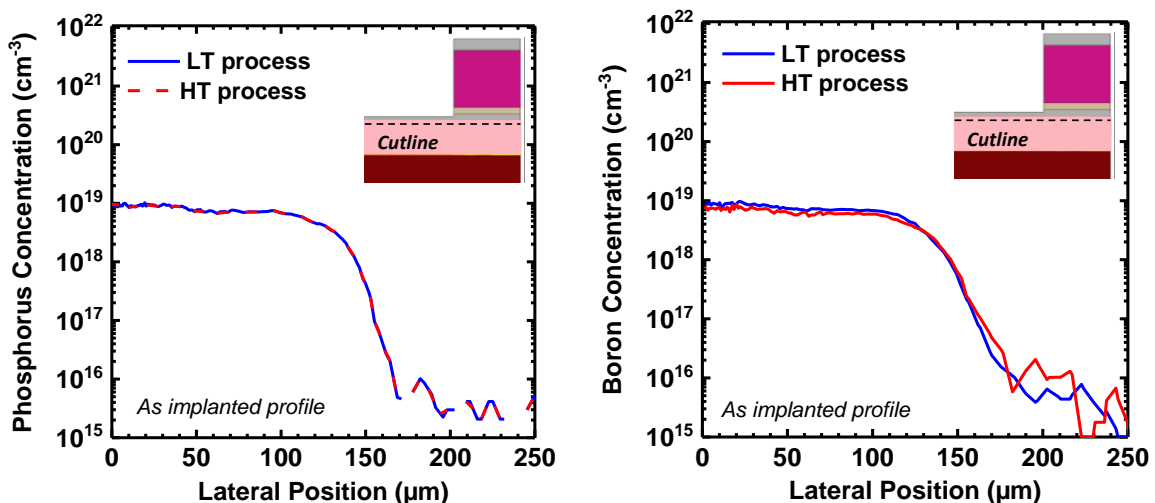


Figure 4.10 KMC simulation depicts the LDD (as implanted) junction profile of Phosphorus and Boron for low temperature splits as compared to the high temperature case. The inset shows the position of the cutline.

It is highlighted that the low temperature process fits well the high temperature case. More importantly, it is observed that the LDD profile of 2.5V devices have a maximum doping concentration around $1 \times 10^{19} \text{at/cm}^3$, hence the first step consists in choosing the best activation method for this region. This is because, given the now so high dopant concentration, maybe the LDD can be thermally activated @500°C.

If the LDD can be thermally activated, this means that SPER for LDD is not needed. As a reminder, in the SPER process, only dopants present in the pre-amorphized region will be further activated by SPER mechanism. This means that by suppressing the SPER for the LDD, one may as well suppress the pre-amorphization implant. As highlighted from the recapitulation table below of C_{SPER} and C_{SOL} for Phosphorus and Boron, thermal activation C_{SOL} @500°C could be sufficient.

	C_{SOL} @500°C	C_{SPER} @500°C
Phosphorus	$1e19 - 8e19$	$6e20$
Boron	$1e17 - 2e18$	$3e20$

Table 4.1 Recapitulation table of C_{SOL} and C_{SPER} extrapolated at 500°C, as seen on Figure 4.2. All values are given in cm^{-3}

Even though it is believed that the SPER for LDD can be suppressed (hence also suppressing the PAI), some splits with a PAI for LDD are also used, for two distinct reasons:

- Compare the activation levels between the splits with PAI (activated by SPER) and without PAI (thermally activated @500°C).
- Address the impact of the PAI on the junction leakage. It is possible that by suppressing the PAI (that creates EOR defects described on section 4.3) would lead to lower leakage current.

The pre-amorphization implantation conditions for PMOS and NMOS devices has been chosen by KMC simulations, with the goal to create an amorphous layer of a certain desired thickness.

Given that the LDD conditions were not sufficient to create the amorphous layer, the PAI was performed with a Germanium implant. It can be seen from Fig. 4.11 that by tuning the implantation condition, different amorphization thicknesses are achieved.

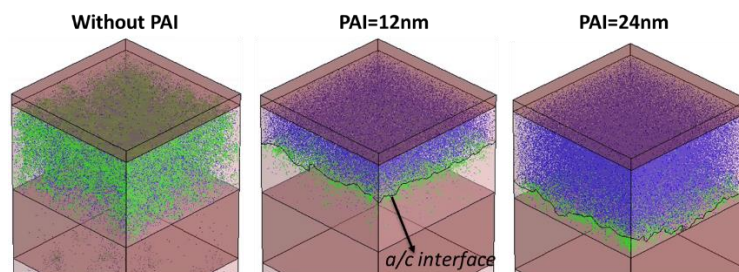


Figure 4.11 KMC amorphization thickness for a Si channel of 30nm. The blue dots represents the silicon point defects, indicating that an amorphous region is created once a certain threshold value is reached.

Therefore, this work shall analyze LDD splits where (1) no PAI prior to LDD has been performed, PAI leading to an amorphous thickness of (2) 12nm and (3) 24nm of the 30nm-thick Si-film.

4.5.3 HDD implants

Concerning the High Doped Drain (HDD), a high and constant doping profile with doping concentration $>1 \times 10^{20}$ at/cm³ is targeted. In this case, this concentration is well above the expected achievable C_{SOL} value at 500°C (please refer to C_{SOL} recapitulation table 1), the latter being especially true in the case of Boron. Hence *for HDD, SPER activation is mandatory*.

As previously discussed, a crystalline seed must be preserved to enable recrystallization. From previous works [25], it has been demonstrated that the choice of seed thickness is crucial, mainly for two reasons.

If the remaining crystalline seed is too thick, it means that there is a significant percentage of the film where the dopants are limited to thermal activation at 500°C (since activation by SPER is efficient only in the amorphous region). On the other hand, since a seed is required for SPER, the implantation condition needs to be carefully chosen in order to not amorphized the entire film.

It must be mentioned that for digital devices, it has been noticed that in the absence of a crystalline seed at the bottom of the S/D, the crystalline channel at the very edge of the gate could act as a seed, and thus in some cases, the recrystallization proceeds from the channel edge [18]. However, this lateral recrystallization is slower. Nevertheless, this would not be the case in the high voltage device architecture, seeing that there is a large set of spacers separating the HDD implant from the channel.

Therefore, for the HDD of LT devices, a maximum amorphization depth of ≈ 27 nm was defined, leaving a crystalline seed thickness of 3nm. The implantations were carried out through a native SiO₂ layer of ≈ 2 nm. In the case of NMOS, no PAI was required, given that only a Phosphorus implant with 20° tilt was sufficient to create an amorphous layer of 26.5nm. As for the PMOS, a PAI step + Boron implant was needed, leading to an amorphous region of 24.5nm.

In the end, the amorphous thickness is smaller than 27nm, which can be beneficial either if (1) there was an absence of SiO₂ prior to implantation or (2) if the channel was a bit smaller than 30nm. Similarly to what have been done for the LDD case, the HDD conditions were defined by KMC simulation and are shown below on Fig. 4.12.

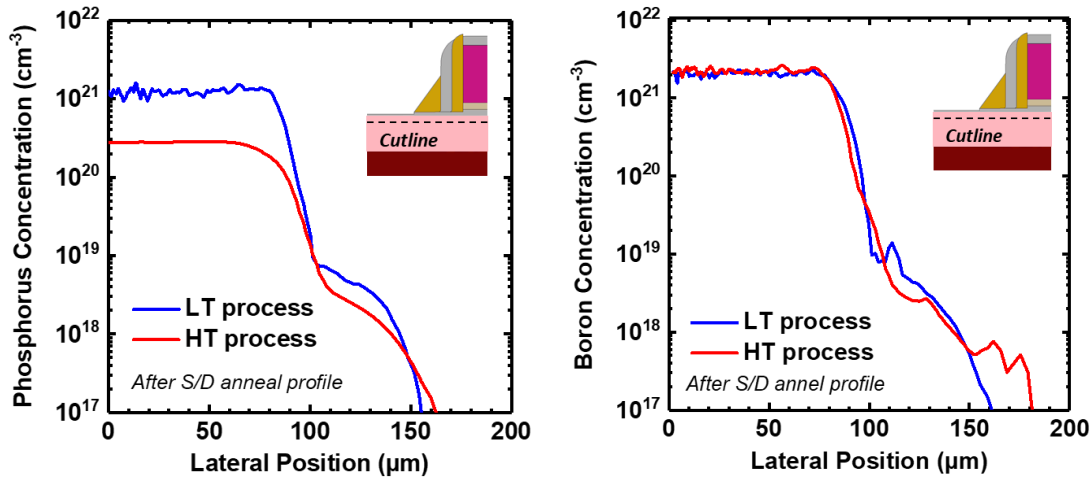


Figure 4.12 KMC simulation depicts the HDD junction profile of Phosphorus and Boron for both low and high temperature process. The inset shows the position of the cutline.

Additionally, Fig. 4.13 also depicts the 2D iso-concentrations along the Silicon channel for N&PMOS, respectively. In both cases, it can be seen that the chemical dopant concentration near the channel edge is inferior to 1×10^{19} at/cm³.

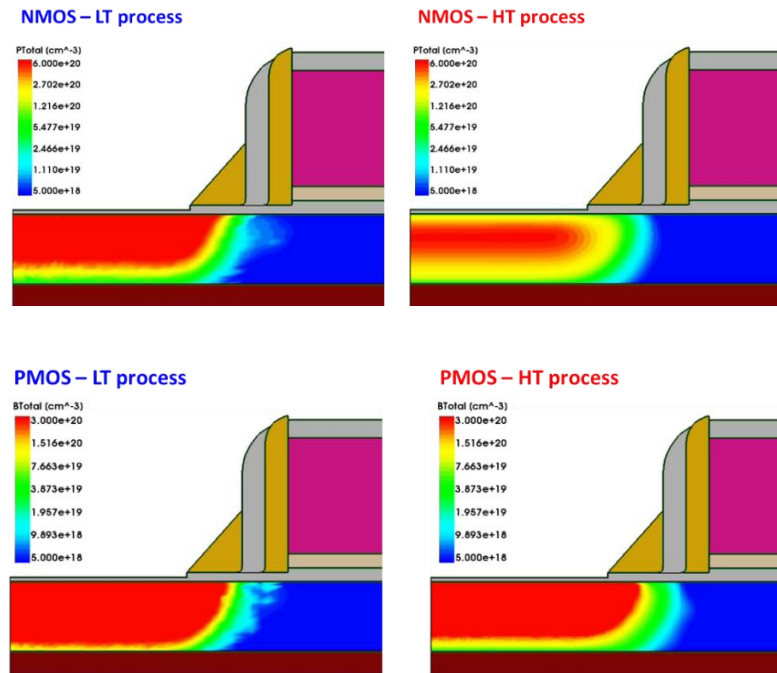


Figure 4.13 Iso-concentration of the phosphorus and boron chemical-concentration profile after anneal of low and high temperature process.

4.5.4 SPER activation temperature

In order to define the temperature for the SPER activation, one needs to consider the SPER velocity, which has been already extensively studied [18]. From those results, deduced values for the recrystallization were used, where 500°C for 30min should suffice to recrystallize $\approx 26.5\text{nm}$ of the S/D regions for the given concentration and channel orientation. It must be highlighted that this is also the targeted thermal budget for 3D sequential integration.

4.6 ELECTRICAL RESULTS @500°C 30MIN

Both high temperature and low temperature devices with width of $2\mu\text{m}$ and channel lengths of $0.45\mu\text{m}$ have been characterized for the nominal $V_{DD}=2.5\text{V}$. The I_{ON} - I_{OFF} trade-off comparison between process of reference and low temperature devices is shown on Fig. 4.14 for both NMOS and PMOS.

It is observed that while all the NMOS splits present similar I_{ON} - I_{OFF} as the high temperature reference, the PMOS suffers from I_{ON} penalty of -30% for all splits at constant

$I_{OFF} = 10^{-7}$ A/ μ m. It is also worth mentioning that contrarily to what was expected, no I_{ON} boost is observed for the splits with a PAI for the LDD region.

