Methodology for Electrical Characterization of MOS Devices with Alternative Gate Dielectrics

Vom Fachbereich 18 - Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte

Dissertation

von

Mag. Andreas Kerber

geboren am 11. Okt. 1973
in Schnann am Arlberg / Tirol

Referent:	Prof. Dr. Udo Schwalke
Korreferent:	Prof. Dr. Herman Maes
Tag der Einreichung:	16. Okt. 2003
Tag der mündlichen Prüfung:	19. Jan. 2004

D17
Darmstädter Dissertationen

Eidesstattliche Erklärung laut §9 PromO

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation allein und nur unter Verwendung der angegebenen Literatur verfasst habe. Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Andreas Kerber

Darmstadt, 16. Okt. 2003

Acknowledgment

After 2 1/2 years of intense studies on high-ε gate dielectrics the moment has come to finalize this chapter of life. Moreover, it is the time to recognize the support from many colleagues I have received throughout that period of time. Without their contribution the comprehensive study presented here, would not have been possible.

First of all, I would like to thank my promotor Prof. Udo Schwalke from the TU-Darmstadt and my co-promotor Prof. Herman E. Maes from the KU-Leuven for supporting the idea of carrying out a PhD within an industrial affiliation program at the Inter-University Microelectronic Center (IMEC) in Leuven, Belgium. Their confidence has been the base for an effective and fruitful collaboration. In addition, my daily advisor Prof. Guido Groeseneken encouraged me to address many of the aspects discussed here. It has been a great pleasure working with you.

The research was conducted within the IMEC Industrial Affiliation Program (IIAP) on high-ε dielectrics for future CMOS technologies. Representatives for the project Stefan DeGendt and Marc Heyns are mentioned here for the successful collaboration. Without their encouragement and the support of the IMEC high-ε team many of the results presented here would not have been collected.

As an Infineon Technologies assignee to International SEMATECH I would like to mention the project management, represented by Dr. Ed Strickland from International SEMATECH, for the support I've received during the last 2 1/2 years.

Furthermore, Infineon Technologies is acknowledged for allowing me to conduct the research towards a PhD in electrical engineering during my stay at IMEC. Representative for the Infineon Technologies management I would like to thank Dr. Reinhard Mahnkopf for his support. Beside the Infineon management I also want to acknowledge my co-workers Georg, Amal, Franz, Jens, Koen, Michael and Nickolay for the good collaboration.

As a member of the electrical characterization team my colleagues Ben, Felice, Guilherme, Jörgen, Luca, Luigi, Lars, Maarten, Rick, Robin, Salvatore, Philippe, Thomas and Xu are greatfully acknowledged for the many stimulating discussions in their respective fields of interest.

Furthermore, I want to thank the colleagues from the IMEC prototyping line (P-lijn) for device fabrication and the AMSIMEC lab for accommodating a flexible environment for the development of new measurement techniques.

Representatives for device integration, high-ε material deposition and physical characterization the corresponding work-package leaders Stefan Kubicek, Matty Caymax and Wilfried Vandervorst are mentioned. Their contributions have been essential for the successful fabrication of the devices studied here.

On the other hand, our counterparts at International Sematech are also acknowledged for discussion and sample fabrication. A special thanks to Yudong Kim (Intel), George Brown and Chadwin Young for their contribution.

Beside the contribution from the IMEC high-ε team members my colleagues from International SEMATECH at IMEC are greatfully acknowledged for their contribution in the respective fields and, moreover, for taking the lead in several high-ε dielectric related integration issues. The success of this work is to a large extent a result of their effort. Edward, Jerry, Jon, Marty, Se-Aug, Steven, Vincent, Vidya and Wilman thanks for your contribution to the IIAP program.

The additional IIAP members are also acknowledged for their contribution to the high-ε program (Yukiko and Yasuhiro from Hitachi, Masaaki and Shigenori from Matsushita and Kwak from Samsung).

A few collaborators outside the IMEC IIAP program I would like to thank for their contribution. Yuegang Zhao from Keithley Instruments for providing the software for Charge Trapping measurements using the Keithley semiconductor parameter analyzer and Thomas Pompl from Infineon Technolgies for the program used to calculate the tunneling current through arbitrary potential barriers.

Finally, there is one colleauge who deserves a distinct acknowledgement. Ed Cartier from IBM, he was not only contributing to many of the aspects discussed here, but also guiding me in the difficult moments of my research phase. Ed, thanks to you and your family for the time we could spend together.

There are a few people outside IMEC, International SEMATECH and Infineon I would like to mention. A special thanks to mom and dad for educating me and providing me with the necessary base to start my PhD. Your faith and support was of great help throughout the years. Beside the parents I also want to acknowledge my siblings and in particular my brother Martin, who introduced me to the exciting field of semiconductor devices. Without your help I would not have been able to accomplish this in such short period of time.

Andreas Kerber

Leuven, Oct. 2003

Abstract

The great commercial success of the microelectronics industry over more then 30 years is to a large extent based on the unique properties of SiO_2, which is grown by thermal oxidation. However, the aggressive scaling of Complementary Metal Oxide Semiconductor (CMOS) devices is driving SiO_2 based gate dielectrics to its physical limits as stated in the International Technology Roadmap for Semiconductors (ITRS). Currently several alternative dielectric materials like Al_2O_3, ZrO_2, HfO_2, Y_2O_3, La_2O_3 and Pr_2O_3, as well as mixed oxides containing SiO_2 (silicates) or Al_2O_3 (aluminates) and metals like Hf, Zr, Y or La are being studied extensively. In order to successfully integrate a gate dielectric material with higher dielectric constant (ε) into a state-of-the-art CMOS technology, its electrical device performance (carrier mobility, stability, reliability) has to match or exceed that of SiO_2.

This work mainly discusses charge trapping effects and the dielectric reliability of SiO_2 / Al_2O_3 and SiO_2 / HfO_2 dual layer gate dielectrics. It will be shown, that both gate stacks suffer from severe charge trapping, which prevents their successful application in MOSFET devices at present.

In SiO_2 / Al_2O_3 dual layer stacks with TiN gate electrodes, polarity dependent charge trapping and defect generation is observed. Due to differences in workfunction between the TiN gate electrode and the Si substrate and the inherent asymmetry of a dual layer gate dielectric, significantly lower injection fields at fixed current densities are being extracted for substrate injection than for gate injection. This is reflected in a strongly asymmetric charge trapping behavior. Electron trapping in the bulk of the Al_2O_3 film dominates for substrate injection, whereas positive charge trapping near the Si substrate is observed in case of gate injection. For substrate injection, no distortion of the quasi static C-V characteristics was observed, whereas a rapid build up of interface states occurs for gate injection. This asymmetry in defect creation also causes an asymmetry in dielectric reliability. For gate injection, reliability is limited by defect generation near the Si substrate, yielding a low Weibull slope (β), independent of the Al_2O_3 thickness. In case of substrate injection, reliability is limited by electron trap generation in the bulk of the Al_2O_3 film, yielding a strong thickness dependence of the β values, as expected from the percolation model.

Furthermore the impact of Post Deposition Annealing (PDA) on the charging effects in SiO_2 / Al_2O_3 gate stacks has been investigated. The qualitative trapping behavior after high temperature PDA in N_2 did not change. For substrate injection, differences in the voltage dependence of the electron trapping are seen and from "current versus time" traces the formation of shallow defects are evident, which cause transient charge loss after dielectric stress. The reliability parameters of SiO_2 / Al_2O_3 gate stacks are also being affected by the modified trapping behavior due to the high temperature PDA, however, the general trends observed in *as deposited* layers remain significant.

The studies on SiO_2 / HfO_2 dual layer stacks were mainly carried out on samples with poly-Si electrodes. Charging instabilities were observed similar to the ones of SiO_2 / Al_2O_3 gate stacks with a high temperature PDA. Electron trapping

for positive gate bias is the dominant charging mechanism in n-channel MOSFETs. To better quantify the fast-transient charging instability in high-ε gate dielectrics time resolved electrical characterization techniques in the μs time range were introduced and compared. Both charge pumping and pulsed I_D-V_D measurements yield similar results when the device geometry and the measurement parameters are chosen carefully. The impact of high temperature annealing in N_2 and O_2 on the trapping behavior of SiO_2 / HfO_2 dual layer gate dielectrics was studied. No significant improvement in the charge trapping was obtained by the PDA.

Based on the experimental results a defect band in the HfO_2 layer is postulated, which is located above the conduction band of the Si. For samples with a thin interfacial SiO_2 layer efficient charging and discharging of the defects is obtained for positive and negative bias, respectively, which is caused by the rapid motion of the defect band in energy due to the large differences of the dielectric constant in SiO_2 and HfO_2.

Furthermore, an alternative technique was proposed to directly measure the inversion charge in MOSFETs. In Inversion Charge Pumping (ICP) the geometric component in charge pumping is maximized and when the device symmetry is taken into account the full inversion charge can be extracted. By combining the ICP technique together with pulsed I_D-V_G measurements the carrier mobility was extracted under conditions where the influence of charge trapping is significantly reduced. The results suggest that charge trapping and the fixed charge in SiO_2 / HfO_2 dual layer gate stacks are not the primary cause for the severe mobility degradation compared to the SiO_2 reference.

The assessment of the breakdown behavior of SiO_2 / HfO_2 dual layer stacks was strongly aggravated by the poly-Si integration issues. Significant yield loss was observed on large area capacitors which shows a strong dependence on the device fabrication process. A preliminary reliability assessment indicates that similar degradation processes compared to SiO_2 / Al_2O_3 gate stacks apply. For gate injection interface degradation is dominant whereas for substrate injection bulk defects in the HfO_2 are being formed.

In order to successfully integrate high-ε materials into future CMOS technologies strong improvements in the device performance (carrier mobility) and the charge trapping (V_T-instability) behavior are required based on the results shown here. The improvements may come from varying the deposition process, the deposition temperature and possibly the post-deposition treatment. Nitrogen and silicon incorporation into the high-ε dielectric may also improve the stability and performance of CMOS devices, based on recent results reported in the literature. To correctly quantify the magnitude of the instability and its impact on the device performance pulsed measurement techniques were introduced and demonstrated. The same characterization tools are applicable to monitor progress of high-ε development and optimization towards the application in future CMOS technologies.

Zusammenfassung

Der grosse Erfolg der Mikroelektronik in den vergangenen 30 Jahren beruht großteils auf den hervorragenden Eigenschaften von SiO_2 welches durch thermische Oxidation hergestellt wird. Aufgrund der raschen Miniaturisierung von komplementären Metall-Oxyd-Halbleiter (CMOS) Schaltkreisen werden jedoch in naher Zukunft die physikalischen Grenzen von Dielektrika auf SiO_2 Basis erreicht, wie es in der Planung für zukünftige Halbleitertechnologien vorhergesagt wird. Verschiedene dielektrische Materialien, wie Al_2O_3, ZrO_2, HfO_2, Y_2O_3, La_2O_3 und Pr_2O_3, als auch gemischte Oxide welche SiO_2 (Silikate) oder Al_2O_3 (Aluminate) und Metalle wie Hf, Zr, Y oder La beinhalten, werden derzeit intensiv untersucht. Für eine erfolgreiche Integration von Materialien mit hoher Dielektrizitätskonstante (ε) in eine zukünftige CMOS Technologie müssen vergleichbare oder bessere elektrische Eigenschaften (Ladungsträgerbeweglichkeit, Stabilität, Zuverlässigkeit) im Vergleich zu SiO_2 Schichten erzielt werden.

In dieser Arbeit wird vorwiegend der Ladungseinfang und die Zuverlässigkeit von SiO_2 / Al_2O_3 und SiO_2 / HfO_2 Schichten diskutiert. Es wird gezeigt, dass die elektrischen Eigenschaften beider Materialsysteme ernsthaft vom Ladungseinfang beeinträchtigt sind, welcher deren erfolgreiche Anwendung derzeit verhindern.

In SiO_2 / Al_2O_3 Schichten mit TiN Metall Elektroden wird polaritätsabhänger Ladungseinfang und Defektgeneration beobachtet. Aufgrund der Differenz in der Austrittsarbeit zwischen TiN und dem Si Substrat und der inhärenten Asymmetrie mehrschichtiger Dielektrika, sind signifikant niedrigere Injektionsfelder für Substratinjektion im Vergleich zur Gateinjektion extrahiert worden, welches sich in einem stark asymmetrischen Ladungseinfang widerspiegelt. Elektroneneinfang im Al_2O_3 ist dominant für Substratinjektion, während für Gateinjektion positive Ladungen in der Nähe der Grenzfläche zum Si Substrat generiert werden. Für Substratinjektion wird keine Störung der quasi-statischen C-V Charakteristik beobachtet. Gateinjektion hingegen führt zu einer raschen Generation von Grenzflächenzuständen. Die Asymmetrie in der Defektgeneration spiegelt sich auch in der Zuverlässigkeit wider. Wenn Ladungsträger vom Gate injiziert werden, bestimmt die Defektgeneration in der Nähe der Grenzfläche zum Si Substrat die Zuverlässigkeit, was einen niedrigen β Wert unabhängig von der Dicke der Al_2O_3 Schicht zur Folge hat. Im Fall von Ladungsträgerinjektion vom Si Substrat wird die Zuverlässigkeit von der Defektgeneration im Al_2O_3 bestimmt, was zu einer Dickenabhängigkeit des Formfaktors (β) in der Weibullverteilung führt. Die Dickenabhängigkeit ist in Übereinstimmung mit dem "percolation model".

Des weiteren wurde der Einfluss von Temperschritten in Stickstoff (N_2) Atmospäre nach der Abscheidung der Al_2O_3 Schicht auf den Ladungseinfang in SiO_2 / Al_2O_3 Schichten untersucht. Das qualitative Ladungsverhalten wird durch den Temperschritt in Stickstoff nicht verändert. Für Substratinjektion wird ein modifiziertes Ladungsverhalten als Funktion der Gate-Spannung beobachtet und von den Strom - Zeit Kurven kann auf die Generation von flachen Störstellen geschlossen werden. Die Generation von flachen Störstellen verursacht einen transienten Ladungsverlust wenn der Stress unterbrochen wird. Der Temperschritt zeigt auch Auswirkungen auf die Zuverlässigkeitsparameter von SiO_2 / Al_2O_3 welche auf die Generation

von flachen Störstellen und das damit modifizierte Ladungsverhalten zurückgeführt werden können.

Untersuchungen von SiO_2 / HfO_2 Schichten wurden hauptsächlich an Proben mit Poly-Si Elektroden durchgeführt. Ähnliche Ladungsinstabilitäten wurden im Vergleich zu getemperten SiO_2 / Al_2O_3 beobachtet. Elektroneneinfang für positive Gate-Spannung ist der dominante Ladungsmechanismus in n-Kanal MOSFETs. Um die schnellen, transienten Ladungseffekte in Materialien mit hoher Dielektrizitätskonstante besser beurteilen zu können wurden Messverfahren entwickelt, welche es erlauben Instabilitäten im μs Bereich zu messen. Mit beiden Messverfahren (Charge Pumping und gepulste I_D-V_G Charakterisierung) werden vergleichbare Ergebnisse erzielt wenn die Messparameter und Strukturen entsprechend gewählt werden. Außerdem wird eine alternative Methode entwickelt, mit welcher die Ladungsträgerdichte im Inversionskanal des Transistors direkt gemessen werden kann. Bei Inversion Charge Pumping (ICP) wird der Beitrag des geometrischen Effekts in Charge Pumping maximiert, und unter Hinzunahme der Struktursymmetrie kann die gesamte Inversionsladung bestimmt werden. Wenn ICP mit der gepulsten I_D-V_G Charakterisierung kombiniert wird, kann die Ladungsträgerbeweglichkeit bei Bedingungen mit stark reduziertem Ladungseinfang bestimmt werden. Die Ergebnisse zeigen, dass der Ladungseinfang und die festen Oxidladungen nicht für die starke Reduktion der Ladungsträgerbeweglichkeit im Inversionskanal verantwortlich sind.

Der Einfluss von Temperschritten in Stickstoff und Sauerstoff auf den Ladungseinfang in SiO_2 / HfO_2 Schichten wurde ebenfalls untersucht. Die Temperschritte haben keine signifikante Verbesserung im Ladungsverhalten ergeben. In Bezug auf die elektrischen Ergebnisse wurde eine Defektband in der HfO_2 Schicht postuliert, welches sich oberhalb der Leitungsbandkante von Si befindet. Das Defektband im HfO_2 kann effizient mit Elektronen vom n-Typ Substrat oder vom Inversionskanal im n-Kanal MOSFET unter positiver Gate-Spannung geladen und unter negativer Gate-Spannung entladen werden. Der effiziente Lade- und Entladevorgang wird durch die rasche Energieverschiebung des Defektbands aufgrund der stark unterschiedlichen Dielektrizitätskonstante zwischen SiO_2 und HfO_2 hervorgerufen.

Die Beurteilung des Durchbruchverhaltens von SiO_2 / HfO_2 Schichten wurde durch die Intgrationsprobleme mit Poly-Si Elektroden stark erschwert, welche eine signifikante Reduktion der Ausbeute an großflächigen Strukturen zur Folge hatte. Durch Veränderungen im Herstellungsprozess konnte die Ausbeute stark verbessert werden. Vorläufige Untersuchungen zeigen, dass ähnliche Degradationsechanismen wie für SiO_2 / Al_2O_3 Schichten zutreffen. Die Generation von Grenzflächenzuständen ist dominant für die Ladungsträgerinjektion von der Gate-Elektrode und bei der Substratinjektion werden Defekte in der HfO_2 Schicht erzeugt.

Für eine erfolgreiche Integration von Materialien mit hoher Dielektrizitätskonstante in zukünftige CMOS Technologien ist, basierend auf den hier gezeigten Ergebnissen, eine starke Verbesserung der elektrischen Eigenschaften (Ladungsträgerbeweglichkeit, Ladungseinfang) zwingend erforderlich. Die Verbesserungen können entweder von Änderungen im Abscheideprozess, der Abscheidetemperatur oder den Temperschritten einsetzen. Der Einbau von Stickstoff und Silizium in Metalloxide kann auch zur Stabilität und zur Verbesserung der Ladungsträgerbeweglichkeit beitragen, wie jüngste Ergebnisse in der Literature belegen. Für eine korrekte Abschätzung der Ladungsinstabilitäten und deren Einfluss auf die Ladungsträgerbeweglichkeit wurden gepulste Messverfahren vorgestellt und erfolgreich angewandt. Dieselben Messverfahren sind auch zur Prozesskontrolle notwendig um Fortschritte auf dem Weg zur Anwendung dieser Materialien in zukünftigen CMOS Technologien zu verifizieren.

