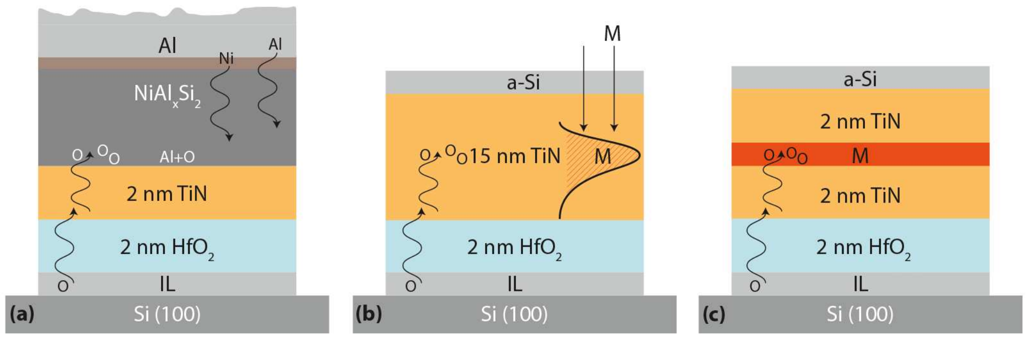


Information / Information
Band / Volume 28
ISBN 978-3-89336-898-3

JÜLICH
FORSCHUNGSZENTRUM

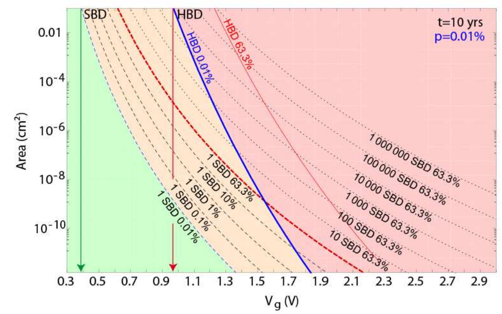
HfO$_2$ and LaLuO$_3$ high-κ/metal gate stacks for CMOS applications

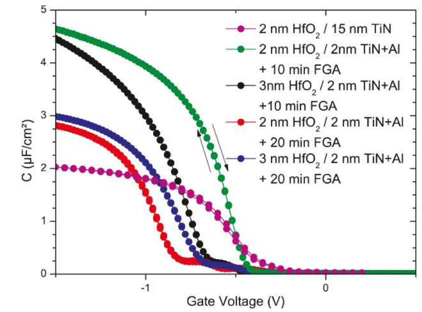
Alexander Nichau

Information

28

Member of the Helmholtz Association





Characterization, integration and reliability of HfO$_2$ and LaLuO$_3$ high-κ/metal gate stacks for CMOS applications

Alexander Nichau

JÜLICH
FORSCHUNGSZENTRUM

Schriften des Forschungszentrums Jülich
Reihe Information / Information Band / Volume 28

Forschungszentrum Jülich GmbH
Peter Grünberg Institute (PGI)
Semiconductor Nanoelectronics (PGI-9)

Characterization, integration and reliability of HfO_2 and $LaLuO_3$ high-κ/metal gate stacks for CMOS applications

Alexander Nichau

Schriften des Forschungszentrums Jülich
Reihe Information / Information Band / Volume 28

ISSN 1866-1777 ISBN 978-3-89336-898-3

Bibliographic information published by the Deutsche Nationalbibliothek.
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Publisher and Distributor:	Forschungszentrum Jülich GmbH Zentralbibliothek 52425 Jülich Tel: +49 2461 61-5368 Fax: +49 2461 61-6103 Email: zb-publikation@fz-juelich.de www.fz-juelich.de/zb
Cover Design:	Grafische Medien, Forschungszentrum Jülich GmbH
Printer:	Grafische Medien, Forschungszentrum Jülich GmbH

Schriften des Forschungszentrums Jülich
Reihe Information / Information, Band / Volume 28

D 82 (Diss., RWTH Aachen, University, 2013)

ISSN 1866-1777
ISBN 978-3-89336-898-3

The complete volume is freely available on the Internet on the Jülicher Open Access Server (JUWEL) at www.fz-juelich.de/zb/juwel

Abstract

The continued downscaling of MOSFET dimensions requires an equivalent oxide thickness (EOT) of the gate stack below 1 nm. An EOT below 1.4 nm is hereby enabled by the use of high-κ/metal gate stacks. $LaLuO_3$ and HfO_2 are investigated as two different high-κ oxides on silicon in conjunction with TiN as the metal electrode.

$LaLuO_3$ and its temperature-dependent silicate formation are characterized by hard X-ray photoemission spectroscopy (HAXPES). The effective attenuation length of $LaLuO_3$ is determined between 7 and 13 keV to enable future interface and diffusion studies. In a first investigation of $LaLuO_3$ on germanium, germanate formation is shown.

$LaLuO_3$ is further integrated in a high-temperature MOSFET process flow with varying thermal treatment. The devices feature drive currents up to $70\,\mu A/\mu m$ at $1\,\mu m$ gate length. Several optimization steps are presented. The effective device mobility is related to silicate formation and thermal budget. At high temperature, the silicate formation leads to mobility degradationdue to La-rich silicate formation. The integration of $LaLuO_3$ in high-T processes delicately connects with the optimization of the TiN metal electrode. Hereby, stoichiometric TiN yields the best results in terms of thermal stability with respect to Si-capping and high-κ oxide.

Different approaches are presented for a further EOT reduction with $LaLuO_3$ and HfO_2. Thereby the thermodynamic and kinetic predictions are employed to estimate the behavior on the nanoscale. Based on thermodynamics, excess oxygen in the gate stack, especially in oxidized metal electrodes, is identified to prevent EOT scaling below 1.2 nm. The equivalent oxide thickness of HfO_2 gate stacks is scalable below 1 nm by the use of thinned interfacial SiO_2. The prevention of oxygen incorporation into the metal electrode by Si-capping maintains the EOT after high temperature annealing. Redox systems are employed within the gate electrode to decrease the EOT of HfO_2 gate stacks. A lower limit found was EOT = 5 Å for Al doping inside TiN. The doping of TiN on $LaLuO_3$ is proven by electron energy loss spectroscopy (EELS) studies to modify the interfacial silicate layer to La-rich silicates or even reduce the layer.

The oxide quality in Si/HfO_2/TiN gate stacks is characterized by charge pumping and carrier mobility measurements on 3d MOSFETs a.k.a. FinFETs. The oxide quality in terms of the number of interface (and oxide) traps on top- and sidewall of FinFETs is compared for three different annealing processes. A high temperature anneal of HfO_2 improves significantly the oxide quality and mobility. The gate oxide integrity (GOI) of gate stacks below 1 nm EOT is determined by time-dependent dielectric breakdown (TDDB) measurements on FinFETs with HfO_2/TiN gate stacks. A successful EOT scaling has always to consider the oxide quality and resulting reliability. Degraded oxide quality leads to mobility degradation and earlier soft-breakdown, i.e. leakage current increase.

KURZFASSUNG

Die fortschreitende Skalierung der Dimensionen des Feldeffekttransistors (MOSFET) benötigt eine äquivalente Oxiddicke (EOT) des Gatters unterhalb eines Nanometers. EOT-Werte unterhalb von 1,4 nm können hierbei durch die kombinierte Verwendung von Oxiden höherer relativer Permittivität und Metallelektroden für das Gatter erreicht werden (engl. “high-κ/metal gate stacks”). Unter Verwendung von TiN als Metallelektrode werden $LaLuO_3$ und HfO_2 als zwei unterschiedliche Oxide hoher Permittivität untersucht.

$LaLuO_3$und dessen temperaturabhängige Silikatbildung werden mittels harter Röntgenstrahlung und Fotoemissionsspektroskopie (HAXPES) charakterisiert. Die effektive Abschirmlänge von $LaLuO_3$ wurde dabei zwischen 7 und 13 keV bestimmt um zukünftige Studien der Grenzschicht und des Diffusionsverhaltens zu ermöglichen. In einer ersten Untersuchung von $LaLuO_3$ auf Germanium wird eine Germanatbildung gezeigt.

$LaLuO_3$ wurde des Weiteren erfolgreich in einen Hochtemperatur-fertigungsprozess für MOSFETs integriert. Die hergestellten Bauelemente weisen Anströme bis zu $70\,\mu A/\mu m$ bei $1\,\mu m$ Gatterlänge auf. Hierzu werden mehrere Optimierungsschritte präsentiert. Die effektive Mobilität der Bauelemente kann dabei mit der Silikatbildung und dem thermischen Prozessbudget erklärt werden. Bei hoher Temperatur führt die Silikatbildung zu einer Mobilitätsabnahme, welche durch La-reiche Silikatbildung erklärt wird. In diesem Hochtemperaturprozess ist die Integration von $LaLuO_3$ eng mit der Optimierung der Metallelektrode TiN verbunden. Hierbei zeigt stöchiometrisches TiN die höchste thermische Stabilität in Bezug auf die Verwendung einer Si-Schutzschicht und high-κ Oxiden.

Verschiedene Ansätze für die EOT-Reduktion mittels $LaLuO_3$ und HfO_2 werden aufgezeigt. Thermodynamische und kinetische Berechnungen werden dabei für die Abschätzung des Verhaltens auf der Nanoskala angewendet. Den thermodynamischen Erwägungen zufolge verhindert übermäßiger Sauerstoff innerhalb der Gatterschichten die EOT-Skalierung unterhalb von 1,2 nm. Durch die Verwendung einer dünneren SiO_2-Schicht ist die äquivalente Oxiddicke von HfO_2 bis unter 1 nm skalierbar. Die Abdeckung der Metallelektrode mittels Siliziums ermöglicht gleiche EOT-Werte auch nach Heizen auf hohe Temperatur. Die weitere Skalierung von HfO_2-basierten Gatterschichten erlauben Redox-Systeme, wobei das EOT mittels Al dotiertem TiN bis unterhalb 5 Å skaliert wurde. Die Dotierung des TiN in Verbindung mit $LaLuO_3$ führt zu einer Modifikation oder Reduktion der Silikatgrenzschicht zwischen Si-Substrat und Oxid. Dieses Verhalten wird mittels Elektronenenergieverlustspektroskopie gezeigt.

Die Oxidqualität in Si/HfO_2/TiN Gatterschichten wurde mittels Ladungspumpen und Mobilitätsmessung der Ladungsträger auf 3d-Feldeffekttransistoren ("FinFETs") bestimmt. Die Oxidqualität als Anzahl der Oxidund Grenzschichtdefekte wird dabei auf der Oberfläche und der Seitenflanke für drei verschiedene Temperaturprozesse verglichen. Dabei wird gezeigt, dass ein Hochtemperaturheizen des HfO_2 signifikant die Oxidqualität und die Mobilität verbessert. Die Oxid-Zuverlässigkeit für Gatterschichten unterhalb von 1 nm EOT werden mithilfe der Messung des zeitabhängigen dielektrischen Zusammenbruchs von FinFETs mit HfO_2/TiN Gatterschichten bestimmt.

Die erfolgreiche Skalierung des EOT muss immer die Oxidqualität und die resultierende Zuverlässigkeit berücksichtigen. Eine geringere Oxidqualität führt hierbei zu einer Mobilitätsabnahme und früherem Anstieg des Gatter-Leckstroms.

Acknowledgements

It is my pleasure collaborating in such strong research environments like Forschungszentrum Jülich and imec, Leuven.

The herein presented work was achieved in collaboration with plenty of highly skilled and trained colleagues. I would like to address my special thanks to some of them, even if there is danger to miss anybody:

First, I would like to thank my PhD supervisor Prof. Dr. Siegfried Mantl for the supervision of my thesis and many fruitful discussions. Particularly, I would like to thank him for the excellent research conditions and his continuous effort to enable cutting-edge semiconductor research in Germany. I would also like to thank my second promoter Prof. Dr. Detlev Grützmacher for the evaluation of my thesis. I thank both for the trust they gave me in the past years of our research.

My special thanks go to my daily supervisor Dr. Jürgen Schubert for the shared passion in our daily research, many fruitful discussions and all support during the preparation of my thesis. He is a great teacher and helped me to focus and streamline my results down to finished paper output.

I am also indebted to Dr. Thomas Kauerauf, Dr. Lars-Åke Ragnarsson and Prof. Dr. Guido Groeseneken for allowing me to carry out research at imec and their hospitality. I am grateful for all that you taught me, and the time you invested to answer dozens of questions.

With Dr. Qing-Tai Zhao, Dr. Tom Schram (imec) and Dr. Karl Jacobs (diploma thesis) I would like to thank the people who taught me every tiny detail about cleanroom experiments and (semiconductor) processing.

Dr. Philippe Roussel (imec) was the best teacher for all statistical and analytical real-world problems, they did not teach us in detail at university — and Mathematica, of course. Dr. Robin Degraeve was one of the most inspiring and creative people I had the pleasure to work with and ask questions to. Dr. Bernd Holländer and Dr. Dan Buca were great teachers in solid-state physics and its analytical methods. Especially, Dr. Astrid Besmehn is acknowledged for the shared beamline time and for all I learned about XPS analytics. Dr. Michael Müller is acknowledged for his support with the thermodynamics software. Patric Bernardy was always a great help whenever I did not know sth. about chemistry or XPS.

I would like to thank my colleagues Anna Schäfer, Eylem Durgun Özben, Lars Knoll, Bao Tang (imec), Dr. Moonju Cho (imec), Dr. María Luque Toledano (imec), Jacopo Franco (imec), Dr. Hiroaki Arimura (imec), Simon Richter, Stephan Wirths, Dr. Renato Minamisawa, Dr. Stefan Habicht, Dr. Roman Luptak, and Matthias Schmidt for a lot of fun doing research in the labs. Particularly, I thank all my coauthors for the successful collaborations.

I would also like to thank Prof. Dr. Christoph Buchal and Prof. Dr. Hermann Kohlstedt for the effort they took in teaching us semiconductor physics in university and the later DAAD evaluation for my stay in Belgium.

Successful experiments are not possible without good engineering. Therefore, I would like to thank our technical crew for all work and technical support: Alfred Steffen, Andreas Tiedemann, Christian Scholtysik, Hans Wingens, Jürgen Müller, Karl-Heinz Deussen, Steffi Lenk, and Willi Zander. Especially, Andre Dahmen was the biggest support in all experimental stages with his gift for improvisation in all technical contexts.

Special thanks go to my parents and my parents-in-law for their support. Especially for the time in Leuven, I thank you for Lola's daycare and all your help.

I dedicate this thesis to my beloved wife Katja and my wonderful daughter Lola for your happiness and loving support. Without you, this journey would not have been possible.

Jülich, August 2013 *Alexander Nichau*

Contents

1 Introduction . . . 1

2 High-κ/metal gate stacks for scaled CMOS . . . 3
- 2.1 The Metal-Insulator-Semiconductor (MIS) gate stack . . . 3
  - 2.1.1 Classical MIS capacitance . . . 3
  - 2.1.2 The silicon MOS capacitor . . . 3
  - 2.1.3 From quantum mechanical contribution to Equivalent Oxide Thickness (EOT) . . . 5
  - 2.1.4 Leakage current mechanisms in high-κ gate stacks . . . 6
- 2.2 Scaling of MOSFETs . . . 8
  - 2.2.1 Introduction of high-κ/metal gate stacks . . . 9
- 2.3 Introduction to CMOS Processing . . . 10
  - 2.3.1 Gate-First and Metal-Inserted Poly-Silicon (MIPS) processes 10
  - 2.3.2 Gate-last and replacement gate processes . . . 10

3 Studies of temperature effects in gate stacks . . . 13
- 3.1 Thermodynamic approach . . . 13
  - 3.1.1 Gibbs free energy change and Sanderson's group electronegativity . . . 13
  - 3.1.2 High-κ materials in contact with Si . . . 14
  - 3.1.3 TiN as metal gate on Si/HfO_2 . . . 16
  - 3.1.4 TaN as metal gate on Si/HfO_2 . . . 17
  - 3.1.5 Increased scavenging by TiN-nanolaminates . . . 19
  - 3.1.6 Summary . . . 22
- 3.2 Spectroscopic study of $LaLuO_3$ gate stack interface reactions . . . 23
  - 3.2.1 X-ray photoemission spectroscopy . . . 23
  - 3.2.2 Hard X-ray photoemission spectroscopy . . . 24
  - 3.2.3 Effective attenuation length in $LaLuO_3$ for hard X-rays . . . 25
  - 3.2.4 Thermal stability of the TiN/$LaLuO_3$ gate stack . . . 31
  - 3.2.5 $LaLuO_3$ integration on Ge . . . 51
  - 3.2.6 Summary . . . 54

3.3 Further EOT scaling concepts 57
3.3.1 Use of thinner high-κ dielectric 57
3.3.2 Use of higher-κ dielectrics 57
3.3.3 Use of thinner SiO_2 interfacial layer 59
3.3.4 Thin TiN metal gates 62
3.3.5 Doped TiN 67
3.3.6 Summary 77

4 Integration of high-κ/metal gate stacks into MOSFETs 79
4.1 Optimization of TiN as gate electrode for the integration with $LaLuO_3$ and HfO_2 79
4.1.1 Oxygen barrier layers on TiN 83
4.1.2 Film stress measurements on TiN 86
4.1.3 Boron penetration and effective work function in MIPS gate stacks with TiN 88
4.1.4 TiN as metal diffusion barrier 93
4.1.5 Summary 95
4.2 Device integration of $LaLuO_3$ and HfO_2 96
4.2.1 Device fabrication 96
4.2.2 Device results 97
4.2.3 Summary 104

5 Oxide quality and reliability in high-κ/metal gate stacks 107
5.1 Investigation of oxide traps in high-κ/ metal gate stacks 107
5.1.1 Introduction to charge-pumping 108
5.1.2 Device fabrication and data extraction 111
5.1.3 Impact of interface states on mobility measured by charge pumping 112
5.1.4 Origin of mobility reduction 117
5.1.5 Summary 121
5.2 TDDB reliability of advanced FinFET devices 122
5.2.1 TDDB reliability of HfO_2/TiN on Si-FinFETs 124
5.2.2 The correct determination of the SBD information 136
5.2.3 SILC slope for SBD determination in short-channel MOSFETs 136
5.2.4 Outlook on test time optimization 138
5.2.5 Summary 148

6 Conclusion 153

A MIPS Fabrication Scheme for $LaLuO_3$/TiN Gate Stacks 155

B Etch-Back Process for SiO_2 159

C Publications 161

D **Glossary** 165
References 167

To Katja and Lola

Chapter 1
Introduction

The first bipolar transistor was realized on Ge by John Bardeen and Walter H. Brattain under leadership of William B. Shockley at Bell Labs in 1947. Their work was awarded the Nobel Prize in Physics 1956 "for their researches on semiconductors and their discovery of the transistor effect".[1] In 1960 Khang and Atalla realized the first metal-oxide-semiconductor field effect transistor (MOSFET) on Si[2] The ability to form integrated circuits (IC) on Si has paved the way to continued downscaling of these IC: Gordon Moore predicted in his famous Moore's Law[3] a doubling of transistors per area every two years based on economic considerations back in 1965. The costs involved in the introduction of new process generations into production required a continued downscaling. More than 50 years Moore's Law has been predicting the increase in computational power. The economic needs unleashed creativity and innovations to maintain the predictions. Nowadays the digital society demands evermore increasing computational power and evolves at the heartbeat of Moore's Law.

A recent caesura was the investigation of gate oxides with higher relative permittivity than Silicon dioxide, so called "high-κ" oxides.[4,5] In the past 40 years the gate control of the MOSFET was ever improved by the use of even thinner SiO_2 and the accompanying capacitance increase to avoid the dominance of fringing fields at the boundaries.[6–8] The change to high-κ oxides became necessary since the continued down-scaling of SiO_2 is limited to approx. $\sim$1.4 nm physical thickness.[9] Below this value, gate leakage current becomes a limiting factor[10] as leakage current density approaches several A/cm^2. High-κ gate oxides offer higher capacitance at larger physical thickness, which reduces the gate leakage current at identical equivalent oxide thickness (EOT).[4]

The first high-κ studies favored the introduction of a SiON gate oxide doped with Hafnium at the beginning of the last decade. Since then, a plethora of oxide materials have been screened for the introduction with Si or Ge[11–16] Recent research concentrated on the understanding and integration of Hafnium-based oxides with focus on HfO_2.[17] The use of Hafnium-based oxides utilizes a thin interfacial layer of SiO_2.[18] Several studies state a necessary minimum thickness of SiO_2 around 4 Å to keep its isolating properties.[19–21] For continued EOT scaling below 5 Å the direct contact of Si and high-κ oxide could be beneficial. This property is matched by the Lanthanide group oxides, which tend to silicate formation in conjunction with Si.[15,22] Consequently, a silicate can replace the native SiO_2 and lead to even lower EOT.[23]

Additionally, the introduction of metallic gate electrodes became necessary to avoid the depletion capacitance and high impedance of highly-doped poly-Si electrodes, which became now a limiting factor for EOT scaling.[12] Before, Al had already served as a gate electrode during initial studies in the 1950–1960s.[24] The combination of high-κ oxide and metal gate is usually understood as "high-κ/metal gate" (HKMG) technology.

The advantage of metallic gates is the excellent conductivity and minimum screening length for potential variations following the Thomas-Fermi model.[25,26] However, the material choice is limited if high temperatures in contact with Si are mandatory to withstand.[14,27–29] Annealing can influence the thermal stability and may alter the gate work function especially for today's gate-first processes, where the whole gate stack undergoes the dopant activation. A possible alternative to this high thermal budget is the replacement of the gate stack after activation in a so-called replacement gate process (or quasi-damascene process).

In this, study the integration of Lanthanum Lutetium oxide as a promising higher-k material is investigated on the material and processing side. The high relative permittivity in combination with silicate formation in the interfacial layer is promising for a new gate oxide in direct contact to Si or Ge. Several optimization steps are necessary to make the path from first material characterization to the introduction of $LaLuO_3$ in a classical gate-first process.[30–32] The integration results are furthermore compared to a low-T replacement gate process of another study.[33] A comparison to an HfO_2 gate-first process is pursued. The study deals also with the optimization of TiN as a metal electrode for both studied high-κ oxides.

Several approaches are theoretically and experimentally studied for the continued EOT scalingdown to a few Angstroms. A conservative EOT scaling is intuitively found by the use of thinner interfacial layer. A more aggressive EOT reduction is enabled by the doping of TiN to increase its oxygen solubility. Interestingly, in a first study the doping is found to enable the manipulation of the formation of interfacial silicate in Si/$LaLuO_3$/TiN.

This study closes with a chapter on the impact of (high-κ) oxide quality on both MOSFET device performance and device reliability studied on short channel FinFETs. The oxide quality for HfO_2 as determined by charge pumping measurements is compared to the effective carrier mobility of the FinFETs. A separation of top- and sidewall oxide defects was possible. Furthermore, the time-dependent dielectric breakdown (TDDB) reliability in terms of soft and hard breakdown was analyzed to address the role of EOT scaling and interfacial layer reduction on device reliability.

Chapter 2
High-κ/metal gate stacks for scaled CMOS

2.1 The Metal-Insulator-Semiconductor (MIS) gate stack

The object of interest in this study is the CMOS gate stack and its scalability in terms of oxide capacitance. This section introduces the parameters for successful scaling.

2.1.1 *Classical MIS capacitance*

The classical gate stack for theoretical considerations consists of a metal-insulator-semiconductor (MIS) gate stack. In this study, all major considerations are done with respect to silicon as the semiconductor. The capacitance is then given as the one of a plate capacitor

$$C = \varepsilon_0 \varepsilon_r \frac{A}{d}, \tag{2.1}$$

with ε_0 the vacuum permittivity, ε_r the relative permittivity of the dielectric, d its thickness and A the area of the capacitor.

2.1.2 *The silicon MOS capacitor*

For Silicon and SiO_2 as the dielectric of the gate stack ($\varepsilon_{SiO_2} = 3.9$) the capacitance is commonly expressed on the length scale of SiO_2 thickness at normalized area. The measured quantity is called capacitance equivalent thickness (CET) and means accordingly, that the current gate stack has a capacitance equal to a SiO_2 capacitor of the given thickness.[17,34] This convention is especially useful for the introduction of high-κ oxides, where the direct comparison to SiO_2 is the benchmark.

Often, CET and inversion thickness t_{inv} are used synonymously, whereby t_{inv} is mostly used for the measured capacitance of a MOS field effect transistor in inversion. Both values should be extracted from the capacitance relative to flatband voltage plus a defined offset to allow for a comparison of different gate stacks. Otherwise, differences in flatband voltage or induced charges can lead to incorrect comparison.

2.1.2.1 Flatband voltage V_{fb}

The flatband voltage of the MOS gate stack is defined as the work function difference ϕ_{ms} between semiconductor ϕ_s and gate electrode ϕ_m:[35]

$$V_{\mathrm{fb}} = \phi_{\mathrm{ms}} = \phi_{\mathrm{m}} - \phi_{\mathrm{s}}, \tag{2.2}$$

with the semiconductor potential ($\pm$ for n-type and p-type, respectively)

$$\phi_{\mathrm{s}} = \chi_{\mathrm{s}} + \frac{E_{\mathrm{g}}}{2\mathrm{q}} \pm \phi_{\mathrm{b}} \tag{2.3}$$

The bulk potential ϕ_{b} is calculated as

$$\phi_{\mathrm{b}} = \frac{k_{\mathrm{B}}T}{\mathrm{q}} \ln(N/n_{\mathrm{i}}), \tag{2.4}$$

where N has to be replaced by the acceptor or donor concentration for p-type or n-type substrates, respectively. The intrinsic carrier concentration n_{i} is given by the density of states in the valence D_{v} and conduction band D_{c} as:

$$n_{\mathrm{i}} = (D_{\mathrm{c}}D_{\mathrm{v}}) \exp(-E_{\mathrm{g}}/k_{\mathrm{B}}T) \approx 3.02 \cdot 10^{15} T^{3/2} \exp(-0.55q/k_{\mathrm{B}}T) \tag{2.5}$$

and

$$D_{\mathrm{c}} = 12(2\pi m_{\mathrm{e}} h^2 k_{\mathrm{B}} T)^{3/2} \tag{2.6a}$$

$$D_{\mathrm{v}} = 2(2\pi m_{\mathrm{h}} h^2 k_{\mathrm{B}} T)^{3/2}, \tag{2.6b}$$

where the numeric prefactor is given by the number of equivalent minima in the conduction or valence band, respectively. m denotes the geometric mean of the hole or electron effective masses along the principal axis[24] E.g. for Si: $m_{\mathrm{e}} = (m_{\mathrm{l}} m_{\mathrm{t}}^2)^{1/3}$.

In the case of oxide charges within the gate stack, the flatband voltage is shifted by the potential difference over the gate oxide ϕ_{iso}:

$$V_{\mathrm{fb}} = \phi_{\mathrm{ms}} - \phi_{\mathrm{iso}} = \phi_{\mathrm{ms}} - \frac{Q_{\mathrm{ox}}}{\varepsilon_{\mathrm{ox}}} \cdot EOT = \phi_{\mathrm{m,eff}} - \phi_{\mathrm{s}}, \tag{2.7}$$

Thus, the fixed charges within the (high-κ) oxide can be extracted with a sample set of different (high-κ) oxide thicknesses and a plot of V_{fb} vs. EOT.[35] An experimental extraction of the flatband voltage is thereby given by Hillard and Schroder.[35,36] The magnitude $\phi_{m,\mathrm{eff}}$ is often denoted as the **effective work function (EWF)** and will be used as this term throughout this study. The EWF depends on the bulk potential, the oxide charges, and the (vacuum) work function of the gate electrode.

Equation (2.2) can be further expanded to include all oxide charges by a spatial integration:

$$V_{\mathrm{fb}} = \phi_{\mathrm{ms}} - \phi_{\mathrm{iso}} = \phi_{\mathrm{ms}} - \frac{Q_{\mathrm{it}}(\phi_{\mathrm{s}})}{\varepsilon_{\mathrm{ox}}} \cdot EOT - \frac{1}{C_{\mathrm{ox}}} \int_0^{t_{\mathrm{ox}}} \frac{x}{t_{\mathrm{ox}}} \rho_m \,\mathrm{d}x - \frac{1}{C_{\mathrm{ox}}} \int_0^{t_{\mathrm{ox}}} \frac{x}{t_{\mathrm{ox}}} \rho_{\mathrm{f}} \,\mathrm{d}x\,, \quad (2.8)$$

where by ρ_{f} and ρ_{m} are the charge densities of fixed and mobile charges within the gate stack and Q_{it} are the charges at the semiconductor interface. Those depend on the surface potential ϕ_{s}. The remaining constants have their usual meaning.

2.1.2.2 MOSFET threshold voltage V_{th}

Based on the introduction of the flatband voltage, the threshold voltage of the MOSFET can be introduced as the sum of twice the bulk potential, the flatband voltage and the voltage built-up by the semiconductor charges:

$$V_{\mathrm{th}} = V_{\mathrm{fb}} + 2\Phi_{\mathrm{b}} + \frac{|Q_s|}{C_{\mathrm{ox}}} = V_{\mathrm{fb}} + 2\Phi_{\mathrm{b}} + \frac{\sqrt{2\varepsilon_0 \varepsilon_{\mathrm{Si}} q N_{\mathrm{A}} 2\Phi_{\mathrm{b}}}}{C_{\mathrm{ox}}}, \quad (2.9)$$

Since V_{th} depends on V_{fb}, it depends also on the effective work function and thus on the eventually present oxide charges in the gate oxide.

2.1.3 From quantum mechanical contribution to Equivalent Oxide Thickness (EOT)

When the oxide thickness is scaled, capacitance contributions of the gate electrode and the semiconductor inversion layer cannot be neglected anymore. The overall, measured capacitance becomes then a series system of those capacitances:

$$\frac{1}{C_{\mathrm{total}}} = \frac{1}{C_{\mathrm{gate}}} + \frac{1}{C_{\mathrm{high}-\kappa}} + \frac{1}{C_{\mathrm{IL}}} + \frac{1}{C_{\mathrm{sub}}}, \quad (2.10)$$

where the total capacitance is the sum of the gate capacitance (e.g. poly-Si depletion[17,37] $\sim +3$ Å) and the capacitances of an assumed high-κ oxide and the interfacial oxide (SiO_2). For a metallic gate electrode, the gate depletion capacitance can be neglected (in contrast to a preliminary calculation[37] for Cu of $\sim$1 Å). Additionally, the inversion layer capacitance of the substrate C_{sub} leads to a lower measured capacitance as expressed in CET.

The theoretically achievable capacitance is then obtained from the correction of the CET value by subtraction of the equivalent thickness t_{QM} of the quantum capacitance contributions of the inversion layer:

$$EOT = CET - t_{\text{QM}} \approx CET - 0.4\,\text{nm} \tag{2.11}$$

The substrate contribution of the inversion layer capacitance C_{sub} can be calculated e.g. by the triangular well approximation[38–43] and originates from the quantization of the inversion charge. This leads to a finite distance of the inversion layer to the semiconductor surface, a shift of the charge centroid, and therefore additional charge at the interface to the capacitor. The lowest sub band energies are given in the triangular well approximation as[39]

$$E_n = \left(\frac{\hbar^2}{2m_e}\right)^{1/3} \left[\frac{3\pi}{2} qF \left(n + \frac{3}{4}\right)\right]^{2/3}, \tag{2.12}$$

where m_e is the effective electron mass, q the electron charge and F the electric field at the semiconductor surface. Consequently, the quantum capacitance depends also on the effective mass and should accordingly differ for strained substrates.[44,45]

The capacitance was first modeled by Van Dort et al.[46 1] based on calculations of Ohkura,[42] which calculated the difference in inversion layer thickness to $\sim$1.2 nm in Si depending on the effective field. The thickness difference translates then to $t_{QM} =$ 4 Å according to their different relative permittivity. Usually, a difference of 3–4 Å is given as obtained by the popular C-V fit program of Hauser.[47] However, as stated in Hauser's original article the effective mass in the triangular well approximation is a fit parameter to obtain higher numerical convergence.

2.1.4 Leakage current mechanisms in high-κ gate stacks

As we will later address the gate oxide quality and integrity, the understanding of the band diagram of high-κ/metal gate stacks is mandatory. In the following, a thin interfacial layer (IL) of SiO_2 will be assumed to lie between substrate and high-κ oxide. As two dielectrics are now involved, the voltage drop across the gate stack will be divided by these two layers:

$$V_g = V_{IL} + V_{\text{high}-\kappa} \tag{2.13}$$

The voltage ratios can be calculated from Maxwell's equation and the boundary conditions at the high-κ/IL interface[48,49] with D the dielectric displacement field and E the electric field across the gate oxide(s):

[1] the report contains also a model of the surface potential modification and the modified band gap.

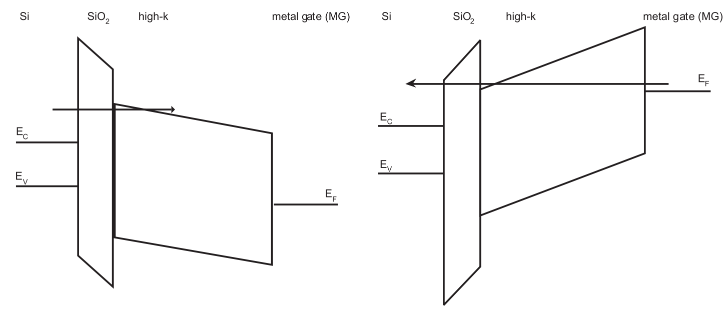


Fig. 2.1 Schematic band diagrams for positive and negative gate bias. (Left) At positive gate voltage electrons from the substrate's conduction band can enter the high-κ conduction band by direct tunneling and release energy in the high-κ oxide (substrate injection). (Right) At negative gate bias (gate injection) electrons must tunnel through the high-κ oxide and SiO_2. At very high bias ($> 5\,\mathrm{V}$) also a Fowler-Nordheim tunneling into the high-κ conduction band is possible. Trap assisted tunneling mediated by defects in the high-κ oxide is also possible at low bias. (not to scale)

$$\nabla \cdot \mathbf{D} = \nabla \cdot \mathbf{D}_{IL} = \nabla \cdot \mathbf{D}_{\text{high}-\kappa} \tag{2.14}$$

$$D = \varepsilon_{\text{SiO2}} E_{IL} = \varepsilon_{\text{high}-\kappa} E_{\text{high}-\kappa} \Leftrightarrow \frac{E_{IL}}{E_{\text{high}-\kappa}} = \frac{\varepsilon_{\text{high}-\kappa}}{\varepsilon_{\text{SiO2}}} \tag{2.15}$$

$$E_{IL} = \frac{V_{IL}}{t_{IL}}; \quad E_{\text{high}-\kappa} = \frac{V_{\text{high}-\kappa}}{t_{\text{high}-\kappa}} \tag{2.16}$$

The voltage across high-κ oxide and interfacial thickness are called $V_{\text{high}-\kappa}$ and V_{IL}, respectively. The letter t denotes the thickness of the respective layer.

Combining equations (2.15) and (2.16) with (2.13) yields an expression for the voltage drop across high-κ oxide and interfacial layer:

$$V_{IL} = \underbrace{\frac{\varepsilon_{\text{high}-\kappa}}{\varepsilon_{IL}} \cdot \frac{t_{IL}}{t_{\text{high}-\kappa}}}_{=:\,\alpha} V_{\text{high}-\kappa} = \left[\frac{1}{\alpha} + 1\right]^{-1} V_g \tag{2.17}$$

$$V_{\text{high}-\kappa} = [\alpha + 1]^{-1} V_g \tag{2.18}$$

Thus, for $\varepsilon_{\text{high}-\kappa} > \varepsilon_{IL}$ and the usual thickness ratios ranging from 1:6 to 1:2: $\alpha > 1$ and the voltage drop across the interfacial layer will be larger than over the high-κ film. This important finding leads to the band diagram[48,49] of a high-κ/metal gate stack as displayed in Figure 2.1.

At positive gate bias, electrons can be injected into the conduction band of the high-κ oxide (**substrate injection** for nMOS). The thin IL and the band bending in the high-κ oxide lead to an effective triangular potential barrier. At negative bias electrons have to tunnel through high-κ and IL and thus the tunneling probability is significantly lowered (**gate injection** for pMOS). At very high gate bias ($> 5\,\mathrm{V}$), a Fowler-Nordheim[50] tunneling into the high-κ conduction band is possible. At lower voltage defects in the high-κ layer can lead to trap-assisted tunneling.[51] A comprehensive overview of tunneling mechanisms is given by Kim and Iwai.[52] The band diagram with oxide charges in place can be simulated with the software package by Southwick.[53,54]

2.2 Scaling of MOSFETs

As stated in the introduction, the scaling of MOSFETs is inevitably connected to the gate capacitance and thus gate control over the field effect. The most intuitive scaling is the *constant field scaling*,[24] where all dimensions and fields are scaled by the same scaling factor κ to achieve scaled MOSFETs. As Sze[24] points out, this intuitive approach is physically limited e.g. by not scaling of subthreshold slope and threshold voltage (due to off-current requirements). The result is a concentration on constant-voltage scaling and enables an individual scaling of the MOSFET parameters in order to preserve its behavior with scaling.

R. Dennard and coworkers at IBM Research developed a set of scaling rules,[6,7,55] which are thought to be most influential for today's scaling.[8] The basic idea is the rescaling[7] of the Poisson equation:

$$\nabla\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = -q/\varepsilon_s(p - n + N_D - N_A)$$

with a set of linear coordinate transformations, the scaling rules:

$$\begin{aligned}
\phi' &= \phi/\kappa \\
(x', y', z') &= (x, y, z)/\lambda \\
(n', p', N_D', N_A') &= \frac{(n, p, N_D, N_A)\lambda^2}{\kappa} \\
\xrightarrow{\text{yields}} \frac{\partial^2(\kappa\phi')}{\partial(\lambda x')^2} + \frac{\partial^2(\kappa\phi')}{\partial(\lambda y')^2} + \frac{\partial^2(\kappa\phi')}{\partial(\lambda z')^2} &= -q/\varepsilon_s(p' - n' + N_D' - N_A')\frac{k}{\lambda^2} \\
\Rightarrow \frac{\partial^2\phi'}{\partial x'^2} + \frac{\partial^2\phi'}{\partial y'^2} + \frac{\partial^2\phi'}{\partial z'^2} &= -\frac{q}{\varepsilon_s(p' - n' + N_D' - N_A')} \qquad (2.19)
\end{aligned}$$

The latter equation (2.19) is interpreted as the Poisson equation for a device with scaled dimensions and its solution differs only by a prefactor from the non-scaled device. As Dennard pointed out, the shape of the electric field in the device stays the same, but the field intensity varies by λ/κ. Thus, important for the gate oxide scaling, a field increase is given if $\lambda > \kappa$.

The *current density* scales then with λ^3/κ^2 and the *power density* on the chip with λ^3/κ^3. Thus, a *constant-voltage scaling* $(l^{\ulcorner}-1)$ would result in the power and current density to be raised by λ^3 with problems on connection lines (electromigration) and heat dissipation, according to Dennard *et al.*[7] Consequently, a more relaxed approach is the scaling of the supply voltage in parallel to the device dimensions in adaption of most parts of the constant-field approach. Interestingly, this scaling is also limited by $V_{th}(I_{\text{off}})$ and subthreshold slope considerations and can only be overcome by novel device concepts like e.g. Tunnel FETs.[25,56–58]

2.2.1 Introduction of high-κ/metal gate stacks

Coming back to the gate oxide and other capacitive components, scaling has to proceed with $1/\lambda$ in order to keep the same field distribution and channel control. Thus, at a design of 45/65 nm the limit for the use of SiO(N) gate oxides was found.

In order to maintain t_{mv} or EOT scaling below -1.4 nm and maintain the channel control in scaled devices, gate oxides with higher relative permittivity compared to SiO_2 have to be introduced. The higher permittivity enables the use of physically thicker layers to obtain the gate capacitance necessary for the adapted constant-field scaling at a parallel reduction of gate leakage.

In parallel, the introduction of metallic gates improves the high-frequency impedance of the gate and leads to higher cut-off frequencies as recently demonstrated.[59,60] Moreover, as prior mentioned, the use of metallic gates in favor of poly-Si gate electrodes decreases the dielectric screening length and avoids potential variations on the lateral scale larger than 1 nm. The dielectric screening length in the dielectric can be calculated by the Thomas-Fermi approximation.[26,61] In this approximation the static screening of local charge variations can be described in the *k*-space with the dielectric screening length

$$k_s = \sqrt{4(3/\pi)^{1/3} n_0^{1/3}/a_0} \tag{2.20}$$

with the Bohr radius $a_0 = 0.53$ Å and n_0 the atomic density of the metal (cm^{-3}). In the case of TiN, the atomic density amounts to $n_0^{\text{TIN}} = 1.0569 \cdot 10^{22}/\text{cm}^3$ and the dielectric screening length is calculated to be 0.53 Å$(\approx a_0)$. A corresponding interface or defect charge is then faster screened, as the unscreened Coulomb potential would decrease:

$$\sim \exp(-k_s r)/r.$$

The minimum screening length in metals like TiN is in contrast to the minimum achievable Debye length in a degenerate semiconductor like poly-Si of $\sim$ 1–2 nm.[24] Thus, especially at the boundaries of the transistor gate, additional charges can be screened proportional to $\sim \exp(4) \approx 55x$ more effective.

The current study investigates several possibilities for the continued scaling of different high-κ oxides and appropriate metal gates.

2.3 Introduction to CMOS Processing

State-of-the-art CMOSFETs with high-κ/metal gate stacks demand for advanced, self-aligned[55] processes. Two basic methodologies were established over the past years. The classical gate-first process originates from the processing steps for SiO_2/poly-Si gate stacks as used for over 4 decades. A more recent approach to HKMG integration is the use of a gate-last process.[59,62]

2.3.1 Gate-First and Metal-Inserted Poly-Silicon (MIPS) processes

The Gate-First (GF) goes back to the establishment of the SiO_2/poly-Si gate stack, which was and is used in CMOS processing since the 1960s.[24] This material system allowed the use of high temperatures for S/D (and gate) dopant activation with an already patterned gate stack. The patterned gate stack served thereby as an implantation mask for the S/D implants, which made the process self-aligned in respect to the gate finger and resulted in lower S/D resistance.[55]

Consequently, the horizon of HKMG stacks was pursued with minimum changes to the established processes. In the beginning nitrided SiO_2 was used to avoid dopant diffusion through ultrathin SiO_2.[10,63,64] In a next step, the SiON gate dielectric was increased in relative permittivity by the doping with Hafnium.[65,66] Over the last years research focused on the full integration of HfO_2.[12,34,67–72] In parallel, a metallic gate electrode was inserted[27,28,65,73–76] between gate dielectric and poly-Si to decrease the potential variations and the corresponding screening length.[7,25,26] The resulting process is called metal-inserted poly-Si (MIPS) and its characteristic is the dopant activation at temperatures above 1000°C with HKMG already patterned and in place.

A few groups experimented also with the use of fully silicided (FuSi) poly-Si gates, where a Ni or W silicide was formed after doping activation. By the control of poly-Si to metal ratio (i.e. gate stack height) one can control the resulting silicide phase and the corresponding work function.[66,77] However, this approach demands a tight process control over the thickness determining steps (poly-Si CVD growth, gate etch, poly-Si CMP, etc.).

2.3.2 Gate-last and replacement gate processes

In recent years the metal nitrides TiN/TaN were recognized to feature excellent thermal stability in conjunction with HfO_2, but the work function after annealing was shifted to midgap position with respect to the Si band gap.[17,27,28,73,75,78–81] Capping layers were introduced to tune the work function to n/p-type band edge behavior.[17,29,67,70,82–86] However, another idea was proposed in parallel to progress in gate stack patterning: so-called gate-last (GL) or damascene or replacement gate (RPG) process integration.[59,60,62]

The main idea for the gate-last approach is the replacement of a damascene gate ("dummy gate") after high-T doping activation with an appropriate HKMG combination with work functions suitable for nMOS and pMOS.[28,29] As Schaeffer[29] stated, most metals with *n*-type work function (around 4.05–4.1 eV) are highly reactive and cannot withstand the classical doping activation temperatures. The integration of these metals at lower thermal budget, i.e. later in the process flow, allows suppressing unwanted reactions and ideally the adoption of the targeted work functions.

Nevertheless, the realization remains challenging. The damascene gate is usually a SiO_2/poly-Si gate stack. After gate patterning, implantation, and activation the whole wafer is covered by e.g. SiO_2. This passivation layer has to be planarized by CMP in a following step opening the cover layer to the poly-Si gate stack (ILDO poly-Si opening CMP').[87] The poly-Si can be removed by a short HF dip and TMAH wet etching with high selectivity to SiO_2.[88] The surface preparation to the channel follows by an ozone oxidation. The obtained trenches can then be filled by several steps of HfO_2 (ALD)/metal gate (ALD) and fill metal (e.g. W (CVD)).[62]

On the one hand, the advantage of a GL process over the GF integration manifests itself in a more flexible choice of the metal gate. A second advantage is the ability to integrate high mobility substrates like (Si) Ge or III/V layers as local epitaxially grown layers with the replacement gate process as those layers demand lower process temperatures as compared to the typical 1000°C dopant anneal on pure Si.[89–91] A third possibility is the use of compressive and tensile stress liners for *p* and nMOS, respectively, which is more complex to achieve at high-T.[45,62]

A comprehensive overview of the most important GF and GL process steps is given in Figure 2.2. The gate-last process was adapted from a sketch by Veloso and Mertens.[87]

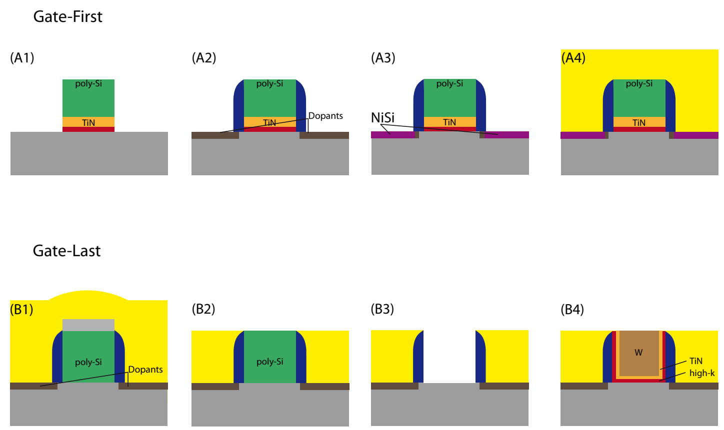


Fig. 2.2 Summary of the most important process steps commonly used in gate-first and gate-last processes. (A1) The GF gate stack is deposited and patterned. (A2) Afterwards a first thin spacer is deposited, followed by ion implantation and activation. The complete gate stack undergoes the high-T annealing. (A3) NiSi formation provides high conductive S/D areas. (A4) Several passivation steps and stress liners are deposited prior to metallization. In a gate-last process, a damascene gate (poly-Si) is patterned and serves as a mask for ion implantation (B1). Gray on top of poly-Si: SiO_2 serves as a hard mask for reactive ion etching. A dielectric like SiO_2 fills the complete wafer (yellow). (B2) A CMP step follows to planarize the wafer and opens the trenches to poly-Si. (B3) a wet etch using HF/TMAH can be employed to remove the poly-Si sacrificial gates. (B4) After poly-Si removal a conformal deposition of high-κ, metal gate and fill metal (here W) forms the final gate stack. A second planarization of W must be carried out to planarize the wafer again. Finally, the S/D area can be opened again for silicidation (after Veloso and Mertens.[59,62,87]

Chapter 3
Studies of temperature effects in gate stacks

3.1 Thermodynamic approach

The typical CMOS processing temperatures range from depositions at 300°C to annealing steps over 1000°C. A key element for the reduction of SiO_2 interfacial layer and continued EOT scaling is the control of oxygen reactions at every processing step.

In this chapter the interaction of interfacial layer, high-κ oxides ($LaLuO_3$ and mainly HfO_2) and metal electrodes TiN/TaN is discussed based on thermodynamic considerations. These considerations suggest an experimental route, which was followed in the experimental part of this work. The reduction of interfacial oxide is examined on base of the theory of redox systems and the theoretical evaluation of reactants.

3.1.1 Gibbs free energy change and Sanderson's group electronegativity

The pioneering work of Hubbard and Schlom[13] employed standard thermodynamic calculations to investigate the thermal stability of gate oxides on silicon; later the approach was extended by Schlom and Haeni.[14] This numerously cited work founded the base of the high-κ/metal gate introduction into CMOS. Moreover, it is a very efficient approach to assess material compatibility. Normalized to one mole of SiO_2, Hubbard and Schlom calculated the Gibbs free energy change for most elements in direct contact with Si. Likewise, one obtains the thermal stability, i.e. the direction of reaction, of gate stacks on silicon:

$$\Delta G^o_T = \sum_{\text{Products}} \nu G^o_{f,T} - \sum_{\text{Reactants}} \nu G^o_{f,T}, \tag{3.1}$$

here ν is the balancing coefficient in the chemical reaction formula and $G^{o}_{f,T}$ denotes the free energy formation of each constituent element in its standard state, as denoted by the superscript o, and at reaction temperature T. A *negative value of ΔG^{o}_{T} indicates thermodynamic equilibrium on the product side, while a positive value marks a reaction to the reactant's side*.

The thermodynamic approach provides indications for the actual reactions within the gate stack. Nevertheless, this first order approach neglects kinetic and dynamic diffusion at the atomic level. The latter should also not be ignored, as we will later see.

Values for the free energy are given at 1000°C but can be assumed to have the same sign (RT to 1000°C) if not denoted otherwise. Thermodynamical data are taken from the compilations of Barin[92] and the CRC Handbook of Chemistry & Physics.[93] The Gibbs free energy was calculated manually or with the software package Outotec HSC Chemistry 5.11.[†] Due to the large number of possible reactions, we concentrate herein on the most certain phase formations. If reactions between metal gate and interfacial layer are given, the high-κ layer in between is thought to remain stable. Then this layer was neglected for the sake of readability if not denoted otherwise.

In the absence of calorimetric data, the concept of Sanderson's group electronegativity (EN) can be applied as an estimation of possible redox systems.[29,94] Thereby a metal featuring an oxide with lower group electronegativity can possibly reduce a higher group EN metal oxide. The group EN is obtained by taking the geometric mean of the weighted elemental EN χ_i. The EN is weighted by the number of atoms n in the compound molecule:

$$EN_{\text{Compound}} = \sqrt[n]{\prod_{i}^{n} \chi_i}. \tag{3.2}$$

The group EN of $LaLuO_3$ is then obtained as

$$EN_{\text{LaLuO}_3} = \sqrt{(EN_{Lu} \cdot EN_{Lu} \cdot EN_{O}^{3})} = \sqrt[5]{1.1 \cdot 1.27 \cdot 3.44^{3}} = 2.24. \tag{3.3}$$

The following calculations mostly address metal gates on Si/HfO_2. $LaLuO_3$ could not be assessed by thermodynamic calculations due to the absence of calorimetric data in the literature for the oxide and its silicates.

3.1.2 High-κ materials in contact with Si

For the case of metal oxides MO_x as high-κ oxides on silicon, the thermal stability can be obtained from the phase diagram M-O-Si. Schlom and Hubbard[13] developed a flowchart, reproduced in Figure 3.1, for the determination of stable phases in the

[†] http://www.outotec.com

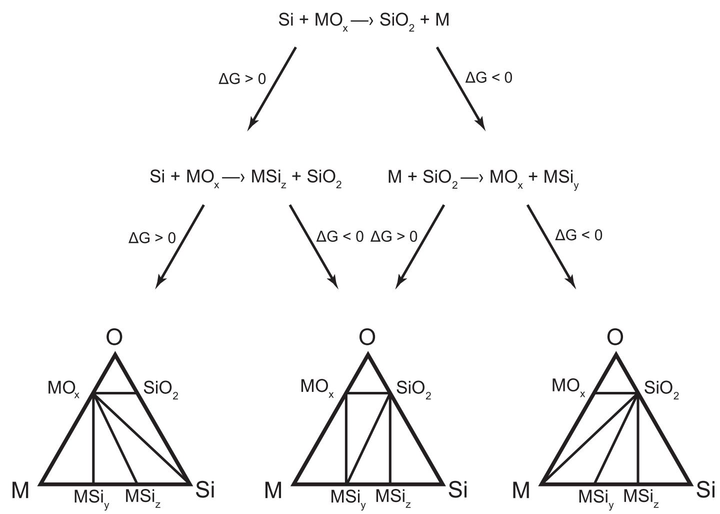


Fig. 3.1 (reproduced after Schlom and Hubbard[13]) for binary oxides on Si three different types of phase diagrams can be distinguished (bottom). The flowchart above indicates the possible reactions and the applicable phase diagram. Tie lines between two reactants mark stable phases.

system M-O-Si. The formation of ternary phases is neglected. However, the relevant ternary phases, i.e. silicates, were mostly unknown to the work of Schlom and are still non-described as the relevant literature might suggest.[93,95]

The flow chart in Figure 3.1 has two meanings. On one side, the thermal stability of metal oxide (high-κ) on Si can be determined with the appropriate thermodynamic data. Specifically, HfO_2 on Si is found to be stable (RT to 1000°C) regarding the initial reaction:

$$\mathrm{Si} + \mathrm{HfO_2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = +193\,\mathrm{kJ/mol}} \mathrm{SiO_2} + \mathrm{Hf} \tag{3.4}$$

However, the second reaction involving Hf silicide (2$^{\mathrm{nd}}$ step in Figure 3.1) may be assumed to be largely positive according to recent data.[96,97] The quality of those simulation data must still be considered premature since they are only based on two measurement points. However, those support calculation leading to stable HfO_2 on Si in agreement with recent experimental data, besides some minor silicate formation. In theory, the thermal stability of HfO_2 and La_2O_3 is not well understood specifically the silicate formation due to insufficient data.

Nowadays both, La_2O_3 and HfO_2, are used or investigated as high-κ material The silicate formation is claimed to be controllable for La_2O_3 and HfO_2 was experimentally found to feature high thermal stability.[17] A third candidate ZrO_2 was experimentally and thermodynamically shown to be stable in contact with Si.[14]

3.1.3 *TiN as metal gate on* Si/HfO_2

For a known-stable system on Si, like $TiN/HfO_2/Si$, the oxidation of the metal electrode in conjunction with the reduction of Si is a promising way to further reduce the amount of interfacial SiO_2 between Si and high-κ material and has been proven experimentally.[84,85,98,100,101] In the temperature range RT-1000°C Ti and TiN are stable on HfO_2.

$$2\ \mathrm{TiN} + 2\ \mathrm{HfO_2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}}=+415\,\mathrm{kJ/mol}} 2\ \mathrm{TiO_2} + 2\ \mathrm{Hf} + \mathrm{N_2}\uparrow \tag{3.5}$$

$$\mathrm{TiN} + \mathrm{HfO_2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}}=+157\,\mathrm{kJ/mol}} \mathrm{TiO_2} + \mathrm{HfN} \tag{3.6}$$

$$\mathrm{Ti} + \mathrm{HfO_2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}}=+206\,\mathrm{kJ/mol}} \mathrm{TiO_2} + \mathrm{Hf}. \tag{3.7}$$

In consequence, defect generation in HfO_2, i.e. oxygen vacancy generation, by oxygen gettering from TiN is less expected. Also the stoichiometry of Ti:N is not determining the thermal stability of this system with regard to thermodynamics. However, Ti silicide formation is thermodynamically favorable and Ti would react with a Si-cap influencing the work function of the metal in contact with high-κ. Stoichiometric TiN, however, prevents the silicide formation:

$$\mathrm{Ti} + \mathrm{Si} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}}=-31\,\mathrm{kJ/mol}} \mathrm{TiSi}, \tag{3.8}$$

$$2\ \mathrm{TiN} + 2\ \mathrm{Si} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}}=+86\,\mathrm{kJ/mol}} 2\ \mathrm{TiSi} + \mathrm{N_2}\uparrow. \tag{3.9}$$

If no Si cap is employed, there is still the possibility of phase mixture with the subsequent contact metal to the gate (diffusion). Diffused species might also cause unwanted reactions with the high-κ layer.

As Schlom states, the thermally stable oxide can act as oxygen conductor in a diffusion reaction:

$$\mathrm{Ti} + \mathrm{HfO_2(stable)} + \mathrm{SiO_2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}}=-31\,\mathrm{kJ/mol}} \mathrm{TiO_2} + \mathrm{HfO_2(stable)} + \mathrm{Si}, \tag{3.10}$$

$$\begin{aligned} 2\ \mathrm{TiN} + \mathrm{HfO_2(stable)} + 2\ \mathrm{SiO_2} &\xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}}=+373\,\mathrm{kJ/mol}} 2\mathrm{TiO_2} \\ &+ \mathrm{HfO_2(stable)} + 2\ \mathrm{Si} + \mathrm{N_2}\uparrow. \end{aligned} \tag{3.11}$$

Thus, the reduction of interfacial layer can be promoted by the metal electrode if the diffusion gradient is sufficient and excess Ti atoms are present. If prior TiO_2 is present, there is an indication for SiO_2 growth under reduction of TiO_2. In addition, oxygen in the gate stack can be transferred to the interfacial layer through the stable HfO_2 layer during processing:

$$\mathrm{Si} + \mathrm{HfO_2(stable)} + \mathrm{O_2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = -717\,\mathrm{kJ/mol}} \mathrm{SiO_2} + \mathrm{HfO_2(stable)}. \tag{3.12}$$

Additionally, Ti(N) oxidizes very easily and uncontrolled oxidation should be avoided if the oxidation was later used for controlled oxygen scavenging and interfacial layer reduction:

$$\mathrm{Ti} + \mathrm{O_2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = -704\,\mathrm{kJ/mol}} \mathrm{TiO_2}, \tag{3.13}$$

$$2\ \mathrm{TiN} + 2\ \mathrm{O_2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = -977\,\mathrm{kJ/mol}} 2\ \mathrm{TiO_2} + \mathrm{N_2}\uparrow, \tag{3.14}$$

where TiO_2 (anastase grade) was chosen for analysis, but reactions involving rutile Ti dioxide are found close in magnitude. Note that: in oxidation studies of TiN the oxidized species were primarily interpreted to lie along grain boundaries.[102, 103] Hence, a complete transformation of TiN to oxide is not expected and is likely diffusion limited; the theory of Fromhold[104] models the metal oxidation as the oxygen radical reaction at the underlying MO_x/M interface in agreement with Fick's law.

Yet, some first conclusions can be made upon these calculations. TiN as metal electrode is only stable in contact with Si capping if stoichiometric. TiN oxidizes also in the stoichiometric form and should be protected by appropriate capping. Excess oxygen in the high-κ layer/gate stack (ALD water process, surface humidity, carbon contaminations) first can oxidize TiN, second produces interfacial SiO_2, and should be avoided.

Some researchers found good scaling trends for very thin layers (2–5 nm) of TiN[59, 70, 72] in high temperature gate-first processes. At high temperatures ($>900^\circ$C), SiO_2 desorption to gaseous SiO radicals might help to scavenge oxygen from the interfacial layer. A possible explanation is the combination of oxidation within TiN with a strong oxygen gradient and a reduction of Si:

$$2\ \mathrm{TiN} + 4\ \mathrm{SiO}\uparrow \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = -161\,\mathrm{kJ/mol}} 2\ \mathrm{TiO_2} + 2\ \mathrm{N_2}\uparrow + 4\ \mathrm{Si}. \tag{3.15}$$

In summary, the oxygen scavenging is mainly reproduced with the thermodynamic data by careful prevention of TiN oxidation during deposition and *ex-situ* treatment. Excess Ti (like in PVD depositions) or high-temperature annealing might be the main origin of oxygen scavenging.

3.1.4 TaN as metal gate on Si/HfO_2

Leaving TiN and considering TaN as the candidate for metal electrodes different properties in contact with Si or HfO_2 are found. Ta(N) can easily be oxidized, but is stable on HfO_2, and should be used in conjunction with suitable capping layers:

$$2\ \mathrm{Ta} + \frac{5}{2}\ \mathrm{O_2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = -1481\,\mathrm{kJ/mol}} \mathrm{Ta_2O_5}, \tag{3.16}$$

$$2\ \mathrm{TaN} + \frac{5}{2}\ \mathrm{O_2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = -1184\,\mathrm{kJ/mol}} \mathrm{Ta_2O_5} + \mathrm{N_2} \uparrow, \tag{3.17}$$

$$2\ \mathrm{Ta} + \frac{5}{2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = +792\,\mathrm{kJ/mol}} \mathrm{Ta_2O_5} + \frac{5}{2}\ \mathrm{Hf}, \tag{3.18}$$

$$2\ \mathrm{TaN} + \frac{5}{2}\mathrm{HfO_2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = +1090\,\mathrm{kJ/mol}} \mathrm{Ta_2O_5} + \mathrm{Hf} + \mathrm{N_2} \uparrow. \tag{3.19}$$

In contact with SiO_2 (separated by stable HfO_2) Ta(N) does not reduce Si by oxygen scavenging and rather may release oxygen to Si. The oxygen scavenging at moderate temperatures as observed for TiN is not indicated:

$$4\ \mathrm{Ta} + 5\ \mathrm{SiO_2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = +620\,\mathrm{kJ/mol}} 2\ \mathrm{Ta_2O_5} + 5\ \mathrm{Si}, \tag{3.20}$$

$$13\ \mathrm{Si} + 2\ \mathrm{Ta_2O_5} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = -979\,\mathrm{kJ/mol}} 4\ \mathrm{TaSi_2} + 5\ \mathrm{SiO_2}. \tag{3.21}$$

SiO desorption may equilibrate Tantalum pentoxide reduction:

$$4\ \mathrm{TaN} + 5\ \mathrm{SiO_2} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = +595\,\mathrm{kJ/mol}} 2\ \mathrm{Ta_2O_5} + 5\ \mathrm{Si} + 2\mathrm{N_2} \uparrow, \tag{3.22}$$

$$2\ \mathrm{TaN} + 5\ \mathrm{SiO(g)} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = -152\,\mathrm{kJ/mol}} \mathrm{Ta_2O_5} + 5\ \mathrm{Si} + \mathrm{N_2} \uparrow. \tag{3.23}$$

To clarify the predominant phase in the oxidation and reduction of TaN, equilibrium calculations were done with the software package HSC chemistry 5.1. The calculations reveal equilibrium between SiO_2 and TaN between RT and 1000°C, if no prior existence of Ta_2O_5 is assumed which shifts the equilibrium.

