Springer Series in Advanced Microelectronics 43

Samares Kar *Editor*

High Permittivity Gate Dielectric Materials

Springer

Springer Series in Advanced Microelectronics

Volume 43

For further volumes:
http://www.springer.com/series/4076

The Springer Series in Advanced Microelectronics provides systematic information on all the topics relevant for the design, processing, and manufacturing of microelectronic devices. The books, each prepared by leading researchers or engineers in their fields, cover the basic and advanced aspects of topics such as wafer processing, materials, device design, device technologies, circuit design, VLSI implementation, and subsystem technology. The series forms a bridge between physics and engineering and the volumes will appeal to practicing engineers as well as research scientists.

Samares Kar
Editor

High Permittivity Gate Dielectric Materials

Springer

Editor
Samares Kar
Department of Electrical Engineering
Indian Institute of Technology
Kanpur
India

ISSN 1437-0387
ISBN 978-3-642-36534-8 ISBN 978-3-642-36535-5 (eBook)
DOI 10.1007/978-3-642-36535-5
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013940446

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Welcome to the world of high permittivity gate dielectric materials and its environment.

Genesis. I began writing a book on ultrathin gate dielectrics all by myself but was making poor progress; it was not clear to me when I would be able to complete the project. Then, at one of our annual high-k symposia, Dr. Claus E. Ascheron, Executive Editor Physics, Springer Science, Heidelberg, suggested that we undertake the writing of this book, consisting of some 15 chapters on different areas of the high-k gate stacks with different chapters being written by different authors with the coordination done by me as an editor. This is how this book was conceived. A number of books of this type (edited with chapters written by different authors) exist on this subject. These books were written quite a few years ago, before MOSFETs with high-k gate stacks and metal electrodes were introduced into the market in the year 2007 by Intel and other manufacturers for the 45 nm technology node. Since then the high-k world has seen significant changes in terms of technological maturity, high-k gate stack materials, and our scientific understanding of the different aspects and issues. Hence one could argue that this book is timely.

Multiple authorship. There are different intrinsic advantages and disadvantages built into a book authored by one person and a book written by many authors. A single-author book may enjoy a higher degree of cohesion, continuity, and readability, whereas a multiple-author book will have a larger pool of scientific knowledge, wisdom, experience, and expertise. A multiple-author book to be effective will require diligence, tenacity, and finesse in coordinating the needs, activities, and the response of a large number of busy and preoccupied individuals (the authors), and the cohesion and continuity of the book will depend upon the quality of this coordination and the mutual cooperation between all the actors. Which of these two options (modes) of writing a book on this subject would have taken a longer or a shorter time is difficult to predict. One interesting point to ponder in this connection is why no single-author book has been published on the subject of this book? Is it because the complexity of the subject is beyond the effective reach of a single author? Most of the chapters have been written by a single author, except Chaps. 6, 10, and 12. In the case of Chap. 12, the part dealing with the Ge channel has been authored by Michel Houssa, the part dealing with the

GaAs and InGaAs channel has been authored by Peide Ye, and the part dealing with the III–V interface characterization has been authored by Marc Heyns. All of Chap. 6 has been written jointly by Akira Toriumi and Toshihide Nabatame; all of Chap. 10 has been jointly written by Akira Toriumi and Koji Kita.

Readership. We have tried to make the readership of the book as wide as possible—as a reference book for researchers as well as a text book for graduate students with electrical engineering, chemical engineering, materials science, or physics background.

Introduction. An overview of this book and all its chapters is presented in Chap. 1 entitled "Introduction to High-k Gate Stacks"; this introductory chapter provides simple description of the concepts and defines the terminology to be encountered in the various chapters. Chapter 1 also provides the missing links to enhance the continuity and the readability.

Coverage of the topics and comprehensiveness. We have tried to see that each chapter begins with the definitions and the basics, reviews the current literature, and ultimately graduates to the current status of the technology, and our understanding. In addition to the introduction presented in Chap. 1, we have designed Chap. 2 to provide coverage of the basics, a theoretical foundation of the high-k gate stacks, the MOS structure and the MOSFET, and the missing links, to enhance the cohesion, the continuity, and the readability of the book. In the ten chapters—Chaps. 3–12, we have tried to cover the topic as completely and comprehensively as possible. In my view all the important current and emerging areas (in addition to those covered in Chaps. 1 and 2) have been treated including physical properties of the high-k materials, hafnium-based, and lanthanide-based high-k gate stacks—their processing, characterization and characteristics, and transistor performance, properties of ternary and doped higher-k materials, crystalline high-k oxides, high-k gate stack degradation and reliability, and high-mobility channels.