List of abbreviations

Abbreviations	Description
AC	Alternating Current
ALD	Atomic Layer Deposition
Al	Aluminum
Al_2O_3	Aluminum Oxide
CCS	Constant Current Stress
CHCD	Channel Hot Carrier Degradation
CHCI	Channel Hot Carrier Injection
CMOS	Complementary Metal Oxide Semiconductor
CMP	Chemical Mechanical Polishing
C-P	Charge Pumping
C-V	Capacitance Voltage Characteristic
CVD	Chemical Vapor Deposition
CVS	Constant Voltage Stress
Cl	Chlorine
DC	Direct Current
DT	Direct Tunneling
EOT	Equivalent Oxide Thickness
FNT	Fowler-Nordheim Tunneling
HBD	Hard-Breakdown
HF	Hydrofluoric Acid
HfO_2	Hafnium Oxide
HfSiO	Hafnium Silicate
HfSiON	Hafnium Silicon Oxynitride
HP	High Performance
H_2O	Water
H_2O_2	Hydrogen Peroxide
IC	Integrated Circuit
ICP	Inversion Charge Pumping
ILD	Inter Level Dielectric
I-V	Current Voltage Characteristic
I_D-V_D	Drain Current - Drain Voltage Characteristic
I_D-V_G	Drain Current - Gate Voltage Characteristic
K	Potassium
LCR	Gain / Phase Analyzer for Extraction of Inductance or Capacitance and Resistance
LDD	Lightly Doped Drain
LOCOS	Local Oxidation Of Silicon
LSP	Low Standby Power
MBE	Molecular Beam Epitaxy
MOCVD	Metal Organic Chemical Vapor Deposition
MOS	Metal Oxide Semiconductor
MOSFET	Metal Oxide Semiconductor Field Effect Transistor

Continued on next page

Continued from previous page

Abbreviations	Description
Na	Sodium
NBTI	Negative Bias Temperature Instability
O_3	Ozone
OH	Oxygen - Hydrogen complex
PBL	Poly Buffered LOCOS Isolation
PBTI	Positive Bias Temperature Instability
PDA	Post Deposition Anneal
PMA	Post Metallization Anneal
PSG	Phosphosilicate Glass
PVD	Physical Vapor Deposition
Q_{BD}	Charge To Breakdown
RIE	Reactive Ion Etch
RTO	Rapid Thermal Oxidation
RTP	Rapid Thermal Processing
SiON	Silicon Oxynitride
SiO_2	Silicon Dioxide
SBD	Soft-Breakdown
T_{BD}	Time To Breakdown
TDEAH	Tetrakis-Diethylamido-Hafnium
TDMAS	Tetrakis-Dimethylamido-Silicon
TEM	Transmission Electron Microscopy
TEOS	Tetraethyl Orthosilicate (SiO_2)
Ti	Titanium
TiN	Titanium Nitride
$TiSi_2$	Titanium Silicide
VLSI	Very Large Scale Integration
W	Tungsten
ZrO_2	Zirconium Oxide

List of symbols

Symbol	Unit	Description
A	cm^2	Area
β		Shape parameter of the Weibull distribution
C	$\mathrm{F/cm}^2$	Gate capacitance per unit area
C_{OX}	$\mathrm{F/cm}^2$	Oxide capacitance per unit area
C_S	$\mathrm{F/cm}^2$	Substrate capacitance per unit area
C_{acc}	$\mathrm{F/cm}^2$	Total gate capacitance per unit area measured in accumulation at a specified V_G
C_{inv}	$\mathrm{F/cm}^2$	Inversion capacitance per unit area
C_{it}	$\mathrm{F/cm}^2$	Interface state capacitance per unit area
D	cm^2/s	Diffusion constant
$\mathcal{D}$		Dissipation factor
$D(\mathcal{E})$		Transmission probability
$\overline{D_{it}}$	$1/(\mathrm{cm}^2\mathrm{eV}^1)$	Average interface trap density per unit area and energy
$D_{it}(\mathcal{E})$	$1/(\mathrm{cm}^2\mathrm{eV}^1)$	Energy distribution of interface traps
d_{int}	m	Physical distance of the positive from the interface between SiO_2 and Al_2O_3
ΔV_{FB}	V	Flatband voltage shift
ΔV_{I-V}	V	Shift in the gate current - voltage characteristic
ΔV_{C-V}	V	Shift in the capacitance - voltage characteristic
ΔV_{MG}	V	Midgap voltage shift extracted at $C_{acc}/2$
ΔV_T	V	Threshold voltage shift
$\mathcal{E}$	J	Energy
E_{OX}	V/m	Oxide field
E_{Si_max}	V/m	Maximum field in the Si substrate during a hysteresis measurement
e_e	s^{-1}	Electron emission rate
$\varepsilon_{Al_2O_3}$		Dielectric constant of Al_2O_3
ε_{HfO_2}		Dielectric constant of HfO_2
ε_{OX}		Dielectric constant of the gate oxide
ε_{Si}		Dielectric constant of Si
ε_{SiO_2}		Dielectric constant of SiO_2
ε_{hk}		Dielectric constant of the high-ε material
ε_{int}		Dielectric constant of the interfacial layer
ε_0	F/m	Vacuum permittivity
$F(t)$		Cumulative failure distribution
f	s^{-1}	Frequency
$f(t)$		Failure probability density function
$\hbar$	Js	Reduced Planck's constant
η		Scale parameter of the Weibull distribution
G_e	$1/\mathrm{C\,cm}$	Trap generation rate per unit charge
G_P	A/V	Conductance of the parallel circuit

Continued on next page

Continued from previous page

Symbol	Unit	Description
G_t	$1/(\mathrm{cm}^3\sqrt{s})$	Trap generation rate per unit time
g_m	A/V	Channel transconductance
γ	1/V	Voltage acceleration factor
I_{CP}	A	Recombination current measured at the substrate in a C-P measurement
I_D	A	Drain current
I_G	A	Gate current
I_S	A	Source current
I_{S+D}	A	Current measured at Source and Drain in a C-P measurement
J_{DT}	A/m^2	Current density for direct tunneling
J_{FN}	A/m^2	Current density for Fowler-Nordheim tunneling
J_e	A/cm^2	Electron current density
k	J/K	Boltzmann constant
I_{Sub}	A	Substrate current
L	m	Channel length of the MOSFET
L_D	m	Debye length
m^*	kg	Effective electron mass
$m^*_{Al_2O_3}$	kg	Effective electron mass in the bandgap of Al_2O_3
$m^*_{HfO_2}$	kg	Effective electron mass in the bandgap of HfO_2
$m^*_{SiO_2}$	kg	Effective electron mass in the bandgap of SiO_2
m_e	kg	Electron mass
μ_{CS}	cm^2/Vs	Coulomb scattering mobility
μ_{PS}	cm^2/Vs	Phonon scattering mobility
μ_{SRS}	cm^2/Vs	Surface roughness scattering mobility
μ_{eff}	cm^2/Vs	Effective carrier mobility
N_{CP}	cm^{-2}	Number of carriers per cycle and unit area measured in the substrate by C-P (N_{CP}=I_{CP}/(f·A·q))
N_{S+D}	cm^{-2}	Number of carriers per cycle and unit area measured at Source / Drain by C-P
N_0	cm^{-3}	Initial trap density per unit volume
N_A	cm^{-3}	Acceptor concentration per unit volume
N_D	cm^{-3}	Donor concentration per unit volume
$N_{OT}(t)$	cm^{-3}	Number of generated traps as a function of time
N_{dep}	cm^{-2}	Depletion carrier concentration
N_{inj}	cm^{-2}	Number of injected carriers per unit area
N_{inv}	cm^{-2}	Inversion carrier concentration
N_{it}	cm^{-2}	Number of interface traps measured with charge pumping
N_s	cm^{-2}	Surface carrier concentration per unit area
N_{tr}	cm^{-3}	Number of trapped carriers per unit volume
$\overline{P}$		Apparent trapping probability
P		Trapping probability
Φ_B	eV	Barrier height
ϕ_F	V	Fermi level in the Si substrate
ϕ_S	V	Surface potential at the interface between the oxide and the Si substrate
$\mathcal{Q}$		Quality factor
Q_{C-V}	$\mathrm{C/cm}^2$	Charge density per unit area extracted from C-V sensing
Q_{inj}	$\mathrm{C/cm}^2$	Injected charge per unit area
Q_{it}	$\mathrm{C/cm}^2$	Trapped interface charge per unit area

Continued on next page

Continued from previous page

Symbol	Unit	Description
Q_{I-V}	C/cm^2	Charge density per unit area extracted from I-V sensing
Q_{OX}	C/cm^2	Effective trapped oxide charge per unit area
$\mathcal{Q}_P$		Quality factor of the parallel circuit
Q_S	C/cm^2	Substrate charge per unit area
$\mathcal{Q}_S$		Quality factor of the serial circuit
q	C	Electronic charge
$R(L)$	V/A	Load resistance of the inverter circuit
R_S	V/A	Resistance of the serial circuit
R_P	V/A	Resistance of the parallel circuit ($R_P = 1/G_P$)
$R(t)$		Reliability or survival function
σ_e	cm^2	Electron Capture Cross Section
σ_h	cm^2	Hole Capture Cross Section
T	K	Temperature
$t_{Al_2O_3}$	m	Aluminum dioxide thickness
t_{HfO_2}	m	Hafnium oxide thickness
T_{OX}	m	Oxide thickness
t_{hk}	m	Thickness of the high-ε material
t_{int}	m	Interfacial layer thickness
t_{SiO_2}	m	Silicon dioxide thickness
V_{Amp}	V	Amplitude of the voltage pulse in a conventional base level C-P measurement
V_{Base}	V	Base level of the voltage pulse in a conventional basel level C-P measurement or in a C-P measurement using the amplitude sweep
V_D	V	Drain bias
V_{FB}	V	Flatband voltage
V_G	V	Gate bias
V_{IN}	V	Input voltage applied to Source, Drain and Substrate of a MOSFET in pulsed C-V measurement
V_{OFFSET}	V	Offset voltage of the current-voltage amplifier in pulsed C-V measurement
V_{OUT}	V	Output voltage of the current-voltage amplifier in pulsed C-V measurement
V_{Peak}	V	Peak voltage level in a C-P measurement using the amplitude sweep
V_S	V	Source bias
V_{Sub}	V	Substrate bias
V_T	V	Threshold voltage
W	m	Channel width of the MOSFET
ω	s^{-1}	Angular frequency
$\overline{x}$		Normalized distance of the charge centroid from the gate electrode
Y_S	A/V	Admittance of the serial circuit
Y_P	A/V	Admittance of the parallel circuit
Z_{AMP}	V/A	'Gain' of the current-voltage amplifier
Z_S	V/A	Impedance of the serial circuit
Z_P	V/A	Impedance of the parallel circuit

Contents

1 Introduction

1.1 Brief summary of the history of semiconductor devices

In 1947 W. Shockley, J. Bardeen and W.H. Brattain invented the bipolar transistor at Bell Laboratories, Murray Hill, New Jersey. The invention of the transistor, a solid state amplifier, resulted in great efforts in the field of semiconductor devices. The integration of semiconductor devices on a single chip was one of the consequences of the combined efforts. J. Kilby at Texas Instruments first demonstrated the concept of Integrated Circuits (IC) in 1959. This concept together with the fabrication of the first Metal Oxide Semiconductor Field Effect Transistor (MOSFET) by D. Kahng and M.M. Attala in 1960 provided the basis for the evolution of the microelectronics industry. The principle of a surface field effect transistor was already proposed in the early 1930's by Lilienfeld and Heil. The experimental verification of the surface field effect, however, could not be demonstrated for more than 30 years. Since then, the MOSFET has become by far the most important electronic device for very large scale integrated (VLSI) circuits such as microprocessors and semiconductor memories. The minimum device dimension since the first demonstration of Integrated Circuits has been reduced from $\sim$10 μm in the 1960's to sub μm features in the 1990's. Currently, research is being conducted towards integration of sub 100 nm CMOS devices. The introduction of alternative gate dielectrics with a high dielectric constant (ε) is part of this research effort.

1.2 Scaling of CMOS devices

The early years of integrated circuits were strongly affected by process related issues like dopant diffusion, junction passivation and SiO_2 growth. The initial set of challenges related to the use of SiO_2 included the control of the growth process, the reduction of mobile charges, fixed charge and interface trap centers. Post Deposition Anneals (PDA) and Post Metallization Anneals (PMA) were developed in the 1960's to minimize both fixed and interface charge. The mobile charge carriers, such as Na and K, required stringent control. Phosphosilicate glass (PSG) was found to be an effective getter for mobile ions. The introduction of HCl in the oxidizing am-

bient resulted in strong improvements in the carrier lifetime by reducing metallic impurities and mobile ions.
Once the initial issues in growing oxide and controlling charge were solved, attention was directed to manufacturing and reliability issues. In modern technologies reliability considerations and direct tunneling gate leakage are the two main factors limiting the minimum dielectric thickness in CMOS technologies.

In the 1970's scaling of the device dimensions (channel length, oxide thickness and junction depth) was introduced to increase the device density and reduce the transistor costs. From the early 70's until the mid 90's the industry followed the so-called constant voltage scaling. In the constant voltage scaling mode the oxide voltage is kept constant whereas oxide thickness and device dimensions are reduced which degraded the oxide integrity due to increasing oxide fields and process improvements were soon required. Advances in cleanroom technology and wafer cleaning, along with optimized oxidation conditions, were implemented to improve the time-zero breakdown voltage distribution. The migration from aluminum gate electrodes to poly-Si gates resulted in dramatic improvements in reliability and later allowed dual workfunction gate electrodes for optimal CMOS performance. The constant voltage scaling in the late 1970's resulted in hot electron injection and trapping. Substantial improvements in the hot carrier charge trapping were made by introducing the Lightly Doped Drain (LDD) device structure. As the oxide quality improved, process induced degradation became more important. Nevertheless, the microelectronics industry managed to decrease the feature size and increase the device density per chip continuously for more then 3 decades, as shown in Fig. 1.1 and 1.2.

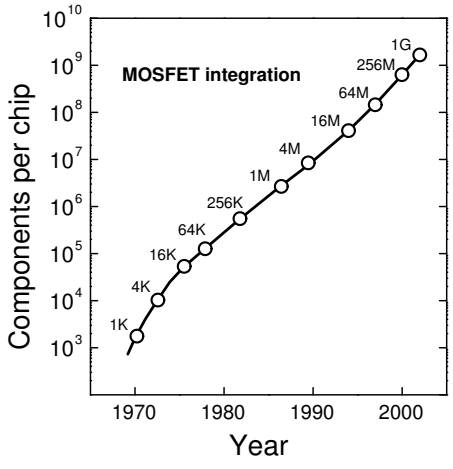

Figure 1.1: Number of components per chip versus year of introduction. Since the early 70's the device density increased by $\sim$ 100x per decade.

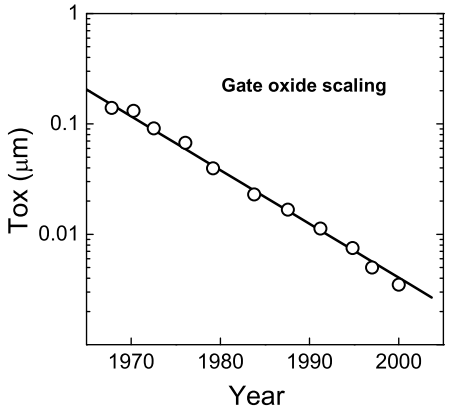

Figure 1.2: Evolution of gate oxide thickness versus year of introduction.

The scaling requirements for future CMOS technologies is generally guided by the International Technology Roadmap for Semiconductors (ITRS) [1], where the introduction of alternative gate dielectrics is predicted for 2005 to 2007 depending on the technology application as shown in Fig. 1.3 and 1.4. For the High Performance (HP) technologies dielectric scaling is more aggressive and will reach the sub 1 nm regime in the near future. On the other hand the leakage current requirements are more relaxed and the introduction of high-ε dielectrics is not expected before 2007.

Due to the stringent leakage requirements for Low Standby Power (LSP) devices the leakage current specifications can not be met with conventional gate dielectrics in the sub 1.5 nm Equivalent Oxide Thickness (EOT) regime. Therefore, it's more likely that high-ε dielectrics are first introduced in LSP technologies. The dielectric of choice will likely be deposited by either a chemical vapor deposition technique or one of the many other methods under investigation like physical vapor deposition, jet vapor deposition or molecular beam epitaxy.

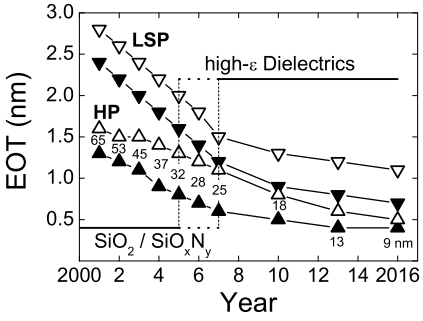

Figure 1.3: Equivalent Oxide Thickness (EOT) versus year of introduction for both High Performance and Low Standby Power technologies. For comparison the minimum (solid symbols) and maximum (open symbols) EOT is shown for each technology generation. The year of introduction for high-ε dielectrics is predicted between 2005 and 2007.

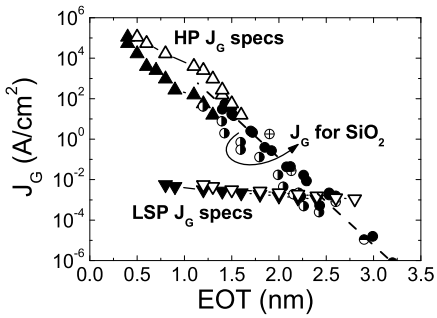

Figure 1.4: Gate leakage current specification versus EOT for High Performance and Low Standby Power technologies. The same symbols for the minimum and maximum EOT are used as in Fig. 1.3. Experimental data for SiO_2 from different companies are included to demonstrate the necessity for high-ε dielectrics for LSP technologies in the near future.

1.3 Requirements for integration of alternative gate dielectrics

Historically, the requirements for gate dielectrics are based on the unique properties of thermally grown SiO_2. For integration of alternative gate dielectrics with a high dielectric constant (ε) into future CMOS technologies these stringent requirements have to be satisfied. The entry point for high-ε gate dielectrics remains a moving target. Depending on the technology application an EOT of 1 nm or less will be required. The SiO_2-equivalent oxide thickness can be expressed as $EOT = (\varepsilon_{SiO_2}/\varepsilon_{int})t_{int} + (\varepsilon_{SiO_2}/\varepsilon_{hk})t_{hk}$, where the index 'hk' refers to the high-ε layer and the index 'int' refers to a possible interfacial layer between the Si substrate and the high-ε film. The formation of such interfacial layer prior, during and after deposition need to be minimized to obtain an EOT of less than 1 nm. Therefore, it is of high interest to deposit the high-ε material either on a hydrogen terminated Si surface as obtained after an HF clean or on a 'thin' intentionally grown SiO_2 layer. The latter leaves less room for an additional high-ε film to reduce the gate leakage current.

Another requirement for high-ε gate dielectrics is the thermodynamic stability in contact with Si up to temperatures required for CMOS integration. The maximum temperature during device fabrication depends on the integration scheme, however,

for a conventional CMOS process a dopant activation temperature of $\sim 1000°C$ is likely required. It is essential to avoid the formation of silicides during the deposition and further device processing. Furthermore, it is also desirable that these materials are compatible with a poly-Si gate process. The dopant penetration issue related to a poly-Si gate process has also to be considered. Nitrogen incorporation might be a viable solution to suppress dopant penetration.

1.4 Electrical requirements for alternative gate dielectrics in future CMOS technologies

The primary aim of introducing high-ε gate dielectrics is to reduce the leakage current. The main factors which determine the leakage current through an insulator are the barrier height and the physical thickness of the layer. To obtain low leakage currents in high-ε dielectrics barrier heights of 1.5 eV or higher combined with a significantly higher dielectric constant compared to SiO_2 are required. Theoretical calculations predict that significant benefit in gate leakage can be expected for high-ε dielectrics like Al_2O_3, ZrO_2 and HfO_2.