The silicide formation between Ta(N) and Si-cap is comparable to TiN. Ta tends to form a silicide in contact with Si, whereby all silicide phases are thermodynamically accessible between RT and 1000°C. The predominant silicide phase depends on the concentrations. TaN is again stable in contact with Si and should therefore be the predominant phase in the TaN deposition in conjunction with a Si-cap:

$$\mathrm{Ta} + 2\ \mathrm{Si} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = -115\,\mathrm{kJ/mol}} \mathrm{TaSi_2}, \tag{3.24}$$

$$2\ \mathrm{TaN} + 4\ \mathrm{Si} \xrightarrow{\Delta G^0_{1000^\circ\mathrm{C}} = +72\,\mathrm{kJ/mol}} 2\ \mathrm{TaSi_2} + \mathrm{N_2} \uparrow. \tag{3.25}$$

Summarizing, TaN is a very stable metal electrode on HfO_2 and prevents the interfacial layer growth over the typical temperature range during processing under the assumption of oxygen free, in-situ capped TaN. A scavenging effect as discussed for TiN could not be identified on base of equilibrium calculations. For EOT values below 8 to 10 Å the interface preparation of ultrathin SiO_2 ($\sim$ 4–8 Å) and prevention of further oxidation might not be sufficient if a minimum thickness of HfO_2 around 2 nm is mandatory to prevent excess leakage. Thus, thermodynamic considerations recommend TiN in favor of TaN for ultimate scaling of the interfacial layer and resulting EOT.

3.1.5 Increased scavenging by TiN-nanolaminates

As calculated in section 3.1.3, the scavenging by excess Ti in TiN is a viable way to control the interfacial layer reduction. However, most scavenging is reached for very thin layers (2 nm HfO_2) and high-temperature annealing, likely due to SiO desorption. Even then, additional scaling of the high-κ thickness is necessary to achieve ultra-low EOT values[7] in comparison to other methods, which can leave the high-κ thickness untouched.[85,98,101]

In the last years several constrains led to an increased demand for gate-last device integration processes. These processes feature per definition a lower thermal budget for the gate stack, depending on the literature between 600 and max. 800°C[59] The necessity for new ideas to reduce the interfacial layer has therefore become a prominent demand for the further downscaling of the MOSFET dimensions.

The use of TiN is preferred due to its superior thermal stability and basic scavenging properties. To maintain its work function and avoid additional diffusing species, the direct contact of pure TiN on HfO_2 is favorable.[60,98] A key element for increased scavenging over the whole temperature range can be inserted on top of thin TiN ($\sim$ 2 nm): A metallic, conductive element, which offers a higher reactivity with oxygen, i.e. higher negative Gibbs free energy change, and enhances the oxygen solubility. In this model, TiN is assumed to scavenge oxygen and reduce Si as discussed before. Now the additional top layer reduces TiO_2 again and increases the overall oxygen solubility within the gate stack. Such an element M must exceed TiN in negative reaction enthalpy:[98]

$$2\frac{x}{y}\,\mathrm{M} + \mathrm{SiO_2} \xrightarrow{\Delta G_T^0 < 0} \mathrm{Si} + \frac{2}{y}\mathrm{M}_x\mathrm{O}_y. \tag{3.26}$$

The relevant metals M should also be considered CMOS compatible. Based on Pauling's electronegativity scale and the free energy of formation, relevant elements are identified. Figure 3.2 shows the resulting Ellingham diagram, i.e. Gibbs free energy of formation vs. temperature, for several metal oxides in comparison to SiO_2 and TiO_2. The energies were normalized to the reaction of a single oxygen atom to allow for comparison among the different stoichiometry. An intersection with the study of rare-earth oxides becomes visible. Eu_2O_3, La_2O_3 and Lu_2O_3 are possible candidates for a nanolaminate of TiN/M/TiN to increase scavenging. However, it is also visible, that HfO_2 lies above Strontium oxide, Yttrium oxide, Lanthanum and Lutetium oxide. If the scavenging is too aggressive, oxidation of these elements might consume not only SiO_2 but also HfO_2 leading to oxygen vacancies. However,

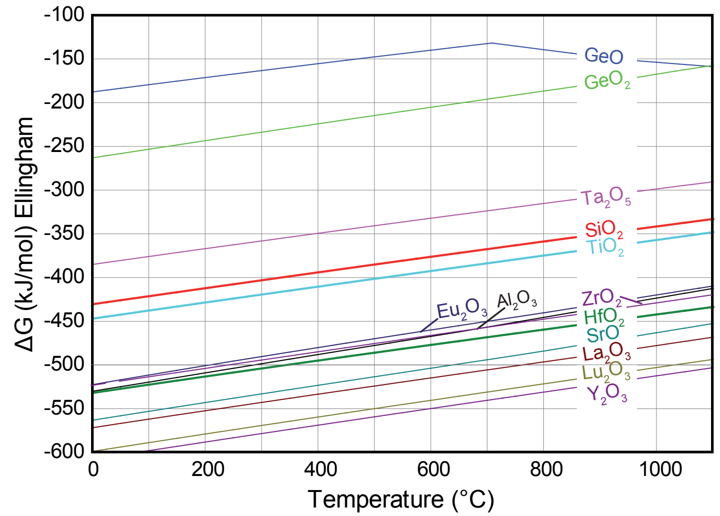


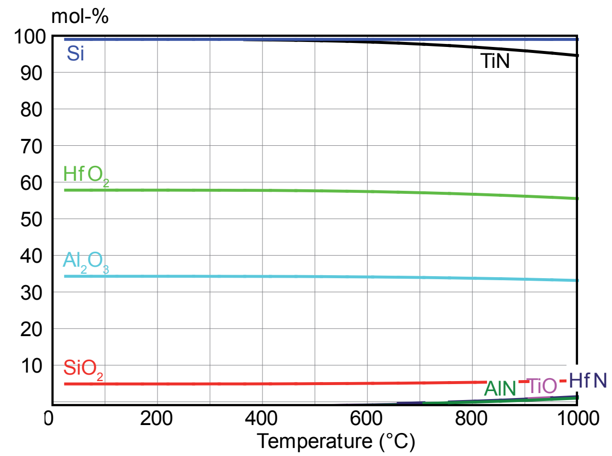
Fig. 3.2 Ellingham diagram of the Gibbs free energy of formation vs. temperature at 1013 hPa for several metal oxides in comparison to SiO_2 and TiO_2. Elements with their oxides at more negative level can reduce above lying oxides to elemental form under oxidation of the reactant, e.g. *Y* can reduce Al_2O_3 to pure Al under formation of Y_2O_3. Si cannot reduce TiO_2.

this mechanism depends on the oxygen gradient within HfO_2/TiN/M/TiN and has to be optimized. Note that the scope is the reduction of the silicon dioxide interlayer, but not to completely remove it. Some lower-κ oxide has to remain for device mobility. Several studies have shown a detrimental effect on electron mobility for high-κ in direct contact with the Si substrate.[12, 20, 71, 105]

Among the possible options, Aluminum has the reactivity to reduce TiO_2 and SiO_2 but leaves HfO_2 unreacted. Moreover, Al is widely used in semiconductor processes and can be easily introduced for such a scavenging process. The annealing to activate the oxygen scavenging has to be performed at moderate temperatures for short period and best for stoichiometric TiN since Al-Ti form binary alloys.[95] For stoichiometric TiN, alloying with Al is thermodynamically unfavorable (though diffusion of Al through TiN is possible):

$$2\,\mathrm{TiN} + 2\,\mathrm{Al} \xrightarrow{\Delta G_T^0 = +310\,\mathrm{kJ/mol}} 2\,\mathrm{TiAl} + \mathrm{N_2} \uparrow . \tag{3.27}$$

Figure 3.3 shows the equilibrium compositions of oxides/nitrides and elements in the model system SiO_2/HfO_2/TiN/Al in mol-% of each species. One kmol of each compound is assumed while varying the Al fraction. The remaining fraction of SiO_2 for 1.2 kmol Al is shown l.h. *s.*, whereby the r.h. *s.* assumes 1.34 kmol Al, slightly



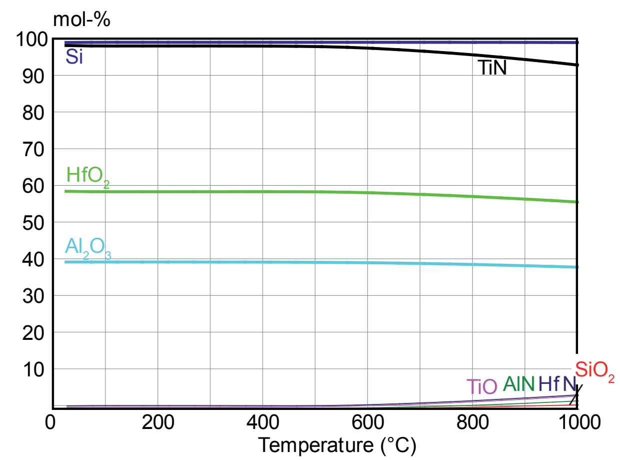


Fig. 3.3 Thermodynamic equilibrium compositions vs. temperature in mol-% for oxides/nitrides/elements in the model system 1 kmol SiO_2/1 kmol HfO_2/1 kmol TiN/x kmol Al. (left) 1.2 kmol Al consume nearly all SiO_2 as expected from the redox-reaction between Al and SiO_2. (right) 1.34 kmol Al consume all SiO_2 below 600°C. Above, minor fractions of SiO_2 and nitride compounds are formed under reduction of HfO_2 and Al_2O_3.

above the stoichiometric redox-reaction between Al and SiO_2. As expected, the material system behaves in equilibrium like the initial, simple redox-reaction. Above 600°C, unwanted reactions to nitrides occur at small fractions, which seem to have the potential to reduce HfO_2 and Al_2O_3. These reactions should be kept in mind for process optimization, since damage of the high-κ layer in terms of oxygen vacancies is unwanted.[69] The Ellingham diagram suggests also Zr, an alternative to Al, to reduce SiO_2. Finally, also the diffusivity of the materials within TiN should be considered. In this study, Al was found to diffuse fast even at low T.

If Al or other listed candidates are introduced as dopants for increased scavenging, the (high temperature) anneals should be implemented after gate stack structuring, as usually the case, to avoid observed difficulties in patterning due to hardened metal layers (i.e. Al_2O_3 in a TiN matrix).

For the case of SiGe and Ge substrates, the Ellingham diagram indicates a favorable reduction of Ge oxide compared to SiO_2 reduction. Two consequences arise: First, the scavenging principle can be adapted to Ge substrates and second, the reduction is stronger for Ge substrates as on Si. Thus, a Gedanken experiment would be to use SiGe without the usual cap-layer and use the scavenging in a peculiar way to reduce the low quality Ge oxides but remain the better passivating SiO_2 as interfacial layer between gate stack and channel. This principle has the potential to avoid the search for an interfacial layer on SiGe and use ultrathin Si capping layers on Ge in combination with low EOT.

3.1.6 Summary

The thermodynamic calculations and collected experimental work indicate the thermal stability of HfO_2 on Si/SiO_2. The reduction of SiO_2 at the interfacial layer is possible by the reduction of SiO_2 and oxidation of TiN. This reaction can be achieved by minimization of oxygen incorporation into TiN prior to high-T processing and oxygen control in subsequent processes. A suitable method to prevent TiN oxidation is the in-situ capping of TiN by a-Si. The reduction of the interfacial layer can be achieved with thin TiN and high-T desorption of SiO or oxygen gettering of bulk TiN. TaN oxidizes as well ex-situ, but cannot reduce SiO_2 as TiN does. An appropriate use of TaN for nMOS work functions would require other methods for the reduction of interfacial layer (e.g. back-etch of the native oxide).

A method to increase the oxygen solubility in TiN is the doping of the metal electrode with other metals with higher tendency to reduce SiO_2. Suitable candidates are found among the rare-earth elements and Al/Zr. These redox systems remain functional also at lower temperatures and thus are promising for future replacement gate processes. The reduction of the interfacial layer is even more effective on (Si)Ge than on pure Si due to the higher Gibbs free energy of the Ge oxides. A possible integration of high-κ oxides on SiGe could employ such a well-tuned redox system to reduce the interfacial layer to pure SiO_2 at low EOT.

3.2 Spectroscopic study of $LaLuO_3$ gate stack interface reactions

The integration of a new material into an existing CMOS process flow demands a detailed knowledge of its thermal stability in contact with the underlying silicon and other materials used. Second but not last this knowledge can prevent possible contamination during processing and side effects on downstream operations. The temperature applied in today's CMOS process flows varies from 400°C to 1000°C[59,70] in respect to the integration scheme, i.e. the sequence the final gate stack is deposited. Even for a so-called high-κ first replacement gate process[59] a higher temperature applied to the high-κ itself can be beneficial to improve the oxide quality or reduce SiO_2 in the interfacial layer.[85,98,101,106]

The thermal stability of the gate stack comprising $LaLuO_3$ and TiN as metal gate was analyzed by means of photoemission spectroscopy. Thereby the meaningfulness of results was checked in comparison to mass spectroscopy and electron microscopy.

3.2.1 X-ray photoemission spectroscopy

The chemical binding states of the studied materials can be assessed by X-ray photoemission spectroscopy (XPS, sometimes also referred to as Electron Spectroscopy for Chemical Analysis (ESCA)). This technique uses the kinetic energy of emitted electrons from photo-absorption processes to characterize the inter-atomic binding states. The penetrating X-rays are absorbed by an atom in the solid and lead to the emission of an electron from a core-level or valence level of the atoms. Under vacuum, the emitted electrons can then be analyzed by an electrostatic detector in terms of kinetic energy and count rate. The binding states of the solid are obtained by scanning the kinetic energy E_{kin} of the electrons in the range from zero to the photon energy E_{Ph} emitted from the X-ray source (e.g. $E_{\text{Ph}}(\text{A1 } K_\alpha) = 1486.6\,\text{eV}$). Photoemission takes place, where[107]

$$E_{\text{kin}} = h\nu - E_{\text{bin}} - \phi \tag{3.28}$$

i.e. the emitted electrons equal in kinetic energy the energy difference of the absorbed photon and the originating electron binding level. Thereby the binding level is referenced to the Fermi level E_F in the solid as $E_F = 0$. An additional energy difference ϕ can arise from the work function difference of detector and sample plus additional charges present in insulating.[108] A common approach for energy calibration ($\phi = 0$) is the measurement of a grounded, metallic reference with known binding energy (e.g. 4f-level in $Au^{\pm 0}$). The emission process itself is kinetically limited and thus only the first few nanometers of the solid are contributing to photoemission.

This basic photoemission process is referred to as the "one-electron picture", which expresses the simplified concept of emitted electrons being not influenced by any surrounding potential. Especially, for the herein analyzed Lanthanides certain deviations from this picture are observed.

XPS is a relatively new experimental technique originally developed with main achievements[109] of Kai Siegbahn, who received the Nobel Prize for his pioneering work in 1981.

3.2.2 Hard X-ray photoemission spectroscopy

The advent of powerful synchrotron sources worldwide allowed using these sources for photoemission spectroscopy (PES). Synchrotron facilities offer increased X-ray brilliance, i.e. photons per solid angle, area, and second, exceeding standard laboratory equipment by many orders of magnitude. The experiments described herein were mainly carried out at the European Synchrotron Radiation Facility (ESRF). Today (2012), this light source offers a brilliance of $8 \cdot 10^{20}$photons/s/mm^2/mrad^2/0.1% bandwidth. For comparison, a standard X-ray tube offers a brilliance of 10^8 at much lower energy.

Hard X-ray photoemission spectroscopy (HAXPES) at synchrotron beamlines offers increased X-ray brilliance and the availability of photon energies exceeding the limits of laboratory XPS. Pioneering work[110] for the instrumentation in this research area has been done at the Spanish beamline, European Synchrotron Radiation Facility (ESRF) in Grenoble, France or at Spring-8 in Riken, Japan.

The relevant interfaces in thermal stability studies are often buried by tens of nanometers of e.g. the gate stack. Traditional laboratory characterization for chemical reactions is governed by mass spectroscopy (i.e. ToF-SIMS), electron microscopy techniques (i.e. EELS/EDX), and the aforementioned photoemission spectroscopy. These methods are either spatially limited (e.g. electron microscopy) or depth limited by the mean escape depth of electrons from the sample (XPS). Material removal by sputtering and subsequent analysis can increase the depth resolution of XPS and is the working principle of ToF-SIMS. Nevertheless, a diffusion profile is thereby convoluted with sputter effects (preferential sputtering especially at layer interfaces), which lead to a virtual layer broadening. The idea behind HAXPES analysis of buried surfaces is to provide depth information in the solid by varying the photon energy or by the choice of an orbital at different binding energy-a clear asset of synchrotron radiation sources.

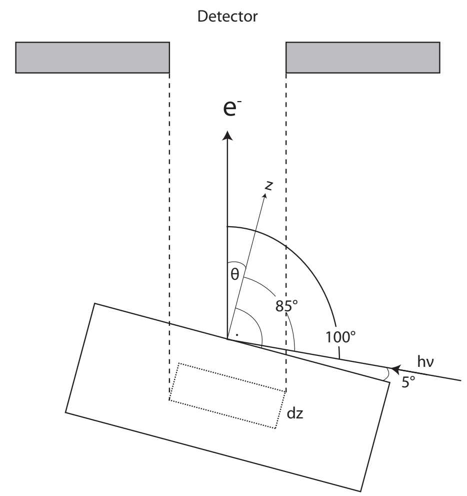


Fig. 3.4 Angle setup in the HAXPES experiment. The angle θ between surface normal and detector is 15°, the angle between beam and detector is 100°.

3.2.3 Effective attenuation length in $LaLuO_3$ for hard X-rays

3.2.3.1 Theoretical background

The study of buried surfaces by means of varying photon energy needs to be converted to a depth profile on the length scale. The effective attenuation length (EAL) of a solid is an experimentally obtained quantity, which relates the kinetic energy of the emitted electrons in PES to the information depth (nm) is originating. Thereby an approximate exponential decay is assumed.[107,111] The practical EAL decay covers thereby both elastic and inelastic processes. In practice, the EAL can be obtained[107,112,113] by measuring the intensity I_{layer} from a thin layer of known thickness t and referencing it to its bulk intensity $I^{\infty}_{\mathrm{layer}}$ by

$$I_{\mathrm{1ayer}} = I^{\infty}_{\mathrm{1ayer}}(1 - \exp(-t/(EAL\cos\theta))), \tag{3.29}$$

with the angle θ between surface normal and analyzer, $\cos\theta$ projecting the EAL on the surface normal. Alternatively, but not used in this work, one can also obtain the EAL from the intensity I_s of a substrate buried by a layer of known thickness t and again referencing it to its bulk intensity I^{∞}_s as

$$I_s = I^{\infty}_s \exp(-t/(EAL\cos\theta)). \tag{3.30}$$

The amount of electrons $N(z)$, passing through a layer of thickness z without inelastic scattering follows[107] the Poisson distribution and can thereby be described as

$$N(z) = N_0 e^{-\frac{z}{\lambda}}. \tag{3.31}$$

Thereby N_0 describes the incoming electron flux. Consequently, mean value and standard deviation (i.e. 2^{nd} moment) of the distribution are λ, which is called the Inelastic Mean Free Path (IMFP) in a solid. Hence, ISO[111] describes the IMFP as the "average distance that an electron with a given energy travels between successive inelastic collisions". Comparing the definitions of IMFP and EAL, the directly accessible quantity is the EAL and it should lie below the theoretically obtained values for the IMFP.

A third term, the information depth (ID), e.g. the depth 95% of information is originating,[111] can be interpreted with the EAL: As the EAL decays approx. exponentially, 95% of the signal are obtained at a depth $z = -\ln(0.95)$ EAL.

3.2.3.2 Experimental EAL-determination in $LaLuO_3$

The effective attenuation length in $LaLuO_3$ was determined by the bulk/thin layer method[114] between 7 and 13 keV kinetic energy. The photon energy was chosen as 10, 12, and 14 keV and the intensity of several orbitals for La, Lu, and O were measured.

$LaLuO_3$ was deposited by pulsed-laser deposition (PLD) on Si from a stoichiometric target under ambient oxygen. Prior to deposition, the low-doped Si substrates were cleaned with a standard RCA clean, leaving a thin chemical oxide on the Si substrate (1 nm). Two samples were prepared: A thin film with 12.4 nm thickness (determined by XRR) and a bulk layer of 124 nm. The mass density for the PLD grown films was $\sim 8.62\,\mathrm{g/cm^3}$ as also found for samples deposited by molecular beam deposition (MED).[115] No thermal treatment was applied.

HAXPES measurements were carried out at the CRG Spanish beamline (SpLine) at the European Synchrotron Radiation Facility (ESRF), Grenoble. The kinetic energy of the incoming photons were chosen as 10, 12, 14 keV. At 14 keV, no Si substrate signal could be recorded for the thick sample. Thus, bulk properties can be assumed for this film at the chosen photon energies. The spectra were measured by a high voltage cylindrical sector analyzer.[110] Focus CSA 300/15 and calibrated against the chemical shift of the Si ls signal from Si substrate, visible for the thin sample (binding energy $\mathrm{BE} = 1839.2\,\mathrm{eV}$[116]). The Gaussian peak width is given by the beam line optics and corresponds to 1.5 eV, 1.8 eV, and 2.1 eV for 10, 12, and 14 keV, respectively.

The intensities for bulk material and thin overlayer on Si were recorded together with the photon flux of the beamline. Subsequently the measured spectra were normalized to the mean beamline current. Spectra obtained during ring injection were discarded. Due to the same transmission and photon cross section, bulk and overlayer spectra were not needed to be normalized any further. The effective attenuation length can then be determined[113] energy-dependent from the intensity ratio $I_{\mathrm{overlayer}}/I_{\mathrm{bulk}}$ of overlayer and bulk film as given by (3.2):

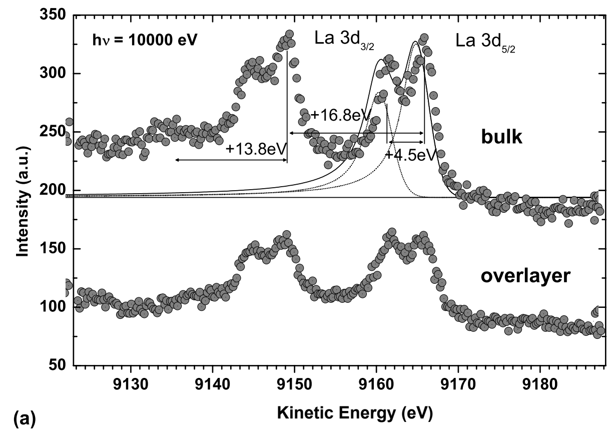
$$EAL = \frac{t_{\mathrm{overlayer}}}{\ln\left(1 - \left(\frac{I_{\mathrm{overlayer}}}{I_{\mathrm{bulk}}}\right)\right) cos\theta} \tag{3.32}$$

where $t_{\mathrm{overlayer}}$ is the thickness of the thin overlayer and $\theta = 15^\circ$ is the collection angle formed between the analyzer and the sample normal. Error analysis for the peak fits was done by Monte-Carlo simulation within the software CasaXPS 2.3.17. Deviations from a pure Poissonian noise distribution due to the $2d$ event counting system were thereby neglected. Error estimation for the EAL values was done by subsequent application of the Gaussian propagation law.

The measured lanthanide orbitals show strong plasmon features. The intensity increases for the bulk material suggesting both surface and bulk plasmon contributions. The La $3d$ orbitals possess strong satellites shifted by 4.5 eV with intensities of 70% of the main peaks, as displayed in Figure 3.5a. Similar structures were before observed for La_2O_3 by Teterin.[117,118] For higher photon energy satellites and peaks were found to merge due to the broader energy distribution. La $3d_{5/2}$ and La $3d_{3/2}$ are separated by 16.8 eV, comparable to results found for La_2O_3 by Sunding.[119] Increasing plasmon contributions are visible at lower and higher binding energy if comparing bulk and overlayer samples. These loss structures were previously described by Crecelius *et al.*[120] and increase for bulk material compared to thin layers. One finds low energy satellites in the peak onset and bulk plasmon features around $+13.8$ eV apart from La $3d_{3/2}$. These plasmon features most prominent for the bulk material lead to an increasing background signal between La $3d_{5/2}$ and La $3d_{3/2}$ exacerbating the background subtraction during peak fitting.

The best agreement of EAL value with the theoretical predictions[121–123] of the TPP-2M formula (Tanuma, Powell, Penn[122,124]) were found when a Doniach-Sunjic (DS) profile was used for peak fitting (cf. Figure 1) as suggested by Crecelius *et al.*. A Shirley background type with a pure Voigt profile lead to random scatter in the EAL prediction most likely dependent on the loss structures as described previously. This was also the case if the Gaussian width determined for the beamline optics in combination with a Lorentzian width as tabulated[125] was used. The calculation ignored thereby the loss structure by the choice of the DS fit. As the DS profile is only asymptotically vanishing its use becomes problematic. To circumvent these limitations a convolution of a Voigt-type and DS profile was employed. This profile type results in a much faster decay. Moreover, the deconvolution of bulk and overlayer intensity leads to comparable decays in the shown case, so that the error imposed by the profile is small in comparison to the noise in HAXPES measurements. The deconvolution resulted in a main peak for La $3d_{5/2}$ at a binding energy (BE) of 834.4 eV.

Lu $3d_{5/2}$ is shown in Figure 3.5b together with a DS-type fit. The absence of a satellite structure is attributed to the trivalent configuration Lu^{3+} ([Xe] $4f^{14}$) with fully occupied f-shell. Also less shake-down[120] background is observed. Here, a constant background signal at lower binding energy was chosen for background



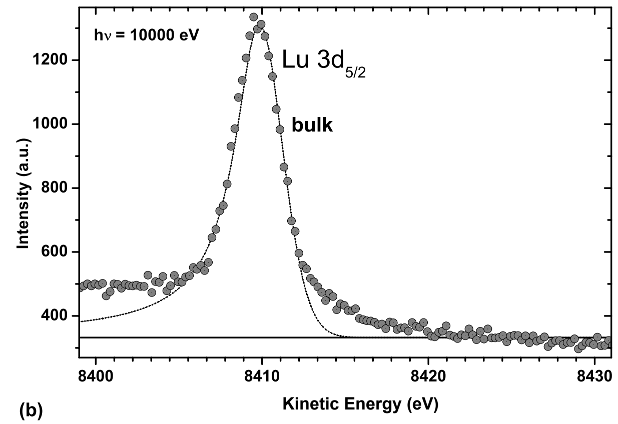


Fig. 3.5 (a) Bulk and overlayer intensity of La 3*d* in $LaLuO_3$. A strong satellite is present at higher binding energy (keV) accompanied with satellite loss at low binding energy. Plasmon loss is increased for the bulk sample. A Doniach-Sunjic profile convoluted with a Voigt profile was used for deconvolution. (b) Lu $3d_{5/2}$ shows much less loss compared to La. No satellite structure was observed due to the completed electron configuration. FWHM is 3.2 eV.

subtraction, thereby revealing the low energy loss structure. The employed constant background depends on the accumulated signal during measurements. Fluctuations

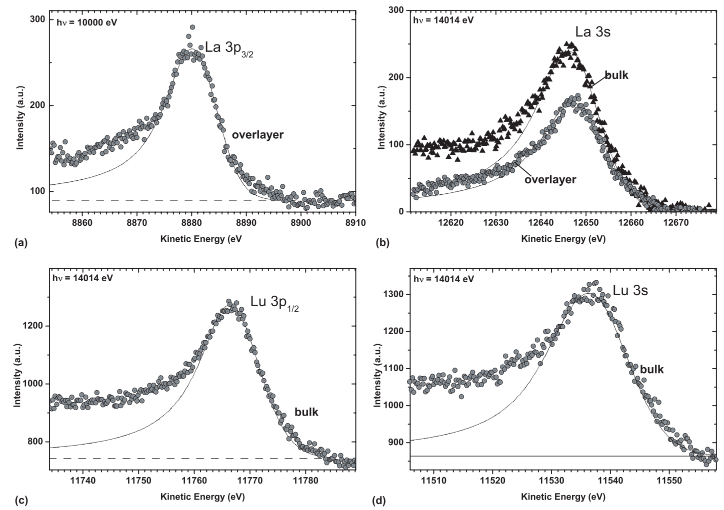


Fig. 3.6 Selection of oxide orbitals collected during EAL determination. (a) La $3p_{3/2}$ for a thin layer (PE = 10 keV). This orbital shows loss structures to lower and higher binding energy. FWHM is 10 eV for the Voigt profile of the fit. (b) La 3s comparing bulk and overlayer loss structure after constant background subtraction: Higher binding energy loss structures for bulk material are much more pronounced than for the overlayer and increasing with photon energy (14 keV). FWHM is 16.3 eV. (c) Lu $3p_{1/2}$ shows a similar structure as observed for La (FWHM = 12 eV). (d) Lu 3s is a broad peak with distinct loss structure at higher binding energy (PE = 10 keV, FWHM = 14 eV).

in ring current or injections can thereby lead to artificial patterns or intensity drops. A thorough inspection of every single scan must be employed in advance.

A selection of peaks used for determination of the EAL is displayed in Figure 3.6. All structures show a common loss structure at lower binding energy[120] and have a skewed structure due to the plasmon loss. The loss structure observed is mainly due to bulk plasmon loss as observed when comparing measurements for bulk and overlayer (Figure 3.6b). Thereby the determination of the correct peak area is strictly dependent on the loss structure and subject to error. The best results were achieved if the constant background could be determined precisely. Likewise, the energy range had to be set accordingly wide to screen this background (10–20 eV before rise of the first energy loss). However, for very broad loss and peak structures as in Figure 3.6d the setup of the appropriate energy window is difficult and limited by scheduled beam time.

The resulting effective attenuation length is plotted in Figure 3.7 and summarized in Table 3.1 together with the peak positions found. As mentioned, the error bars are based on an error calculation for the peak models based on a Monte-Carlo simulation of possible peak models. Errors between 9 and 40 Å were found. The

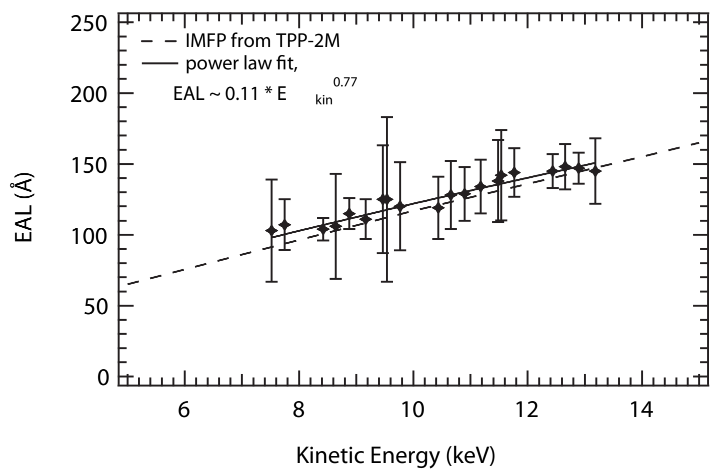


Fig. 3.7 Effective attenuation length for $LaLuO_3$ as determined by the depth profile method for 7 to 14 keV. The result is in agreement with the theoretical predictions of the TPP-2M model. Error bars are obtained by the Gaussian propagation law on the bases of peak fit and experimental errors (analyzer angle, XRR thickness). An error calculation based on min/max or standard deviation of the fitted values would clearly underestimate the errors due to deconvolution.

EAL was fitted by a power law EAL $\sim 0.11 E_{\text{kin}}^{0.77}$. This result is in close agreement with the theoretical prediction for the inelastic electron mean free path (IMFP) obtained by the TPP-2M formula. The IMFP should exceed the EAL in the solid.[107] The complicated loss structures in La and Lu leads to uncertainty in the fit limits for the background function. Likely, this uncertainty causes systematic deviation and let both prediction and experiment collapse with large scatter. The most accurate way to prevent such deviations would be an extended energy scan width. Thereby one could clarify the real background function in respect to the loss structure. However, since time slots at synchrotron beamlines have to be tightly packed with experiments, here a bigger uncertainty on the background function seems reasonable.

The uncertainty in EAL determination was higher for O 1s due to the overlayer signal consisting both of SiO_2 from the substrate. Lu 3s could also be determined with higher uncertainty due to the broad loss structure at lower binding energy (max. ± 58 Å). A possible solution for O 1s would be the growth of amorphous $LaLuO_3$ on an oxide-free substrate (i.e. not Silicon). Further improvements are likely possible for the overlayer method,[113] if orbitals with a large cross section and less electron loss structures would be used. However, the second requirement seems to be more difficult for La than for Lu because of the observation of loss structures in every orbital.

Table 3.1 Summary of the peak positions and EAL values found by using the depth profile method.

	Orbital	$h\nu$(eV)	E_{bind} (eV)	EAL(Å)	ΔEAL(Å)
Lu	3s	10000	2494.9	103	36
Lu	$3p_{1/2}$	10000	2265.6	107	19
Lu	$3d_{5/2}$	10000	1590.3	104	9
La	3s	10000	1371.3	106	37
La	$3p_{3/2}$	10000	1128.4	115	11
La	$3d_{5/2}$	10000	834.3	111	15
O	1 s	10000	530.3	125	38
Lu	3s	12017		126	58
Lu	$3p_{1/2}$	12017		120	31
Lu	$3d_{5/2}$	12017		120	22
La	3s	12017		129	24
La	$3p_{3/2}$	12017		130	19
La	$3d_{5/2}$	12017		135	19
O	1s	12017		139	30
Lu	3s	14017		143	32
Lu	$3p_{1/2}$	14017		145	17
Lu	$3d_{5/2}$	14017		146	13
La	3s	14017		148	17
La	$3p_{3/2}$	14017		147	12
La	$3d_{5/2}$	14017		145	24

3.2.4 Thermal stability of the TiN/$LaLuO_3$ gate stack

3.2.4.1 Experimental setup

The analyzed gate stack consists of 11 nm $LaLuO_3$ dielectric layers deposited by molecular beam deposition (MBD) at 450°C[22] and 16 nm TiN_x metal layers deposited by physical vapor deposition (PVD). Prior to deposition, the Si (100) substrates were RCA cleaned which forms a thin silicon chemical oxide ($\sim$ 1 nm). The metal layer stoichiometry was determined by Rutherford backscattering spectrometry in channeling direction (RES/C) to reduce the Si background signal.[126] RBS was performed with 1.4 MeVHe^+ at an incident angle of 0° and a scattering angle of $-170°$. A stoichiometry of Ti:N $\approx 1:0.95$ was determined. The layer thickness was determined by XRR.

To study the interface reactions in the gate stack, the samples were annealed in a rapid thermal processing (RTP) system under N_2 ambient in the temperature range $400-1000$°C (post deposition anneal (PDA)). For electrical characterization, before MOS capacitors fabrication similar samples are annealed at 1000°C for 5 or 60 s. Using a 200 nm Al hard mask the TiN is removed by reactive ion etching stopping on the high-κ layer. Forming gas annealing (FGA) in 90% $N_2/10\%H_2$ at a temperature of 450°C for 10 min was carried out after MOS capacitor fabrication.

For the HAXPES experiments, the kinetic energy of the incoming photons was 10, 12, and 14 keV at pass energy of 100 eV. The spectra, measured with a high voltage cylindrical sector analyzer Focus CSA 300/15[110] were calibrated against the chemical shift of the N 1s signal in TiN (binding energy BE = 397.0 eV[116]). The incident angle (beam respect to sample surface) was 5 degrees, hence, take-off angle (beam angle respect to sample normal) was 85 degrees, and beam respect to analyzer was $85+15=100°$. The intensity was corrected for measurement intensity, photon flux and beam line transmission.[127] The calibration was confirmed by the signal position of the Si ls orbital from the Si substrate at 1839.2 eV.[116] The Gaussian peak width is given by the beam line optics and corresponds to 1.5 eV, 1.8 eV, and 2.1 eV for 10, 12, and 14 keV, respectively. For peak fitting, a convolution of Lorentzian and Gaussian peak contributions was used. Thereby the beamline optics are the limiting factor for the identification of features in the deconvolution. The Gaussian width as determined[113] was held fixed and the Lorentzian peak width was chosen constant where lifetime broadening was known.[125] The spectra were normalized with respect to photon flux, the photoionization cross sections, and detector transmission.

After HAXPES analysis, the unstructured stacks were analyzed by XPS, time-of-flight secondary ion mass spectrometry (ToF-SIMS), RBS, and high-resolution TEM (HRTEM).

3.2.4.2 HAXPES results

An advantage of the HAXPES method is the ability for excitation of large cross section's orbitals, like Si 1s, to screen the chemical state of buried interfaces. This allows an efficient information access, because many observed reactions could already be identified by the chemical shift in the reactant Si.

Si 1s orbital analyses

The silicon Si ls spectra of the as-deposited and annealed sample obtained by HAXPES at 14 keV are presented in Figure 3.8: HAXPES Si ls spectra of the as-deposited and different annealed samples for photon energy of 14 keV. All spectra were normalized to the as-deposited sample intensity of $Si^{\pm 0}$ and its contribution was then subtracted from the annealed sample spectra. No difference could be obtained between PDA 600°C/60s and the as-deposited sample. At 800°C, the first interface change into silicate is visible. All spectra were normalized to the as-deposited sample intensity of $Si^{\pm 0}$, which was subsequently subtracted from the spectra. Specifically, this means that the resulting plots emphasizes the change with respect to further thermal treatment. For the as-deposited stack neither Si chemical oxide nor a peak asymmetry is visible. The large intensity for $Si^{\pm 0}$ indicates an information

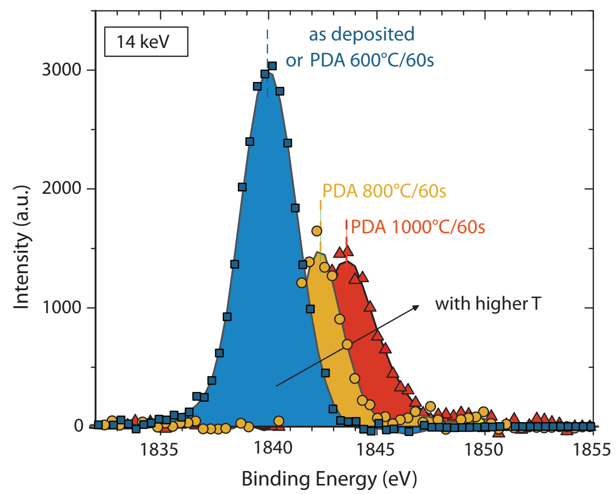


Fig. 3.8 HAXPES Si 1s spectra of the as-deposited and different annealed samples for photon energy of 14 keV. All spectra were normalized to the as-deposited sample intensity of $Si^{\pm 0}$ and its contribution was then subtracted from the annealed sample spectra. No difference could be obtained between PDA 600°C/60 s and the as-deposited sample. At 800°C, the first interface change into silicate is visible.

depth (ID) much larger compared to the information which comes from the thin interfacial layer (RCA clean substrate, $t_{SiO2} = 1\,nm$). Thereby the ID is the specimen thickness from which a specified percentage, e.g. 95%, of the detected signal originates.[127]

The reaction between Si and $LaLuO_3$ and the final reaction products are strongly dependent on the annealing duration and temperature. For PDA between 400°C and 600°C and up to 60 s no additional silicate formation could be detected within the measurement resolution. The nomenclature for the formation of silicates was chosen as in literature.[128]

First at 800°C for 60 s Si-rich silicate bonds could be observed in the Si 1s orbital spectra,[128] whereby the major contribution can be identified as La-rich silicate (dashed line at $\sim 1841.5\,eV$). Finally, a 60 s PDA at 1000°C leads to the formation of an Si-rich compound, positioned at 1844 eV, most likely Si- and O-rich compounds, e.g. $La_2Si_2O_7$, and a La-rich silicate like La_2SiO_5 ($1839\,eV < EE < 1844\,eV$), less pronounced than at lower temperatures.

The integrated Si intensity of the annealed samples (spectrum of elemental Si and oxidation states) becomes larger compared to the signal of the as-deposited sample. Due to the constant ID in the sample, one can argue that Si diffuses in reacting with the high-κ layer above.

La $3p_{3/2}$ orbital analyses

In Figure 3.9 the HAXPES spectra of the La $3p_{3/2}$ orbitals, taken at 14 keV, before and after annealing are plotted.

Directly after deposition, several binding states are attributed to Si ls and Lu 3d orbitals by peak deconvolution: La-O at-1126.0 eV, a La exchange satellite (La-O sat) at-1129.0 eV and La-O-Si at 1132.6 eV. In addition, plasmon features are observed at $+12.5$ eV from the main peaks. The chemical shift indicates the reaction with a reactant of higher electronegativity (EN), like Si and Lu (Pauling $EN_{Si} = 1.9$, $EN_{Lu} = 1.27$, $EN_{La} = 1.1$) demonstrating the silicate formation at the high-κ/Si interface.

A further partition in several silicate contributions as indicated for Si 1s cannot be assessed here due to the aforementioned Gaussian peak width. Therefore, the La silicate signal is assumed as a single contribution. Accordingly, the BE is constraint within $3\,\mathrm{eV}(1841-1844\,\mathrm{eV})$ to account for different amounts of bridging and non-bridging oxide,[128] BO and NBO, respectively. The upper limit was chosen with the reference value for SiO_2.[116] A change in BE marks then a change for NBO to BO. Lower BE is interpreted as La-richer silicate and higher BE as oxygen and Si-richer SiO_2-like bonds. Please note also that due to the higher Gaussian FWHM for hard x-ray optics, the peaks in our measurement added up to one broad convolution. Moreover, the spectra for $LaLuO_3$ are more difficult to deconvolute than for pure Lanthanide, Ln_2O_3, because of the slight stoichiometric variations within the layer with different EN for La and Lu.

According to Teterin *et al.*[117,118] and Crecelius *et al.*[120] the occurrence of the satellite La-O-Ln in the La orbital spectra can be ascribed to many-body perturbation, i.e. shake-up processes. Missing charge transfer over the NBO in La-O-Si explains the absence of a satellite in the La-O-Si bonds. The measured satellite position ($\Delta E_{\mathrm{sat}} \approx 3\,\mathrm{eV}$) is within the expectations of Teterin's study for lower BE orbitals.

After annealing, the peak intensities change clearly and the peak convolution broadens. The observed contribution at BE-1121.3 eV, Whose intensity is less than 10% of the total La signal, indicates the occurrence of either additional shake-down processes, as described by Teterin *et al.*[117] for La 4p or metallic La-bonds. The significance of this peak has to be investigated by complementary measurements and is highly depending on background subtraction. For the annealed sample, the peak deconvolution revealed La-O and La-O-Lu bonds with $a + 0.8$ eV higher BE compared to the as-deposited samples. This energy shift is correlated with a reduced amount of La-O bonds per atom in favor of increasing atomic bond density to Si (La-O-Si). Thereby the electron bonds are closer correlated to La, resulting in higher BE for the replacement of La-O-La by La-O-Si bonds.

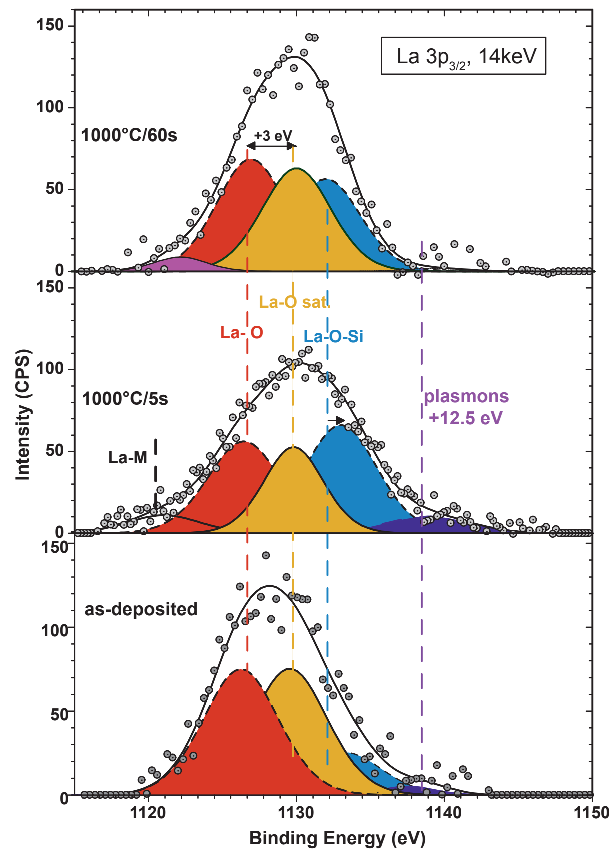


Fig. 3.9 HAXPES La $3p_{3/2}$ spectra of the as-deposited and annealed samples for photon energy of 14 keV. The following contributions are identified: La O at 1126.3 eV, a satellite La O sat. at 1129.3 eV and La O Si at 1132.5 eV plus plasmon features at +12.5 eV away from the main peaks. The as-deposited sample shows ∼1.4 eV lower BE for La-O and La-O-Lu bonds.

The silicate peak at 1132.8 eV for the 1000°C/5s anneal is found within 0.2 eV at the same position as the non-annealed sample (1132.6 eV). This peak contribution increased after annealing in parallel to a reduction of La-O bonds. After a longer anneal (1000°C/60 s), the silicate component is lowered in BE to 1132.0 eV, which may be explained by a larger amount of La-rich silicate in respect to SiO_2-like bonds.

Lu $3d_{5/2}$ orbital analyses

Lu-O and Lu-O-Si are present already after deposition. Silicate formation at the interface Si/$LaLuO_3$ is proved by the additional peak component at higher BE. The boundary silicate formation from Si orbital signals was also observed in prior annealing studies.[22] Here, Lu $3d_{5/2}$ orbital signal (Figure 3.10) of the as-deposited samples consists of a main contribution from Lu-O at $1589.9\,\mathrm{eV}(\pm 0.5\,\mathrm{eV})$ and a minor silicate contribution (11%) at 1591.9 eV.

After annealing a variety of additional peaks are detected: For the 1000°C anneal for 5s a clear trend to form more Lu silicates and orthosilicates is present (Figure 3.10, dotted lines): the component at 1591.8 and the one at 1593.8 eV are identified as silicate and orthosilicate complexes, respectively. The orthosilicates refer to a silicate containing the group SiO_4 in which the ratio of silicon to oxygen is 1 to 4, e.g. $Lu_{10}(SiO_4)_6O_3$. Interestingly, these peak contributions are positioned at rather high BE (Figure 3.10). The oxygen rich complex shifts the BE to even higher values compared to normal bridging oxygen between Ln-O-Si. These (Si-rich) Ln-orthosilicates complexes are claimed by other authors to form high quality silicate interfaces between Ln oxide and silicon layer as referred to La-rich silicates.[129] Despite, in La no orthosilicates were detected, possible due to the broad (Lorentzian) lifetime for the La 3p hampering the deconvolution for hard X-rays.

A minor peak contribution (6% of the integrated signal) at 1595.3 eV is only present for the 5s anneal. The additional component is interpreted with Lu-O-Ti bonds due to the argument of different electronegativity, which vanish again for the full silicate formation after 60 s anneal.

However, the peak position of the orthosilicates phase can also be identified via Ln-O-Ti bonds based on their electronegativity. A separation of Si or Ti to O bonds is difficult because of reduced energy resolution at high energy. Only for the 1000°C/5s sample a minor contribution at 1586.7 eV for intermetallic bonds (Lu-M) is indicated. As for La, the Lu-M contribution is significant for the 1000°C/5s anneal.

After 1000°C anneal for 60 s the Lu-O component decreases compared to the as-deposited sample. Thereby, the orthosilicate compound at 1593.8 eV decreases in density and is replaced by a strong silicate contribution, located at 1591.8 eV. The difference to the 5s anneal is, therefore, significant: The Lu-O contribution is screened already at 10 keV (not shown here) and increases only slightly with ID. Despite this little contribution left from the oxide, a large amount of Lu compound is transformed into a silicate. The ID in the material stack remains the same for fixed photon energy (cf. Figure 3.10). Therefore, this transformation is equivalent with an up-diffusion of Si into the Ln-oxide. In former studies the amorphous structure in $LaLuO_3$ was shown to be maintained during high temperature anneal.[130] The silicate formation is an important second parameter to control in gate stack engineering.

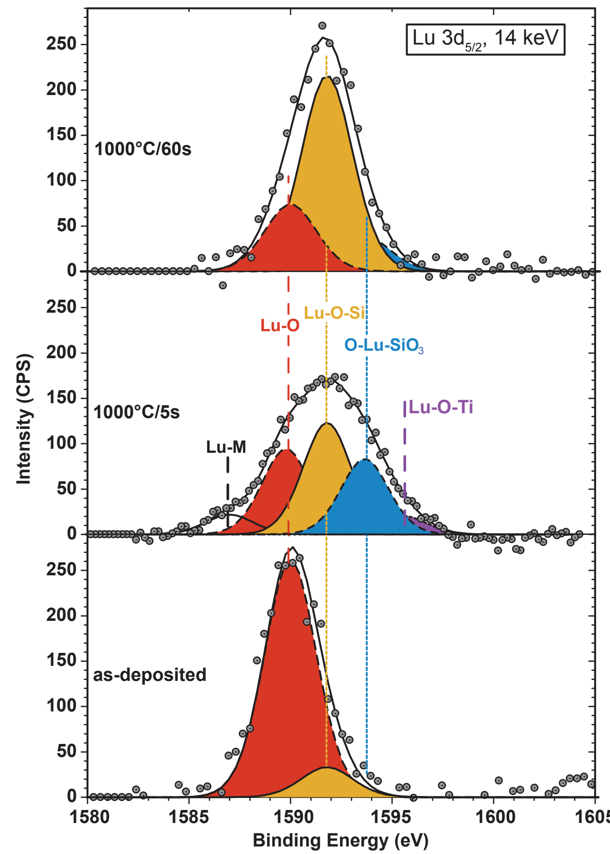


Fig. 3.10 HAXPES Lu $3d_{5/2}$ spectra of the as-deposited and annealed samples for photon energy of 14 keV. The following bonds are identified: Lu = O at 1589.9 eV, Lu-O-Si at 1591.8 eV. After annealing additional binding states occur. Those are attributed to orthosilicate Lu-O-SiO3 at 1593.8 and 1595.5 eV [31]. Within the scan range of 1580–1600 eV no clear plasmon features were detected.

However, based on the higher relative permittivity of the silicates compared to SiO_2, it has been shown that the control of the silicate formation can lead to low EOT and high capacitances.[99,128,129,131] For Lu $3d_{5/2}$ no significant plasmon features were found within 40 eV scan range.

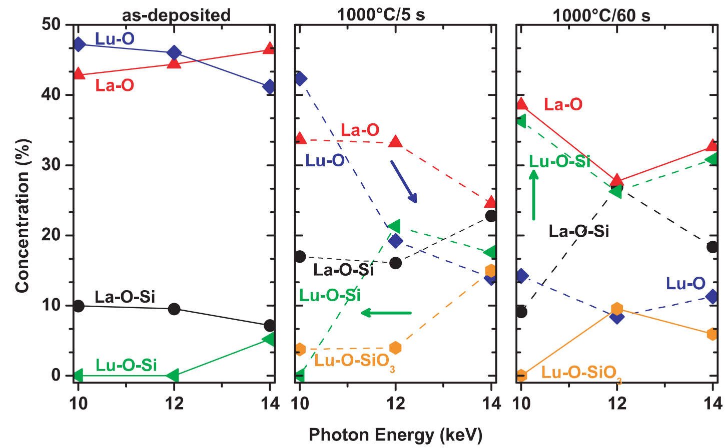


Fig. 3.11 HAXPES depth profile vs. EAL using the integrated peak areas obtained by deconvolution of the orbitals. For the as-deposited sample, La silicate is found through the layer and Lu silicate only at the interface. The amount of La silicate increases during annealing. More pronounced is the up-diffusion and reaction of silicon to form Lu silicate. Based on this qualitative analysis the diffusion of Si seems to be mainly promoted by the generation of Lu-O-Si bonds and less by La bonds.

HAXPES qualitative depth analyses

A non-destructive but qualitative depth profile analyses can be obtained by varying the incoming X-ray energy and applying the orbital deconvolution approach (Figure 3.11). The integrated peak areas of Figure 3.9 and Figure 3.10 are compared on a relative scale. The information depth on the x-axis was obtained from the EAL fit of $LaLuO_3$. The ID differs slightly for La3$p_{3/2}$ and Lu $5d_{5/2}$ due to the difference in binding energy.

The energy shift observed for La silicate was summarized as a single silicate formation to increase readability. Remarkably, the as-deposited sample contains already La silicate while Lu silicate is only present at the interface to Si (i.e. detected only at 14 keV photons energy). For the 5s annealed sample, the La silicate amount increases but remains homogeneously distributed. More important, Lu silicate contribution increases rapidly from the substrate into the gate oxide layer on the cost of a reduced amount of Lu-O bonds.

Comparing the 5s and 60 s annealing the orthosilicates observed for Lu decreased again (cf. also Figure 3.10). A maximum of only 13% Lu-O bonds density remains after longer temperature treatment. Both Lu and La silicate compounds density increases along with a reduction of La-O bonds. However, at the upper interface $(\mathrm{LaLuO_3}/|hboxTiN)$ a higher La oxide signal compared to La silicate is obtained:

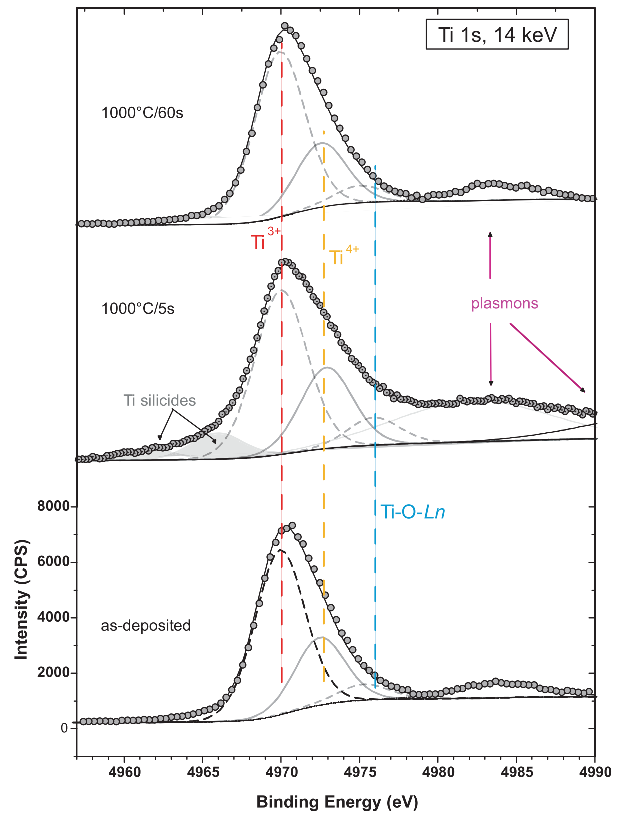


Fig. 3.12 HAXPES Ti 1s spectra of the as-deposited and annealed samples at photon energy of 14 keV). The as-deposited sample shows contributions from Ti^{3+}(TiN) and Ti^{4+}(TiO)$_x$. The long time between sample preparation and beam line experiment has led to oxidation of the TiN layer (Shirley background subtraction, ratio 64:32% for TiN:TiO_x). A minor contribution at higher BE than Ti^{4+} could be identified due to the peaks asymmetry.

the La silicate accumulates in the gate oxide layer in contrast to Lu silicate, which accumulates at the top and bottom interfaces. The reduction of Lu oxide to Lu silicate is likely the driving force behind the up-diffusion of silicon. This effect is less pronounced for La and, for long annealing, the relative amount of La oxide increases again at the top and the bottom interfaces.

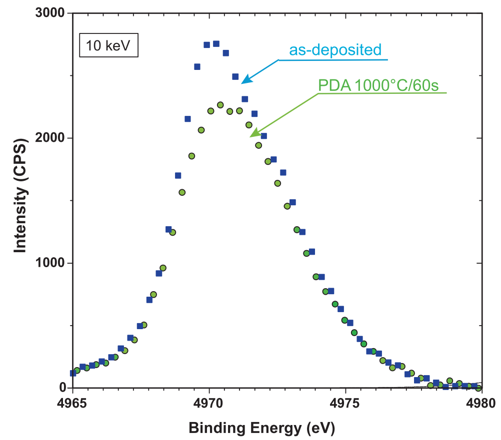


Fig. 3.13 HAXPES Ti 1s spectra (background subtracted) after 1000°C/60 s anneal and as-deposited for photon energy of 10 keV. The lower information depth at 10 keV compared with 14 keV leads to lower intensities after annealing meaning Ti diffusion into the gate stack.

Metal gate analyses and Ti diffusion study

The reactions taking place in the top TiN layer have been investigated by HAXPES via the Ti ls orbital. The normalized total intensities are shown in Figure 3.12 for photon energy of 14 keV. The total intensity of Ti 1s after annealing equals, within the measurement errors, the intensity before annealing. For the samples undertaking a 1000°C for 5s annealing an ex-situ 2 nm Si-capping was deposited prior annealing to prevent additional TiO_x formation on the surface. However, this ex-situ treatment does not prevent the prior oxidation of TiN. At the time of this experiment, in-situ deposited Si capping of TiN was not available, but was upgraded for PVD and AVD later. The comparison of the spectra in Figure 3.12 shows that no further oxidation during annealing takes place, for the present stoichiometry of TiN. A percentage ratio of 64 to 32 for TiN to TiO_x compounds was extracted by XPS for the uncapped, as-deposited sample. Since this large amount of TiO_x is already present prior annealing, no reduction of $LaLuO_3$ or interfacial SiO_2 (scavenging effect) is observed and would even be thermodynamically unfavorable for La reduction by Ti.

A minor Ti-O-Ln bond density is present at 4975.5 eV for 14 keV. This binding state amounts to 6–7% before and after annealing but not significant to prove diffusion of TiN into $LaLuO_3$ layer. Here, the observed Ln-O-Ti peak in Ti 1s differs in magnitude from observations in La and Lu. Since the information depth and effective attenuation length for Lu3d, La3p, and Ti ls differ, the bonds at the upper interface may be more significant in the Ti ls peak. For the 1000°C/5s sample the background signal was measured up tO 5115 eV to incorporate the role of the plasmon features on peak deconvolution. Two huge plasmon features at 4982 eV and 4995 eV at a FWHM = 11.6 eV were detected. The incorporation of the additional information leads to similar peak fits besides a minor shift of the Ti-O-Ln peak from 4975.5 eV (as deposited) to 4975.9 eV. This deviation remains acceptable on base of the wider definition of the background function and the peak width (FHWM $\sim$ 5.3 eV). Thus, future experiments can be limited to a scan range within the plasmon decay.

Comparing the Ti ls spectra at 10 keV (Figure 3.13) an overall intensity decrease by 9% is observed for the annealed sample, originating from an intensity decrease of the original Ti^{3+} peak of TiN. Therefore, a slight diffusion of Ti deeper into the oxide layer can be attested. The slight shift to higher BE present, is attributed to more Ti = O binding states after annealing. It is worth to note that, neither at 10 keV nor at 14 keV additional peaks arose from an interaction of Ti with La or Lu. Likewise, more oxygen is bond to Ti but few bindings to Ln are present. A limitation for the further quantification is the use of a Shirley background and the consideration of the large plasmon features in Ti(N). The Shirley background subtraction is still under debate for Ti[132] but used in the absence of better solutions.