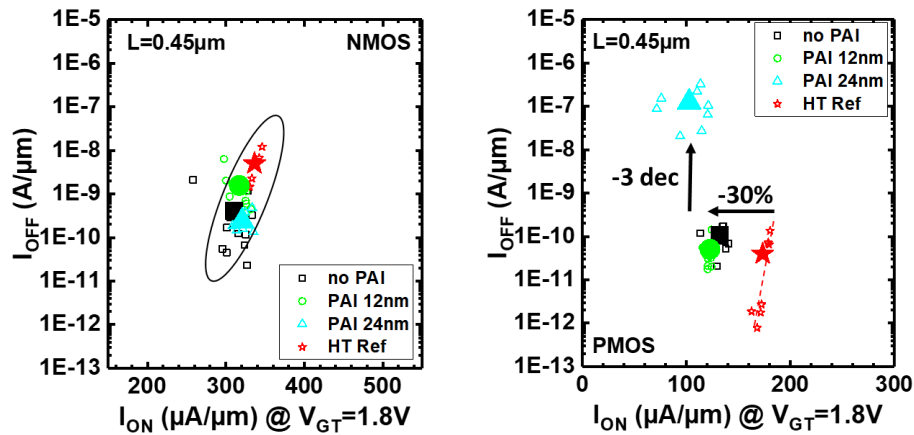


Figure 4.14 I_{ON} - I_{OFF} trade-off for high temperature reference and low temperature splits for the different PAI thicknesses @ $V_{DD}=2.5V$. Whereas each empty symbol represents a data point, the filled symbols represents the mean.

Concerning the O_{OFF} regime, the NMOS does not seem to be affected by a PAI step. As for the PMOS, a LDD PAI=24nm appears to be detrimental to the I_{OFF} of these devices, leading to 3 decades increase. The next sections aim at understanding these results.

Lastly, it is also important to highlight that overall, the data points are very scattered. This can especially be seen in the case of the I_{OFF} of the reference wafer. In this lot, it is believed that a contamination occurred during the deposition of the HTO. As a result, many dies presented a high gate leakage, and were excluded from the analysis. This also explains the dispersiveness and low amount of data points.

4.6.1 Understanding the I_{ON} results

Overall, two main questions remain to be answered:

- where is the I_{ON} degradation coming from?
- and why there is no performance difference between the splits with and without LDD PAI, knowing that a higher activation level is expected with SPER ($C_{SPER} > C_{SOL}@500^{\circ}C$).

The discussion shall focus on the most extreme cases, such as a comparison between LDD no PAI and PAI=24nm.

Firstly, it is important to validate the channel thickness (T_{Si}). If a thinner channel were present, it could be possible that all the Si film became amorphous at the moment of the

PAI, thus impacting growth direction and recrystallization speed. Fig. 4.15 depicts a Transmission Electron Microscopy (TEM) image of the finalized flow, where a $T_{Si}=30\text{nm}$ is highlighted.

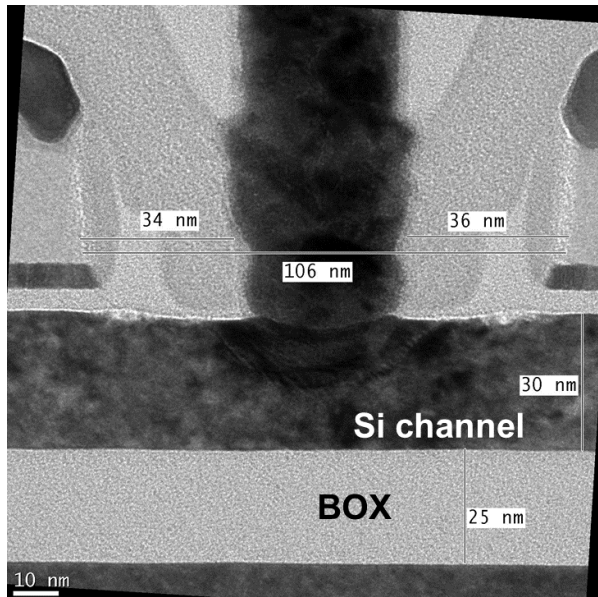


Figure 4.15 TEM image of the studied device, thus highlighting that the channel thickness has indeed 30nm.

Other than that, several parameters could potentially influence the performance of the device and shall be briefly described below.

The first parameter to be mentioned is the threshold voltage (V_{TH}), where a higher V_{TH} in the low temperature wafer could explain the lower performance. However, the $I_{ON}-I_{OFF}$ results are already shown at a constant $V_G - V_{TH}$ ($V_{GT}=1.8V$), meaning that any V_{TH} variation should not be considered.

Another parameter that should be discussed is the mobility. However, the latter could not be extracted due to the HTO contamination issues previously described. Overall, since the same gate stack was used, it could be believed that the mobility would remain unchanged for both low temperature and high temperature device. However, the gate stack of the high temperature split also “sees” the S/D activation temperature ($>1000^\circ\text{C}$), which could act in favor of a higher mobility value w.r.t. the low temperature case. As a result, it is believed that the mobility is the same for high temperature and low temperature devices (for example, as seeing the I_{ON}/I_{OFF} of NMOS that is not degraded), but it is not possible to conclude on that for sure. Likewise, no variation in the oxide thickness is expected, but the latter could also not be extracted due to HTO contamination issues. With

that said, there are still two things that could impact the performance, the electrical gate length, and the access resistance.

One way to address the electrical gate length is by analyzing the DIBL. Starting from the DIBL of PMOS devices, it is observed from Fig. 4.16 that there is no significant change between the low temperature and high temperature splits. In addition, the DIBL seems unchanged despite amorphization thickness (PAI splits). These are reassuring results, especially for the case with $PAI_{(LDD)}=24nm$. This particular split, as observed on the I_{ON} - I_{OFF} trade-off, has an increased leakage, which could be attributed to an increased number of EOR defects.

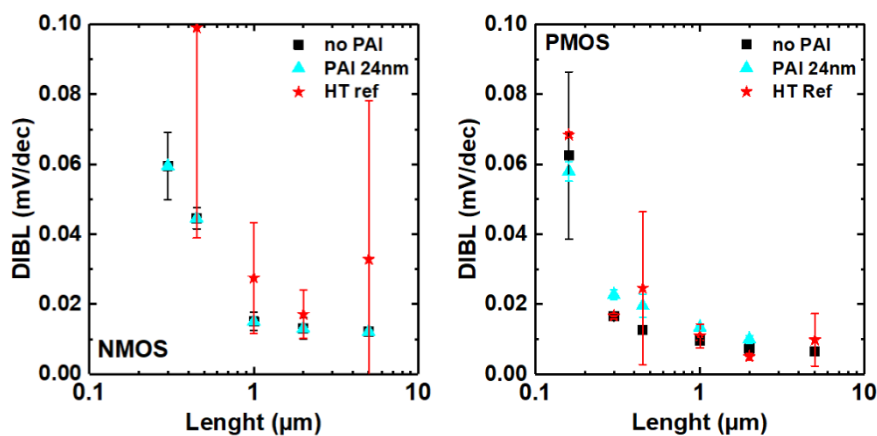


Figure 4.16 DIBL for high temperature reference and the different low temperature splits.

The fact that DIBL seems stable, could be an indication that no significant additional Boron TED occurs.

Concerning the NMOS, the low temperature splits with and without $LDD_{(PAI)}$ present the same results. In the case of the high temperature reference, it is observed that (1) there is a significant variation between the dies and that (2) for smaller gate lengths ($L_G \leq 0.3\mu m$) the DIBL could not be extracted.

These results allow one to conclude that (1) there is no variation in the electrical gate length due to a PAI, and that (2) PMOS devices do not experience any profile modification due to Boron diffusion. In addition, even though the electrical gate length seems stable when comparing high and low temperature processes, it is hard to conclude due to the variability of the high temperature data points.

At this point, based on the previous results, the dopant activation levels must be investigated. Fig. 4.17 depicts the I_{ON}/I_{OFF} trade-off for a range of gate lengths (from $L_G=5\mu\text{m}$ to $0.16\mu\text{m}$).

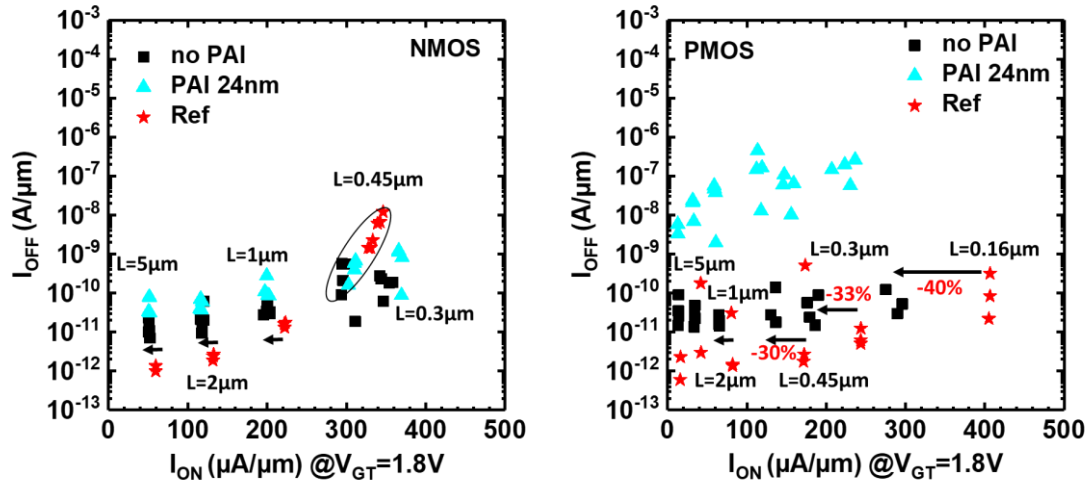


Figure 4.17 I_{ON} - I_{OFF} trade-off for high temperature reference and low temperature splits for the different PAI thicknesses @ $V_{DD}=2.5\text{V}$ for different gate lengths.

In the case of the NMOS low temperature case, for the largest L_G , there is already a slight I_{ON} degradation of $\approx 8\%$ w.r.t the high temperature process. As L_G is reduced, the I_{OFF} starts to increase, such that for $L_G=0.45\mu\text{m}$ the I_{OFF} of the high temperature case becomes superior to that of the low temperature one. *No degradation between $PAI_{(LDD)}=24\text{nm}$ and no $LDD_{(PAI)}$ is highlighted.* For smaller gate lengths ($L_G=0.3\mu\text{m}$), it was not possible to extract the I_{ON} of the reference wafer. Despite, it is possible to compare the $PAI_{(LDD)}$ of the low temperature splits. In that case, it seems that as the access resistance starts to become more important a better activation is obtained with an $LDD_{(PAI)}=24\text{nm}$ (in comparison to no $LDD_{(PAI)}$).

As for the PMOS, it is evidenced that the I_{ON} degradation becomes even more important for smaller dimensions, where up to -40% is seen for $L_G=0.16\mu\text{m}$.

Lastly, R_{TOT} vs. DIBL is depicted on Fig. 4.18. In the case of the NMOS, it could explain the I_{ON} - I_{OFF} results, where a slightly higher I_{ON} is observed for the high temperature split at the expense of a higher DIBL. It must however be highlighted that the error margin of the high temperature split is quite important. As for PMOS devices, there is a degradation observed for the LDD PAI 24nm split that is not observed on the I_{ON} degradation.