The influence of the atomic structure on the electrical performance is being debated as well. An amorphous gate dielectric like thermally grown SiO_2 is considered to be more suitable compared to a poly-crystalline material. Issues of a poly-crystalline dielectric are related to the structural anisotropy e.g. in the dielectric constant and dopant diffusion along grain boundaries. In terms of leakage current the poly-crystalline structure of the gate dielectric may be tolerable.

Fixed charge and threshold voltage control are another important aspect when considering high-ε gate dielectrics. For a standard CMOS process, n- and p-degenerated poly-Si gate electrodes with work functions of 4 eV and 5 eV are used to control the threshold voltage, V_T, respectively. In MOS devices with high-ε dielectrics and poly-Si electrodes V_T is often found to deviate significantly from the values measured in devices with a SiO_2 control oxide. The observed V_T shifts are commonly attributed to fixed charge in the dielectric either located at the interface or distributed throughout the film. In order to control V_T a low fixed charge ($<10^{11}cm^{-2}$) is of importance.

Charge trapping and V_T instabilities are other points of attention for high-ε gate dielectrics. Recent experiments indicated that Negative Bias Temperature Instability (NBTI) and charge trapping due to homogenous gate stress is of concern for most of these materials. At this time little is known about Channel Hot Carrier Degradation (CHCD), however, due to the reduced bandgap of high-ε dielectrics they are expected to be prone to CHCD.

Achieving high carrier mobility is considered to be essential for integration. Again, SiO_2 devices serve as a baseline for devices with high-ε materials. From literature it is known that most high-ε devices suffer from severe mobility degradation. The cause for the degradation is still under debate. Scattering due to fixed charge in the dielectric or remote phonon scattering are proposed as origin for the mobility reduction. The experimental verification of the origin of the low carrier mobility, however, is still missing. In any case, control of fixed charge and understanding of the impact of remote phonons are essential for improving the carrier mobility in high-ε devices.

Finally, reliability is considered as a further challenge for high-ε materials. Initial experiments suggest that high gate reliability may be obtainable with these materials. However, detailed studies are required to investigate the impact of the atomic structure (amorphous versus crystalline), the interfacial layer, the stress polarity and the gate electrode on the reliability. Early learning is important to

understand the reliability limitations of high-ε gate dielectrics.

This work focuses on electrical properties of high-ε gate dielectrics. In particular charge trapping, defect generation and dielectric reliability of SiO_2 / Al_2O_3 and SiO_2 / HfO_2 dual layer gate stacks are being discussed.

2 Basic terms and definitions

To facilitate the discussion of the experimental data presented in this work the relevant basics of MOS devices are discussed in this theoretical chapter.

2.1 The MOS capacitor

2.1.1 Gate Current-Voltage (I-V) characteristic

The Current-Voltage characteristic of conventional gate dielectrics (SiO_2) is determined by the tunneling process through either a trapezoidal or a triangular potential barrier. To simulate the Current-Voltage characteristic of high-ε gate dielectrics tunneling through multi-layer stacks and possibly also a trap assisted conduction mechanism need to be considered. In the following, different concepts are briefly introduced.

Direct and Fowler-Nordheim tunneling

The Current-Voltage characteristics of MOS structures with conventional gate dielectrics is commonly analyzed using the simple band diagrams shown in Fig. 2.1 and 2.2 for direct tunneling (DT) and Fowler-Nordheim tunneling (FNT), respectively.

The difference between DT and FNT is the shape of the tunnel barrier. If the tunnel barrier is triangular, FNT occurs (see Fig. 2.2) while DT takes place through a trapezoidal barrier (see Fig. 2.1). The transition from FNT to DT occurs when the difference in potential energy over the oxide layer qV_{OX} becomes smaller than the tunnel barrier height Φ_B at the injecting interface. Below 4 nm oxide thickness (T_{OX}) DT becomes the dominant conduction mechanism.

The relationship between current density and oxide field during FNT is given by Eq. 2.1 [2].

$$J_{FN} = AE_{OX}^2 \exp\left(-\frac{B}{E_{OX}}\right) \tag{2.1}$$

The parameters A and B depend on the tunnel barrier height Φ_B and the effective

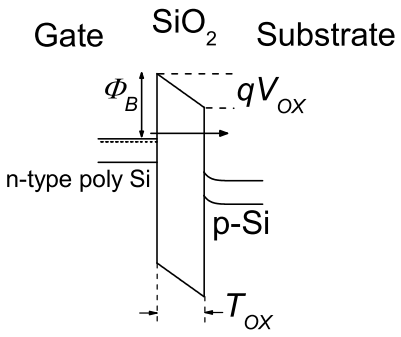

Figure 2.1: Energy band diagram for direct tunneling.

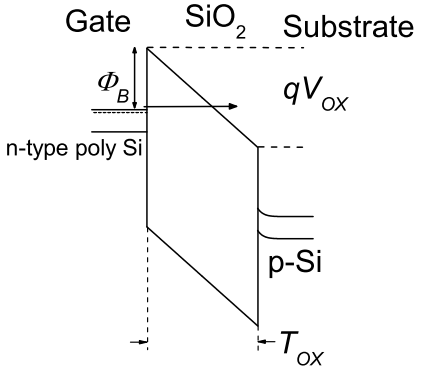

Figure 2.2: Energy band diagram for Fowler-Nordheim injection.

mass m^* given by:

$$A = \frac{q^3}{16\pi^2\hbar\Phi_B} \tag{2.2}$$

$$B = \frac{4}{3}\frac{(2m^*)^{1/2}}{q\hbar}\Phi_B^{3/2} \tag{2.3}$$

These parameters can be derived experimentally from a Fowler-Nordheim plot where $ln(J_{FN}/F_{OX})$ is plotted versus $1/F_{OX}$. The slope of this straight line gives B while A is the intercept.

A simplified expression for the DT current is given by Eq. 2.4 [3].

$$J_{DT} = \frac{AE_{OX}^2}{\left[1-\left(\frac{\Phi_B+qV_{OX}}{\Phi_B}\right)^{1/2}\right]^2}\exp\left[-\frac{B}{E_{OX}}\frac{\Phi_B^{3/2}-(\Phi_B-qV_{OX})^{3/2}}{\Phi_B^{3/2}}\right] \tag{2.4}$$

Tunneling through stacked dielectrics

The tunneling current through stacked dielectrics can be calculated based on the transmission probability across arbitrary potential barriers [4]. The potential barrier is split into a sequence of small segments as shown in Fig. 2.3 where the potential function U, the effective mass m^* and the permittivity are approximated by multi-step functions. The general wave function Ψ_j for an electron with energy E is given by

$$\Psi_j = A_j\exp(ik_jx) + B_j\exp(-ik_jx) \tag{2.5}$$

where

$$k_j = \sqrt{2m_j^*(\mathcal{E}-U_j)}/\hbar \tag{2.6}$$

and $\hbar$ is the reduced Planck's constant. From the continuity of $\Psi_j(x)$ and $(1/m^*)(d\Psi_j/dx)$ at each boundary, the determination of A_j and B_j in Eq. 2.5 is reduced to the multiplication of the following $N+1$ (2x2) matrices:

$$\begin{pmatrix}A_j\\B_j\end{pmatrix} = \prod_{l=0}^{j-1} M_l\begin{pmatrix}A_0\\B_0\end{pmatrix} \tag{2.7}$$

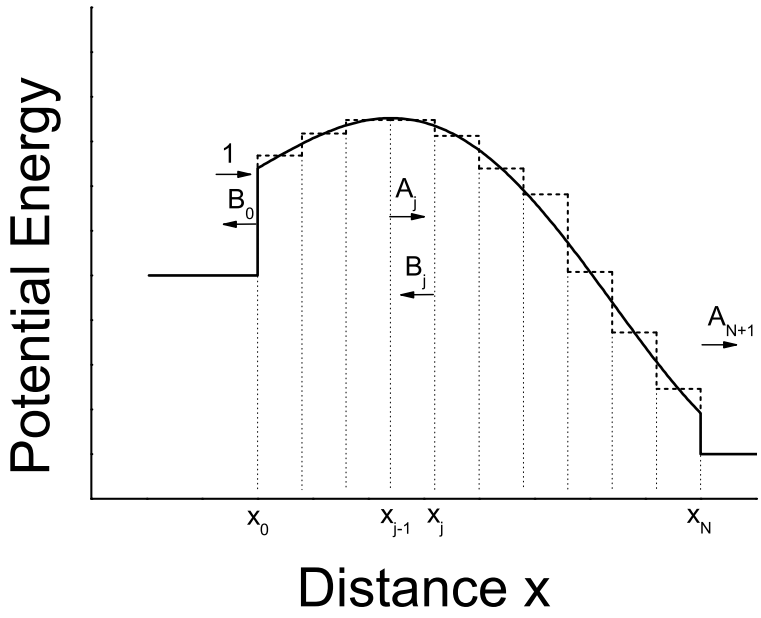

Figure 2.3: Energy band diagram (solid line) and approximated potential function (dashed line) for the potential barrier.

where

$$M_j = \frac{1}{2}\begin{bmatrix} (1+S)\exp\left[-i(k_{l+1}-k_l)x_l\right] & (1-S)\exp\left[-i(k_{l+1}+k_l)x_l\right] \\ (1-S)\exp\left[i(k_{l+1}+k_l)x_l\right] & (1+S)\exp\left[i(k_{l+1}-k_l)x_l\right] \end{bmatrix} \quad (2.8)$$

and

$$S_l = \frac{m^*_{l+1}}{m^*_l}\frac{k_l}{k_{l+1}}. \quad (2.9)$$

By setting $A_0 = 1$ and $B_{N+1} = 0$ in Eg. 2.7 for $j = N+1$ the transmission amplitude A_{N+1} and the transmission probability $D(\mathcal{E})$ are calculated as follows:

$$A_{N+1} = \frac{m^*_{N+1}}{m^*_0}\frac{k_0}{k_{N+1}}\frac{1}{M_{22}} \quad (2.10)$$

and

$$D(\mathcal{E}) = \frac{m^*_0}{m^*_{N+1}}\frac{k_{N+1}}{k_0}\left|A_{N+1}\right|^2 \quad (2.11)$$

where

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = \prod_{l=0}^{N} M_l. \quad (2.12)$$

The potential function $U(x)$ in the barrier is composed of superposition of the zero-bias potential distribution and the potential drop $\phi(x) = \int_0^x [qN_s/\varepsilon(x)]dx$ with $V_{OX} = \phi(x) - \phi(0)$ and N_s being the surface carrier concentration. The current density is given by

$$J(V_{OX}) = \frac{qm_0}{2\pi^2\hbar^3}\int_0^\infty \int_{\mathcal{E}_x}^\infty D(\mathcal{E}_x, V_{OX})\left[f_0(\mathcal{E}) - f_{N+1}(\mathcal{E})\right] d\mathcal{E}_x d\mathcal{E} \quad (2.13)$$

where $f_0(\mathcal{E})$ and $f_{N+1}(\mathcal{E})$ are the carrier distribution functions in the cathode and anode respectively. Examples for calculated tunneling currents through stacked dielectrics will be given in Section 4.1.

2.1.2 Capacitance-Voltage (C-V) characteristic

The Capacitance-Voltage behavior of a MOS device can be described using the equivalent circuit shown in Fig. 2.4 [5] where C_{OX} is the oxide capacitance, C_S the substrate capacitance, C_{it} the interface state capacitance, R_S the series resistance and $1/R_P$ the parallel conductance.

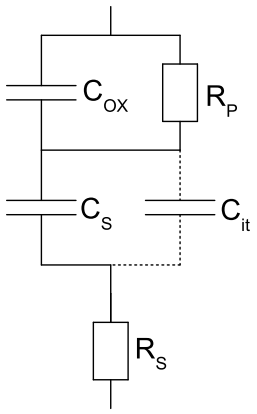

Figure 2.4: Equivalent circuit of an MOS structure including the oxide capacitance (C_{OX}), the substrate capacitance (C_S), interface state capacitance (C_{it}), series resistance (R_S) and parallel conductance ($1/R_P$)

The capacitance of a MOS capacitor is defined as

$$C = \frac{dQ_G}{dV_G}. \tag{2.14}$$

Based on charge neutrality $Q_G = -(Q_S + Q_{it})$ with Q_S being the substrate charge and Q_{it} being the trapped interface charge. This assumes no charge trapping in the dielectric. The gate voltage is partially dropped across the dielectric and partially across the semiconductor substrate. This gives $V_G = V_{FB} + V_{OX} + \phi_S$, where V_{FB} is the flatband voltage, V_{OX} the voltage drop across the oxide and ϕ_S the Si surface potential allowing Eq. 2.14 to be written as

$$C = \frac{dQ_S + dQ_{it}}{dV_{OX} + d\phi_S}. \tag{2.15}$$

Depending on the the Si surface potential either majority, minority or depletion charge is contributing to the substrate charge. The total gate capacitance can now be written as

$$C = \frac{1}{\frac{1}{C_{OX}} + \frac{1}{C_S + C_{it}}}. \tag{2.16}$$

The $low - frequency$ substrate capacitance is given by

$$C_{S,lf} = \hat{U}_S \frac{\varepsilon_{Si}\varepsilon_0}{2L_D} \frac{\left[e^{U_F}(1 - e^{-U_S}) + e^{-U_F}(e^{U_S} - 1)\right]}{F(U_S, U_F)} \tag{2.17}$$

where the dimensionless surface electric field $F(U_S, U_F)$ is defined by

$$F(U_S, U_F) = \sqrt{e^{U_F}(e^{-U_S} + U_S - 1) + e^{-U_F}(e^{U_S} - U_S - 1)}. \tag{2.18}$$

U_S and U_F are normalized potentials, defined as $U_S = q\phi_S/kT$ and $U_F = q\phi_F/kT$. The Fermi potential is calculated by $\phi_F = (kT/q)\ln(N_A/n_i)$ where N_A is the acceptor concentration and n_i the intrinsic carrier concentration in the Si substrate. The symbol $\hat{U}_S$ stands for the sign of the surface potential and is given by

$$\hat{U}_S = \frac{|U_S|}{U_S} \tag{2.19}$$

where $\hat{U}_S = 1$ for $U_S > 0$ and $\hat{U}_S = -1$ for $U_S < 0$. The extrinsic Debye length L_D is

$$L_D = \sqrt{\frac{\varepsilon_{Si}\varepsilon_0 kT}{2q^2 N_A}} \tag{2.20}$$

For scaled MOS devices with highly doped poly-Si electrodes a capacitance contribution from the gate electrode needs to be considered as well (e.g. gate depletion). Furthermore, when the oxide thickness is reduced below 10 nm quantum mechanical effects in the substrate and gate electrode have to be taken into account. Therefore, a C-V simulation tool from the North Carolina State University was used in this work to extract the EOT where a first order quantum mechanical correction is included [6].

High Frequency C-V measurement

In a High Frequency Capacitance-Voltage (HF C-V) measurement a small AC signal superimposed on a DC bias is applied to the MOS device and its response analyzed with respect to gain and phase. A schematic drawing of the HF C-V measurement setup is shown in Fig. 2.5.

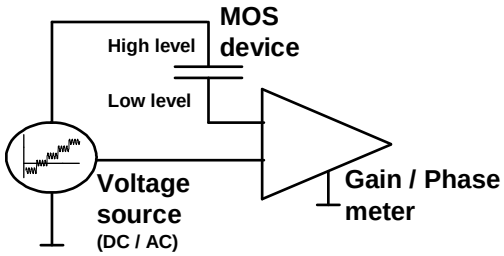

Figure 2.5: Schematic drawing of High Frequency setup. The DC and AC bias is applied to one terminal of the capacitor whereas the gain and phase of the AC signal is measured on the second terminal.

The gain / phase analysis is based on the equivalent circuits given in Figs. 2.6 and 2.7 where in addition to the measured capacitance either a parallel conductance term or a series resistance is considered. The impedance Z_S of the serial circuit is given by

$$Z_S = R_S + \frac{1}{i\omega C_S} \tag{2.21}$$

where R_S is the series resistance, $\omega = 2\pi f$ the angular frequency and C_S the serial capacitance. For the parallel circuit the admittance Y_P can be written as

$$Y_P = G_P + i\omega C_P \tag{2.22}$$

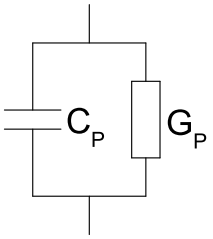

Figure 2.6: Equivalent circuit of a HF C-V taken in the parallel mode.

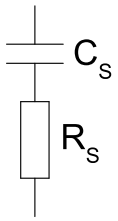

Figure 2.7: Equivalent circuit of a HF C-V taken in the serial mode.

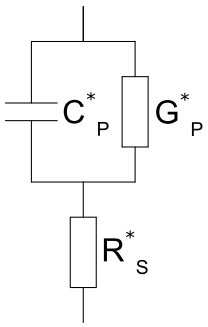

Figure 2.8: Simplified equivalent circuit of a MOS device with parallel conductance and serial resistance.

where G_P is the parallel conductance and C_P the parallel capacitance. To discuss the limitations of the equivalent circuits we transform the serial circuit into a parallel circuit and vice versa. The inverse of the serial impedance is given by

$$\begin{aligned} Y_S = \frac{1}{Z_S} = \frac{1}{R_S + \frac{1}{i\omega C_S}} &= \frac{1/R_S}{1+\mathcal{Q}_S^2} + \frac{i\omega C_S}{1+1/\mathcal{Q}_S^2} \equiv Y_P \\ G_P &= \frac{1/R_S}{1+\mathcal{Q}_S^2} \\ C_P &= \frac{C_S}{1+1/\mathcal{Q}_S^2} \end{aligned} \tag{2.23}$$

with $\mathcal{Q}_S = 1/\omega C_S R_S$ being the quality factor of the serial circuit. For the inverse of the parallel admittance we can write

$$\begin{aligned} Z_P = \frac{1}{Y_P} = \frac{1}{G_P + i\omega C_P} &= \frac{1/G_P}{1+\mathcal{Q}_P^2} + \frac{1/i\omega C_P}{1+1/\mathcal{Q}_P^2} \equiv Z_S \\ R_S &= \frac{1/G_P}{1+\mathcal{Q}_P^2} \\ C_S &= C_P\big(1+1/\mathcal{Q}_P^2\big) \end{aligned} \tag{2.24}$$

with $\mathcal{Q}_P = \omega C_P/G_P$ being the quality factor of the parallel circuit. For the case where $G_P \ll \omega C_P$ and $R_S \ll 1/\omega C_S$ both measurement modes yield equivalent results and the extracted capacitance becomes $C_P \equiv C_S$. On the other hand, if we consider the simplified equivalent circuit of the MOS devices given in Fig. 2.8 the impedance Z^* and admittance Y^* of the circuit are given by

$$\begin{aligned} Z^* = R_S^* + \frac{1}{G_P^* + i\omega C_P^*} &= R_S^* + \frac{1/G_P^*}{1+\mathcal{Q}_P^{*\,2}} + \frac{1/i\omega C_P^*}{1+\frac{1}{\mathcal{Q}_P^{*\,2}}} \\ \mathcal{Q}_P^* &= \omega C_P^*/G_P^* \\ Y^* = \frac{1}{Z^*} &= \frac{R_S^* + \frac{1/G_P^*}{1+\mathcal{Q}_P^{*\,2}} + \frac{i\omega C_P^*}{\omega^2 C_P^{*\,2}\left(1+1/\mathcal{Q}_P^{*\,2}\right)}}{\left(R_S^* + \frac{1/G_P^*}{1+\mathcal{Q}_P^{*\,2}}\right)^2 + \left(\frac{1/\omega C_P^*}{1+1/\mathcal{Q}_P^{*\,2}}\right)^2} \equiv Y_P \end{aligned}$$

$$C_P = \frac{\frac{C_P^*}{\omega^2 C_P^{*2}\left(1+1/\mathcal{Q}_P^{*2}\right)}}{\left(R_S^* + \frac{1/G_P^*}{1+\mathcal{Q}_P^{*2}}\right)^2 + \left(\frac{1/\omega C_P^*}{1+1/\mathcal{Q}_P^{*2}}\right)^2} \tag{2.25}$$

where C_P is the equivalent capacitance measured in the parallel mode [7]. The effect of the series resistance and the parallel conductance on the equivalent capacitance is illustrated in Fig. 2.9. For the parameters considered the capacitance is independent of the frequency over a large range of the spectrum. The limiting factor at higher frequencies is the series resistance whereas at low frequencies the parallel conductance has to be taken into consideration. In order to ensure accurate results from the instrument the dissipation factor $\mathcal{D} = 1/\mathcal{Q}$ should be below 0.1 [8].