3.2.4.3 ToF-SIMS and RBS analyses

ToF-SIMS measurements were also performed and compared with the HAXPES measurements (Figure 3.14). At increased temperature, TiN is found to diffuse slightly into $LaLuO_3$. This is consistent with the HAXPES analysis. Regarding the $LaLuO_3$ layer, the SIMS spectra indicate a broadening of the initial $LaLuO_3$ thickness. The La and Lu signals (blue diamonds and green crosses, respectively) do not change their homogenous distribution through annealing. The silicate formation is supporting the strong Si diffusion into the oxide layer, proved by the $^{30}Si^-$ signal change. Comparing the $^{30}Si^-$ signal with the above HAXPES results, a temperature dependent Si up-diffusion into the oxide layer accompanies silicate formation through the whole oxide layer after 60 s annealing. This Si diffusion stops at the interface to TiN. For Ti-rich TiN, it could be shown by EELS that Si diffusion even continues in TiN due to nitrogen deficiency.[31]

RBS measurements obtained from equally processed samples are presented in Figure 3.15. After annealing at 1000°C/60 s, broadening in the signals from La and Lu supports the thickness increase of the oxide layer. Moreover, the signal intensity decrease is explained by a change of the stoichiometry by adding Si in the oxide layer. A change in the silicon signal due to silicate formation could not be proved by either simulation or measurement due to overlapping with the strong Si substrate signal. This limitation is due to the scattering of He^+ ions deep in the substrate. For the 1000°C/5s anneal and lower temperature processes the stoichiometry of the stack determined by RBS stays almost constant, i.e. the effects evaluated in HAXPES cannot be further quantified by RBS. The initial stoichiometry was determined to be La:Lu:O = 1 : 1 : 2.9.

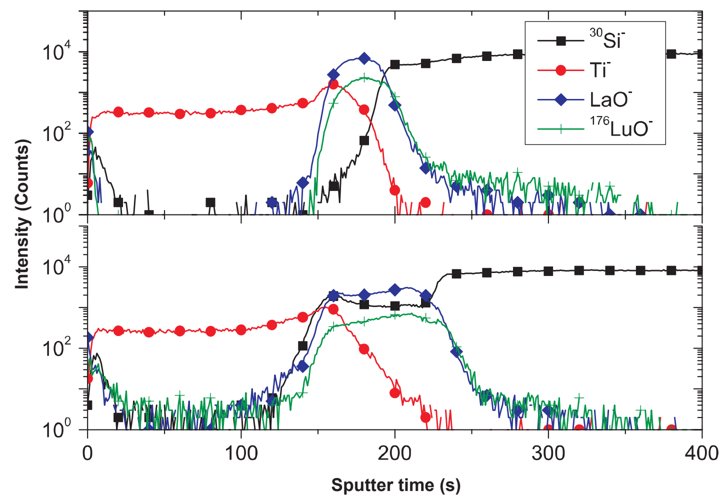


Fig. 3.14 ToF-SIMS measurements of the (top) as-deposited and (bottom) 1000°C/60 s annealed sample. A clear plateau of $^{30}Si^-$ (full squares) is apparent after RTP, whereby the Si diffusion stops at the interface to TiN. A slight intermixing of Ln silicate into Ti(N) is present. Thereby, the Ti(N) signal remains unchanged within the measurement accuracy of ToF-SIMS. Since the sputter rate was held constant between both measurements the silicate formation into $LaLuO_3$ leads to a significant thickness increase.

3.2.4.4 Depth profiles comparison of HAXPES and XPS analyses

The comparison of XPS depth profiles of the differently treated sample was performed using Ar sputtering. The resulting depth profiles are presented in Figure 3.16. The TiN stoichiometry of the as-deposited sample (Figure 3.16a) does not match the RBS calibration (1:0.95) even by adding the intensity of the post-deposition oxide (∼1:0.83). Similar disagreement was also reported in the literature[126] and can be explained by the complicated satellite and plasmon loss structure in Ti 2p. Additionally, gaseous species like N and O tend to be preferentially sputtered. For the energy range 450 to 500 eV or 450 to 470 eV both Shirley and Tougaard background functions, used in spectra analyses, violate the quantum mechanical intensity ratio of Ti $2p_{1/2}$ and Ti $2p_{3/2}$.[132] To solve this problem, a Shirley background region between 450–500 eV is applied and the background intensity is adjusted to match the peak intensity ratio of 2:1. The obtained background region was truncated to match the area of interest (Figure 3.17). Plasmon loss features were fitted at the energies of 470 eV and 485 eV. Especially close to the sample surface, additional surface plasmons were detected which have to be modeled to match the

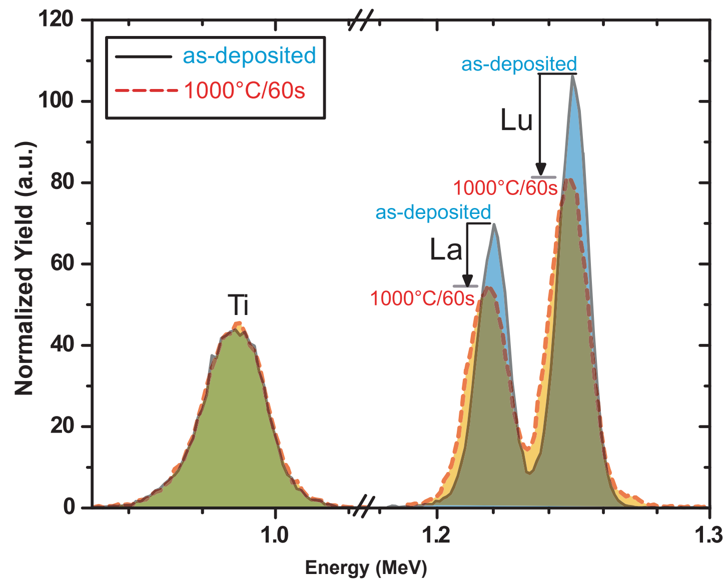


Fig. 3.15 RBS measurements of the annealed samples compared to as-deposited. The signals of La and Lu decrease and broaden after annealing, which corresponds to a relative decrease of the Ln concentration in the oxide layer.

intensity ratios. Moreover, the Ti:N ratio was found constant over the TiN layer depth.

At longer high temperature annealing, Ti and N signals overlap the La and Lu signals. Although the sputter duration for the as-deposited sample is lower, still some effect may be attributed to the diffusion of Ti(N) into $LaLuO_3$; the decreasing profile of Ti is crossing close to the maximum of the La and Lu profiles, meaning that diffusion during annealing is exceeding the matrix effects. This is in agreement with the concentration decrease of Ti measured by HAXPES at 10 keV (cf. Figure 3.13). The La and Lu spectra show profile broadening and clear Si diffusion through the oxide is proven depending on the annealing conditions. In addition, Lu atoms are equally distributed within the oxide, the profile decreases shallower than for La as approaching the interface to Si. This new information fully supports the SIMS data and the physical interpretation of the HAXPES data. In parallel to the in-diffusion of Si, the oxygen content in TiN is increasing from the top towards the bottom interface.

The XPS spectra show also reduced N concentration within TiN during annealing. Diffusion of N into the oxide is observed, whereby sputter migration can be

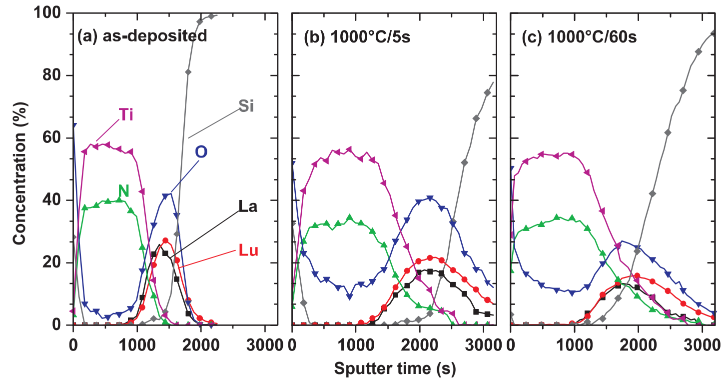


Fig. 3.16 XPS depth-profile spectra comparing different sample treatments. Surface oxidation of the Si-cap and Ti(N) is observed. For higher temperature annealing, the Si atoms diffuse into the $LaLuO_3$ and a layer broadening is visible as increased sputter duration. The oxygen diffuses into TiN and silicate formation via Si intermixing occurs.

excluded due to the low mass. This effect is still found in conjunction with the presence of the aforementioned preferential sputtering of gaseous species.

3.2.4.5 Electrical characterization of MOS capacitors

The capacitance-voltage characteristics (CV) for 10 nm thick $LaLuO_3$ layers are shown in Figure 3.18. The capacitance equivalent thickness (CET) was extracted at flat-band voltage $V_{fb} + 1.5\,V$. After 450°C forming gas anneal ofthe MOS stack a CET $= 3.1$ nm is extracted, as expected for an oxide with $k \sim 30$ and 1 nm interfacial oxide. The knowledge of interface chemistry and the gate oxide thickness is important for the determination of the relative permittivity.

For a typical source/drain activation temperature of Si MOSFETs at 1000°C for 5s, comparable CET was found without *in-situ* Si-capping of TiN.

In contrast, the capacitance after 1000°C/60 s reduced drastically. Based on the HAXPES information the relative permittivity of the silicate is calculated to be $\kappa \sim 11$ if only a single silicate layer is assumed for the 1000°C/60 s PDA. This value matches the literature reports[99] and the assumption seems reasonable in the light of the thin film analysis. The hysteresis indicates dipole charges within the gate stack.

Additionally, when a CVD step is simulated by adding an additional temperature step at 650°C/10 min (in ambient forming gas, magenta) V_{fb} shifts even towards lower voltages by 2 V under presence of large hysteresis and late recovery. These large shifts cannot be explained by a simple change of the TiN workfunction ($\max. < 1.2\,V$) and have primarily to be related to fixed charges in the gate stack.

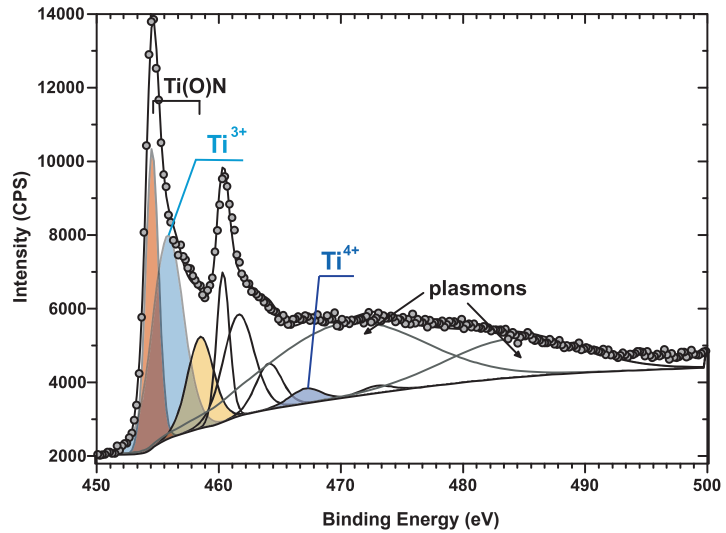


Fig. 3.17 XPS signal of Ti 2p including plasmon loss structure after 810s sputtering. Several defect states, i.e. sub-stoichiometric compounds, are assumed to fit the Ti = O and Ti-N main structures. A truncated and adjusted Shirley background function was used to match the intensity ratio in Ti 2p. The components of Ti 2p3/2 are plotted in colors.

Two possibilities arise: the interaction of $LaLuO_3$ with Si (lower interface) or defects at the upper interface to TiN as observed with HAXPES/XPS. To separate both influences, a sample with CVD simulation and RTA (magenta) is compared to a sample with single anneal at 650°C and final FGA 400°C/10 min (blue). The capacitance for 650°C (blue curve) is drastically reduced, but features comparable flatband voltage at slightly increased hysteresis.

Combining the CVD simulation with the RTA step, large flatband voltage shifts occur. This effect is again reduced for the 1000°C/60 s RTA sample. The accumulation capacitance is maintained at lower level for all three samples. Based on the complete transformation to silicate as found by SIMS for 1000°C/60 s anneals, a similar effect may reduce the capacitance during 650°C thermal budget. However, first the combination of silicate transformation *and* subsequent high-temperature annealing leads to strong positive defect charges induced within the gate stack. Nevertheless, the silicate formation was found to depend on thickness and temperature.[128] Thus, the observed charges should be less pronounced for thin layers. The capacitance reduction is problematic since it is also observed for long anneals at intermediate temperatures (650°C). Temperatures around 600°C – 650°C are typical for film deposition in CMOS processing, e.g. W CVD is used for filling the nano-

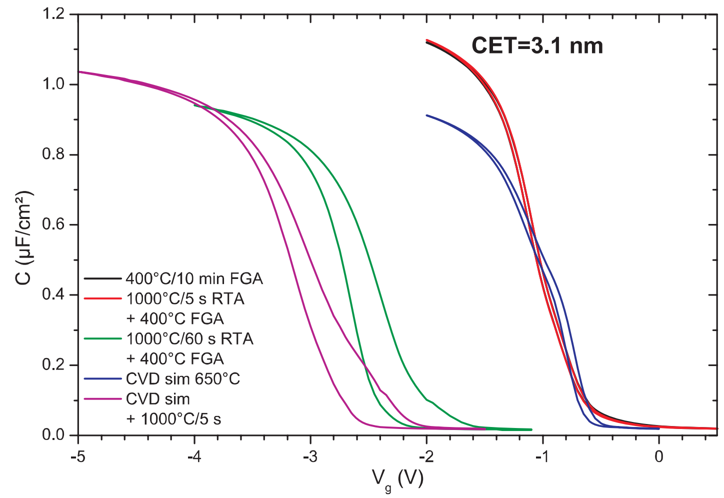


Fig. 3.18 C-V characteristics for different anneals. The CET was extracted at $V_{fb} + 1.5\,V$. Assuming complete silication of $LaLuO_3$ after 1000°C for 60 s annealing the LaLu silicate has a relative permittivity $k \sim 11$.

scaled trenches in full replacement gate integration schemes. The same method and comparable temperatures are chosen for amorphous silicon deposition nowadays.

3.2.4.6 HRTEM analysis

HRTEM analyses were performed employing an image-corrected FEI Titan 80–300 microscope to assess detailed information on the material morphology and layer thickness increase. In Figure 3.19 the as-deposited sample is shown in comparison to the sample annealed at 1000°C for 5 s. The thickness of as-deposited $LaLuO_3$ (11 nm) is in reasonable agreement with the RBS data. The TiN thickness is determined to be-16 nm. Lattice planes visible in the top layer confirm the polycrystalline structure of TiN in contrast to the amorphous $LaLuO_3$. After annealing (1000°C/5 s) a clear thickness increase from 11 nm to 14 nm ($\sim$+27%) is measured in agreement with the qualitative SIMS and XPS results. The thickness increase is also in agreement with reports for La silicate (+20%) by Kakushima *et al.*[128] In parallel, the thick interfacial layer transformed to a single silicate layer. This transformation can be explained by the HAXPES results (Figure 3.8). After deposition, the film consists of a Si-rich interfacial layer and $LaLuO_3$. After annealing, the Si-rich phase transformed into a La-rich silicate consuming partly the high-κ oxide. Additionally, the local separation of $LaLuO_3$ into two equal sized layers can also be

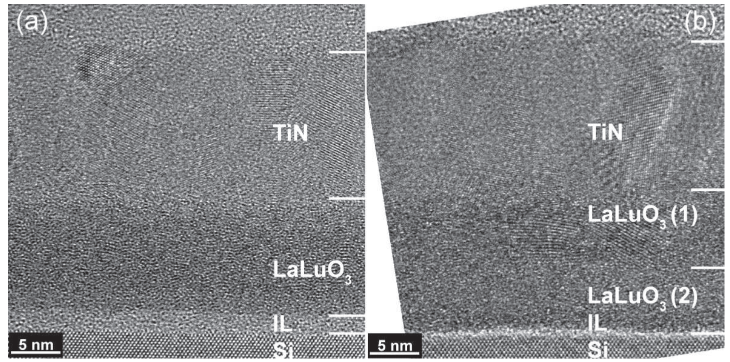


Fig. 3.19 HRTEM of the as-deposited (a) and 1000°C for 5 s annealed sample. The thickness of TiN is about 16 nm. The $LaLuO_3$ layer thickness is $\sim$ 10.4 nm in (a) and 13.3 nm in (b). The separation of $LaLuO_3$ into two layers can be interpreted by the aforementioned silicate formation (Measurements courtesy of Dr. A. Mücklitz, HZDR).

aligned with the photoemission data, which indicate silicate formation in $LaLuO_3$. The nano-crystalline features observed for the $LaLuO_3$ layer (1) can be attributed to diffusion of Ti into the layer with about 9% Ti, as verified by HAXPES. This is further substantiated by literature findings[133] on the formation of crystalline phases in $Lu_2Ti_2O_7$. A clear contrast between high-κ oxide and TiN is still observed in HRTEM, thus the diffusion of Ti does not induce significant density changes in the first approximation.

Additionally, a sample submitted to a 1050°C spike anneal and prior poly-Si deposition (600°C) was analyzed by means of STEM and EELS to assess the silicate formation during high temperature annealing. Figure 3.20 shows the simultaneously recorded STEM micrograph (left) and the corresponding EELS spectra. In the *z*-contrast STEM micrograph, higher atomic mass density appears in bright contrast and light elements in dark contrast. The EELS spectra were recorded in a distance of 4.7 Å. The lattice of the Si crystal allows estimating the drift during measurement to be zero. Starting from the Si substrate, the Si L-edge and La N-edge are shown on the right. Both spectra overlap in the magenta spectrum, whereby plasmon features in both species hampered a further background subtraction and quantification. Hence, quantification is only possible for the Si signal, i.e. the background subtracted EELS counts are integrated over an energy window from 99.4 eV to 103.4 eV and normalized. Here, the interdiffusion of both species is *limited* from the magenta spectrum to the next spectra ($\sim$ 1 nm) and thus less pronounced.

Averaging the EELS signal in scans parallel to the interfaces, the diffusion profiles were measured by the StripeSTEM[134] method, i.e. the simultaneous measurement of HAADF and EELS averaged parallel to the interfaces. The integrated EELS spectra reveal a diffusion gradient of 19 Å in respect of Si into $LaLuO_3$. The La

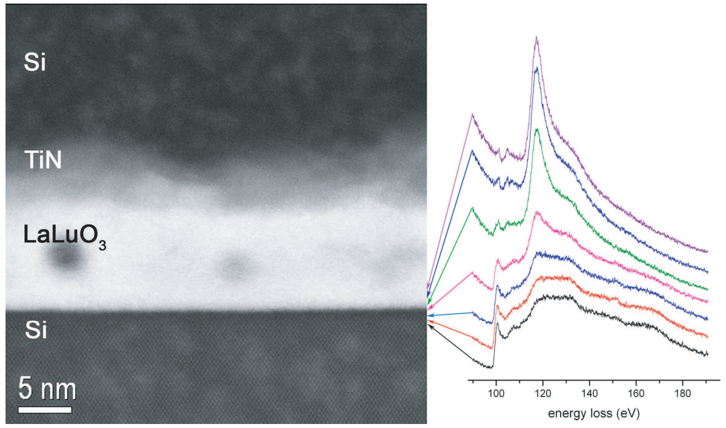


Fig. 3.20 HAADF micrograph (left) from STEM and corresponding EELS spectra (right) $TiN/LaLuO_3$ gate stack after spike anneal at 1050°C (n.b. 10 nm $TiN_{0.8}$). The EELS spectra were isochronously obtained in a lateral distance of 4.7 Å. Starting from the Si substrate, the L-edge of Si is visible (black and red). Further in $LaLuO_3$ (bright area left, i.e. higher density/magenta spectrum right) the La N-edge overlaps with the Si L-edge and plasmon features are present in both materials. Spectra deeper in the oxide show less Si background indicating no severe interdiffusion of both species for a spike anneal (Measurements courtesy of Dr. M. Luysberg, ER-C).

signal vanishes over a width of 9 Å, residual Si in $LaLuO_3$ amounts to-10% and is found over the whole layer. With respect to the XPS results for 1000°C/5 s annealing, the profiles belong to Si-rich silicate at the interface to Si and La-rich silicate/$LaLuO_3$ on top. The z-contrast in poly-Si is around 20% of the maximum density.

The dark areas observed in STEM (Figure 3.20) indicate local density variations. These variations were also found for the sample annealed at 650°C (blue CV curve in Figure 3.18). An EELS analysis (Figure 3.22) of these local density variations revealed similar density deviations, which contain La, O and Ti. Lu could not be identified due to its too high loss spectra. The mechanism introducing Ti into the $LaLuO_3$ layer remains unclear since only locally present. The effect can be interpreted as follows. The 600/650°C deposition or CVD simulation can lead to more local $Ln_2Ti_2O_7$ structures which are highly ionic and have low density[133] compared to the oxide. A subsequent high temperature anneal stabilizes the pyrochlore phase and leads to increasing ionic character. The ionic character leads to the large hysteresis observed for the combination of both anneals. Therefore, the large hysteresis can be accounted to local defect generation at the top interface $TiN/LaLuO_3$, whereby the overall reduction in capacitance is likely governed by the silicate formation.

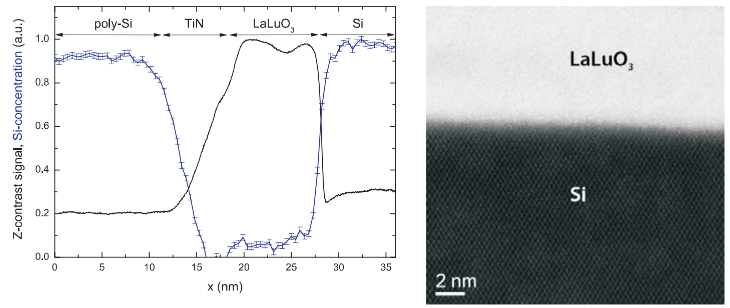


Fig. 3.21 (left) 10 nm $LaLuO_3$ after poly-Si deposition and spike anneal. *Z*-contrast of HAADF overlaid with the Si concentration obtained by the StripeSTEM method and EELS. The Si signal decreases over 19 Å into $LaLuO_3$. In parallel, the oxide layer vanishes to the substrate over 9 Å. In the whole layer Si is found by $\sim 10\%$. (Right) *Z*-contrast scan. The interface Si/$LaLuO_3$ shows a clear gradient without additional SiO_2 interface.

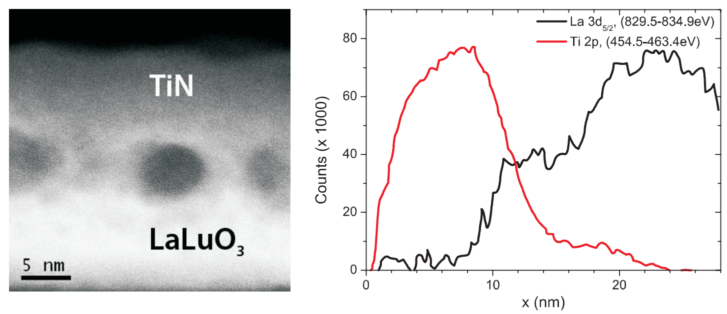


Fig. 3.22 STEM *z*-contrast image (left) and accompanying EELS spectrum averaged along the interfaces. The dark region contains La, O, and Ti at the interface to $TiN_{0.95}$. This interface region is ~ 7 nm wide. The sample with $650°C/10$ min anneal is shown.

3.2.4.7 HAXPES valence band spectroscopy

Very important material properties, with respect to device integration, are the bandgap, and the valence and conduction band offsets to Si. The valence band of $LaLuO_3$ was recently calculated by *ab initio* methods.[135] To compare $LaLuO_3$ oxide qualitatively to the simulated band structure, the valence band of the as-deposited sample consisting of $Si/LaLuO_3/TiN$ was scanned at photon energy of 10 keV. The determined orbitals suggest an Ln^{3+} state in $LaLuO_3$ with empty La 4f and filled Lu 4f orbitals. One important aspect in the determination of the correct valence band structure from HAXPES is the separation of the different layers scanned simultaneously. Fortunately, the Si orbitals Si 3p and Si 3s have a comparable small cross section for photoelectron emission. The same applies to Lu 6s, La 6s and Lu 5d orbitals.[136,137] Therefore, the obtained band structure is a superposition of N2p, Ti 3p and Ti 3d (Fermi edge) with the $LaLuO_3$ orbitals components. To separate the metallic TiN orbitals from the oxide ones the initial spectra were fitted using the orbitals of N, O, Ti, La, and Lu. Additional constraints were applied as the ratio between the known cross sections for La, Lu, Ti, and O were kept constant for each species. A linear increasing Shirley background function was used for background subtraction. The resulting peak fits for Ti and N were subsequently subtracted from the original spectrum. The obtained spectrum solely consists of the emission from the $LaLuO_3$ layer minus a minor contribution from oxygen in surface-oxidized TiN. On a final note, the resulting spectrum was smoothed using a Savitzky-Golay type algorithm. Figure 3.23 shows the initial and the subtracted spectra of $LaLuO_3$.

Table 3.2 Binding energies (BE) for the valence band of TiN and $LaLuO_3$. Due to very small cross section, N 2p was not detected.

Orbital	BE (eV)
Ti 3d	1.4
$Ti3p_{3/2}$	38.0
$Ti3p_{1/2}$	39.8
La 5s	41.8
$La5p_{3/2}$	20.4
$La5p_{1/2}$	22.1
$Lu4f_{7/2}$	9.6
$Lu4f_{5/2}$	11.6
$Lu5p_{3/2}$	31.3
$Lu5p_{1/2}$	31.6
N 2p	–
O 2s	25.1
O 2p	13.3

As known from *ab initio* calculations, the valence band of $LaLuO_3$ consists of La 6s, Lu 6s, Lu 5d, Lu 4f, O 2s and 02p.[135,138] The orbitals La 5s, Lu 5p, La 5p, Lu 4f, and O2p/O2s can be determined in $LaLuO_3$. Ln 6s and Ln 5d could not to be identified, thus, these states can be assumed empty as expected from Ln^{3+} states. Here, the strength of HAXPES valence band spectroscopy is the measurement of stacked thin films. Figure 3.23 summarizes the binding energies determined. For simplicity and due to unknown separation energy, O 2p and Ti 3d were assumed as each one peak. Literature values can only be given for Ti $3p_{3/2}$ in TiO_2 at 37.3 eV and Lu $5p_{3/2}$ at 28.10 eV in Lu_2O_3. These values can only serve as a guide line.

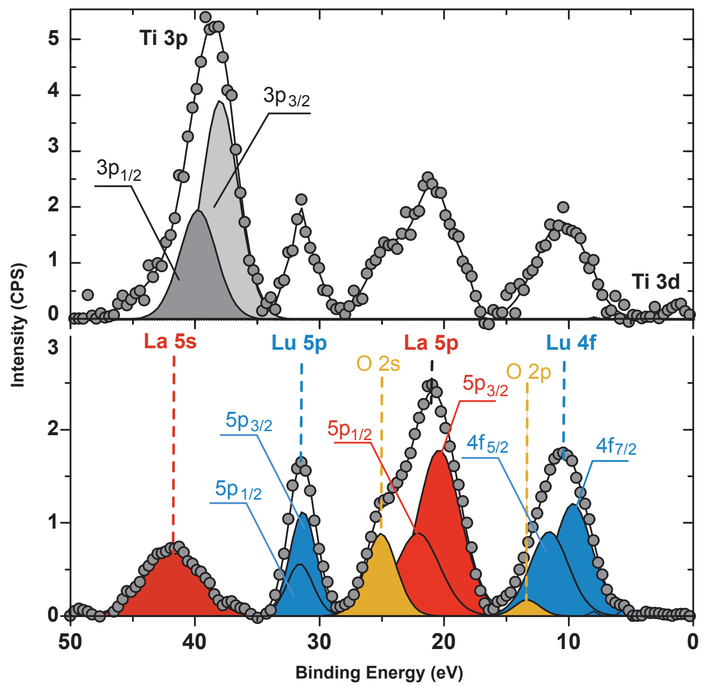


Fig. 3.23 HAXPES (10 keV) measurements of the valence band in $Si/LaLuO_3/TiN$ displayed after Shirley background subtraction. The measured spectrum (top) includes contributions from Ti(N) and $LaLuO_3$. Due to its low cross section, Si is not visible here. From the lower spectrum, the Ti $3p$ and Ti $3d$ spectra were subtracted and the data fitted again.

3.2.5 $LaLuO_3$ integration on Ge

Today several heterogeneous integrations are investigated, namely III-V compound semiconductors on Si as performance boosters for nMOS or local Ge epitaxy for pMOS.[89] The integration of those high mobility semiconductors comes at the price of increased complexity for the surface passivation. Namely on Ge, the native oxide is a mixture of sub- stoichiometric (GeO_{2-x}, often referred to "Ge-suboxide") and GeO_2. In contrast to SiO_2 and SiO_x on Si, GeO_{2-x} growth on Ge leads to much higher density of interface traps D_{it} compared to Si passivation. The earlier described tendency of $LaLuO_3$ to form bridging oxides to the underlying semiconductor (silicates, germanates) can be a promising alternative to oxidation of Ge for passivation.[22, 139–141] Figure 3.24 shows the capacitance-voltage (CV) curves obtained for 6 nm $LaLuO_3$ on Ge. To evaporate GeO_x growing on the Ge surface, the substrate was heated in vacuum to 600°C for 30 min. The high-κ deposition followed at 300°C substrate temperature by MBD. A final forming gas anneal at 400°C/10 min was carried out prior to C-V measurements. The measured CV curves show very low EOT values for the GeO_x free surface. The low depletion capacitance at high frequency suggests a functional semiconductor interface. However, measurement noise and a large difference of *n*- and *p*-type flatband voltage V_{fb}[‡] indicate the presence of frequency dependent, [§]charged defects at the interface to Ge. The conductance method[50] revealed a $D_{it} \sim 1 \cdot 10^{14}/(\mathrm{eV\ cm^2})$, i.e. only every third dangling bond on the (100) surface is passivated. Therefore, the adsorption of GeO_x in combination with a Sulfuric acid etch step and subsequent silicate formation leads to ineffective surface passivation as carried out here by MBD deposition and FGA.

To assess the interface quality, HAXPES measurements were done at the KMC-I beamline/Bessy II, Berlin. The detection angle was $\theta = 85°$ between surface normal and detector. Two samples were compared:

A PLD-grown sample on Ge/GeO_x and an MBD-grown sample with oxygen-desorbed surface processed as the gate stack measured in Figure 3.24. An alternative to desorption is the wet passivation with appropriate chemistry.[89] The PLD grown sample was dipped for 10 min at RT in H_2SO_4 (98%) and subsequently in 2% HF for 240 s prior

for 240 s prior to deposition. Both samples were capped with 2 nm a-Si (e-gun).

In Figure 3.25 the normalized spectra of Ge $2p_{3/2}$ for PLD- and MBD-grown $LaLuO_3$ are shown. The spectra were normalized to the $Ge^{\pm 0}$ substrate peak for comparison of the oxidation states. The MBD sample shows only germanate contributions at 1219 eV, which seems not detrimental for the high accumulation capacitance. However, a metallic component is also visible at lower binding energy (1216.4 eV). These bonds are less pronounced for PLD and indicate direct *Ln-Ge* bonds without bridging oxygen. Therefore, charged defects are also visible in photoemission spectra. The PLD sample shows germanate bonds at 1218.6 eV and 1219.6 eV (black arrows). The higher binding energy (BE) suggests Ln-richer germanate in comparison to the lower BE germanate observed for MBD growth. The binding energy of GeO and GeO_2 are tabulated as 1221.5 eV and 1220.2 eV, respectively. Sub-stoichiometric defective oxides are assumed to lie in between. Therefore, the interface between $LaLuO_3$ and Ge remained free of Ge oxide. The higher binding energy Germanate in the PLD case is likely due to the higher oxygen partial pressure during deposition. No sulfur was found in the stack. Therefore, the passivation with sulfuric acid and hydrofluoric acid is not resulting in an effective surface passivation by sulfur, which is in agreement with the observed D_{it} levels. The surface treatment still avoids Ge oxide growth during ex-situ deposition.

The associated La $3d_{5/2}$ spectrum (Figure 3.27) shows strong satellite features ($I_{sat}/I \sim 60\%$), which are also found in La $3d_{3/2}$ (not shown). These satellite features agree with the shake-up satellites ($70 \pm 10\%$) found by Teterin *et al.*[118] A com-

‡ i.e. $\Delta V_{fb} \approx 1\,\mathrm{V}$ compared to the band gap of $\Delta E^{Ge} = 0.67\,\mathrm{V}$

§ N.b. the difference in V_{fb} between 10 and 100 kHz measurement frequency

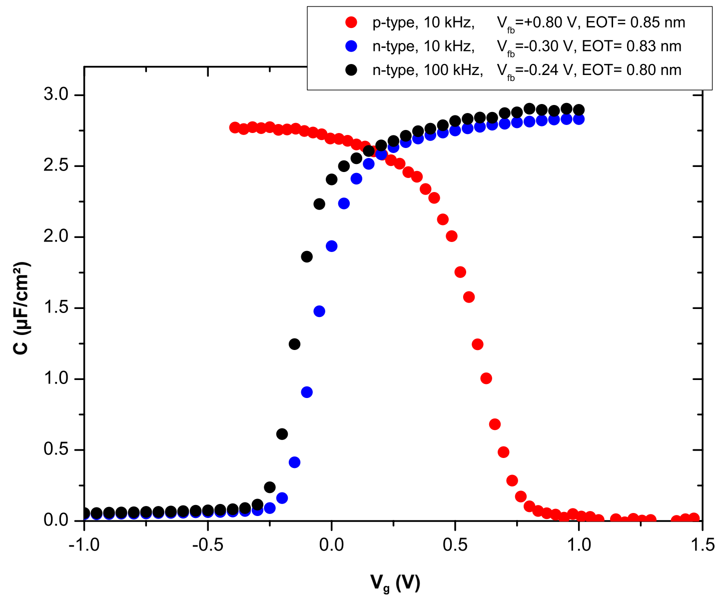


Fig. 3.24 Capacitance-Voltage curves of 6 nm $LaLuO_3$ on n- and p-type Ge(100) after 400°C/10 min FGA. Very low EOT values are reached and confirmed at 10/100 kHz. The difference in flatband voltage V_{fb} in respect to n/pMOS and frequency indicates additional, frequency dependent charges present at the interface.

ponent indicates a chemical shift originating from La bond over oxygen to Ge (germanates). The satellite structure for La-O is ~ 4.3 eV away from the main peak, whereby its germanate counterpart is closer to its main peak ($+2.7$ eV). Teterin[118] finds an energy splitting of 3.9 eV for La_2O_3. Following the argumentation of Teterin[117] the occurrence of different energy splitting can be explained by the O 2p binding state involved in the shake-up process for the La 3d core hole. The binding energy in O 2p for Ln-Germanate and Ln_2O_3 differs and thus leads to different satellite positions. The metallic bonds found in Ge 2p are only weakly pronounced ($< 2\%$) in La 3d. Previously, Sacchi[142] determined the EAL in Ge to be 50 Å at 4500 eV photon energy. For $LaLuO_3$, the EAL is extrapolated to reduce to 50 Å at 4500 eV (cf. section 3.2.3.2, p. 27). Consequently, the bonds in $LaLuO_3$ originate mainly from the interface region to Ge, whereby the Ge information comes mainly from 50 Å deeper in the substrate. However, the presence of these bonding states were more pronounced for MBD than for PLD and confirmed by measurements of the La 3d orbital. A further optimization of $LaLuO_3$ should avoid these bonding states.

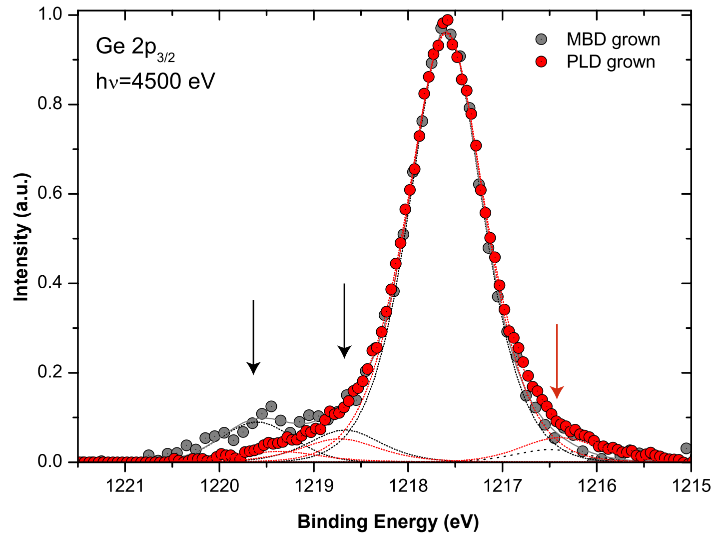


Fig. 3.25 Comparison of Ge $2p_{3/2}$ buried by 6 nm $LaLuO_3$ grown by PLD (gray) and MBD (red). For the MBD grown sample, the growth of metallic Ln-Ge bonds is observed (lower BE, red arrow). Ge-rich germanate bonds at 1219 eV are present. The PLD-grown layer shows two germanate contributions at 1218.6 eV and 1219.6 eV. The higher binding energy can be interpreted as more oxygen and La-richer bonds.

3.2.6 Summary

In previous work the thermal stability of $LaLuO_3$ in terms of crystallization temperature was shown to be sufficient for high-temperature CMOS processing $(\max .1100°C)$.[130, 143] By the combination of synchrotron-based HAXPES measurements and standard laboratory measurements, a transition from oxide to silicate phases could be shown with temperature. The major advantage offered by HAXPES is the possibility to separate the silicate growth and the atomic diffusion without degrading effects during material removal. The comparison to SIMS and XPS depth profiles revealed sputter-induced material degradation especially for the light elements. A clear difference was found between silicate formation coordinated to La and Lu. However, the exact localization of the diffusion profile remains challenging and time- consuming.

TiN might not be the best candidate as metal gate on $LaLuO_3$, since TiN was found to intermix with the underlying oxide layer up to 9%. Locally, Lanthanide Titanate forms, which causes high ionic oxygen transport leading to hysteresis in the capacitance once fixed in the crystalline phase. This mechanism is most likely

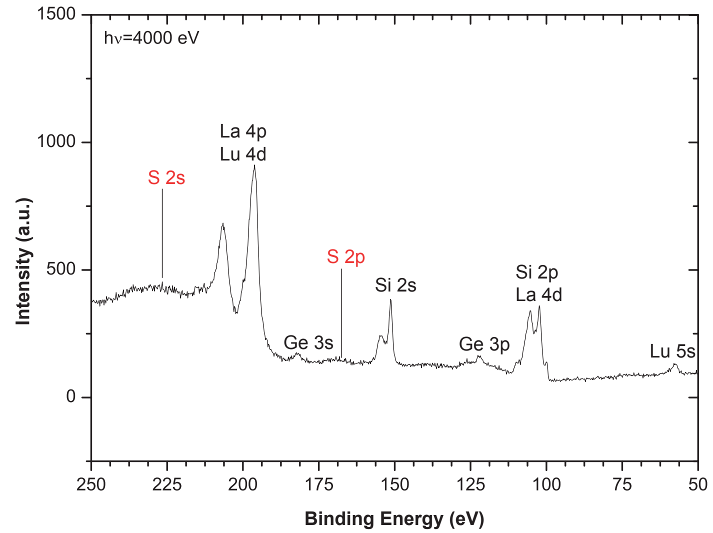


Fig. 3.26 Survey scan of the PLD grown sample plotted from 50–250 eV binding energy. No Sulfur was found within detection limit ($< 1\%$). A curvature at the indicated sulfur peak position is too broad for a photoemission peak. Likely surface plasmons from Si 2p and La 4p are causing these broad and flat contributions. The Si orbitals are present due to 2 nm Si-cap on top of $LaLuO_3$.

observed for a combination of intermediate annealing temperature (like thin film deposition by CVD) and subsequent high temperature annealing. Nowadays processing schemes avoid the high-T annealing and might render this issue obsolete. The more pronounced silicate formation as interpreted from capacitance reductions could be more important, since even in modern integration schemes such temperatures are still common.

La and Lu silicates are already present after deposition and were grown during MBD (fraction max. 10%). During annealing La silicate is constantly formed. Si diffuses significantly into the oxide layer and the majority of the initial Lu-bonds are transformed into Lu silicate. For shorter annealing, Lu orthosilicate formation was found which, as a Lu-rich silicate, could be beneficial for lower equivalent oxide thickness in MOSFET. The quantitative analysis proves that the diffusion of Si is mainly promoted via the generation of Lu-O-Si bonds and less by La bonds. Based on the structure analysis, further gate stack engineering is most promising for La-based oxides, whereby Lu tends to rather fast silicate formation.

The integration of $LaLuO_3$ on Ge was possible with low EOT[22,140,144] values under germanate formation, comparable to the silicates on Si. Photoemission spectroscopy of a thin high-κ oxide (today typically $< 5nm$) and underlying substrate

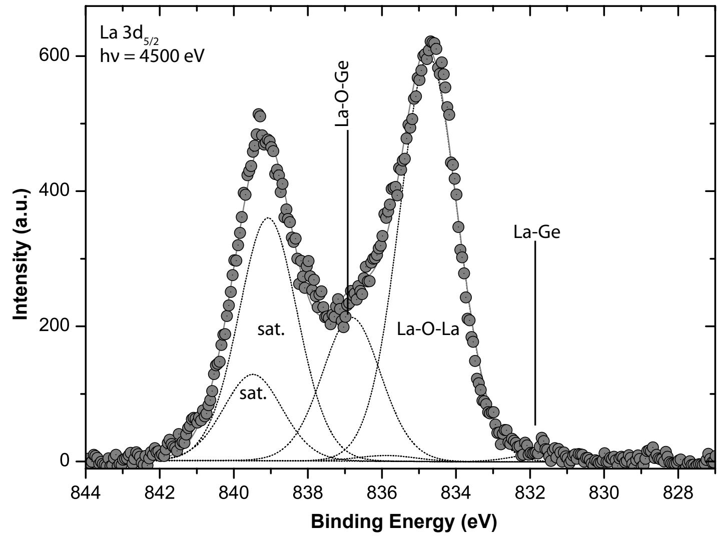


Fig. 3.27 La3d$_{5/2}$ obtained on Ge substrate. Weakly pronounced metallic La-Ge bonds and strong satellite features ($I_{sat} \sim 60\%$) are present. A chemical shift attributed to germanate bonds is also visible.

can be done in standard laboratory XPS. HAXPES measurements shall be preferable if the complete gate stack is to be screened. Therefore, control of germanate formation by XPS is an efficient, i.e. quick and detailed, procedure for the optimization of semiconductor passivation. The avoidance of Ge oxides at the interface seems promising in favor of a better Ge passivation. The influence of the different germanates on the electrical performance has to be investigated in future studies.

3.3 Further EOT scaling concepts

The combination of RCA-cleaned substrates and thin high-κ oxides allows the scaling of the equivalent oxide thickness (EOT) of the gate stack to values around 10–15 Å. The continued scaling demands the control of short-channel effects with an EOT of 4–10 Å.[145] In the following sections, new approaches are discussed for the further reduction ofthe EOT.

3.3.1 Use of thinner high-κ dielectric

The use of scaled high-κ dielectrics is a viable way for main processor units (MPUs) where gate leakage current densities up to tens of A/cm^2 are tolerable according to the ITRS 2011th update specifications[145] and no better solution is at hand. Thereby, gate leakage current will increase the total power consumption in support of equal or better device control with smaller EOT at smaller device dimension. The leakage current depends exponentially on gate oxide thickness and metal-insulator barrier height ϕ_0. Kim *et al.*[52] give a comprehensive overview of tunneling mechanisms.

As power consumption is a great concem of today's data centers[146] and mobile applications, the fastest growing markets, this scaling option can be judged as an intermediate option.

3.3.2 Use of higher-κ dielectrics

The partial replacement of SiO_2 as gate oxide with high-κ dielectrics enables EOT scaling at suitable low SiO_2 interlayer thickness t_{IL}. The high-κ oxide thickness t_{ox}, in measures of equivalent SiO_2 thickness, can be defined as,

$$t_{ox} = \frac{\varepsilon_0 \varepsilon_{\mathrm{SiO_2}} A}{c_{\mathrm{high}-\kappa}} = \frac{\varepsilon_{\mathrm{SiO_2}}}{\varepsilon_{\mathrm{high}-\kappa}} t_{\mathrm{high}-\kappa}, \tag{3.33}$$

here C is the capacitance of a normal plate capacitor and $t_{\mathrm{high}-\kappa}$ the thickness of the high-κ layer. The total EOT is then obtained by

$$EOT = t_{\mathrm{IL}} + t_{ox} = t_{\mathrm{IL}} + \frac{\varepsilon_{SiO_2}}{\varepsilon_{\mathrm{high}-\kappa}} t_{high-\kappa}, \tag{3.34}$$

Accordingly, assuming $\kappa \sim 20$ (e.g. HfO_2), a five times larger high-κ thickness can be employed for the same EOT. Alternatively, as EOT should be scaled down, the EOT can be lowered at constant leakage current due to larger total physical thickness.

The high-κ oxide thickness t_{ox} in terms of thickness SiO_2 and the overall EOT should not be confused with the capacitance equivalent thickness (CET) or inversion layer thickness t_{inv}.

The measured capacitance is expressed in CET. The CET value summarizes the thickness of t_{ox} and t_{IL} with the *measured* inversion layer capacitance of the substrate

$$CET = EOT + t_{QM} = t_{IL} + t_{ox} + t_{QM}, \tag{3.35}$$

where t_{QM} is the capacitance of the Si inversion layer and amounts to ~ 4 Å in Si.[41–43] The CET value is widely used in physical science, whereby t_{inv} originates from the scaling laws for device physics.

As t_{ox} scales reciprocally with the relative permittivity (κ-value"), the same EOT can be reached in a high-κ/metal gate stack by a material with higher relative permittivity compared to e.g. HfO_2. Then, even larger physical thickness consequently leads to reduced leakage current.

$LaLuO_3$, with reported[115] κ-value of 30 and comparable to HfO_2 conduction and valence band offsets,[147] is a possible candidate to outperform HfO_2 in the future technology nodes. In practice, lower EOT values were reported[23] for La_2O_3 at high D_{it} levels and hysteresis. La_2O_3 has the disadvantage to show permittivity degradation due to its hygroscopic[148] nature. Thereby wet processing of these gate stacks becomes difficult even if the surface is in-situ capped by the metal electrode.

The interface studies of $LaLuO_3$ have shown that the realization of low EOT involves optimum annealing conditions for Ln oxides and temperature dependent formation of Ln silicates. With standard annealing conditions (FGA 450°C/10 min), the relative permittivity is also reproduced with PLD-deposited $LaLuO_3$. Figure 3.28 shows the CET-plot for PLD-deposited $LaLuO_3$ under optimum conditions and after final FGA. The reported permittivity of 30 for MBD grown samples is confirmed in this plot for the PLD deposition. The thickness calibration for PLD was done by RBS measurements, which were calibrated against the oxide thickness from TEM measurements.

A possible future integration of Ln oxides differs from the integration of HfO_2 on Si/SiO_2: The integration of $LaLuO_3$ involves the formation of silicate interlayers between Si and high-κ oxide, while HfO_2 does less interact with SiO_2. Consequently, the scaling of SiO_2 thickness and HfO_2 is in principle independent. The silicate formation depends on the initial oxide thickness and temperature.[149] However, the silicates showed in this and other studies[99,129,150] a relative permittivity between 7 and 12 (max. 16). On one hand, the silicate formation adds additional process complexity. On the other hand, the direct contact of Si and silicates/oxides with higher permittivity than SiO_2 enables further downscaling below the minimum 4 Å SiO_2 thickness in the HfO_2 system. This mechanism could be a possible solution to achieve smaller than 5 Å EOT at the 10 nm technology node.

Combining the Si-capping of TiN with the high-temperature annealing of $LaLuO_3$ a further EOT reduction can be achieved. Figure 3.29 shows a C-V measurement for 3 nm $LaLuO_3$/2 nm TiN/a-Si. TiN was in-situ capped by 3 nm a-Si. Then the sample was dipped in diluted HF (1%). Subsequently, deposition of 100 nm a-Si and

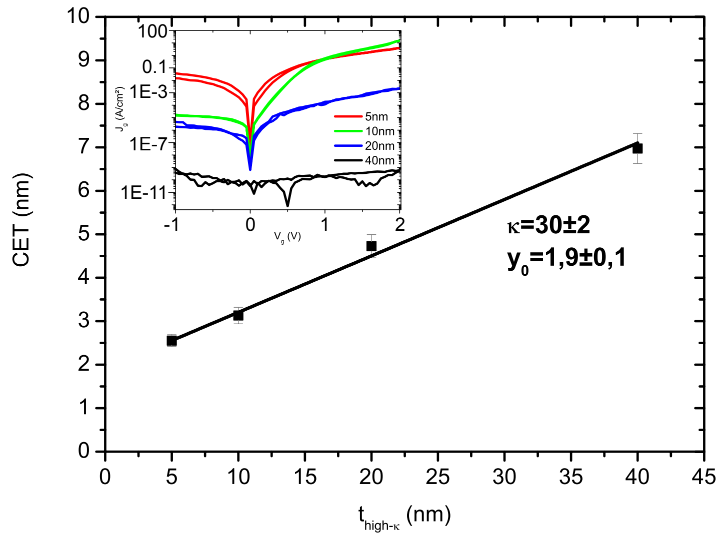


Fig. 3.28 CET-plot (high-κ oxide thickness vs. CET obtained at $V_{fb} + 1\,V$) for $LaLuO_3$ deposited by PLD under 100 Pa oxygen atmosphere. Reducing the oxygen partial pressure below 100 Pa led to smaller permittivity. For optimum deposition parameters, $\kappa = 30$ is confirmed. The interfacial layer for PLD deposited samples is rather high with nearly 2 nm. Inset: leakage current density for the different oxide thicknesses.

B^+ implantation followed. The thermal treatment consisted of a $600°C/10$ min anneal to mimic the CVD growth of 100 nm a-Si and 1050°C spike doping activation. After final FGA $(400°C/10\,\text{min})$, the EOT was extracted at $V_{fb} + 1\,V$ to amount to 1.2 nm. However, as for the devices with poly-Si from LPCVD, the hysteresis of the dielectric much increased. The EOT reduction compared to the non-optimized TiN process flow amounts to 3 Å. Without 600°C step, the device can be expected to provide low hysteresis.

3.3.3 Use of thinner SiO_2 interfacial layer

The EOT scaling is determined by the thickness and permittivity of the interfacial layer (SiO_2 or Silicate) and the high-κ oxide. The use of high-κ oxides was discussed before. In combination with high-κ oxide, the capacitance of the interfacial layer may further increase by lower SiO_2 thickness. One possibility here is the ozone oxidation for ultrathin IL[70,151] or the use of a thick SiO_2 layer, which is

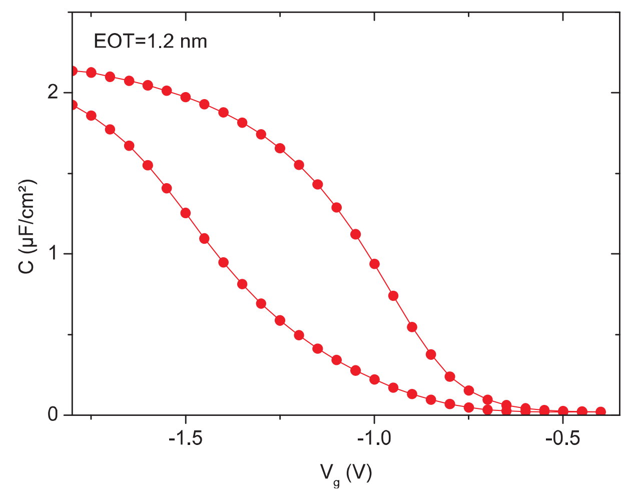


Fig. 3.29 C-V measurement of 3 nm $LaLuO_3$(MBD)/2 nm TiN (PVD)/100 nm a-Si (PVD) after CVD simulation (600°C/10 min) and RTA at 1050°C/spike. Here the optimized and Si-capped TiN was used in conjunction with an RCA-cleaned substrate (1 nm SiO_2 thickness prior to processing). EOT reduces from 1.5 nm to 1.2 nm. The aforementioned 10 min anneal at 600°C leads to large hysteresis.

subsequently etched back.[152] The same starting conditions apply also to the use of a thinner silicate.[128]

In this study, an etch-back process was investigated. The details of the process and process windows are given in Appendix A. The minimum SiO_2 thickness with full bulk properties is reported to be ~ 8 Å[9] and between 4 and 8 Å for bulk-like properties with reduced band-gap can be assumed.[9,153,154] Consequently, the minimum thickness should be 4 Å of high quality oxide after full processing in order to avoid mobility degradation due to additional Coulomb scattering.[20] These considerations limit the process window between 4 and 10 Å, whereby 10 Å is the reference value obtained for normal RCA cleaning. In total, after 40 s an interfacial layer of about 4 Å should remain based on the theoretical assumptions of the process window.

The experiments reveal high quality oxide formed with ozone. A $D_{it} < 1 \cdot 10^{11}/(\mathrm{eV\,cm^2})$ was extracted as determined from the conductance method[50] after final FGA. The D_{it} level stays constant for etch back times up to 40 s indicating no damage of the interfacial layer. This finding is in agreement with the SiO_2 study of Muller *et al.*[9]

Comparing different FGA at 400°C for 10 or 20 min and its position in process order, the FGA after full processing delivers the lowest D_{it} surpassing the PDA after

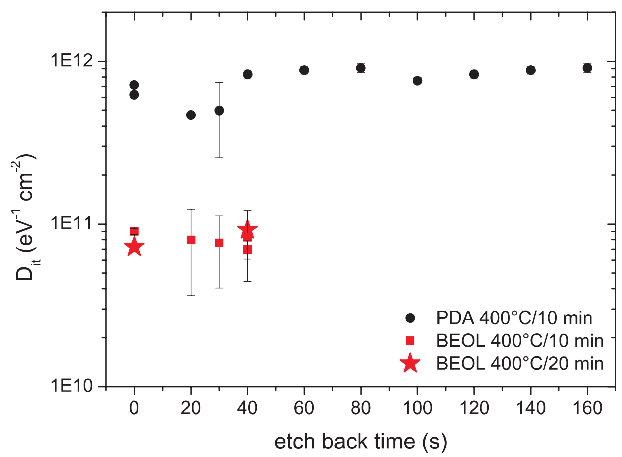


Fig. 3.30 Comparison of D_{it} levels extracted from the conductance method for different annealing and process order. A backend (BEOL) anneal after full processing leads to one order of magnitude lower D_{it} levels. Increasing the annealing duration does not further improve the result.

TiN deposition at 400°C/10 min about one magnitude (Figure 3.30). In parallel, the series resistance was extracted to lie between 350–450Ω and the higher D_{it} for the PDA was not accompanied by series resistance R_s reduction. The decrease in D_{it} is synonym for lower peak conductance of the samples, which is inversely proportionally influenced by series resistance.[155] Accordingly, the lower D_{it} levels after full processing are due to a lower peak conductance, which is not decreased by higher R_s. A possible explanation may be sintering of TiN and Al together during moderate temperature annealing. The native oxide present on TiN might reduce the ac conductance of the samples and final sintering improves the inter-metallic contact.

According to literature small amounts of Ti on the metal surface can reduce the hydrogen dissociation energy from 58 meV[156] (at 400°C) close to zero,[157] which explains the beneficial effect on passivation with metal gate in place. In contrast, the effect of further D_{it} reduction due to FGA after full processing is more likely an artifact of the better inter-metallic contact after sintering and not a sign of better interface passivation. Nevertheless, after complete processing an annealing is mandatory.

The minimum EOT was achieved for 40 s etch back and final FGA at 400°C/10 min as shown in Figure 3.31. The EOT was reduced by 3 Å with etch back. The theoretical expectation would yield a EOT of 6 Å for 3 nm HfO_2 on 4 Å SiO_2. No capping of TiN was employed and accordingly a regrowth of etched-back interface can be assumed: The lowest EOT values were found for PDA in FG directly after TiN deposition instead of FGA after full processing, which indicates an interface regrowth

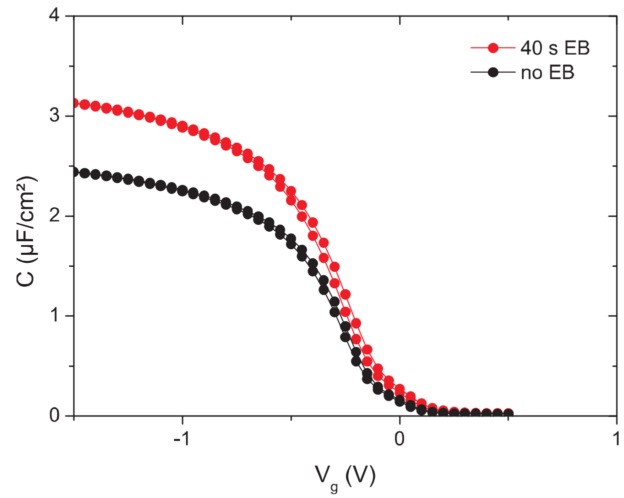


Fig. 3.31 C-V measurements for Si/3 nm HfO_2/TiN MOS Capacitors with and without 40 s etch back of the interfacial layer. The EOT amounts to 11 Å and reduces by etch back to 8 Å. Flatband voltage is 0.05 V and EWF ~ 5.1 eV.

during annealing due to oxygen sources brought in after TiN deposition. Moreover, the process variability over 200 mm wafers was yet larger as for a standard RCA clean and EOT > 12 Å. Possible explanations could be the missing Si-cap of TiN and the absence of moisture control prior to deposition of TiN (PVD). Possibly the in-situ processing of TiN (AVD) and HfO_2 (ALD) combined with Si-capping of TiN or moisture control[91] in PVD improve the process variability and are currently investigated. The EOT scaling below the thermal equilibrium of the involved reactions seems to be much more sensitive to moisture and deposition conditions. First experiments show the possibility to lower the ozone concentration to below 10 mg/1. Thus, a surface preparation without etch back could be possible since the oxide thickness was shown to depend on the ozone concentration.[151,158]

3.3.4 Thin TiN metal gates

The EOT after low-T annealing improves for the thicker and capped TiN. Recent reports suggest the usage of very thin (2 nm) metal gates at high T[59,70,71] which was explained in the section about thermodynamics (cf. section 3.1.3) to be promising in favor of SiO evaporation and TiO_2 formation.

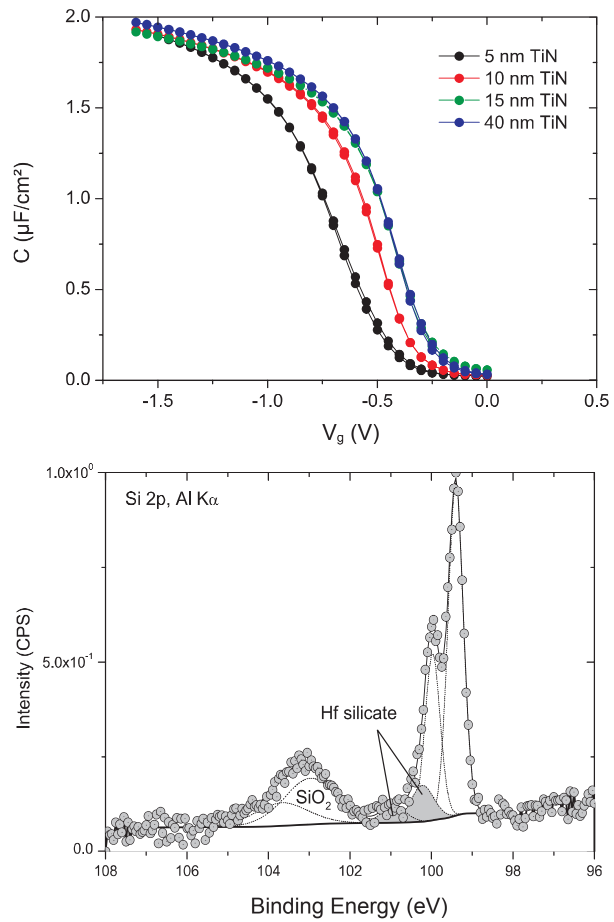


Fig. 3.32 (top) C-V measurements for 3 nm HfO_2 on RCA cleaned Si-substrate and different thicknesses of TiN/3 nm a-Si cap after high-T (MIPS) annealing. (bottom) XPS signal of 3 nm HfO_2 on ozone-cleaned substrates shows 14% silicate bonds after deposition. Similar results were found for RCA clean substrate.