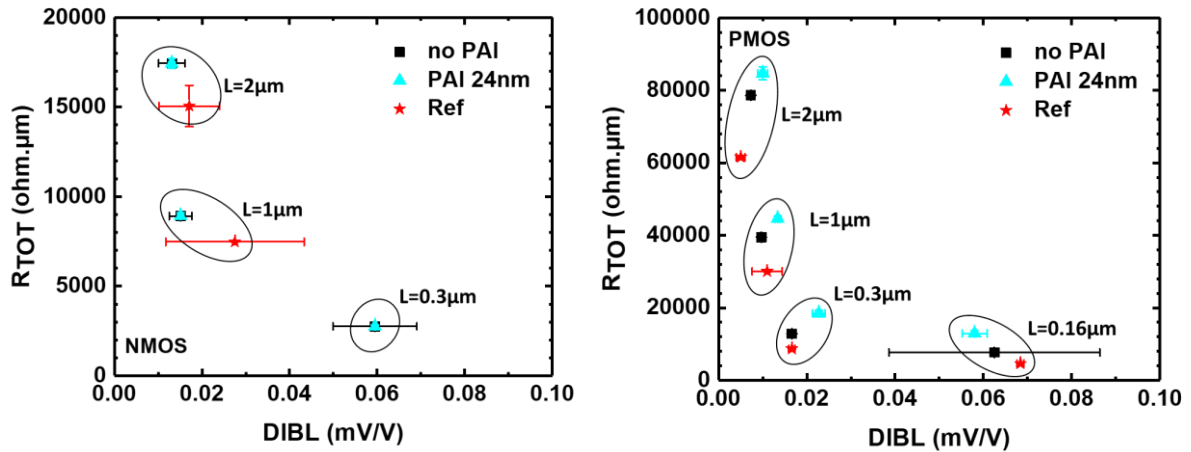


Figure 4.18 R_{TOT} vs DIBL of NMOS and PMOS, respectively, for high temperature reference and low temperature splits with different PAI thicknesses.

Overall, table 2 summarized the parameters shown so far, and compares with the reference. Based on the results, the most probable FOM that could explain the performance loss is a degradation of the access resistance. This means that the activation levels on the low temperature may not be exactly the same as the one of the high temperature reference (especially in the case of PMOS). As a result, the last parameter that needs to be evaluated that could explain the performance results is the access resistance.

Figure of Merit (FOM)	LT (500°C)			
	no PAI(LDD)	PAI(LDD) 24nm	no PAI(LDD)	PAI(LDD) 24nm
	NMOS		PMOS	
I_{ON}	=	=	-30%	-30%
Oxide thickness	= *	= *	= *	= *
Mobility	= *	= *	= *	= *
DIBL	?	?	=	=
R_{TOT}	+	+	++	++

Table 3.2 Compares the results between high temperature reference and low temperature splits. The equal sign means that there is no variation w.r.t. the reference. The + sign means that there is an increase in the value w.r.t. the reference. Concerning gate oxide and mobility, both are considered to be equal but could not be extracted. As for the DIBL of NMOS, it is hard to conclude due to data variability.

4.6.2 Access resistance definition

In a long channel device, the source and drain parasitic resistance becomes negligible compared to the channel resistance. However, as the gate length is decreased, these two resistances become comparable and can therefore cause a significant current degradation. The total transistor resistance is then defined as:

$$R_{TOTAL} = \frac{V_{DS}}{I_{DS}} = R_{ch} + R_{access} \quad (\text{Equation 4.1})$$

Where R_{access} is the access resistance and R_{ch} the channel resistance. R_{TOTAL} is also often referred to as R_{ON} ($R_{ON}=R_{TOTAL}$ extracted at $V_{Dlin}=100\text{mV}$). Seeing that the gate length of a high voltage device ($L_{G,nominal}=0.3\mu\text{m}$) is far superior to those of advanced digital devices ($L_{G,nominal}=30\text{nm}$), its access resistance should be less of an issue. Still, the -30% PMOS degradation observed for $L_G=0.45\mu\text{m}$ needs to be understood.

Equation 4.2 defines the access resistance as the resistance corresponding to the voltage drop in the source and drain regions. The different contributions to the resistance of high voltage devices can be decomposed in several resistance contributions, depicted on Fig. 4.19.

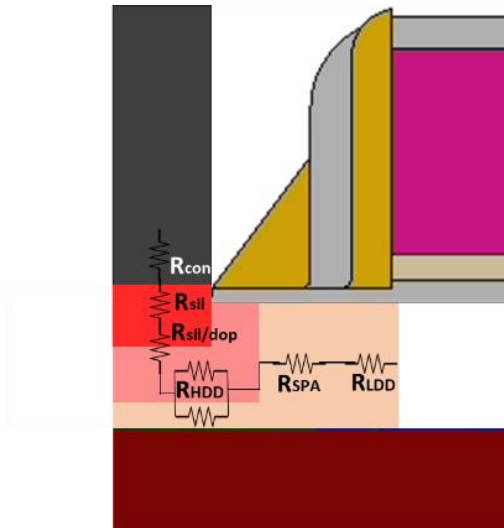


Figure 4.19 Schematic representation of the different contributions to the access resistance of the studied devices. In this example, a low temperature junction is taken into account.

With:

$$R_{access} = R_{con} + R_{sil} + R_{sil/dop} + R_{HDD} + R_{SPA} + R_{LDD} \quad (\text{Equation 4.2})$$

where R_{con} is the contact resistance, R_{sil} the resistance corresponding to the silicided region, $R_{sil/dop}$ the interface resistance between silicided region and doped region, R_{HDD}

and R_{LDD} are the resistances related to highly and lightly doped region respectively, and R_{SPA} the resistance corresponding to the zone below the second set of spacers.

R_{con} and R_{sil} can be considered equivalent for high temperature and low temperature devices. Concerning $R_{sil/dop}$, if there is a dopant variation in the HDD region, this could eventually be problematic for the access resistance.

In the case of a low temperature junction, the R_{HDD} can be decomposed into 2 parallel contributions, attributed to the zone activated by SPER (with a high activation values, as shown on table 1), and the crystalline seed region, which activation will be limited by C_{SOL} @500°C (Fig. 4.20). In this work, as previously mentioned, the amorphization depth varies between 26.5nm and 24.5nm, for N&PMOS respectively. For advanced digital applications, where the channel thickness=7nm, this region can strongly contribute to R_{SD} . In this work, given that the channel is thick (30nm), and these contributions are in parallel, R_{HDD} should not be crucial.

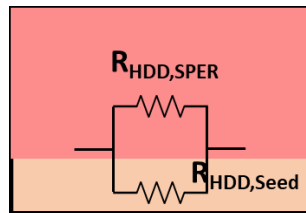


Figure 4.20 Zoomed schematic depicting the contributions to R_{HDD} .

The two other remaining resistance contributions (R_{LDD} and R_{SPA}) are very dependent on the doping profile and shall be discussed separately.

4.6.2.1 Access resistance extraction

Fig. 4.21 depicts R_{ON} as a linear function of gate length, where the access resistance (R_{SD}) corresponds to the intercept [26] in the y-axis. In the case of NMOS, no R_{SD} difference is evidenced for the split with or without $PAI_{(LDD)}$. Despite, there is almost a factor 3 between low temperature and high temperature split, even though not evidenced on I_{ON}/I_{OFF} trade-off. Concerning PMOS devices, there is now an important increase in R_{SD} for the split with $PAI_{(LDD)}=24\text{nm}$, also not evidenced on I_{ON}/I_{OFF} . As for the high temperature reference, R_{SD} extraction was not possible, where the intercept leads to negative values.

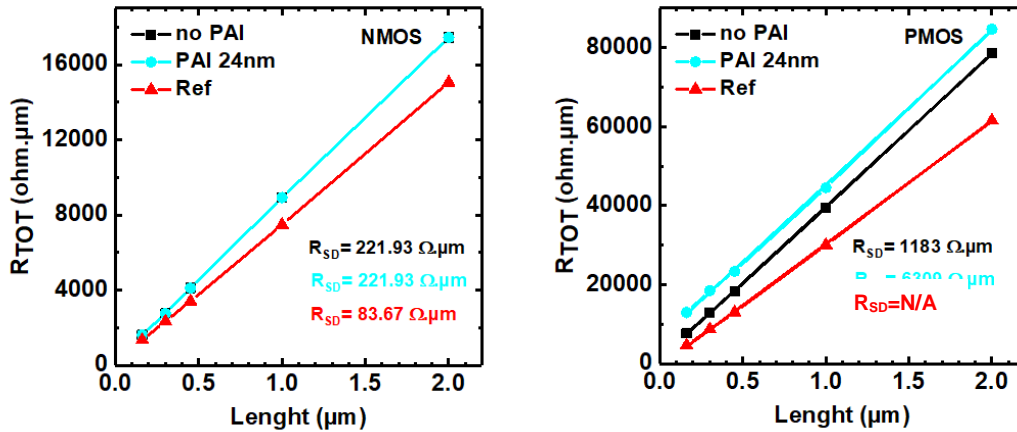


Figure 4.21 $R_{TOT} \times L_G$ of NMOS and PMOS, respectively. RSD

It then becomes hard to quantify the R_{SD} different between high and low temperature processes, especially because the number of tested dies was not sufficient to have better statistics. R_{SD} was also extracted with the Y function method [27], and is compared with the values obtained from the total resistance method (Fig. 4.22). Despite the device type, there are inconsistencies in both methods, where the extracted values do not correspond to the observed I_{ON} degradation.

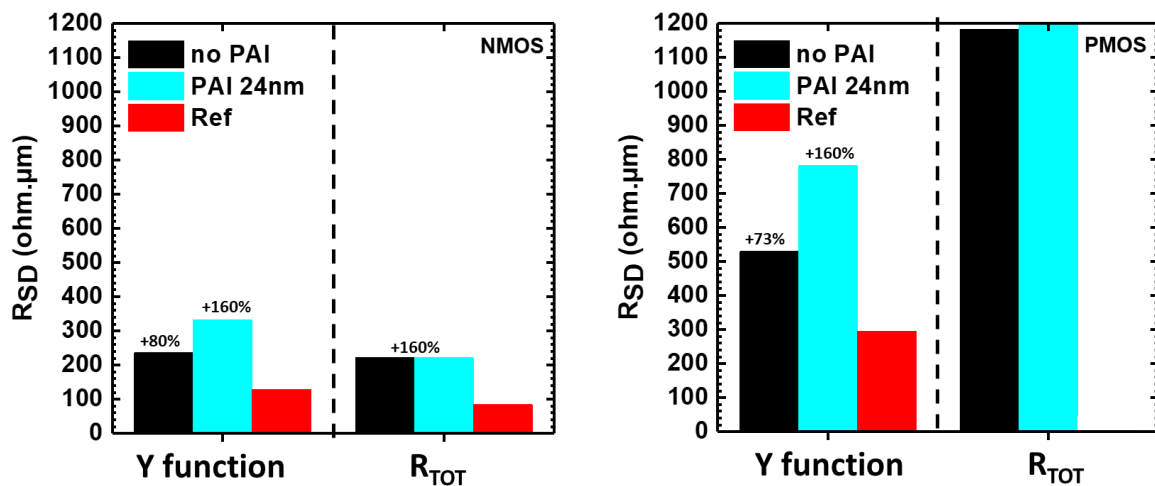


Figure 4.22 R_{SD} values extracted from Y function and total resistance method. The percentages are given w.r.t. high temperature reference.

As such, further investigation is necessary in order to address the junction of the devices.

4.6.2.2 Discussion on junction activation mechanisms

The goal of this section is to describe, step by step, the activation mechanisms that took place in the studied device. Dopants chemical concentrations and amorphization position of the proposed junction architectures shall be displayed, obtained through 2D Kinetic Monte Carlo (KMC) simulations. For illustration, only Phosphorus has been shown, but the same idea is applied in the case of Boron.

As previously described, splits with and without Pre-Amorphization Implants for LDD ($\text{PAI}_{(\text{LDD})}$) activation have been tested and these splits shall always be compared (no $\text{PAI}_{(\text{LDD})}$ depicted on the left, whereas $\text{PAI}_{(\text{LDD})}=24\text{nm}$ shown on the right).

To start, the LDD is first implanted (step 1, Fig. 4.23). In the case of no $\text{PAI}_{(\text{LDD})}$, only Phosphorus is employed. Contrarily, in the case of $\text{PAI}_{(\text{LDD})}=24\text{nm}$, the Phosphorus implant is preceded by a Ge implant, thus creating an amorphous template (required for SPER) as shown on the amorphous/crystalline interface.