Search, *reference* (*look-up*), *and readability-enhancement aids*. To facilitate search and look-up and to enhance readability, we have provided aids such as a comprehensive subject index at the end of the book, and an exhaustive table of contents, a list of abbreviations, a complete list of acronyms, and a complete list of symbols at the beginning of the book. In addition, we have provided in the seven appendices values of the fundamental constants, the periodic table, and experimental data on the physical properties of a large number of semiconductors, and high-k materials. The reader may need to look up or wish to consult these values and data while going through the theory, experimental results, and their interpretation. In other words, an attempt has been made to make the book complete with all the necessary elements under one roof.

Readability, *continuity*, *and cohesion*. Each chapter begins with an abstract and ends with a summary and a list of references. To promote readability, we have tried to see that the treatment at the beginning of the chapter is simple; and that the characteristic terms are defined and the fundamental and the basics are covered first, to help a beginner to pick up the new material. In addition, Chap. 1 (in text) and Chap. 2 (in theory) should facilitate easier entry into the world of high-k gate stacks by providing an introduction and an insight into the topics of the other

chapters and the important links missing in those chapters. We have also incorporated cross-references to sections in the other chapters dealing with the same and/or similar topic. Our list of abbreviations and our list of acronyms have no duality; however, our list of symbols is not completely cohesive in the sense that for the same parameter, we have multiple symbols in a few cases. We tried but could not completely avoid in the case of a few parameters the use of different symbols by different authors for the same parameter. Acronyms are fortunately uniform and one could say standardized by virtue of the process of their formation or by definition. Abbreviations have also near-standard forms. In contrast, symbols do not enjoy standard representations. In the list of symbols, in the few cases where multiple notations occur, we have listed all the different notations used in the different chapters of this book for those parameters. Since the notations in almost all cases are suggestive, it should not cause any great problems for the reader in recognizing what the symbol stands for.

Acknowledgment. A book of this kind would not have seen the light of day without the untiring cooperation and perseverance of all the authors and the entire publishing team—in particular, Claus E. Ascheron, Executive Editor Physics, and S. A. Shine David who executed an excellent job of setting the figures, the tables, the text, and the references to a fine form. It will take too long to elaborate on the assistance and help they extended to me. I remain deeply grateful to all of them.

Kolkata, India Samares Kar

Contents

Contributors

Gennadi Bersuker SEMATECH, 257 Fuller Road, Albany, NY 12203, USA, e-mail: gennadi.bersuker@sematech.org

Marc Heyns IMEC, Kapeldreef 75, Leuven 3001, Belgium; Department of Metallurgy and Materials Engineering, K.U. Leuven, 3001 Leuven, Belgium, e-mail: heyns@imec.be

Michel Houssa Semiconductor Physics Laboratory, Department of Physics and Astronomy, University of Leuven, Celestijnenlaan 200D, Leuven 3001, Belgium, e-mail: michel.houssa@fys.kuleuven.be

Samares Kar Indian Institute of Technology, Kanpur 208016, India, e-mail: skar@iitk.ac.in

Koji Kita Department of Materials Engineering, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan, e-mail: kita@adam.t.u-tokyo.ac.jp

Daniel J. Lichtenwalner Cree, Inc., 4600 Silicon Drive, Durham, NC 27703, USA, e-mail: daniel_lichtenwalner@cree.com

Toshihide Nabatame Association of Super-Advanced Electronics Technologies, MIRAI, Tsukuba West 7, Tsukuba, Ibaraki 305-8569, Japan, e-mail: toshihide.nabatame@aist.go.jp

Akira Nishiyama Corporate R&D Center, Toshiba Corporation, 1 Komukai-toshiba-cho, Saiwai-ku, Kawasaki 212-8582, Japan, e-mail: akira.nishiyama@toshiba.co.jp

Masaaki Niwa Center for Innovative Integrated Electronic Systems, Tohoku University, Sendai 980-0077, Japan, e-mail: niwa.masaaki@cies.tohoku.ac.jp