Figure 3.32 (left) shows C-V measurements for different thickness of TiN and 3 nm a-Si cap on 3 nm HfO_2. The EOT (extracted at $V_{fb} + 1\,V$) is the same for all TiN thickness. Thus, within this work no results showed as large reduction of the interfacial layer after high-T annealing as reported in literature. In a recent report, Ragnarsson *et al.* showed a dependence of EOT reduction on the interfacial layer preparation. For thermally grown or chemical oxide the reduction is less pronounced (1–2 Å) compared to an ozone passivation ("imec cleaned").[72]

This result agrees with Figure 3.31, where an ozone process delivered 3 Å less EOT compared to Figure 3.32 for RCA cleaned substrate even at low T. Both batches were deposited with the same recipe of TiN (PVD). This comparison is astonishing as the physical thickness of ozone produced SiO_2 exceeded the chemical oxide of RCA cleaning by 5–6 Å as measured by ellipsometry. Either the real thickness is larger in the case of ozone passivation or the SiO_2 layers deviate in density. Mack[159] pointed out, that the chemical oxide (in their case using ozone) tends to silicate formation with high OH^- concentration on the surface. This silicate changes its thickness with HfO_2 thickness, whereby a thermal oxide (less OH^- on surface) does not change in presence of an ALD HfO_2 layer. Moreover, Green *et al.* described the HfO_2 growth kinetics to change in presence of these different interfaces.[160] They pointed out also, that the use of HF-last treated surface inhibits the growth of HfO_2. Similar results of inhibited growth and island nucleation on HF-last surfaces were also observed for $A1_2O_3$ and ZrO_2.[161–164]

The interface passivation and HfO_2 deposition were therefore investigated by means of XPS under exploitation of the information depth, which allows a screening of the buried Si interface below 3 nm HfO_2 (30 ALD cycles). Samples with 3 nm HfO_2 deposited on either RCA cleaned substrate or ozone passivated Si were compared.

The Si 2p signal of the ozone sample directly after deposition in Figure 3.32 (right) shows a shoulder at 100.2 eV for Si $2p_{3/2}$ attributed to Hf silicate formation[78] and amount to 14% of the bonds to oxygen. The low binding energy indicates an Hf-rich silicate, i.e. more Hf-Si than Hf-O-Si bonds. The further analysis of Hf and O in Figure 3.33 shows two distinct peak contributions for O 1s, whereby the small peak at 533 eV is attributed to SiO_2 and the Hf bonds are found at 531.2 eV. Unfortunately, the silicate formation cannot be further quantified from the Hf 4f spectrum, because the high symmetry of Hf $4f_{7/2}$ and Hf $4f_{5/2}$ do not justify further refinements of the peak model.

The RCA cleaned substrate showed the same characteristic Si 2p signal if the intensity was normalized to unity. Therefore, a possible difference in silicate formation between RCA cleaned or ozone cleaned Si, as reported by Mack,[165] is qualitatively not indicated. The investigated surface passivation is qualitatively in agreement with the chemical states reported by Mack[165] for chemical oxide but different for thermal oxide. Interestingly, the Hf:O ratio was determined by XPS to amount to 1:1.6 for both surface passivation methods. From RBS the stoichiometry was determined to be 1:2 ignoring any contribution of ultrathin SiO_2. Thus, XPS can quantify the concentration of thin, layered stacks more accurate as the method allows differentiating between the chemical states. The sub-stoichiometric $HfO_{1.6}$ explains possibly lower κ-value and thus, a decreased EOT scaling ultimately, but it does not explain the poor EOT scaling in respect to TiN.

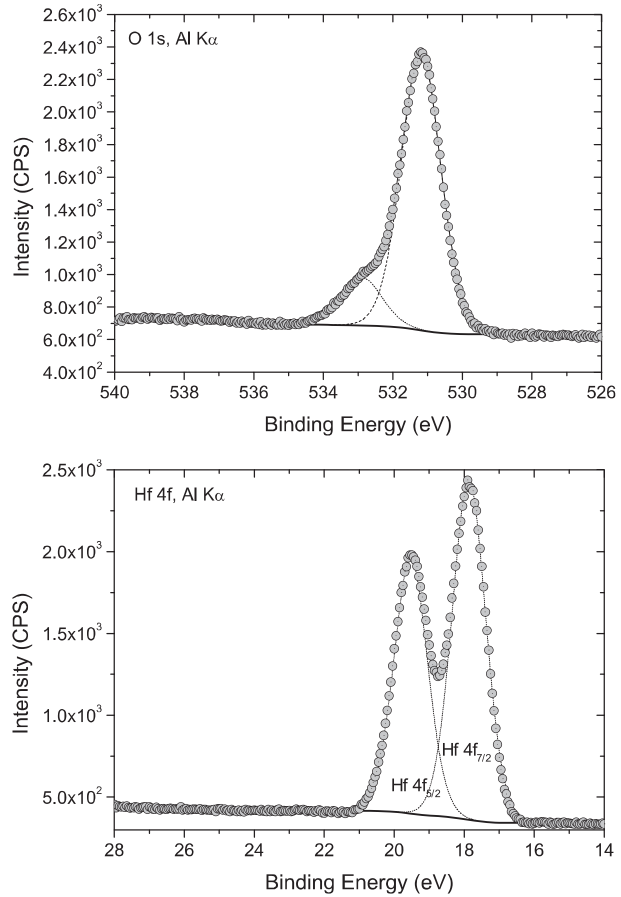


Fig. 3.33 XPS spectra of O 1s (top) and Hf 4f (bottom) for 30 cycles ALD-deposited HfO_2 on Si.

To assess the role of TiN the material was analyzed as deposited, after 400°C/10 min FGA and after a complete gate-first (MIPS) thermal budget. The latter mimics CVD growth for 10 min at 600°C, followed by 1050°C spike anneal and 400°C/10 min FGA. The annealing was again done under oxygen control ($<$ 0.1 ppm) in an RTP. Figure 3.34 compares the Ti 2p signal measured from a stack Si (100)/2 nm HfO_2/2 nm TiN/3 nm a-Si. The spectra were normalized to unity in order to com-

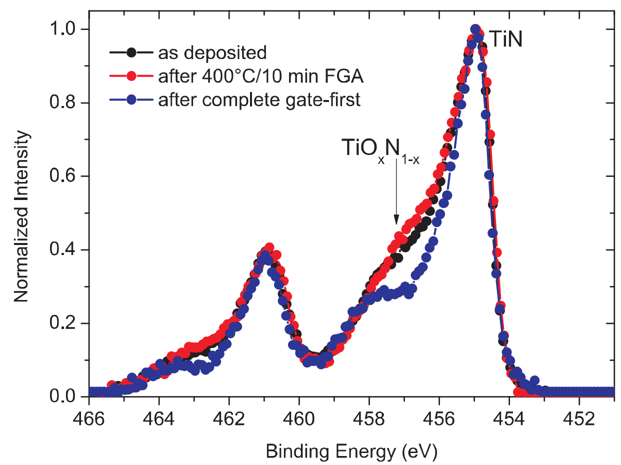


Fig. 3.34 Normalized Ti 2p spectra of 2 nm TiN with 3 nm a-Si cap after different process steps.

pare the qualitative peak shapes. The intensities will not be compared. As visible for 15 nm TiN, the spectrum shows directly after deposition a shoulder around 457 eV. The shoulder can be interpreted as a TiO_xN_{1-x} component next to the TiN peak at 455 eV BE. This contamination increases slightly after a single 400°C FGA, which may explain a minor capacitance increase after FGA. The spectrum after complete processing looks different: The component at 457 eV decreases clearly in favor of the TiN component. The decrease can be due to the reduction of Ti oxygen bonds and oxidation of SiO_2 as expected from thermodynamic considerations. Hence, the oxygen in TiN hampers an EOT scaling, most obvious after 1000°C annealing.

Interestingly, the total amount of Si after 1050°C spike anneal decreased drastically by $\sim 35\%$ compared to the as-deposited or 400°C processed samples (Figure 3.35). In parallel, the amount of silicate/SiO_2 and Hf 4f from HfO_2 below remained almost constant. This result is explained by the desorption of Si(O) during high-T annealing and thinning of the Si-cap, while the remaining Si oxidized again after RTA. The loss of Si has to be considered for the choice of Si-cap thickness.

In summary, ozone and RCA cleaning do not indicate different IL formation during processing in the form of different chemical states. Thus, the discrepancy between the results reported by Ragnarsson *et al.* and the results of this work are likely due to differences in the TiN metal gates used or extrinsic oxygen sources. RBS results showed a TiN layer with negligible oxygen content in the bulk. Incorporation at the interface to the Si-cap is then visible in XPS and cannot be excluded by RBS. The oxygen source between both in-situ steps has still to be clarified.

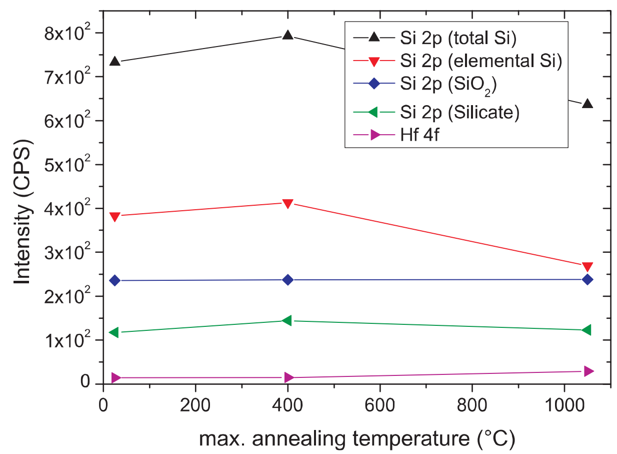


Fig. 3.35 Integrated peak intensities for Si 2p and Hf 4f after annealing at 400°C and 1050°C compared to the as-deposited gate stack. The high-T annealed sample lost Si.

3.3.5 Doped TiN

The thermodynamic calculations recommend the combination of a thermally stable gate electrode, preferentially TiN, with a scavenging element like La, Y or Al to increase the oxygen solubility in the gate stack. This idea is well known for the deposition of metallic superconductors, where the oxygen partial pressure of the deposition chamber is lowered by Ti sputtering prior to deposition. Several ideas have been proposed by one research group[60, 85, 98, 100, 101] to increase the scavenging of TiN. The best results were obtained for thin TiN between scavenging element and high-κ oxide in order to keep the TiN work function.[98] In this work, several concepts for EOT scavenging were evaluated.

Al scavenging in contact with TiN was realized by the NiSi-mediated diffusion of Al to the interface Si/TiN. The process was chosen according to the MIPS flow with 1050°C dopant activation with 100 nm a-Si (PVD) deposited in-situ directly on TiN. The Boron doping concentration for CVD and PVD poly-Si was determined by SIMS to be $2-3\cdot 10^{20}/\mathrm{cm}^3$. *After activation* of the gate stack, MOS capacitors with Si (100)/2.3 or 3.5 nm HfO_2/x nm TiN/B^+ poly-Si were covered by 10 nm Ni and 200 nm Al. The thickness of TiN was chosen 2 or 5 nm, which was determined by RBS and TEM. Subsequent forming gas anneal at 400°C for 10 or 20 min leads to the formation of NiSi within the gate stack. Most important is the intermixing of Al and Ni. In a previous study of our research group[166] the Ni-Al system in Si was found to produce high quality $NiSi_2$ on the active S/D area.

Thus, the use of the Ni-Al system on poly-Si offers two advantages: First, it serves as a low resistive contact to the gate stack. Second, the segregation of Ni and Al leads to the uniform distribution of Al down to the interface to TiN as confirmed by ToF-SIMS. Moreover, the silicidation of S/D and gate area is common for low-resistive gate-first processes, where the gate stack is formed prior to S/D activation. The diffusion length d can be estimated from the given diffusivity Das[167–169]

$$d = \sqrt{D \cdot t} \tag{3.36}$$

where D is the diffusivity for diffusing species in TiN and t is the processing time at temperature T. The concentration decay at the interface is then obtained by the one-dimensional law of Fick:

$$n(x,T) = n_0 \cdot \operatorname{erfc}\left(\frac{x}{\sqrt{D \cdot t}}\right), \tag{3.37}$$

where n_0 the concentration at the interface $(x = 0)$ and erfc () the complementary Gaussian error function. The short temperature ramps during processing are neglected for the sake of simplicity.

The diffusion length of Ni and Al, as calculated from literature data,[167–169] amounts to 1.4 nm and 4 nm, respectively, for a processing temperature of 400°C and 10 min duration. The concentration diffused at the interface TiN/HfO_2, i.e. 2 nm away from the top interface NiSi/TiN, is approximately

$$n_{\mathrm{Al}}(2\,\mathrm{nm}, 600s) \approx 63\%$$
$$n_{\mathrm{Ni}}(2\,\mathrm{nm}, 600s) = 5\%.$$

The diffusivity is expected to decrease with nitrogen content and denser TiN.[167,169] Thus, the use of optimized TiN guarantees the suppression of significant diffusion and thereby induced work function changes. The use of a fully silicided gate stack is not at the peril of significant Ni diffusion to HfO_2 according to the calculations. For 20 min FGA, the approximate diffused Ni concentration is significantly present at the oxide ($n_{\mathrm{Al}}(2\,\mathrm{nm}, 1200\,\mathrm{s}) \approx 73\%$ and $n_{\mathrm{Ni}}(2\,\mathrm{nm}, 1200\,\mathrm{s}) = 48\%$. Thereby work function changes are likely for 20 min FGA if the diffusivity is as given in the literature.

La, Eu, Sr, Lu, and Y were not available as sputter targets. Thus, doping of the metal gate (15 nm TiN) by ion implantation of the desired species was carried out. TiN was in-situ capped with 3 nm a-Si prior to implantation to prevent most oxidation. The estimation of diffusion was not possible due to the lack of appropriate literature data. After ion implantations, the samples were HF-dipped to remove SiO_2 from surface and subsequently covered by 100 nm a-Si. Due to our implanter limitations, the implantations were only possible with acceptable currents if the implanter was operated at energies above 10 keV. This limits the lowest thickness of TiN to 15 nm ($\rho_{\mathrm{TiN}} \approx 5.44\,\mathrm{g/cm^3}$ and $\rho_{a-Si} \approx 2\,\mathrm{g/cm^3}$) with e.g. a projected range of 80(+23) Å for La and 79(+19) Å for Y. The parameters were chosen to have minimum implantation into HfO_2 to avoid defect generation. The remaining process

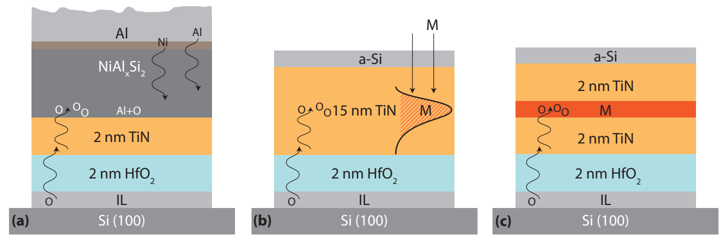


Fig. 3.36 Different integration schemes for increased oxygen scavenging in TiN. (a) The use of a MIPS gate stack allows the formation of NiSiAl during BEOL anneals and lets Al and Ni diffuse to the interface TiN. Here the thermal treatment and the Al concentration may influence unwanted work function changes (b) The implantation of appropriate scavengers into 15 nm thick TiN is a prototypical screening method. The implantation straggle can cause damage to the high-κ oxide. However, ΔR_p is small in TiN. (c) The best integration for low-T processes seems the nanolaminate since it allows a precise thickness and concentration control. For (b) and (c) difficulties can arise during gate stack etch if the doped TiN was subjected to a high temperature anneal.

flow was as in the MIPS process flow. Figure 3.36 gives an overview of possible integrations of doped TiN. The temperature history of the sample can hamper the gate stack etch if Ti-M alloys were formed at high temperature. For replacement gate processes, this problem is avoided by design.

N.b. the La and Y doped TiN gate stacks underwent the full thermal budget of a gate-first process in contrast to Al, which was only subjected to a low-T $(400°C)$ anneal.

The gate stack analysis was done by XPS depth profiles and ToF-SIMS. Figure 3.37 shows a SIMS analysis of 2 nm and 5 nm TiN within the gate stack after NiSiAl diffusion during FGA $400°C/10$ min. As observed before, Al and Ni diffuse. Al distributes uniformly over the gate stack and Ni piles up in front of TiN, whereby TiN blocks further Ni diffusion. 5 nm TiN (bottom) blocks also the diffusion of Al. For 2 nm TiN no clear pictures is obtained due to the sputter mixing zone, which leads to signal decay of both signals at the same time.

To assess the chemical states of the elements, also sputter depth profiling by XPS was carried out. The difficulty here arises from the information depth (ID) in the sample ($3^*EAL \sim 5$–10 nm, effective attenuation length (EAL)). The high Al concentration within poly-Si leads to an underestimation of the 2 nm TiN/2 nm HfO_2 layers and the ID broadens the signal.[170] Based on the XPS measurements, NiSi formation was also found to stop prior to TiN and the ratio NiSi to $NiSi_2$ (not shown) was $\sim 10:1$. The XPS depth profile shown in Figure 3.38 reveals SiO_2 on top of TiN and HfO_2 below TiN. At the interface to TiN, Si is found to vanish almost in favor of a high concentration of Al and is accompanied with a peak concentration NiSi, which is interpreted to block further Al diffusion to TiN.

Thus, the NiSiAl phase propagated to the interface to TiN with a highly diluted Al concentration. The further analysis of Si and Al is hampered by the relatively huge concentration. Averaging over the information depth ($\sim 3^*$ effective attenua-

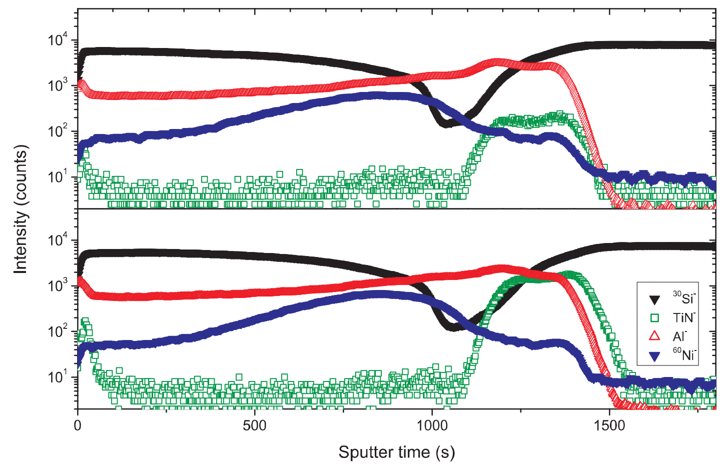


Fig. 3.37 ToF-SIMS depth profile for the Al-doped gate stack. NiSi formation leads to uniform Al diffusion into poly-Si. Ni accumulates at the interface to TiN and is blocked. (top) 2 nm TiN lead to Al diffusion close to the assumed interface to 2 nm HfO_2. The sputter mixing zone prevents a better depth resolution. (Bottom) 5 nm TiN on 2 nm HfO_2 is surely indicated to block Al and Ni diffusion.

tion length) leads to less signal for 2 nm thin TiN compared to Al as ID is greater than the film thickness. N and O were also measured but suffer from preferential sputtering and are excluded. Based on XPS and SIMS, the Al diffusion likely stops within 2 nm TiN, close to HfO_2, but is at the depth resolution limit of both measurement methods. Moreover, Si and Al oxide are found above

TiN indicating a possible oxygen transport within the gate stack away from the interface to Si.

The electrical characterization was done by leakage current $(I_g - V_g)$ and C-V measurements. Compared with a MIPS reference process (2 nm HfO_2 on RCA cleaned Si), a well-pronounced increase in accumulation capacitance can be achieved for 10 min annealing at 400°C and the use of 2 nm HfO_2 with an EOT = 4.6 Å extracted at $V_g = -1$ V (Figure 3.39). The minimum EOT achieved with 3.4 Å extracted at maximum accumulation is surprisingly small. The curves show nearly no hysteresis. However, the processing is sensitive to annealing and high-κ thickness. For 3 nm HfO_2, the CV curve becomes leaky (one order of magnitude) and D_{it} increases to $10^{13}/(\text{eV cm}^2)$, which manifests in a large D_{it} hump and a -200 mV shift of the accumulation capacitance. A possible explanation might be the higher crystallinity of 3 nm HfO_2 and a diffusion of Al along the grains to the interfacial layer causing defects due to reduction of SiO_2. The details could not be resolved with the available resolution.

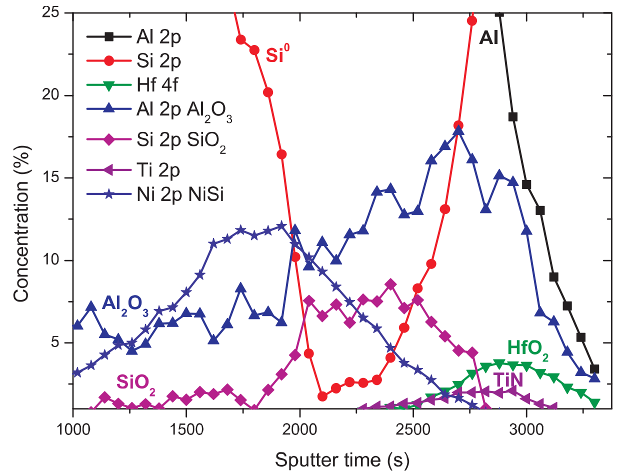


Fig. 3.38 XPS depth profile for the Al-doped TiN. NiSi formation accumulates at the interface to TiN and blocks further Si from top. Below SiO_2 and Al_2O_3 is found to increase directly on TiN. The high Al concentration leads to an underestimation of the 2 nm thin TiN/HfO_2 layers and hinders a further analysis of the Al gradient.

For 20 min annealing the results deteriorate as the D_{it} increases further and capacitance drops by 50%. Here, the 3 nm high-κ layer shows less shift as for 10 min anneal. The leakage current density as shown in Figure 3.40 exceeds the leakage of classical SiO_2 gate stacks (reference 1line[100]) due to the EOT extracted below accumulation but at a value important for device performance. Of course, corrected extraction of the EOT at V_{fb} plus an offset would shift the result to an EOT ~ 9 Å and to reasonable EOT-leakage scaling. Nevertheless, the huge D_{it} renders usable device performance impossible. A comparison with state-of-the-art gate stack results[84] reveals good $J_g - EOT$ scaling for pure HfO_2. Ando *et al.*[84] showed even higher capacitance results (no hysteresis shown) in combination with La-cap layers, which are known to transform the interfacial layer to a silicate.[83] A possible combination of scavenging and interface silicate formation remains open, due to no known thermodynamic data for the La silicates: If La silicate is more stable than HfO_2, as the case for La_2O_3, the scavenging would preferentially reduce HfO_2 and lead to defects in the high-κ layer.

The C-V curves for La and Y ion implantation into TiN are shown in Figure 3.41. Both results show a capacitance increase with increasing doping level. The best results are obtained for Y doping with a capacitance increase of 17% for a level of $5 \cdot 10^{14}/\mathrm{cm}^2$ Y in TiN. The doping level is equal to 5% Y in respect to the density of TiN. The achieved EOT decrease exceeds the prediction of thermodynamic reactions. It could be that the favorable oxidation of dopants in the upper part of TiN

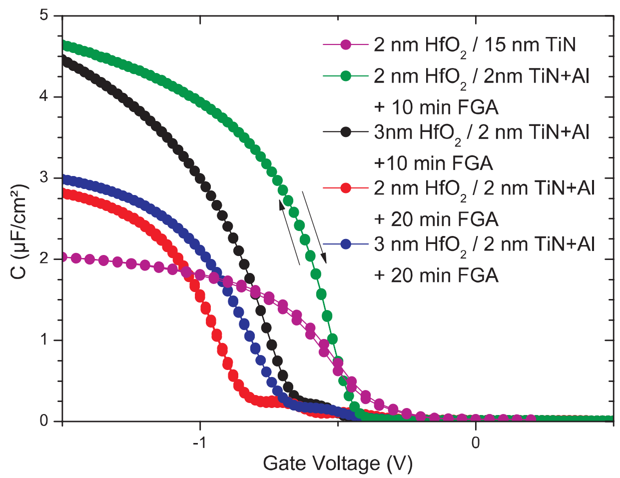


Fig. 3.39 Dual-sweep C-V measurements for TiN/HfO_2 MIPS gate stacks after low-T diffusion of Al. For undoped TiN EOT is 14 Å. With thin TiN, the EWF decreases after high-T annealing as described before to midgap (4.6 eV). The EOT for thin TiN and 2 nm HfO_2 was found to decrease to 4.6 Å at $V_g = -1\,\mathrm{V}$.

leads to a cap effect for not consumed TiN. However, this discrepancy cannot be explained by the current measurements. Higher Y concentrations lead to irremovable TiN with respect to the standard TiN etch step. Already for the lower concentration in some curves a large D_{it} increase $(> 10^{12}/(\mathrm{eV\ \ cm}^2))$ is visible as a hump around flatband voltage. Likely, the ion implantation lead to defects in the high-κ oxide and increases there the density of oxide defects D_{ot}, which can be also visible for thin layers and high defect density in the D_{it} level. Characteristic for the large D_{ot} curves were deviations from the ideal MOS shape due to the convolution with the capacitance of defect state.

It is concluded, that ion implantation can be a suitable way to screen materials for TiN doping. However, certain process variability masks the results and conclusions become more difficult. E.g. also experiments with Eu, Sr, and Lu were carried out at high dose $(10^{15}–10^{16}/\mathrm{cm}^2)$. Therein TiN was realized, like for Y at $10^{15}/\mathrm{cm}^2$, to be not removable by standard etch steps. Surely, this problem is purely extrinsic due to the MOS capacitor process flow, where TiN was structured after high-T annealing. A gate patterning before any annealing could promise less process variability. However, the intrinsic high-κ defect increase is due to the high-energy ion implantations. Clearly, a better way is shown in Figure 3.36c as the nanolaminate avoids all process-induced damage to the high-κ layer.

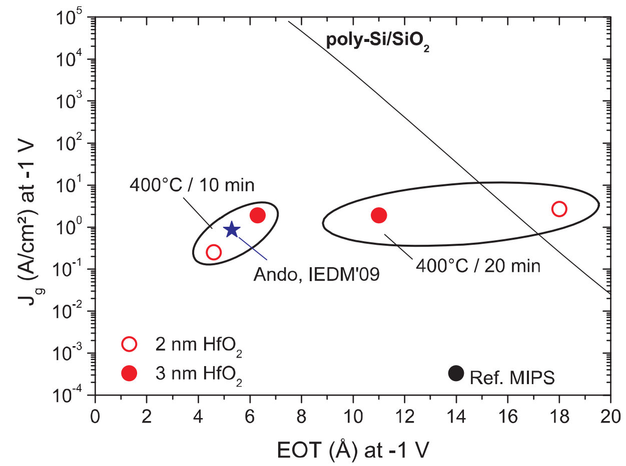


Fig. 3.40 EOT vs. leakage current density for two different annealings and two different thicknesses of HfO_2. For 10 min annealing and 2 nm HfO_2, Al diffusion is beneficial for EOT scaling.

The ion implanted La and Y samples underwent the full thermal treatment of a gate-first process (CVD simulation 600°C, 1050°C spike anneal and final FGA at 400°C). The important difference between these dopants and the Al diffusion process is the diffusivity. For La and Y negligible flatband voltage shifts are observed in Figure 3.41 (except for the D_{ot} increase) in contrast to the annealing study for Al. For the ion implantation experiments, also a 10 nm TiN barrier layer between poly-Si and Al metallization was used. Thus, the diffusivity of Y and La within TiN is negligible for the utilized thermal budget.

The scavenging effect was also investigated for $LaLuO_3$ using TEM, STEM and EELS analysis. Samples comprising Si(100) substrate/3 nm $LaLuO_3/22$ nm TiN (ALD) were compared after 400°C/10 min FGA. A Sr implanted and a Y implanted sample was analyzed in comparison to a control sample without implantation. Sr and Y were chosen according to the thermodynamic calculations: Sr can reduce TiO_2 but not La_2O_3 and Lu_2O_3, which were chosen in default of missing thermodynamic data for $LaLuO_3$. Y is capable of reducing La and Lu oxide and was chosen to identify possible effect on $LaLuO_3$.

The implantations were carried out into TiN at energy of 20 keV to a dose of $5 \cdot 10^{15}\,\mathrm{cm}^{-2}$ The implantation energy was chosen carefully on base of TiN thickness measurements to avoid a projected range exceeding TiN thickness into high-κ material. The corresponding TEM micrographs are shown in Figure 3.42. Clear differences in the interface to the Si substrate are visible. The sample without implantation shows relatively rough interface with bright sections indicating low-density areas. However, the sample is relatively thick, which is indicated in poorer high-resolution

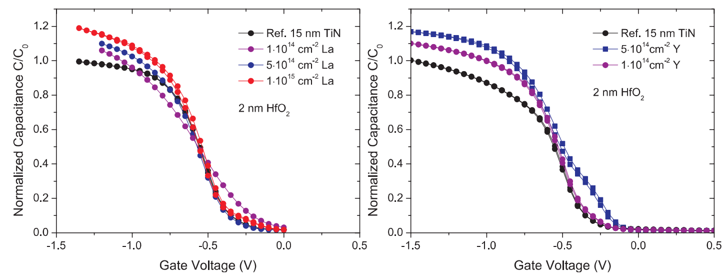


Fig. 3.41 C-V measurements for doped 15 nm TiN capped by 3 nm a-Si and processed in a MIPS process flow. Measurements are normalized to the accumulation capacitance in the TiN reference wafer of each batch. (left) results for different La dose implanted at 12 keV. A clear difference between zero and $1\cdot 10^{15}/\mathrm{cm}^2$ La doping in TiN is visible. The difference between 1 and $5\cdot 10^{14}/\mathrm{cm}^2$ is visible but close to process variability. (right) Y-implanted samples show a little stronger capacitance increase with doping level. The capacitance increased 17% for $5\cdot 10^{14}/\mathrm{cm}^2 Y$.

contrast in the Si substrate. The result match the low-density Si-rich silicate and high density $LaLuO_3$ found in the interface studies. With Sr implantation, the interface roughness decreases and a bright area indicates the presence of a Si-rich/SiO_2 interlayer between Si and $LaLuO_3$. The sample with Y implantation features no visible interfacial layer and reduced interface roughness.

To quantify the qualitative picture of the TEM analysis, STEM and EELS measurements were employed. Common in all analysis is the ADF scan mainly following the EELS signal of La/Lu. Therefore, an analysis based on ADF delivers also the gradient at the interface.

The STEM analysis (Figure 3.43) confirms the early results of TEM. ADF for the Sr implanted sample shows a density gradient of 9 Å (ADF scan in Figure 3.44) compared to a 3 nm shallow decreasing and rough interface for the non-implanted sample (not shown). The EELS spectra (Figure 3.44) indicate an Ln-Silicate interface to Si at that position. Lu was not detectable due to the poor signal-to-noise ratio for the K and L-shell, which was also true for O K in this sample. The ADF scan confirms a 20% lower density in that region compared to $LaLuO_3$ bulk, which vanishes to the background intensity at the same position the La signal increases. The interface TiN/$LaLuO_3$ is ~ 1 nm wide after FGA.

The EELS study of the Y-implanted sample shows a different interface structure as confirmed by the TEM micrograph. The corresponding STEM image and EELS analysis (Figure 3.45, right) reveals a much decreased interface width of Si/$LaLuO_3$ below 5 Å. The signal was averaged along the interface of the stripe displayed to the left of Figure 3.45. The density obtained by ADF is nearly constant within $LaLuO_3$. Interestingly, the oxygen content increases to TiN, which supports the argument of (increased) oxygen solubility in TiN (with dopants). The interface to TiN shows a width around 2 nm (FWHM).

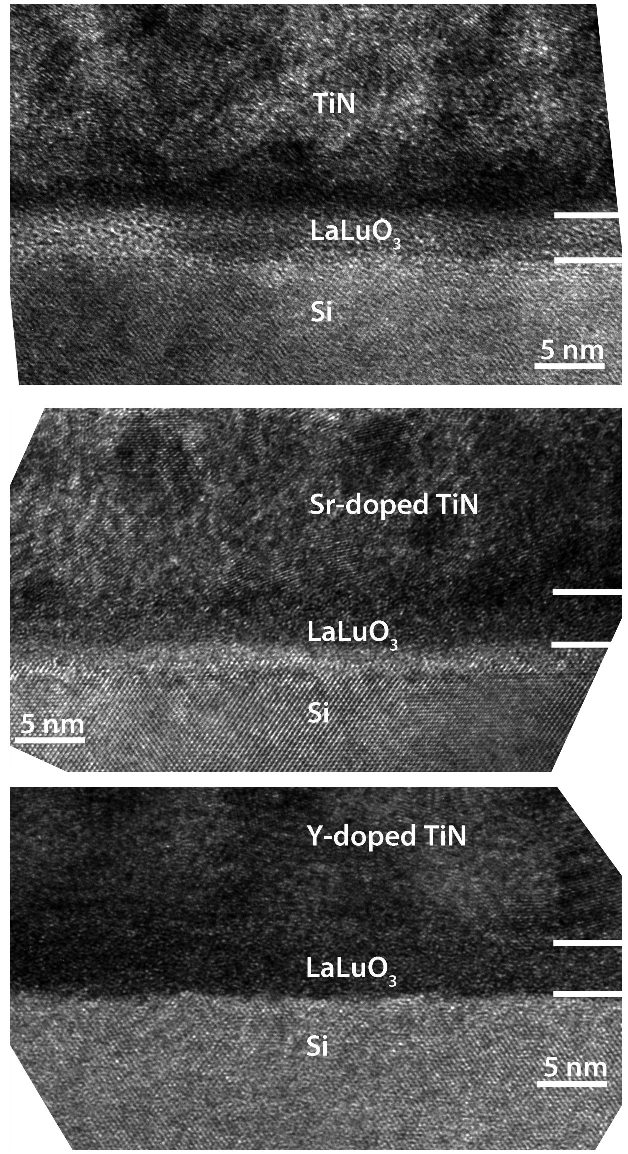


Fig. 3.42 TEM analysis of $LaLuO_3/TiN$ gate stacks with optional doping of TiN.

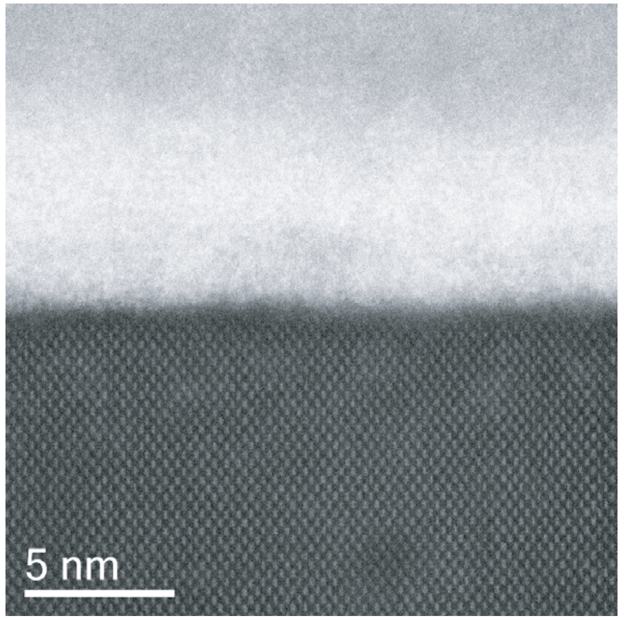


Fig. 3.43 STEM micrograph of the Si/$LaLuO_3$ interface with Sr implantation into TiN.

The interface study on $LaLuO_3$ shows a modulation of the interfacial silicate in dependence of the implanted species. With Y implantation, the reduction of the Si-rich silicate to La-rich silicate or even $LaLuO_3$ seems possible. Therefore, a further investigation of metal gate modifications seems promising, not only for HfO_2 but also with $LaLuO_3$, in order to control the interfacial layer to Si. The large amount of oxygen found in TiN could indicate oxygen vacancies resulting from the reduction of $LaLuO_3$ to metal and oxidation of Y. Such a redox system depends strongly on Y dose and is more difficult to control than the Sr-implanted system. Damage to the high-κ oxide has truly to be avoided to keep its high quality properties. A weakness is the missing capability to measure also the electron loss originating from the implanted species. Consequently, the study implies a modification of the silicate structure, but cannot fully clarify if the dopants are present in the oxide due to ion implantation. As the silicate structure is modified by increased oxygen solubility in the metal gate, a key aspect for lower EOT might be the metal electrode.

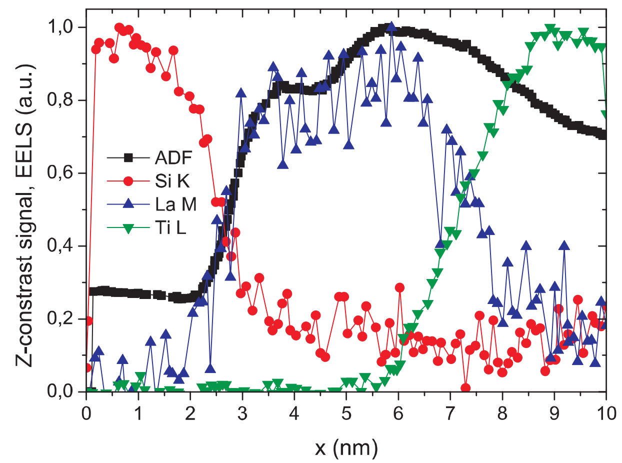


Fig. 3.44 EELS line scans averaged along the interfaces show a Ln-Silicate interface with ~20% less La than in the $LaLuO_3$ on top. The mix zone between TiN and oxide is about 1 nm. The associated ADF line scan along the interface to Si shows a width of 9 Å.

3.3.6 Summary

Several experiments for further downscaling of the equivalent oxide thickness were presented. Among these, the most intuitive solution is the continued scaling of the interfacial layer. The IL reduction was shown to be possible with ultra-diluted HF acid and an ozone passivated Si. The ozone passivation leads already to lower EOT in comparison to RCA cleaned substrates, although the thickness for RCA cleaning is thinner (10 Å) as for ozone passivation (16 Å). With etch back the EOT can be further reduced down to 8 Å.

The use of higher-κ oxides, like $LaLuO_3$, was shown to be scalable to 12 Å but with lower oxide quality and hysteresis as compared to HfO_2. The combination of etch back and $LaLuO_3$ deposition could not successfully be evaluated.

The scavenging of very thin TiN as propagated by some research groups could not be reproduced presumably due to extrinsic oxygen sources during processing. Scaling of SiO_2 was successful by increased oxygen scavenging in the metal gate. Al doped TiN reached EOT values < 5 Å but the process depends strongly on TiN thickness and thermal budget. The results for Sr and Y doped TiN are promising since the thermal budget, considering diffusivities, can be much higher than for Al. However, in EELS the examined elements often interfered with the Si K-shell at 99 eV and hampered the analyses. All evaluated materials may cause difficulties in gate patterning, if structured *after* annealing. The concept of redox systems seems

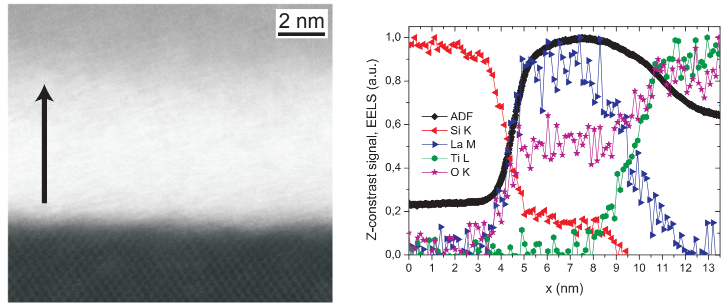


Fig. 3.45 The $LaLuO_3/TiN$ with implanted Y into TiN shows an interface width Si/$LaLuO_3$ of $\sim$ 5 Å. The oxygen content is constant within $LaLuO_3$ and increases to TiN, where Y was implanted. Y itself could not be determined due to large interference of the K-shells. The L-shell was out of detector energy range. The ADF spectrum (black) shows a near constant density over the oxide layer in contrast to the control and Sr implanted samples. The sample drift was below 1%. The arrow indicates the x-axis of the EELS scan (Si to TiN)

also promising for other rare-earth based oxides: The interface study showed capabilities to modify the interfacial layer to rare-earth metal-rich silicates possibly enabling a higher permittivity and thus a lower EOT.

The doping of TiN with Y and Sr was achieved by ion implantation. An adaption of the sandwich structure for doped metal electrodes via *in-situ* thin film deposition and replacement of the implantations is mandatory for scaled device structures.

Chapter 4
Integration of high-κ/metal gate stacks into MOSFETs

4.1 Optimization of TiN as gate electrode for the integration with $LaLuO_3$ and HfO_2

This chapter addresses the necessary optimization steps for the optimization of TiN as a gate electrode on $LaLuO_3$ and HfO_2high-κ gate oxides. The optimization consists of an adjustment of the stoichiometry and the change of its (effective) work function during annealing, the in-situ capping to avoid oxidation of the electrode, stress measurements, and the control of dopant diffusion through ultra-thin gate electrodes. The use of TiN as a blocking layer for other metals is addressed as well.

In a systematic study, Cabral *et al.* showed that most metals with low work function are thermally unstable on Silicon or even agglomerate during high temperature annealing.[27] In contrast, metal nitrides, especially TaN and TiN, were found to be favorable gate electrodes for today's CMOS technology due to their high thermal stability with respect to Si and suitable work functions: as-deposited TaN presents an *n*-type work function ($\sim 4.1\,\mathrm{eV}$) and TiN offers a work function close to *p*-type, i.e. valence band edge E_V $(4.95-5.17\,\mathrm{eV})$.

During high-temperature annealing, the TiN work function is lowered to close to mid-gap energy E_M (cf. Figure 4.1, $\sim 4.61\,\mathrm{eV}$[101,171,172]) of Si. On the advent of the high-κ/metal gate era people thought the binary nitrides can be tunable work function metals,[173] which is only true in the absence of high-temperature annealing.[75,78] The metal nitrides were also believed to be variable scavenging layers due to their tunable stoichiometry.[17,173,174] The high oxygen solubility in TiN is indicated to be maintained for any stoichiometry.[17]

In this work, the first MIPS integration of 10 nm TiN with $LaLuO_3$ or HfO_2 is presented with layers deposited by physical vapor deposition (PVD) in reactive nitrogen plasma. The integration was tested on MOS capacitors (MOSCAPs). The stoichiometry of these layers was determined by RBS to be Ti 1:N 0.8. After high-κ and TiN deposition, the subsequent deposition of 100 nm poly-Si followed by low-pressure chemical vapor deposition (LPCVD, ~ 30 min at 600°C). The activation of poly-Si was performed by a 1050°C spike anneal. The MOS capacitors were fi-

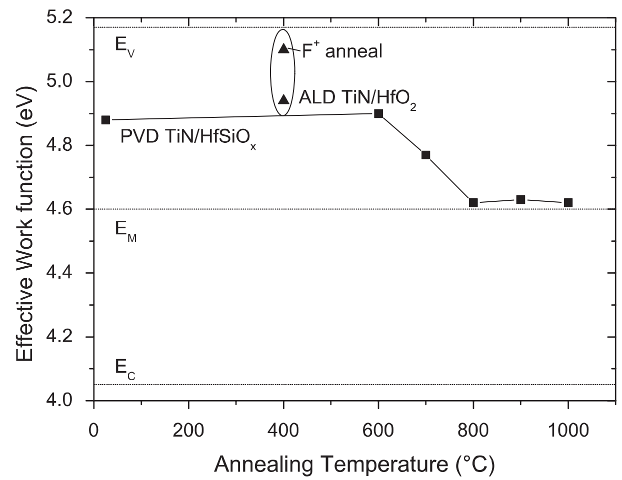


Fig. 4.1 Effective work function vs. annealing temperature as reported in literature (Graph adapted from T. Ando).

nally patterned by RIE using a hard mask of 100 nm Al. The C-V measurements obtained for the substoichiometric TiN and TaN revealed a drastically reduced capacitance and EOT increase, most pronounced for TaN. Moreover, at the time of this study nitrogen optimized deposition of TaN was not possible due to the absence of nitrogen-controlled depositions for TaN (RF sputtering or ALD).

To investigate the origin of increasing EOT, the interfaces were analyzed by means of STEM and EELS. Figure 4.2 shows the simultaneously recorded HAADF signal (black) and the corresponding EELS spectrum of Si (blue). The diffusion profiles were measured by the StripeSTEM[134] method with the simultaneous measurement of HAADF and averaging of EELS spectra parallel to the interface.

The *z*-contrast of the poly-Si is around 20% of the maximum density. At the interface to TiN, the Si signal vanishes within 6 nm. In parallel the density increases within TiN up to the maximum density found in the oxide layer. The nominal density of TiN with respect to $LaLuO_3$ is around 64%, which matches the shoulder measured in the ADF scan at 17.5 nm within 5% error bar. Based on ADF $LaLuO_3$ and TiN are assumed to intermix by ~ 2.5 nm (17.5 to 20 nm). The inter diffusion was interpreted on the basis of thermodynamics to be mainly promoted by the nitrogen deficiency of TiN.

To reduce intermixing of TiN and $LaLuO_3$ or poly-Si the nitrogen content of TiN was optimized to a stoichiometry close to 1:1, i.e. above 1:0.95 as measured by RBS. Therefore the nitrogen flow during PVD deposition was doubled to 32 sccm (standard cubic centimeter per minute). To increase also the nitrogen ion current density

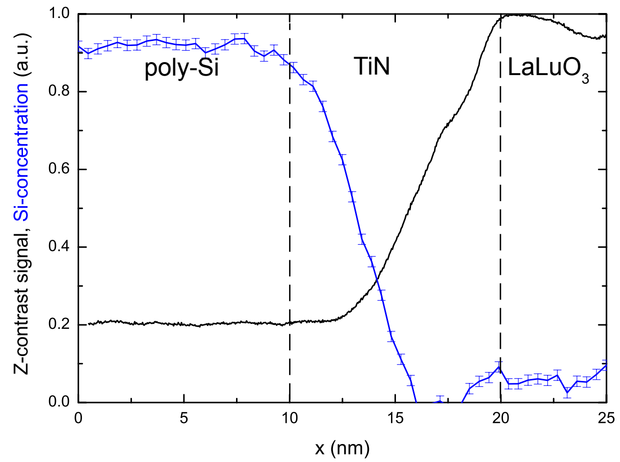


Fig. 4.2 Simultaneous measurement of z-contrast and EELS at the interface poly-Si/$TiN_{0.8}$. A significant intermixing of Si and Ti is found, which stops within TiN. TiN and $LaLuO_3$ can be assumed to intermix according to the density change measured in the ADF scan. Dashed lines indicate the estimated interfaces before annealing.

during reactive sputtering, the RF power was adjusted from 1 kW to 1.5 kW. The experiments were carried out on a 200 mm wafer PVD tool Oerlikon LLS EVO II. As described by Pritschow,[175] an exact film stoichiometry of 1:1 is hardly achievable without active pressure control during deposition due to hysteresis effects. The target periodically changes according to the sputter frequency between an isolating (nitrided) surface and a metallic surface causing this hysteresis effect[176] Hence, the obtained nitrogen fraction of above 0.95 fraction nitrogen without capping seems reasonably high for PVD and matches literature reports[76] As described later, the capping of PVD-deposited TiN can even increase the nitrogen content avoiding nitrogen out-diffusion and oxidation of Ti. The sheet resistance increased from 61 $\mu\Omega$cm for $TiN_{0.8}$ to 90 $\mu\Omega$cm for $TiN_{>0.95}$ and offers still low-resistive metallic properties. The TiN of comparable stoichiometry by ALD deposition features a sheet resistance of 300–400 $\mu\Omega$cm, which is explained by lower density or higher impurity concentrations. Nevertheless, the sheet resistance of the thin metal gate layers is nearly negligible compared to 100 or more nm contact metal or poly-Si on top.

The optimized recipe offers relatively conformal depositions on nanostructures due to the high pressure ($p = 0.33$ Pa, before optimization $p = 0.25$ Pa) during deposition combined with the wafer rotation perpendicular to the sputter targets. If the pitch between structures is not too narrow, conformal deposition of TiN is possible as shown in Figure 4.3 and tested for a pitch of 200 nm.

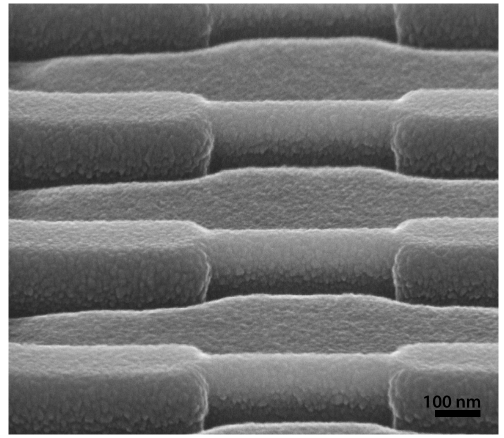


Fig. 4.3 Nanowire array with a pitch of 400 nm between the wires after deposition of 60 nm TiN. The wafer was mounted perpendicularly to the rotation direction, which leads to conformal deposition under relaxed pitch requirements. (Data kindly provided by L. Knoll).

Comparing C-V characteristics of 15 nm TiN/10 nm $LaLuO_3/p$-Si before and after optimization of the nitrogen content in TiN (Figure 4.4), the capacitance is found to decrease with increasing annealing temperature for the Ti-rich TiN (left). It is stable for the nitrogen-rich TiN (right). N.b. 40 nm TiN were used for the red curve, therefore comparable flatband voltage is observed after annealing. Moreover, the Ti-rich TiN suffers from generation of charged defects in the oxide layer, which becomes mainly visible in larger hysteresis, i.e. movable charges within the oxide. A silicate formation takes place in both cases. Ti can be assumed, based on the ADF scan, to intermix with $LaLuO_3$. Hence, the increase of hysteresis for Ti-rich TiN can be accounted to an interaction at the top interface $TiN/LaLuO_3$. In both cases the D_{it} was extracted by the conductance method[50] to amount to $6 \cdot 10^{11}/(\mathrm{eV\,cm}^2)$, which indicates comparable interface quality. The aforementioned study of silicate generation has been identified to be the key aspect to the capacitance reduction in stoichiometric TiN. For the Ti-rich TiN capacitance is reduced already after a single RTA step. Two possibilities exist, which were not further investigated: Either the Ti-rich TiN promotes a stronger oxygen gettering, as interpreted in the XPS depth profile for a sample annealed 60 s at 1000°C. This could cause oxygen deficiency in the high-κ layer. Alternatively, the intermixing between $TiN/LaLuO_3$, indicated by EELS, is already present at low-T. The EWF of *N*-rich TiN in a MOSFET was found to feature *n*-type band-edge behavior for 3 nm $LaLuO_3$. Thus, a more *p*-type EWF seems difficult to achieve, since TiN offers already high work functions on the *N*-rich side. A possible alternative for *p*-type could be the use of W(N) as investigated by other groups.[177]

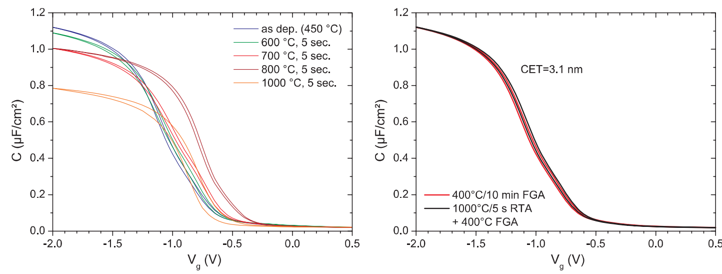


Fig. 4.4 (left) C-V measurements on gate stacks with PVD $TiN_{0.8}/10\,nm$ $LaLuO_3$ after different annealing temperatures and subsequent FGA at $400^\circ C/10$ min. V_{fb} is mostly $-0.65\,V$ and shifted for $800^\circ C$ to $V_{fb} = -0.4\,V$; movable charges increase the hysteresis observed for $1000^\circ C$. (Right) 10 nm $LaLuO_3$ capacitor with optimized PVD $TiN_{0.95}$ behaves stable for $1000^\circ C/5\,s$ anneal. The flatband voltage $V_{fb} = -0.65\,V$ is identical for the RTA and pure FGA sample.

4.1.1 Oxygen barrier layers on TiN

The high oxygen solubility in TiN is reduced by its oxidation during ex-situ processing (cf. XPS depth profiles). Therefore, in-situ capping of TiN by a material blocking oxygen is an effective way to avoid the oxygen saturation of TiN.[70] The in-situ capping can be achieved by either PVD-deposited or ALD/AVD-deposited amorphous Si with a thickness of 2–5 nm.[72] During ex-situ treatment, the first nanometer of a-Si oxidizes while elemental a-Si remains to protect TiN. Figure 4.5 shows a comparison of TiN (*N*-rich if not otherwise noted) without Si-cap after 2 month of air exposure. The 46 nm film is oxidized to 14% with respect to the Ti fraction. In parallel, the film has slightly lost nitrogen (fraction 0.93). The width of N and O peaks match the Ti peak, hence a complete oxidation of TiN in total depth is observed. Literature reports the columnar structure of the poly-crystalline grains to be sources of this oxidation behavior.[103, 175] The sample r.h.s of Figure 4.5 was in-situ capped with 3 nm a-Si. After the same air-exposure, the sample shows only an oxidized surface (SiO_2). The bulk of TiN shows nearly zero oxygen content (below RBS resolution limit of 2%). The stoichiometry was determined from the peak integration of the total count rate to amount to $Ti_1N_{1.1}$. The nitrogen fraction exceeds the values for uncapped samples, measured directly after deposition. Consequently, the Si-cap blocks oxygen from entering the stack and out-diffusion of nitrogen. For stress engineering of the deposited films, the over-stoichiometric layers might induce additional stress compared to uncapped layers.

To assess the quality and efficiency of the Si-cap as protective layer, XPS measurements were performed. 15 nm TiN were deposited by PVD on Si (100) substrates and capped by 3 or 6 nm Si. Under the use of the maximum information depth of the laboratory XPS (~ 50–$100\,Å$), the determination of the oxidation states in the capped material system is possible. After air exposure, the sample was trans-

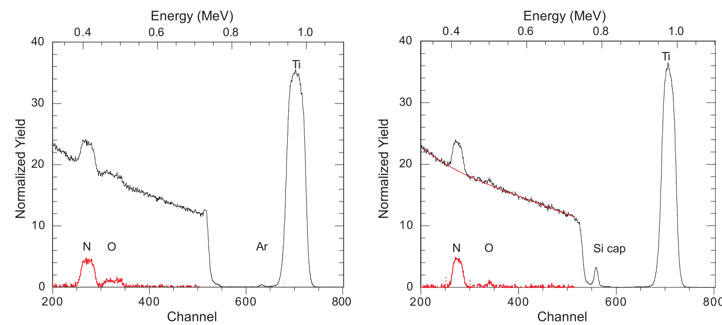


Fig. 4.5 RBS/C channeling measurements of TiN (He^+ 1.4 MeV, 170° scattering angle, 0° incident angle, 100 μC), Si background signal has been subtracted to integrate the areal coverage of N and O (red curve). The beam was aligned along the Si(100) surface normal. (Left) After two-month air-exposure 46 nm TiN without Si-cap is oxidized over the complete layer indicated by the width of N, O, and Ti peaks. The stoichiometry was determined as Ti:N:O = 1:0.93:0.14. A poly-crystalline Si layer with Ar content is visible below the sputtered TiN film. (right) 20 nm TiN with 3 nm a-Si cap shows nearly no oxidation apart from the surface peak of SiO_2. The stoichiometry is Ti:N:O = 1:1.1.

ferred into the XPS chamber and 50% oxidized Si was measured in several (sub-) stoichiometric bonding states (cf. Figure 4.6, bottom right). In parallel, the sample with a 6 nm cap thickness showed the same peak shapes but at lower intensity for the capped layers. Since 50% of the 3 nm Si-cap remain non-oxidized, the Si-cap is proven to be effective in oxygen blocking, as determined in the investigations over a long-time shown in Figure 4.5. Nevertheless, a shoulder in Ti 2p was found +2 eV away from the TiN contribution, which is still a topic of ongoing, controversial discussion. The scientific arguments range from the existence of TiON[178] to the presence of shake-up processes dependent on the nitrogen content.[103,179] The existence of TiON positioned between the chemical shift for TiO_2 (459–460 eV) and TiN (454–455 eV[116]) is doubtless the most intuitive argument. However, as Saha[103] and coworkers pointed out, the presence of surface TiO_2 is reported in many studies at a higher binding energy ($\sim$ 459 eV) independent of its detailed stoichiometry. Moreover, the samples processed for this study and in the study of Saha *et al.* do not show characteristic chemical shifts in O 1s, where mostly Si-O bonds are visible.

The strong satellite features observed by Saha for close-to-stoichiometric TiN are not found in the studied samples. Here, the feature is much less pronounced. It seems more likely that the observed features in the cited study originate from contaminations present in the nitrogen used. To check this source of error for the depositions used here, new high purity gas bottles (Ar 99.9999%, N_2 99.9995%) were used, a He leakage test was performed without indication for leaks. Residual gas analysis showed the partial pressures of water, carbon dioxide and oxygen to be six orders of magnitude lower than the sputter pressure (10^{-9} vs. 10^{-3} mbar).

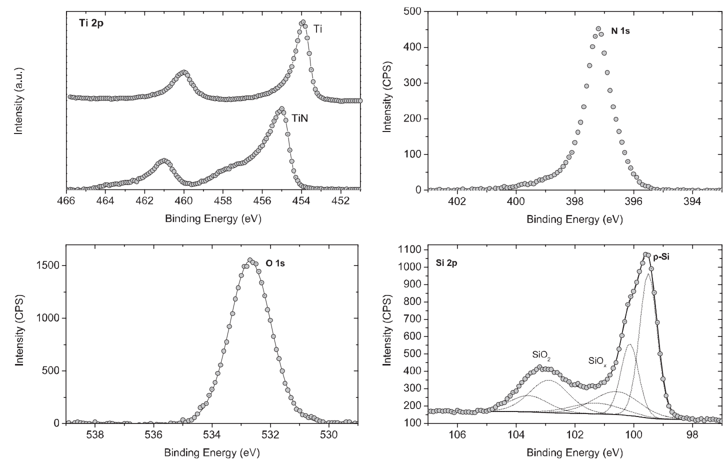


Fig. 4.6 XPS spectra obtained from 15 nm TiN/3 nm a-Si cap. Ti 2p from TiN is plotted in comparison to a layer of pure Ti capped by a-Si. The nitrogen content in the layer leads to a chemical shift of about 1 eV to 455 eV and a shoulder at 457 eV (explanation see text). The asymmetric N 1s spectrum indicates the contribution of several binding states within TiN. O 1s is highly symmetric and can be fitted by one or three components, which are also visible as sub-stoichiometric SiO_x in the Si 1s signal.

The gas flows of Ar/N_2 lead not to a rise of those partial pressures. Since at least three binding states can be identified from the asymmetry in the N 1s spectrum and a small shoulder in O 1s to lower binding energy, a small oxygen contamination of TiN is still possible. If a further reduction of TiON is possible remains unclear.

The importance of Si capping on the TiN, already at low processing temperatures, is shown by the C-V measurements in Figure 4.7. The capacitive test structures consist of RCA-cleaned substrates, cleaned in the same run, 2 nm ALD HfO_2 deposited on all substrates in parallel and TiN (+a-Si) deposited by PVD. A standard process flow for MOS capacitors was employed. The a-Si capped samples show an improved total capacitance, whereby this improvement is higher for the thick TiN layer at low temperature (max. $400°C/10$ min FGA). The flatband voltage changes from 10 mV for the non-capped sample to 70 and 80 mV with Si-cap. The hysteresis for all samples is around 10 mV and the D_{it}-level is 1 to $6 \cdot 10^{11}/(\text{eV cm}^2)$, lowest for the sample without cap.

At room temperature and at 400°C reaction temperature the enthalpy for oxidation of unsaturated Ti and TiN is negative, meaning thermodynamically favorable. If the oxidation due to air exposure can be prevented, oxidation within the gate stack becomes possible. At moderate temperature this process can be even increased by the use of thicker metal gates, Under the assumption of oxygen diffusion inside the

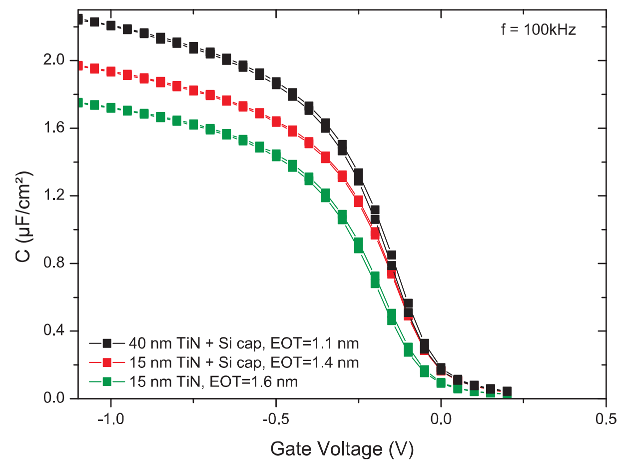


Fig. 4.7 C-V measurements on 2 nm HfO2/Si(100) with different TiN thicknesses and optional in-situ Si-capping of TiN. The capping improves the total capacitance of the gate stack. Max. thermal budget was $400°C/10$ min FGA. The effective work function is ~ 5.1 eV and 5.0 eV for the uncapped sample.

gate stack[104] a higher oxygen solubility is achieved. An oxidation of SiO_2 by a reduction of possibly existing TiO_2 within TiN is not thermodynamically favorable at this temperature. Hence, the capacitance increase can be explained by an oxidation of TiN and reduction of interfacial SiO_2 under a slight increase in D_{it}. N.b. the total capacitance of the RCA-cleaned Si-substrate and 2 nm HfO_2 can reach an EOT of 1.4 nm as found previously for MOSFETs and reported in literature.[70] This would be also the theoretical limit, if the interfacial layer had not scaled during processing. However, the deviation in the total capacitance are due to a certain run-to-run variability in the ALD process, but the overall comparison is still valid for TiN since the substrate passivation and dielectrics deposition were done each in the same run.