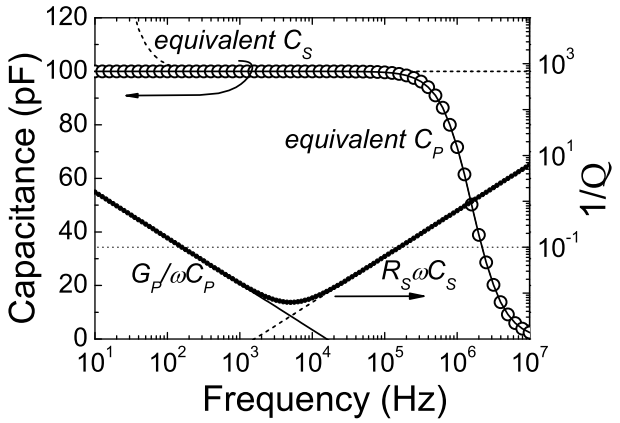

Figure 2.9: Frequency dependence of the equivalent capacitance using the circuit models given in Figs. 2.6, 2.7 and 2.8. The frequency dependence was calculated for $R_S = 1\,\mathrm{k\Omega}$, $G_P = 10\,\mathrm{nS}$ and a capacitance of 100 pF. For a wide frequency range the equivalent capacitance is independent of the circuit mode.

Quasi static C-V measurement

An alternative technique to measure the Capacitance-Voltage characteristic of MOS devices is the quasi static C-V measurement. This technique either measures the displacement current or displacement charge, which directly reflects the capacitance of the test structure. The strength of the quasi static C-V technique is its ability to measure interface states at the SiO_2 / Si interface for which the measurement has been used in the literature extensively throughout the past decades. However, since device scaling progresses the gate leakage current has become a severe problem for the quasi static C-V measurement.

Nevertheless, the different quasi static measurement techniques are going to be introduced and applied to thick high-ε gate dielectrics.

Quasi static C-V using a linear voltage ramp: One of the commonly used quasi static C-V measurement techniques applies a linear voltage ramp to one terminal of the device and measures the current as a function of time at the second

terminal as shown in Fig. 2.10. The capacitance is then given by the relation

$$C = \frac{J}{dV/dt}. \tag{2.26}$$

The voltage ramp dV/dt can be varied over a range from ∼10V/s down to <0.01V/s. The practical limitation of the voltage ramp is given by the sensitivity of the amperemeter and the leakage current of the measurement setup. The basic gate leakage requirements can be estimated as follows

$$J_G < C_{OX} \cdot dV/dt. \tag{2.27}$$

where for an EOT of $\sim 2\,\text{nm}$ ($1.5\,\mu F/cm^2$) and a voltage ramp of 1 V/s a leakage current of $J_G < 1.5\,\mu A/cm^2$ would be required.

Equilibrium controlled quasi static C-V: An alternative quasi static C-V technique is given by the equilibrium controlled quasi static C-V measurement [9]. In contrast to the linear voltage ramp technique discrete voltage steps are applied and the displacement charge is measured using the electrometer, as illustrated in Fig. 2.11. The capacitance is calculated from the measured displacement charge and the voltage step following the relationship in Eq. 2.28.

$$C(V) = \frac{\Delta Q}{\Delta V} = \frac{Q(V + \Delta V/2) - Q(V - \Delta V/2)}{\Delta V} \tag{2.28}$$

The limitation of the equilibrium controlled quasi static C-V measurement is similar to the one of the linear voltage ramp technique. Leakage currents of $J_G < 10nA/cm^2$ are required for accurate C-V measurements.

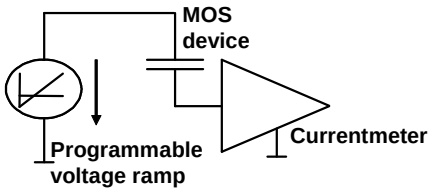

Figure 2.10: Schematic drawing of quasi static measurement setup using a linear voltage ramp.

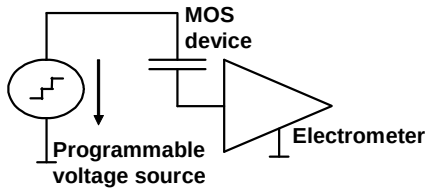

Figure 2.11: Schematic drawing of equilibrium controlled quasi static measurement setup where discrete voltage steps are applied and the displacement charge is measured using an electrometer.

2.2 The MOSFET

The Metal Oxide Semiconductor Field Effect Transistor (MOSFET) is based on the surface field effect. In Fig. 2.12 the basic structure of the four-terminal device is shown for a n-channel MOSFET. The basic device parameters are the channel length (L), which is the distance between the two metallurgical n^+-p junctions, the channel width (W), the dielectric thickness (T_{OX}), the junction depth and the substrate surface doping (N_A). When the gate bias (V_G) is below the threshold voltage (V_T) the source-to-drain current path includes two reversed biased p-n junctions and therefore the current is limited by the reverse leakage current. For $V_G \geq V_T$ a surface inversion layer is formed connecting source and drain through which a large current can flow. The channel conductance can be modulated by varying the gate bias.

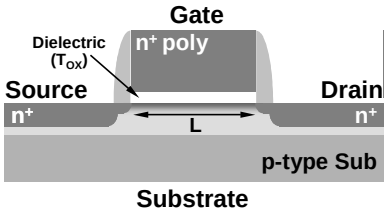

Figure 2.12: Schematic drawing of n-channel MOSFET

2.2.1 I_D-V_G characteristic

The drain current (I_D) versus gate voltage (V_G) characteristic of a MOSFET can be split into a linear and a saturation region. In the linear region the channel acts as a resistance and the drain current is proportional to the drain voltage, where the channel resistance is modulated by the gate voltage. Whereas for $V_D \geq (V_G - V_T)$ the drain saturates due to pinch-off in the channel at the drain side.

The drain current versus gate voltage characteristic in the linear regime $V_D \ll (V_G - V_T)$ can be written as

$$I_{Dlin} \simeq \frac{W}{L} \mu_{eff} C_{OX} (V_G - V_T) V_D \tag{2.29}$$

with μ_{eff} being the carrier mobility, C_{OX} the oxide capacitance and where V_T is given by

$$V_T = V_{FB} + 2\phi_F + \frac{\sqrt{2\varepsilon_{Si} q N_A (2\phi_F)}}{C_{OX}} \tag{2.30}$$

In the saturation regime the drain current follows

$$I_{Dsat} \simeq \frac{mW}{L} \mu_{eff} C_{OX} (V_G - V_T)^2 \tag{2.31}$$

where m is a function of the doping concentration and approaches $\frac{1}{2}$ for low doping levels [10]. In Figs. 2.13 and 2.14 typical I_D-V_G and I_D-V_D characteristics of n-channel MOSFETs with a 2 nm SiO_2 gate dielectric are shown for a selected device geometry. In addition to the drain current, the gate and substrate current are shown for comparison. The data in Fig. 2.13 clearly show the presence of gate leakage in inversion for the selected device geometry weakly depending on the drain bias. The substrate current, however, significantly increases when a large drain bias is applied due to hot carrier effects [11].

The channel transconductance at constant drain bias is further defined as

$$g_m = \left. \frac{\partial I_D}{\partial V_G} \right|_{V_D = const} \tag{2.32}$$

where in the linear regime g_m can be written as

$$g_m = \frac{W}{L} \mu_{eff} C_{OX} V_D \tag{2.33}$$

The carrier mobility μ_{eff} depends on the surface electric field, the substrate dopant concentration and temperature.

The drain current in the sub-threshold regime $V_G < V_T$ is dominated by diffusion of carriers from source to drain following the diffusion equation

$$I_D = -qAD \frac{dn}{dx} \tag{2.34}$$

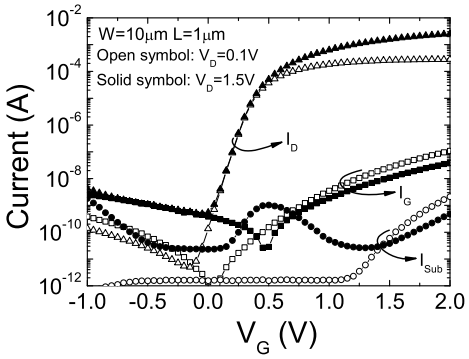

Figure 2.13: I_D versus V_G characteristic of a conventional n-channel MOSFET measured with a drain bias of 0.1V and 1.5V, respectively. In addition, gate (I_G) and substrate (I_{Sub}) current are shown.

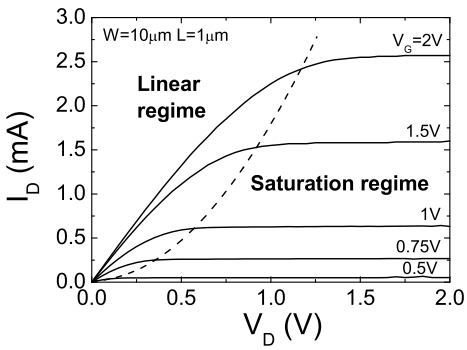

Figure 2.14: I_D versus V_D characteristic of a conventional n-channel MOSFET measured for different gate bias as indicated in the figure. The dashed line indicates the transition from the linear to the saturation regime.

where A is the cross section of current flow, D the diffusion constant, n the carrier concentration in the channel at the source side and x the distance from the source. Due to the exponential dependence of the carrier concentration on the surface potential the drain current drops exponentially with gate bias in weak inversion. The sub-threshold swing is defined as

$$S = \ln 10 \frac{dV_G}{d(\ln I_D)}. \tag{2.35}$$

2.2.2 Carrier mobility

The carrier mobility in the inversion layer of n- and p-channel MOSFETs is extracted by combining the I_D-V_G characteristic in the linear regime with C_{Inv}-V_G data. The effective carrier mobility is given by

$$\mu_{eff} = \frac{L}{W} \frac{I_D(V_G)}{V_D q N_{inv}} \tag{2.36}$$

where the inversion charge is either estimated using

$$N_{inv} = \frac{C_{OX}}{q}(V_G - V_T) \tag{2.37}$$

or directly obtained by integrating the gate to channel capacitance as derived from a HF C-V measurement where the substrate is grounded. This procedure is called the split C-V method.

$$N_{inv} = \frac{1}{q}\int^{V_G} C_{inv}(V_G')dV_G' \tag{2.38}$$

In the presence of depletion effects in the gate electrode the split C-V technique is a more reliable procedure to extract the inversion charge. The carrier mobility given in Eq. 2.36 [12] is usually plotted versus the effective surface field in the Si substrate which is given by

$$E_{eff} = \frac{q}{\varepsilon_{Si}\varepsilon_0}(N_{dep} + \eta N_{inv}) \tag{2.39}$$

where qN_{dep} is the surface concentration of the depletion charge and qN_{inv} the inversion charge. The factor η is given as 1/2 for electron mobility and 1/3 for hole mobility, respectively [12]. The depletion charge can either be determined experimentally, or for constant substrate doping, can be calculated using

$$N_{dep} = \sqrt{4\varepsilon_{Si}\varepsilon_0\phi_F\frac{N_A}{q}} \tag{2.40}$$

where

$$\phi_F = \frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right). \tag{2.41}$$

When plotting the carrier mobility versus the effective surface field the universal mobility behavior is attributed to different scattering mechanisms as shown in Fig. 2.15 [12]. At low effective fields Coulomb scattering (μ_{CS}) is dominant whereas at high fields surface roughness (μ_{SRS}) has the strongest impact. At moderate fields the carrier mobility is limited by phonon scattering (μ_{PS}), which has a strong temperature dependence. According to the Matthiessen rule the effective carrier mobility (μ_{eff}) can then be written as

$$\frac{1}{\mu_{eff}} = \frac{1}{\mu_{CS}} + \frac{1}{\mu_{PS}} + \frac{1}{\mu_{SRS}}. \tag{2.42}$$

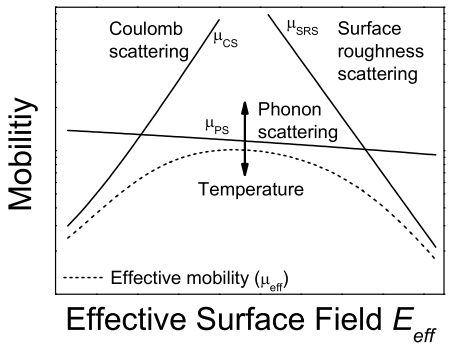

Figure 2.15: Universal mobility behavior versus the effective Si surface field. Coulomb scattering, phonon scattering and surface roughness scattering are the relevant scattering mechanism.

2.2.3 Charge Pumping (C-P)

The charge pumping effect in MOSFETs was first reported in the late 60's [13] long before a reliable interpretation of the experimental results could be given [14].

A schematic drawing of the setup used here is shown in Fig. 2.16. In charge pumping a pulse train is applied to the gate of a MOSFET which drives the surface potential from accumulation into inversion and vice versa. The charge pumping technique is based on the recombination process of charged interface (surface) states at the Si / SiO_2 interface which contributes to the substrate current. In Fig. 2.17.a) to d) the energy band diagram is shown for a n-channel MOSFET for one charge pumping cycle. In strong inversion, Fig. 2.17.a), interface traps are charged by electrons as indicated by the solid symbols. When the gate pulse changes from

positive to negative the MOS device switches from inversion to accumulation, Fig. 2.17.d). In the finite transition of the gate pulse inversion carriers drift back to the source / drain junction. In addition, trapped electrons near to the conduction band get thermally emitted to the conduction band and also drift to the source / drain junction (see Fig. 2.17.b), whereas electrons trapped deeper in the bandgap remain captured in the interface traps. When the hole barrier is reduced, holes (indicated by the open symbols) are accumulated at the Si surface and recombine with the trapped electrons, Fig. 2.17.c) and d).

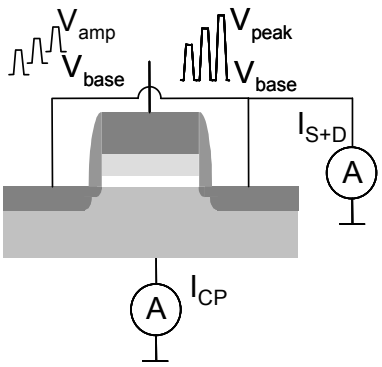

Figure 2.16: Schematic measurement setup for a charge pumping measurement.

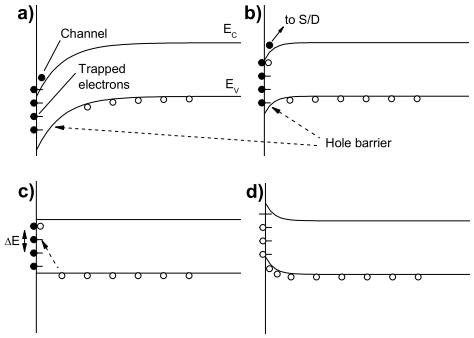

Figure 2.17: Energy band diagram for charge pumping measurements in a n-channel MOSFET [15].

The relation between substrate current (I_{CP}) and surface states density ($D_{it}(\mathcal{E})$) is given in Eq. 2.43, where q is the electronic charge, f the frequency of the gate pulse and A the channel area of the MOSFET.

$$I_{CP} = qfA \int D_{it}(\mathcal{E})\, d\mathcal{E} \tag{2.43}$$

The measured charge pumping current depends on the rise (t_r) and fall (t_f) time of the gate pulse due to thermal emission of carriers and the capture cross section for electrons σ_e and holes σ_h. For the Si / SiO_2 interface with a uniform surface states density throughout the bandgap the charge pumping current follows Eq. 2.44 where $\overline{D_{it}}$ is the average interface state density, v_{th} the thermal velocity of the carriers, n_i the intrinsic carrier concentration and ΔV_G the amplitude of the gate pulse.

$$I_{CP} = 2qfA\overline{D_{it}} \left[\ln \left(v_{th} n_i \sqrt{\sigma_e \sigma_h} \frac{V_{FB} - V_T}{\Delta V_G} \sqrt{t_r t_f} \right) \right] \tag{2.44}$$

The capture cross section for electrons and holes were extracted using a three-voltage level charge pumping technique [16] and found to be $\sigma_e \approx 10^{-14} cm^2$ and $\sigma_h \approx 10^{-16} cm^2$ for conventional Si / SiO_2 interface.

When rise and fall times or ramp rates (V/s) are kept constant the normalized charge per cycle N_{CP} is frequency independent.

$$N_{CP} = \frac{I_{CP}}{qfA} \tag{2.45}$$

Beside charged surface states also the recombination of inversion carriers can contribute to the substrate current. This effect is known in the literature as the geo-

metric component [17]. Therefore the device geometry together with rise and fall time need to be chosen carefully.

2.3 Dielectric reliability

Reliability of MOS devices is generally defined as: "The probability that the device will perform its required function under defined conditions for a specified period of time". The required function for gate dielectrics is to maintain the insulating properties at the operation voltage and temperature. In addition to the gate dielectric requirements the MOS device also has to maintain its characteristic device parameters like V_T, I_{Dsat} and g_m for the specified period of time. In the following sections the basic charge trapping models are discussed and the terms of gate oxide reliability are defined. Furthermore, commonly used lifetime prediction models are summarized.