H. Jörg Osten Institute of Electronic Materials and Devices, Leibniz University, Schneiderberg 32, D-30167 Hannover, Germany, e-mail: osten@mbe.uni-hannover.de

Jamie K. Schaeffer GLOBALFOUNDRIES, Dresden 01109, Germany, e-mail: jamie.schaeffer@globalfoundries.com

Akira Toriumi Department of Materials Engineering, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan, e-mail: toriumi@material.t.u-tokyo.ac.jp

Peide Ye School of Electrical and Computer Engineering, Birck Nanotechnology Center, Purdue University, West Lafayette, IN 47907, USA, e-mail: yep@purdue.edu

Chadwin Young Materials Science and Engineering Department, University of Texas at Dallas, 800 W. Campbell Road, RL10, Richardson, TX 75080, USA, e-mail: chadwin.young@utdallas.edu

Acronyms

AC	Alternating Current
AFM	Atomic Force Microscopy
ALD	Atomic Layer Deposition
ALE	Atomic Layer Epitaxy
BTI	Bias Temperature Instability
CB	Conductance Band
CBM	Conduction Band Minimum
CET	Capacitance Equivalent Thickness
CM	Clausius–Mossotti
CMOS	Complimentary Metal Oxide Semiconductor
CMOSFET	Complimentary Metal Oxide Semiconductor Field Effect Transistor
CMP	Chemical Mechanical Polishing
CN	Coordination Number
CNL	Charge Neutrality Level
CP	Charge Pumping
CRN	Continuous Random Network
CVD	Chemical Vapor Deposition
CVS	Constant Voltage Stress
DC	Direct Current
DCIV	Direct Current IV
DIBL	Drain Induced Barrier Lowering
DT	Direct Tunneling
EELS	Electron Energy Loss Spectroscopy
EN	Electro-Negativity
EOT	Equivalent Oxide Thickness
ESR	Electron Spin Resonance
EWF	Effective Work Function
EXAFS	Extended X-ray Absorption Fine Structure Spectroscopy
FEOL	Front End of Line
FET	Field Effect Transistor
FGA	Forming Gas Annealing
FLP	Fermi Level Pinning
FNT	Fowler–Nordheim Tunneling
FOM	Figure of Merit

FT	Fourier Transform
FTC	Fast Transient Charging
FTCE	Fast Transient Charging Effect
FUSI	Fully Silicided
GIDL	Gate Induced Drain Leakage
HAADF	High-Angle Annular Loss Spectroscopy
HBD	Hard Break-Down
HBT	Hafnium-Tetra-Tertiary-Butoxy
HEMT	High Electron Mobility Transistor
HF	High Frequency
HOT	Hybrid Orientation Technology
HTTB	Hafnium-Tetra-Tertiary-Butoxy
IC	Integrated Circuit
IL	Intermediate Layer
IPE	Internal Photo Emission
IPES	Internal Photo-Emission Spectroscopy
IR	Infra Red
ISSG	In Situ Steam Generation
ITO	Indium Tin Oxide
ITRS	International Technology Roadmap for Semiconductors
JVD	Jet Vapor Deposition
LF	Low Frequency
LSI	Large Scale Integration
MBE	Molecular Beam Epitaxy
MEIS	Medium Energy Ion Scattering
MESFET	Metal Semiconductor Field Effect Transistor
MG	Metal Gate
MIGS	Metal Induced Gap States
MIM	Metal Insulator Metal
MIS	Metal Insulator Semiconductor
MISFET	Metal Insulator Semiconductor Field Effect Transistor
MO	Metal Oxide
MO	Metal-Organic
M–O	Metal–Oxygen
MOCVD	Metal Organic Chemical Vapor Deposition
MOS	Metal Oxide Semiconductor
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
MS	Metal Semiconductor
NBTI	Negative Bias Temperature Instability
NC	Nano-Cluster
NFET	N-Channel Field Effect Transistor
NMOS	N-Channel Metal Oxide Semiconductor
NMOSFET	N-Channel Metal Oxide Semiconductor Field Effect Transistor
PBD	Progressive Break-Down
PBTI	Positive Bias Temperature Instability