4.1.2 Film stress measurements on TiN

Metal nitrides like TiN can have significantly different properties based on the deposition method, deposition parameters, and film thickness. The aforementioned stoichiometry control of TiN is one possibility to tune them. The EELS analysis of TiN and the thermodynamic approach have shown that stoichiometric TiN is beneficial for the thermal stability of the gate stack. On the other hand, different thermodynamic reactions during annealing depend on stoichiometry, i.e. amount of radical Ti

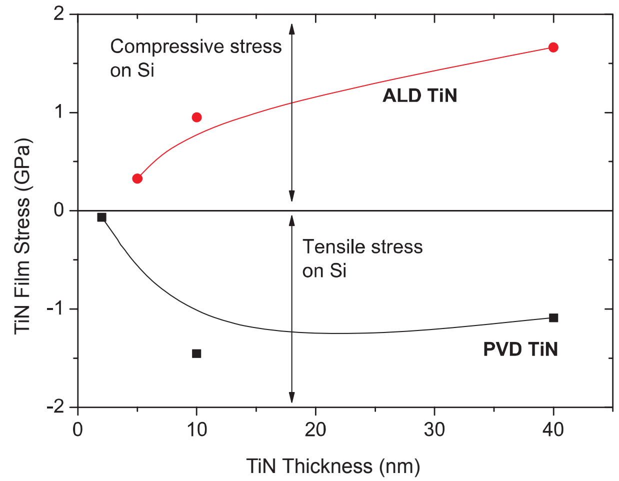


Fig. 4.8 Thin film stress of TiN vs. film thickness. Remarkable differences between AVD and PVD deposition are found. PVD films deliver high tensile stress on Si (compressive TiN). In contrast, AVD TiN is less but tensile strained, leading to compressive stress in the underlying Si.

within TiN, and film thickness (SiO desorption vs. TiN oxygen scavenging). At a consequence,stoichiometric TiN should be used. Hence, if the EOT scavenging depends also on thickness, the intrinsic film stress and its behavior with film thickness must be known.

The stress in the metal film was calculated from the wafer bow measured by white-light interferometric surface probing (MPI Halle), whereby the wafer bow before and after coating is measured. Possibly different stress in coating and wafer leads to strain in the wafer of opposite sign with respect to the thin film stress. The sign of the film's stress can then be obtained by the measured curvature direction: Concave (positive curvature) indicates tensile film and convex (negative curvature) marks compressive stress. The exact film stress is then obtained from the difference in wafer curvature as[180]

$$\sigma = \frac{1}{6}\left(\frac{1}{R_{\text{post}}} - \frac{1}{R_{\text{pre}}}\right)\frac{E_{\text{Si}}}{(1-\nu)}\frac{t_{\text{sub}}^2}{t_f}, \tag{4.1}$$

where by R_{post} and R_{pre} are the substrate curvature before and after coating, ν is the Poisson ratio, E_{Si} is Young's modulus for Si, t_{sub} and t_f are the substrate and film thickness, respectively.

Figure 4.8 shows the TiN film stress at different film thickness for two different deposition methods. Both deposition methods and recipes deliver stoichiometric TiN. PVD TiN was additionally capped by 3 nm a-Si to prevent oxidation and nitrogen out-diffusion. The measurement with a-Si cap should still be valid due to the low density of a-Si in general and particularly for this a-Si deposition at high pressure (0.5 Pa). PVD deposition leads to compressive TiN with this particular recipe, which even increases for thinner films. The measurements are thereby in qualitative agreement with the findings of Kang[181] (industry report, no scale provided) and quantitatively below the measurements of Lim[182] for pulsed PVD depositions. Kang *et al.* explain the film stress decrease below 3 nm with island growth for PVD TiN and excluded measurements below 5 nm thickness. Although island growth was not observed in our TEM investigations,an increased roughness for 2 nm TiN layers was detected in comparison to 5 nm thin layers. Likewise, Stranski-Krastranov growth could be possible but this investigation was outside the scope of the present work. However, these thickness variations can certainly serve as explanation for the decrease of compressive stress for 2 nm TiN films. Addressing the ALD TiN, the deposition method delivers tensile strained films, which in turn induces compressive stress on the Si substrate. Here a special type of ALD, called Aixtron AVD® was used, where the precursor is injected by pulses under constant NH_3 flow. In contrast to the PVD films, the stress of ALD TiN decreases for decreasing film thickness and is much lower in magnitude than the corresponding compressive stress for PVD TiN.

According to the stress measurements, the nMOS mobility on Si is expected to increase with thinner PVD TiN due to the tensile strain effects[44] on Si as reported by Kang *et al*. For ALD TiN the choice of very thin TiN prevents the degradation of the electron mobility.

4.1.3 Boron penetration and effective work function in MIPS gate stacks with TiN

For decreasing the layer thickness of PVD TiN the tensile stress on Si increases and thus mostly enhances the electron mobility in scaled devices.[65,181] In parallel, the scavenging by SiO desorption becomes possible for very thin TiN, whereby thick layers show normal oxygen gettering. The remaining question is the influence of dopants on the TiN work function when the TiN thickness is scaled down. Ion implantation is planar2d Poisson-distributed. Therefore, work function variations could be caused by the presence of several few dopants at the metal gate/high-κ interface. Such effects were already observed for random dopant fluctuations[183] and may cause additional variations in the device threshold voltage.[79,80] Bae *et al.*[65] observed steeper cumulative distribution functions (CDF) for 10 nm TiN compared to 5 nm indicating possible variations due to Boron penetration. This problem might mainly arise for pMOS in conjunction with a MIPS process flow due to its much higher diffusivity in Si compared to e.g. as for nMOS.

Figure 4.9 shows the Boron depth distribution after full processing of an HfO_2/TiN MIPS gate stack for 15 nm (left) and 2 nm (right) TiN metal gates. The relative sensitivity factor due to matrix effects for Boron in TiN differentiates one order

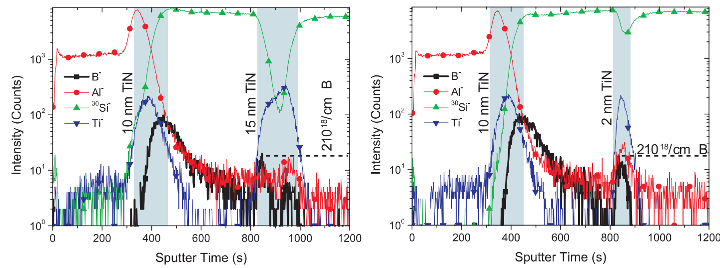


Fig. 4.9 Observed Boron distribution (black) in MIPS gate stacks after full processing (RTA 1050°C spike + 400°C FGA). The final gate stack consists of Si/2 nm HfO_2/x nm TiN/100 nm a-Si (3 keV B implantation, $D = 3 \cdot 1015/\text{cm}^2)/10$ nm TiN as diffusion barrier for Al/200 nm Al (M1). The top barrier layer TiN plus M1 were deposited after doping activation and prior to final FGA; the top barrier effectively blocks the diffusion of Al into Si. The peak concentration of B in TiN was calculated from prior calibration to amount to $2 \cdot 10^{18}/\text{cm}^3$. Matrix effects are found within calibration, such that the count rate of Al and B in TiN exceeds the rate in a-Si. Left plot shows 15 nm TiN, right shows 2 nm TiN.

of magnitude to the one in a-Si. Boron was found within TiN to distribute over the layer. However, the peak concentration of B after spike annealing and FGA was found to amount to $2 \cdot 10^{18}/\text{cm}^3$. Hence, the boron penetration into TiN is a weak effect, which should not influence the work function of bulk TiN (density $\sim 1.06 \cdot 10^{23}/\text{cm}^3$).

The effective work function (EWF) of a MIPS gate stack after full processing was extracted for different PVD TiN thicknesses on HfO_2 and a substrate doping of $1 \cdot 10^{16}/\text{cm}^2$ Boron from C-V measurements. The substrate passivation was performed for all samples by an RCA cleaning and high-κ deposition followed by 3.5 nm (30 cycles) ALD HfO_2. The resulting EWF ranges from 4.8 eV down to 4.6 eV for thin TiN. The nearly constant EWF for 15 nm and 40 nm TiN is interpreted as the onset of bulk properties in the layer. For p-type band-edge (BE) work functions, a value of 5.1 to 5.17 eV is favorable and is not indicated to be reached by high-T TiN processing. Also for low-T, an EWF $\sim$ 5.0 eV was obtained (cf. Figure 4.7). An explanation is the low doping of the p^+ substrate used for MOS capacitors. With a highly doped p^{++} substrate, the flatband voltage is increasing and close to p-type band-edge operations become possible. The results show the obstacles in the use of TiN as dual (n/pMOS) work function metal; midgap work functions are within reach either by high-T annealing or low-T plus a change of stoichiometry. BE functionality with respect to HfO_2 is only given for p-type devices and moderate temperatures.

Additionally to C-V measurements, an XPS study was performed for a gate stack of $\text{Si(100)}/SiO_2/2$ nm $HfO_2/2$ nm TiN/3 nm a-Si to assess the band structure of the gate stack and possible EWF changes. Three different wafers were compared: The as-deposited case, a sample with 400°C/10 min FGA after deposition, and a sample,

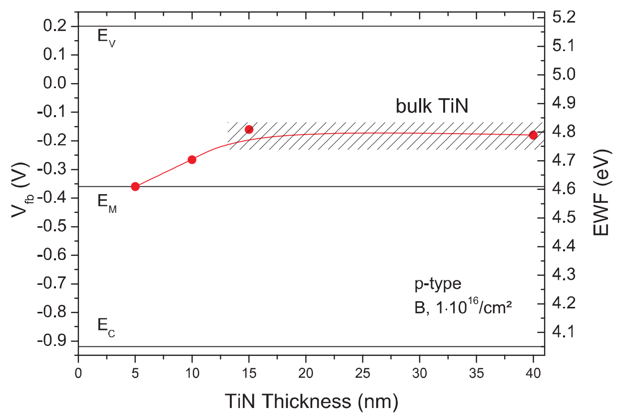


Fig. 4.10 V_{fb} and EWF vs. thickness of TiN obtained from C-V measurements. After high temperature annealing the EWF of 15 and 40 nm TiN (PVD) are comparable and attributed to bulk behavior. For thinner TiN the flatband voltage is reduced and near midgap work functions are obtained.

which underwent a simulation of a gate-first process. The latter sample underwent $600°C/10$ min (CVD sim.), 1050°C RTA, and final $400°C/10$ min FGA. All wafers were cleaned by an ozone process prior to ALD HfO_2 deposition.

During XPS measurements the substrate was grounded and the spectrometer work function is 4.218 eV referenced to the same ground potential. The amorphous Si-cap is expected to be conductive due to a Boron doping of $10^{18}/cm^3$. However, the difference in work functions will lead to a contact voltage ϕ between sample surface and spectrometer.[184] As we saw earlier in the XPS chapter, the binding energy can be written as:

$$E_{\text{bin}} = h\nu - E_{\text{kin}} - \phi. \tag{4.2}$$

The contact voltage ϕ includes effects from additional charges within the gate stack as also the work function difference. The XPS measurement is performed with respect to the Fermi level E_F, which is the same between spectrometer and grounded gate stack. If the gate oxide is not assumed to change its chemical state, a change in binding energy can be solely attributed to a change in ϕ, i.e. a relative change in effective work function of the gate stack. The change in BE was observed for both O 1s and Hf 4f. Moreover, the absolute value of the EWF can be determined, if the vacuum level BE_0 of the oxide (here HfO_2) is known. Then the EWF is obtained from the difference of both binding energies:

$$\text{EWF} = \text{BE}_0 - \text{BE}. \tag{4.3}$$

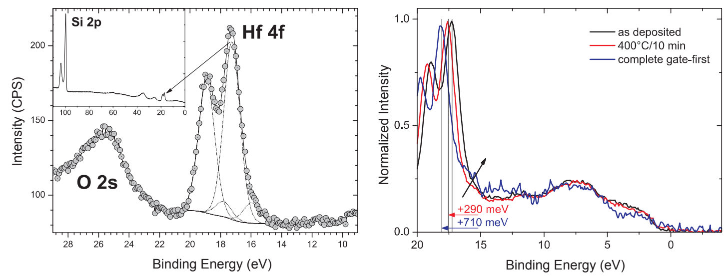


Fig. 4.11 (left) Hf 4f spectrum (as deposited) buried by a-Si/TiN. The two smaller peaks indicate a silicate formation at the interface. Inset: To avoid shifts of the kinetic energy axis and to determine the necessary information with minimum instrumental error bar, the spectrum was measured from Fermi level (0 V) up to the Si binding level (99–105 eV). (right) Detail window of the wide scan for three different thermal budgets ranging from Fermi level to Hf 4f. The valence band maximum (VBM) of Si was extracted for the as-deposited sample to be ~ 0.6 eV. $VBM_{Hf} \approx 4.13$ eV. For the two annealing steps the BE_{Hf} shifts 290 and 810 meV.

The experiment was carried out with single detail scans for O 1s, C 1s, N 1s, Hf 4f, Si 2p, and Ti 2p. Two avoid any instrumental error due to energy shifts in the high-voltage electronics of the XPS, the important window from Fermi level to Si 2p was measured in one continuous scan (cf. inset of Figure 4.11 (left)). Fortunately, the O 1s energy scan (520-540 eV) shows the same shifts in BE, thus the instrumental error seems negligible for future experiments (in the case of the PHI Versaprobe II XPS).

Figure 4.11 shows the Hf 4f spectrum for the as-deposited sample. Typically, a small silicate contribution is indicated by the two peaks (7/2 and 5/2) at smaller BE. The energy splitting between Hf $4f_{7/2}$ and Hf $4f_{5/2}$ was extracted by a peak deconvolution to amount to 1.7 eV. After annealing, the BE of Hf 4f has changed visibly as displayed in Figure 4.11 (right). The binding energy changed $+290$ meV for the 400°C FGA and $+810$ meV for the gate-first simulation. *The valence bands and the Si 2p substrate peaks are still aligned with each other*. The small deviations between 0 and 10 eV BE are caused by the normalization; the gate-first sample has less Si on top (at comparable C 1s surface contamination), which leads to a decrease of the Si 3s and Si 3p signals. The silicate fraction obtained from deconvolution increases for the high-T anneal from 10% to 15% of the total intensity (cf. arrow). The Hf silicate signals were found at 15.8 eV. An overview of the changes in BE is tabulated in Table 4.1. The change in binding energy allows a conclusion in terms of metal work function ± 100 meV and depends strongly on the experimental conditions. The C-V measurements obtained earlier are much more precise (± 10 meV) for a known substrate.

The advantage of the XPS measurement is the direct observation of EWF change and control of chemical reactions within the gate stack. The measurements allow also an estimation of the band diagram of the gate stack. The valence band consists

Table 4.1 Overview of the orbitals before and after gate stack annealing as obtained by peak deconvolution. WF is the estimated work function from XPS measurements.

Sample	$BE_{Ti\ 2p3/2}$	$BE_{Si\ 2p_{3/2}}$	$BE_{Hf\ 4f_{7/2}}$	WF	ΔEWF (ΔBE_{Hf})
as dep.	454.95 eV	99.55 eV	17.26 eV	–	0 eV
400°C/10 min	455.95 eV	99.53 eV	17.55 eV	4.9 eV	−0.29 eV
gate-first	455.68 eV	99.52 eV	17.97 eV	4.5 eV	−0.71 eV

of the convoluted valence bands of SiO_2, Si, HfO_2, and TiN. Si and SiO_2 are present at both top- and bottom interface, which hampers the accuracy of the extraction. By using a step background and extrapolating the valence band edge to zero, the valence band maximum (VBM) of Si can be obtained, as explained by Miyazaki.[185] Subsequent subtraction of the current valence band and iteratively repeating the extraction for SiO_2 and HfO_2 the VBM of both materials can be determined from a single measurement. The VBM of Si and its offset to SiO_2 was calculated from the difference of core level and valence band maximum (98.95 eV) as given in literature.[83]

The Si/SiO_2 system was determined on a sample with thermal SiO_2 without gate stack but with the same substrate doping. The use of two samples is necessary to exclude any contribution of the higher-doped Si-cap: The higher doped Si-cap was found to have a +150 meV larger VBM as the lower doped substrate, which agrees with the doping-dependent Fermi-level calculations by Sze.[24] Thus, one has to take extreme care in applying XPS energy calibration by a fixed, tabulated Si orbital, because those values are often unreliable in the fact that doping levels are mostly neglected.[116] The band gaps of HfO_2 and SiO_2 were taken from literature.[83,147,186,187] The VBM_{Si} is then obtained from Table 4.1 as $99.52 - 98.95 = 0.57$ eV. The iterative extraction of the VBMs and calculation of energy offsets yield the band diagram as shown in Figure 4.12. The conduction band offset (CBO) from TiN to HfO_2 was determined to ~ 1.6 eV and is close to XPS measurements reported by Copel.[83] However, a comparison to internal photoemission (IPE) experiments done by Afanasev *et al.* reveals a certain deviation from the IPE measurements. The offsets for Si/SiO_2 agree within ± 150 meV, but the band offsets for HfO_2 after annealing deviate for constant band gap 1 eV from their IPE results for deposited HfO_2.

Hence, the band structure determination by XPS is getting closer to reference measurements but is still a proof of concept with respect to instrumental uncertainty. Increased measurement accuracy in terms of better signal to noise ratio for the VBM could still improve the result below 100 meV. Moreover, the shift in the HfO_2 band offsets is larger than instrumental uncertainty. The origin can mostly be explained by the shift of 710 meV observed in the HfO_2 signal (Figure 4.11). The annealing leads to a shift of the band offsets compared to the as-deposited sample. Likewise, the barrier height increases for higher EWF in agreement with usual calculations. For the as-deposited sample, the band structure is in agreement with the IPE measurements of Afanasev. The advantage of the gate stack analysis by XPS is surely its capability to assess buried structures relevant for semiconductor physics.

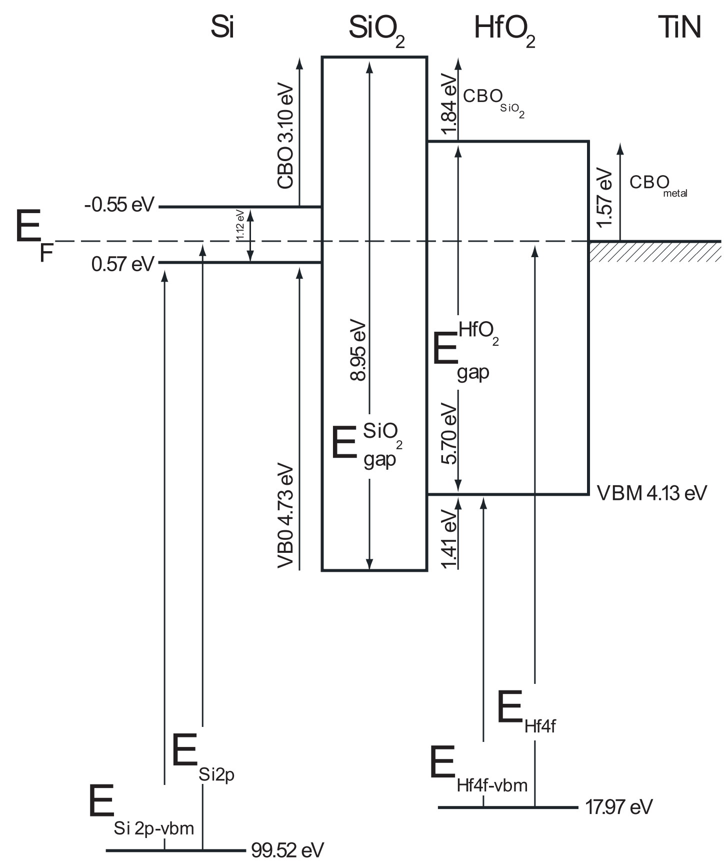


Fig. 4.12 (not to scale) Band energy diagram for the 1000°C annealed gate stack. The Si-cap is not drawn to the right, because the CBO extraction for TiN-HfO_2 depends only on BE_{Hf4f} and the VBM_{Hf}.

4.1.4 TiN as metal diffusion barrier

The left TiN layer in Figure 4.9 was introduced between gate stack and metallization because Al spikes were found to penetrate a-Si even at low-T annealing $(400°C/10\,\mathrm{min})$ and confirmed by literature as a vacancy-mediated mechanism[156] in Si. Furthermore, the diffusion increased, if a NiSi interlayer between Al and a-Si

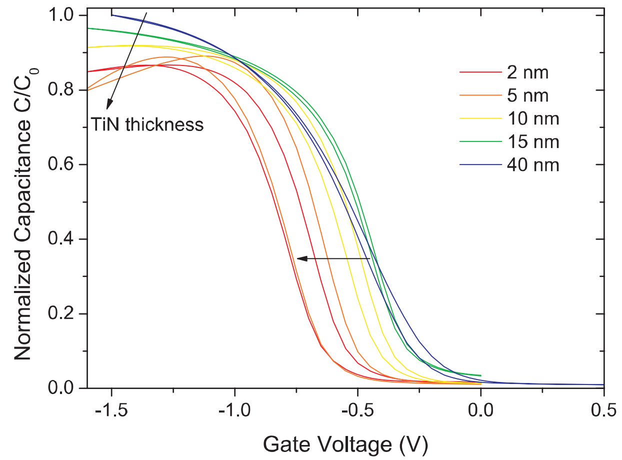


Fig. 4.13 Normalized capacitance for 2 nm HfO_2 with different TiN thicknesses. TiN was annealed in direct contact with Al without poly-Si or barrier layer. Normalized to the accumulation capacitance of the MOS stack with 40 nm TiN.

was used. A 10 nm TiN barrier layer between a-Si and Al blocks effectively Al diffusion into the stack, because below 600°C the intermixing of TiN and Al is diffusion limited.[169]

The diffusion length of Al in TiN at 400°C/10 min was estimated to amount to 2.9 nm and the concentration after a 10 nm TiN barrier to 2% according to literature values[156] for the diffusivity and Fick's law. However, this can only serve as an estimation due to possible deviations between reference and own materials. The shallow decay in the Al signal after the TiN barrier is mainly caused by the overlapping background of BO^- present in a-Si and Al^-. The pile-up in front of top TiN and in the lower TiN might be overestimated due to a change in the sputter matrix.

Nevertheless, contacting TiN directly by Al can lead to work function shifts due to Al diffusion to the interface TiN/high-κ layer.Therefore, the thermal budget should be checked by calculating the diffusion kinetics. Figure 4.13 expresses the need for this additional barrier layer: The MOS capacitors were processed with different thicknesses of TiN (cf. label), 3 nm a-Si capping, and 2 nm HfO_2 and underwent a spike anneal prior to 5 nm TiN (adhesion mediation)/200 nm Al metallization. The MOS stack was subsequently tempered in FGA (400°C/10 min).

N.b., in contrast to other results shown in this work no additional layer was between TiN and metallization. The plot shows a decrease of accumulation capacitance with decreasing TiN thicknesses due to increasing conductance. In parallel, the hysteresis increases from 20 mV to 150 mV. For the thinnest TiN compound of about 8 nm, the diffused Al concentration is only 5% according to Fick's law.

In contrast, this degradation was not observed for Al/Si alloys or pure poly-Si gate stacks, which both were characterized in this work. Thus, the high concentration of Al leads to a degradation, whereby gate stack properties are improved in the case of highly diluted NiAlSi alloy in contact with TiN.

Similar diffusion barrier effects are also found for Cu. Copper diffuses for the used anneal in the millimeter-range[156] in Si. With a TiN barrier layer between Cu and Si, the diffusion is asymptotically close to zero. According to reports of successful integrations of TiN/W,[59,60] this likely holds true for rapid formation of tungsten silicide in Si.[156]

Consequently, the designing of suitable barrier layers is necessary to prevent unwanted diffusion into the high-κ layer or even into the channel. The design has to be chosen according to the thermal budget, which is applied in backend-of-line (BEOL) anneals, i.e. after contact layer formation.

4.1.5 Summary

The optimization of the stoichiometric of TiN deposited by PVD was presented to lead to metal gates stable in contact with a-Si and HfO_2. In contact with $LaLuO_3$, inter diffusion could be mostly prevented as proven by photoemission experiments. In addition, capacitance reductions observed for non-stoichiometric TiN could be avoided. An in-situ deposited oxygen barrier on TiN prevents direct and long-term oxidation of TiN and leads to better process control. The effective work function of TiN was found to be 5.1 eV after a 400°C anneal with a-Si cap and 5.0 eV without capping. The high-T treatment within a MIPS process lowers the EWF depending on the metal thickness to 4.6 eV (midgap, thin layers) and 4.8 eV (bulk TiN), respectively. The stress induced by TiN on Si is compressive for ALD TiN and tensile for PVD TiN. The band diagram of the $Si/HfO_2/TiN$ system was successfully measured by laboratory XPS on thin layers. The results are in agreement with other methods like internal photoemission and allow the observation of process-induced changes to the band-alignment. Work function modulation by diffusion of metallization (Al) could be suppressed by 10 nm TiN barrier layer. The WF modulation by Boron dopants is negligible. The design of suitable barrier layers depends on the reaction kinetics and diffusion. The barrier layer should be adjusted to the thermal budget applied after metallization.

4.2 Device integration of $LaLuO_3$ and HfO_2

4.2.1 Device fabrication

To investigate the device performance with $LaLuO_3$ and HfO_2high-κ dielectrics in MOSFETs, long-channel transistors were fabricated with a gate length L_g ranging from 1 to 20 μm on Silicon-on-Insulator (SOI) with a body thickness of 15 nm and a 145 nm Buried OXide (BOX) layer. This chapter presents successful integrations of $LaLuO_3$ and HfO_2 in high-T CMOS process flows and addresses the role of Si-rich and La-rich silicate formation on device performance.

The transistors were processed in a classical high-temperature integration scheme, often referred to Metal-Inserted-Poly-Si (MIPS) process. The MIPS integration scheme is historically originating from the replacement of poly-Si as gate electrode with a combination of metallic gate and poly-Si on top. This extension improves the electrostatic control over the channel region (electrostatic screening length) and avoids the EOT increase due to depleted poly-Si in the on-state of the transistor.[17] Immanent of the MIPS integration scheme is the patterning of the gate before Source/Drain (S/D) implantation into Si and subsequent activation.[70]

In general, the integration prior to S/D doping activation is the so-called gate-first (GF) integration, which does not necessarily imply the use of poly-Si within the gate stack. The gate-first integration is the most demanding process in terms of thermal stability, but offers also possibilities to utilize the thermal budget for a reduction of interfacial layer thickness.[17,70,71,85,98,101] Additionally, the study of a gate-first process complements our studies[33,188,189] of replacement gate (gate-last) processes, where dielectric and gate electrode are deposited after S/D doping activation.

The fabrication was manually performed; the process steps are shown in appendix A. The first devices were fabricated without NiSi contacts and with direct Al contact metallization to S/D area. Silicidation was found to be most effective to increase the on-currents of the devices and was therefore integrated. Moreover, the implantation was initially done through the high-κ oxide to improve straggling and solid phase epitaxial regrowth. However, the implantations led to very large variations in the on-currents of the devices. Most likely thickness variations and roughness of high-κ resulting from prior process steps caused different oxide thicknesses and therefore different implantation depths. The removal of high-κ prior to implantation avoids these difficulties. Moreover, the integration of $LaLuO_3$ with activation anneals leads to the described silicate formation. The silicates formed on the S/D area were found to be not removable by the standard wet etch recipe. Especially, contact formation on the active area was only possible with high bias during RIE etching, whereby the yield dropped due to imprecise stopping on the SOI layer when no *silicide* was in place. Hence, it is necessary, for a good process control (silicidation and contact formation on active area), to remove the Lanthanide oxides prior to activation.

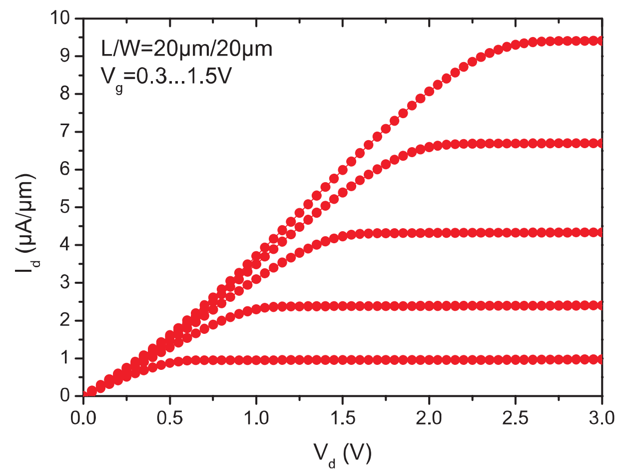


Fig. 4.14 $I_d - V_d$ output characteristics of the device with 3 nm $LaLuO_3$ and oxygen anneal. A well-behaved output characteristic is obtained. The output current shows a high S/D resistance R_{sd}, which was extracted to $14\,k\Omega$, mainly due to high sheet resistance of the only-implanted active region.

4.2.2 Device results

The devices were characterized by means of current-voltage (I-V) and (C-V) measurements. Mobility data were extracted by the split-CV method[190,191] and corrected for series resistance[192,193] and overlap capacitances.[194–196] Devices with implanted S/D area show acceptable long-channel sub-threshold slope (SS) of $\sim 70\,\mathrm{mV/dec}$ (cf. Figure 4.15, left). Important for the integration of the new material, the split-CV characteristics (cf. Figure 4.14, right) showed negligible hysteresis $(< 40\,\mathrm{mV})$ and normal accumulation for devices without poly-Si deposition. Moreover, the effective work function aligns nicely with the Si conduction band at $\mathrm{EWF} = 4.05\,\mathrm{eV}$. However, the EOT is large with 1.9 nm after full processing. The EOT increase can be accounted to the oxygen anneal. Moreover, the split-CV measurements for 3 nm $LaLuO_3$ confirm the results obtained on capacitor test structures (3, 6, and 10 nm $LaLuO_3$) prior to device integration.

The S/D resistance of the devices was extracted to amount to $14\,k\Omega$, which indicated a contact problem on the active area. This contact problem is not measured-with the series resistance of the gate contact during C-V measurements. Moreover, the devices exhibited initially a strong positive bias-tem perature instability (PBTI) component (60 mV/V) for the first sweep in the transfer characteristics $(I_d - V_g)$. This charge trapping in the oxide shifts V_{th} for all following curves (Shockley-Read-

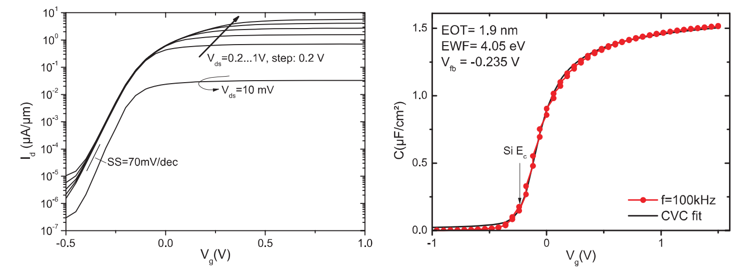


Fig. 4.15 (left) $I_d - V_g$ transfer characteristics of MOSFET with 15 nm TiN/3 nm $LaLuO_3$ gate stack and O2 anneal after high-κ deposition $(400°C/10\,min)$. Gate-Induced Drain Leakage (GIDL) is also observed below $-0.4\,V$. The device exhibits small series resistance but the S/D resistance R_{sd} was estimated to be $4\,k\Omega$. The subthreshold slope of the device is 70 mV/dec. (right) The C-V measurements obtained on the same device shows negligible hysteresis. However, the EOT with O_2 anneal amounts to 1.9 nm.

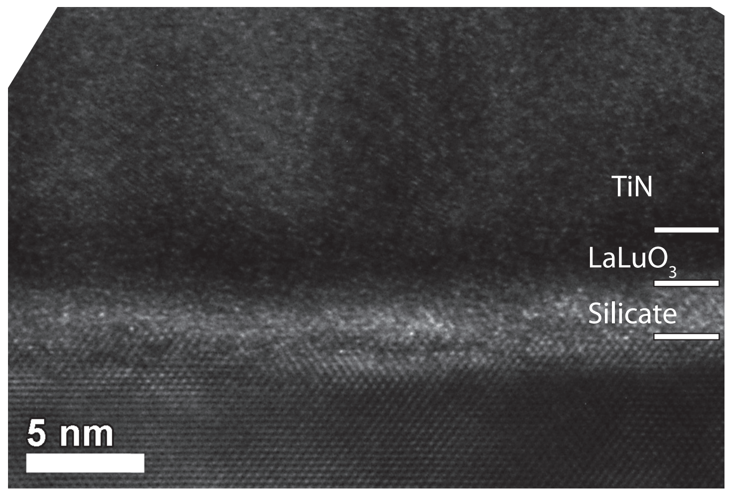


Fig. 4.16 TEM micrograph of the interface Si/3 nm LaLuO3/15 nm TiN after full device processing. The interface between $LaLuO_3$ (dark gray) and Si is continuous $(\sim 1\,nm)$ and indicates a silicate formation. No typical SiO_2 interface could be observed.

Hall trap generation[51]). Due to the absence of a sufficient number of devices for a robust statistical treatment, the PBTI component in $LaLuO_3$ will not be discussed any further. The output characteristic is well behaved and shown in Figure 4.14. The large S/D resistance R_{sd} decreases the drive current.

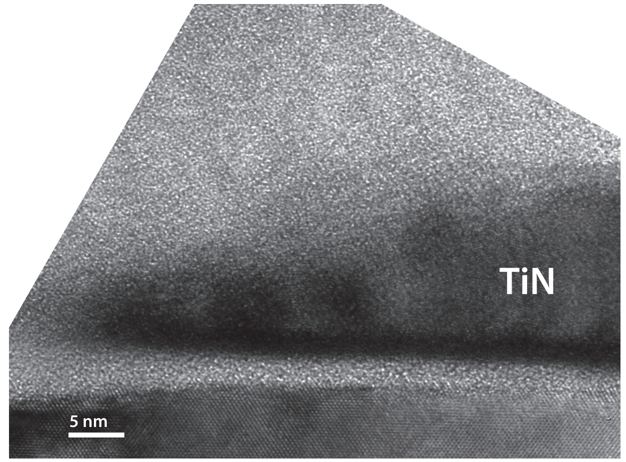


Fig. 4.17 At the side of the gate, the interface $Si/LaLuO_3$ (dark gray) is bending up ("bird's beak", left) due to oxidation from the side (bright) in this spacer-less process. No contrast difference between oxidized Si and the herein used passivation layer is observed.

A cross sectional TEM micrograph of one of the devices shows the intact $LaLuO_3$ layer after processing (Figure 4.16). Defocussing of the TEM could not show a bright SiO_2 interlayer between $LaLuO_3$ and Si substrate. As expected from the photoemission studies, the interface appears as a continuous gray gradient, whereby both layers are of equal thickness. The $LaLuO_3$ layer is visible in dark gray, silicate (assumed from the PES studies) is shown in light gray. The layer has increased in thickness as observed in the material studies, more precisely; $LaLuO_3$ has lost ca. 25% thickness, whereby the interfacial layer (1 nm after RCA cleaning) has grown to ~ 23 Å. Kakushima[128] reported an EOT of 1.2 nm for 3 nm La_2O_3 after 500°C anneal with unknown partial pressure of oxygen. Consequently, the current oxygen anneal must be further optimized to reach the performance of La_2O_3.

By comparing the TEM image of the side of the gate (Figure 4.17) with the inner part, a difference in interfacial layer was found, which becomes important for short-channel devices. On a width of 50 nm the gray interfacial layer becomes brighter and continuously thicker. However, the mentioned oxygen anneal was done prior to gate patterning. Hence, another source of oxygen must be accounted for this "bird's beak".[69] Prior to activation anneal the device was passivated with 30 nm Si_3N_4, visible in light gray. Since silicon nitride is reported[11,69] to be an effective oxygen barrier, most likely the deposition (at 350°C) itself incorporated the oxygen source, which lead to oxidation of the sides.

To increase the drive current of the devices NiSi contacts were introduced. The resulting output characteristics show much improved drain currents as shown in Fig-

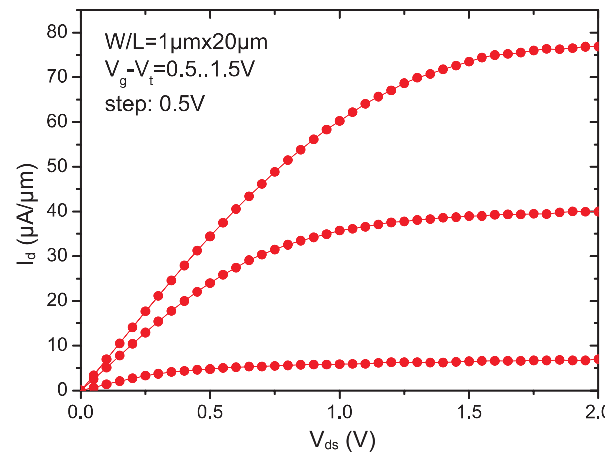


Fig. 4.18 MOSFET with NiSi contacts on active S/D area show much increased drive currents The gate stack comprises 3 nm $LaLuO_3/15$ nm TiN/100 nm poly-Si.

ure 4.18. Compared to the purely implanted devices, the drain current improves by a factor 10 at same gate length, drain, and gate voltage. Without oxygen anneal, the EOT reduces by 20% and an EOT $= 1.5$ nm is obtained for 3 nm $LaLuO_3$ (Figure 4.19) without any post deposition anneal. The deposition of 100 nm poly-Si by LPCVD (580°C, > 30 min) leads to large hysteresis (n.b. TiN was capped ex-situ by 5 nm a-Si prior to CVD deposition and HF was used to remove SiO_2 from the Si-cap). Without LPCVD deposition, no hysteresis was measured (e.g. Figure 4.15).

Comparing the various $LaLuO_3$ integration schemes with the first realization of a 2.5 nm HfO_2 high-κ layer with MIPS gate stack (split-CV in Figure 4.20), the hysteresis after full processing incl. LPCVD deposition is below 10 mV. Moreover, the EOT obtained on an RCA cleaned substrate matches some literature reports[70] for the integration of HfO_2 with uncapped TiN.

EOT and interfacial layer transformation are determining factors for the effective mobility in the device.[71, 197–199] Therefore the effective electron mobility was calculated from the electrical measurement data as[185, 191, 193, 194, 194, 196, 200–203]

$$\mu_{\text{eff}} = \frac{L}{W}\frac{I_{\text{DS}}}{V_{ds}\,Q_{\text{inv}}}. \tag{4.4}$$

Low $V_{\text{DS}} = 50$ mV was chosen in order to not distort the mobility data[200] Q_{inv} is the inversion charge integrating the C-V curve from accumulation to inversion. W and L is the gate width and gate length. All data were measured on the same device to obtain most accurate results.

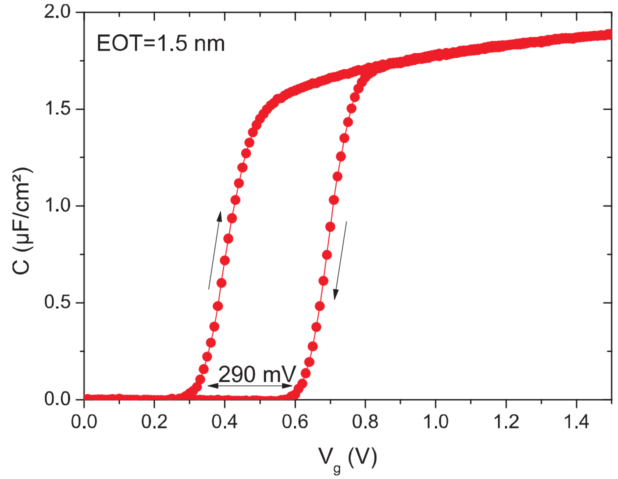


Fig. 4.19 Split-CV measurement on a MOSFET with 3 nm $LaLuO_3$ without PDA in O_2. The EOT reduces from 1.9 to 1.5 nm (at $V_{fb} + 1\,\mathrm{V}$).

An interesting comparison for mobility data is the (approximate) determination of additional mobility components μ_{add} via the Matthiessen's rule[71, 198, 199, 204] (after work of Augustus Matthiessen[205]) from the effective device mobility μ_{eff} and the reference mobility μ_{ref} to compare with:

$$\frac{1}{\mu_{\mathrm{add}}} = \frac{1}{\mu_{\mathrm{eff}}} - \frac{1}{\mu_{\mathrm{ref}}}. \tag{4.5}$$

Applying this rule, one has to keep in mind that the extraction is subject to error and increases with temperature. Stern's calculations[40] point out, that even at low-T (80 K) the error might exceed $\sim 35\%$. Nevertheless, this method gives a good approximation at room temperature for the comparison with known universalities (Takagi,[198, 199, 204, 206] Ragnarsson,[71] Tatsamura[197]), which were extracted with the same methodology.

Figure 4.21 (left) shows the effective electron mobility vs. inversion charge for the aforementioned device (gate-first) of this work in comparison to a gate-last device with 6 nm $LaLuO_3$ taken from Durgun Özben *et al.*[33] Both devices (this work and paper of coworker) received the same oxygen anneal at $400°\mathrm{C}/10$ min. The mobility for the GF device is around half of the GL mobility with the same gate oxide. For HfO_2 the mobility was previously shown to be lowered for lower EOT,[20, 71] which was directly explained with closer distance of high-κ to channel due to thinner interfacial (SiO_2) layer. In $LaLuO_3$, the situation is different: As found in the

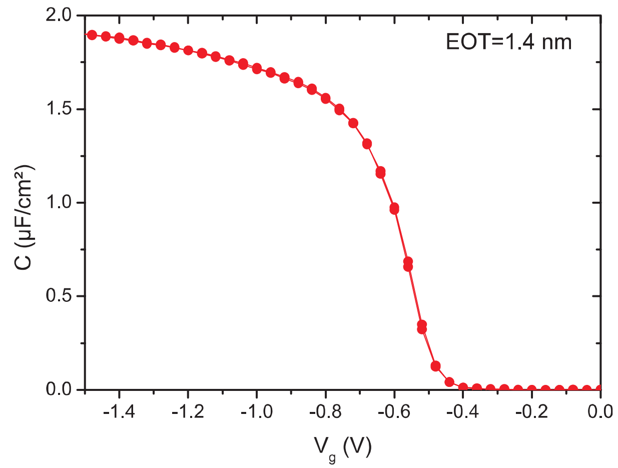


Fig. 4.20 Dual sweep split C-V measurement of 2.5 nm HfO_2/15 nm TiN/100 nm poly-Si gate stack in pMOS.

material analysis, the interface can consist of a Si-rich and La-rich interfacial silicate. For the high mobility of the gate-last device (oxygen anneal + FGA), we can assume the existence of a Si-rich silicate interlayer, based on the XPS spectra shown in Figure 2 in Durgun Özben *et al.*[33] This interpretation aligns also with the results of Kakushima,[23] who showed increased mobility with both Si- and La-rich silicate interlayer and decreased mobility for mainly La-rich silicate. However, the minimum EOT = 0.4 nm was only shown for La-rich silicate in combination with drastically reduced mobility. In this work, the high-temperature anneal was shown to produce mainly La-rich and less Si-rich silicate (for 10 nm oxide thickness), which was shown to reduce device mobility. The mobility reduction for the gate-first device is therefore in agreement with the results of Kakushima[23] and Durgun Özben[33] in our work group.

Figure 4.21 (right) shows the additional mobility component for the gate-first device in comparison to the gate-last reference. The universality for the inversion charge N_{inv} dependence is adapted in agreement to the work of Takagi[198,199] and additions for high-κ/metal gate by Tatsumura.[197] Two asymptotic regions depending on N_{inv} by a power-law can be identified. At low inversion charge, the mobility decays with $N_{inv}^{-0.15}$. The universality of -0.15 is interpreted as additional mobility reduction due to interaction of TiN and $LaLuO_3$ at the top high-κ interface in adaption of Tatsamura's interpretation.[197] The influence is weak and might be negligible because the device operates at high N_{inv} More important is the power law of $N_{inv}^{-0.15}$ Cartier[20] as soft-optical (SO) phonons of the high-κ layer. As Fischetti *et al.* pointed

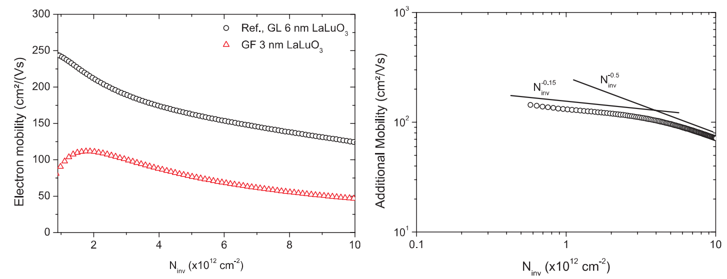


Fig. 4.21 (left) Effective electron mobility μ_{eff} for the aforementioned gate-first (GF) device with 3 nm $LaLuO_3$ (EOT $= 1.9$ nm, red curve) compared to the gate-last (GF) integration of 6 nm $LaLuO_3$ (EOT $= 1.6$ nm, data reproduced from Durgun Özben *et al.*) (right).

out; the impact of the SO phonons on the damping of the electron mobility depends exponentially on the thickness of the interfacial layer. For HfO_2 the interface consists of SiO_2, which explains that Tatsamura and Takagi found decay factors around -0.3.

In consequence, the gate-last (GL)-device has a Si-rich interfacial silicate, which dampens the impact of SO phonons on mobility. This damping is mostly absent in the gate-first device and carriers in the channel are directly screened by the high-κ optical phonons. The oxygen post deposition anneal was identical for GF and GL devices. The activation anneal for GF followed after gate patterning and implantations. Hence, the oxygen anneal does not improve the mobility if followed by the activation anneal. The isolated activation anneal in nitrogen itself was found by XPS and HAXPES to produce mainly La-rich silicate in the temperature range of 600–1000°C, 60 s and for $1000°\mathrm{C}/5\,\mathrm{s}$. Based on Kakushima's results the La-rich silicate leads to a lower EOT for cold egun deposition and subsequent low-T anneal (300–400°C). A comparison of the leakage current density for the GF MOSFETs and their replacement gate counterparts[33] leads to higher leakage for the La-rich Silicate IL in comparison to the Si-rich Silicate IL in the RPG MOSFETs. Hence, the La-rich silicate could offer a stronger reduction in EOT but leads to more gate leakage than the Si-rich IL. This can be explained with the more SiO_2-like IL for the latter. However, exactly this property also hampers the scaling of the Si-rich IL well below 1 nm EOT. In order to scale the EOT further with La-based oxides, the higher permittivity in La-rich silicates can be again a competitive alternative to the Si-rich IL.[177]

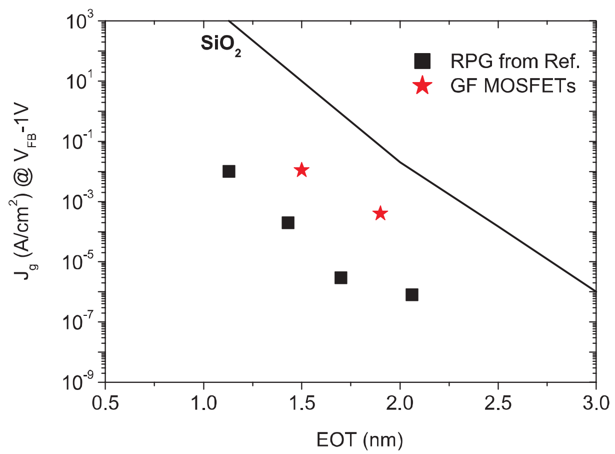


Fig. 4.22 Leakage current density for the GF $LaLuO_3$ MOSFETs vs. $LaLuO_3$ RPG MOSFETs (max. $T = 400°C$). The Si-rich RPG FETs show reduced leakage current in comparison to the GF FETs.

4.2.3 Summary

The gate-first approach leads to a lower mobility due to soft-optical phonons of the La-rich silicate screening the channel. This result was obtained by the analysis of the universality in the additional mobility for gate-first vs. gate-last. In conjunction with oxygen anneal after high-κ deposition, EOT increases and subsequent activation anneals lead to a more pronounced La-rich silicate than Si-rich silicate during oxygen anneal. The silicate formation led to *n*-type band-edge functionality with respect to the effective work function. The introduction of NiSi S/D contacts significantly improved the drive current. The combination of CVD deposited a-Si with $LaLuO_3$ leads to large hysteresis in the C-V characteristics as characterized by HAXPES. Therefore, an implementation of $LaLuO_3$ in a standard MIPS process might be difficult.

A possible alternative for ultra-low EOT devices could be the combination of both GL and GF approach: Prior to activation a thin $(< 1\,\mathrm{nm})$ $LaLuO_3$ can be deposited on a *HF-treated* substrate and subjected to activation temperature. La-silicate could then serve as an excellent stopping layer for the removal of the damascene gate. After removal, an optimized oxygen annealing can be carried out to form a Si-rich silicate for improved mobility due to the suppression of SO phonons. Finally, the $LaLuO_3$ layer must be deposited at low-temperature (e.g. ALD $< 300°C$) to prevent inter diffusion.

The first in-house realization of an HfO_2-based long-channel MOSFET with a MIPS integration scheme is promising for further process optimization in respect to EOT scaling.

Chapter 5
Oxide quality and reliability in high-κ/metal gate stacks

5.1 Investigation of oxide traps in high-κ/ metal gate stacks

In the last chapters, the materials for high-κ/metal gate stacks and their thermal stability were investigated. The successful integration of these new materials is not only connected to suitable low EOT and thermal stability. In the device integration chapter we have already learned, that in the case of $LaLuO_3$ different interfacial layers (La-rich or Si-rich) can lead to huge differences in effective electron mobility. In this chapter, the focus is now placed on the investigation of oxide defects in HfO_2 and their role on device performance using the example of gate integration with FinFETs.

Multi-gate architecture devices like FinFETs are viable successors for CMOS integration beyond the 28 nm technology node.[207,208] The FinFETs exhibit a better short-channel immunity compared to their planar counterparts at comparable performance and equal footprint.[209–214] A crucial step for the integration, already for the technology above the 32 nm node, is the integration of the high-κ/metal gate (HKMG) stack. Two competing approaches are presently investigated: The high-temperature source/drain (S/D) activation with HKMG in place (aka gate-first, GF) and a process replacing a damascene gate after annealing (aka gate-last, GL). The gate-last approach offers a better control of the threshold voltage (V_{th}) distribution for both n- and p-MOSFETs.[213,214] The GL approach can be further subdivided into a high-κ first (HKF) and high-κ last (HKL) process, where the high-κ film either undergoes the full thermal budget (HKF) or is deposited afterwards (HKL).

To illustrate the influence of the gate stack on device performance, a mature gate-first process is compared with two non-optimized gate last approaches (HKF/HKL) for Si-bulk FinFETs. The impact of interface quality and number of defect states N_{it} on the overall FinFET performance is examined.

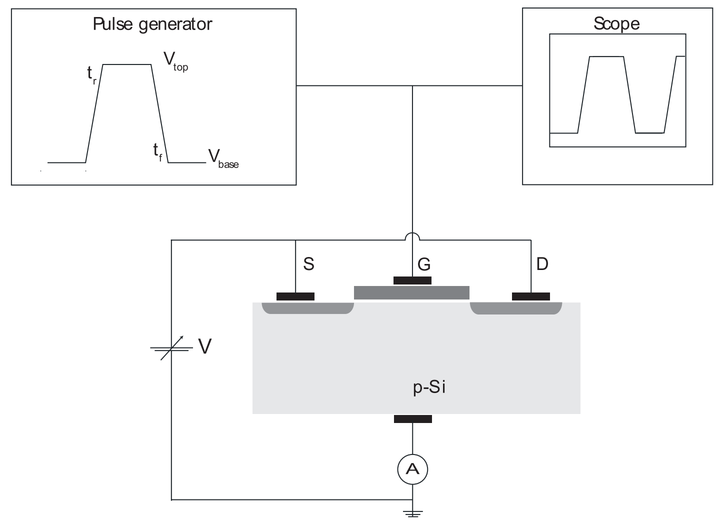


Fig. 5.1 Measurement setup for charge-pumping of a MOSFET. S/D terminals are grounded and the current response on a pulse train applied to the gate is measured at the bulk terminal. A scope is used to monitor the applied pulse shape.

5.1.1 Introduction to charge-pumping

Several techniques have been proposed to measure the interface quality, aka the density of interface D_{it} and oxide states D_{ot}. An indirect method to the interface characterization is the aforementioned extraction of D_{it} from a conductance measurement,[35,50,215] which can be used for both MOS capacitors and MOSFETs. A direct measurement of interface traps on MOSFETs is possible by the so-called Charge-Pumping (CP).[216,217] Thereby, a pulse train is applied to the gate of the device, S/D terminals are grounded, and a body current is measured on the bulk terminal. The pulse train is characterized by its amplitude V_{ampl}, the periodicity and rise/fall time of the single pulse, called t_{r} and t_{f}, respectively. A schematic measurement setup is shown in Figure 5.1. Notably, the bulk terminal can be replaced with P-I-N junctions[218] between S/D. Therefore we employed CP also to characterize Tunnel-FETs[56] on SOI. A second possibility to CP on SOI is the use of a T-gate,[219] where a prolonged gate finger is used to provide an ohmic contact to the SOI.

The charge-pumping current registered at the bulk terminal is caused by trapping of electrons in the oxide and at its interface to the channel. The present traps are filled with electrons from S/D during the rise time of the pulse sweeping the band diagram from inversion to accumulation (Figure 5.2 (1.-2.), graph adapted

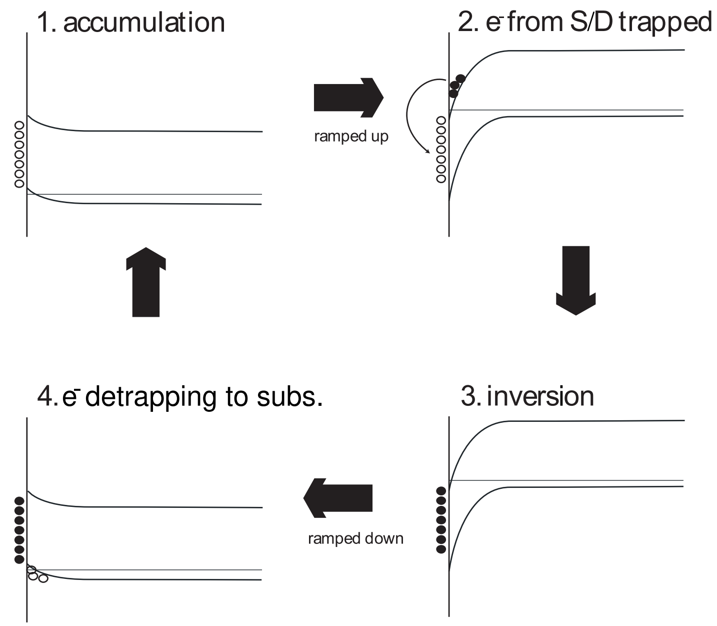


Fig. 5.2 First-order model of charge pumping ignoring the characteristic capture-emission time of the SRH statistics. Electrons from S/D are trapped in defect states and filled during sweep from accumulation to inversion. During pulse ramp-down the trapped electrons recombine with holes from substrate and lead to hole current, which is measured.

from Groeseneken[220]). During inversion, the traps are occupied by electrons (Figure 5.2 (3.-4.)). The capture/emission time of the traps follows the Shockley-Read-Hall (SRH) statistics[51] and can be characterized by variation of the frequency, i.e. charge and discharge time t_{charge} and $t_{\text{discharge}}$.[217] Moreover, by a variation of the base-level potential V_{base} of the pulse different defect levels can be profiled as explained by Toledano *et. al.*[217]

The subsequent ramp-down with time t_f leads to recombination of the trapped electrons with holes from bulk resulting in a hole current from the substrate (Figure 5.2, (4.-1.)). This hole current is measured as shown in Figure 5.1 and is a measure for the charge trapping i.e. the defect densities D_{it} and D_{ot}:

$$I_{\text{cp}} = f\,Q_{\text{cp}} = q\,f\,A\int D_{\text{t}}(E)\,\mathrm{d}E = q\,f\,A\,D_{\text{t}}\,\Delta E, \tag{5.1}$$

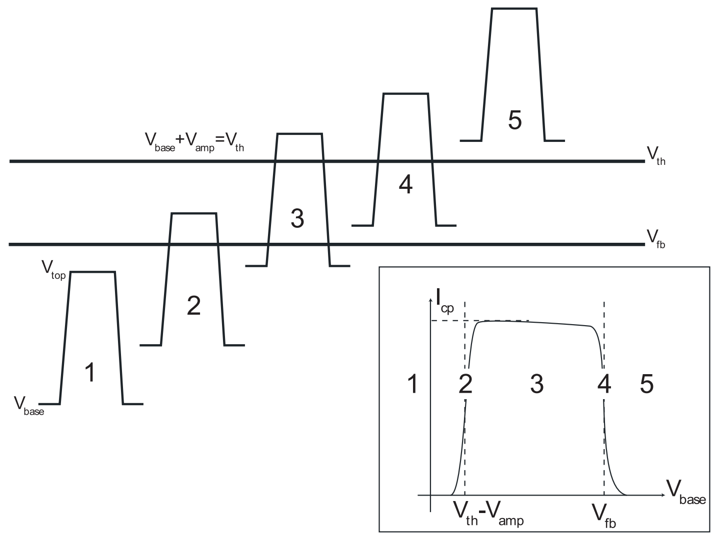


Fig. 5.3 Working principle of base-level charge pumping CP_{base} (cf. text).Inset: resulting CP current.

f denotes the measurement frequency, *q* is the electron charge, A the gate area and D_t the density of interface and oxide traps (per energy level or averaged, respectively).[216] As Toledano[217] pointed out, the CP current in high-κ gate stacks is a combination of D_{it} and D_{ot} of the bulk (high-κ) oxide. The individual trap levels can then be assessed by a variation of the discharge time (spectroscopic CP). A change in discharging time enables moreover the spatial profiling between $Si/SiO_2/HfO_2$ interfaces if the prior charging time was long enough to charge defects in HfO_2. The spatial profiling is important for the current experiment, since the measurements on different EOT gate stacks lead to a partial charging of high-κ defects and the emission time depends exponentially on the trap distance from the interface (shown in Figure 5.4). Cho *et al.* have shown the use of on-chip ring oscillators for CP to provide very high measurement frequencies and thus enabling a separation between high-κ and SiO_2 interface defects.[221]

A common measurement routine of charge pumping is the application of a pulse train with constant amplitude V_{ampl} and variation of the base level voltage V_{base}, called base-level charge pumping (CP_{base}). The working principle is shown in Figure 5.3 (after Toledano[217]):

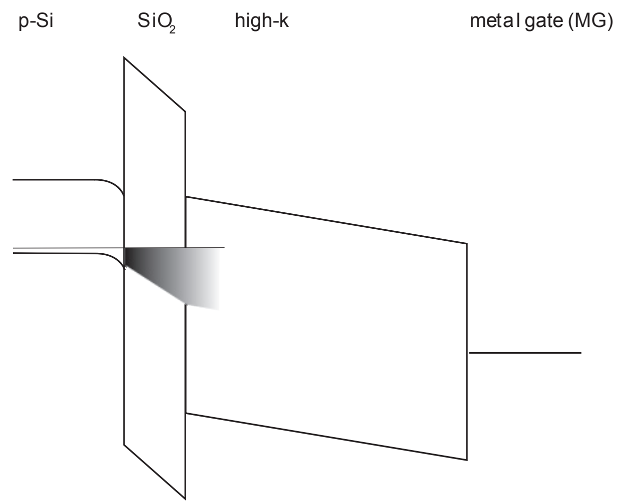


Fig. 5.4 Band diagram in inversion with traps in high-κ and SiO_2. The emission time of the traps depends exponentially on the distance from the interface. Thus, partial trapping/detrapping of electrons in high-κ defects is observed depending on the capture and emission time during CP.

Starting well below flatband voltage V_{fb} no traps in the band gap can be screened (1) and no CP-current is present. Above V_{fb} traps within the band gap can be filled and the recombination leads to a current (2) and inset. Reaching the complete band gap (3) with the pulse, all traps can be filled. Leaving partially the CP-regime (4, $V_{fb} < V_{base} < V_{th}$) less current is generated. Above threshold (5) no CP current will flow.

A common problem for ultrathin EOT devices $(\text{EOT} \sim 1\,\text{nm})$ is the intrinsic gate leakage, which adds on the CP current. The leakage current can then be compensated by associated low frequency CP measurements $(\sim 1\,\text{kHz})$, which is dominated by gate leakage ($I_{cp} \sim f$, equation (5.1). The higher frequency measurement can be corrected by this low frequency sweep (subtraction).