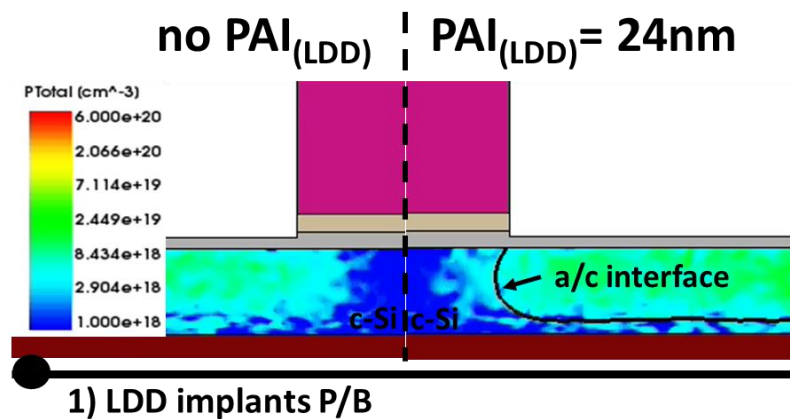


Figure 4.23 Step 1: LDD implantation for the splits without (on the left) and with a PAI prior to LDD.

The next step is to form the first spacer (step 2, Fig. 4.24). It has been observed that due to the spacer deposition, wafers withstood an additional thermal budget of 500°C 2h. It is highlighted that the latter is sufficient to recrystallize an amorphous region. As a consequence, the spacer formation led to a LDD dopant activation. However, the activation mechanisms are not the same; in the case of no $\text{PAI}_{(\text{LDD})}$, the wafers withstood a thermal activation @ 500°C . The electrical concentration of these dopants shall thus be limited to C_{SOL} . On the other hand, for the split with $\text{PAI}_{(\text{LDD})}=24\text{nm}$, given the amorphous region created by the PAI, a SPER mechanism took place. Hence, the electrical concentration of these dopants are equivalent to C_{SPER} .

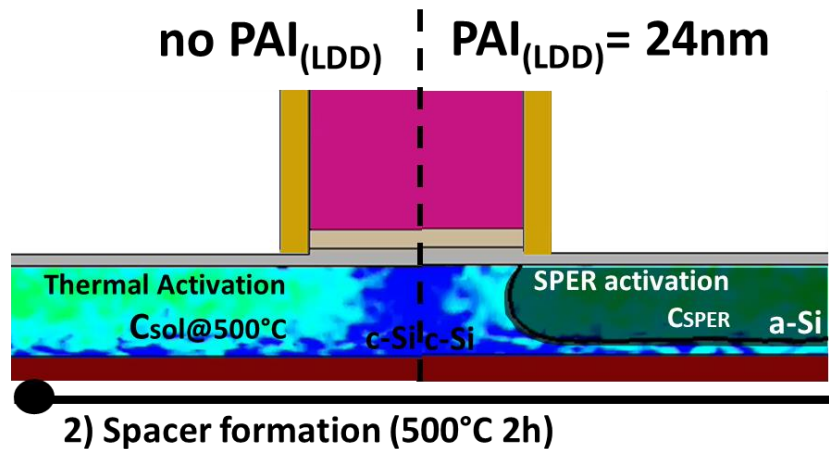


Figure 4.24 Step 2: First spacer formation @500°C 2h. This thermal budget is sufficient enough to recrystallize an amorphous region, thus leading to a SPER activation for the split $PAI_{(LDD)}=24nm$. On the other hand, the no $PAI_{(LDD)}$ split underwent a thermal activation @500°C.

That said, it is now important to understand how this spacer's thermal budget influenced the junction activation. By proceeding with the process flow, the second set of spacers is formed (this time at 400°C), followed by the HDD implants (step 3, Fig. 4.25). A comparison is made between the $PAI_{(LDD)}$ splits, and with or without a spacer's thermal budget.

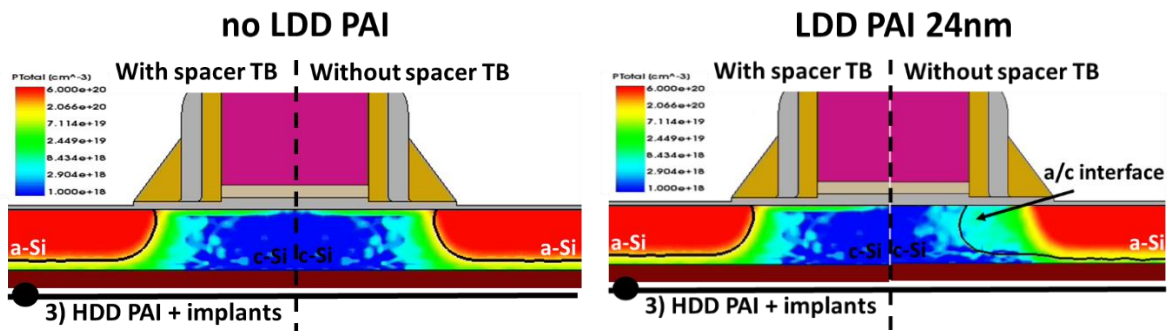


Figure 4.25 Step 3: HDD implantation. It must be highlighted that whilst the spacer's thermal budget has no impact for the split with no $PAI_{(LDD)}$, the latter impacts the amorphous region under the spacer.

Starting from the split $PAI_{(LDD)}=24nm$, it can be noticed that the position of the a/c interface is different. This is because the spacer's thermal budget recrystallized the LDD region. Once more, the dopants are only activated by SPER if prior to this activation, the region was amorphous. As a result, because of the spacer's thermal budget, the HDD dopants are only going to be activated by SPER in the amorphous region created by the PAI prior to HDD ($PAI_{(HDD)}$). In this case, the a-c interface is limited at a few nanometers below the last triangular shape spacer. On the other hand, not taking into account the

spacer's thermal budget, it can be seen that the amorphous region would extend all the way up to the gate edge. The implications of that is that all the region below the spacer's would have been amorphous, hence activated by SPER.

As for the split with no PAI_(LDD), a spacer's thermal budget has absolutely no effect since the region was not amorphous before. Nevertheless, it is important to highlight that because of the spacer's thermal budget, the no PAI_(LDD) and PAI_(LDD)=24nm had a similar activation mechanism. This could then explain why there was no performance improvement (as observed on the I_{ON}-I_{OFF}) for the splits with a PAI_(LDD).

Lastly, a SPER @500°C 30min is performed to activate the HDD implants (step 5, Fig. 4.26). Once more, it is observed that the spacer's thermal budget had no impact in the final active junction profile for the split with no PAI prior to LDD (PAI_(LDD)). Once again, since the only PAI step was performed AFTER the spacer's thermal budget (PAI_(HDD)), no amorphous region had been recrystallized before.

On the other hand, for the case with PAI_(LDD)=24nm, there is an important difference between the activation of the HDD dopants.

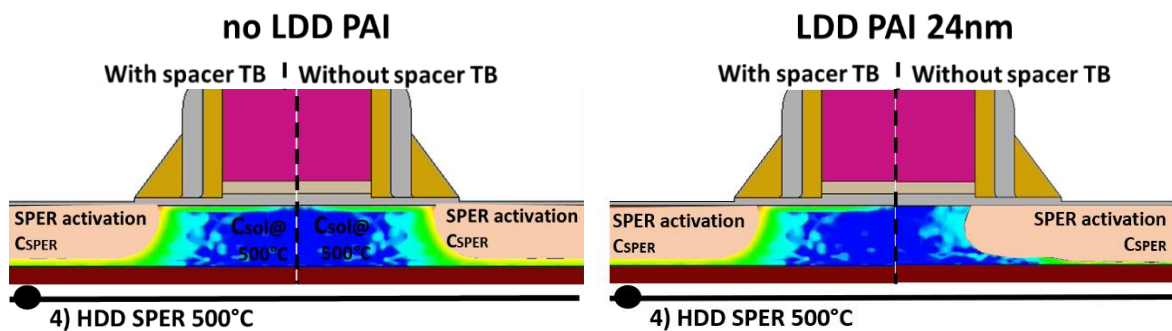


Figure 4.26 Step 4: SPER activation @500°C 30min.

4.6.2.3 Discussion on chemical and electrical profile

To properly address the impact of this additional thermal budget (from spacer formation) on the electrical performance of the device, one needs to observe the electrical dopant profile instead of the chemical one. This is because not necessarily all dopants will become active, as described on the previous sections. In order to ease the understanding, all splits shall be analyzed.

Fig. 4.27 depicts only the Phosphorus cases, and analyses the activated junction profile of both splits with or without $PAI_{(LDD)}$. In addition, the impact of the spacer's thermal budget shall also be highlighted.

Phosphorus no $LDD_{(PAI)}$ (Fig. 4.27-a) with the spacer's thermal budget: everything that is not in an amorphous region, shall be thermally activated @500°C. The range of this activation level is depicted by the blue dashed lines, where the C_{SOL} values are taken from table 1. Since Phosphorus has a relatively high C_{SOL} threshold (with a maximum of 8×10^{19} at/cm³), almost all the profile below the spacer was activated (red dashed line).

The amorphous-crystalline interface, created by the PAI prior to HDD ($PAI_{(HDD)}$), is also depicted (HDD A/C interface). From that point, SPER mechanism takes place, at the maximum activation level is now limited by C_{SPER} (green dashed line).

Phosphorus with no $LDD_{(PAI)}$ (Fig. 4.27-b) without the spacer's thermal budget: when comparing with the case with spacer's thermal budget it can be seen that there is no difference in the estimated activated profile. The dopants shall only be activated by SPER if within an amorphous template.

Phosphorus $LDD_{(PAI)}=24\text{nm}$ (Fig. 4.27-c) with the spacer's thermal budget: Because of the spacer's thermal budget, the region below the spacers gets recrystallized/activated at this moment by SPER. However, since Phosphorus has a relatively high C_{SOL} threshold (with a maximum of 8×10^{19} at/cm³), almost all the profile below the spacer was activated (red dashed line). At this stage, it is highlighted that the profile is equivalent to the one obtained with no $PAI_{(LDD)}$ (Fig. 4.27-a).

Phosphorus $LDD_{(PAI)}=24\text{nm}$ (Fig. 4.27-d) without the spacer's thermal budget: all the region gets activated by SPER, because there was no recrystallization/activation if there was no spacer's thermal budget. In this case, all the chemical profile below the spacer gets activated.