2.3.1 Charge trapping

Charge trapping in conventional SiO_2 gate dielectrics has extensively been studied in the past [18, 19, 20] and simple 1^{st} order trapping models were introduced to explain the experimentally observed behavior.
Let us first consider electron trapping in pre-existing defect sites. The 1^{st} order rate equation describing the build up of negative charge is given in Eq. 2.46 where N_{tr} is the volume density of trapped charges, N_0 the number of initial trapping centers per unit volume, J_e the electron current density and σ_e the electron capture cross section per unit area.

$$q\frac{dN_{tr}}{dt} = J_e\sigma_e\left(N_0 - N_{tr}\right) \tag{2.46}$$

The solution of the 1^{st} order rate equation with the boundary condition $N_{tr}(0) = 0$ is given as

$$N_{tr}(t) = N_0\left(1 - \exp\Big(-\frac{J_e\sigma_e}{q}t\Big)\right) \tag{2.47}$$

When an emission term is included the rate equation can be written as

$$q\frac{dN_{tr}}{dt} = J_e\sigma_e\left(N_0 - N_{tr}\right) - qe_eN_{tr} \tag{2.48}$$

where e_e is the emission rate. In this case the solution for the rate equation is

$$N_{tr}(t) = \frac{N_0J_e\sigma_e}{qe_e + J_e\sigma_e}\left(1 - \exp\Big(-\big(e_e + J_e\sigma_e/q\big)t\Big)\right) \tag{2.49}$$

If, in addition, trap generation is also considered then the rate equation has to be written as

$$q\frac{dN_{tr}}{dt} = J_e\sigma_e\big(N_0 + N_{OT}(t) - N_{tr}\big) - qe_eN_{tr} \tag{2.50}$$

where $N_{OT}(t)$ is the volume density of generated trapping centers as function of time. In the following the case of linear trap generation (t) and non-linear trap generation ($t^{1/2}$) are discussed separately.

For the case of linear trap generation $N_{OT}(t)$ is written as $N_{OT}(t) = G_eJ_et$ where G_e is the trap generation rate per unit charge. The solution for linear trap generation is given by.

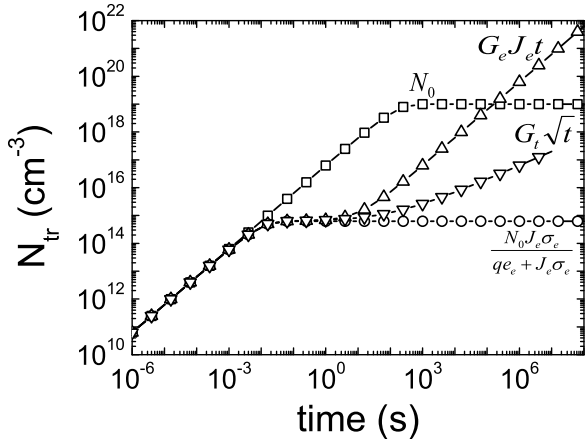

Figure 2.18: Number of trapped carriers n as a function of time considering the 1st order rate equation including emission as well as trap generation.

$$N_{tr}(t) = \frac{J_e\sigma_e}{(qe_e + J_e\sigma_e)^2}\left[G_e J_e t\big(qe_e + J_e\sigma_e\big) + \Big(qG_e J_e - N_0\big(qe_e + J_e\sigma_e\big)\Big) \times\left[\exp\Big(-\big(e_e + J_e\sigma_e/q\big)t\Big) - 1\right]\right] \quad (2.51)$$

For non-linear trap generation the case of square-root dependent generation rate is of particular interest. Then $N_{OT}(t)$ is written as $N_{OT}(t) = G_t\sqrt{t}$ where G_t is the trap generation rate per unit time and the solution for the rate equation is given by

$$N_{tr}(t) = \frac{J_e\sigma_e}{(qe_e)^{3/2}}\left[\Big(N_0\sqrt{qe_e} + A\sqrt{qe_e t}\Big) - \Big(N_0\sqrt{qe_e} + \frac{A\sqrt{\pi q}}{2}\ Erfi\Big(\sqrt{e_e t}\Big)\Big)\exp\big(-e_e t\big)\right] \quad (2.52)$$

with $Erfi\big(\sqrt{e_e t}\big)$ being the imaginary error function given as

$$Erfi\big(\sqrt{e_e t}\big) = \frac{2}{\sqrt{\pi}}\int_0^{\sqrt{e_e t}} \exp\big(x^2\big)\,dx \quad (2.53)$$

Typical results for the 1^{st} order rate equation are given in Fig. 2.18 with $N_0 = 10^{19} cm^{-3}$, $J_e = 10^{-6} A/cm^2$, $\sigma_e = 10^{-15} cm^2$, $e_e = 100/s$, $G_e = 10^{24}/(C\,cm)$ and $G_t = 10^{18}/(cm^3\sqrt{s})$. As can be seen the first order rate equation saturates to N_0, whereas when an emission term is included saturation to $(N_0 J_e\sigma_e)/(qe_e + J_e\sigma_e)$ is observed. In case of trap generation the saturation level either follows $G_e J_e t$ or $G_t\sqrt{t}$.

2.3.2 Dielectric breakdown

In order to study the reliability behavior of high-ε gate dielectrics the basics of dielectric breakdown typical for conventional gate dielectrics are briefly summarized.

Definition of dielectric breakdown

Dielectric breakdown has been defined as sudden loss of the insulating properties of the dielectric leading to a large leakage current increase through a localized spot caused by ohmic conduction, which is called hard-breakdown (HBD). The breakdown event in ultra-thin gate dielectrics has been studied extensively in the recent past and the soft-breakdown (SBD) phenomenon has been reported [21]. A soft-breakdown in the dielectric only causes a current increase and the current-voltage characteristic follows a power law relation ($I_G \sim A\,V^{\delta}$) [22], which may be tolerable from device point of view and might not have to be considered as a circuit failure [23]. The soft-breakdown is also associated with an increase in the noise of the stress current traces, which can be used as a breakdown monitor [24]. In this study, depending on the gate stack and the device geometry either SBD or HBD are considered as dielectric failure.

Test methodology

In practice different measurement techniques are applied to investigate the reliability of MOS devices [25]. In the following the Ramp I-V, Constant Voltage Stress (CVS) and Constant Current Stress (CCS) method will be discussed.

Ramp I-V: In a ramp I-V the gate voltage is increased rapidly from zero until a sudden increase in the gate leakage current is measured. In this measurement the breakdown is monitored as a function of gate bias from which a breakdown field is determined. In principle different failure modes can be detected with this technique. The ramp I-V test was used in this study only for gate dielectric screening.

Constant Voltage Stress (CVS): During a CVS, the gate voltage is kept constant and the gate current is measured as a function of time. A breakdown event is recorded either when a sudden change in the gate current is measured (HBD) or when the gate current noise exceeds a specified current noise level which corresponds to a SBD event [24]. Charge trapping during CVS stress can either lead to a continuous increase in current, which is related to positive charge, or decrease due to trapping of negative charge. These effects are more pronounced in thick dielectrics.

Constant Current Stress (CCS): During a CCS, the gate current is kept constant and the gate voltage is monitored as a function of time. A breakdown event is recorded when a sudden change in the gate voltage is measured. This procedure is well suited to study the reliability behavior of thick dielectrics. The CCS procedure is hardly used in ultra-thin gate dielectrics due to the strong correlation between the gate voltage and time-to-breakdown [26]. In addition, the presence of soft-breakdown events cause further difficulties in the breakdown detection. Charge trapping is also seen during a CCS stress. It leads to a gate voltage decrease for build up of positive charge and a voltage increase in case of negative charge trapping.

In absence of charge trapping CCS and CVS stress yield equivalent results.

Statistical evaluation of reliability data: The reliability of MOS structures is described by $R(t)$ which gives the probability that a single device will retain its required function under certain operating condition for a given time period. The probability that a device will be defective in the time interval $[0, t]$ is given by the cumulative failure distribution

$$F(t) = 1 - R(t) \tag{2.54}$$

This function is related to the failure probability density $f(t)$ via the integral

$$F(t) = \int_0^t f(t')\,dt' \tag{2.55}$$

It is well accepted in the literature that dielectric breakdown follows the Weibull statistic. Then $f(t)$ can be written as

$$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \tag{2.56}$$

where β and η represent the shape parameter and the scale parameter, respectively. The cumulated failure distribution calculated by integration according to 2.55 is

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \tag{2.57}$$

To analyze reliability data the cumulative failure distribution is commonly plotted using

$$\ln\big(-\ln\big(1-F\big)\big) = \beta\ln\big(t\big) - \beta\ln\big(\eta\big) \tag{2.58}$$

where β is given by the slope of the Weibull plot and η reflects the time when the failure percentile of 63% is reached. For an accurate determination of the reliability parameters β and η the reliability data are fitted using the maximum likelihood method [27]. For thin SiO_2 (sub 10 nm) it has been found that β strongly decreases with decreasing dielectric thickness. This effect is explained by the percolation theory [28] which is based on uniform trap generation in the bulk of the dielectric layer. The number of traps required to form a conducting path between the anode and the cathode decreases with decreasing dielectric thickness and therewith the failure probability distribution broadens.

Area scaling law: Based on the assumption that the breakdown spot in MOS devices happens at a random localized position it has been shown that the time to the 63% failure (η_1 and η_2) measured on two different areas (A_1 and A_2), respectively, follows the relationship given in Eq. 2.59.

$$\eta_2 = \eta_1 \cdot \left(\frac{A_1}{A_2}\right)^{1/\beta} \tag{2.59}$$

Since β strongly decreases with decreasing oxide thickness this effect became evident in sub 10 nm gate dielectrics.

2.3.3 Lifetime prediction

The aim of a reliability study is to predict the lifetime of MOS circuits at operation conditions, as shown in Fig. 2.19. The specified reliability targets are a low failure percentile for a specified device area. The failure distribution at operation condition is predicted based on accelerated reliability measurements which are extrapolated according to mostly 'semi-empirical' acceleration laws. The extrapolation is based on the assumption that the same degradation mechanism apply at test and operation condition. The experimental verification of the acceleration model requires long-term reliability tests which are time consuming and costly and therefore not always feasible. Since the era of ICs different acceleration laws have been used to predict the product lifetime which are briefly discussed in the following paragraphs.

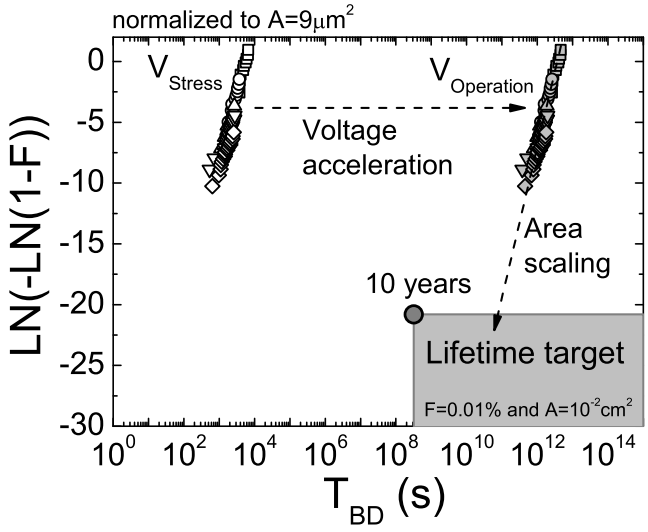

Figure 2.19: Lifetime extrapolation according to voltage acceleration and area scaling. Product target is indicated by the shaded area using a failure percentile of F=0.01% and device area of $A = 10^{-2}cm^2$.

Field acceleration model, $1/E$ versus E: In the early days of IC manufacturing the empirical E model was commonly used to determine the gate oxide lifetime of MOS circuits. The first physical acceleration model ($1/E$) was introduced in the mid 80's [29] including a physical interpretation of the field dependence. In the late 90's the underlying physics for the thermochemical E model was provided as well [30, 31].

The extrapolation law for $1/E$ and E are given in Eq. 2.60 and 2.61, respectively. In both cases the measured time to breakdown t_0 at a given oxide field E_{OX} is used for extrapolation to the operation field. The acceleration factor G in Eq. 2.60 consists of contributions from the $1/E_{OX}$ dependence of the FN tunneling current and the field dependence of the impact ionization coefficient. The latter causes field dependent hole generation and trapping near the injecting interface [29].

$$T_{BD} = t_0 \cdot \exp\big((G/E_{OX}\big) \tag{2.60}$$

The thermochemical model predicts the time-dependent dielectric breakdown based on thermal and field dependent bond-breakage in the dielectric. The extrapolation law given in Eq. 2.61 only describes the field dependent term of the electrochemical model.

$$T_{BD} = t_0 \cdot \exp\big(-\gamma E_{OX}\big) \tag{2.61}$$

The linear E_{OX} model is frequently used in practice since it presents a worst case assessment.

Voltage acceleration model: Due to the change in the conduction mechanism in ultra-thin gate dielectrics from Fowler-Nordheim tunneling to direct tunneling the acceleration law had to be adjusted. Since the charge transport through such thin dielectrics is ballistic, the electric field does not determine the carrier energy at the anode but the applied gate bias is responsible for the defect generation process

in the dielectric. Therefore the gate voltage rather than the electric field appears in the exponent of the acceleration formula in Eq. 2.62 [32].

$$T_{BD} = t_0 \cdot \exp\big(-\gamma V_{OX}\big) \tag{2.62}$$

Power law model: Recently it has been demonstrated that the voltage dependence of time-to-breakdown in ultra-thin gate dielectrics does not follow the simple relationship given in Eq. 2.62 but can be described using a power law expression as given in Eq. 2.63.

$$T_{BD} = t_0 \cdot V^{(-n)} \tag{2.63}$$

The constant n of the power law extrapolation model was experimentally verified over a large voltage range and found to be in the order of ~ 44 [33]. The physical origin for the power law dependence, however, is still under debate.

3

Dielectric deposition techniques and device fabrication process

In this section, various deposition techniques for high-ε gate dielectrics are described together with the device fabrication process. The discussion focuses on the **Atomic Layer Deposition technique** and two device fabrication schemes. The **conventional self aligned process** which is commonly used to fabricate CMOS circuits and the **non-self aligned gate last process**. The latter enables MOSFET fabrication at a strongly reduced thermal budget.

3.1 Deposition techniques for high-ε gate dielectrics

3.1.1 Atomic Layer Deposition

The principle of the Atomic Layer Deposition (ALD) technique is based on a self-limiting growth process [34]. In a typical ALD deposition process alternating reaction gases are injected into the reaction chamber separated by purge cycles. The self-limitation in the growth arises from the saturation in the surface reaction during each reaction cycle. The self-limiting deposition process enables precise control of the layer thickness, ensures uniformity over large area substrates and good reproducibility in the sub-nm thickness regime. Therefore, ALD is considered as one of the most promising deposition techniques for manufacturing of high-ε gate dielectrics in future CMOS technologies.

Metal chlorides and water are most often used as precursors in ALD of metal oxides. The chlorides are often good precursors because they are stable, sufficiently volatile, reactive and not too difficult to handle. Beside chlorides also organometallic precursors (methyl-, ethyl-compounds) are used in ALD metal oxides. Water as source of oxygen is not necessarily always the best choice. Alternative precursors for oxygen are O_2, H_2O_2 and O_3.

Some of the issues with ALD as technique for gate dielectric deposition are the low deposition temperature and the growth inhibition on hydrogen terminated Si. This results in fairly high impurity level such as e.g. chlorine concentration and

the island like growth behavior [35]. The growth kinetics of ALD is much more favorable on an OH terminated Si surface as can be obtained by a wet chemical pre-gate clean [36].

In this work mainly two metal oxide processes were used. Al_2O_3 was formed using $Al(CH_3)_3$ and H_2O as precursor for Al and O. In the second process $HfCl_4$ and H_2O were used as Hf and O sources in the HfO_2 deposition process. Both ALD processes were carried out at a deposition temperature of 300°C.

3.1.2 Alternative high-ε deposition techniques

Metal Organic Chemical Vapor Deposition (MOCVD) is also considered a promising deposition technique for gate dielectric application. In a metal-oxide or metal-silicate deposition process by MOCVD Hf-(N-(CH_2-CH_3)$_2$)$_4$ (tetrakis-diethylamido-hafnium (TDEAH)) or Si-(N-(CH_3)$_2$)$_4$ (tetrakis-dimethylamido-silicon (TDMAS)) are frequently used as metal or Si precursors, respectively. The basic reactions in a MOCVD process are:

$$\text{TDEAH} + O_2 \rightarrow HfO_2 + CO_2 + NO_x + H_2O$$
$$\text{TDMAS} + O_2 \rightarrow SiO_2 + CO_2 + NO_x + H_2O$$

Typical deposition temperatures range from 300°C to 600°C. The fairly high level of carbon impurities is a known complication of the MOCVD process, which can be reduced by increasing the deposition temperature. Due to the availability of both Hf and Si precursors, HfO_2 and $HfSiO_x$ films with a wide range of Si concentrations can be fabricated.

Initial results on high-ε gate dielectrics were often reported using Physical Vapor Deposition (PVD) techniques. In a PVD process either a metal film is deposited by nonreactive sputtering or a metal oxide layer is formed by reactive sputtering in an oxygen containing plasma. In case of nonreactive sputtering the metal film is subsequently converted into a metal oxide by an oxidizing anneal or plasma oxidation step. Precise thickness control, process uniformity and process repeatability are of major concern for PVD as gate dielectric deposition technique. However, a few very promising results have also be obtained by PVD [37].

A further technique used to deposit mostly rare-earth metal oxides is Molecular Beam Epitaxy (MBE). In MBE a molecular beam is deflected to the heated substrate where epitaxial growth occurs. A modified version of MBE exposes the substrate to both metal and oxygen species separately. Using these techniques MOS devices with rare-earth metal oxides have been fabricated in the past which showed some potential for future application in CMOS technologies [38, 39]. The MBE technique has advantages over the CVD and PVD technique in terms of film purity and Si substrate damage. However, the required ultra-high vacuum for MBE is the cause for a low throughput which is of concern from the manufacturing point of view. Furthermore, in MBE the material deposition is non-conformal which results in a poor step coverage. Therefore it is not considered a very promising deposition technique for future CMOS technologies with high-ε gate dielectrics.

3.2 Non-self aligned gate last process

As briefly mentioned, the major advantage of a gate last process is the strongly reduced thermal budget to which the gate stack is exposed during the device fabrication. This is achieved by forming the junctions prior to the gate dielectric and

gate electrode deposition. Furthermore, a less stringent contamination protocol applies for the gate stack patterning.

The first process steps in the non-self aligned gate last process are related to the isolation scheme. In this fabrication process a standard poly-buffered Local Oxidation Of Silicon (LOCOS) isolation was used followed by an reoxidation step. The source and drain junctions were formed by ion implantation using a dummy gate structure as shown in Fig. 3.1 and 3.2. After the dummy gate removal an annealing step was performed to activate the source and drain dopants. Prior to the gate opening a SiO_2/Si_3N_4 dual layer dielectric was deposited. The gate opening was achieved by using a dry etch process for Si_3N_4 and a wet etch process for the SiO_2. The use of a wet etch process to remove the underlaying SiO_2 resulted in an undercut in the transistor process, which is of concern for yield related issues.

Prior to the gate dielectric deposition a pre-deposition treatment was carried out predominantly using thermal or wet-chemical treatments. As gate dielectric either Al_2O_3 or HfO_2 was used, but also other gate dielectric materials like ZrO_2, $ZrAlO_x$, $HfAlO_x$, HfO_2/Si_3N_4 and laminate stacks were investigated using this fabrication process. In a variety of experiments the gate dielectric deposition was followed by a post-deposition anneal prior to the gate electrode deposition. As gate material either poly-Si or metal (TiN) was used. The gate stack formation was completed by the gate patterning process. Due to the use of a thick SiO_2/Si_3N_4 dual layer dielectric for isolation the gate patterning process was rather uncritical. This enabled the use of a backend tool for gate etch, which significantly relieved the contamination protocol. In Fig. 3.3 the device structure is shown prior to contact hole opening. In the contact hole opening process the gate area was protected using a photo-resist mask, as shown in Fig. 3.4. After contact hole etch a resist removal step was carried out followed by a wet chemical clean.