5.1.2 Device fabrication and data extraction

The GF devices were fabricated in the imec pilot line using a standard metal-inserted polysilicon (MIPS) process by the patterning of the full HKMG stack and subsequent 1000°C S/D activation anneal.[70] For the case of the HKF, the high-κ was deposited before S/D activation, covered by an amorphous Si (a-Si) damascene gate.

Table 5.1 Extracted device parameters.

Process	n/p	EOT [Å]	J_g [A/cm^2]
HKL	N	10	$1\cdot10^{-3}$
	P	10	$1\cdot10^{-3}$
HKF	N	13	$5\cdot10^{-4}$
	P	13	$1\cdot10^{-3}$
GF	N	13	$2\cdot10^{-3}$
	P	13	$5\cdot10^{-4}$

After activation, the dummy gate was replaced by 5 nm ALD TiN and 100 nm a-Si. Interface passivation was achieved by an imec clean[151] for the case of GF and HKF. For the HKL process, a classical SiO_2/poly-Si gate stack was used and removed after S/D activation by HF and TMAH. Subsequently, an ozone oxidation step and 1.8 nm HfO_2ALD deposition followed. All wafers were then processed up to the first metallization layer (M1).

The equivalent oxide thickness (EOT), extracted by split-CV measurements at 100 kHz, and the gate leakage current density (J_g) are shown in Table 5.1 for $V_g = 1$ V. Base-level charge pumping (CP_{base}) was used to extract the D_{it} levels from 1 μm long FinFETs on bulk-Silicon (100). Different fin widths were used to extrapolate the top- and sidewall N_{it}, as described by Kapila *et. al.*[222] V_{th} was extracted using method.[223] All parameters were screened at different fin-width to evaluate the influence of parasitics at increasing gate control.

5.1.3 Impact of interface states on mobility measured by charge pumping

The impact of interface states on V_{th}, subthreshold slope (SS) and effective carrier mobility is shown for long-channel (LC) FinFETs with different fin width (W_{fin}). The threshold voltage plotted as a function of the fin width, presented in Figure 5.5, shows for the three processes a clear roll-off behavior with decreasing *fin width*. This effect is ascribed to an increased electrostatic control. The same thermal budget was used for the HKF and HKL processes after TiN deposition. In contrast, the GL process, with the same TiN deposition properties and thickness, shows strong V_{th}-shifts for *both n-* and pMOS devices. The gate-first 16 nm fin width and nMOS FinFETs show 180 mV and 300 mV V_{th} shifts in comparison to HKF and HKL devices, respectively. The V_{th}-shift of 120 mV between the HKF and HKL processes at narrow fin width cannot be accounted to work function changes in TiN metal gate due to the absence of an annealing step. More likely, it is a change of the effective work function, i.e. a potential change induced in the channel due to additional charges within the gate stack.

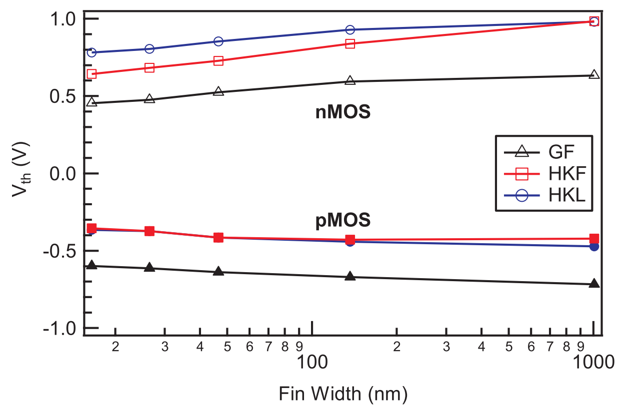


Fig. 5.5 V_{th} vs. fin width for different processes measured for long-channel FinFETs ($Lg = 1\,\mu$m) For GF nMOS and GF TiN shows near *n*-type band-edge behavior, whereas the effective work function decreases for the low temperature processes (180 mV/300 mV shift for HKF/HKL at 16 nm fin width). This trend is more pronounced for nMOS than for pMOS.

For the pMOS both gate-last processes show almost identical V_{th} at narrow fins width and slight variation for wider fins while the GF devices show nearly symmetrical behavior to nMOS. It may be then argued that the V_{th}-shift difference between nMOS and pMOS at narrow fin width relates to the relative position of the charged defects levels within the band gap rendering pMOS less affected.

The subthreshold slope is less sensitive to GF and HKF processing, where high-κ on imec clean interface passivation was used, showing only a max 5% difference between the two (Figure 5.6). Most of the long-channel nMOS devices show subthreshold slopes of 70–75 mV/dec with weak dependence on fin width, below 3%. The HKL *n*/pMOS, however, lies with about 15% above these values presenting SS of 80–90 mV/dec. For all wafers, the slope reaches a maximum at 46 nm fin width and decreases again with increased channel control for narrow fins. For pMOS the slope increases $(+18\%)$ with increasing sidewall to topwall ratio, whereby the sidewall transport itself is beneficial for pMOS. It decreases again for increased electrostatics at 46 nm fin width.

The effective carrier mobility is considered as an overall measure of the device performance. Figure 5.7 shows the extracted effective mobility for $L_g = 1000\,\text{nm}$ and $W_{\text{fin}} = 16\,\text{nm}$ for both *n*- and pMOS and all splits. High field mobility was extracted at $N_{\text{inv}} = 10^{13}\,\text{cm}^{-2}$, peak mobility was extracted at maximum mobility. The reduction in electron mobility for the HKL split compared to GF was peak-to-peak -47% and in high-field -36%. In the HKL split most likely remote-Coulomb scattering (RCS) reduces the onset to peak mobility.[197] The hole mobility is less but also significantly reduced for HKL compared to HKF (-31%). Comparing GF and

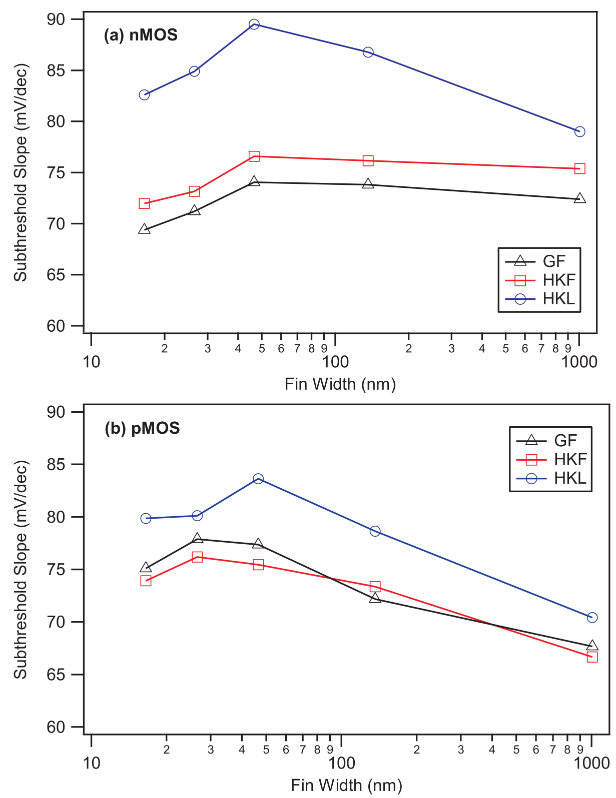


Fig. 5.6 Subthreshold slope (SS) vs. fin width measured for long-channel FinFETs ($L_g = 1\,\mu$m): (a) nMOS/(b) pMOS. The SS distribution for the GF and HKF process are comparable whereby nMOS FinFETs show slightly higher SS with increasing ratio of sidewall to topwall (67–78 mV/dec.). For HKL the SS shows a larger spread and does not scale with fin width.

HKF splits, the mobility reduction is less pronounced as evaluated for the SS and V_{th}: The electron mobility is only slightly reduced, about 10%, and the hole mobility is comparable.

To assess the performance reduction from the processing side, base-level charge pumping was employed. Additionally, the measurement of FinFETs with varying fin-width allows the determination of top- and sidewall N_{it} values. For low-EOT devices a leakage current correction was performed. Unfortunately, this correction was not possible for the nMOS in the HKF split since excess leakage was also present at

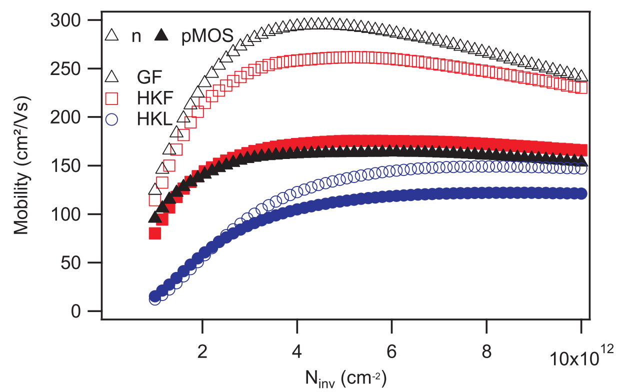


Fig. 5.7 Effective mobility, μ_{eff}, for *n*/pMOS FinFETs with $L_g = 1000\,\mathrm{nm}$ and $W_{\mathrm{fin}} = 16\,\mathrm{nm}$. The gate-first process flow shows the overall highest electron mobility. In summary, the results for electron and hole mobility on the gate-first and HKF wafer are comparable. Remarkable is the much lower HKL carrier mobility: High field mobility drops by -36% and peak mobility by -47%. Comparing HKL and HKF split most likely more RCS is present in HKL due to lower EOT.[13]

measurement frequency of 2 MHz. Therefore, the analyses was limited to pMOS charge pumping and verified, where possible, with nMOS data for consistency. An example curve is shown in Figure 5.8 (left): The hat-shape of the nMOS CP-curve is typical and relates to the trap levels in the band-gap.[217] The device shown was taken from the HKL split with five fins, $L_g = 1000\,\mathrm{nm}$ and $W_{\mathrm{fin}} = 22\,\mathrm{nm}$.[13]

The measured charge-pumping current versus fin-width is plotted in Figure 5.8 (right). The current extracted for all three splits and nine different fin widths (16 nm–1006 nm) gives an accurate linear interpolation for the determination of the *y*-intercept. This *y*-intercept represent the number of defect states at the fin sidewalls, $N_{\mathrm{it,\ sidewall}}$, while the fit slope represents the number of defect states $N_{\mathrm{it,\ top}}$ at the top of the fin.

The N_{it} levels extracted from Figure 5.8 (right) are summarized in Table 5.1.3. The lowest top ($\{100\}$ surface) and sidewalls ($\{110\}$ surfaces) N_{it} levels were found for the gate-first integration. The values for the HKF integration are 7 and 3 times higher for the top and sidewalls, respectively. The clear HKL process differences in *y*-offset and slope (Figure 4) prove one order of magnitude increase of the $N_{\mathrm{it,\ sidewall}}$, and $N_{\mathrm{it,\ top}}$ compared to the lowest levels in the MIPS process. Since EOT and the thermal budget for interfacial layer and high-κ layer are comparable between GF and HKF, an explanation is not obvious. The differences between the two gate-last processes are too large to be explained only by geometric variability due to the different integration schemes. Altering of the high-κ quality (number of oxide traps, N_{ot}) during removal of the damascene gate and the deposition of the new TiN layer might explain this increase.

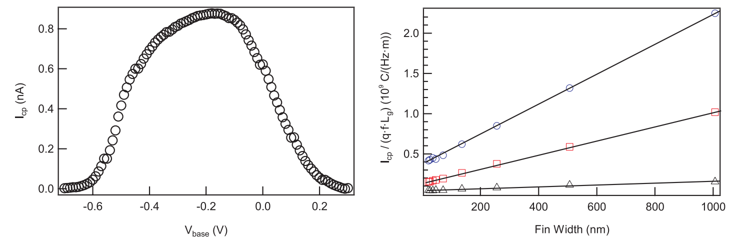


Fig. 5.8 (left) Typical curve of base-level charge-pumping in HKL nMOS FinFETs. Typical is the hat-like charge-pumping curve.(right) Extracted normalized charge-pumping current for the three splits using pMOS devices. The slope of the interpolation is proportional to the Dit level on the top surface, the sidewall Dit is extracted by the *y*-intersept.

Table 5.2 Extracted number of oxide defect per process.

Process	n/p	$N_{it,\,\{100\}}$ [cm^{-2}]	$N_{it,\,\{110\}}$ [cm^{-2}]
GF	P	$1.2 \cdot 10^{10}$	$5.8 \cdot 10^{10}$
	N	$1.6 \cdot 10^{10}$	$7.6 \cdot 10^{10}$
HKF	P	$8.7 \cdot 10^{10}$	$2.1 \cdot 10^{11}$
	N	–	–
HKL	P	$1.9 \cdot 10^{11}$	$5.0 \cdot 10^{11}$
	N	$2.7 \cdot 10^{11}$	$6.6 \cdot 10^{11}$

The above results demonstrate the highest passivation efficiency for the optimized gate-first process even if a high thermal budget step follows.

To visualize the impact of N_{it} level on device performance, the effective peak mobility is plotted against the fin width in Figure 5.9. As a general trend, increased electron and hole mobility is found for narrow fins. In general, thinner fins present increased electrostatic control.[224] The performance increase at narrow fin width can be explained by higher tensile strain induced by SiN spacers in scaled nMOS devices for $\{110\}$ direction on both (110) and (100) crystal planes. For pMOS, the stress is neither beneficial nor detrimental. The effective hole mobility benefits, however, from the contribution of the sidewall ($\{110\}$ surfaces) when the fins are fabricated along $\{110\}$ direction.[44,225] It is well known that the $\{110\}$ surfaces possess the highest hole mobility regarding all Si crystal planes.

Figure 5.9 shows that the carrier mobility of GF devices is the highest compared to HKF and HKL. Neglecting R_{s} influence on the mobility extraction, two effects can be identified: First, HKL shows, for lower EOT $= 1.0$ nm, a stronger mobility degradation[20] compared to HKF and GF, both with EOT $= 1.3$ nm. The mobility reduction due to EOT reduction is superimposed to HKL as compared to GF. As smaller EOT means thinner interfacial layer for HfO_2, the decreased distance of channel to high-κ oxide can also lead to increased charge pumping current due to

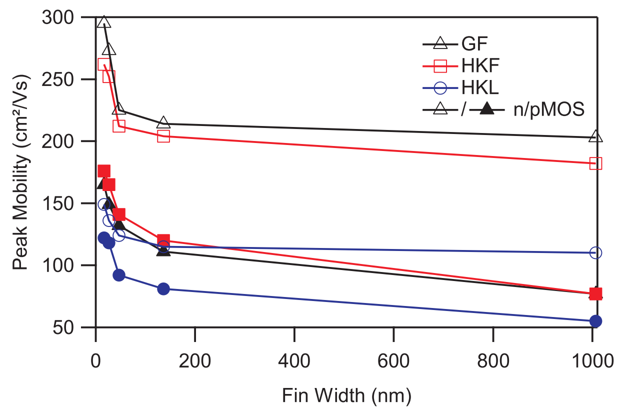


Fig. 5.9 Peak mobility vs. fin width at $L_g = 1000\,\mathrm{nm}$ for all three process flows. A general mobility improvement is observed for narrow fin. However, the improvement between HKF and GF (same EOT $= 1.3\,\mathrm{nm}$) is less pronounced than for HKL with lower EOT (1.0 nm).

defect screening in the high-κ oxide[217] (spatial profiling). Nevertheless, the overall larger field at lower EOT should reduce the mobility by roughly 23%[71] The differences found here exceed the ones given by Ragnarsson[71] for the peak and high-field mobility. The additional mobility reduction can therefore be explained by the higher number of defect states for HKL.

The origin of the mobility reduction is the interface scattering or high-κ inherent scattering: remote Coulomb scattering and remote phonon scattering, RCS and RPS, respectively.[20,197] Comparing the measured mobility values for GF with the mobility reduction due to EOT scaling by Ragnarsson,[71] one finds a much higher reduction for HKL as expected from RCS for the peak mobility and from RPS at high field ($200\,\mathrm{cm}^2/\mathrm{Vs}$ by Ragnarsson[71] at EOT $= 13\,\text{Å}$ compared to $153\,\mathrm{cm}^2/\mathrm{Vs}$ at EOT $= 10\,\text{Å}$). Nevertheless, the FinFET mobility for GF is higher as compared.[71]

5.1.4 Origin of mobility reduction

The impact of fin width on the experimentally observed decrease of V_{th} is related to the increased electrostatic control, where the ratio W/H is the crucial parameter rather than the gate length.[208] The natural screening length or also named screening parameter for FinFETs or tri-gate architecture[224] is given by

$$\lambda = \sqrt{\frac{\varepsilon_{\mathrm{Si}}}{3\varepsilon_{ox}} t_{ox} t_{\mathrm{Si}}} \tag{5.2}$$

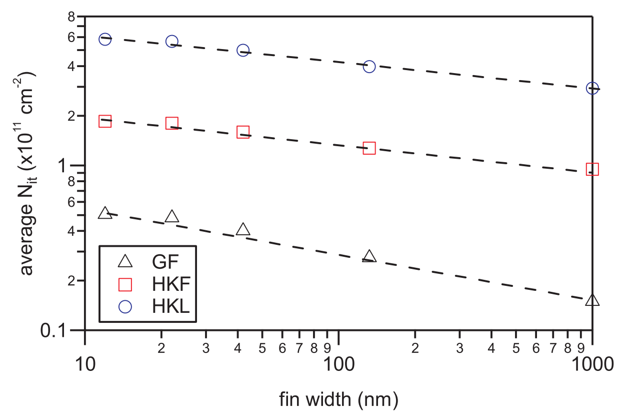


Fig. 5.10 Average N_{it} vs. fin width. The N_{it} value increases with the weight of the sidewall. The average was obtained by taking the top- and sidewall Nit levels at their W_{fin}:$2H_{\mathrm{fin}}$ ratio.

with a Si film thickness t_{Si}, i.e. fin width w_{Si}, and oxide film thickness t_{ox}. Its decay for narrow fins indicates improved overall channel control. The predicted metal grains orientation induced work function variation at nanoscale[79,80] may be also considered to explain the change of V_{th} with fin width due to the increasing fraction of sidewall area. However, this effect has not yet been experimentally proven as literature might suggest. Shao *et al.*[226] calculate a drop in threshold voltage increasing quantum confinement in the wires or fins. However, the classical theory predicts a V_{th} increase.[224]

The discussion regarding the effect of the fin width on subthreshold slope can be started from the definition of (average) *SS*

$$SS = 2.3\,\frac{k_B T}{q}\left(1 + \frac{C_{\mathrm{depl}} + C_{\mathrm{it}}}{C_{\mathrm{ox}}}\right) \tag{5.3}$$

q denotes the electron charge, k the Boltzmann constant and T the temperature.

The decrease of the fin width induces an increase of the average N_{it} due to the increasing weight of the sidewall (Figure 5.10). Thereby, this parasitic capacitance increases having a detrimental impact on SS. However, for the narrower fins the subthreshold slope decreases again, as visible in Figure 5.6. Here, the main role belongs to the sidewall quantum capacitance. Due to its heavier effective mass,[41,225] the depletion capacitance on (110) surface is smaller compared to (100) quantum capacitance. C_{depl} is thereby decreasing. In the case of pMOS the SS increase is less pronounced, whereby the screening of the holes by interface and oxide traps is less effective (heavier eff. mass) and the overall C_{it} taken into account has to be decreased.

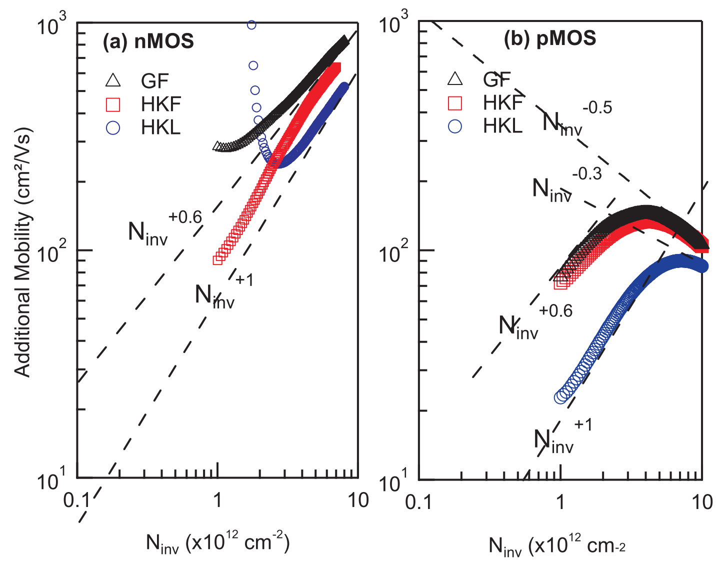


Fig. 5.11 Additional mobility gain when going from 1 μm fin width to 16 nm width. Electron/hole mobility behaves differently: In nMOS, mostly charges lead to (remote) Coulomb scattering. Hole mobility is reduced by charges and phonon scattering.

To assess the mobility decrease for HKL and HKF indicated in Figure 5.7 and Figure 5.9, Matthiessen's rule is applied to compare the additional mobility components when shrinking the fin width from 1 μm (μ_{LW}) to 16 nm (μ_{ref}):

$$\frac{1}{\mu_{\mathrm{add}}} = \frac{1}{\mu_{\mathrm{LW}}} - \frac{1}{\mu_{\mathrm{ref}}}. \tag{5.4}$$

For nMOS in Figure 5.11a, mostly sidewall charges lead to increased Coulomb scattering. The scattering process depends on the EOT value. It is most detrimental for the HKL integration route, where the highest N_{it} levels were measured. For the GF integration an universality factor in agreement with the results of Tatsumura *et al.*[197] is extracted as $N_{\mathrm{inv}}^{+0.6}$. They explained this universality by oxide defects close to the interface HK/IL. In comparison, this means that narrower fins would have led to less passivated sidewalls, higher D_{it} and higher sidewall oxide traps, D_{ot}. The aforementioned charge-pumping results support this assumption and additionally emphasize the importance of the interfacial layer quality for the high-κ growth.

For the hole mobility (Figure 5.11b) a clear influence of Coulomb scattering is observed in the HKL devices. For the HKF and GF splits the higher quality passivation is confirmed by CP and the same universality of $+0.6$ is found at medium

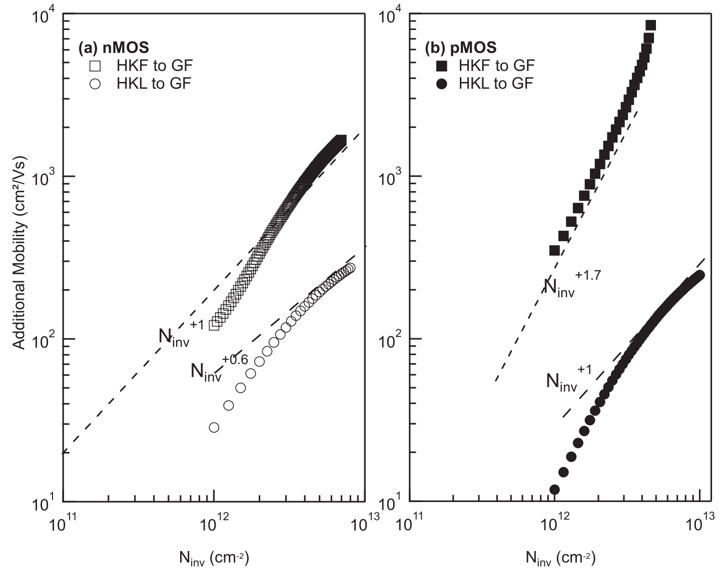


Fig. 5.12 Additional mobility components for the GF split in comparison to HKF and HKL for planar-like (a) nMOS and (b) pMOS with $L_g = 1000\,\mathrm{nm}/\mathrm{FinW.} = 1000\,\mathrm{nm}$. (a) The additional component for HKF-GF is small and Coulomb interaction-based. The mobility decrease for HKL is mostly mediated by the additional charges present and the overall lower EOT. (b) for pMOS following the HKF route, a steeper slope at midrange Ninv was found.

inversion levels N_{inv} i.e. the mobility is slightly reduced by the imperfect sidewall passivation. The high field hole mobility in HKL is reduced by phonon scattering in the case of narrow fin having a $N_{inv}^{-0.3}$ dependence[198,199] For GF and HKF integration routes, another mobility component with universality of -0.5 was found. This is interpreted as an increased scattering due to surface roughness on the $\{110\}$ sidewall beginning with decreased fin width. The deviation to Takagi's results can be attributed to N_{inv} splitted into top- and sidewall fractions.

The additional mobility components for planar-like test structures (fin width 1000 nm/gate length 1000 nm) are shown in Figure 5.12 *in reference to the gate-first mobility*. *It allows the judgment of mobility affecting mechanisms strictly related to the top (100) interface.* For nMOS devices where Coulomb scattering is the main mobility degradation mechanism, the HKL split shows lower mobility due to its lower EOT. However, the high-field mobility (quantum limit) shows an universality as described by Tatsumura.[197]

The pMOS mobility dependence deviates from the Coulomb type behavior, for the HKF split, having a universality $N_{\mathrm{inv}}^{+1.7}$, possibly due to dipole-like interactions. However, additional efforts are needed to clarify this effect.

5.1.5 Summary

N_{it} levels and mobility data for three different integration schemes for the FinFET architecture were shown. V_{th} differences with and without high-T anneals up to 300 mV were demonstrated for the same deposition method and thickness of TiN. Most likely oxygen and nitrogen diffusion within the gate stack cause the potential variations as discussed for the TiN characterization. Increased sub-threshold slope for high N_{it} levels were found in the non-optimized HKL integration. The N_{it} levels were assessed using different fin-widths to extract top- and sidewall N_{it}. N_{it} levels up to one order of magnitude difference were found between gate-first and high-κ last integration of the gate stack. Additionally, sidewall N_{it} levels were up to one order of magnitude higher than the topwall values. The Coulomb scattering increases with decreased EOT as also found in earlier experiments. The electron mobility for both planar-like FinFETs and narrow-width FinFETs is governed by Coulomb scattering. The scattering is higher for high-κ last than for high-κ first integration. Coulomb scattering affects also the hole mobility. For high inversion charge the hole mobility decreases due to surface roughness scattering at the sidewalls of the fins.

From an integration point of view, the high-κ first scheme offers nearly identical mobility values and interface quality as a mature MIPS integration. The results show clearly, that the mobility of as-deposited (HKL) HfO_2 is not comparable to the optimized gate-first results with anneals. For defect reduction, additional annealing steps are beneficial as shown for HKF. As Guha and Narayanan[17] pointed out, the regrowth of interfacial layer can improve the mobility at high-T annealing. A combination with oxygen control to control the resulting EOT remains necessary. The accuracy of the extracted universality factors can further be increased by the use of temperature-dependent mobility measurements.

5.2 TDDB reliability of advanced FinFET devices

The successful integration of a new high-κ/metal gate stack or new device structure like a FinFET should also consider its lifetime under operating conditions. In the last decades, the reliability physics developed a set of tools to assess different phenomena of device degradation over lifetime (Front-End-Of-Line (FEOL) reliability). Three key mechanisms can be identified:

- Bias-Temperature-Instability (BTI) leading to V_{th}-shifts over lifetime and thus shifting the operating point of p-/nMOS or both (lowered drive current). Accordingly, positive BTI (PBTI) affects nMOS and negative BTI (NBTI) affects pMOS. The measurements are carried out on many devices at specified operating temperature (e.g. 125°C) with S/D and bulk terminal grounded. Thereby different sets of stress voltages are applied. The V_{th}-shifts under voltage acceleration follow a power law and can be used to estimate the V_{th}-shift over lifetime. The reason for BTI is de-passivation of oxide defects.[35,227]
- Hot Carrier Injection (HCI) introducing also defects at the Si/SiO_2 or high-κ interface due to high energetic carriers entering through the channel the space-charge region at the drain. There the carriers can cause defects, charge trapping, or leakage current increase (bulk or gate leakage) depending on their energy. The stress is carried out with applied gate and drain potential and after stress the MOSFET characteristics are compared.[35]
- Gate Oxide Integrity (GOI) characterizes the reliability of the gate oxide under applied gate voltage. Typical tests for thin gate oxides are ramped voltage (V-ramp) tests and Time-Dependent Dielectrics Breakdown (TDDB), here under constant voltage stress (CVS).[35,228] The first method applies a constant voltage ramp and traces the gate current until the breakdown condition is fulfilled. The latter applies a constant voltage until the breakdown condition is reached. During gate voltage stress, all other terminals are grounded. The reliability over lifetime is typically extrapolated from the acceleration parameters (voltage and temperature).

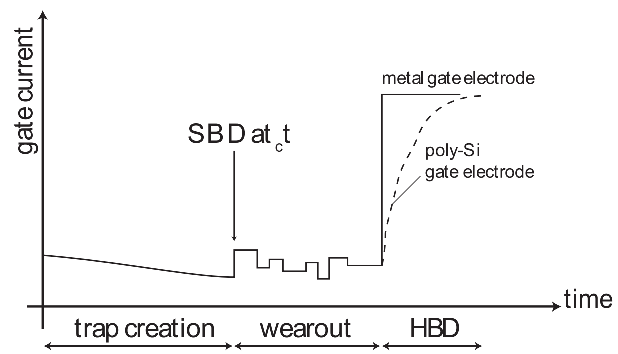


Fig. 5.13 The evolution of gate current with time under CVS features a first SBD with first current increase and the begin of a digital wearout phase. The final HBD is marked by an instantaneous current increase for metal gate electrodes and gradually increasing leakage current for poly-Si electrodes ("analog phase").

In the following section, the TDDB lifetime extraction by application of constant voltage stress on HfO_2/TiN gate stacks will be addressed.

The individual MOSFET under constant voltage stress shows an initial, small current decrease due to the charging of trap levels. The accumulated charges lead to the generation of a first conducting leakage path through the oxide, the soft-breakdown (SBD) at creation time t_c. Due to the current transport through this weak spot, wearout of this spot is observed, which was found to lead to material migration inside the gate stack.[49] The continued wearout with material transport leads to a digital variation of the gate current prior to final hard-breakdown (HBD). The HBD is instantaneous for metal gates, but with gradually increasing current for poly-Si electrodes. Figure 5.13 displays the aforementioned sequence of degradation under CVS. The current increase for the first SBD and final HBD can then be found by analysis of the recorded curves and leads to two statistics of the time-to-breakdown (t_{BD}) for SBD and HBD.

Once t_{BD} is extracted for the complete experimental set, the obtained lifetime population is built using the Bernard or Kaplan-Meier estimator. The so-obtained lifetime population can be fitted by the Weibull distribution.[229] The extracted 63% values of TDDB lifetime are subsequently extrapolated to operating conditions and lower percentiles assuming voltage acceleration and Poissonian area scaling.

However, this area scaling can be challenging for short-channel devices with low EOT due to process variability and an often non constant thickness of the interfacial layer between high-κ and Si substrate ("bird's beak".[69]). This effect can be even more pronounced in 3d MOSFETs.[72] Moreover, for low EOT nMOS the gate leakage current density can exceed one $\mathrm{A/cm^2}$. Thus, μm-scale test structures for TDDB can be affected by series resistance.[230] Complementary to the first FinFET

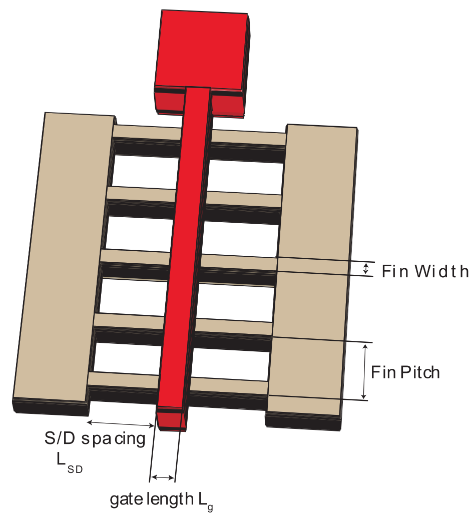


Fig. 5.14 The typical FinFET structure in this investigation comprises 40 nm gate length, 20 nm fin width, 30 nm fin height and 5 fins. The pitch was 200 nm. (not to scale).

results described[231] these limitations are circumvented by measuring TDDB on devices of the actual technology node (both SBD and HBD). The all-in-one methodology[232–234] is applied to short-channel FinFETs to obtain the lifetime on real devices of the technology node and to show all-in-one plots for both *n*/pMOS FinFETs with two different EOTs.

Furthermore, earlier studies[230] indicate a relationship between EOT and SBD lifetime (cf. Figure 1 in ref.). Therefore, the all-in-one reliability of different EOT gate stacks is compared in terms of SBD and wearout both for *n*- and pMOS FinFETs.

5.2.1 TDDB reliability of HfO_2/TiN on Si-FinFETs

The TDDB measurements were carried out under constant voltage stress on bulk-Si FinFETs of dimension $40\,\text{nm} \times 20\,\text{nm} \times 30\,\text{nm} \times 5$ ($L_g \times W_{\text{fin}} \times H_{\text{fin}} \times NumberFins$, fin pitch 200 nm, $A = 1.6 \cdot 10^{-10}\,\text{cm}^2$), cf. Figure 5.14. The gate stack comprised 1 nm interfacial oxide/1.8 nm $HfO_2/3$ nm TiN on 300 mm substrate. For the metal gate, TiN deposited by atomic layer deposition (ALD) and physical vapor deposition (PVD) are compared. The devices were processed up to metallization level M1.

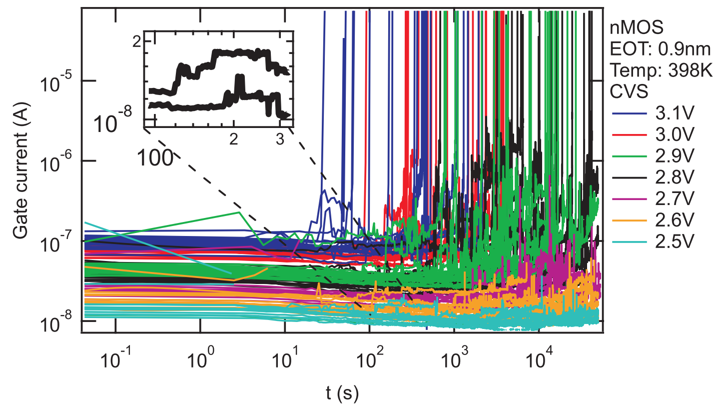


Fig. 5.15 Current-voltage traces under CVS for $(40 \times 20 \times 30\,\text{nm}) \times 5$ FinFETs with TiN (ALD) metal gate (EOT 0.9 nm): A long wearout phase is present. Inset: Current increase for first SBD of $\sim 1\,\text{nA}$.

After processing the equivalent oxide thickness was extracted to be 0.9 nm for the ALD deposited wafer and 0.8 nm for the PVD case.

5.2.1.1 nMOS FinFET results

The I-t traces obtained from short channel nMOS FinFETs with TiN (ALD) are shown in Figure 5.15. They show a large wear-out phase before the final hard breakdown. As suggested by Kauerauf,[230] additional lower stress voltages were chosen for the extraction of the SBD (Figure 5.17). The use of high V_g for SBD extraction would result in an underestimation of the wearout phase. As displayed in Figure 5.15, the four lowest stress voltages lead mostly not to HBD within measurement time. Those CVS data are suitable for the SBD extraction. Then the first significant current increase in the I-t traces was chosen as the current step trigger (digital behavior, inset in Figure 5.15). At higher stress voltages the SBD was triggered within the wearout phase due to samples with time-zero SBD, which had to be prevented. Otherwise, the wearout phase and number of SBD events would have been underestimated. However, the high V_g sets are needed for HBD extraction, whereby they are not used for SBD extraction (Figure 5.17).

Figure 5.16 shows the average leakage current at $V_g = 1\,\text{V}$ grouped by the dies used for the different stress conditions used afterwards in CVS. Although the leakage current is in the low pA-regime, certain variability is visible; please note the log scale compared to large area devices. Going to long fins (e.g. 1000 nm), this variability is not visible anymore due to averaging of the leakage current density. Using

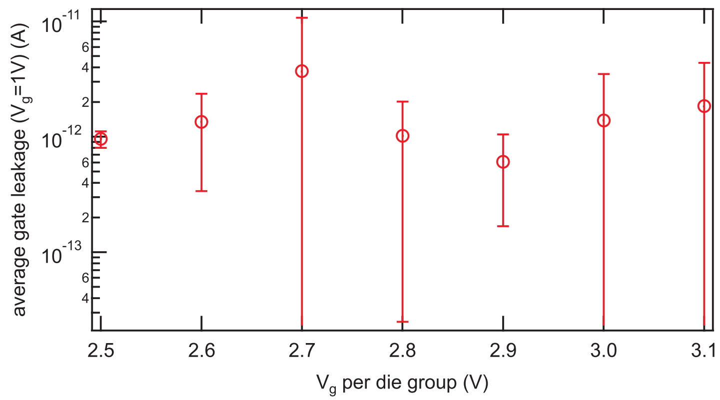


Fig. 5.16 Average gate leakage at $V_g = 1\,\mathrm{V}$ prior to CVS, grouped by later stress groups. In the short channel devices, variations cause a bigger absolute spread in leakage current, which can be problematic in subsequent TDDB analysis. Local measurements of TDDB can limit the influence of wafer non-homogeneity on the obtained Weibull η-values. Note also the low leakage values.

the short channel devices the variability impact was limited by measuring TDDB locally.[49]

Finally, a clear common Weibull slope for all four stress voltages could be fitted to the distribution by means of a maximum likelihood estimator (MLE) to the Weibull distribution. The SBD distributions were extracted using a current step criterion of $\Delta I = 1\,\mathrm{nA}$ (inset of Figure 5.15). Ashape parameter of $\beta_{\mathrm{SBD}} = 0.63$ was obtained (Figure 5.17). This value lies between the value usually addressed as one-trap path (0.35–0.4) and two-trap paths[230] (0.7–0.8). The main cause is likely the observed variability in leakage current even if staying locally to prevent against wafer inhomogeneity. Since wearout is a local phenomenon, this inhomogeneity could also cause stronger wearout acceleration in the device.

The shape parameter for HBD is found to be $\beta_{\mathrm{HBD}} = 1.0.t_{\mathrm{BD}}$ for hard breakdown was extracted using a current step criterion with $dI = 10\,\mu\mathrm{A}$ and the corresponding Weibull distributions are shown in Figure 5.18. The deviations around the extracted Weibull slope and the common η-spacing can be attributed to die-to-die variation of the device parameters. The result for HBD is smaller than obtained for large area test devices (~ 1.2) with comparable EOT and probably due to the variability for the small area devices. Imposing an inverse power law on the obtained η_{BD}-values the specification voltage for 10 years lifetime without HBD is obtained.[235] This straightforward lifetime extrapolation towards specification limits is referred to as the *apparent apparent* HBD Weibull slope.

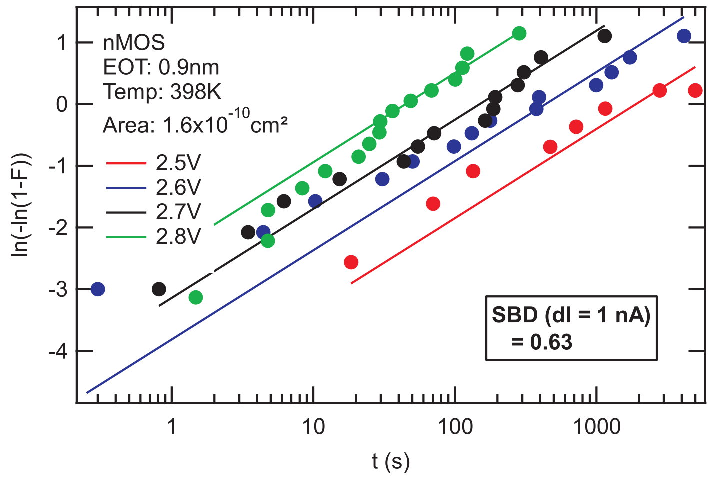


Fig. 5.17 SBD Weibull distributions for the nMOS FinFETs with EOT 0.9 nm (TiN ALD). A common slope is fitted to the four distributions. Device-to-device variations in leakage current cause the visible extra variability on the ranks.

The all-in-one TDDB lifetime extrapolation

The conventional lifetime extraction is pursued[230] by the extraction of soft- or hard-breakdown events from a set of $I-t$ traces measured at different gate voltages V_g. This extraction can be done by a current-step algorithm necessary for SBD extraction or a constant current algorithm for HBD detection.

The Weibull CDF (Figure 5.18) itself is estimated from the SBD/HBD data by applying a weighted Kaplan-Meier estimator (KME). The advantage of this weighted Kaplan-Meier estimator over the commonly used Bernard approximation is the ability to mix data series of different timeouts:

$$F_i = 1 - \frac{n+0.7}{n+0.4} \cdot \prod_{j \in F, j \leq i} \frac{n-j+0.7}{n-j+1.7} \tag{5.5}$$

where F_i is the CDF estimate of i-th ranked sample within n number of samples. A series of timeouts at an arbitrary position in the result set does not increase the rank, whereby in the Bernard approximation the rank would be artificially increased. However, for the subsequent MLE fit the result is not influenced by the offset in rank, as only the timing (t_{BD}) data are taken into account. The use of an adapted likelihood function that includes the timeouts in a proper way allows a reduced test time with increased number of timeouts within experimental groups to optimize test

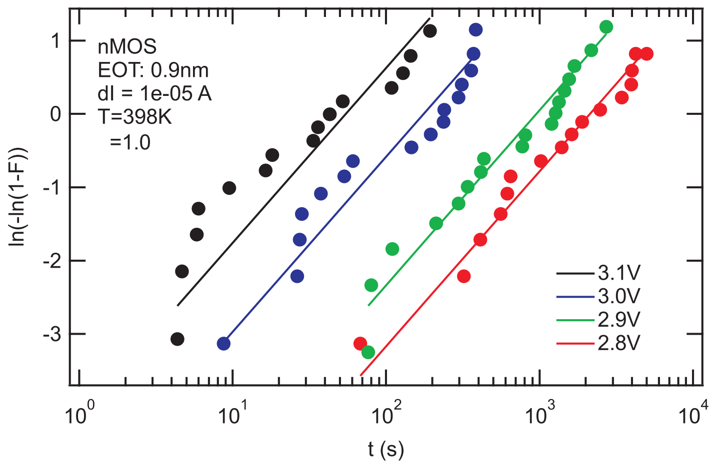


Fig. 5.18 Weibull distributions for the measured HBD times of the nMOS FinFETs with EOT 0.9 nm (TiN ALD). A common slope is fitted to the four distributions. Certain deviations in η-spacing can be attributed to die-to-die variations.

time and *to stabilize information on the Weibull slope in the distribution's lower tail*. Moreover, it allows for using information from datasets after e.g. a parallel reliability test had to be stopped before catastrophic failure.

The area-scaled Weibull distributions are then characterized by their common Weibull slope (β_{BD}) and the 63% failure value (η_{BD}) for the different V_g. Imposing a power law on the η-values, the lifetime at operating condition V_{spec} is obtained. Secondly, the obtained η is area scaled to the desired total gate oxide area, e.g. 0.1 cm^2. Thirdly, this result is scaled to low failure percentiles. This approach results in two area and probability scaled lifetime extrapolations, one each for SBD and the apparent HBD, with common information: The maximum applicable V_g without SBD and without HBD for 10 years lifetime. A power law fit for the apparent HBD is shown in Figure 20 as an example for pMOS.

The purpose of the all-in-one TDDB prediction is to account for the competition effect of SBDs during wearout and combine information in a convenient manner. Thereby SBD and HBD information is convoluted in the peculiar way leading to the *Roussel distribution*,[230,232,233] which is fitted data:

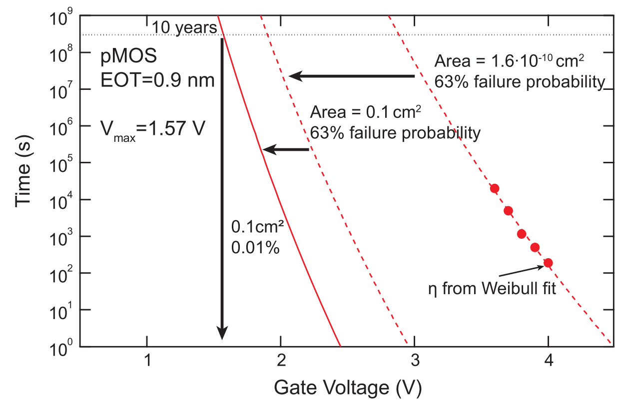


Fig. 5.19 Example for pMOS of a power law fit and subsequent area and probability scaling to obtain the max applicable gate voltage for ten years lifetime. The fit is obtained as a power law. Absolute voltages are shown and the max. voltage for 10 years lifetime is obtained from the apparent HBD distribution as $V_g = -1.57\,\mathrm{V}$.

$$F_{\mathrm{HBD}}(t) = 1 - \exp\left[-\left(\frac{\eta_w}{\eta_C}\right)^{\beta_C} \int_0^{1-\exp\left(-\frac{t-\tau_C}{\eta_W}\right)^{\beta_W}} \times \left(\frac{t-\tau_C}{\eta_w} - (-\ln(1-Ftw))^{1/\beta_w}\right)^{\beta_C} \mathrm{d}Ftw \right], \tag{5.6}$$

where Ftw is the integral substitution for the area scaled Weibull distribution and τ_c can be used to model initial damage in the oxide (here equals zero). All w-indices are related to wearout and are fit parameters. The c-indices (short for creation) mark the initial SBD distribution. The purpose of the formula is given in detail by Roussel[232] and can be summarized; an area-downscaling of the first SBD such, that no 2nd SBD can happen before wearout of the first event. The scaled SBD distribution is then convoluted with the wearout phase (integral) to lead to the HBD distribution. This distribution is scaled back to the originally measured area to obey the correct HBD distribution.

Coming back to the FinFET results, the resulting plot is marked by a SBD free region (e.g. Figure 5.20 (left), green) enclosed by the 0.01% failure value for a single SBD and the 10 years lifetime at given total gate oxide on the chip (y-axis). Second, the region with multiple SBD and wearout follows (orange), whereby the dashed lines correspond to the higher percentiles of a single SBD up to 63% failures (thick dashed, red). From 63% failures on, the following lines mark the occurrence

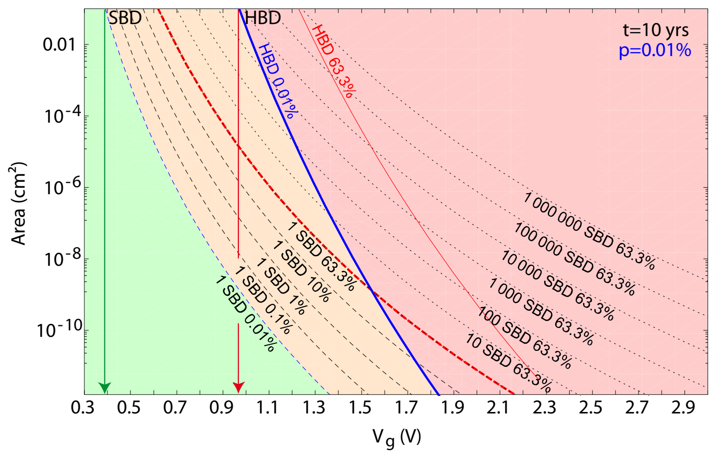


Fig. 5.20 All-in-one lifetime extrapolation of SBD and HBD information for nMOS high-κ/TiN (ALD) stack on FinFET (EOT 0.9 nm). After rescaling, a gate voltage of 0.97 V is obtained for 10 years operation without HBD (0.01% failure/0.1 cm^2 chip area). Green (left) area is SBD free, orange (middle) marks wearout phase, red (right) is the limit to HBD. Dotted lines: percentiles of SBD distribution. $V_{g,\,\mathrm{SBD}} = 0.38\,\mathrm{V}$. The green and red arrows mark the maximum operating voltage without SBD and HBD, respectively.

of multiple SBD (by powers of ten). For HBD the 0.01% and 63% CDF value are plotted in the same graph (blue and red).

Strength of this plot arrangement is the area scale on the *y*-axis; based on the total gate dielectric area of the product, the lifetime and SBD data can be read off.[230] This is valuable since the FEOL modules and the corresponding TDDB lifetime might be used for different products. Thereby the extrapolated lifetime on a certain chip area is directly accessible as a decision aid.

The all-in-one lifetime extrapolation of SBD and HBD information for the devices with TiN (ALD) are shown in Figure 5.20. At 0.01% failure and 0.1 cm^2 chip area (upper limit of left axis), a gate voltage of 0.97 V is extrapolated for 10 years operation without HBD.

The first left dotted line in the plot shows the 0.01% failure probability for a single SBD (log-scale). The next lines show higher percentiles for one SBD (n.b. labels). From the red dotted line on, the lower percentiles are replaced by the 63.3% value of the Gamma Regularized distribution for increasing number of SBD. Again, the number of SBD is log-scaled (starting from unity).

For the case of PVD deposited TiN, a lower lifetime is expected due to the overall lower EOT and consequently higher leakage current density J_g. For the detection of SBD and HBD the same triggers as for the ALD TiN wafer were used. The applied stress conditions had to be chosen lower to account for the higher leakage

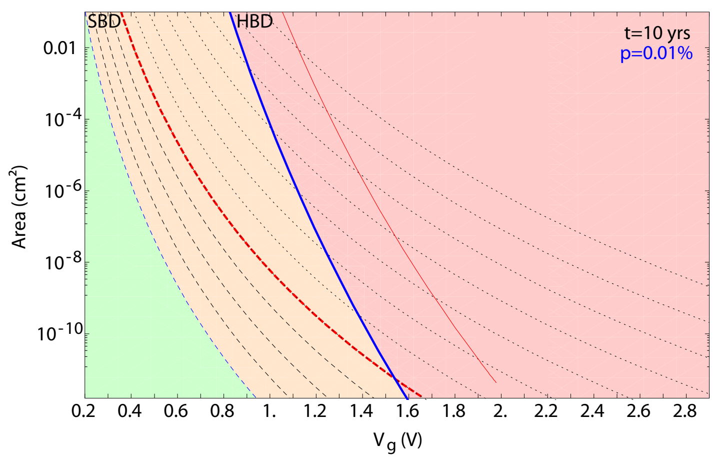


Fig. 5.21 All-in-one plot lifetime extrapolation of SBD and HBD information for nMOS high-κ/TiN (PVD) stack on FinFET (EOT 0.8 nm): $V_{g,\ \mathrm{SBD}} = 0.35\,\mathrm{V}$ and $V_{g,\ \mathrm{HBD}} = 0.82\,\mathrm{V}$ is extracted for 10 yrs. lifetime.

current. The SBD and HBD information is plotted in Figure 22. Due to the increased leakage current in this gate stack, a decreased overall operating voltage of 0.82 V is extracted ($\beta_{\mathrm{HBD}} = 0.93, m_{\mathrm{HBD}} = -40$). The shape parameter obtained from the SBD information amounts to $\beta_{\mathrm{SBD}} = 0.52$, $\beta_{wo} = 0.42$ and the absolute voltage acceleration during wearout $m_{wo} = 41$.

Moreover, the voltage acceleration was observed as $m_{\mathrm{SBD}} \approx m_{\mathrm{WO}}$ to be consistent with the findings of Sahhaf.[233,234] Lower EOT and faster wearout can be linked by this model: Comparing WO for the different gate stacks (PVD and ALD TiN) *higher voltage acceleration during wear-out* for the PVD electrode (lower EOT) is observed. Apparently, comparing the 1st SBD in Figure 5.20 and Figure 5.21 the WO phase starts also earlier in the PVD gate stack (as expected for lower EOT): Although the shape parameters β_{SBD} are comparable, the 63% value η_{SBD} differs and leads to earlier wearout. Then the resulting lifetime reduction is found to be less remarkable than the much increased wearout phase. *The wearout starts already at 0.2 V for ten years lifetime.* The lower lifetime in the PVD case could be compensated by e.g. 0.6 nm thicker HfO_2 layer, which would result in an additional reduction of leakage current.

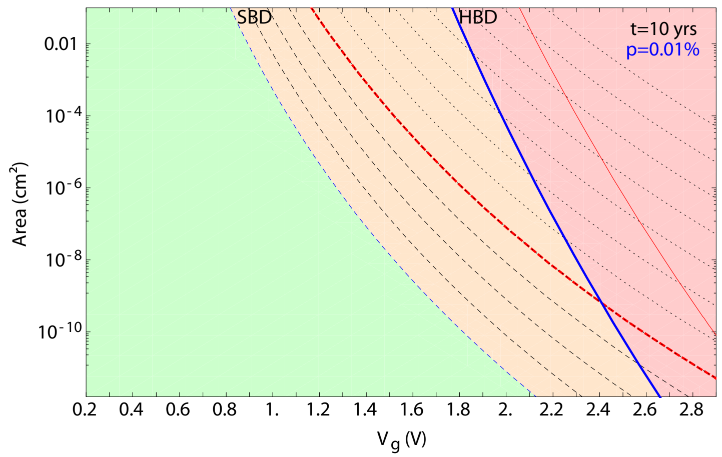


Fig. 5.22 All-in-one plot lifetime extrapolation of SBD and HBD information for pMOS high-κ/TiN (ALD) stack on FinFET (EOT 0.9 nm): $V_{g,\,\mathrm{SBD}} = 0.82\,\mathrm{V}$ and $V_{g,\,\mathrm{HBD}} = 1.77\,\mathrm{V}$ are extracted for 10 years lifetime without chip failure.

5.2.1.2 pMOS FinFET results

For the pMOS with 40 nm nominal gate length and ALD TiN electrode SBD and HBD extraction were obtained using a current step criterion with $dI = 0.25\,\mathrm{nA}$ and $dI = 100\,\mathrm{nA}$, respectively. The current step triggers for pMOS were chosen lower compared to nMOS to account for the different current each breakdown conducts. A maximum operating voltage of 1.77 V was extrapolated for 10 years. *The pMOS shows drastically increased lifetime compared to the nMOS due to larger values in β and η in the Weibull distribution for HBD* (Figure 5.22).

The distributions show a common Weibull slopes of $\beta_{\mathrm{SBD}} = 0.80$ and $\beta_{\mathrm{HBD}} = 1.29$. The soft-breakdown value lies between a typical 2- or 3-trap path value.[230] Most likely, the observed device variability is lowering the resulting slope also for pMOS. In the pMOS the voltage acceleration (exponent) during wearout ($m_{\mathrm{WO}} = 52$) is comparably larger than for nMOS ($m_{\mathrm{WO}} = 47$). Nevertheless, more SBD until HBD will likely happen for the pMOS device (Figure 5.22): The extrapolated lifetime is much higher in pMOS, therefore, the device undergoes more SBD and relative leakage current. Consequently, the absolute voltage acceleration for wearout m_{WO} is larger. Here the reference $\eta_{\mathrm{SBD}} \approx \eta_{\mathrm{WO}} \approx 23\,\mathrm{s}$ and, therefore, the sum of the shape parameter β_{SBD} and β_{WO} is different from β_{HBD}.[234] In this case the wearout and SBD are convoluted in a peculiar way[232] to give the HBD distribution. In agreement, a power law exponent was found for the voltage acceleration m_{SBD} that differs from the exponent during WO acceleration.

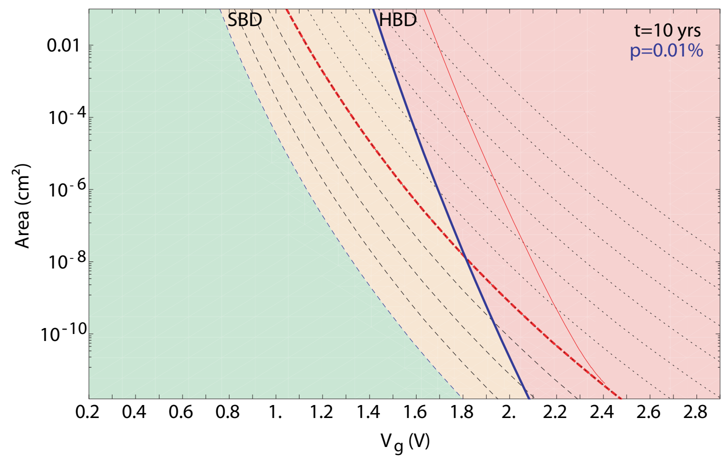


Fig. 5.23 All-in-one plot lifetime extrapolation of SBD and HBD information for pMOS high-κ/TiN (PVD) stack on FinFET (EOT 0.8 nm): $V_{g,\,\mathrm{SBD}} = 0.76\,\mathrm{V}$ and $V_{g,\,\mathrm{HBD}} = 1.42\,\mathrm{V}$ is extracted for 10 yrs. lifetime without chip failure.

For the pMOS devices with TiN (PVD) SBD and HBD extraction were done with a current step criterion of $dI = 5\,\mathrm{nA}$ and $dI = 100\,\mathrm{nA}$, respectively. pMOS with TiN (PVD) show also a decreased 10 years HBD gate voltage as expected from the lower EOT. The all-in-one plot for the PVD TiN pMOS is shown in Figure 5.23 estimating the 10 years gate voltage without HBD to be 1.42 V. The characteristics for SBD and HBD were calculated to amount to $\beta_{\mathrm{SBD}} = 0.9$, $\beta_{\mathrm{HBD}} = 1.0$. The wearout phase can be described by $\beta_{\mathrm{WO}} = 0.76$, $m_{\mathrm{WO}} = 47$. Thus, here the cause of the soft-breakdown is identified as a 3-trap path, whereby the imposed variability is again lowering the resulting slope β_{SBD}. Thereby the results for SBD in the pMOS with two different EOT gate stacks remain comparable. Moreover for the pMOS with lower EOT the small difference of the shape parameter β_{BD} indicates that SBD is the dominant effect which leads to HBD as first described by Sahhaf.[234] In contradiction to Sahhaf,[234] in this case of $\eta_{\mathrm{SBD}} \gg \eta_{\mathrm{WO}}$ and small area, the voltage acceleration of SBD $(m_{\mathrm{SBD}} = 32)$ was smaller than the absolute power law exponent of the WO acceleration $(m_{\mathrm{WO}} = 47)$. Here: $\eta_{\mathrm{SBD}}/\eta_{\mathrm{WO}} \approx 10$.

The dominance of SBD for the distinction of the HBD is apparent in the shape parameters of SBD and HBD. It leads to the merging of both 63% percentiles at 2.4 V in Figure 5.23. The apparent Weibull distribution as the standard lifetime extrapolation would lead to nearly comparable distributions for SBD and HBD. Thus, the wearout would be negligible. However, in the special case of the results shown in Figure 5.23 and the constraints mentioned above, the Roussel distribution leads to a lifetime extrapolation with more wearout.

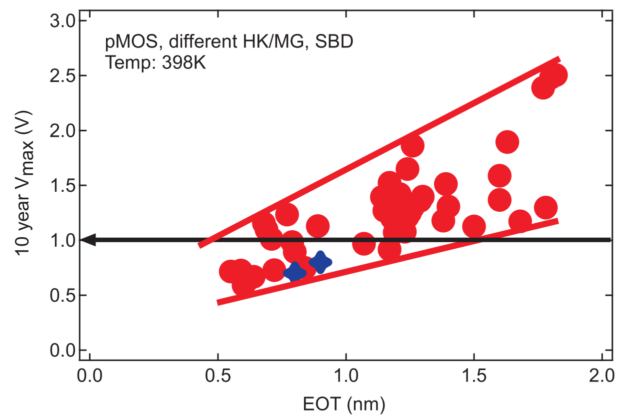


Fig. 5.24 10 yrs. SBD lifetime extrapolation for different high-κ/metal gate stacks on planar $1 \times 1\,\mu m^2$ pMOS devices (data courtesy of T. Kauerauf). The soft-breakdown was triggered. This experiment: (blue) stars indicating the SBD lifetime in FinFET devices. Here, the SBD lifetime is also found well below the typical target specification of 1 V but matches the results obtained on planar test structures.

In Figure 5.24 the SBD lifetime (Figure 5.21 and Figure 5.23) are compared with those found on planar test structures (data for planar devices courtesy of T. Kauerauf[230]). For pMOS FinFETs, the maximum operating voltage without SBD is comparable to its planar counterpart. Nevertheless, the SBD lifetime extrapolation remains well-below 1 V if going to sub-1 nm EOT gate stacks.

5.2.1.3 Summary

TDDB results for short-channel FinFETs combined in the all-in-one reliability were presented. The results were obtained under constant voltage stress. *For low-EOT (FinFET) devices* $(EOT < 1\,nm)$ *the soft-breakdown and wearout phase start well below operating voltage*. The following wearout is intrinsic part of the degradation and should not be neglected. In the all-in-one TDDB reliability analysis,this wearout phase can be addressed together with the final hard-breakdown and *allows quantifying the increase in leakage current during lifetime*.

Comparable SBD shape parameters are extracted for the nMOS (substrate injection)and two different EOT. Earlier onset of wearout was found for lower EOT. Apparently, the lower EOT leads to a 10 year HBD free gate voltage below $V_{dd} = 1\,\mathrm{V}$ ($V_{g,\,\mathrm{HBD}} = 0.82\,\mathrm{V}$ for EOT 0.8 nm vs. 0.97 V for EOT 0.9 nm).