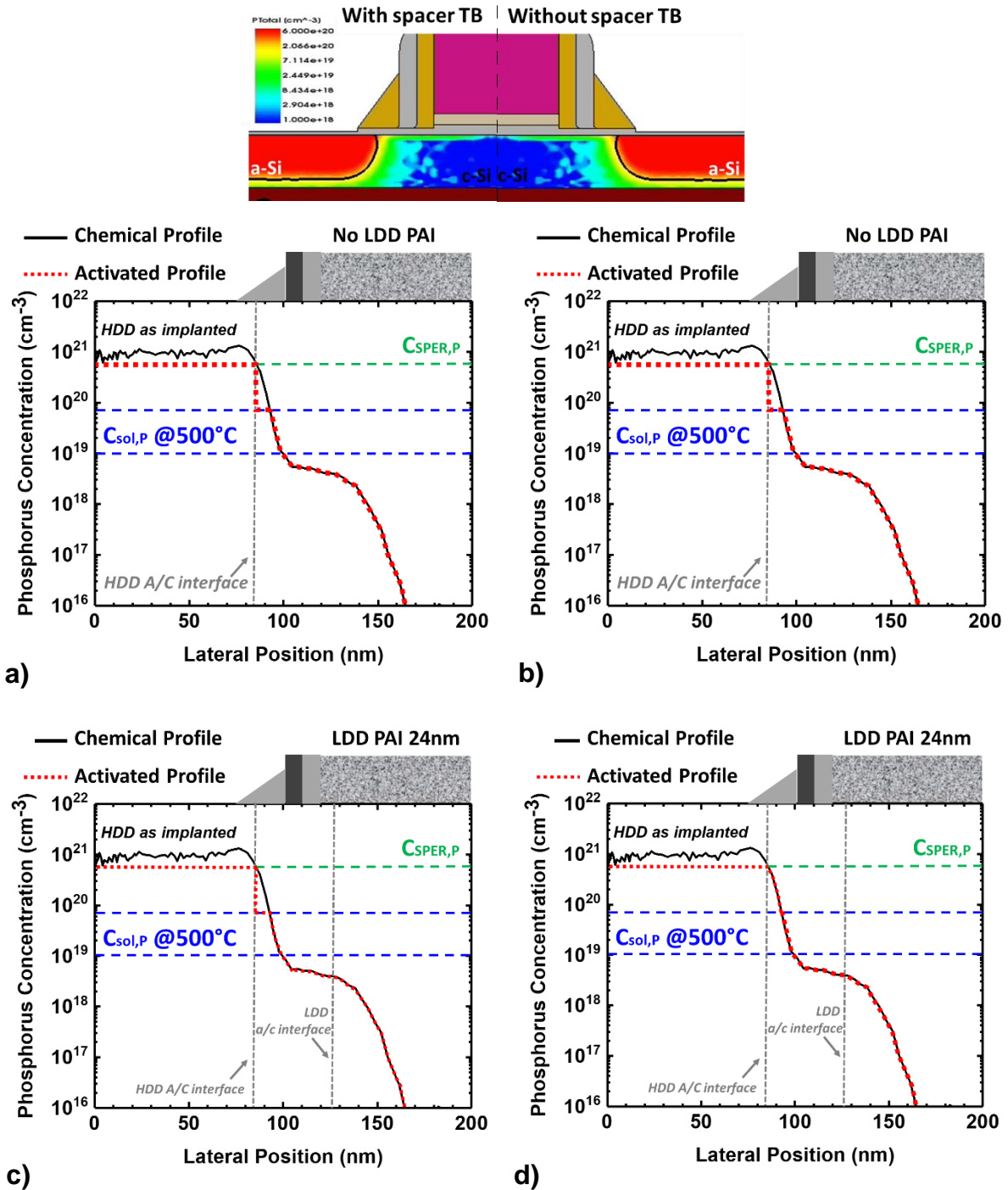


Figure 4.27 Estimated activated Phosphorus profile (red dashed line) as compared to its respective chemical concentration (black straight line), for the split with and without PAI_(LDD). The amorphous/crystalline interface (a/c) is also depicted (grey dashed line). Lastly, the activation thresholds of thermal activation at 500°C and SPER are displayed.

To conclude this section, it has been observed that by suppressing the PAI_(LDD), there is a small fraction ($\approx 4\text{nm}$) of dopants below the spacer which activation is limited by C_{SOL} @500°C. Furthermore, due to the spacer’s thermal budget, the case with / without PAI_(LDD) end up having the same activated profile.

The same analysis shall now be done for Boron. Fig. 4.28 depicts only the Boron cases, and analysis the activated junction profile of both splits with or without PAI_(LDD). In addition, the impact of the spacer's thermal budget shall also be highlighted.

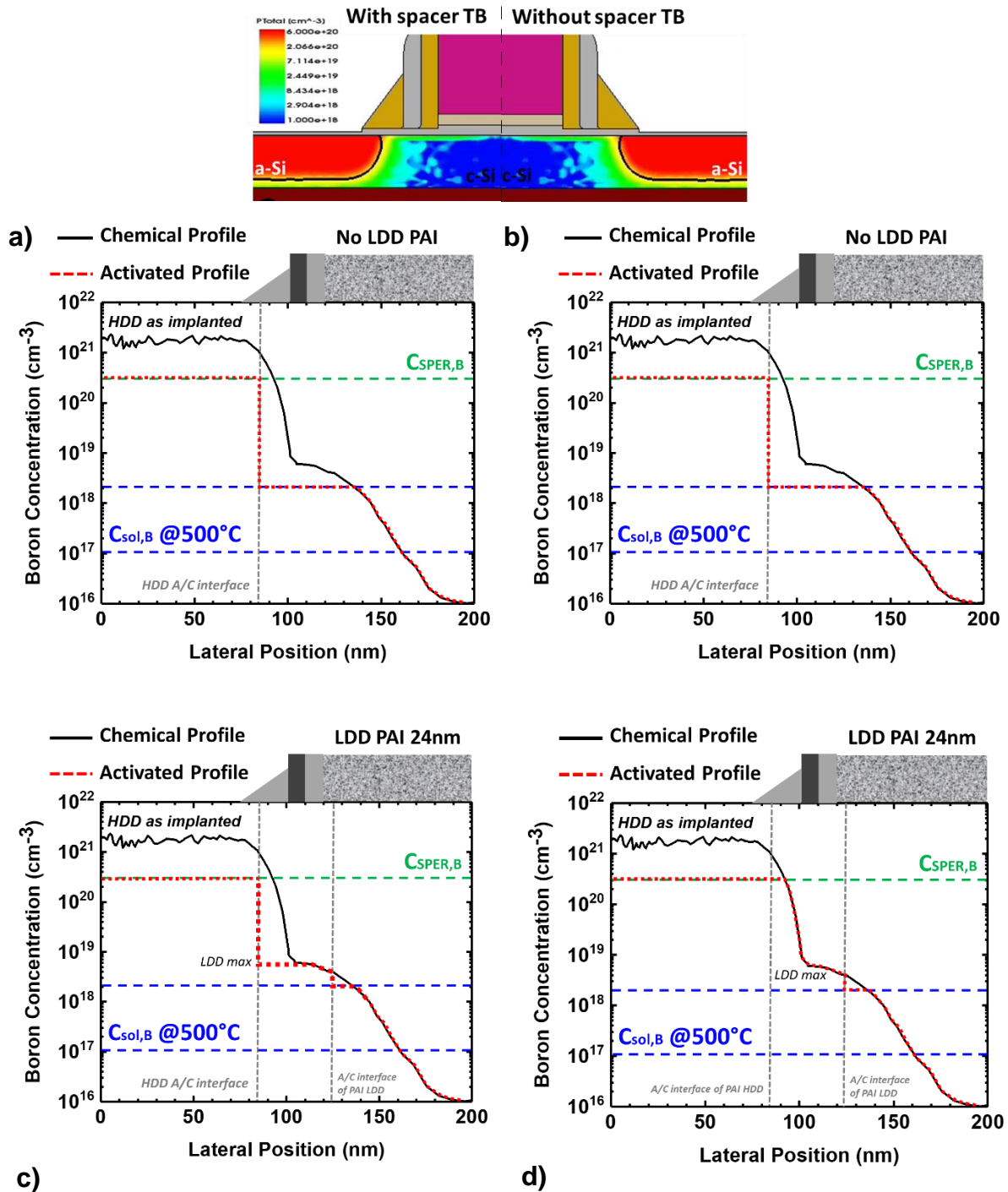


Figure 4.28 Estimated activated Boron's profile (red dashed line) as compared to its respective chemical concentration (black straight line), for the split with and without PAI_(LDD). The amorphous/crystalline interface (a/c) is also depicted (grey dashed line). Lastly, the activation thresholds of thermal activation at 500°C and SPER are displayed.

Boron with no LDD_(PAI) (Fig. 4.28-a)) with the spacer's thermal budget: once again, everything that is not in an amorphous region, shall be thermally activated @500°C. The range of this activation level is depicted by the blue dashed lines, where the C_{SOL} values are taken from table 1. It can be observed that contrarily to Phosphorus, Boron presents a relatively low C_{SOL} threshold (with a maximum of 2×10^{18} at/cm³). Consequently, there was a significant region below the spacer that was not properly activated. This is because if the region was not previously amorphous, the mechanism in place is thermal activation, where $C_{\text{SOL}} < C_{\text{SPER}}$. From the a-c interface (HDD A/C interface), the activation level thus increases, because now the activation is dominated by SPER. It must also be highlighted that the maximum of the HDD profile is 2×10^{21} at/cm³, whereas the maximum of C_{SPER} is 3×10^{20} at/cm³. This means that a fraction of the HDD dopants shall not be activated, because the SPER limit (C_{SPER}) has been reached.

Boron with no LDD_(PAI) (Fig. 4.28-b)) without the spacer's thermal budget: similarly to Phosphorus, when comparing with the case with spacer's thermal budget it can be seen that there is no difference in the estimated activated profile. The dopants shall only be activated by SPER if within an amorphous template.

Boron LDD_(PAI)=24nm (Fig. 4.28-c)) with the spacer's thermal budget: at the a-c interface created by the PAI_(LDD) (A/C interface of PAI LDD), the LDD dopants (delimited by LDD max) start to get activated by SPER at the spacer formation. As a result, afterwards, that zone is no longer amorphous. At the a-c interface created by the PAI_(HDD) (A/C interface of PAI HDD), the HDD dopants start to get activated by SPER. However, there is a share of the HDD dopants in the zone below the spacer that are limited by thermal activation @500°C, because of the previous recrystallization imposed by the spacer's formation. C_{SOL} of Boron being too low, results in a region where the dopants were not properly activated.

Boron LDD_(PAI)=24nm (Fig. 4.28-d)) without the spacer's thermal budget: basically all the region gets activated by SPER, because there was no recrystallization/activation if there was no spacer's thermal budget. Only a small fraction is limited by C_{SOL} , delimited by the a-c interface created by the PAI_(LDD). Despite, all the chemical profile below the spacer gets activated.

To conclude this section, it has been observed that impact of the spacer's thermal budget for the splits with PAI prior to LDD (PAI_(HDD)) is different depending on the dopant type. In the case of Phosphorus, thanks to its high C_{SOL} value, this effect is negligible. Contrarily,

in the case of Boron, almost all the dopants could have been activated if there was no additional thermal budget due to spacer's formation.

However, there is still one remaining question (when taking into account the spacer's thermal budget); it seems that the $PAI_{(LDD)}$ should yield a better activation than the split without. But how is it that this is not seen on the $I_{ON}-I_{OFF}$?

The latter can be explained by the C_{SOL} of Boron. As stated in the beginning of the chapter, there is a strong discrepancy between the extrapolated values @500°C. Furthermore, the final activation is strictly related to the number of defects. It can be said that the silicon-interstitial concentration in the region below the spacer can be particularly low. As a conclusion, it is suspected that superior activation levels could have been obtained ($\approx 1.10^{19}$ at/cm³ at 500°C) for thermal activation @500°C. As a result of the $C_{SOL,B}$ being higher than expected, it would explain why there is no difference in the case with or without $PAI_{(LDD)}$. The latter is summarized on Fig. 4.29, depicting the possible higher C_{SOL} threshold.

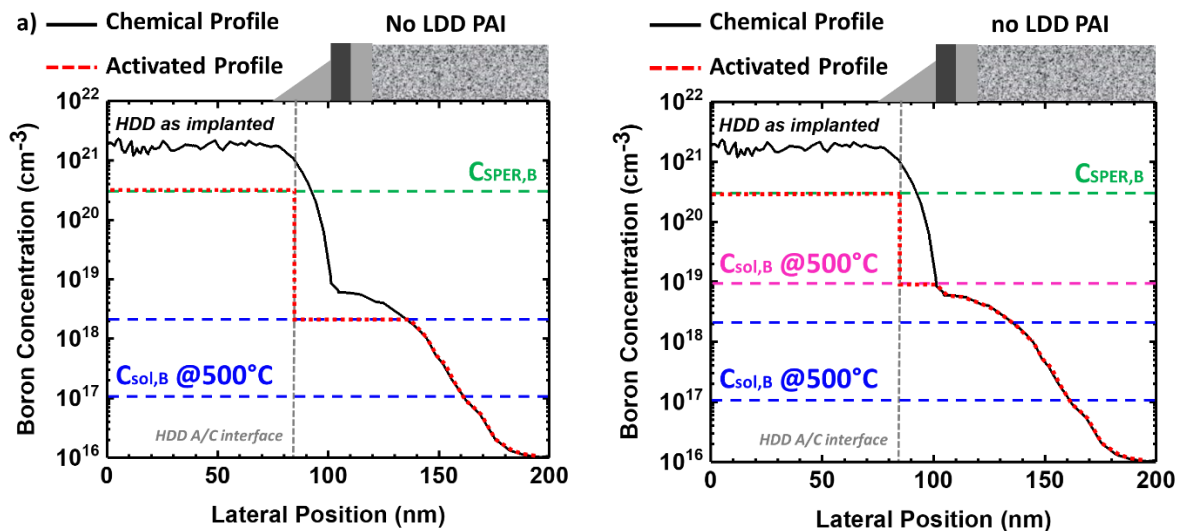


Figure 4.29 Estimated activated Phosphorus profile (red dashed line) as compared to its respective chemical concentration (black straight line), for the split without $PAI_{(LDD)}$. Lastly, the activation thresholds of thermal activation at 500°C and SPER are displayed.