For metallization a multi layer layer stack of Ti / TiN / Al / Ti / TiN was deposited and patterned. The device fabrication process was completed using a final passivation anneal in forming gas, which is a standard mixture of N_2 and H_2. The aim of the forming gas anneal is to passivate electrically active defects similar to the ones known from the Si / SiO_2 interface. However, the use of Si_3N_4 in the device fabrication process for isolation together with the introduction of alternative gate dielectrics and metal gate electrodes may require changes in this process step. The final MOSFET device structure is shown schematically in Fig. 3.5.

At this point it should be mentioned that a non-self aligned process does not allow to fabricate short channel devices due to alignment issues. Therefore only device structures with a gate length of more than $1\,\mu$m are available. The overlap capacitor structure is shown in Fig. 3.6 for comparison. This structure does not suffer from the gate undercut on the active area as mentioned above, which makes it a more suitable device structure for reliability evaluations.

Figure 3.1: PBL is used as isolation scheme followed by a sacrificial oxidation step to protect the active device area. Photo-resist was used for dummy gate patterning.

Figure 3.2: The channel region is masked by the dummy gate during the junction implantation. After the ion implantation the dummy gate is removed followed by the source, drain activation anneal.

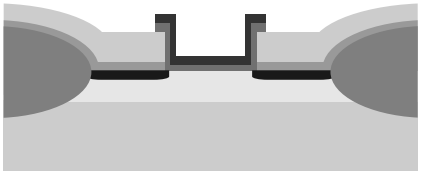

Figure 3.3: A SiO_2/Si_3N_4 dual layer dielectric is deposited prior to the active area formation, which is opened by a combined dry / wet etch. The gate stack is then deposited including the pre- and post-deposition treatment followed by the deposition of the gate electrode. Gate patterning is completing the gate stack formation.

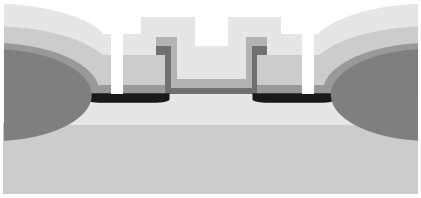

Figure 3.4: Photo-resist is protecting the gate area during the contact hole etch. A resist removal is performed which finalizes the contact hole opening.

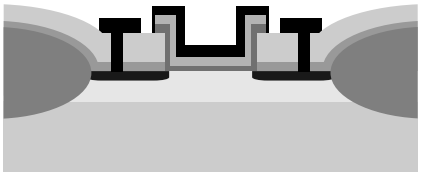

Figure 3.5: The metallization process consists of Ti/TiN/Al/Ti/TiN deposition and metal patterning. The device fabrication process is completed by a final passivation anneal in forming gas (N_2/H_2).

Figure 3.6: Structure of the overlap capacitor in the non-self aligned gate last process where the wafer backside is used as 2^{nd} device terminal.

3.3 Conventional self aligned process

For integration of high-ε gate dielectrics into a CMOS process a conventional integration scheme is generally favored. On the other hand, the higher thermal budget and the compatibility with a poly-Si gate electrode increase the requirements for the high-ε material. However, to demonstrate compatibility of high-ε gate dielectrics with a conventional processing scheme a standard MOSFET process was adapted.

The processing sequence of the conventional self aligned process starts similar to the gate last process with the isolation scheme. Again poly-buffered LOCOS

was used for isolation as shown in Fig. 3.7. After the isolation process, a pre-gate clean was performed using hydrofluoric (HF) acid followed by the pre-deposition treatment. Either a thermal treatment or a wet chemical process was used as surface preparation for the high-ε deposition. Then the high-ε film is deposited. Post-deposition anneals in N_2, NH_3 or O_2 were carried out prior to the deposition of an amorphous Si electrode followed by ion implantation. N-type electrodes for NMOS transistors were obtained by using phosphorus as gate dopants (see Fig. 3.8). For gate patterning photo resist was used in combination with a reactive ion etch (RIE) which stopped on the high-ε material. The device structure after an ideal gate etch is shown in Fig. 3.9. However, the post gate etch clean process did not completely remove the high-ε material either. Therefore, the Lightly Doped Drain (LDD) implantation had to be done through the high-ε dielectric which required a modification of the implant energy. Prior to source and drain implantation a dual layer spacer was formed using a seal nitride layer followed by a standard Tetraethyl Orthosilicate (TEOS) spacer. The activation of the junction and gate dopants was achieved by a 10 s anneal at 1000°C. The ideal device structure after spacer and junction formation is illustrated in Fig. 3.10. However, due to issues related to the high-ε removal process the real device structure differs somewhat from Fig. 3.10. The high-ε material was basically removed after the spacer formation and therefore the high-ε dielectric extends underneath the spacer. This fact is of concern in view of channel hot carriers which may be trapped in the vicinity of the drain extensions. Process modifications focusing on the high-ε removal prior to the spacer formation significantly improved the V_T behavior.

In order to minimize the junction and gate resistance the standard self aligned Ti silicide process was used, where Ti is deposited on the wafer first and then a thermal process step is applied to react the metal and the Si forming $TiSi_2$. The unreacted Ti is selectively removed by a wet chemical process step followed by a 2^{nd} thermal treatment. The device structure after salicide formation is shown in Fig. 3.11.

For device contact, first an Inter Level Dielectric (ILD) was deposited followed by a planarization step using Chemical Mechanical Polishing (CMP). The contact hole process includes lithography, etch, clean and tungsten (W) deposition. The final metallization was achieved using Al similar as in the gate last process. A passivation anneal in forming gas completed the MOSFET fabrication.

The final device structure using the conventional process flow is shown in Fig. 3.12. The major advantage of the self aligned process is the rather simple fabrication of short channel devices, but on the other hand a higher thermal budget and compatibility with poly-Si gates are more stringent requirements for the high-ε material.

The transmission electron micrographs (TEM) of the fabricated devices using the non-self aligned gate last process and the conventional self aligned process are shown in Figs. 3.13 and 3.14, respectively. Note the presence of the gate undercut in the gate last process and the removal of the high-ε material after the spacer formation for the conventional structure.

Figure 3.7: PBL is used as isolation scheme similar as in the non-self aligned gate last process.

Figure 3.8: Prior to the gate stack deposition an HF clean is performed followed by the pre-deposition treatment. After dielectric deposition and a possible post-deposition treatment the poly-Si deposition completed the gate stack deposition.

Figure 3.9: The gate stack patterning is achieved using a RIE etch followed by a post etch clean. Preferably the high-ε material should be removed during the gate etch process, which turned out to be rather difficult.

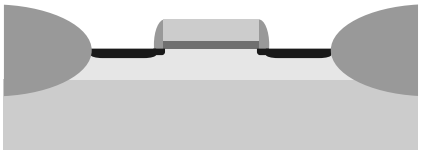

Figure 3.10: LDD implantation was done prior to the spacer formation followed by source, drain implantation. The cleaning step after dopant activation was used to remove the remaining high-ε material.

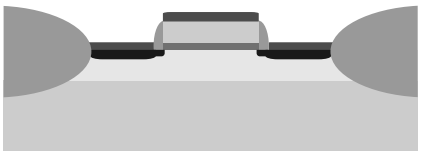

Figure 3.11: Self aligned silicidation process was used to form junction and gate contact.

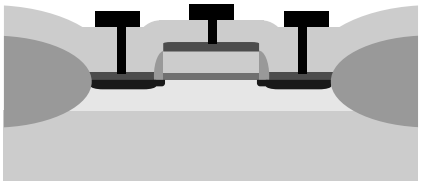

Figure 3.12: Prior to metallization ILD was deposited followed by contact hole formation. Al was used for metallization similar as for the non-self aligned gate last process. Device fabrication was completed by final passivation anneal. Note: Contact of the gate electrode is located outside of the active channel area.

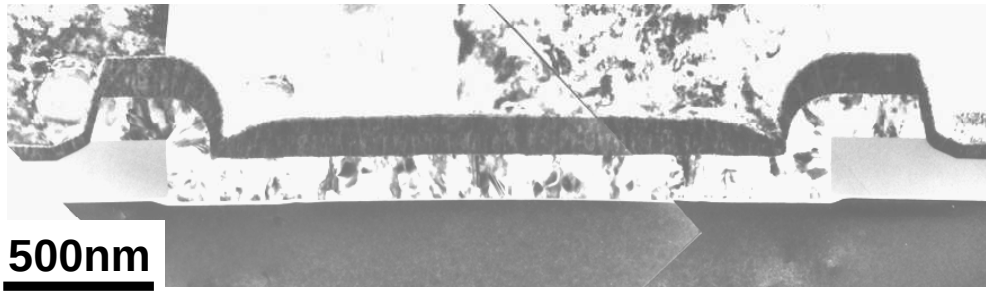

Figure 3.13: TEM image of a MOS capacitor fabricated using the non-self aligned gate last process. Note the presence of an undercut for the non-overlapping capacitor similar as observed for the transistor structure.

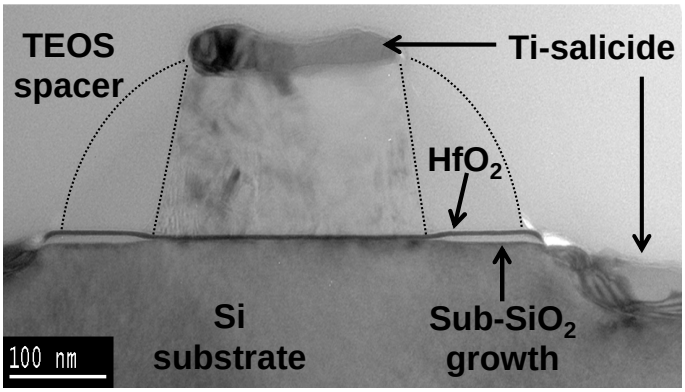

Figure 3.14: TEM image of a MOSFET structure fabricated using the conventional self aligned process. Note the high-ε material is removed after the spacer formation.

4

Conventional electrical characterization of MOS devices

4.1 Gate I-V analysis

Reducing the gate leakage current of MOS devices is the primary aim of introducing high-ε gate dielectrics into future CMOS technologies. Towards integration of these materials it is important to understand the charge transport through such gate dielectrics. Recently, it was reported that depending on the injection polarity the conduction mechanism in high-ε gate dielectrics can either be explained by trap-assisted tunneling [40] or by Frenkel-Poole tunneling and Schottky emission [41].

In this section the gate leakage current for gate and substrate injection will be discussed for SiO_2 / Al_2O_3 gate stacks with TiN as well as for SiO_2 / HfO_2 stacks with conventional n-type poly Si gate electrodes.

Gate leakage currents in SiO_2 / Al_2O_3 gate stacks with TiN electrodes: Planar capacitors ranging in gate area from 9 to $10^4 \mu m^2$ were fabricated first by growing a nominal 1 nm thermal SiO_2 layer followed by 7 to 20 nm Al_2O_3 film deposited by ALD. Samples with and without post deposition treatment are compared. PVD TiN was used as gate electrode.

The leakage current through 7, 15, and 20 nm thick *as deposited* Al_2O_3 layers can be described reasonably well by tunneling through a stacked dielectric using the method described in Section 2.1.1. The parameters used to calculate the tunneling current are summarized in Table 4.1 where Φ_B is the barrier height for the injection from the TiN gate electrode and Φ_B^* the barrier for injection from the Si conduction band as illustrated in Fig. 4.1.

The parameter m_x^* is the effective electron mass in the bandgap of either Al_2O_3 or SiO_2, respectively. The gate bias for the calculated I-V characteristics was determined using $V_G = V_{OX} + V_{FB}$, where V_{FB} is the extracted flatband voltage (see Fig. 4.13). Band-bending in the Si substrate and barrier distortion due to the presence of fixed charge in the dielectric was neglected to first order. The fitted

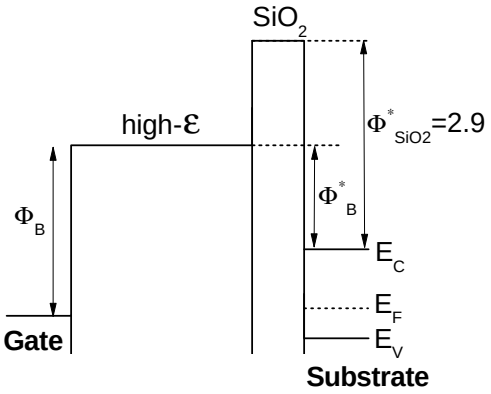

Figure 4.1: Schematic drawing of the band diagram indicating the barrier height for injection from the TiN gate electrode (Φ_B) and the Si substrate (Φ_B^*).

Table 4.1: Parameters used to calculate the tunneling current through the SiO_2 / Al_2O_3 gate stacks shown in Fig. 4.2

$t_{Al_2O_3}$ (nm)	t_{SiO_2} (nm)	$V_{FB}(V)$	Injection	Φ_B^* (eV)	$m^*_{Al_2O_3}$ / $m^*_{SiO_2}$ (m_e)
7	1.3	1.4	Substrate	1.3	0.42 / 0.5
15	1.3	1.7	Substrate	1.3	0.42 / 0.5
20	1.3	1.9	Substrate	1.3	0.42 / 0.5
$t_{Al_2O_3}$ (nm)	t_{SiO_2} (nm)	$V_{FB}(V)$	Injection	Φ_B (eV)	$m^*_{Al_2O_3}$ / $m^*_{SiO_2}$ (m_e)
7	1.3	0.57	Gate	2.7	0.42 / 0.5
15	1.3	1.05	Gate	2.7	0.42 / 0.5
20	1.3	1.2	Gate	2.7	0.42 / 0.5

curves together with the experimental data are shown in Fig. 4.2. As can be seen a good parameterization of the tunneling currents over a large range in thicknesses is possible in this way. The deviations at high current densities are due to electron and hole tapping for substrate and gate injection, respectively. This will be discussed in detail in Section 6.1.4.

A schematic drawing of the energy band diagram for SiO_2 / Al_2O_3 gate stacks with TiN electrodes is shown in Fig. 4.3 for gate and substrate injection, respectively. As can be seen, the strong asymmetry in the gate leakage current is due to the inherent asymmetry of the dual layer stack (band structure and dielectric constant) and due to the work-function difference between the injecting electrodes (TiN versus n-type Si).

The asymmetry in the gate current with respect to the injection polarity cannot only be explained by the difference in barrier height between n-type Si substrate and TiN. For gate injection the interfacial layer acts as capacitive voltage divider which significantly reduces the voltage drop across the Al_2O_3 layer. The presence of the interfacial SiO_2 layer is detrimental for injection from the Si substrate. In this case the larger voltage drop across the interfacial layer, due to the large differences in the dielectric constant, enables electron tunneling into the conduction band of the Al_2O_3 layer.

These general considerations remain true for high-temperature annealed gate stacks, as can be seen by comparing Figs. 4.4 and 4.5. However, a more detailed analysis for gate injection shows that the injection voltage is significantly reduced by

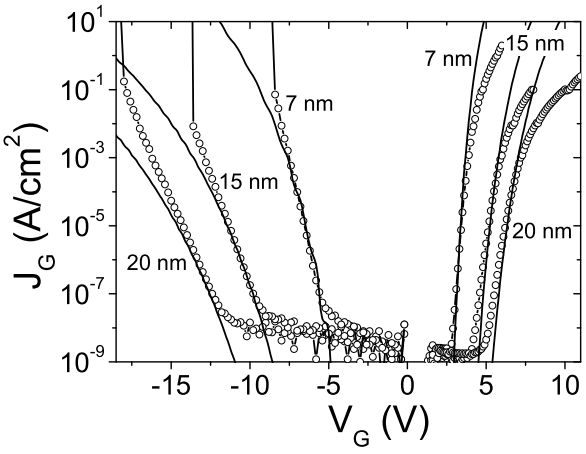

Figure 4.2: Comparison between experimental (open symbols) and calculated I-V characteristics of *as deposited* Al_2O_3 layers on 1 nm SiO_2 interface with TiN electrodes. Parameters used to calculate the tunneling currents are given in Table 4.1.

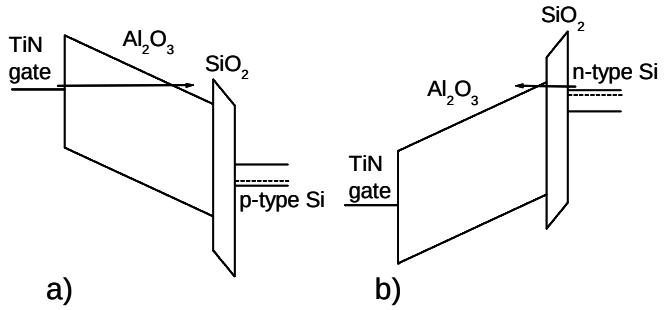

Figure 4.3: Schematic band diagram of SiO_2 / Al_2O_3 gate stacks with TiN electrodes.

high-temperature annealing. Transmission electron micrographs (TEM) taken for these gate stacks indicate a thickness reduction of $\sim 10\%$. This change in thickness cannot account for the reduced injection voltage. A decrease of the barrier height between the Al_2O_3 and the TiN gate electrode is a more likely explanation for this change.

For each condition in Figs. 4.4 and 4.5, 29 capacitors were measured across the 8-inch wafer. The observed spread in the gate current arises from a small thickness variation ($\sim 5\%$) across the wafer, and it will be taken into account in the reliability studies presented in Section 7.1. The data in Figs. 4.4 and 4.5 also show that there exists a strong asymmetry in the breakdown current densities irrespective of PDA condition.

When scaling the Al_2O_3 and the interfacial SiO_2 layer to obtain a lower EOT the strong asymmetry in the leakage current gets reduced. On the other hand, also the work-function difference between the gate electrode and the Si substrate contributes to the asymmetry in leakage current and has to be taken into account. In general, from gate leakage point of view a symmetric gate stack would suppress the leakage current in accumulation and inversion and therefore be a favorable choice.

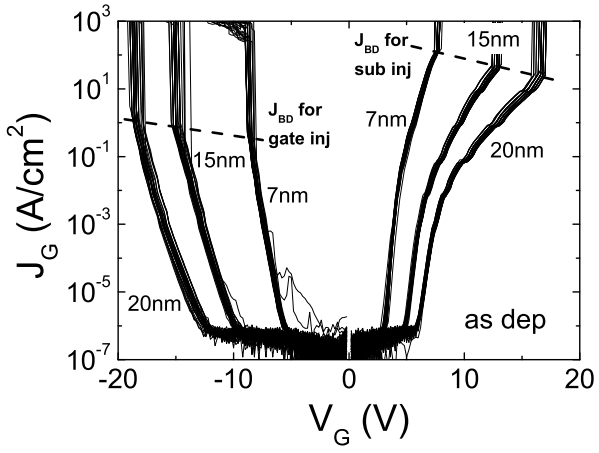

Figure 4.4: Current-voltage characteristic for *as deposited* SiO_2 / Al_2O_3 gate stacks with TiN electrodes on either p- or n-type Si substrates. The strongly asymmetric gate current is due to the asymmetry of the gate stack and the work-function difference between the n-type Si substrate and the TiN gate electrode.