Second, but not less important the accelerated wearout phase will also increase the leakage current over product lifetime for these lower EOT gate stacks.

For both ALD and PVD TiN pMOS (gate injection) close SBD lifetimes and acceleration were extracted, but also 15–20% different HBD lifetimes with $V_{g,\,\mathrm{HBD}} = 1.42\,V$ (EOT 0.8 nm) and $V_{g,\,\mathrm{HBD}} = 1.77\,\mathrm{V}$ (EOT 0.9 nm). In parallel, the absolute wearout acceleration m_{WO} for pMOS was larger than for nMOS and leads to longer wearout phase.

The results for SBD are in agreement with those for sub-1 nm EOT gate stacks on planar devices. However, the results for the Weibull slope β_{SBD} showed slight variations from the well-accepted trap quantization as found in literature. Comparing the leakage current before CVS this deviation can mainly be accounted to the device variability observed for small area devices.

The different EOT values were yet achieved with constant thickness of HfO_2 but different strength of scavenging. The scaled IL might lead to comparable onset of SBD. The well-known fact of reduced lifetime at reduced EOT[230] can be argued within the all-in-one reliability as caused by higher voltage acceleration during the wearout phase. Therefore, the reliability in general and especially the wearout of sub-1 nm high-κ gate stacks increases in importance for the leakage current increase and overall device lifetime.

In general, the device variability observed for nm-scaled FinFETs, i.e. small area devices,exacerbate the TDDB parameter extraction. Nevertheless, a measurement and judgment of small area test structures is mandatory to predict the lifetime of the field devices.

5.2.2 The correct determination of the SBD information

The information extracted from SBD data has advanced in importance as the previous section showed. The MOSFETs with sub-1 nm EOT feature an onset of wearout well below 1 V operating voltage (10 yrs. lifetime). The accurate extraction of both SBD and HBD data is therefore mandatory for the meaningfulness of the all-in-one methodology.

One obstacle herein is the correct determination of the initial SBD events. If the SBD formation occurs within measurement time resolution, one will easily underestimate the wearout phase in the framework of the all-in-one methodology.

Figure 5.25 (bottom) for example shows the same dataset for HBD as in Figure 5.25 (top) and Figure 5.20 but the SBD information on lower voltages was not taken into account. The apparent HBD Weibull slope as originally depicted in Figure 5.18 does not change. However, the convolution of SBD and HBD information within the Roussel distribution changes, resulting in a virtual decreased wearout phase with only 322 SBD events before HBD compared to 10^6 in the correct plot. Moreover, the 10 years HBD voltage is virtually increased from 0.97 to 0.99 V. The creation of the first SBD leads to a different shape parameter than during the wearout phase.[230] Thus, the correct determination of the first creation of soft-breakdown events remains crucial and similar results were found for pMOS.

5.2.3 SILC slope for SBD determination in short-channel MOSFETs

The utilization and extraction of the stress-induced leakage current (SILC[35]) slope for soft-breakdown in short-channel MOSFETs is limited if the device architecture demands for testing on smaller dimensions (e.g. nanowire MOSFETs or FinFETs). Since stress-induced leakage current is a stochastic effect scaling with device area, single devices of small area ($\sim 10^{-10}\,\mathrm{cm}^2$) will only show the individual SBD events (Figure 5.26). Additionally, stress-induced leakage current cannot be used for pMOS due to the too small current increase during wearout.

The onset of increasing slope at $V_g = 2.25\,\mathrm{V}$ is the onset of wearout in the small area and not equal to a common SILC slope as known from large area devices. A SILC slope extraction from this data would result in an underestimation of the true initial SBD. Therefore, extreme care has to be given to the design of new test structures for SBD determination. To achieve an overall sufficient area for the use of the SILC slope method, many devices should be connected in parallel test structures to achieve an area comparable to that of large area devices ($> 10^{-8}\,\mathrm{cm}^2$). Compared to the previous stress-induced leakage current results[236] the amount of stress-induced leakage current is further increasing with decreasing EOT. Therefore, traditional trigger at 100% current increase or constant current levels will fail for ultra-thin EOT devices and possibly underestimate the lifetime.

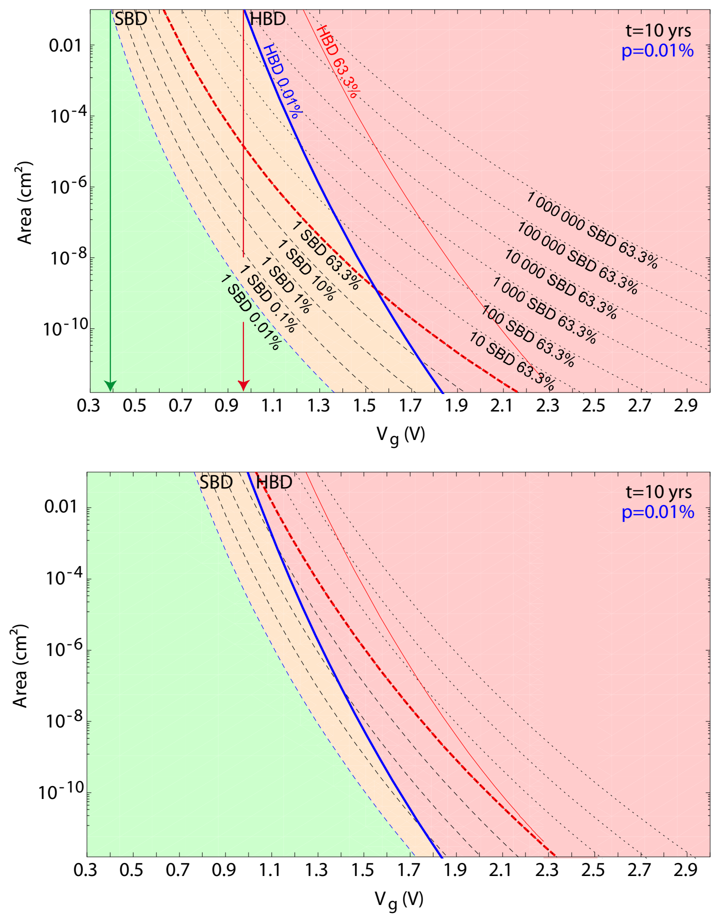


Fig. 5.25 Top: All-in-one plot for SBD, wearout and HBD of the measured nMOS bulk FinFETs; Bottom: The same HBD information using only the SBD information from the four voltages used for HBD extraction.

The SILC slope in nMOS large area devices with an EOT value below 1 nm may be equivalent to a current one would judge as HBD and set a trigger to a current level within the wear-out phase (dashed line in Figure 5.27). Additionally, leakage current densities approach $J_g \sim 10^{-4}\,\mathrm{A/cm^2}$ in pMOS bulk FinFETs. Thus, in scaled

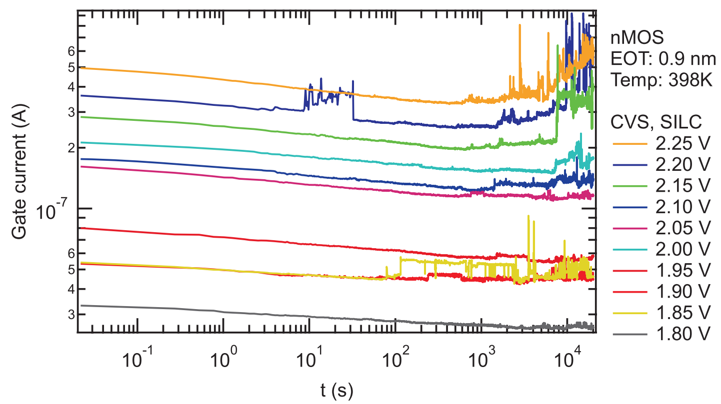


Fig. 5.26 CVS traces for SILC determination on small area devices $(1.6 \times 10^{-10}\,\mathrm{cm}^2)$; no common SILC slope is present, only individual SBD events are present. The present slope for increased V_g is the onset of wearout in the device and not SILC.

SC devices leakage currents of around 10 fA at 1 V demand for parallel-connected devices to allow a meaningful leakage current measurement on an increased test area. Current step detection algorithms and use of parallel test structures can help in reducing the variations in initial leakage currents as those observed in Figure 5.26 at 1.85 V.

5.2.4 Outlook on test time optimization

The TDDB results suggest the use of CVS in small area devices. Thereby SBD and HBD are likely to be extracted at different stress voltages, which imply extended measurement times for the determination of SBD and wearout. In the worst case, the number of stress voltages has to be doubled. Lower stress voltages combined with small variability in the initial leakage current can exacerbate the termination of the experiment: Too tight boundaries for the constant current trigger may cause an increased number of failures within measurement time resolution. The aforementioned use of a relatively high constant current trigger for SBD can lead to increased number of suspensions, i.e. devices without failure within measurement time.

A judgment of measurement conditions and their impact on measurement error is presented next for the case of a bootstrap experiment,[237] i.e. repeated draws from a known stochastic population, which assumes different censoring times for an increasing number of DUTs. The choice of the censoring time leads to an obvious reduction of measurement time. In parallel to this reduced measurement time, the

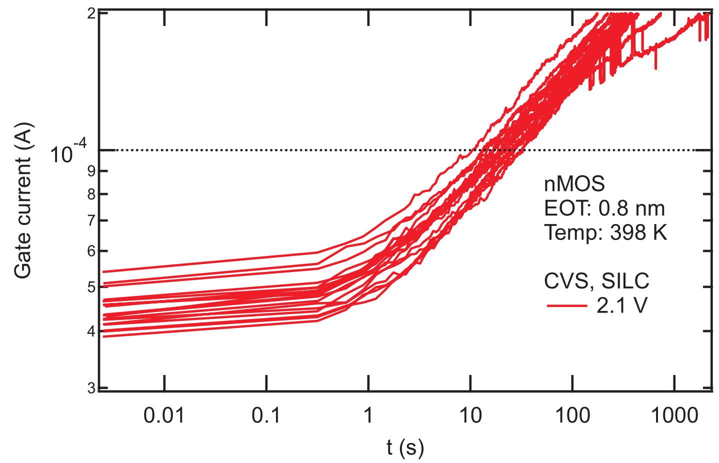


Fig. 5.27 SILC traces during CVS for nMOS long-channel FinFETs with $A = 4 \cdot 10^{-9}\,\mathrm{cm}^2$, traditional constant current triggers (e.g. $100\,\mu$A) fail due to SILC. With a stop criterion chosen below R_s dependence no HBD is observed.

error bars on the parameter extraction could be maintained. The influence of the number of DUTs on the evolution of measurement error is also discussed.

5.2.4.1 Mathematical framework

The bootstrap experiment was carried out using the apparent Weibull slope for left- and right censored data. One speaks of left-censored data, if a device fails within the experimental time resolution (here 0.1 or 0.3 s) – or is already broken prior to measurement. The right-censorship is reached if a device does not fail within the experimental time window.

The Log Likelihood function yields then the theoretically assumed distribution function, which needs to be fitted to the population. This Log Likelihood function for the Weibull distribution is extended to left- and right censorship and voltage-scaling based on prior work of Roussel[232] and Kauerauf[49] for the area-scaled function. A solution to the Log Likelihood function is then given by maximization. To maximize the Log Likelihood function, the first derivative is deduced under the requirement of minimal covariance between the reference η_r and the area- and voltage acceleration, a_A and a_V, respectively. The latter leads to significant simplifications in the equations and can be imposed due to statistical independence of reference scale and acceleration parameters (hence later, covariance in between equals zero). Further derivation of the 2$^{\text{nd}}$ derivatives is needed to get the variance and covariance elements as the error estimates (in analysis called the Hessian matrix). Here, the minimal covariance argument leads also to simplifications.

The mathematical framework is subsequently applied in the bootstrap experiment to describe the influence of censorship on measurement time and error estimates.

The Log Likelihood function

The narrow experimental window observed for CVS on low-EOT gate stacks leads to the observation of both left- (pseudo-initial fails) and right-censored (suspensions) Weibull-distributed data sets. In this paragraph, the Likelihood equations are conducted for the Weibull distribution under these censors.

The censorship is accounted by the use of the non-normalized Multinomial Bayesian likelihood:

$$L = \prod_{i,k=1}^{P} (F_{ik}(\delta,\beta,\eta) - F_{ik}(0,\beta,\eta)) \prod_{i,k=1}^{R} f_{ik}(t_{ik},\beta,\eta) \cdot \prod_{i,k=1}^{s} (1 - F(t_{ik},\beta,\eta)) \quad (5.7)$$

f_{ik}: Weibull probability density function (pdf)
F_{ik}: Weibull cumulative distribution function (cdf)
P is the product over the probability of the interval failures between time zero and lower detection limit, i.e. that one device fails within time resolution. R is the product over all real failures and S the product over the reliability function of suspensions, i.e. survived until experiment end. The product runs over k experiments (e.g. different accelerations) and i devices. The total number of DUTs N is then given as $N = P + R + S$.

The corresponding Log likelihood function Λ for the censored Weibull distribution transforms accordingly:

$$\Lambda = \sum_{j,k=1}^{P} \ln\left(1 - \exp\left[-\left(\frac{\delta}{\eta_{jk}}\right)^{\beta}\right]\right) + \sum_{j,k=1}^{R} [\ln\beta + z_{jk} - \ln t_{jk} + \ln R_{jk}(t_{jk})] + \sum_{j,k=1}^{s} \ln R_{jk}(t_{jk}) \quad (5.8)$$

with

$$R_{jk}(t_{jk}) = 1 - F_{jk}(t_{jk}) = \exp\left[-\left(\frac{t_{jk}}{\eta_{jk}}\right)^{\beta}\right] \quad \text{and Weibit} \quad z_{jk} = \beta(\ln t_{jk} - \ln\eta_{jk}).$$

R_{jk} is the Weibull reliability function. As long as the time resolution is small compared to the scale parameter value at the stress condition considered, the first sum can be approximated with $1 - \exp(-x) \approx 1 - x$:

$$\begin{aligned}\Lambda &= \sum_{j,k=1}^{P} \beta(\ln t_{jk} - \ln \eta_{jk}) + \sum_{j,k=1}^{R} [\ln \beta + z_{jk} - \ln t_{jk} + \ln R_{jk}(t_{jk})] \\ &\quad + \sum_{j,k=1}^{S} \ln R_{jk}(t_{jk}) \\ &= \sum_{j,k=1}^{P} z_{j,k} + \sum_{j,k=1}^{R} [\ln \beta + z_{jk} - \ln t_{jk} + \ln R_{jk}(t_{jk})] + \sum_{j,k=1}^{S} \ln R_{jk}(t_{jk}) \end{aligned} \tag{5.9}$$

The three sums can be contracted using the symbol notation δ_{jk} and the unity step function $U(x)$:

$$\delta_{jk} = \begin{cases} -1 & \text{pseudo initial fails} \\ 1 & \text{for real fails} \\ 0 & \text{suspensions} \end{cases}$$

Thus, the Log likelihood function for a vector or list of flagged δ_{jk} time results t_{jk} follows in a very compact form:

$$\begin{aligned}\Lambda(\beta, \eta_{jk}, t_{jk}, \delta_{jk}) = \sum_{j,k=1}^{N} [&Z_{jk} \cdot |\delta_{jk}| + \ln \beta \cdot \delta_{jk} \cdot U(\delta_{jk}) - \ln t_{jk} \cdot \delta_{jk} \cdot U(\delta_{jk}) \\ &+ \ln R_{jk}(t_{jk}) \cdot U(\delta_{jk})]\end{aligned} \tag{5.10}$$

The maximum Likelihood equations

The maximum likelihood equations are obtained as the extremes of the Log likelihood equation Λ. First, the equations for arbitrary scaling are derived. The consistency of the results with the inverse power law (IPL) scaling for the acceleration voltage and the area scaling[238] will be shown last.

The scaling is introduced into the parameter η_k of the Weibull distribution for k different experimental conditions by:

$$\eta_k(\eta_r, a_A, a_V) := \eta_r \left(\frac{A_k}{A_r}\right)^{a_A} \left(\frac{V_k}{V_r}\right)^{a_V} \tag{5.11}$$

$$\ln \eta_k(\ln \eta_r, a_A, a_V) := \ln \eta_r + a_A \ln\left(\frac{A_k}{A_r}\right) + a_V \ln\left(\frac{V_k}{V_r}\right) \tag{5.12}$$

The η_k of the k-th experiment with area A_k and voltage stress V_k becomes then a function of the reference η_r, the exponents of area scaling a_A and voltage scaling a_V. Both are referenced to a common reference area A_r and a reference voltage V_r. The maximum Likelihood equations (MLE) are obtained under the requirement of minimal covariance between each of the acceleration parameters and the reference scale parameter η_r:

$$\partial_\beta \Lambda = \sum_{j,k=1}^{N} [\ln(t_{jk}) - \ln(\eta_k)]|\delta_{jk}| + \ln R_{jk}(t_{jk}) \cdot (\ln t_{jk} - \ln \eta_k) U(\delta_{jk})$$
$$+ \frac{1}{\beta} \delta_{jk} U(\delta_{jk}) \tag{5.13}$$

$$\partial_{\eta_r} \Lambda = -\frac{\beta}{\eta_r} \sum_{j,k=1}^{N} [|\delta_{jk}| + \ln R_{jk}(t_{jk}) U(\delta_{jk})] \tag{5.14}$$

$$\partial_{\eta_r} \Lambda = 0 \Longrightarrow \sum_{j,k=1}^{N} R_{jk}(t_{jk}) \cdot U(\delta_{jk}) = -\sum_{j,k=1}^{N} |\delta_{jk}| \tag{5.15}$$

$$\partial_{a_A} \Lambda = -\beta \sum_{j,k=1}^{N} \ln\left(\frac{A_k}{A_r}\right) \left[|\delta_{jk}| + \underbrace{\ln R_{jk}(t_{jk}) U(\delta_{jk})}_{=\partial_{\eta_r, a_A} = 0} \right] \tag{5.16}$$

$$\partial_{a_V} \Lambda = -\beta \sum_{j,k=1}^{N} \ln\left(\frac{V_k}{V_r}\right) \left[|\delta_{jk}| + \underbrace{\ln R_{jk}(t_{jk},) U(\delta_{jk})}_{=\partial_{\eta_r, a_V} = 0} \right] \tag{5.17}$$

By further deriving (5.15) and requiring the two FIM elements mentioned above to be zero expressions for the reference area A_k and voltage V_k are obtained:

$$0 \stackrel{!}{=} \partial_{\eta_r, a_A} \Lambda = \sum_{j,k=1}^{N} \ln\left(\frac{A_k}{A_r}\right) \ln R_{jk}(t_{jk}) U(\delta_{jk}) \tag{5.18}$$

As this sum is part of the Max. Likelihood equation (5.16) in a_A, the latter collapses into:

$$\sum_{j,k=1}^{N} \ln\left(\frac{A_k}{A_r}\right) |\delta_{jk}| = 0 \tag{5.19}$$

and is independent from the distribution parameters, so that it is handier to use this equation for computing A_r first:

$$\ln A_r = \frac{\sum_{j,k=1}^{N} \ln(A_k) \cdot |\delta_{jk}|}{\sum_{j,k=1}^{N} |\delta_{jk}|} \tag{5.20}$$

so that in effect, (5.18) becomes the ML equation for a_A. Thanks to the symmetry in the parameters, the expression for V_r and a_V is deduced in the same manner:

$$\ln V_r = \frac{\sum_{j,k=1}^{N} \ln(V_k) \cdot |\delta_{jk}|}{\sum_{j,k=1}^{N} |\delta_{jk}|} \tag{5.21}$$

$$\sum_{j,k=1}^{N} \ln\left(\frac{V_k}{V_r}\right) \ln R_{jk}(t_{jk}) U(\delta_{jk}) = 0 \tag{5.22}$$

Solution of the maximum Log-Likelihood equations

The equation $\partial \eta_r \Lambda = 0$ can be solved in η_r^β by the use of:

$$\ln R_{jk}(t_{jk}) = -t_k^\beta \eta_r^{-\beta} A_k^{-a_A\beta} A_r^{a_A\beta} V_k^{-a_V\beta} V_r^{a_V\beta} \tag{5.23}$$

and

$$s_k := \left(\frac{A_k}{A_\Gamma}\right)^{a_A} \left(\frac{v_k}{v_r}\right)^{a_V} \tag{5.24}$$

$$\eta_r^\beta = \frac{\sum_{j,k=1}^{N} (t_{jk}/s_k)^\beta U(\delta_{jk})}{\sum_{j,k=1}^{N} |\delta_{jk}|} \tag{5.25}$$

Equations (5.18) and (5.22) can be transformed to obtain the MLE in a_A and a_V:

$$0 = \sum_{j,k=1}^{N} (\ln A_k - \ln A_r) t_{jk}^\beta A_k^{-a_A\beta} V_k^{-a_V\beta} U(\delta_{jk}) \tag{5.26}$$

$$0 = \sum_{j,k=1}^{N} (\ln V_k - \ln V_r) t_{jk}^\beta A_k^{-a_A\beta} V_k^{-a_V\beta} U(\delta_{jk}) \tag{5.27}$$

Thus, the deduction of (5.26) and (5.27) is based on the minimal covariance requirements so that the 2$^{\text{nd}}$ derivative becomes the MLE for a_A and a_V.

The MLE in β is obtained by plugging (5.15) into the MLE for a_A(5.26) and a_V(5.27) and their reference values (5.20) and (5.21):

$$0 = \partial_\beta \Lambda = \cdots = \sum_{j,k=1}^{N} \ln t_{jk} |\delta_{jk}| + \frac{1}{\beta} \delta_{jk} U(\delta_{jk}) - \left(\frac{t_{jk}}{s_k}\right)^\beta \frac{1}{\eta_{r^\beta}} \ln t_{jk} U(\delta_{jk}) \tag{5.28}$$

By plugging in (5.25) this equation becomes independent of η_r:

$$\Longrightarrow 0 = \partial_\beta \Lambda$$
$$= \sum_{j,k=1}^{N} \left[\ln t_{jk} |\delta_{jk}| + \frac{1}{\beta} \delta_{jk} U(\delta_{jk}) - \frac{\ln t_k (t_{jk}/s_k)^\beta U(\delta_{jk})}{\sum_{j,k=1}^{N} (t_{jk}/s_k)^\beta U(\delta_{jk})} \sum_{j,k=1}^{N} |\delta_{jk}| \right] \tag{5.29}$$

With (5.25), (5.26), (5.27) and (5.29) the MLE are obtained for reference scale parameter η_r, the area scaling a_A, the voltage acceleration a_V and shape parameter β.

By assuming Poissonian area scaling[238] the area scaling parameter becomes: $a_A = -\frac{1}{\beta}$. This result is consistent with the earlier equation found[49] for $a_V = 0$, i.e. without voltage scaling. For the case of the widely accepted power law scaling[49,239] the scale parameter η becomes then:

$$\eta(\eta_r, \beta, n) := \eta_r \left(\frac{Ak}{A_r} \right)^{-1/\beta} \left(\frac{V_k}{V_r} \right)^{a_V} \tag{5.30}$$

The Fisher information matrix of the Weibull distribution

The error estimation for the prior deduced MLE of the Weibull distribution parameters can be derived from the inverse of the Fisher Information Matrix (FIM).

The FIM, also known in analysis as the Hessian of a function is the matrix of 2^{nd} derivatives in the parameters of the (Weibull) distribution. The parameters are $\{\beta, \eta_r, a_A, a_V\}$ giving a square matrix in four dimensions $(\partial_{i,j}\Lambda(\beta, \eta_r, a_A, a_V))_{i,j=\{\beta,\eta_r,a_A,a_V\}}$. The parameters' variances are then given as the diagonal elements, the covariances are found as off-diagonal elements.

Some noteworthy simplifications of this FIM are achieved by applying the aforementioned argument of minimal covariance between reference and voltage a_V and area acceleration a_A, respectively:

$$\partial_{\eta_r,a_V}\Lambda = \partial_{\eta_r,a_A}\Lambda = 0$$

Moreover, the results for reference voltage V_r and area A_r can be applied for rearranging the derivatives. By iteratively plugging in both simplifications the 2^{nd} derivatives become much more compact (summation over $j,k = 1$ to N removed for readiness):

$$\begin{aligned}
\partial^2_\beta \Lambda = \cdots = &\quad \sum \frac{z_{jk}^2}{\beta^2} \ln R_{jk}(t_{jk}) U(\delta_{j,k}) - \frac{\delta_{j,k}}{\beta^2} U(\delta_{j,k}) \\
\partial^2_{\eta_r} \Lambda = \cdots = &-\sum \frac{\beta^2}{\eta_r^2} |\delta_{j,k}| \\
\partial^2_{a_V} \Lambda = \cdots = &\quad \sum \beta^2 (\ln V_k - \ln V_r)^2 \ln R_{jk}(t_{jk}) U(\delta_{j,k}) \\
\partial^2_{a_A} \Lambda = \cdots = &\quad \sum \beta^2 (\ln A_k - \ln A_r)^2 \ln R_{jk}(t_{jk}) U(\delta_{j,k}) \\
\partial_{\beta,a_V} \Lambda = \cdots = &-\sum (\ln V_k - \ln V_r) z_{jk} \ln R_{jk}(t_{jk}) U(\delta_{j,k}) \\
\partial_{\beta,a_A} \Lambda = \cdots = &-\sum (\ln A_k - \ln A_r) z_{jk} \ln R_{jk}(t_{jk}) U(\delta_{j,k}) \\
\partial_{\beta,\eta_r} \Lambda = \cdots = &-\sum \frac{z_{jk}}{\eta_r} \ln R_{jk}(t_{jk}) U(\delta_{j,k}) \\
\partial_{a_A,a_V} \Lambda = \cdots = &\quad \sum \beta^2 (\ln V_k - \ln V_r)(\ln A_k - \ln A_r) \ln R_{jk}(t_{jk}) U(\delta_{j,k})
\end{aligned}$$

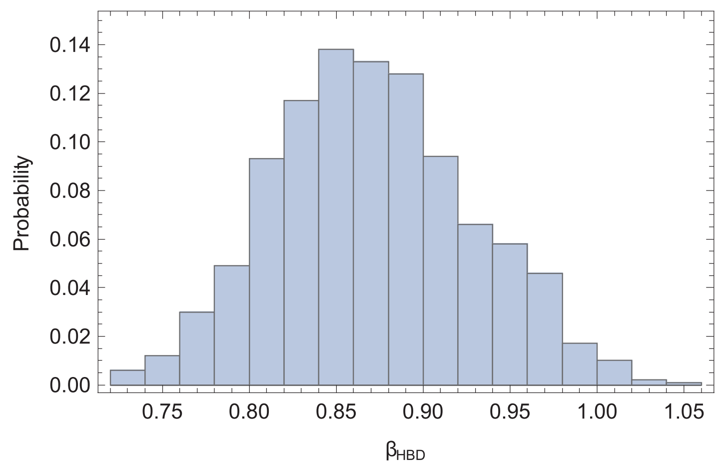


Fig. 5.28 Probability plot of the shape parameter of the Weibull distribution for $N_{\mathrm{DUTs}} = 245$ and #Draws $= 1000$. The distribution of the parameter's values is Normal after bootstrapping from the known distribution.

5.2.4.2 Influence of measurement time and right-censorship on the error estimation for the MLE

The error estimation for different right-censor times was done by bootstrapping[237] known TDDB datasets. First, data sets of different cardinality (varying number of DUTs) were drawn from the known population of pMOS HBD data (HBD trigger 100 nA, $T = 125°\mathrm{C}$). Then the maximum likelihood estimator (MLE) was calculated depending on cardinality of the data sets. This algorithm was repeated for each number of DUTs (N_{DUTs}) each 1000 times, whereby at maximum 248 devices were used. In the end, one ends up with Normally distributed MLE parameters for each N_{DUTs} (Figure 5.28).

Four experimental voltages $\{-3.3\,\mathrm{V}; -3.4\,\mathrm{V}; -3.5\,\mathrm{V}; -3.6\,\mathrm{V}\}$ were used and a reference voltage $V_r = -3.49\,\mathrm{V}$ was calculated. The reference area was not varied and amounts to $4 \cdot 10^{-9}\,\mathrm{cm}^2$. The draws from each group started at $N_{\mathrm{DUTs}} = 4$ and were increased by unity in each voltage group. Thereby one ends up with draws $N_{\mathrm{DUTs}} = 4 \times 4,\ 5 \times 4 \ldots$ for which the MLE is calculated. Each voltage group is fitted with the same shape parameter β. However, the bootstrap approach is rather computationally demanding; drawing 1000 times the Weibull distributed data sets for $N_{\mathrm{DUTs}} = 16 \ldots 248$ lasted up to 55 000 s on a modern four core CPU. The advantage of this algorithmic approach is the possibility to parallelize the draws and fits using a random number generator stable towards multi-core processing. By using

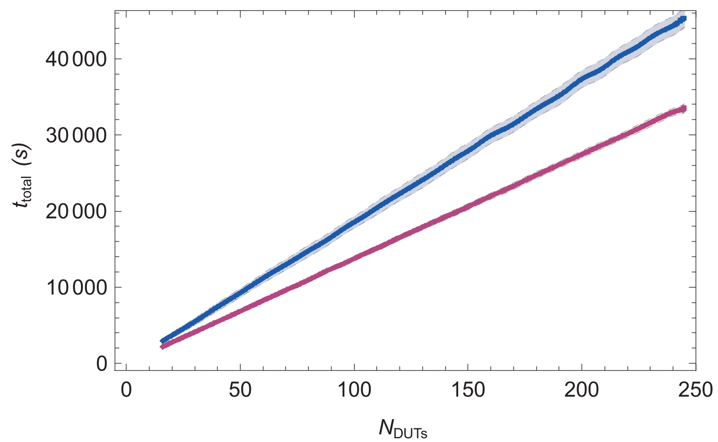


Fig. 5.29 Measurement time vs. number of DUTs for $t_{rc} = 300$ s (red) and 10 000 s (blue). At a fixed number of DUTs the measurement time can be reduced by lowering the maximum measurement time. With possible 248 DUTs more than 10 000 s measurement time is less used with lower timeout.

more than one computation kernel, the computation time reduced almost linearly since no communication between the single threads (i.e. draws) is necessary.

The bootstrapping of the Weibull distributed data sets was done for known populations with right-censorship at $t_{rc} = 300$ s and $t_{rc} = 10000$ s. Bootstrapping the known population approximates the real distribution parameters according to the law of giant numbers.

With the bootstrap results for $N_{DUTs} = 16 \ldots 245$ the Weibull parameters and their error bars were calculated subsequently for each N_{DUTs}. Figure 5.29 shows the total measurement time t_{total} used for the two different t_{rc}. The aim is to reduce the measurement time without detrimental effects on the extraction of the Weibull parameters.

The parameters' median and standard deviation were then plotted vs. total number of DUTs N_{DUTs}. The resulting evolution of the parameters' values vs. N_{DUTs} are shown in Figure 5.30 for two different right-censor times $t_{rc} = 300$ s and $t_{rc} = 10\,000$ s. It is clearly visible that for a low number of samples the parameters are quite off from their real values. After ~ 100 samples both β_{HBD} and η_r are approaching the real values of the distribution as found also in the experiment. However, it is also clear that η_r remains much lower for lower right-censorship. This offset is compensated for lower t_{rc} by higher voltage acceleration (exponent) of η_r.

The error bars for the Weibull parameters were accordingly extracted and analyzed. The measurement error for each Weibull parameter was obtained from the Fisher Information Matrix (FIM, in analytics more familiar as the Hessian). The parameter errors can be estimated from the inverse of this matrix, wherein the diagonal elements describe the corresponding variance. The off-diagonal elements mark the covariance between the distribution parameters. Here the logarithmic form of the error bar was chosen to emphasize the relative error in the asymptotic limit. This form can be obtained by a variable transformation in the derivatives.

The resulting error bars for $\Delta \ln \eta_r$, $\Delta \ln \beta$, Δa_V are shown in Figure 5.31. Again higher $\Delta \ln \eta_r$ for lower timeouts is compensated by lower error in the voltage acceleration Δa_V. In addition, a quite important number can be extracted: The error for $\ln \beta$ asymptotically approaches 0.06, i.e. a minimum error of 7% was found for this (bootstrap) experiment.

For applications where ramped voltage test are not sufficient for quality control or process qualification, the use of a significant lower t_{rc} value can be most beneficial since there the average Weibull parameters are commonly known. Then a statement for the analyzed wafers can be drawn with maximum timesaving. For measurements in a research environment with varying gate stacks the maximum benefit is difficult to obtain since stress conditions cannot be known beforehand. Then the benefit of lower t_{rc} reduces to the avoidance of measuring devices until timeout, most likely for the lowest stress conditions.

A possible way out of this dilemma is the measurement of a basic (e.g. 10 samples) TDDB experiment. Then an additional experiment to stabilize the Weibull distribution can be carried out using a lower timeout i.e. right censor. As a guiding principle, Figure 5.32 shows the error estimates for $\Delta \ln \eta_{\mathrm{r}}$ for right-censors $t_{\mathrm{rc}} = 300 \ldots 1000\,\mathrm{s}$ and $N_{\mathrm{DUTs}} = 16 \ldots 250$. The error estimate looks quite constant with increasing right-censor. The error reduces with the number of samples but little with a change in the timeout.

Since SBD and HBD parameters β and η_r are decreasing with scaled gate stacks, the gain of measuring long-time experiments seems to be low based on the bootstrap results (for EOT $= 9\,\text{Å}$).

5.2.5 Summary

The TDDB measurements on small area bulk FinFETs were shown. Below 1 nm equivalent oxide thickness the nMOS FinFETs show hard-breakdown voltage for 10 years lifetime below 1 V. The pMOS devices showed better lifetime for hard-breakdown. Soft-breakdown starts also at a much lower operating voltage than 1 V for both *n*/pMOS and leads to leakage current increase. The wearout acceleration is smaller for pMOS leading to a longer wearout phase than for nMOS.

Therefore, the current increase over lifetime has to be characterized and included for a reliability assessment of the devices. A convenient way to present both, breakdown and leakage current increase vs. lifetime in a single plot, is the all-in-one methodology. This requires an accurate knowledge of the soft-breakdown data. Inaccurate soft-breakdown extraction was shown to lead to erroneous predictions of the wearout phase and consequently to an underestimation of the leakage current increase.

However, the extraction of soft-breakdowns at low voltage and small area devices is subject to certain process variability at the nanoscale. The initial leakage current variations can be assessed by a constant current trigger for soft-breakdown extraction. Nevertheless, these variations can hamper the definition of suitable stop conditions for the constant voltage stress. Increased experiment durations are resulting because most devices finish with the predefined experimental timeout. Thus, the role of timeout and initial failure (left- and right censorship) on the error of the Weibull distribution was assessed by a bootstrap experiment. The simulation suggests the use of more devices at lower timeout. At comparable error rates, the significance of the TDDB experiment may be kept at much decreased total measurement time.

In summary, the EOT scaling below 1 nm leads to decreased TDDB reliability. Especially for nMOS devices the current gate stacks can likely not reach a 10 years lifetime and responsive circuit design should address the reduced lifetime as Groeseneken suggested.[240]

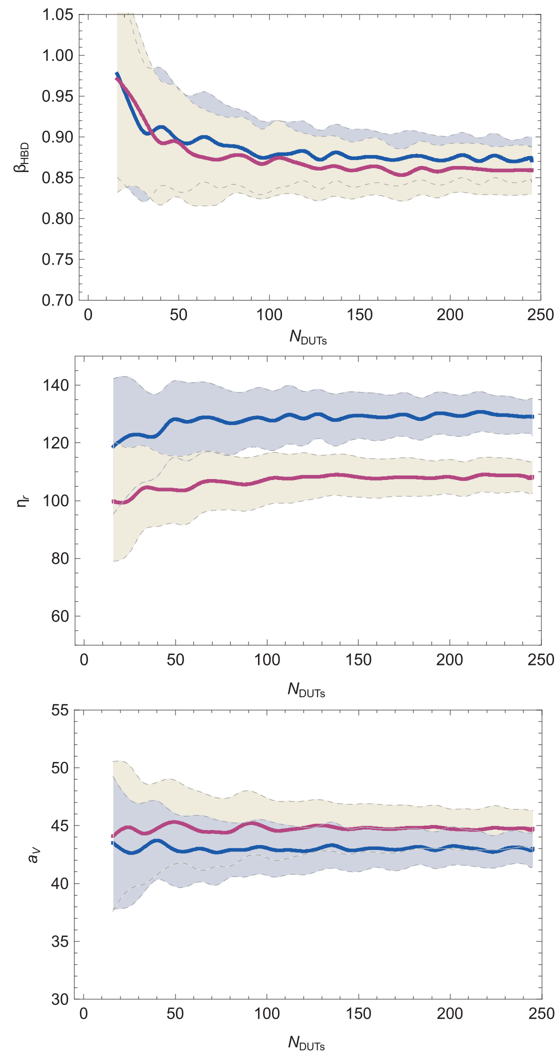


Fig. 5.30 Bootstrap results for β_{HBD} and η_r and their corresponding standard deviation (shaded region with dashed boundaries) for two different right-censor times $t_{\mathrm{rc}} = 300\,\mathrm{s}$ (violet) and $t_{\mathrm{rc}} = 10\,000\,\mathrm{s}$ (blue). The shape parameter β_{HBD} approaches the values for longer t_{rc} for ~ 100 samples. In parallel the 63% value η_r is here $\sim 20\%$ lower for lower t_{rc} and will not approach the values for higher t_{rc}. This offset is compensated by higher voltage acceleration a_V for lower t_{rc}.

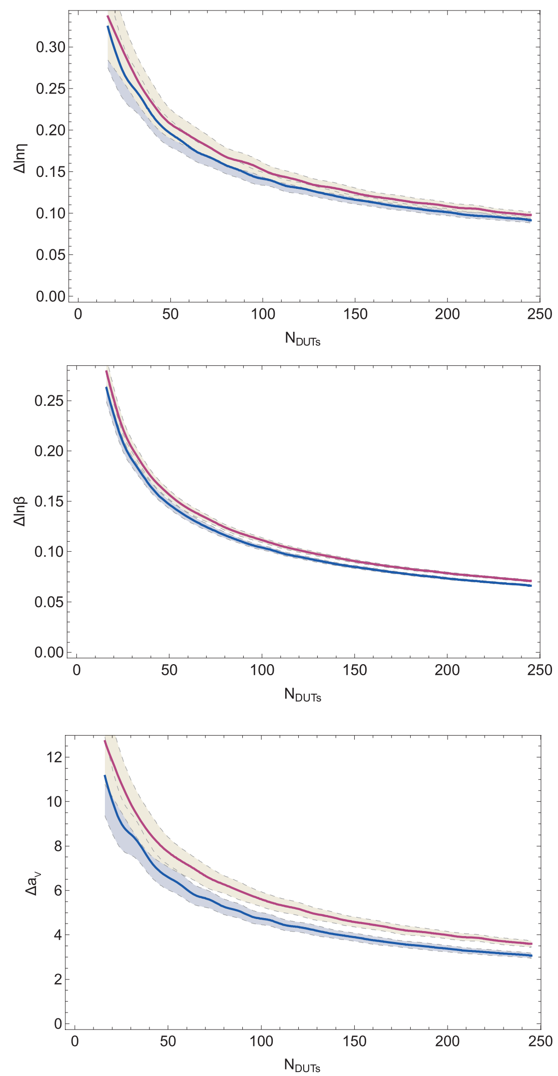


Fig. 5.31 Parameter error vs. N_{DUTs} for the two different right-censor times $t_{\mathrm{rc}} = 300\,\mathrm{s}$ (violet) and $t_{\mathrm{rc}} = 10\,000\,\mathrm{s}$ (blue). The errors are reduced rapidly with increasing N_{DUTs}. A higher error bar $\Delta \ln \eta_r$ is compensated by a lower Δa_V.

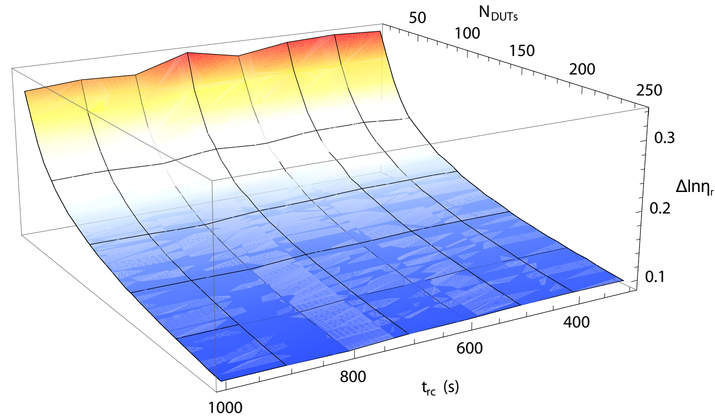


Fig. 5.32 Parameter error $\Delta \ln \eta_r$ vs. N_{DUTs} for the right-censor times $t_{\mathrm{rc}} = 300 \ldots 1000\,\mathrm{s}$. Extending the right-censor beyond 300 s does not significantly change the error estimate for the bootstrapped data set. Increasing the number of DUTs N_{DUTs} greatly reduces $\Delta \ln \eta r$.

Chapter 6
Conclusion

The continued downscaling of MOSFET dimensions requires an equivalent oxide thickness (EOT) below 1 nm. An EOT below 1.4 nm is enabled by the use of high-κ/metal gate stacks. In this study, $LaLuO_3$ and HfO_2 were investigated as two different high-κ oxides in conjunction with TiN as the metal electrode.

For $LaLuO_3$, the material characterization by hard X-ray photoemission spectroscopy was successfully employed to show the temperature dependent formation of La-rich and Si-rich silicate phases at the interface to Si. The HAXPES measurements allowed the study of a complete gate stack with metal electrode in place. At moderate temperatures or short high temperature annealing, a La-rich phase was found changing to Si-rich silicate for high temperature and long annealing. The silicate formation at high temperature was found to consume even a 10 nm thick film. For short annealing (5 s) the defect trapping and hysteresis in electrical measurements was negligible, while even 600°C for long time led to increased defects in $LaLuO_3$. The effective attenuation length of $LaLuO_3$ was determined between 7 and 13 keV to be employed in future experiments for interface and diffusion studies. The investigation of $LaLuO_3$ on Ge by HAXPES showed germanate formation at the interface in the absence of GeO_x, which could be a promising alternative to the problematic Ge oxide interface passivation.

The successful integration of $LaLuO_3$ in a high-temperature MOSFET process flow was shown. Here, additional annealing steps at 600°C (poly-Si deposition) were found to cause an equivalent oxide thickness increase and large trap density. The material characterization could explain these observations with the consumption of high-κ film in favor of La-rich silicate formation. Oxygen anneals were shown to reduce defects in the high-κ film, but could also increase the equivalent oxide thickness. The obtained mobility for MOSFETs with La-rich interfacial layer is below the result for Si-rich interface in a comparing study and in agreement with theoretical predictions. A further optimization should focus on thinner Si-rich interlayers to retain high effective device mobility.

The integration of $LaLuO_3$ in high-T processes was found to be delicately connected with the optimization of the TiN metal electrode. An optimization to stoichiometric TiN yields the best results in terms of thermal stability with respect to

Si-capping and high-κ oxide. Stress measurements revealed compressive and tensile stress in Si for PVD-deposited TiN and ALD TiN, respectively. TiN was characterized to be an effective diffusion barrier for Boron dopants from a poly-Si stack and depending on thickness as for most metallization used in CMOS processing.

A further EOT reduction was successfully shown with different approaches on $LaLuO_3$ and HfO_2. Thereby the thermodynamic and kinetic predictions were employed to estimate the behavior on the nanoscale. Based on thermodynamics, excess oxygen in the gate stack, especially in oxidized metal electrodes, was identified to prevent EOT scaling below 1.2 nm. The equivalent oxide thickness of HfO_2 gate stacks was shown to be scalable below 1 nm by the use thinned interfacial SiO_2. The limitation of oxygen in metal electrodes by Si-capping was found to keep the EOT after high temperature annealing. Literature reports even a reduction of EOT based on high temperature anneals. The employment of redox systems within the gate electrode was successfully used to decrease the EOT of HfO_2 gate stacks. A lower limit found was EOT $= 4.6$ Å for Al doping. The doping of TiN on $LaLuO_3$ was found in microscopy studies to modify the interfacial silicate layer or even reduce the layer. Future work should investigate the interface modifications in order to control the Si-/La-rich silicate phases at reduced EOT.

The oxide quality in high-κ/metal gate stacks was investigated by charge pumping and carrier mobility measurements. The oxide quality on top- and sidewall of FinFETs was determined by the method of Kapila *et al.* In agreement with literature, the mobility was degraded lower EOT was reached. The best oxide quality and mobility was obtained after a high temperature anneal. A much lower mobility was obtained if no post-deposition anneal for the high-κ layer was employed. The integration of such an annealing resulted in a similar mobility as in the high-T anneal for the complete gate stack. Precise EOT scaling has to be performed concerning mobility reduction since higher polarized dielectrics are closer to the channel.

The oxide reliability of gate stacks below 1 nm EOT was determined by TDDB measurements on short-channel FinFETs with HfO_2/TiN gate stacks. EOT scaling was shown to lead to comparable onset of soft-breakdown (compared 8 and 9 ÅEOT) but stronger voltage acceleration and shorter wearout of the weak spots. For nMOS also hard-breakdown close to the targeted 10 years' operating voltage were found. For both *n*-and pMOS, the onset of soft-breakdown was found to lie well below 1 V in agreement with the scaling trend in planar devices resulting in leakage current increase over device lifetime. The all-in-one methodology was employed to characterize both, soft- and hard-breakdown, in a common plot. The optimization of measurement time is enabled by the use of shorter right-censors and the use of the lower tail in the Weibull distribution. The error bars on the extracted results remain comparable.

A successful EOT scaling has always to consider the oxide quality and the resulting reliability. Degraded oxide quality leads to mobility degradation and earlier soft-breakdown, i.e. leakage current increase over lifetime. Post deposition anneals can improve the oxide quality after thin film deposition at comparable EOT. The shown EOT scaling mechanisms have to be further evaluated in respect to their reliability.

Appendix A
MIPS Fabrication Scheme for $LaLuO_3/TiN$ Gate Stacks

- **Mesa patterning** by optical lithography and RIE. The mesa for the MOSFET operation is patterned in a positive process and structured by RIE to isolate the device area.
- **Si_3N_4 deposition** (LPCVD/PECVD). Deposition of a 50 nm thick Si_3N_4 hard mask by LPCVD or PECVD for lateral mesa oxidation.
- Patterning of the hard mask, CF_4 process (RIE)
- **Sidewall oxidation** of the mesa (vertical furnace). Wet oxidation delivers a high quality SiO_2 on the mesa edges to prevent leakage paths due to non-conformal high-κ deposition by MBD.
- Removal of nitride hard mask by H_3PO_4 (wet etch)
- **RCA cleaning** as Si cleaning and passivation (wet etch). ~ 1 nm SiO_2 is left on mesa after RCA cleaning.
- **High-κ deposition** (MBD/ALD). $LaLuO_3$ is deposited by MBD, HfO_2 by ALD.
- (Oxygen anneal in RTP).
  (Optional) passivation anneal to annihilate defects in $LaLuO_3$ in 100% O_2 at $400°C/10$ min.
- 15 nm TiN deposition (PVD).
- 100 nm $n++$ poly-Si (LPCVD).
- **Gate patterning** (RIE). Optical Lithography with UV6.06 photo resist with 1 μm minimum line feature size, subsequent gate patterning in ICP-RIE, a Cl_2/Ar process was established to etch anisotropically poly-Si/TiN.
- **High-κ removal** on S/D area (wet etch). 180 s HF 1% for HfO_2, 120 s HCl (buffered with NaOH to pH = 3) for $LaLuO_3$.
- **Ion implantation**. Dose $1 \cdot 10^{15}/cm^2$, 7° tilt, 0° twist, 1 Q, 6 keV. As (n-type)/3 keV BF_2^+ (p-type).
  The high implantation dose in conjunction with thin SOI allows fully-depleted device operation.
- **Passivation layer** (PECVD). 30 nm Si_3N_4 (PECVD) for surface passivation during annealing.
- **RTA** 1000°**C, 5 s** in Ar under O_2 control (< 1 ppm). After doping activation the active S/D area has a resistivity of $\sim 200\,\mu\Omega$cm.

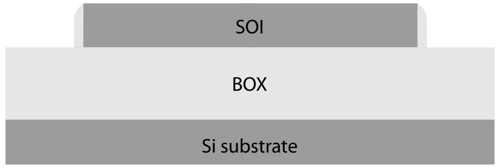


Fig. A.1 Isolated mesa after removal of the nitride hard mask.

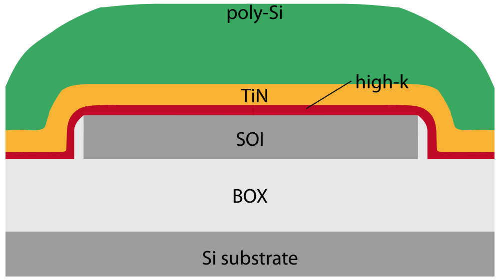


Fig. A.2 Isolated mesa after deposition of the gate stack.

- **Spacer formation** (RIE). Etching of 30 nm Si_3N_4 by Ar/CHF_3 RIE, leaving side-wall spacers at the gate.
- **Ni deposition** (PVD). 30 s HF 1% dip and PVD deposition of 5 nm Ni.

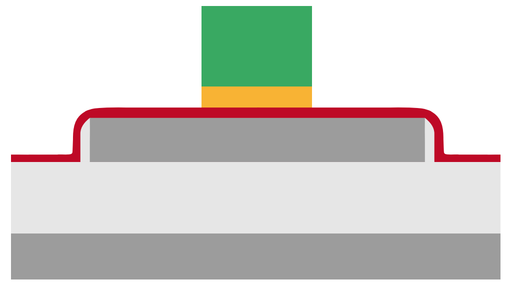


Fig. A.3 Gate patterning by RIE stops on the high-κ layer, which is subsequently removed by wet etching.

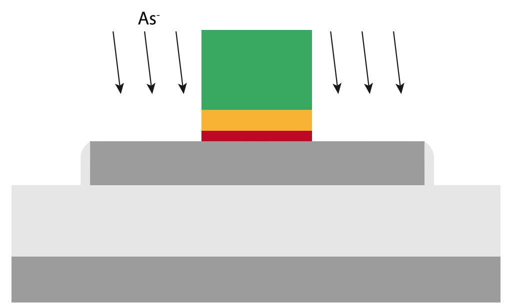


Fig. A.4 After complete gate patterning, implantation is carried out into SOI.

- **NiSi formation** (RTP). 500°C/30 s in forming gas. The non-reacted Ni was selectively etched by H_2SO_4 for 1 min and a 30 s HF 1% dip. The NiSi formation on gate and active area lead to a resistivity $\sim 20\,\mu\Omega$cm for 15 nm SOI, i.e. a resistivity reduction of $x10$ compared to simply implanted devices. Optional anneals for the high-κ layer are done before silicide formation. The silicidation after implantation and activation leads to the drive-out of dopants to the interface NiSi – channel area (dopant segregation by "snow plow effect"). The resulting structure has dopants segregated directly at the channel in contact with ohmic NiSi contacts.[241]
- **Passivation layer** (PECVD). 50 nm SiO_2 for surface passivation. With Si_3N_4 spacers in place an additional, short NH_4OH/H_2O_2 clean is possible to remove sulfur contaminants on the surface. Without spacers, TiN would be removed here.

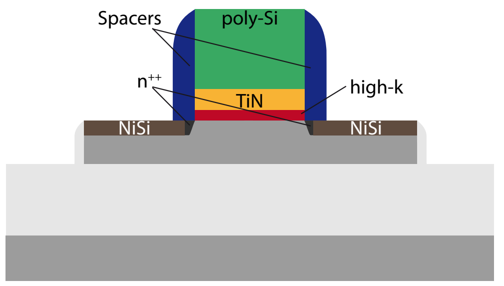


Fig. A.5 Long channel device high-κ/metal gate stack and NiSi S/D contacts after front-end-of-line processing (FEOL) on SOI. After FEOL processing a final passivation layer follows and contact holes are etched through the passivation onto active S/D and gate area (*n.b.*: dimensions not to scale).

- **Contact hole formation** (RIE). Optical lithography for the contact etch to the devices.
- **Metallization** (PVD). 30 s HF 1% and deposition of 5 nm Ti/200 nm Al.
- **Final FGA**, 400°C/10 min.

Appendix B
Etch-Back Process for SiO_2

In this work, the use of an etch-back process was investigated. The thick oxide can be formed by an ozone concentration of 55 ml/l DI water, whereby the water is spun on the rotating wafer in a single wafer tool and an immersion of water/ozone mixture oxidizes the wafer. The oxidation is self-limiting[104, 151] depending on the ozone concentration (55 ml/l) and yields to 16 Å SiO_2 after 120 s oxidation. The process flow in detail consists of

- 45 s ozone oxidation to reduce organic particles
- Rinse with DI water
- complete etch back with HF (100 ml 50% HF in 18 l DI water buffer)
- Rinse in DI water/HCl[152]
- 120 s ozone oxidation at 55 mg/l ozone
- Rinse with DI water
- *x* seconds etch back in (Spray process, 55 ml 50% HF in 18 l DI water buffer)
- Rinse with DI water
- spin dry in N_2

The HF concentration for etch back was determined from the change hydrophilic to hydrophobic surface around 60 s for 55 ml (0.3%). The etch rate for 1% HF was determined to be 6 nm/min and suggests a linear etch rate of $\sim 1.8\,\mathrm{nm}/\,\mathrm{min}$ for 0.3% as the etching depends approximately linear on the HF solution.[153] This leaves a target process window of 1 min to remove all SiO_2 if assuming a linear etch rate down to zero SiO_2.

Appendix C
Publications

Published articles and conference contributions

1. A. Nichau, M. Schnee, A. Besmehn, U. Breuer, J. Rubio-Zuazo, G. R. Castro, P. Bernardy, B. Holländer, J. Schubert and S. Mantl, J. Chem. Phys. **138**, 154709 (2013).
2. L. Knoll, Q. Zhao, A. Nichau, S. Trellenkamp, S. Richter, D. Esseni, L. Selmi, K.K. Bourdelle and S. Mantl, IEEE Electronic Device Letters **34**, 813 – 815 (2013).
3. S. Mantl, L. Knoll, M. Schmidt, S. Richter, A. Nichau, S. Trellenkamp, A. Schafer, S. Wirths, S. Blaeser, D. Buca, Q.-T. Zhao, in *Proceedings of the 14th International Conference on Ultimate Integration on Silicon (ULIS)*, 15–20 (2013).
4. L. Knoll, S. Richter, A. Nichau, A. Schaefer, K.K. Bourdelle, Q.-T. Zhao and S. Mantl, *Proceedings of the 14th International Conference on Ultimate Integration on Silicon (ULIS)*, 97–100 (2013).
5. A. Nichau, A. Schäfer, L. Knoll, S. Wirths, T. Schram, L.-Å. Ragnarsson, J. Schubert, P. Bernardy, M. Luysberg, A. Besmehn, U. Breuer, D. Buca and S. Mantl, Microelectronics Engineering **109**, 109–112 (2013).
6. L. Knoll, S. Richter, A. Nichau, S. Trellenkamp, A. Schäfer, K.K. Bourdelle, Q.-T. Zhao and S. Mantl, in *Proceedings of 9th Workshop of the Thematic Network on Silicon-on-Insulator Technology (EUROSOI)* (2013).
7. A. Nichau, J. Rubio-Zuazo, M. Schnee, G.R. Castro, J. Schubert and S. Mantl, Applied Physics Letters **102**, 031607 (2013).
8. A. Nichau, E. Durgun Özben, M. Schnee, J.M.J. Lopes, A. Besmehn, M. Luysberg, L. Knoll, S. Habicht, V. Mussmann, R. Luptak, S. Lenk, J. Rubio-Zuazo, G.R. Castro, D. Buca, Q.-T. Zhao, J. Schubert and S. Mantl, Solid-State Electronics **71**, 19–24 (2012).
9. A. Nichau, J. Schubert, J. Rubio-Zuazo, A. Besmehn, M. Schnee, A. Schäfer, G.R. Castro and S. Mantl, in *SpLine Users' Meeting, Madrid, April 16/17th 2012* (2012).
10. A. Nichau, J. Schubert, J. Rubio-Zuazo, A. Besmehn, M. Schnee, A. Schäfer, G.R. Castro and S. Mantl, in Wilhelm and Else Heraeus Physics School: Mi-

croelectronics for Society: “More Than Moore” Expands “More Moore”, Bad Honnef/Germany, June 10th-16th, 2012 (2012).

11. A. Schäfer, A. Tiedemann, A. Nichau, R. Luptak, V. Mussmann, J. Schubert and S. Mantl, in Wilhelm and Else Heraeus Physics School: Microelectronics for Society: “More Than Moore” Expands “More Moore”, Bad Honnef/Germany, June 10th–16th, 2012 (2012).
12. S. Richter, C. Sandow, A. Nichau, S. Trellenkamp, M. Schmidt, R. Luptak, K.K. Bourdelle, Q.T. Zhao and S. Mantl, IEEE Electron Device Letters **33**, 1535–1537 (2012).
13. E. Durgun Özben, M. Schnee, A. Nichau, V. Mussmann, R. Luptak, J.M.J. Lopes, S. Lenk, K.K. Bourdelle, Q.-T. Zhao, J. Schubert and S. Mantl, Microelectronic Engineering **88**, 1323–1325 (2011).
14. W. Yu, B. Zhang, Q.-T. Zhao, J.-M. Hartmann, D. Buca, A. Nichau, R. Luptak, J.M.J. Lopes, S. Lenk, M. Luysberg, K.K. Bourdelle, X. Wang and S. Mantl, Solid-State Electronics **62**, 185–188 (2011).
15. E. Durgun Özben, J.M.J. Lopes, A. Nichau, R. Luptak, S. Lenk, A. Besmehn, K.K. Bourdelle, Q.-T. Zhao, J. Schubert and S. Mantl, Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures **29**, 01A903 (2011).
16. E. Durgun Özben, J.M.J. Lopes, A. Nichau, M. Schnee, S. Lenk, A. Besmehn, K.K. Bourdelle, Q.-T. Zhao, J. Schubert and S. Mantl, IEEE Electron Device Letters **32**, 15–17 (2011).
17. A. Nichau, E.D. Özben, M. Schnee, J.M.J. Lopes, A. Besmehn, M. Luysberg, L. Knoll, S. Habicht, V. Mussmann, R. Luptak, S. Lenk, J. Rubio-Zuazo, G.R. Castro, D. Buca, Q.T. Zhao, J. Schubert and S. Mantl, in *Ulis 2011 Ultimate Integration on Silicon* (IEEE, 2011), pp. 1–4.
18. E. Durgun Özben, J.M.J. Lopes, A. Nichau, R. Luptak, S. Lenk, A. Besmehn, K.K. Bourdelle, Q.-T. Zhao, J. Schubert and S. Mantl, IEEE Transactions on Electron Devices **58**, 617–622 (2011).
19. J.M. Lopes, E. Durgun Özben, M. Schnee, R. Luptak, A. Nichau, A. Tiedemann, W. Yu, Q.-T. Zhao, A. Besmehn, U. Breuer, M. Luysberg, S. Lenk, J. Schubert and S. Mantl, ECS Transactions **35**, 461–479 (2011).
20. Q.-T. Zhao, W. Yu, B. Zhang, J.M.J. Lopes, D. Buca, R.A. Minamisawa, E. Durgun Özben, R. Luptak, S.F. Feste, A. Nichau, J. Schubert, B. Holländer, J.M. Hartmann, K.K. Bourdelle, N. Kernevez and S. Mantl, in *5th International SiGe Technology and Device Meeting (ISTDM 2010) Stockholm, Schweden: 24.05.2010–26.05.2010, 2010.* (2010).
21. E. Durgun Özben, A. Nichau, J.M.J. Lopes, S. Lenk, A. Besmehn, K.K. Bourdelle, Q.-T. Zhao, J. Schubert and S. Mantl, in *ECS Transactions* (2010), pp. 195–202.
22. W. Yu, E. Durgun Özben, B. Zhang, A. Nichau, J.M.J. Lopes, R. Luptak, S. Lenk, J.-M. Hartmann, D. Buca, K.K. Bourdelle, J. Schubert, Q.-T. Zhao and S. Mantl, in *2010 10th IEEE International Conference on Solid-State and Integrated Circuit Technology* (IEEE, 2010), pp. 875–878.