PDA	Post Deposition Annealing
PFET	P-Channel Field Effect Transistor
PLD	Pulsed Laser Deposition
PMA	Post Metallization Annealing
PMOS	P-Channel Metal Oxide Semiconductor
PMOSFET	P-Channel Metal Oxide Semiconductor Field Effect Transistor
PVD	Physical Vapor Deposition
PW	Pulse Width
QS	Quasi Static
RC	Resistor Capacitor
RCA	Radio Corporation of America
RCP	Random Close Packed
REELS	Reflection Electron Energy Loss Spectroscopy
RHEED	Reflection High Energy Electron Diffraction
RIE	Reactive Ion Etching
RPO	Remote Plasma Oxidation
RTA	Rapid Thermal Annealing
RTD	Resonant Tunneling Diodes
RTN	Rapid Thermal Nitridation
RTO	Rapid Thermal Oxidation
RTP	Rapid Thermal Processing
SAD	Selected Area Diffraction
SBD	Soft Break-Down
SCE	Short Channel Effect
SD	Single Domain
SDH	Silicon Doped Hafnia
SDR	Spin Dependent Recombination
SE	Spectroscopic Ellipsometry
SHC	Substrate Hot Carrier
SILC	Stress Induced Leakage Current
SIMS	Secondary Ion Mass Spectroscopy
SOI	Silicon on Insulator
SPE	Solid Phase Epitaxy
SS	Subthreshold Swing
STI	Shallow Trench Isolation
TAT	Trap-Assisted Tunneling
TDDB	Time-Dependent Dielectric Breakdown
TDEAH	Tetrakis Di-Ethyl Amino Hafnium
TDMAH	Tetrakis-Di-Methyl Amino Hafnium
TDMAS	Tetrakis Di-Methyl Amino Silane
TEM	Transmission Electron Microscope/Microscopy
TEM/EDX	Transmission Electron Microscope/Energy Dispersive X-ray Spectroscopy
TEMAH	Tetrakis-Ethyl-Methyl-Amino-Hafnium
TEOS	Tetra Ethoxy Silane

TMA	Tri-Methyl-Aluminum
UHF	Ultra High Frequency
UHV	Ultra High Vacuum
ULSI	Ultra Large-Scale Integration
VB	Valence Band
VBM	Valence Band Maximum
VL	Vacuum Level
WF	Work Function
XPS	X-ray Photoelectron Spectroscopy
XRD	X-ray Diffraction
XRR	X-ray Reflectivity
XTEM	Cross-Sectional TEM
YDH	Yttrium-Doped Hafnia
YSZ	Yttria-Stabilized Zirconia

Symbols

α	Constant
α	Coeffecient of thermal expansion
α_{acc}	Pre-factor for C_p^{acc}
α_e	Electronic polarization
α_{ion}	Ionic polarization
α_m	Microscopic polarizability
α_{total}	Static polarizabilities
β	Constant
β	Inverse thermal voltage = q/kT
β	MOSFET quality factor; transconductance coefficient
β	Weibull slope
β_{acc}	Accumulation surface potential quotient in the exponent
γ	Body factor
ε_{di}	Dielectric permittivity
$\varepsilon_{di,high\text{-}k}$	High-k layer permittivity
$\varepsilon_{di,IL}$	IL permittivity
ε_{HfO_2}	Permittivity of HfO_2
ε_s	Semiconductor permittivity
ε_{SiO_2}	Permittivity of the SiO_2 layer
ε_0	Permittivity of vacuum
η	Constant
$\theta/2\theta$	Bragg angle
κ	Attenuation constant for electron wave function
κ	Dielectric permittivity
κ	Dielectric constant
κ_h	Attenuation constant for hole wave function
λ	Characteristic thickness of bottom high-k layer
μ	Mobility
μ_0	Permeability of vacuum
μ_{ch}	Mobility of carriers in channel
μ_{coul}	Mobility related to Coulomb scattering
μ_e	Electron mobility
μ_{eff}	Effective carrier mobility
μ_h	Hole mobility