Because of this higher $C_{SOL,B}$ ($\approx 1.10^{19}$ at/cm³ instead of 2.10^{18} at/cm³) @500°C, there is no difference in the profiles with or without PAI prior to LDD, hence explaining why both splits present the same performance.

4.6.2.4 Access resistance evaluation

In order to verify the aforementioned hypothesis, TCAD electrical simulation must be performed. This section shall focus on the Boron case only because it is the one suffering from performance degradation.

The estimated activated profile (considering the spacer's thermal budget) was reproduced on TCAD electrical simulation. The structures are depicted as followed: the chemical profile obtained from KMC (circle) was reproduced on TCAD electrical simulation (straight line). Likewise, the estimated activated profile was also reproduced. Additionally, the access resistance was extracted with total resistance method @ $V_{GT}=500\text{mV}$. The profiles are altogether shown with the extracted R_{SD} values.

Boron (Fig. 4.30) with no $PAI_{(LDD)}$: It can be seen that with this configuration, there is more than a factor 2 on the resistance values between the chemical and the activated profile.

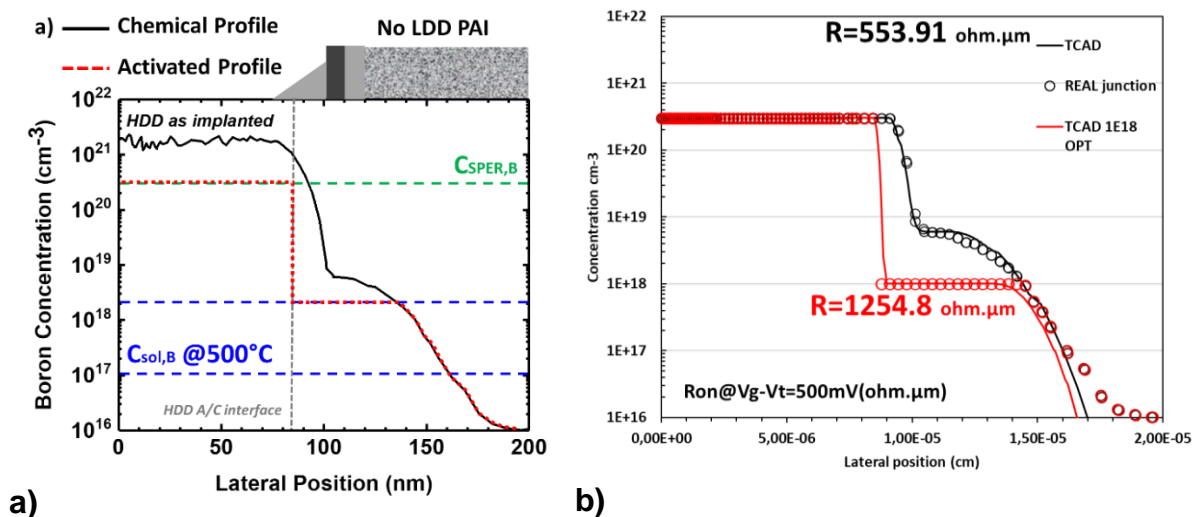


Figure 4.30-a) Estimated activated Boron profile (red dashed line) as compared to its respective chemical concentration (black straight line), for the split without $PAI_{(LDD)}$. In addition, the aforementioned profiles were replicated on TCAD electrical simulation (b) as follow: the straight black line represents the chemical concentration, fitted with TCAD by the black circles. Likewise, the estimated active profile is depicted as a straight red line, reproduced on TCAD electrical simulation as the red circles.

This would result in more than 30% degradation on the I_{ON} results, which only increases the suspicious that $C_{SOL,B}$ is actually higher than what was extrapolated.

Boron (Fig. 4.31) with $PAI_{(LDD)}=24\text{nm}$: by now analyzing the estimated profile with a $PAI_{(LDD)}$, it can be seen that there is also a degradation in R_{ON} , which this time seems more coherent with what has been observed on the I_{ON} - I_{OFF} trade-off. These simulations

highlight the importance of the activated dopants below the spacer region, allowing one to conclude that *the I_{ON} performance decrease is mainly attributed to R_{SPA} degradation.*

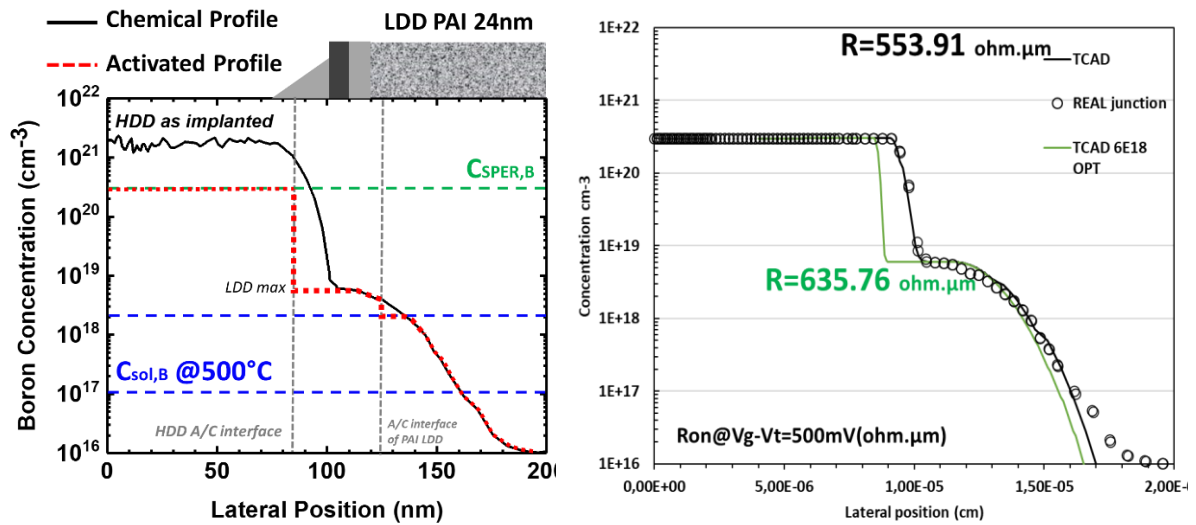


Figure 4.31-a) Estimated activated Boron profile (red dashed line) as compared to its respective chemical concentration (black straight line), for the split with $\text{PAI}_{(\text{LDD})}=24\text{nm}$. In addition, the aforementioned profiles were replicated on TCAD electrical simulation (b) as follow: the straight black line represents the chemical concentration, fitted with TCAD by the black circles. Likewise, the estimated active profile is depicted as a straight red line, reproduced on TCAD electrical simulation as the red circles.

4.6.3 Understanding I_{OFF} Results

The I_D - V_G curves for $W=1\mu\text{m}$ and $L=0.45\mu\text{m}$ is shown below for $V_D=V_{DD}=2.5\text{V}$ (Fig. 4.32). Even though the curves are shown for both devices, this section shall focus on understanding the leakage of PMOS, due to its superior degradation once a deep $\text{LDD}_{(\text{PAI})}$ is applied.

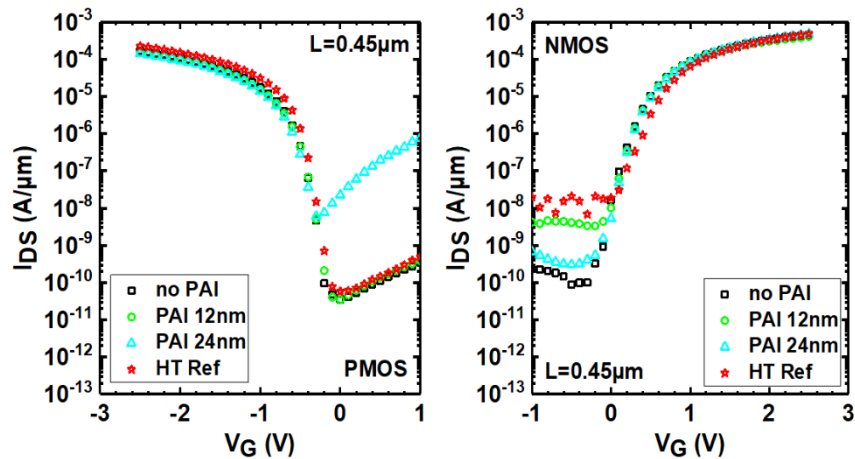


Figure 4.32 $I_D V_G$ curves of N&PMOS devices for $L_G=45\mu\text{m}$ for the different LDD splits and the high temperature process. Each curve represents the mean of a few dies.

2D profiles of interstitials trapped in $\{311\}$ defects at the end of the process (after SPER anneal) is shown on Fig. 4.33. It is observed that for the split without LDD_(PAI), there are no defects located near the gate edge, considered a critical region for GIDL [28]. In the case of a PAI prior to LDD, even though some defects are observed in this region, these are not as deep as when a LDD_(PAI)=24nm is used.

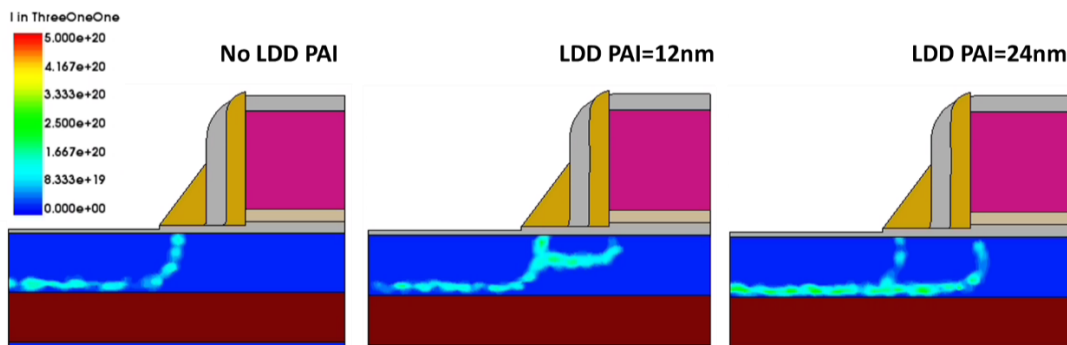


Figure 4.33 2D profiles of interstitials trapped in $\{311\}$ defects at the end of the process (after S/D activation).

In order to further investigate the leakage, Fig. 4.34 depicts the drain current extracted at different V_D at $V_G=-1\text{V}$ for different gate lengths. In general, the addition of defects induced by the LDD_(PAI)=24nm near the gate edge has increased the leakage current density, which agrees with the electrical results, and seems to be independent of gate length. However, this increase between no LDD_(PAI) and LDD_(PAI)=24nm appears to be greater at high E_{FIELD} ($V_D=2.5\text{V}$) where 3 decades increase has been observed, compared to low E_{FIELD} ($V_D=100\text{mV}$), presenting a 2 decades increase. This is because the thermal emission from

traps may be field dependent so the TAT contribution to the leakage current may change with bias [29].

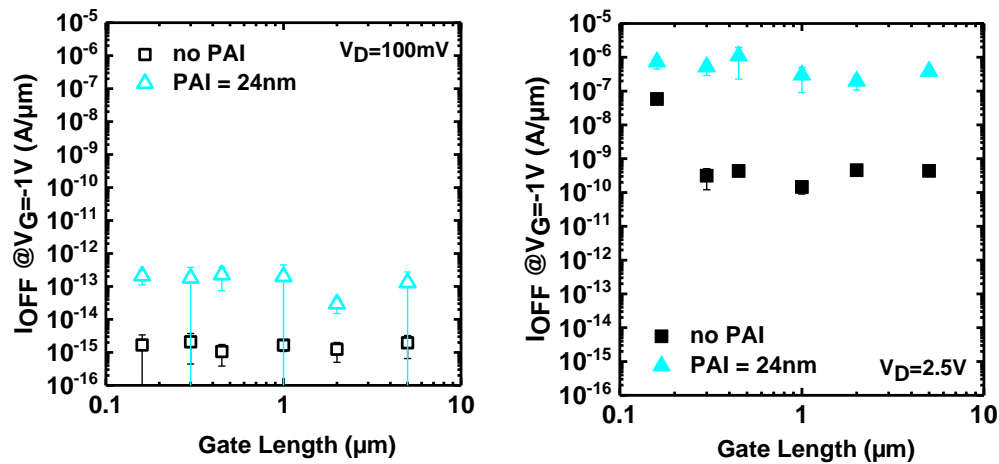


Figure 4.34 Drain current extracted at different V_D : empty symbols correspond to $V_D=100mV$, whereas filled symbols corresponds to $V_D=2.5V$ extracted at $V_G=-1V$.