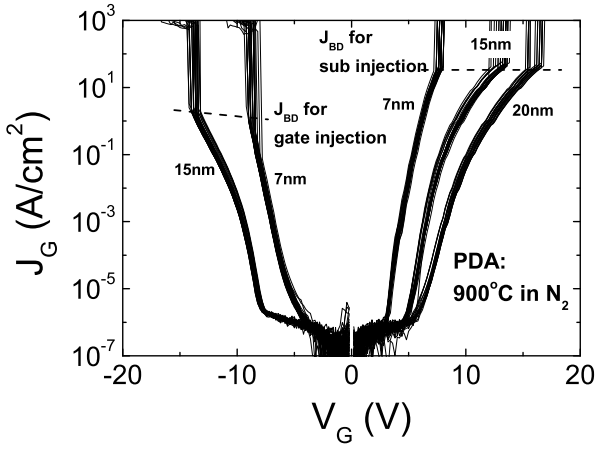

Figure 4.5: Same gate dielectric stacks as in Fig. 4.4, but with a post deposition anneal in N_2 at 900°C for 5 min. The asymmetry in the leakage current remains but is less pronounced compared to the *as deposited* stacks.

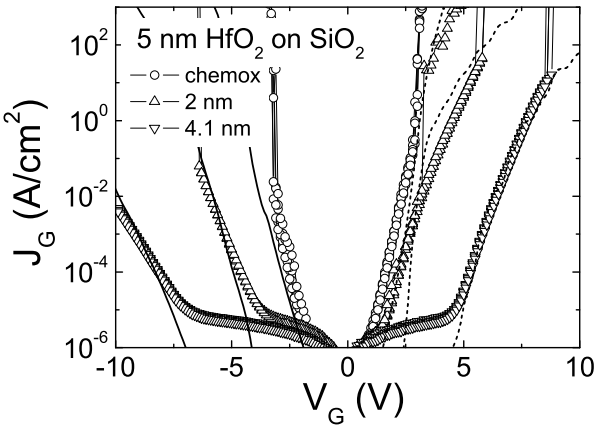

Figure 4.6: Comparison between experimental (open symbols) and calculated I-V characteristics for HfO_2 layers on various SiO_2 interfaces with poly-Si electrodes. Parameters used to calculate the tunnening currents are given in Table 4.2.

Table 4.2: Parameters used to calculate the tunneling current through the SiO_2 / HfO_2 gate stacks shown in Fig. 4.6.

t_{HfO_2} (nm)	t_{SiO_2} (nm)	$V_T(V)$	Injection	Φ_B^* (eV)	$m^*_{HfO_2}$ / $m^*_{SiO_2}$ (m_e)
5	0.8	0.5	Substrate	1.6	0.2 / 0.5
5	1.95	0.9	Substrate	1.6	0.2 / 0.5
5	4.1	1.2	Substrate	1.6	0.2 / 0.5
t_{HfO_2} (nm)	t_{SiO_2} (nm)	$V_{FB}(V)$	Injection	Φ_B (eV)	$m^*_{HfO_2}$ / $m^*_{SiO_2}$ (m_e)
5	0.8	-0.75	Gate	1.6	0.2 / 0.5
5	1.95	-0.7	Gate	1.6	0.2 / 0.5
5	4.1	-0.8	Gate	1.6	0.2 / 0.5

Gate leakage currents in SiO_2 / HfO_2 gate stacks with poly-Si electrodes: A similar parameterization of the tunneling currents through SiO_2 / HfO_2 gate stacks with n-type poly-Si electrodes was carried out using various interfacial SiO_2 thicknesses. The gate leakage current measured on n-channel MOSFETs either in accumulation ($V_G < 0$) or in inversion ($V_G > 0$) together with the fitted I-V characteristics are shown in Fig. 4.6. The parameters used to calculated the tunneling current through HfO_2 are given in Table 4.2.

As can be seen from Fig. 4.6 the parameterization of the tunnel currents for gate injection yields reasonable agreement whereas for substrate injection the calculated current densities significantly deviate from the experimental results, in particular for the thin interfacial layers. The discrepancy between calculated tunneling currents and the experimental results most likely results from the underlaying conduction mechanism. Indeed in literature a strong temperature dependence of the gate leakage current through HfO_2 layers has been reported which is an indication for a defect related charge transport [41].

4.2 C-V analysis

The analysis of the Capacitance-Voltage (C-V) characteristic of MOS devices provides information on various parameters which are of importance from integration point of view. One of these parameters is the EOT with respect to SiO_2 gate dielectrics, which is derived from the accumulation capacitance. In this section the High Frequency and quasi static C-V measurement techniques and their limitations are discussed.

4.2.1 High Frequency and quasi static C-V measurements

Process monitor using HF C-V analysis

Capacitance measurements are commonly used for process monitoring during process development as well as in mass-production. The parametric measurements were expanded to full C-V traces and stored for 29 capacitors across the 8-inch wafer. In this way, beside the variation in the accumulation capacitance, information on the flatband voltage spread can be obtained. Typical results for SiO_2 / Al_2O_3 gate stacks are shown in Fig. 4.7. The data were taken on large area device structures ($10^4 \mu m^2$) to minimize the parasitic effects. As can be seen, beside the spread in the accumulation capacitance a significant variation is also evident in the flatband voltage. In Table 4.3 the extracted parameters are summarized.

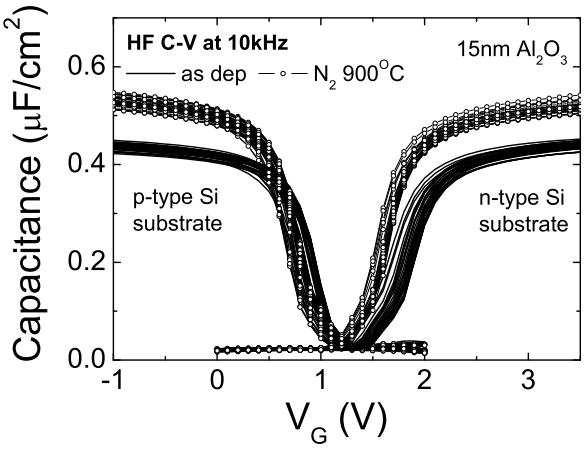

Figure 4.7: Typical HF C-V data of *as deposited* and high temperature annealed SiO_2 / Al_2O_3 gate stacks with TiN electrodes. Data were taken on large area devices ($10^4 \mu m^2$) to minimize parasitic effects.

When plotting the accumulation capacitance as given in Fig. 4.7 versus the wafer position a strong correlation between capacitance variation and wafer location becomes evident, as shown Fig. 4.8. The capacitance variation comes either from a variation in the interfacial layer or the Al_2O_3 thickness itself. The ALD deposition technique is known for precise thickness control, however, the capacitance increase along the gas flow direction within the deposition chamber indicates that the observed pattern arises from an Al_2O_3 thickness variation. Furthermore, the capacitance spread is consistent with the leakage current variation mentioned in Section 4.1. For wafer locations with higher accumulation capacitance the gate

Table 4.3: Summary of process parameters determined from HF C-V data as shown in Fig. 4.7

$t_{Al_2O_3}$ (nm)	Sub type	PDA	$\overline{C_{Acc}}$ ($\mu F/cm^2$)	ΔC_{Acc} ($\mu F/cm^2$)	ΔC_{Acc} (%)	$\overline{V_{MG}}$ (V)	ΔV_{MG} (V)
7	n	*as dep*	0.738	0.011	1.5	1.57	0.037
15	n	*as dep*	0.438	0.007	1.6	1.97	0.068
20	n	*as dep*	0.346	0.006	1.7	2.06	0.076
7	n	N_2, *900°C*	0.890	0.019	2.1	1.51	0.063
15	n	N_2, *900°C*	0.518	0.010	1.9	1.68	0.060
20	n	N_2, *900°C*	0.418	0.008	1.9	1.66	0.028
7	p	*as dep*	0.743	0.011	1.5	0.41	0.058
15	p	*as dep*	0.443	0.007	1.6	0.77	0.061
20	p	*as dep*	0.360	0.006	1.7	1.03	0.089
7	p	N_2, *900°C*	0.880	0.016	1.8	0.46	0.044
15	p	N_2, *900°C*	0.537	0.010	1.9	0.701	0.046

leakage current is also enhanced. In the following the average accumulation capacitance and flatband voltage are going to be used to extract the fixed charge and the dielectric constant.

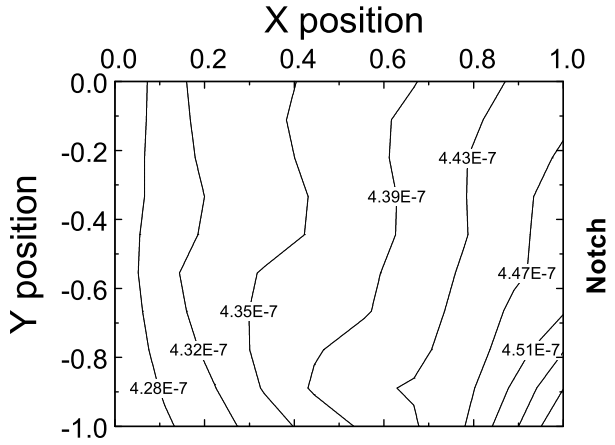

Figure 4.8: Contour plot generated from the capacitance data for 15 nm *as deposited* Al_2O_3 shown in Fig. 4.7. Clear evidence for a systematic capacitance variation across the 8-inch wafer is found.

Comparison of High Frequency and quasi static C-V

A comparison between the High Frequency and quasi static C-V measurement was carried out using a rather conservative high-ε gate stack. As can be seen in Fig. 4.9 good agreement between the two techniques was obtained for 15 nm *as deposited* Al_2O_3 layers. The samples with a high-temperature PDA anneal, however, show a strong discrepancy between HF and quasi static C-V. Following the discussion in Section 2.1.2 on the influence of series resistance and the parallel conduction no significant frequency dispersion is being expected. In the case of the *as deposited* layer

no frequency dispersion is measured. Therefore the observed effect is attributed to the high-temperature process step introducing the frequency dispersion.

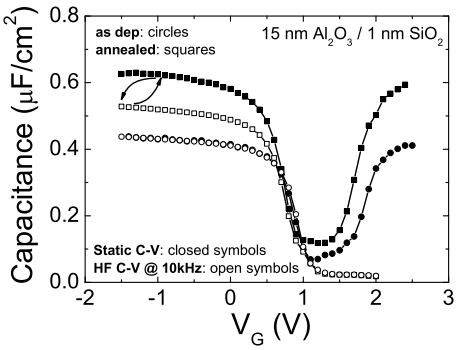

Figure 4.9: Comparison of HF C-V and equilibrium controlled static C-V measurement for 15 nm Al_2O_3 layers on p-type Si substrates. Good agreement was only obtained for the *as deposited* samples.

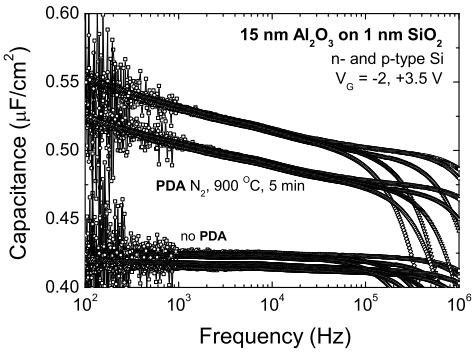

Figure 4.10: Capacitance versus frequency measured on 15 nm Al_2O_3 layers on n- and p-type Si substrates. Samples with and without PDA are compared.

To clarify this point the gain / phase analyzer was used in the frequency scan mode and samples with an Al_2O_3 layer on n- and p-type Si substrates were measured varying the frequency from 100 Hz up to 1 MHz. The capacitor area was ranging from $100\mu m^2$ to $9{\cdot}10^4\mu m^2$ to illustrate the effect of the series resistance on the measured HF capacitance. As expected, in the high frequency range and more pronounced on large area capacitors the series resistance leads to a roll-off in the measured capacitance at higher frequencies (>100 kHz). Furthermore, as shown in Fig. 4.10, the accumulation capacitance of the *as deposited* layer is frequency independent up to $\sim$ 100 kHz on both substrate types. For high-temperature annealed samples a significant frequency dependence is observed below 100 kHz again on both substrate types. The origin of this frequency dispersion remains unclear, however, from the experimental data it is evident that either slow charging or polarization related effects could contribute to the displacement current or the AC response of the MOS structure.

4.2.2 Interfacial layer thickness and dielectric constant

From the C-V data shown in Fig. 4.7 the average EOT of the gate stack was extracted using the C-V simulation tool described in Section 2.1.2 accounting for quantum mechanical effects. In Fig. 4.11 the EOT is plotted versus the nominal Al_2O_3 thickness. When plotting EOT versus the physical thickness of the Al_2O_3 layer the interfacial layer thickness and the dielectric constant of the Al_2O_3 can be obtained from the intercept and the slope, respectively.

An interfacial layer thickness of 1.5 ± 0.1 nm was extracted, indicating a considerable increase compared to the target thickness of 1 nm. For *as deposited* Al_2O_3 layers, a dielectric constant of 10.3 is obtained which is in good agreement with data reported in literature [42], while the the value for the annealed layers is 12.8. This increase results from a 20% increase in the accumulation capacitance (see Fig. 4.7). Based on the TEM analysis the Al_2O_3 layer thickness is reduced by the PDA in N_2 at 900°C for 5 minutes by $\sim$10%, as shown in Fig. 4.12. If this densification is

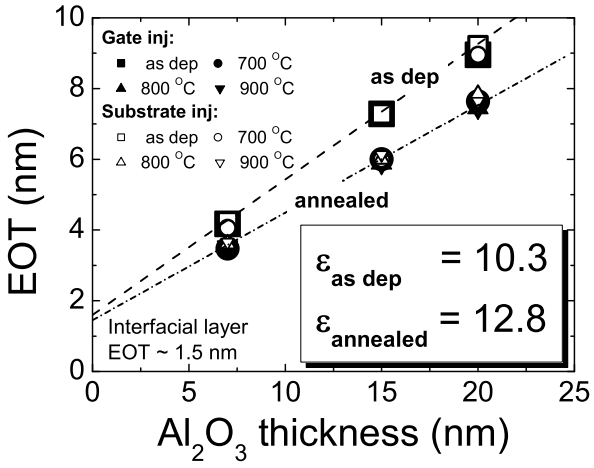

Figure 4.11: EOT versus nominal Al_2O_3 thickness. The slope reflects the dielectric constant of the Al_2O_3 film, whereas the interfacial SiO_2 thickness is given by the intercept.

taken into account, the dielectric constant of annealed films is reduced to a value of 11.5. The remaining difference might be caused by a change in the microstructure of the Al_2O_3. In fact, the TEM image of the annealed layer shows the presence of a crystalline phase, while the *as deposited* layer is fully amorphous.

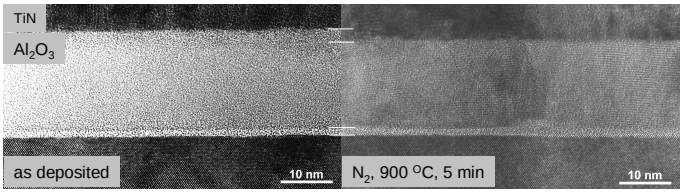

Figure 4.12: TEM images for a nominal 20 nm *as deposited* and high-temperature annealed Al_2O_3 layer. A physical thickness of $\sim$ 19 nm is extracted for the as deposited layer, whereas $\sim$ 17 nm is measured for the high-temperature annealed film due to densification.

4.2.3 Fixed oxide charge and work-function extraction

The fixed oxide charge and the work-function of the gate electrode was determined by plotting the average flatband voltage as obtained from C-V fitting versus the EOT of the SiO_2 / Al_2O_3 gate stacks with TiN electrodes, as shown in Fig. 4.13.

The slope of the V_{FB} versus EOT plot for *as deposited* Al_2O_3 layers yields a value for the fixed charge of approximately $-2.5 \cdot 10^{12} cm^{-2}$, independent of substrate type. No significant reduction of the fixed charge was achieved by the PDA in N_2, however, an increased scatter in the flatband voltages is observed. The value for the fixed charge is consistent with previous reports. The linear dependence of V_{FB} versus EOT also supports the notion that the nature of fixed charge is a negative sheet charge located at the interface between SiO_2 and the Al_2O_3. In case

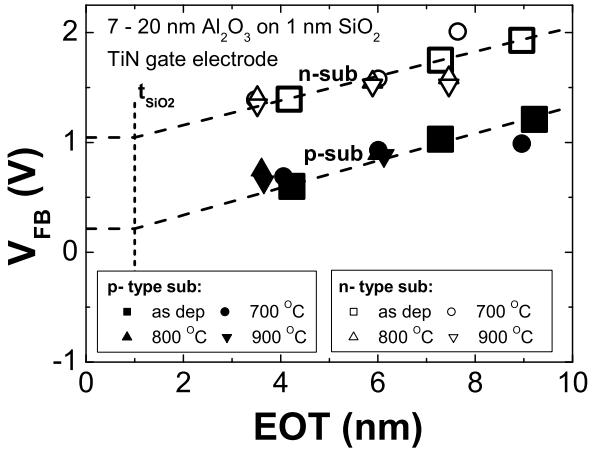

Figure 4.13: V_{FB} is plotted versus EOT. The positive slope indicates negative fixed charge in the gate dielectric. From the intercept the work-function of the gate electrode is extracted.

of homogenous fixed charge in the Al_2O_3 layer, V_{FB} versus EOT would follow a parabolic relation, which is not supported by the experimental data.

For the extrapolated V_{FB} values at an EOT of $\sim 1\,\text{nm}$ for *as deposited* layers, a flatband voltage of 0.2 and 1.05V can be extracted for p- and n-type Si substrates, respectively, yielding values of 5.2 and 5.4 eV for the work function of TiN, respectively. The extrapolation to 1 nm EOT assumes that the interfacial oxide layer is free of charge. The average value of 5.3 eV is significantly higher than the work function reported for PVD TiN on SiO_2 (4.8 eV)[43]. For annealed stacks the work function cannot easily be extracted due to increased scatter in the flatband voltage versus EOT.

4.3 MOSFET characterization

Successful MOSFET device fabrication has been demonstrated in the literature for a variety of high-ε materials like Al_2O_3 [42, 44] and HfO_2 [45], HfSiO [46] and HfSiON [37] in the past. However, beside the compatibility with a conventional poly Si gate process two major integration challenges for high-ε gate dielectrics have been identified and are going to be discussed here. The initial V_T control of n- and p-channel MOSFET is one of the issues and the device performance which is either expressed in terms of channel transconductance or carrier mobility is the other.

Typical examples for I_D-V_G and I_D-V_D characteristics of a n-channel MOSFETs with a SiO_2 / HfO_2 dual layer gate dielectric and n-type poly Si electrode are shown in Figs 4.14 and 4.15, respectively. As can be seen, well behaved I_D-V_G characteristics are obtained for both linear (V_D =0.1V) and saturation (V_D =1.5V) regime with a sub-threshold slope of $\sim 70\,\text{mV/dec}$. Furthermore, a significantly higher substrate current (I_{Sub}) is measured when a drain bias of V_D =1.5V is applied. This indicates the presence of channel hot carrier effects in high-ε materials similar to conventional gate stacks as briefly discussed in Section 2.2.1. Also the

I_D-V_D characteristic of SiO_2 / HfO_2 dual layer stacks shows a similar behavior compared to the reference stacks where linear and saturation regime can clearly be separated for the considered device geometry, as indicated in Fig. 4.15 by the dashed line.