μ_{phonon}	Mobility related to phonon scattering
μ_{rough}	Mobility related to surface roughness scattering
ρ	Volume space charge density
ρ	Specific density
$\rho_{high\text{-}k}$	Volume charge density of the traps in the high-κ layer
ρ_{IL}	Volume charge density of the traps in the intermediate oxide layer
ρ_1	HfO_2 bulk charge density
σ	Number of oxygen atoms per unit area
σ, σ_T, σ_t, σ_{it}	Trap capture cross-section
σ_{ch}	Channel conductivity
σ_e	Cross-section for capture of electrons
σ_h	Cross-section for capture of holes
$\sigma_{high\text{-}k}$	Number of oxygen atoms per unit area in high-k layer
σ_s	Trap emission cross-section
σ_{SiO_2}	Number of oxygen atoms per unit area in SiO_2
τ	Detrapping time constant
τ	Mean free time
τ	Time constant
$\tau_{channel}$	Channel traverse time
τ_{drift}	Drift time through the space charge layer
$\tau_{generation}$	Minority carrier generation time
τ_h	Hole time constant
τ_{inv}	Inversion layer time constant
τ_{it}	Interface trap time constant
$\tau_{relaxation}$	Majority carrier relaxation time
$\tau_{R,maj}$	Majority carrier interface recombination time
τ_T, τ_t, $\tau_{tunneling}$	Tunneling time
ϕ	Energy barrier
ϕ_b	Barrier height
ϕ_b	Schottky barrier height
$\phi_{bn,0}$	Schottky barrier height for n-type semiconductor at zero bias
$\phi_{b,c}$	Conduction band offset
$\phi_{b,v}$	Valence band offset
ϕ_{CNL}	Charge neutrality level
$\phi_{CNL,d}$	Charge neutrality level
ϕ_f	Fermi potential in Si
ϕ_m, ϕ_M	Metal work function
$\phi_{m,eff}$	Effective metal work function
$\phi_{m,vac}$	Metal work function in vacuum
$\phi_{m,eff,high\text{-}k}$	Effective work function of metal on high-k layer
ϕ_{m,eff,SiO_2}	Effective work function of metal on SiO_2
ϕ_{MS}	Metal-semiconductor work function difference

ϕ_p	Fermi potential in p-type silicon
ϕ_s	Fermi level of the silicon
$\phi^{C}_{Si,eff,high\text{-}k}$	Effective work function of polysilicon on high-k after correction for bottom dipole contribution to V_{FB}
ϕ_{si,eff,SiO_2}	Effective work function of polysilicon on SiO_2
ϕ^{Top}_{FLP}	Threshold voltage correction due to Fermi level pinning at the top interface
ϕ^{Bottom}_{Dipole}	Bottom interface dipole voltage
$\delta\phi_X$	Work function anomaly
χ_d, χ_{di}	Electron affinity of the dielectric
χ_s	Semiconductor electron affinity
φ	Potential in the space charge region
$(\varphi_s)_p$	Value of φ_s, at which (G_p/ω) peaks
$\varphi_{s,inv}$	Surface potential in strong inversion
$\varphi_{s,inv,0}$	Surface potential at source in strong inversion
$\varphi_{s,inv,th}$	Surface potential at the onset of strong inversion
$\varphi_{s,inv,th,0}$	Surface potential at source at the onset of strong inversion
$\varphi_{s,inv,th,photo}$	Surface potential at the onset of strong inversion under illumination
$\varphi_{s,0}$	Surface potential at source
$\Delta\varphi_{s,inv}$	Excess surface potential in strong inversion
ω	Angular frequency
ω_o	Characteristic phonon frequency
Φ_b	Barrier height
Φ_b	Band offset
Φ_{ms}	Metal-semiconductor work function difference
X	Electronegativity
Ψ	Electron wave function
Ω	Resistivity
A	Richardson constant
A	Area
a	Lattice constant
a_{film}	Lattice constant of film
a_{Si}	Lattice constant of Si
C	Concentration
C	Capacitance
C	Total MOS capacitance density
c	Force constant
c	Speed of light in vacuum
C_{acc}	Accumulation capacitance density
C_{dep}	Depletion capacitance density
C_{di}	Gate stack capacitance density
$C_{di,high\text{-}k}$	Areal density of the plane-parallel capacitance of the high-k layer