It is still unclear why only PMOS devices experienced these elevated leakage values for $LDD_{(PAI)}=24nm$, given that EOR are not dependent on the implanted dopant (should be the same for Phosphorus and Boron). It could be said that given that N&P present different active profiles, their corresponding depletion region might not be the same (knowing that the defects located within the junction depletion are the ones that magnify leakage effects). However, in order to properly characterize the leakage of the devices, (1) more dies should be tested, and (2) further studies should be implemented also for different temperatures.

4.7 OPTIMIZATION GUIDELINES

Based on the electrical results, a schematic of the optimized junction process for high voltage ($V_{DD}=2.5V$) devices is depicted on Fig. 4.35, with the goal to have a closer activation as the one obtained from the high temperature devices. The following guidelines are proposed: (1) deep PAI prior to LDD should be suppressed to limit junction leakage; (2) the thermal budget of the spacers should be limited to $400^\circ C$ to prevent recrystallization under the spacer region (which is believed to be the source of the performance degradation), and (3) the HDD dose should be reduced to avoid B clustering for concentrations $>3.10^{20}at/cm^{-3}$.

In case 1, a PAI prior to LDD should be limited in depth (12nm for a $T_{Si}=30nm$) if performed at the gate edge. A second PAI prior do HDD is then performed, finishing by SPER activation @500°C 30min. As an alternative, a PAI prior to LDD could also be performed after the spacer 1 formation, thus shifting the junction position in 10nm (case 2). This variation is only interesting if proven that the I_{OFF} increase takes place at the gate edge. As such, a thick PAI could be used, thus suppressing the need of a double PAI implant. This could potentially lead to the creation of less Si interstitials.

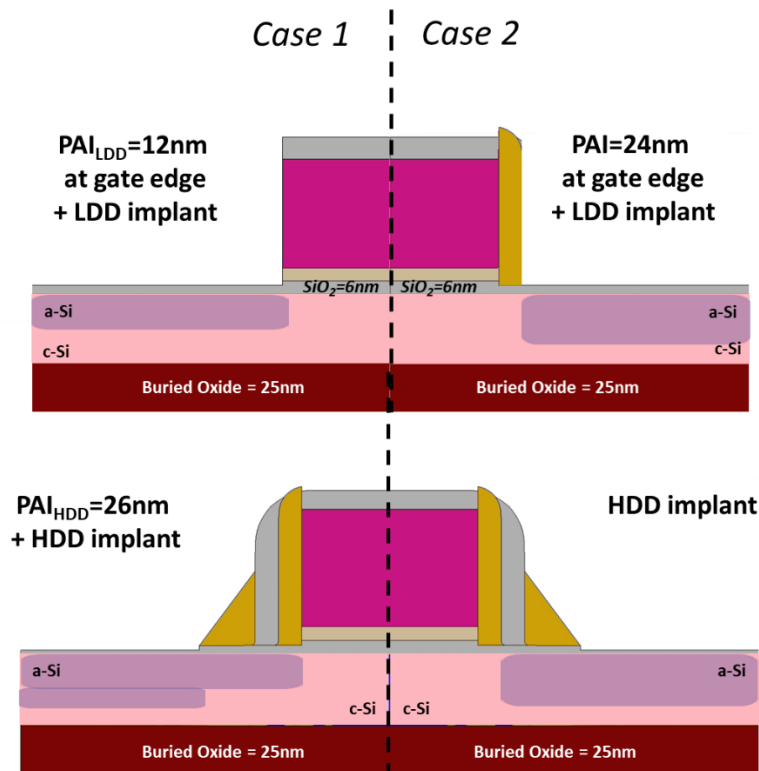


Figure 4.35 Schematic representation of the proposed optimization guidelines.

4.8 CONCLUSION OF THIS CHAPTER

In this chapter, a detailed electrical characterization, supported by TCAD process and electrical simulation results has been carried out on FDSOI devices activated at low temperature and fabricated on a thick Si channel for high voltage applications ($V_{DD}=2.5V$).

Due to the lower doping values of the LDD implants ($<1.10^{19}cm^{-3}$), two methods have been proposed to activate the junction profile, thermal activation (C_{SOL}) at $500^{\circ}C$ and SPER at $500^{\circ}C$ 30min. The main benefit from a thermal activation would be the suppression of the PAI step prior to LDD, potentially reducing junction leakage.

Despite the relaxed gate length of the device ($L_{G,NOM}=0.45\mu m$), an access resistance degradation has been observed, particularly important on PMOS devices (up to -40% on ion). By carefully analyzing the chemical and activated profiles, it has been noticed that an additional TB due to spacer formation recrystallized the Si region under the spacer. This meant that a region that should have been activated by SPER, ended up being activated by $C_{SOL}@500^{\circ}C$, which can be particularly low especially in the case of Boron. As such, the access resistance degradation was mainly explained to the difficulty to activate the dopants in the region below the spacer, leading to a degradation of the R_{SPA} resistance component.

Concerning the O_{FF} regime, an important leakage increase of 3 decades was highlighted on PMOS devices once a LDD PAI=24nm was used. Even though the leakage mechanisms need yet to be understood, it can be concluded that the LDD PAI should have a limited thickness.

Lastly, guidelines have been provided to ensure the best I_{ON}/I_{OFF} performance for the junction of high voltage devices.

- [1] P. Pichler, *Intrinsic Point Defects, Impurities, and Their Diffusion in Silicon*. Vienna: Springer Vienna, 2004.
- [2] P. Pichler, *Intrinsic Point Defects, Impurities, and Their Diffusion in Silicon*. Vienna: Springer Vienna, 2004.
- [3] G. L. Vick et K. M. Whittle, « Solid solubility and diffusion coefficients of boron in silicon », *J. Electrochem. Soc.*, vol. 116, n° 8, p. 1142-1144, 1969.
- [4] B. J. Pawlak *et al.*, « Chemical and electrical dopants profile evolution during solid phase epitaxial regrowth », *J. Vac. Sci. Technol. B Microelectron. Nanometer Struct.*, vol. 22, n° 1, p. 297, 2004, doi: 10.1116/1.1643053.
- [5] B. Colombeau *et al.*, « Electrical deactivation and diffusion of boron in preamorphized ultrashallow junctions: Interstitial transport and F co-implant control », in *IEDM Technical Digest. IEEE International Electron Devices Meeting, 2004.*, 2004, p. 971-974.
- [6] M. Aboy, L. Pelaz, P. López, L. A. Marqués, R. Duffy, et V. C. Venezia, « Physical insight into boron activation and redistribution during annealing after low-temperature solid phase epitaxial regrowth », *Appl. Phys. Lett.*, vol. 88, n° 19, p. 191917, mai 2006, doi: 10.1063/1.2203334.
- [7] S. F. Guo, « A simple model for the deposition of boron in silicon by using a BN diffusion source », *J. Electrochem. Soc.*, vol. 127, n° 11, p. 2506, 1980.
- [8] J. S. Williams, « Solid phase epitaxial regrowth phenomena in silicon », *Nucl. Instrum. Methods Phys. Res.*, vol. 209, p. 219-228, 1983.
- [9] B. Sklénard, « Physical modeling of junction processing in FDSOI devices for 20 nm node and below », Université Grenoble Alpes, 2014.
- [10] R. Duffy, T. Dao, Y. Tamminga, K. van der Tak, F. Roozeboom, et E. Augendre, « Groups III and V impurity solubilities in silicon due to laser, flash, and solid-phase-epitaxial-regrowth anneals », *Appl. Phys. Lett.*, vol. 89, n° 7, p. 071915, août 2006, doi: 10.1063/1.2337081.
- [11] L. Pasini, « Optimisation des jonctions de dispositifs (FDSOI, TriGate) fabriqués à faible température pour l'intégration 3D séquentielle », Grenoble Alpes, 2016.
- [12] A. Claverie *et al.*, « Thermal evolution of extended defects in implanted Si:: impact on dopant diffusion », *Mater. Sci. Semicond. Process.*, vol. 3, n° 4, p. 269-277, 2000.
- [13] C. Xu, « Optimisation du procédé de réalisation pour l'intégration séquentielle 3D des transistors CMOS FDSOI », PhD Thesis, Grenoble, 2012.
- [14] F. Cristiano *et al.*, « Defects evolution and dopant activation anomalies in ion implanted silicon », *Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. At.*, vol. 253, n° 1-2, p. 68-79, déc. 2006, doi: 10.1016/j.nimb.2006.10.046.
- [15] S. Mirabella *et al.*, « Dissolution kinetics of boron-interstitial clusters in silicon », *Appl. Phys. Lett.*, vol. 83, n° 4, p. 680-682, juill. 2003, doi: 10.1063/1.1594264.
- [16] F. Severac *et al.*, « Influence of boron-interstitials clusters on hole mobility degradation in high dose boron-implanted ultrashallow junctions », *J. Appl. Phys.*, vol. 107, n° 12, p. 123711, juin 2010, doi: 10.1063/1.3446844.
- [17] A. T. Fiory *et al.*, « Transient-enhanced diffusion in shallow-junction formation », *J. Electron. Mater.*, vol. 31, n° 10, p. 999-1003, 2002.
- [18] J. Micout, « Fabrication et caractérisation de transistor réalisée à basse température pour l'intégration 3D séquentielle », PhD Thesis, Grenoble Alpes, 2019.
- [19] R. Duffy *et al.*, « Quantitative prediction of junction leakage in bulk-technology CMOS devices », *Solid-State Electron.*, vol. 54, n° 3, p. 243-251, mars 2010, doi: 10.1016/j.sse.2009.09.007.

- [20] B. El-Kareh et L. N. Hutter, *Silicon Analog Components*. New York, NY: Springer New York, 2015.
- [21] M. Aboy *et al.*, « Role of silicon interstitials in boron cluster dissolution », *Appl. Phys. Lett.*, vol. 86, n° 3, p. 031908, janv. 2005, doi: 10.1063/1.1852728.
- [22] B. Sklenard *et al.*, « Influence of device architecture on junction leakage in low-temperature process FDSOI MOSFETs », *Solid-State Electron.*, vol. 88, p. 9-14, oct. 2013, doi: 10.1016/j.sse.2013.04.018.
- [23] P. Batude *et al.*, « 3D Sequential Integration: Application-driven technological achievements and guidelines », in *2017 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, déc. 2017, p. 3.1.1-3.1.4, doi: 10.1109/IEDM.2017.8268316.
- [24] B. Sklenard *et al.*, « Low temperature junction formation by solid phase epitaxy on thin film devices: Atomistic modeling and experimental achievements », in *Junction Technology (IWJT), 2014 International Workshop on*, 2014, p. 1-6.
- [25] L. Pasini *et al.*, « High performance low temperature activated devices and optimization guidelines for 3D VLSI integration of FD, TriGate, FinFET on insulator », in *2015 Symposium on VLSI Technology (VLSI Technology)*, 2015, p. T50-T51.
- [26] F. Monsieur *et al.*, « The importance of the spacer region to explain short channels mobility collapse in 28nm Bulk and FDSOI technologies », in *Solid State Device Research Conference (ESSDERC), 2014 44th European*, 2014, p. 254-257.
- [27] G. Ghibaudo, « New method for the extraction of MOSFET parameters », *Electron. Lett.*, vol. 24, n° 9, p. 543-545, 1988.
- [28] J.-H. Chen, S.-C. Wong, et Y.-H. Wang, « An analytic three-terminal band-to-band tunneling model on GIDL in MOSFET », *IEEE Trans. Electron Devices*, vol. 48, n° 7, p. 1400-1405, 2001.
- [29] R. Duffy et A. Heringa, « Characterisation of electrically active defects: Characterisation of electrically active defects », *Phys. Status Solidi C*, vol. 11, n° 1, p. 130-137, janv. 2014, doi: 10.1002/pssc.201300144.