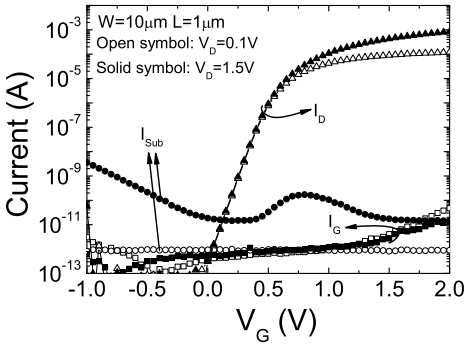

Figure 4.14: I_D versus V_G characteristic of a SiO_2 / HfO_2 n-channel MOSFET measured with a drain bias of 0.1V and 1.5V, respectively. Gate (I_G) and substrate (I_{Sub}) current are shown for comparison.

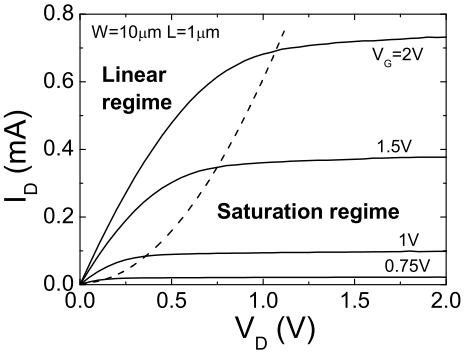

Figure 4.15: Drain current I_D versus drain voltage V_D characteristic of a SiO_2 / HfO_2 n-channel MOSFET measured for different gate bias as indicated in the figure. The dashed line indicates the transition from the linear to the saturation regime.

Beside the similarities mentioned above also strong differences in the MOSFETs characteristics are observed when comparing the n- and p-channel SiO_2 reference devices with the SiO_2 / HfO_2 dual layer stacks as shown in Fig. 4.16. In a conventional CMOS process the V_{FB} and therewith the V_T of long channel MOSFETs is determined by the Fermi-level difference between the poly Si gate electrode and the channel region. When the same type of electrodes are used to fabricate high-ε MOSFETs a significant shift in the initial V_T compared to its ideal position is typically observed. This can be seen when comparing the I_D-V_G characteristics for n- and p-channel MOSFETs in Fig. 4.16 for the SiO_2 / HfO_2 stacks with the SiO_2 reference. For n-channel FETs a positive V_T shift is observed whereas the shift is negative for p-channel FETs. This effect cannot be attributed to the presence of fixed charge in the dielectric, which would cause the same shifts (either positive or negative) for both n- and p-channel devices. The asymmetry in the V_T shift is attributed in the literature to Fermi level pinning at the poly Si / metal oxide interface [47] which is caused by metal-induced electronic states due to the presence of Hf-Si bonds at the gate electrode interface. Depending on the energy position of the defect states (either closer to the conduction or the valence band) V_T shifts of different magnitude are expected for n- and p-channel devices and indeed are experimentally observed.

Beside the inital V_T control the device performance is of major concern for integration of high-ε materials. This becomes clearly visible when the I_D-V_G characteristic measured with a V_D=0.1V is plotted on a linear scale. In Fig. 4.17 the drive current of a conventional stack (EOT $\sim$ 1.5 nm) is compared with the SiO_2 / HfO_2 stack (EOT $\sim$ 1.8 - 2.0 nm). Even though similar EOTs are extracted the MOSFET performance is significantly lower for the n-channel device whereas the p-channel device seems not to be affected.

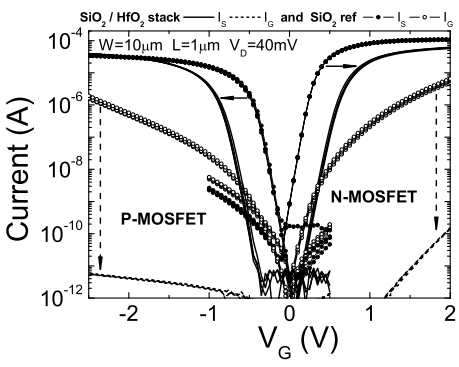

Figure 4.16: Comparison of I_D-V_G for n- and p-channel MOSFETS with SiO_2 / HfO_2 and conventional gate dielectrics. Gate current is shown to illustrate the leakage current reduction obtained with high-ε materials.

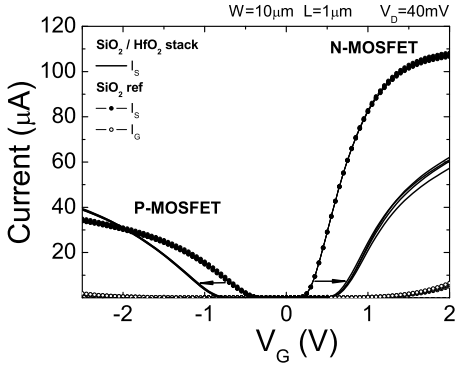

Figure 4.17: Comparison of the I_D-V_G characteristic measured in the linear regime for SiO_2 / HfO_2 and conventional gate stacks. Note the positive V_T shift for n-channel and negative V_T shift for p-channel MOSFET with respect to the SiO_2 control.

When comparing the channel transconductance of conventional gate dielectrics with the SiO_2 / HfO_2 stack, as shown in Fig. 4.18, the strong degradation of the n-channel device is again confirmed. The fact that the p-channel FET shows basically no degradation in the device performance can, to some extent, be attributed to the inherent lower carrier mobility for holes compared to electrons. Therefore, additional degradation effects (e.g. remote charge or phonon scattering) are less pronounced. However, to understand the physical mechanism responsible for the performance degradation is an important step towards the integration of such gate dielectrics.

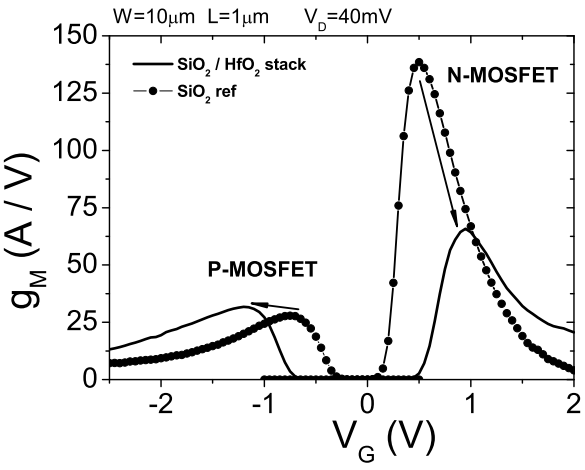

Figure 4.18: Channel transconductance for SiO_2 / HfO_2 and conventional gate stacks derived from I_D-V_G characteristic shown in Fig. 4.17.

4.4 Summary

The basic electrical characteristics of SiO_2 / high-ε gate stack were discussed in this section and can be summarized as follows.

The inherent asymmetry of a dual layer gate stacks consisting of a interfacial layer similar to SiO_2 covered with a high-ε film causes an asymmetry in the gate leakage current. For substrate injection higher leakage currents are measured compared to injection from the gate electrode. In order to meet the leakage requirements for future CMOS technologies this asymmetry needs to be taken into account. However, it is expected that sufficient leakage current reduction will be obtained with high-ε materials compared to conventional SiO_2 and SiON dielectrics.

Based on C-V analysis the impact of the deposition process and the PDA was discussed in detail for SiO_2 / Al_2O_3 gate stacks with TiN electrodes. The EOT extracted from the accumulation capacitance shows clearly a correlation with the deposition process. This results in a wedge shape of the EOT contour plot. The presence of a significant wafer non-uniformity will be considered in the breakdown studies shown in Section 7. In addition, fixed oxide charge and dielectric constant of the Al_2O_3 layer were extracted and found to be in good agreement with values reported in the literature. In general, the high fixed charge in Al_2O_3 is of major concern for integration due to issues related to device performance and V_T control, which makes Al_2O_3 a less promising candidate. Furthermore, the high-temperature anneal showed a remarkable influence on the dielectric constant which can only partially be explained by the observed densification.

The general issues concerning the performance of MOSFETs with high-ε gate dielectrics were demonstrated using SiO_2 / HfO_2 dual layer stacks with conventional poly Si gate electrodes. Two major issues beside the compatibility with poly Si electrodes are the carrier mobility and the initial V_T control. Both are considered potential issues with respect to integration into future CMOS technologies. Possible solutions may come from material engineering (e.g. incorporation of Si and N into the gate dielectric), interface engineering (using capping layers) or switching to metal gate electrodes.

5 Methodology for charge trapping

Charge trapping in conventional SiO_2 gate dielectrics is commonly studied using either hysteresis measurements or 'stress and sense' techniques. In the past, the same methodology has also been applied to study charge trapping in high-ε dielectrics. In a recent study it has been demonstrated that transient charging effects in high-ε materials (e.g. HfO_2) make a reliable assessment of the charge trapping more complicated than in SiO_2. Therefore fast measurement techniques were developed which are well suited to capture the fast transient effects previously not reported.

5.1 Conventional measurement techniques

In a conventional hysteresis measurement a quasi DC ramp is applied to the MOS device using ramp rates ranging from 0.1V/s to 10V/s. Charge trapping is monitored either using the I_D-V_G or the Capacitance-Voltage characteristic. In both cases, when charge trapping / detrapping is present, the initial trace will deviate from the final trace due to build up or loss of charge. Fast transient effects, however, are not captured by the conventional hysteresis measurement due to the inherently slow ramp rates. A schematic drawing of the gate voltage during a hysteresis measurement is shown in Fig. 5.1.

The bias sequence sketched in Fig. 5.1 was applied to MOSFET structures and either the I_D-V_G or the Capacitance-Voltage characteristic was recorded. In the following section typical results obtained for SiO_2 / HfO_2 dual layer stack are going to be discussed.

5.1.1 Multiple C-V traces

The C-V hysteresis measurement provides a fast screening method with respect to instabilities in MOS devices. Additional information on the voltage dependence of the instability can be obtained when the conventional hysteresis measurement is extended using a bias sequence as shown in Fig. 5.1. Furthermore, the use of multiple traces should also allow to address the reversibility of the observed effects.

To illustrate the observed instability in SiO_2 / HfO_2 dual layer stacks the traces are split into groups taken from negative to positive gate bias and vice versa. The

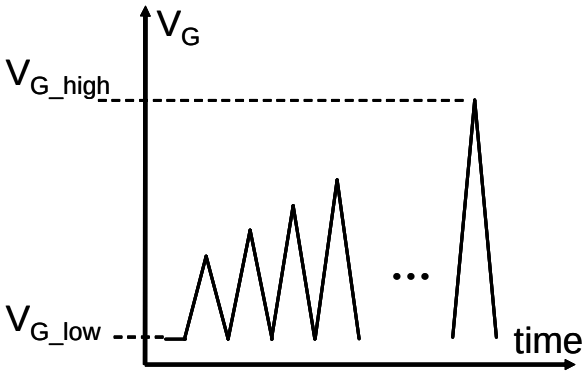

Figure 5.1: Schematic drawing of the gate bias during a hysteresis measurement applied to n-channel MOSFETs. An appropriate negative gate bias (V_{G_low}) is chosen to allow for complete discharging, whereas the high level of the gate bias (V_{G_high}) is continuously increased to monitor charge trapping in inversion.

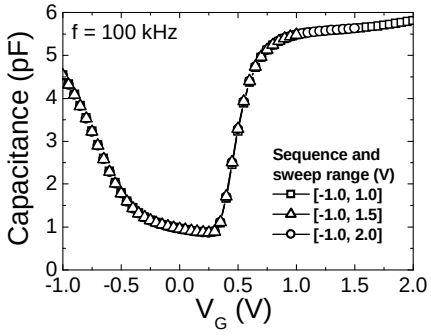

Figure 5.2: C-V characteristic of SiO_2 / HfO_2 dual layer stack measured using a starting voltage of V_{G_low}=-1V.

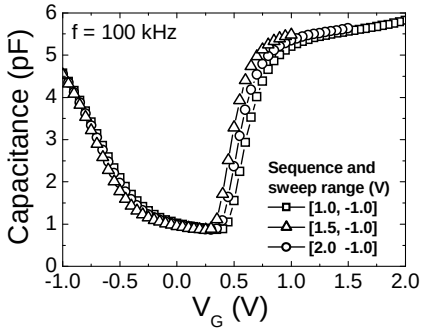

Figure 5.3: C-V characteristic of SiO_2 / HfO_2 dual layer stack measured using a starting voltage V_{G_high}=1, 1.5 or 2V.

results are summarized in Figs. 5.2 and 5.3.

When the gate bias is swept from negative to positive identical C-V traces are obtained. From this it could be concluded that the gate stack does not suffer from an instability. However, when the sweep direction is reversed from positive to negative the instability becomes evident. As can be seen in Fig. 5.3, depending on the maximum gate bias, C-V shifts of $\sim 100\,\mathrm{mV}$ and more can easily be measured. When the device is swept further into accumulation (negative gate bias) the C-V traces merge again indicating that the instability completely recovers.

The major drawback of the C-V hysteresis or the multiple C-V trace technique is the poor control over the amount of injected charge during such an experiment. Therefore the magnitude of the measured instability can significantly vary depending on the ramp rates of the sweep.

5.1.2 Multiple I_D-V_G traces

A complementary technique to the C-V hysteresis measurement is the use of multiple I_D-V_G traces on MOSFETs. Again a sequence of sweeps are taken starting from accumulation either from a negative bias for n-channel devices or from a positive

gate bias for p-channel devices and sweeping the MOSFET towards inversion successively increasing the maximum voltage. In this case the transfer characteristic is used to monitor the instability. A typical example for a n-channel MOSFET with a SiO_2 / HfO_2 dual layer stack is shown in Fig. 5.4.

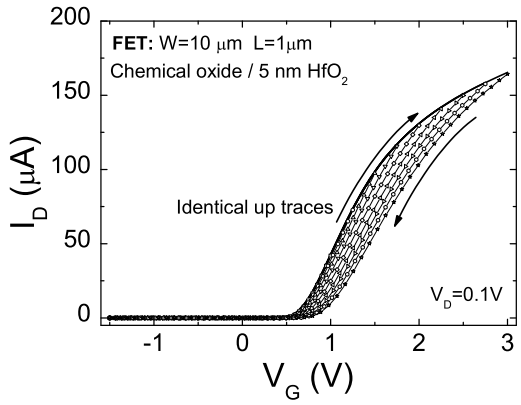

Figure 5.4: I_D-V_G hysteresis measurements were taken using the bias sequence as shown in Fig. 5.1 with a V_{G_low}=-1V.

As can be seen in Fig. 5.4, a shift of $\sim 100\,\mathrm{mV}$ and more in the I_D-V_G characteristic can easily be detected depending on the maximum positive gate bias applied to the gate. The use of the linear I_D-V_G characteristic to monitor the instability also allows to study the recovery during the down ramp of the sweep. When going to large positive gate biases, where larger shifts are observed, a stretch-out on the reverse trace is evident. This indicates that a fraction of the instability already dynamically recovers when the high gate bias is reduced.

The use of the I_D-V_G characteristic limits the monitoring capability to the operation mode of the MOS transistor, which is anyway the focus of interest from application point of view. However, also in the case of I_D-V_G instabilities, when applying a sufficient negative gate bias complete recovery is obtained, consistent with the data shown in Figs. 5.2 and 5.3.

In principle, charge injection during the multiple I_D-V_G measurement can be monitored by measuring the gate current simultaneously. Therewith the instability can be monitored either as a function of maximum gate bias or versus the total amount of injected charge.

5.1.3 Stress and sense measurements

A further procedure, which was used extensively in the literature to study the instabilities in conventional SiO_2 based gate dielectrics, is known as "stress and sense" method. In this case, first an initial sense measurement is carried out prior to stressing the devices. The stress is then interrupted periodically and a sense measurement is performed. A schematic drawing of the bias sequence in the conventional stress and sense procedure is shown in Fig. 5.5.

The instability of the device is extracted by comparing the sense measurement after stress with the initial device characteristic. Predictions to operation conditions are usually made when combining the time dependence with the voltage dependence

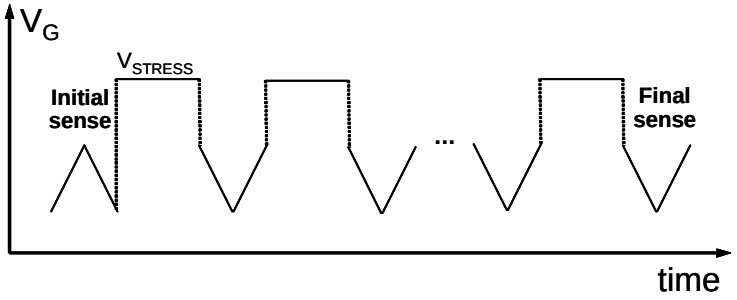

Figure 5.5: Schematic drawing of a "stress and sense" experiment. The stress applied to the device is periodically interrupted to sense the charge state by comparing the V_{FB} or V_T to the initial trace.

of the instability.

One of the known issues of the "stress and sense" procedure is the inherent time delay between stressing and sensing. In case some recovery of the instability occurs at time periods of the order of $\sim$ 10 to 100 ms this procedure will not capture its full extent. Furthermore, the selection of the sense condition requires special attention. In any case, the use of this procedure will help to understand and to improve on the stability of MOSFETs with high-ε gate dielectrics.

5.2 Novel time resolved measurement techniques

As briefly mentioned in the previous section due to the presence of fast recovery of the instabilities measurement techniques significantly faster than the semiconductor parameter analyzers are required for a correct quantification. Therefore, alternative techniques were introduced which will be discussed in the following section and later on extensively applied to high-ε gate stacks.

5.2.1 Pulsed C-V measurements

The pulsed C-V technique is similar to the quasi static C-V measurement using a linear voltage ramp. In this technique the ramp rates can vary from $\sim$0.1 V/s to $>$10 kV/s and as a result the corresponding displacement current is several orders of magnitude higher than in the conventional quasi static C-V measurement. The displacement current is converted into a voltage trace using a current−voltage amplifier and recorded with a digital oscilloscope. From the voltage trace the C-V characteristic can be extracted using the relationship

$$C(V_{IN}) = \frac{V_{OUT} - V_{OFFSET}}{Z_{AMP} \cdot dV/dt}. \tag{5.1}$$

where dV/dt is the voltage ramp, V_{OFFSET} is the offset voltage of the current−voltage amplifier and Z_{AMP} the current−voltage gain. The input (V_{IN}) and output (V_{OUT}) voltage correspond to the terminals of the circuit as illustrated in the schematic measurement setup shown in Fig. 5.6. The measurement setup was calibrated using a n-channel MOS transistor with $\sim$4.5 nm SiO_2 gate dielectric. In Fig. 5.7 the voltage trace applied to source, drain and substrate (V_{IN}) together with the voltage trace (V_{OUT}) measured at the gate due to the displacement current are shown. As can be seen the signal at the output of the current−voltage amplifier is shifted by $V_{OFFSET} \sim$ - 70mV, which need to be considered when converting the voltage traces into a capacitance. A comparison between the HF C-V and the pulsed C-V