$C_{di,IL}$	Areal density of the plane-parallel capacitance of the intermediate-oxide layer bulk
C_{gate}, C_g	Gate capacitance
C_{gb}	Gate-to-bulk capacitance
C_{gc}	Gate-to-channel capacitance
C_{HF}, C_{hf}	High-frequency capacitance density
C_i, C_{ox}	Oxide capacitance
C_{inv}	Inversion capacitance
C_{it}	Interface trap capacitance density
$C_{it,acc}$	Interface trap capacitance density in accumulation
$C_{it,fb}$	Interface trap capacitance density at flat band
C_{LF}, C_{lf}	Low-frequency capacitance density
C_m	Measured capacitance
C_{max}	Maximum capacitance density
CN_a	Anion coordination number
CN_c	Cation coordination number
C_{ox}	Oxide capacitance density
C_p	Parallel capacitance density
$C_{p,acc}$	Parallel capacitance density in accumulation
$C_{p,fb}$	Parallel capacitance density at flat band
$C_{p,inv,photo}$	Parallel capacitance density in strong inversion under illumination
C_{poly}	Capacitance of the poly-Si gate electrode
C_s	Interface state capacitance
C_{sc}	Space charge capacitance density
$C_{sc,acc}$	Space charge capacitance density in accumulation
$C_{sc,fb}$	Space charge capacitance density at flat-band
$C_{sc,hf}$	Space charge capacitance density at high frequency
$C_{sc,photo}$	Space charge capacitance density under illumination
C_{sub}	Capacitance due to quantum mechanical effects which force the centroid of inversion charge in the substrate away from the Si/SiO_2 interface
C_{tot}	Total capacitance
d	Distance from substrate
D	Electric displacement vector
D_{bt}	Bulk trap density
d_{c-a}	Cation–anion distance
D_t, D_{it}, D_{IT}	Density of interface traps
D^A_{it}	Density of acceptor traps at interface
D^D_{it}	Density of donor traps at interface
E	Energy
E	Eigenenergy
E	Electron energy
E	Electric field
e	Magnitude of electronic charge

$E_{applied}$	Applied electric field
E_C, E_c	Conduction band edge
E_{CNL}	Charge neutrality level
E_{eff}	Effective field
E_{FM},$E_{F,m}$	Metal Fermi level
E_{FS}	Silicon Fermi level
E_{FS}^{e}	Electron imref
E_{FS}^{h}	Hole imref
E_g, E_G	Band gap
E_i	Electric field at interface
E_i	Activation energy
E_i	Local field
E_i	Intrinsic level
E_{max}	Maximum electric field
EOT	Effective oxide thickness
EOT_1	Effective oxide thickness of the HfO_2 interface layer
EOT_2	Effective oxide thickness of the SiO_2 layer
$EOT_{threshold,maj}$	Threshold EOT at which majority carrier imref gets pinned to the metal Fermi level
$EOT_{threshold,min}$	Threshold EOT at which minority carrier imref gets pinned to the metal Fermi level
E_T, E_t, E_{it}	Trap energy
E_V, E_v	Valence band edge
E_{vac}	Vacuum level
f	Frequency
F_0	Free energy
f_D	Surface potential fluctuation factor for trap density
f_{FD}^{e}	Fermi–Dirac occupancy for electrons
f_{FD}^{h}	Fermi–Dirac occupancy for holes
f_h	High frequency
f_l	Low frequency
$FM_{high\text{-}k}$	Figure of merit for a high-k gate dielectric
f_σ	Surface potential fluctuation factor for capture cross-section
G	Total MOS conductance density
g_d	Channel conductance
G_{dc}	DC MOS conductance
G_m	Measured conductance density
g_m	Transconductance
G_p	Equivalent parallel conductance density
$(G_p/\omega)_p$	Peak (maximum) value of (G_p/ω)
G_{TE}	Conductance due to thermionic field emission
G_{TS}	Tunneling conductance due to j_{TS}
h	Planck's constant
$\hbar$	Planck's constant/2π
H_{fin}	Height of fin