Conclusion and Perspectives

3D Sequential Integration consists in stacking active device layers on top of each other, in a sequential manner. It offers a unique opportunity to interconnect devices at the transistor level owing to the feasibility of an ultra-small 3D contact pitch ($<100\text{nm}$). Although the initial motivation behind 3D sequential integration was to further Moore's Law, it has also been identified as a key technology for More than Moore applications. For instance, partitioning analog and digital devices on separate levels offers the opportunity of independent tuning of each layer. In this respect, analog devices are needed and need to be fabricated at a low thermal budget, in order to prevent degradation of the underlying tiers. Thus, **chapter 1** provided an overview of the challenges and motivations for a 3D sequential integration. In addition, the proposed process flow for the analog high voltage devices was also presented.

However, before going into the device fabrication, the thermal budget processing window of the top tier had to be defined. As a result, an evaluation of the thermal stability of the bottom tiers (device and interconnects) was presented on **chapter 2**. In order to have flexibility in the partitioning and to provide a complete stability window, different devices were analyzed (28nm FDSOI and BULK), as well as the stability under different types of post-thermal treatments (standard and ultra-short anneal). In addition, the stability of the State-of-the-Art Cu/ULK interconnects was also addressed.

The thermal stability of 28nm devices has been evaluated through the device performance under DC/AC operation and the yield (SRAM cuts and logic gate array) has been set at 500°C 2h. For higher thermal budgets, a frequency degradation was observed, attributed to a dopant deactivation of the S/D in the NMOS device. Furthermore, Cu/ULK interconnects were also evaluated through reliability and yield measurements. The reliability has shown to be stable even up to 525°C 2h, due to a more mature BEOL process. However, in order to keep the integrity of the yield, the thermal stability was set to 500°C 2h. These results pave the way to the introduction of BEOL between tiers in 3D sequential integration.

In addition, the immunity of bottom device and interconnects upon an ultraviolet nanosecond laser annealing was investigated. It was then demonstrated that even for a thin upper active Si layer ($\sim 7\text{nm}$), the Cu resistivity and $I_{\text{ON}}-I_{\text{OFF}}$ remains stable. Therefore, this secured the conclusion that laser energy conditions including melt can be successfully reached without affecting State-Of-The-Art Cu lines for 28nm FDSOI devices.

The thermal stability of high voltage ($V_{\text{DD}}=5\text{V}$) BULK devices was also quantified. The static figure of merit proved unchanged up to 600°C 2h. A slight polysilicon deactivation was observed for the highest thermal budget (600°C 2h), only affecting the gate work-function. However, this effect can be compensated by adjusting the channel dopant concentration. The silicide stability on the source and drain was shown to be stable, the only weak point being the silicide instability of the polysilicon gate starting from 550°C post anneals. Therefore, the thermal budget must be kept below 550°C 2h in order to ensure the gate resistance integrity. These results showed that high voltage SiO_2/Poly devices can be integrated on the bottom level of 3D Sequential Integration with a comfortable thermal budget (at least 500°C 5h) to process the top device.

The thermal budget processing window having been defined at 500°C , the goal was now to fabricate a high voltage ($V_{\text{DD}}=2.5\text{V}$) analog device. This work was centered on two technological bricks that needed to be adapted in order to fulfill the analog's requirements, those being the gate stack and the junction.

Chapter 3 covered the low temperature gate stack for a high voltage ($V_{\text{DD}}=2.5\text{V}$) analog device, where reliable and low noise stacks were envisaged. The gate stack composed of a thick $\text{SiO}_2=6\text{nm}$ was obtained through different low temperature oxide splits and post-deposition treatments (in the 500°C range). An innovative method was proposed, called "OX from BOX", was also proposed and evaluated. Thanks to in-depth characterization, the gate stack features (low field mobility, interface states, noise and reliability) were extensively studied. Additionally, an identification of the defects found in low temperature SiO_2 oxides were provided, with 3 types of traps being identified (deep, border and interface defects). For the NMOS devices, there was a visible effect of the different oxide splits, with the hydrogenated plasma split $\text{Pl-O}_2/\text{H}_2$ showing the best results. Additionally, it was demonstrated that NMOS devices are reliable and ideal for readout circuits with NFETs only. Regarding PMOS devices, a less noticeable effect was

observed for the oxide splits, and the post deposition treatments did not effectively improve its oxide features. Additionally, the best low temperature oxide (hydrogenated plasma PI-O₂/H₂) showed good performance on some analog FOM (gm/Id and gm/gd)

The other proposed oxide solution, the “OX from BOX”, has also been degraded in comparison to the high temperature process, probably due to the chemical treatment used to thin the oxide. Additionally, the post deposition plasma treatment does not seem enough to recover the oxide integrity. However, only a full integration on an electrical lot can determine whether the “OX from BOX” is truly a viable solution or not.

Chapter 4 focused on the low temperature junction dopant activation, aiming to reproduce the same characteristics of the high temperature junction, but on a limited thermal budget. Due to the lower doping values of the LDD implants ($<1.10^{19}\text{cm}^{-3}$), two methods were proposed to activate the junction: thermal activation (C_{SOL}) at 500°C and SPER at 500°C 30min. An access resistance degradation was observed, especially in the case of PMOS devices (up to -40% on I_{ON}). This degradation was mainly due to the difficulty to activate the dopants in the region below the spacer (R_{SPA}), leading to a degradation of the R_{SPA} resistance component. Concerning the I_{OFF} regime, an important leakage increase of 3 decades was highlighted on PMOS devices once a LDD PAI=24nm was used. Although further studies are required for a better understanding of the leakage mechanism, it can be concluded that the LDD PAI should have a limited thickness.

Overall, the results presented on this thesis proves the feasibility of stacking high voltage analog devices in a 3D Sequential Integration for More than Moore applications.

Eventually, several perspectives remain to be explored and should be pending investigation. First, in order to extend the studies concerning the thermal stability (chapter 2) of the interconnects under standard annealing, electromigration tests are encouraged. Although the Cu resistivity seems stable under ultra-short anneals, reliability tests should also be performed.

Regarding the low temperature gate stack (chapter 3), different thermal treatments are proposed in order to improve the quality of the stack, such as laser annealing and atomic hydrogen treatment. Additionally, SiGe channel QW and Junction-less transistors are also viable options for NBTI improvement.

Lastly, the low temperature junction (chapter 4) can be further optimized by reducing the thermal budget of the spacer deposition. The PAI before LDD implantation should either be suppressed or have a limited thickness, thus limiting the junction leakage.

LIST OF PUBLICATIONS

C. Cavalcante et al., "Bottom Tier High Voltage Device Thermal Stability in 3D Sequential Integration for More than Moore Applications", SSDM'2019, Nagoya, Japan, Sept. 2019

C. Cavalcante et al., "Low temperature high voltage analog devices in a 3D sequential integration," 2020 International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA), Hsinchu, Taiwan, 2020, pp. 155-156, doi: 10.1109/VLSI-TSA48913.2020.9203691.

C. Cavalcante et al., "28nm FDSOI CMOS Technology (FEOL and BEOL) Thermal Stability for 3D Sequential Integration: Yield and Reliability Analysis," 2020 IEEE Symposium on VLSI Technology, Honolulu, HI, USA, 2020, pp. 1-2, doi: 10.1109/VLSITechnology18217.2020.9265075.

C. Cavalcante et al., "3D Sequential Integration : towards analog on digital stacking", 2020 ESSDERC Dissemination Workshop

Study, fabrication and characterization for 3D Sequential Integration

Abstract

3D Sequential Integration consists in stacking active device layers on top of each other, in a sequential manner, where its main challenge relies in processing devices at low thermal budget in order to prevent degradation of the bottom layers. This integration has been identified as a key technology for More than Moore applications. In this respect, analog devices are needed and need to be fabricated at a low thermal budget. An evaluation of the thermal stability of the bottom tiers was performed. The thermal stability of FD-SOI 28nm devices was evaluated through the device performance under DC/AC operation and the yield and set at 500°C 2h. Cu/ULK interconnects were also evaluated and set to 500°C 2h. The thermal stability of high voltage ($V_{DD}=5V$) BULK devices was also quantified, and this transistor can be integrated on the bottom level of 3D Sequential Integration with a comfortable thermal budget (at least 500°C 5h). Concerning the device fabrication, two technological bricks needed to be adapted in order to fulfill the analog's requirements, those being the gate stack and the junction. The gate stack composed of a thick $SiO_2=6nm$ was obtained through different low temperature oxide splits and post-deposition treatments (in the 500°C range). Thanks to in-depth characterization, the gate stack features were extensively studied. NMOS devices are reliable and ideal for readout circuits with NFETs only. Additionally, an identification of the defects found in low temperature SiO_2 oxides were provided. The best low temperature oxide (hydrogenated plasma $Pl-O_2/H_2$) showed good performance on some analog FOM (gm/Id and gm/gd). The junction dopant activation by solid phase epitaxial regrowth (SPER) and thermal activation, both at 500°C, was studied through electrical characterization and TCAD simulation. Different splits focusing on the pre-amorphization step were tested. It was shown that by suppressing the pre-amorphization before LDD implantation and decreasing the spacer's thermal budget, performance and junction leakage should be optimized. The results presented on this thesis proved the feasibility of stacking high voltage analog devices in a 3D Sequential Integration for More than Moore applications.

Étude, fabrication and caractérisation pour l'intégration 3D séquentielle

Abstract

L'intégration 3D séquentielle consiste à empiler des couches de dispositifs les unes sur les autres, d'une manière séquentielle. Le défi principal est de fabriquer des transistors à basse température afin d'éviter la dégradation des couches inférieures. Cette intégration a été identifiée comme une technologie clé des applications More than Moore. A cet égard, des transistors analogiques sont requis et doivent être fabriqués à basse température. Dans ce travail, une évaluation de la stabilité thermique des couches inférieures a été réalisée. La stabilité thermique des transistors FD-SOI 28nm a été évaluée sous l'opération AC/DC et le rendement est fixé à 500° 2h. Les interconnexions Cu/ULK ont aussi été évaluées et fixées à 500° 2h. La stabilité thermique des transistors haute tension BULK ($V_{DD}=5V$) a également été quantifiée. Ceux-ci peuvent être intégrés dans la couche inférieure d'une intégration 3D séquentielle avec un budget thermique confortable (500° 5h). Concernant la fabrication des transistors, il a fallu adapter deux briques technologiques afin de remplir les réquis analogiques, à savoir l'empilement de grilles et la jonction. Un empilement de grilles composé d'un épais $SiO_2=6nm$ a été obtenu via différents *splits* d'oxyde basse température et des traitements de densification (500°). Les caractéristiques de l'empilement de grilles ont été étudiées grâce à la caractérisation électrique. Les transistors NMOS sont fiables et adaptés pour des circuits *readout* avec seulement des NFETs. En outre, les défauts trouvés dans les SiO_2 basse température ont été identifiés. Le meilleur oxyde basse température (*hydrogenated plasma* $Pl-O_2/H_2$) a fait preuve d'une bonne performance sur certaines Figures de Mérite Analogiques (gm/Id et gm/gd). L'activation des dopants via la recristallisation épitaxiale en phase solide (SPER) et l'activation thermique, tous deux à 500°, a été étudiée par caractérisation électrique et simulation TCAD. Différents *splits* focalisés sur la pré-amorphisation ont été testés. Il a été établi qu'en supprimant la pré-amorphisation avant l'implantation LDD et en diminuant le budget thermique de l'espaceur, la performance et la fuite de jonction seraient optimisés. Les résultats présentés dans cette thèse démontrent qu'il est possible d'empiler des transistors haute ($V_{DD}=2.5V$) tension analogiques dans une intégration 3D séquentielle pour les applications More than Moore.