23. W. Yu, B. Zhang, Q.-T. Zhao, J.-M. Hartmann, D. Buca, A. Nichau, E. Durgun Özben, J.M.J. Lopes, J. Schubert, B. Ghyselen and S. Mantl, in *Proceedings of 6th Workshop of the Thematic Network on Silicon-on-Insulator Technology (EUROSOI)* (2010), pp. 25–26.
24. C. Radtke, C. Krug, G. V. Soares, I.J.R. Baumvol, J.M.J. Lopes, E. Durgun Özben, A. Nichau, J. Schubert and S. Mantl, Electrochemical and Solid-State Letters **13**, G37 (2010).
25. E. Durgun Özben, J.M.J. Lopes, M. Roeckerath, A. Nichau, R. Luptak, S. Lenk, A. Besmehn, B. Ghyselen, Q.-T. Zhao, J. Schubert and S. Mantl, in *Proceedings of ULIS (Ultimate Integration on Silicon) 17–19 March 2010, Glasgow, Scotland.* (2010), pp. 93–96.
26. R. Luptak, E. Durgun Özben, A. Nichau, J.M.J. Lopes, K.K. Bourdelle, S. Lenk, Q.-T. Zhao, J. Schubert and S. Mantl, JaraFIT Nanoelectronics Days 2010 (2010).
27. J.M.J. Lopes, E. Durgun Özben, A. Nichau, R. Luptak, J. Schubert and S. Mantl, in *Spring Meeting of the German Physical Society, Regensburg: 23.03.2010, 2010.* (2010).
28. E. Durgun Özben, A. Nichau, J.M.J. Lopes, M. Roeckerath, J. Schubert and S. Mantl, in Wilhelm and Else Heraeus Physics School: Nanoscaled Oxides – Big Opportunities in Small Structures, Bad Honnef/Germany, June 2nd-8th, 2009 (2009).
29. J. Schubert, J.M.J. Lopes, E. Durgun Özben, A. Nichau, G. Mussler, M. Roeckerath, S. Mantl, A. Besmehn, U. Breuer and M. Luysberg, in *16th International Workshop on Oxide Electronics Tarragona, Spain: 04.10.2009-07.10.2009, 2009.* (2009).
30. J.M.J. Lopes, M. Roeckerath, E. Durgun Özben, A. Nichau, G. Mussler, C. Radtke, G.G. Marmitt, C. Krug, G.V. Soares, I.J.R. Baumvol, J. Schubert and S. Mantl, in *40th IEEE SISC Arlington, USA: 02.12.2009–05.12.2009, 2009.* (2009).
31. J.M.J. Lopes, E. Durgun Özben, A. Nichau, M. Roeckerath, G. Mussler, U. Breuer, A. Besmehn, J. Schubert and S. Mantl, in *MRS-Fall Meeting, Boston, USA: 30.11.2009–02.12.2009, 2009.* (2009).
32. A. Nichau, Fabrication of AlN Barriers for Superconducting Tunnel Elements Using an Inductively Coupled Plasma, Diploma Thesis, University of Cologne, 2009.

Submitted articles

1. A. Nichau, P. Roussel, T. Kauerauf, R. Degraeve, L.-Å. Ragnarsson, G. Groeseneken and S. Mantl, Submitted to IEEE Transactions on Device and Materials Reliability (2012).
2. A. Nichau, T. Kauerauf, P.J. Roussel, L.-Å. Ragnarsson, T. Chiarella, M. Togo, G. Groeseneken and S. Mantl, Submitted to IEEE Transactions on Device and Materials Reliability (2012).

Accepted conference contribution

1. L. Knoll, Q.T. Zhao, A. Nichau, S. Richter, G.V. Luong, S. Trellenkamp, A. Schfer, L. Selmi, K.K. Bourdelle, S. Mantl, Demonstration of Improved Transient Response of Inverters with Steep Slope Switching strained Si NW C-TFETs, by reduction of TAT in pulsed I-V and Scaling, Accepted for presentation at the International Electron Devices Meeting (IEDM) 2013, Washington D.C., December 9-11 2013.

Appendix D
Glossary

ADF	Angular Dark-Field imaging by STEM.
ALD	Atomic Layer Deposition, today's most common deposition method for nano-scaled structures due to its homogenous, chemically achieved surface coverage. The material is deposited by precursor deposition on the surface and subsequent removal of organic ligands.
AVD	Atomic Vapor Deposition, a patented form of ALD with pulsed injection of precursor during deposition.
CET	Capacitance Equivalent (SiO_2) Thickness, see section 2.1.3.
CMP	chemical mechanical polishing; used for dielectrics and metal layers to planarize the wafer topology.
CVD	Chemical Vapor Deposition.
EELS	Electron Energy Loss Spectroscopy.
EOT	Equivalent Oxide Thickness, see section 2.1.3.
FGA	Forming Gas Annealing in a mixture of max.10% H_2 and N_2.
HAADF	High-Angle Angular Dark-Field imaging from STEM.
HBD	hard breakdown in TDDB measurements: At HBD the gate oxide loses its isolating properties and the leakage current raises above a specified limit.
ILD	Inter-layer dielectric, a dielectric passivation between the different layers of the chip structure.
LPCVD	Low-Pressure Chemical Vapor Deposition.
MBD	Molecular Beam Deposition.
MOS	Metal-Oxide-Semiconductor (gate stack).
MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor.

PDA Post Deposition Anneal after deposition of high-κ material to improve material properties, e.g. to reduce oxygen vacancies.
PECVD Plasma Enhanced Chemical Vapor Deposition.
PLD Pulsed Laser Deposition.
PVD Physical Vapor Deposition, here: sputtering.
RIE Reactive Ion Etching.
RTA Rapid Thermal Annealing.
RTP Rapid Thermal Processing (tool).
SBD Soft BreakDown in TDDB measurements; the creation of additional leakage by defect generation in the gate oxide.
STEM Scanning Tunneling Electron Microscope.
V_{fb} Flatband Voltage of a MOS gate stack i.e. capacitor.
V_{th} Threshold voltage of a MOSFET.
WO Wearout in TDDB; The creation of leakage paths is followed by the wearout of these weak spots. WO is indicated by a "digital" leakage current change (on/off behavior) during measurements until HBD.
μ_{eff} Effective (device) mobility in MOSFETs; the mobility is calculated by the ratio of drain current to inversion charge.

References

1. Nobelprize.org The Nobel Prize in Physics 1956. (1956).at <http://www.nobelprize.org/nobel_prizes/physics/laureates/1956/>
2. Kahng and Atalla: Bell Telephone Lab. Tech. Memorandum MM-61-2821-1, Jan 16, 1961. *Semiconductor Devices: Pioneering Papers* (1991).
3. Moore, G. E. Cramming more compontens onto integrated circuits. *Electronics* **38**, 1–4 (1965).
4. Wilk, G. D., Wallace, R. M. & Anthony, J. M. High-κ gate dielectrics: Current status and materials properties considerations. *Journal of Applied Physics* **89**, 5243 (2001).
5. Economist Plugging the leaks. *The Economist Newspaper Limited* (2011).
6. Dennard, R. H., Gaensslen, F. H., Rideout, V. L., Bassous, E. & LeBlanc, A. R. Design of ion-implanted MOSFET's with very small physical dimensions. *IEEE Journal of Solid-State Circuits* **9**, 256–268 (1974).
7. Baccarani, G., Wordeman, M. R. & Dennard, R. H. Generalized scaling theory and its application to a 1/4 micrometer MOSFET design. *IEEE Transactions on Electron Devices* **31**, 452–462 (1984).
8. Dennard, B. Discussing DRAM and CMOS Scaling with Inventor Bob Dennard. *IEEE Design & Test of Computers* **25**, 188–191 (2008).
9. Muller, D., Sorsch, T. & Moccio, S. The electronic structure at the atomic scale of ultrathin gate oxides. *Nature* **399**, 758–761 (1999).
10. Green, M. L. *et al.* Understanding the limits of ultrathin SiO_2 and SiON gate dielectrics for sub-50 nm CMOS. *Microelectronic Engineering* **48**, 25–30 (1999).
11. Buchanan, D. Scaling the gate dielectric: Materials, integration, and reliability. *IBM journal of research and development* **43**, 245–264 (1999).
12. Guha, S., Gusev, E., Copel, M., Ragnarsson, L.-Å. & Buchanan, D. A. Compatibility Challenges for High-κ Materials Integration into CMOS Technology. *MRS Bulletin* **27**, 226–229 (2002).
13. Hubbard, K. J. & Schlom, D. Thermodynamic stability of binary oxides in contact with silicon. *Journal of Materials Research* **11**, 2757–2776 (1996).
14. Schlom, D. G. & Haeni, J. H. A Thermodynamic Approach to Selecting Alternative Gate Dielectrics. *MRS Bulletin* **27**, 198–204 (2002).
15. Zhao, C. *et al.* Ternary rare-earth metal oxide high-k layers on silicon oxide. *Applied Physics Letters* **86**, 132903 (2005).
16. Coh, S. *et al.* Si-compatible candidates for high-κ dielectrics with the Pbnm perovskite structure. *Physical Review B* **82**, (2010).
17. Guha, S. & Narayanan, V. High-κ/Metal Gate Science and Technology. *Annual Review of Materials Research* **39**, 181–202 (2009).
18. Tsai, W. *et al.* Surface preparation and interfacial stability of high-k dielectrics deposited by atomic layer chemical vapor deposition. *Microelectronic Engineering* **65**, 259–272 (2003).
19. Muller, D. A. *et al.* The electronic structure at the atomic scale of ultrathin gate oxides. *Nature* **399**, 758–761 (1999).
20. Fischetti, M. V., Neumayer, D. A. & Cartier, E. A. Effective electron mobility in Si inversion layers in metal–oxide–semiconductor systems with a high-κ insulator: The role of remote phonon scattering. *Journal of Applied Physics* **90**, 4587 (2001).
21. Cho, M. *et al.* Insight Into N/PBTI Mechanisms in Sub-1-nm-EOT Devices. *IEEE Transactions on Electron Devices* **59**, 2042–2048 (2012).
22. Lopes, J. M. *et al.* Electrical and Structural Properties of Ternary Rare-Earth Oxides on Si and Higher Mobility Substrates and their Integration as High-k Gate Dielectrics in MOSFET Devices. *ECS Transactions* **35**, 461–479 (2011).
23. Kakushima, K. *et al.* Advantage of La_2O_3 gate dielectric over HfO_2 for direct contact and mobility improvment. *ESSDERC 2008 – 38th European Solid-State Device Research Conference* 126–129 (2008).doi:10.1109/ESSDERC.2008.4681715

24. Sze, S. M. & Ng, K. K. *Physics of semiconductor devices*. (John Wiley & Sons: Hoboken, NJ, USA, 2007).
25. Knoch, J., Feste, S. & Mantl, S. Metal-Oxide-Semiconductor Field-Effect Transistors: Novel Materials and Alternative Concepts. *Nanoelectronics and Information Technology* 341–355 (2012).
26. Ashcroft, N. W. & Mermin, N. D. *Solid State Physics*. 848 (Brooks Cole: Toronto, 1976).
27. Cabral, C. *et al.* Evaluation of Thermal Stability for CMOS Gate Metal Materials. *Journal of The Electrochemical Society* **151**, F283 (2004).
28. Schaeffer, J. K. *et al.* Challenges for the integration of metal gate electrodes. *IEDM Technical Digest. IEEE International Electron Devices Meeting, 2004.* 287–290 (2004).doi:10.1109/IEDM.2004.1419135
29. Schaeffer, J. *et al.* Application of group electronegativity concepts to the effective work functions of metal gate electrodes on high-κ gate oxides. *Microelectronic Engineering* **84**, 2196–2200 (2007).
30. Nichau, A. *et al.* $LaLuO_3$ higher-κ dielectric integration in SOI MOSFETs with a gate-first process. *Solid-State Electronics* **71**, 19–24 (2012).
31. Nichau, A. *et al.* Lanthanum Lutetium oxide integration in a gate-first process on SOI MOSFETs. *Ulis 2011 Ultimate Integration on Silicon* 1–4 (2011).doi:10.1109/ULIS.2011.5757952
32. Nichau, A. *et al.* Characterization of Lanthanum Lutetium oxide high-k/metal gate stacks for CMOS process integration. *SpLine Users' Meeting, Madrid, April 16/17th 2012* (2012).
33. Durgun Özben, E. *et al.* Integration of $LaLuO_3$ ($\kappa \sim 30$) as High-κ Dielectric on Strained and Unstrained SOI MOSFETs With a Replacement Gate Process. *IEEE Electron Device Letters* **32**, 15–17 (2011).
34. Robertson, J. High dielectric constant oxides. *The European Physical Journal Applied Physics* **28**, 265–291 (2004).
35. Schroder, D. K. *Semiconductor material and device characterization*. 1–790 (2006).
36. Hillard, R. J., Rai-Choudhury, P., Heddleson, J. M., Zier, D. A. & Schroder, D. K. Direct and Rapid Method for Determining Flatband Voltage from Non-equilibrium Capacitance Voltage Data. *Techniques for Semiconductor Materials and Devices* 261–274–14 (1992).
37. Natori, K. & Kurusu, T. Novel aspects of nanoscale transistors. *2006 International Workshop on Nano CMOS* 261–277 (2006). doi:10.1109/IWNC.2006.4570997
38. Stern, F. & Howard, W. Properties of Semiconductor Surface Inversion Layers in the Electric Quantum Limit. *Physical Review* **163**, 816–835 (1967).
39. Stern, F. Quantum properties of surface space-charge layers. *C R C Critical Reviews in Solid State Sciences* **4**, 499–514 (1973).
40. Stern, F. Calculated Temperature Dependence of Mobility in Silicon Inversion Layers. *Physical Review Letters* **44**, 1469–1472 (1980).
41. Ando, T. Electronic properties of two-dimensional systems. *Reviews of Modern Physics* **54**, 437–672 (1982).
42. Ohkura, Y. Quantum effects in Si *n*-MOS inversion layer at high substrate concentration. *Solid-State Electronics* **33**, 1581–1585 (1990).
43. Davies, J. H. *The physics of low-dimensional semiconductors*. 1–452 (Cambridge University Press: Cambridge, UK, 1997).
44. Feste, S. F., Habicht, S., Zhao, Q.-T., Buca, D. & Mantl, S. Strained and unstrained Si nanowire FETs. *2009 Proceedings of the European Solid State Device Research Conference* **2**, 323–326 (2009).
45. Chu, M., Sun, Y., Aghoram, U. & Thompson, S. E. Strain: A Solution for Higher Carrier Mobility in Nanoscale MOSFETs. *Annual Review of Materials Research* **39**, 203–229 (2009).
46. Van Dort, M. J., Woerlee, P. H. & Walker, A. J. A simple model for quantisation effects in heavily-doped silicon MOSFETs at inversion conditions. *Solid-State Electronics* **37**, 411–414 (1994).
47. Hauser, J. R. & Ahmed, K. Characterization of ultra-thin oxides using electrical C-V and I-V measurements. *The 1998 international conference on characterization and metrology for ULSI technology* 235–239 (1998).doi:10.1063/1.56801

48. Houssa, M., Naili, M., Afanas'ev, V. V., Heyns, M. M. & Stesmans, A. Electrical and physical characterization of high-k dielectric layers. *2001 International Symposium on VLSI Technology, Systems, and Applications. Proceedings of Technical Papers (Cat. No. 01TH8517)* 196–199 doi:10.1109/VTSA.2001.934518
49. Kauerauf, T. Degradation and breakdown of mos gate stacks with high permittivity dielectrics. (2007).
50. Nicollian, E. H. & Brews, J. R. *MOS (Metal Oxide Semiconductor) Physics and Technology*. 1–917 (John Wiley & Sons: New York, Chichester, Brisbane, Toronto, Singapore, 2003).
51. Shockley, W. & Read, W. Statistics of the Recombinations of Holes and Electrons. *Physical Review* **87**, 835–842 (1952).
52. Kim, Y., Ohmi, S., Tsutsui, K. & Iwai, H. Space-Charge-Limited Currents in La_2O_3 Thin Films Deposited by E-Beam Evaporation after Low Temperature Dry-Nitrogen Annealing. *Japanese Journal of Applied Physics* **44**, 4032–4042 (2005).
53. Southwick, R. G. I. & Knowlton, W. B. Stacked Dual Oxide MOS Energy Band Diagram Visual Representation Program. *IEEE Transactions on Device and Materials Reliability* **6**, 136–145 (2006).
54. Southwick, R. G. I., Sup, A., Jain, A. & Knowlton, W. B. An Interactive Simulation Tool For Complex Multilayer Dielectric Devices. *IEEE Transactions on Device and Materials Reliability* **11**, 236–243 (2011).
55. Hu, G., Dennard, R. H., Terman, L. M. & Petrillo, K. E. A self-aligned 1-μm-channel CMOS technology with retrograde *n*-well and thin epitaxy. *IEEE Transactions on Electron Devices* **32**, 203–209 (1985).
56. Richter, S. *et al.*Ω-Gated Silicon and Strained Silicon Nanowire Array Tunneling FETs. *IEEE Electron Device Letters* **33**, 1535–1537 (2012).
57. Knoll, L. *et al.* C-Tunnel FET Inverters with Strained Si Complementary Devices with High I_{on}. 65–66 (2011).
58. Knoll, L. *et al.* Complementary-Tunnel FET Inverters with ultrathin Strained Si. *Proceedings of 9th Workshop of the thematic Network on Silicon-on-Insulator Technology (EUROSOI)* (2013).
59. Veloso, A., Ragnarsson, L.-A., Cho, M. & Devriendt, K. Gate-last vs. gate-first technology for aggressively scaled EOT Logic/RF CMOS. *2011 Symposium on VLSI Technology (VLSIT)* **2011**, 34–35 (2011).
60. Frank, M. M. High-k/metal gate innovations enabling continued CMOS scaling. *2011 Proceedings of the European Solid-State Device Research Conference (ESSDERC)* 25–33 (2011). doi:10.1109/ESSDERC.2011.6044239
61. Kittel, C. *Introduction to Solid State Physics*. 680 (John Wiley & Sons: Hoboken, NJ, USA, 2004).
62. Veloso, A. *et al.* Process control & integration options of RMG technology for aggressively scaled devices. *2012 Symposium on VLSI Technology (VLSIT)* 33–34 (2012).doi:10.1109/VLSIT.2012.6242447
63. Hayashi, T. *et al.* Suppression of Boron Penetration from Source/Drain-Extension to Improve Gate Leakage Characteristics and Gate-Oxide Reliability for 65-nm Node CMOS and Beyond. *Japanese Journal of Applied Physics* **44**, 2157–2160 (2005).
64. Yamashita, T. *et al.* Impact of boron penetration from S/D-extension on gate-oxide reliability for 65-nm node CMOS and beyond. *Digest of Technical Papers. 2004 Symposium on VLSI Technology, 2004.* 136–137 (2004).doi:10.1109/VLSIT.2004.1345438
65. Bae, S. H. *et al.* Thickness optimization of the TiN metal gate with polysilicon-capping layer on Hf-based high-k dielectric. *Microelectronic Engineering* **83**, 460–462 (2006).
66. Son, N.-J. *et al.* A Unique Dual-Poly Gate Technology for 1.2-V Mobile DRAM with Simple In situ n^+-Doped Polysilicon. *IEEE Transactions on Electron Devices* **51**, 1644–1652 (2004).
67. Guha, S. *et al.* Examination of flatband and threshold voltage tuning of HfO_2/TiN field effect transistors by dielectric cap layers. *Applied Physics Letters* **90**, 092902 (2007).
68. Cartier, E. a. *et al.* Fundamental aspects of HfO_2-based high-k metal gate stack reliability and implications on $t_{\rm inv}$-scaling. *2011 International Electron Devices Meeting* 18.4.1–18.4.4 (2011). doi:10.1109/IEDM.2011.6131579

69. Cartier, E. A. The Role of Oxygen in the Development of Hf-Base High-k/Metal Gate Stacks for CMOS Technologies. *ECS Transactions* **33**, 83–94 (2010).
70. Ragnarsson, L.-A. *et al.* Ultra low-EOT (5 Å) gate-first and gate-last high performance CMOS achieved by gate-electrode optimization. *2009 IEEE International Electron Devices Meeting (IEDM)* 1–4 (2009).doi:10.1109/IEDM.2009.5424254
71. Ragnarsson, L.-A. *et al.* On the origin of the mobility reduction in bulk-Si, UTBOX-FDSOI and SiGe devices with ultrathin-EOT dielectrics. *Proceedings of 2011 International Symposium on VLSI Technology, Systems and Applications* 1–2 (2011).doi:10.1109/VTSA.2011.5872255
72. Ragnarsson, L.-Å. *et al.* Ultrathin EOT high-κ/metal gate devices for future technologies: Challenges, achievements and perspectives. *Microelectronic Engineering* **88**, 1317–1322 (2011).
73. Choi, K. *et al.* The effect of metal thickness, overlayer and high-k surface treatment on the effective work function of metal electrode. *Proceedings of 35th European Solid-State Device Research Conference, 2005. ESSDERC 2005.* 101–104 (2005).doi:10.1109/ESSDER.2005.1546595
74. Lemme, M. C. *et al.* Nanoscale TiN metal gate technology for CMOS integration. *Microelectronic Engineering* **83**, 1551–1554 (2006).
75. Westlinder, J. *et al.* On the thermal stability of atomic layer deposited TiN as gate electrode in MOS devices. *IEEE Electron Device Letters* **24**, 550–552 (2003).
76. Wu, L. *et al.* Thermal stability of TiN metal gate prepared by atomic layer deposition or physical vapor deposition on HfO_2 high-K dielectric. *Applied Physics Letters* **96**, 113510 (2010).
77. Tsuchiya, Y., Yoshiki, M., Koga, J., Nishiyama, A. & Koyama, M. Influence of Ni silicide phases on effective work function modulation with Al-pileup in the Ni fully silicided gate/HfSiON system. *Journal of Applied Physics* **106**, 044510 (2009).
78. Callegari, A. *et al.* Materials and Process Integration Issues in Metal Gate/High-k Stacks and Their Dependence on Device Performance. *ECS Transactions* **11**, 265–274 (2007).
79. Dadgour, H. F., Endo, K., De, V. K. & Banerjee, K. Grain-Orientation Induced Work Function Variation in Nanoscale Metal-Gate Transistors—Part II: Implications for Process, Device, and Circuit Design. *IEEE Transactions on Electron Devices* **57**, 2515–2525 (2010).
80. Dadgour, H. F., Endo, K., De, V. K. & Banerjee, K. Grain-Orientation Induced Work Function Variation in Nanoscale Metal-Gate Transistors—Part I: Modeling, Analysis, and Experimental Validation. *IEEE Transactions on Electron Devices* **57**, 2504–2514 (2010).
81. Singanamalla, R. *et al.* On the impact of TiN film thickness variations on the effective work function of poly-Si/TiN/SiO_2 and poly-Si/TiN/HfSiON gate stacks. *IEEE Electron Device Letters* **27**, 332–334 (2006).
82. Takahashi, H. *et al.* Ti-capping technique as a breakthrough for achieving low threshold voltage, high mobility, and high reliability of pMOSFET with metal gate and high-k dielectrics technologies. **5**, 427–430 (2009).
83. Copel, M. *et al.* Interaction of La_2O_3 capping layers with HfO_2 gate dielectrics. *Applied Physics Letters* **95**, 212903 (2009).
84. Ando, T. *et al.* Understanding mobility mechanisms in extremely scaled HfO_2 (EOT 0.42 nm) using remote interfacial layer scavenging technique and V_t-tuning dipoles with gate-first process. *2009 IEEE International Electron Devices Meeting (IEDM)* **2**, 1–4 (2009).
85. Ando, T. Ultimate Scaling of High-κ Gate Dielectrics: Higher-κ or Interfacial Layer Scavenging? *Materials* **5**, 478–500 (2012).
86. Arimura, H. Design of Higher-k Gate Dielectrics and Effective Work Function Control by Interface Dipole Formation. (2011).
87. Suhard, S. *et al.* Development of a Wet Silicon Removal Process for Replacement Metal Gate and Sacrificial Fin. **41**, 51–56 (2011).
88. Williams, K. & Muller, R. Etch rates for micromachining processing. *Journal of Microelectromechanical Systems* **5**, 256–269 (1996).

89. Heyns, M. *et al.* Challenges for introducing Ge and III/V devices into CMOS technologies. *2012 IEEE International Reliability Physics Symposium (IRPS)* 5D.1.1–5D.1.10 (2012).doi:10.1109/IRPS.2012.6241852
90. Wang, G. *et al.* High quality Ge epitaxial layers in narrow channels on Si (001) substrates. *Applied Physics Letters* **96**, 111903 (2010).
91. Ragnarsson, L.-A. *et al.* The Importance of Moisture Control for EOT Scaling of Hf-Based Dielectrics. *Journal of The Electrochemical Society* **156**, H416 (2009).
92. Barin, I. *Thermochemical Data of Pure Substances*. **19A4**, 1–1885 (Wiley-VCH Verlag GmbH: Weinheim, Germany, 1995).
93. *CRC Handbook of Chemistry & Physics*. (Taylor and Francis Group, LLC: 2012).
94. Sanderson, R. T. *Inorganic Chemistry*. (Reinhold: New York, 1967).
95. The Landolt-Börnstein Database. (2012).at <http://www.springermaterials.com>
96. Zhao, J.-C., Bewlay, B. P., Jackson, M. R. & Chen, Q. Hf-Si binary phase diagram determination and thermodynamic modeling. *Journal of Phase Equilibria* **21**, 40–45 (2000).
97. Meschel, S. V. & Kleppa, O. J. Standard enthalpies of formation of some 5d transition metal silicides by high temperature direct synthesis calorimetry. *Journal of Alloys and Compounds* **280**, 231–239 (1998).
98. Ando, T. *et al.* Ultimate EOT Scaling Using Hf-Based High-κ Gate Dielectrics and Impact on Carrier Mobility. *ECS Transactions* **28**, 115–123 (2010).
99. Kakushima, K. *et al.* Selection of rare earth silicate with SrO capping for EOT scaling below 0.5 nm. *2009 Proceedings of the European Solid State Device Research Conference* 403–406 (2009).doi:10.1109/ESSDERC.2009.5331331
100. Choi, K. *et al.* Extremely scaled gate-first high-k/metal gate stack with EOT of 0.55 nm using novel interfacial layer scavenging techniques for 22 nm technology node and beyond. *2009 Symposium on VLSI Technology, VLSIT 2009; Kyoto; 16 June 2009 through 18 June 2009; Category number CFP09VTS-PRT; Code78025* 138–139 (2009).
101. Ando, T. Ultimate Scaling of High-k Gate Dielectrics and Impact on Carrier Transport of Field-Effect Transistors. 122 (2010).
102. Ernsberger, C. Low temperature oxidation behavior of reactively sputtered TiN by x-ray photoelectron spectroscopy and contact resistance measurements. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* **4**, 2784 (1986).
103. Saha, N. C. & Tompkins, H. G. Titanium nitride oxidation chemistry: An x-ray photoelectron spectroscopy study. *Journal of Applied Physics* **72**, 3072 (1992).
104. Fromhold, A. T. *Theory of Metal Oxidation*. 547 (North Holland Publishing Company: Amsterdam, 1976).
105. Maitra, K., Frank, M. M., Narayanan, V., Misra, V. & Cartier, E. a. Impact of metal gates on remote phonon scattering in titanium nitride/hafnium dioxide *n*-channel metal–oxide–semiconductor field effect transistors–low temperature electron mobility study. *Journal of Applied Physics* **102**, 114507 (2007).
106. Arimura, H. *et al.* Maximized Benefit of La–Al–O Higher-κ Gate Dielectrics by Optimizing the La/Al Atomic Ratio. *IEEE Electron Device Letters* **32**, 288–290 (2011).
107. Briggs, D. & Seah, M. P. *Practical surface analysis. 1. Auger and x-ray photoelectron spectroscopy*. 1–657 (John Wiley & Sons: Hoboken, NJ, USA, 1990).
108. Kakushima, K. *et al.* Observation of band bending of metal/high-k Si capacitor with high energy x-ray photoemission spectroscopy and its application to interface dipole measurement. *Journal of Applied Physics* **104**, 104908 (2008).
109. Siegbahn, K. *et al.* ESCA – Atomic, Molecular and Solid State Structure Studied by Means of Electron Spectroscopy. *Nova Acta Regiae Soc. Sci. Upsaliensis* **Ser. IV**, (1967).
110. Rubio-Zuazo, J., Escher, M., Merkel, M. & Castro, G. R. High Voltage-Cylinder Sector Analyzer 300/15: a cylindrical sector analyzer for electron kinetic energies up to 15 keV. *The Review of scientific instruments* **81**, 043304 (2010).
111. ISO ISO 18115: Surface chemical analysis – Vocabulary. *ISO 18115: Surface chemical analysis – Vocabulary* (2001).

112. Rubio-Zuazo, J. & Castro, G. R. Effective attenuation length dependence on photoelectron kinetic energy for Au from 1 keV to 15 keV. *Journal of Electron Spectroscopy and Related Phenomena* **184**, 384–390 (2011).
113. Rubio-Zuazo, J., Ferrer, P. & Castro, G. R. Non-destructive compositional depth profile in the tens-of-nanometer scale. *Journal of Electron Spectroscopy and Related Phenomena* **180**, 27–33 (2010).
114. Nichau, A. *et al.* Effective attenuation length for lanthanum lutetium oxide between 7 and 13 keV. *Applied Physics Letters* **102**, 031607 (2013).
115. Lopes, J. M. J. *et al.* Amorphous ternary rare-earth gate oxides for future integration in MOSFETs. *Microelectronic Engineering* **86**, 1646–1649 (2009).
116. *NIST X-ray Photoelectron Spectroscopy Database, Version 4.0*. (National Institute of Standards and Technology: Gaithersburg, 2008).
117. Teterin, Y. A. & Teterin, A. Y. Structure of X-ray photoelectron spectra of lanthanide compounds. *Russian Chemical Reviews* **71**, 347–381 (2002).
118. Teterin, Y. A., Teterin, A. Y., Lebedev, A. M. & Ivanov, K. E. Secondary electronic processes and the structure of X-ray photoelectron spectra of lanthanides in oxygen-containing compounds. *Journal of Electron Spectroscopy and Related Phenomena* **137-140**, 607–612 (2004).
119. Sunding, M. F. *et al.* XPS characterisation of in situ treated lanthanum oxide and hydroxide using tailored charge referencing and peak fitting procedures. *Journal of Electron Spectroscopy and Related Phenomena* **184**, 399–409 (2011).
120. Crecelius, G., Wertheim, G. K. & Buchanan, D. N. E. Core-hole screening in lanthanide metals. *Physical Review B* **18**, 6519–6524 (1978).
121. Tanuma, S., Powell, C. J. & Penn, D. R. Calculations of electorn inelastic mean free paths. II. Data for 27 elements over the 50–2000 eV range. *Surface and Interface Analysis* **17**, 911–926 (1991).
122. Tanuma, S., Powell, C. J. & Penn, D. R. Calculations of electron inelastic mean free paths. V. Data for 14 organic compounds over the 50–2000 eV range. *Surface and Interface Analysis* **21**, 165–176 (1994).
123. Tanuma, S., Powell, C. J. & Penn, D. R. Calculations of electron inelastic mean free paths. IX. Data for 41 elemental solids over the 50 eV to 30 keV range. *Surface and Interface Analysis* **43**, 689–713 (2011).
124. Tanuma, S., Powell, C. J. & Penn, D. R. Calculations of electron inelastic mean free paths (IMFPs). IV. Evaluation of calculated IMFPs and of the predictive IMFP formula TPP-2 for electron energies between 50 and 2000 eV. *Surface and Interface Analysis* **20**, 77–89 (1993).
125. Krause, M. O. & Oliver, J. H. Natural widths of atomic K and L levels, K_α X-ray lines and several KLL Auger lines. *Journal of Physical and Chemical Reference Data* **8**, 329 (1979).
126. Burrow, B. J. A correlation of Auger electron spectroscopy, x-ray photoelectron spectroscopy, and Rutherford backscattering spectrometry measurements on sputter-deposited titanium nitride thin films. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* **4**, 2463 (1986).
127. Rubio-Zuazo, J. & Castro, G. R. Information depth determination for hard X-ray photoelectron spectroscopy up to 15 keV photoelectron kinetic energy. *Surface and Interface Analysis* **40**, 1438–1443 (2008).
128. Kakushima, K. *et al.* Characterization of flatband voltage roll-off and roll-up behavior in La_2O_3/silicate gate dielectric. *Solid-State Electronics* **54**, 720–723 (2010).
129. Kakushima, K., Okamoto, K. & Tachi, K. Further EOT scaling below 0.4 nm for high-k gated MOSFET. *Technical digest of IWDTF 2008: International Workshop on Dielectric Thin Films for Future ULSI Devices: Science and Technology* 9–10 (2008).
130. Roeckerath, M. *et al.* Characterization of lanthanum lutetium oxide thin films grown by atomic layer deposition as an alternative gate dielectric. *Thin Solid Films* **517**, 201–203 (2008).
131. Kawanago, T. *et al.* Electrical characterization of directly deposited La-Sc oxides complex for gate insulator application. *Microelectronic Engineering* **84**, 2235–2238 (2007).

132. Oku, M., Wagatsuma, K. & Kohiki, S. Ti 2p and Ti 3p X-ray photoelectron spectra for TiO_2, $SrTiO_3$ and $BaTiO_3$. *Physical Chemistry Chemical Physics* **1**, 5327–5331 (1999).
133. Shlyakhtina, A. V., Shcherbakova, L. G., Knotko, A. V. & Steblevskii, A. V. Study of the fluorite-pyrochlore-fluorite phase transitions in $Ln_2Ti_2O_7$ (Ln = Lu, Yb, Tm). *Journal of Solid State Electrochemistry* **8**, 661–667 (2004).
134. Heidelmann, M., Barthel, J. & Houben, L. StripeSTEM, a technique for the isochronous acquisition of high angle annular dark-field images and monolayer resolved electron energy loss spectra. *Ultramicroscopy* **109**, 1447–52 (2009).
135. Xiong, K. & Robertson, J. Oxygen vacancies in high dielectric constant oxides La_2O_3, Lu_2O_3, and $LaLuO_3$. *Applied Physics Letters* **95**, 022903 (2009).
136. Scofield, J. H. Hartree-Slater subshell photoionization cross-sections at 1254 and 1487 eV. *Journal of Electron Spectroscopy and Related Phenomena* **8**, 129–137 (1976).
137. Scofield, J. H. *Theoretical Photoionization Cross Sections from 1 to 1500 keV*. 376 (1973).doi:10.2172/4545040
138. Robertson, J. private communication. *private communication* (2012).
139. Volk, C. *et al.* $LaLuO_3$ as a high-k gate dielectric for InAs nanowire structures. *Semiconductor Science and Technology* **25**, 085001 (2010).
140. Lopes, J. M. J. *et al.* Amorphous ternary high-k oxides on Si and higher mobility substrates. *Spring meeting of the German Physical Society, Regensburg: 23.03.2010, 2010.* (2010).
141. Zhao, Q.-T. *et al.* High mobility Si-Ge channel and high-k materials for NanoMOSFETs. *5th International SiGe Technology and Device Meeting (ISTDM 2010) Stockholm, Schweden: 24.05.2010–26.05.2010, 2010.* (2010).
142. Sacchi, M. *et al.* Quantifying the effective attenuation length in high-energy photoemission experiments. *Physical Review B* **71**, 155117 (2005).
143. Roeckerath, M. Seltenerd-basierte ternäre Oxide als alternative Gatedielektrika. (2008).
144. Radtke, C. *et al.* Physicochemical and Electrical Properties of $LaLuO_3$/Ge(100) Structures Submitted to Postdeposition Annealings. *Electrochemical and Solid-State Letters* **13**, G37 (2010).
145. ITRS ITRS Roadmap 2011^{th} Update. (2011).at <http://www.itrs.net>
146. Green IT. *Wikipedia* (2013).at <http://en.wikipedia.org/wiki/Green_IT>
147. Afanas'ev, V. V. *et al.* Electronic structure of silicon interfaces with amorphous and epitaxial insulating oxides: Sc_2O_3, Lu_2O_3, $LaLuO_3$. *Microelectronic Engineering* **84**, 2278–2281 (2007).
148. Zhao, Y., Toyama, M., Kita, K., Kyuno, K. & Toriumi, A. Moisture-absorption-induced permittivity deterioration and surface roughness enhancement of lanthanum oxide films on silicon. *Applied Physics Letters* **88**, 072904 (2006).
149. Kakushima, K. *et al.* Comprehensive x-ray photoelectron spectroscopy study on compositional gradient lanthanum silicate film. *Journal of Applied Physics* **106**, 124903 (2009).
150. Mitrovic, I. Z. *et al.* Effect of oxygen on tuning the TiN_x metal gate work function on $LaLuO_3$. *Thin Solid Films* **520**, 6959–6962 (2012).
151. Smedt, F. De, Vinckier, C., Cornelissen, I., Gendt, S. De & Heyns, M. A Detailed Study on the Growth of Thin Oxide Layers on Silicon Using Ozonated Solutions. *Journal of The Electrochemical Society* **147**, 1124 (2000).
152. Briend, G., Besson, P., Salvetat, T. & Pollet, O. Advanced Surface Preparation Development on a 300 mm Single Wafer Spin-on Platform. *ECS Transactions* **11**, 249–256 (2007).
153. Miyazaki, S., Nishimura, H., Fukuda, M., Ley, L. & Ristein, J. Structure and electronic states of ultrathin SiO_2 thermally grown on Si(100) and Si(111) surfaces. *Applied Surface Science* **113-114**, 585–589 (1997).
154. Kaneta, C., Yamasaki, T., Uchiyama, T., Uda, T. & Terakura, K. Structure and electronic property of interface. *Microelectronic Engineering* **48**, 117–120 (1999).
155. Vogel, E. M., Henson, W. K., Richter, C. A. & Suehle, J. S. Limitations of conductance to the measurement of the interface state density of MOS capacitors with tunneling gate dielectrics. *IEEE Transactions on Electron Devices* **47**, 601–608 (2000).
156. Springer Materials – The Landolt-Börnstein Database. (2012).at <http://www.springermaterials.com>

157. Pozzo, M., Alfè, D., Amieiro, A., French, S. & Pratt, A. Hydrogen dissociation and diffusion on Ni- and Ti-doped Mg(0001) surfaces. *The Journal of chemical physics* **128**, 094703 (2008).
158. Meuris, M., Mertens, P. & Opdebeeck, A. The IMEC clean: A new concept for particle and metal removal on Si surfaces. *Solid State* (1995).
159. Vitchev, R. G. *et al.* X-ray photoelectron spectroscopy characterisation of high-k dielectric Al_2O_3 and HfO_2 layers deposited on SiO_2/Si surface. *Applied Surface Science* **235**, 21–25 (2004).
160. Green, M. L. *et al.* Nucleation and growth of atomic layer deposited HfO_2 gate dielectric layers on chemical oxide (Si–O–H) and thermal oxide (SiO_2 or Si–O–N) underlayers. *Journal of Applied Physics* **92**, 7168 (2002).
161. Puurunen, R. L. *et al.* Island growth in the atomic layer deposition of zirconium oxide and aluminum oxide on hydrogen-terminated silicon: Growth mode modeling and transmission electron microscopy. *Journal of Applied Physics* **96**, 4878 (2004).
162. Copel, M., Gribelyuk, M. & Gusev, E. Structure and stability of ultrathin zirconium oxide layers on Si(001). *Applied Physics Letters* **76**, 436 (2000).
163. Gusev, E. P., Cabral, C., Copel, M., D'Emic, C. & Gribelyuk, M.Ultrathin HfO_2 films grown on silicon by atomic layer deposition for advanced gate dielectrics applications. *Microelectronic Engineering* **69**, 145–151 (2003).
164. Frank, M. M. *et al.* Enhanced initial growth of atomic-layer-deposited metal oxides on hydrogen-terminated silicon. *Applied Physics Letters* **83**, 740 (2003).
165. Mack, P., White, R. G., Wolstenholme, J. & Conard, T. The use of angle resolved XPS to measure the fractional coverage of high-k dielectric materials on silicon and silicon dioxide surfaces. *Applied Surface Science* **252**, 8270–8276 (2006).
166. Zhang, B. *et al.* Epitaxial growth of Ni(Al)$Si_{0.7}Ge_{0.3}$ on $Si_{0.7}Ge_{0.3}$/Si(100) by Al interlayer mediated epitaxy. *Applied Physics Letters* **98**, 252101 (2011).
167. Grigorov, K. G. *et al.* Diffusion of silicon in titanium nitride films. Efficiency of TiN barrier layers. *Applied Physics A Solids and Surfaces* **55**, 502–504 (1992).
168. Nowak, W. B. Diffusion of nickel through titanium nitride films. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* **3**, 2242 (1985).
169. Grigorov, G. I. *et al.* Aluminium diffusion in titanium nitride films. Efficiency of TiN barrier layers. *Applied Physics A Solids and Surfaces* **57**, 195–197 (1993).
170. Champaneria, R., Mack, P., White, R. & Wolstenholme, J. Non-destructive analysis of ultrathin dielectric films. *Surface and Interface Analysis* **35**, 1028–1033 (2003).
171. Tai, K. *et al.* Threshold Voltage Modulation Technique using Fluorine Treatment through Atomic Layer Deposition TiN Suitable for Complementary Metal–Oxide–Semiconductor Devices. *Japanese Journal of Applied Physics* **47**, 2345–2348 (2008).
172. Takahashi, K. *et al.* High-Mobility Dual Metal Gate MOS Transistors with High-k Gate Dielectrics. *Japanese Journal of Applied Physics* **44**, 2210–2213 (2005).
173. Westlinder, J., Sjöblom, G. & Olsson, J. Variable work function in MOS capacitors utilizing nitrogen-controlled TiN_x gate electrodes. *Microelectronic Engineering* **75**, 389–396 (2004).
174. Kim, H., McIntyre, P. C., On Chui, C., Saraswat, K. C. & Stemmer, S. Engineering chemically abrupt high-k metal oxide/silicon interfaces using an oxygen-gettering metal overlayer. *Journal of Applied Physics* **96**, 3467 (2004).
175. Pritschow, M. Titannitrid- und Titan-Schichten für die Nano-Elektromechanik. (2007).
176. Nichau, A. Fabrication of AlN barriers for Superconducting Tunnel Elements Using an Inductively Coupled Plasma. 74 (2009).
177. Kawanago, T. *et al.* EOT of 0.62 nm and High Electron Mobility in La-silicate/Si Structure Based nMOSFETs Achieved by Utilizing Metal-Inserted Poly-Si Stacks and Annealing at High Temperature. *IEEE Transactions on Electron Devices* **59**, 269–276 (2012).
178. Guillot, J. *et al.* Phase mixture in MOCVD and reactive sputtering TiO_xN_y thin films revealed and quantified by XPS factorial analysis. *Acta Materialia* **54**, 3067–3074 (2006).
179. Porte, L., Roux, L. & Hanus, J. Vacancy effects in the x-ray photoelectron spectra of TiN_x. *Physical Review B* **28**, 3214–3224 (1983).

180. Mayer, J. W. & Lau, S. S. *Electronic Materials Science: For Integrated Circuits in Si and GaAs*. 480 (Prentice Hall: 1989).
181. Kang, C. Y. *et al.* Effects of metal gate-induced strain on the performance of metal-oxide-semiconductor field effect transistors with titanium nitride gate electrode and hafnium oxide dielectric. *Applied Physics Letters* **91**, 033511 (2007).
182. Lim, S. H. N., McCulloch, D. G., Bilek, M. M. M. & McKenzie, D. R. Relation between microstructure and stress in titanium nitride films grown by plasma immersion ion implantation. *Journal of Applied Physics* **93**, 4283 (2003).
183. Toledano-Luque, M. *et al.* From mean values to distributions of BTI lifetime of deeply scaled FETs through atomistic understanding of the degradation. *VLSI Technology (VLSIT), 2011 Symposium on* 152–153 (2011).
184. Lebedinskii, Y., Zenkevich, A. & Gusev, E. P. Measurements of metal gate effective work function by x-ray photoelectron spectroscopy. *Journal of Applied Physics* **101**, 074504 (2007).
185. Miyazaki, S. Photoemission study of energy-band alignments and gap-state density distributions for high-k gate dielectrics. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures* **19**, 2212 (2001).
186. Afanas'ev, V. V., Stesmans, A., Chen, F., Shi, X. & Campbell, S. a. Internal photoemission of electrons and holes from (100) Si into HfO_2. *Applied Physics Letters* **81**, 1053 (2002).
187. Adamchuk, V. K. & Afanas'ev, V. V. Internal photoemission spectroscopy of semiconductor-insulator interfaces. *Progress in Surface Science* **41**, 111–211 (1992).
188. Durgun Özben, E. *et al.* $LaScO_3$ as a higher-κ dielectric for *p*-MOSFETs. *Microelectronic Engineering* **88**, 1323–1325 (2011).
189. Durgun Özben, E. *et al.* Rare-earth oxide TiN gate stacks on high mobility strained silicon on insulator for fully depleted metal-oxide-semiconductor field-effect transistors. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures* **29**, 01A903 (2011).
190. Sodini, C. G., Ekstedt, T. W. & Moll, J. L. Charge accumulation and mobility in thin dielectric MOS transistors. *Solid-State Electronics* **25**, 833–841 (1982).
191. Huang, C.-L. & Gildenblat, G. S. Correction factor in the split C–V method for mobility measurements. *Solid-State Electronics* **36**, 611–615 (1993).
192. Cabon-Till, B., Cristoloveanu, S. & Ghibaudo, G. Experimental determination of short-channel MOSFET parameters. *Solid-State Electronics* **28**, 1025–1030 (1985).
193. Ghibaudo, G. New method for the extraction of MOSFET parameters. *Electronics Letters* **24**, 543–545 (1988).
194. Romanjek, K., Andrieu, F., Ernst, T. & Ghibaudo, G. Improved Split C–V Method for Effective Mobility Extraction in sub-0.1-μ Si MOSFETs. *IEEE Electron Device Letters* **25**, 583–585 (2004).
195. Romanjek, K., Andrieu, F., Ernst, T. & Ghibaudo, G. Characterization of the effective mobility by split C(V) technique in sub 0.1 μm Si and SiGe PMOSFETs. *Solid-State Electronics* **49**, 721–726 (2005).
196. Lime, F. *et al.* Characterization of effective mobility by split C(V) technique in N-MOSFETs with ultra-thin gate oxides. *Solid-State Electronics* **47**, 1147–1153 (2003).
197. Tatsumura, K. *et al.* Clarification of Additional Mobility Components associated with TaC and TiN Metal Gates in scaled HfSiON MOSFETs down to sub-1.0 nm EOT. *2007 IEEE International Electron Devices Meeting* 349–352 (2007).doi:10.1109/IEDM.2007.4418943
198. Takagi, S., Toriumi, A., Iwase, M. & Tango, H. On the universality of inversion layer mobility in Si MOSFET's: Part I-effects of substrate impurity concentration. *IEEE Transactions on Electron Devices* **41**, 2357–2362 (1994).
199. Takagi, S., Toriumi, A., Iwase, M. & Tango, H. On the universality of inversion layer mobility in Si MOSFET's: Part II-effects of surface orientation. *IEEE Transactions on Electron Devices* **41**, 2363–2368 (1994).
200. Huang, C.-L., Faricelli, J. V. & Arora, N. D. A new technique for measuring MOSFET inversion layer mobility. *IEEE Transactions on Electron Devices* **40**, 1134–1139 (1993).

201. Kilchytska, V., Lederer, D., Collaert, N., Raskin, J.-P. & Flandre, D. Accurate effective mobility extraction by split C-V technique in SOI MOSFETs: suppression of the influence of floating-body effects. *IEEE Electron Device Letters* **26**, 749–751 (2005).
202. Koomen, J. Investigation of the MOST channel conductance in weak inversion. *Solid-State Electronics* **16**, 801–810 (1973).
203. Thomas, S. M. *et al.* Improved effective mobility extraction in MOSFETs. *Solid-State Electronics* **53**, 1252–1256 (2009).
204. Takagi, S., Iwase, M. & Toriumi, A. On the universality of inversion-layer mobility in *n*- and *p*-channel MOSFETs. *Technical Digest., International Electron Devices Meeting* 398–401 (1988).doi:10.1109/IEDM.1988.32840.
205. Matthiessen, A. On the electric conducting power of the metals. *Philosophical Transactions of the Royal Society of London* **148**, 383–387 (1858).
206. Takagi, S. & Takayanagi, M. Experimental Evidence of Inversion-Layer Mobility Lowering in Ultrathin Gate Oxide Metal-Oxide-Semiconductor Field-Effect-Transistors with Direct Tunneling Current. *Japanese Journal of Applied Physics* **41**, 2348–2352 (2002).
207. Park, J. T. & Colinge, J.-P. Multiple-gate SOI MOSFETs: device design guidelines. *IEEE Transactions on Electron Devices* **49**, 2222–2229 (2002).
208. Colinge, J.-P. Multiple-gate SOI MOSFETs. *Solid-State Electronics* **48**, 897–905 (2004).
209. Yu, B. *et al.* FinFET scaling to 10 nm gate length. *Technical Digest of the International Electron Devices Meeting, IEDM 2002* 251–254 (2002).
210. Doyle, B. S. *et al.* High performance fully-depleted tri-gate CMOS transistors. *IEEE Electron Device Letters* **24**, 263–265 (2003).
211. Chiarella, T. *et al.* Migrating from planar to FinFET for further CMOS scaling: SOI or bulk? *2009 Proceedings of ESSCIRC* 84–87 (2009).doi:10.1109/ESSCIRC.2009.5325993
212. Bohr, M. Perspective of Past, Present, and Future of Transistors. *IEDM 2011, Washington DC, December 5–7 2011* (2011).
213. Yamashita, T. *et al.* Digest of Technical Papers – Symposium on VLSI Technology. *Digest of Technical Papers – Symposium on VLSI Technology* 14–15 (2011).
214. Chang, J. & Guillorn, M. Scaling of SOI FinFETs down to Fin Width of 4 nm for the 10 nm technology node. *Digest of Technical Papers – Symposium on VLSI Technology* 12–13 (2011).
215. Nicollian, E. H. & Goetzberger, A. The Si-SiO_2 interface-electrical properties as determined by the metal-insulator-silicon conductance technique. *Bell Syst. Tech. J* **46**, 1055–1133 (1967).
216. Groeseneken, G., Maes, H. E., Beltran, N. & De Keersmaecker, R. F. A reliable approach to charge-pumping measurements in MOS transistors. *IEEE Transactions on Electron Devices* **31**, 42–53 (1984).
217. Toledano-Luque, M. *et al.* New Developments in Charge Pumping Measurements on Thin Stacked Dielectrics. *IEEE Transactions on Electron Devices* **55**, 3184–3191 (2008).
218. Elewa, T., Haddara, H., Cristoloveanu, S. & Bruel, M. Charge Pumping in Silicon On insulator Structures Using Gated P-I-N Diodes. *Le Journal de Physique Colloques* **49**, C4–137–C4–140 (1988).
219. Wouters, D. J., Tack, M. R., Groeseneken, G. V., Maes, H. E. & Claeys, C. Characterization of front and back Si-SiO/sub 2/interfaces in thick- and thin-film silicon-on-insulator MOS structures by the charge-pumping technique. *IEEE Transactions on Electron Devices* **36**, 1746–1750 (1989).
220. Groeseneken, G. Introduction to Charge Pumping and Its Applications. *Semiconductor Interface Specialists Conference SISC 2008 – Tutorials* (2008).
221. Cho, M. *et al.* Interface Trap Characterization of a 5.8-Å EOT *p*-MOSFET Using High-Frequency On-Chip Ring Oscillator Charge Pumping Technique. *IEEE Transactions on Electron Devices* **58**, 3342–3349 (2011).
222. Kapila, G., Kaczer, B., Nackaerts, A., Collaert, N. & Groeseneken, G. Direct Measurement of Top and Sidewall Interface Trap Density in SOI FinFETs. *IEEE Electron Device Letters* **28**, 232–234 (2007).

223. Ortiz-Conde, A. *et al.* A review of recent MOSFET threshold voltage extraction methods. *Microelectronics Reliability* **42**, 583–596 (2002).
224. Colinge, J.-P. *FinFETs and Other Multi-Gate Transistors*. (Springer US: Boston, MA, 2008).doi:10.1007/978-0-387-71752-4
225. Young, C. D. & Baykan, M. Critical discussion on (100) and (110) orientation dependent transport: nmos planar and finfet. *2011 Symposium on VLSI Technology (VLSIT)* 18–19 (2011).
226. Shao, X. & Yu, Z. Nanoscale FinFET simulation: A quasi-3D quantum mechanical model using NEGF. *Solid-State Electronics* **49**, 1435–1445 (2005).
227. Jeppson, K. O. & Svensson, C. M. Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices. *Journal of Applied Physics* **48**, 2004 (1977).
228. JEDEC Procedure for the W afer-Level Testing of Thin Dielectrics. (2001).
229. Degraeve, R. Time dependent dielectric breakdown in thin oxides: mechanisms, statistics and oxide reliability prediction. 175 (1998).
230. Kauerauf, T. *et al.* Methodologies for sub-1 nm EOT TDDB evaluation. *2011 International Reliability Physics Symposium* 2A.2.1–2A.2.10 (2011).doi:10.1109/IRPS.2011.5784444
231. Feijoo, P. C. *et al.* Time-Dependent Dielectric Breakdown on Subnanometer EOT nMOS FinFETs. *IEEE Transactions on Device and Materials Reliability* **12**, 166–170 (2012).
232. Roussel, P. J., Degraeve, R., Sahhaf, S. & Groeseneken, G. A consistent model for the hard breakdown distribution including digital soft breakdown: the noble art of area scaling. *Microelectronic Engineering* **84**, 1925–1928 (2007).
233. Sahhaf, S. *et al.* TDDB Reliability Prediction Based on the Statistical Analysis of Hard Breakdown Including Multiple Soft Breakdown and Wear-out. *2007 IEEE International Electron Devices Meeting* 501–504 (2007).doi:10.1109/IEDM.2007.4418984
234. Sahhaf, S. Electrical defects in high-k/Metal gate MOS transistors. 240 (2010).
235. Wu, E. Y. *et al.* Experimental evidence of T_{BD} power-law for voltage dependence of oxide breakdown. *IEEE Transactions on Electron Devices* **49**, 2244–2253 (2002).
236. Wu, E. Y. An Effective Correction Methodology for Interference of Stress-Induced Leakage Current in TDDB Evaluation of High-k Dielectrics. *IEEE Electron Device Letters* **33**, 191–193 (2012).
237. Efron, B. Second Thoughts on the Bootstrap. *Statistical Science* **18**, 135–140 (2003).
238. Nigam, T., Degraeve, R., Groeseneken, G., Heyns, M. M. & Maes, H. E. Constant current charge-to-breakdown: Still a valid tool to study the reliability of MOS structures? *1998 IEEE International Reliability Physics Symposium Proceedings 36th Annual (Cat No 98CH36173) RELPHY-98* 62–69 (1998).doi:10.1109/RELPHY.1998.670444
239. Wu, E. Y. *et al.* Voltage-dependent voltage-acceleration of oxide breakdown for ultra-thin oxides. *International Electron Devices Meeting 2000. Technical Digest. IEDM (Cat. No.00CH37138)* 541–544 (2000).doi:10.1109/IEDM.2000.904375
240. Groeseneken, G., Degraeve, R., Kaczer, B. & Martens, K. Trends and perspectives for electrical characterization and reliability assessment in advanced CMOS technologies. *2010 Proceedings of the European Solid State Device Research Conference* 64–72 (2010).doi:10.1109/ESSDERC.2010.5617735
241. Knoll, L. *et al.* Ultrathin Ni Silicides With Low Contact Resistance on Strained and Unstrained Silicon. *IEEE Electron Device Letters* **31**, 350–352 (2010).

Band / Volume 13
Alternative Systems for Molecular Electronics:
Functionalized Carboxylic Acids on Structured Surfaces
M. C. Lennartz (2010), 183 pp
ISBN: 978-3-89336-667-5

Band / Volume 14
Highly conductive electrodes as diffusion barrier
for high temperature applications
B. Meši (2010), VII, 138 pp
ISBN: 978-3-89336-670-5

Band / Volume 15
Modeling, Fabrication and Characterization
of Silicon Tunnel Field-Effect Transistors
C. P. Sandow (2010), XIII, 112 pp
ISBN: 978-3-89336-675-0

Band / Volume 16
Substituted Coronenes for Molecular Electronics:
From Supramolecular Structures to Single Molecules
P. Kowalzik (2010), ix, 149 pp
ISBN: 978-3-89336-679-8

Band / Volume 17
Resistive switching in TiO2 thin films
L. Yang (2011), VII, 117 pp
ISBN: 978-3-89336-707-8

Band / Volume 18
Crystal- and Defect-Chemistry of Fine Grained Thermistor Ceramics
on $BaTiO_3$ Basis with BaO-Excess
H. Katsu (2011), xxvii, 163 pp
ISBN: 978-3-89336-741-2

Band / Volume 19
Flächenkontakte zu molekularen Schichten in der Bioelektronik
N. Sanetra (2012), XIII, 129 pp
ISBN: 978-3-89336-776-4

Band / Volume 20
Stacked device structures for resistive memory and logic
R. D. Rosezin (2012), 137 pp
ISBN: 978-3-89336-777-1

Band / Volume 21
Optical and electrical addressing in molecule-based logic circuits
M. Manheller (2012), XIV, 183 pp
ISBN: 978-3-89336-810-5

Band / Volume 22
Fabrication of Nanogaps and Investigation of Molecular Junctions by Electrochemical Methods
Z. Yi (2012), 132 pp
ISBN: 978-3-89336-812-9

Band / Volume 23
Thermal Diffusion in binary Surfactant Systems and Microemulsions
B. Arlt (2012), 159, xlvi pp
ISBN: 978-3-89336-819-8

Band / Volume 24
Ultrathin Gold Nanowires - Chemistry, Electrical Characterization and Application to Sense Cellular Biology
A. Kisner (2012), 176 pp
ISBN: 978-3-89336-824-2

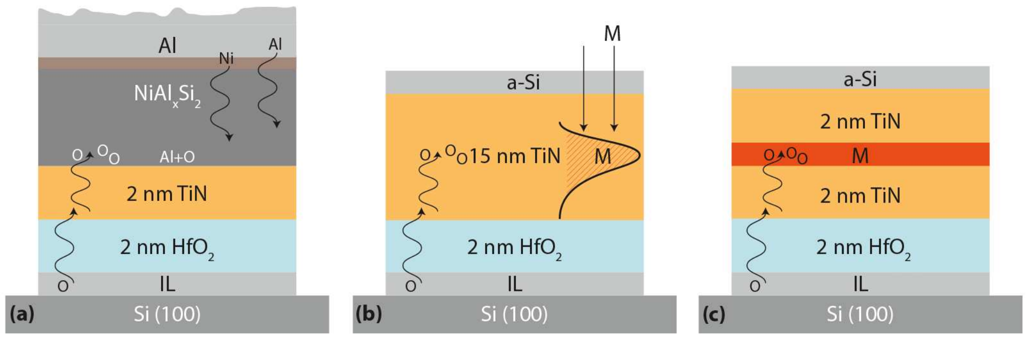
Band / Volume 25
Interaction between Redox-Based Resistive Switching Mechanisms
C. R. Hermes (2012), iii, 134 pp
ISBN: 978-3-89336-838-9

Band / Volume 26
Supported lipid bilayer as a biomimetic platform for neuronal cell culture
D. Afanasenkau (2013), xiv, 132 pp
ISBN: 978-3-89336-863-1

Band / Volume 27
15th European Workshop on Metalorganic Vapour Phase Epitaxy (EWMOVPE XV) June 2-5, 2013, Aachen, Germany
A. Winden (Chair) (2013), ??? pp
ISBN: 978-3-89336-870-9

Band / Volume 28
Characterization, integration and reliability of HfO_2 and $LaLuO_3$ high-κ/metal gate stacks for CMOS applications
A. Nichau (2013), xi, 177 pp
ISBN: 978-3-89336-898-3

Weitere ***Schriften des Verlags im Forschungszentrum Jülich*** unter
http://wwwzb1.fz-juelich.de/verlagextern1/index.asp



Information / Information
Band / Volume 28
ISBN 978-3-89336-898-3

JÜLICH
FORSCHUNGSZENTRUM

JÜLICH
FORSCHUNGSZENTRUM

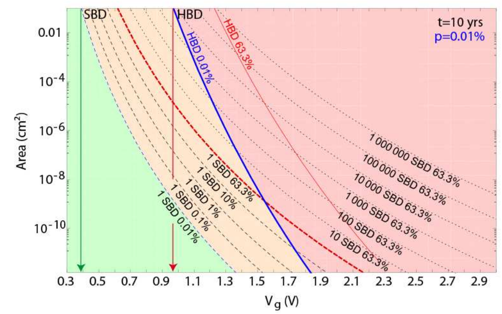
HfO$_2$ and LaLuO$_3$ high-κ/metal gate stacks for CMOS applications

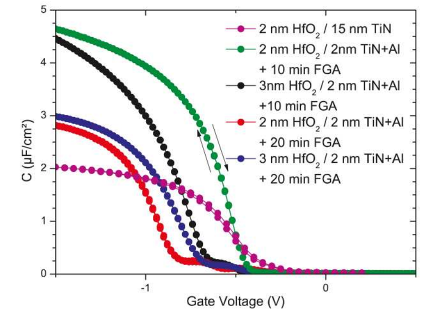
Alexander Nichau

Information

28

Member of the Helmholtz Association





Characterization, integration and reliability of HfO$_2$ and LaLuO$_3$ high-κ/metal gate stacks for CMOS applications

Alexander Nichau

JÜLICH
FORSCHUNGSZENTRUM