I	Current
I_B	Substrate current
I_{cp}	Charge pumping current
I_d, I_D, I_{ds}, I_s, I_{DS}	Drain current
I_{DS}, I_{dsat}	Saturation drain current
I_g	Leakage current
I_{off}	Off current
I_{on}	On current
J_{DT}	Direct tunneling current density
J, J_g	Gate leakage current density
J_s	Flux of electrons emitted from traps by thermal excitation
j_{TS}	Density of tunneling current into the interface states from the metal
k	Boltzmann constant
k	Wave vector
$k_\perp$	Perpendicular wave vector
k, k_d	Dielectric constant
$k_{\|\|}$	Parallel wave vector
k_{cap}	Dielectric constant of the cap layer
k_{eff}	Effective dielectric constant
k_{el}	Dielectric constant component due to electronic polarization
k_{HfO_2}	Dielectric constant of the HfO_2 layer
$k_{high\text{-}K}$	Dielectric constant of the high-k layer
k_{IF}	Dielectric constant at the interface
k_{IL}	Dielectric constant of the IL
k_{ox}, k_{SiO_2}	Dielectric constant of the SiO_2 layer
L	Channel length
L_g, L_G	Gate length
m	Free electron mass
M	Number of fins
m	Microscopically induced polarization
M^-	Anion mass
m*	Effective mass
M*	Reduced mass
M^+	Cation mass
m_e^*	Electron effective mass
m_h^*	Hole effective mass
m^*hh	Heavy hole effective mass
m^*l	Longitudinal electron effective mass
m^*lh	Light hole effective mass
m^*t	Transverse electron effective mass
m_t^*	Effective tunneling mass of the carrier
n	Power law exponent

n	Diode quality factor
n	Electron density
N	Trap density
n	Density of filled fast traps
n	Optical refractive index
n_0	Electron density in neutral region
N_0	Density of pre-existing precursor defects
N_A	Acceptor density
N_A	Avogadro constant
N_A^-	Ionized acceptor density
N_c	Effective density of states in the conduction band
N_D^+	Ionized donor density
n_i	Intrinsic carrier density
N_{inv}	Free carrier density in strong inversion
N_{IT}, N_{it}	Interface trap density per area
N_m	Number density of microscopic polarization
N_{poly}	Doping density in poly-Si
n_s	Interface electron density
N_s	Density of unoccupied secondary traps
N_s	Inversion charge
N_{sub}	Doping concentration in the substrate
N_t	Trap charge
N_t	Hole trap density
N_v	Effective density of states in the valence band
p	Hole density
P	Macroscopic polarization
p_0	Hole density in neutral region
p_{O_2}	Partial pressure of oxygen
p_s	Interface hole density
p_s	Probability of electron capture
q	Fundamental charge
Q	Charge density
Q_1	Fixed charge density at the SiO_2/HfO_2 interface
Q_2	Fixed charge density at the Si/dielectric interface
Q_{ch}	Channel charge density
Q_{dep}	Depletion charge density
Q_{fix}, Q_f, Q_F	Fixed charge density
Q_{gsc}, $Q_{di,gsc}$	Gate stack charge density
Q_{inv}	Inversion charge density
Q_{it}	Interface trap charge density
$Q_{it,fb}$	Interface trap charge density at flat band
$Q_{it,inv}$	Interface trap charge density in strong inversion
Q_M	Metal charge density
Q_{sc}	Areal space charge density

$Q_{sc,inv}$	Space charge density in strong inversion
Q_t	Trapped charge
R	Resistance
r	Position vector
R_B	Bulk resistance
R_{diff}	Differential resistance
R_{it}	Interface trap resistance density
R_s	Series resistance
R_s^{maj}	Interface state resistance for majority carrier recombination
S	Schottky pinning parameter
S	Electrode area
$S_{r,diff}$, $S_{R,diff}$	Slope of differential resistance
T	Absolute temperature
T	Oxidation temperature
T, t	Time
t	Physical layer thickness
t_{bd}	Time to breakdown
t_c	Critical trapping time
t_c, t_{ch}	Channel thickness
T_{ther}	Crystallization temperature
t_{di}	Total physical thickness of the gate stack
$t_{di,cap}$	Thickness of the cap layer
$t_{di,high\text{-}k}$	Thickness of the high-κ layer
$t_{di,IL}$	Thickness of the intermediate oxide layer
$t_{Discharge}$	Discharge time
t_{eq}	Equivalent thickness
t_f	Fall time
$t_{Hi\text{-}k}$, $t_{high\text{-}K}$	Thickness of the high-k layer
t_{IF}	Thickness of the intermediate film
t_{IL}	IL thickness
T_{OX}, t_{ox}	Physical oxide thickness
T_{ph}, T_{phys}	Physical thickness
t_r	Rise time
t_{SiO_2}	Thickness of the SiO_2 layer
t_{STRESS}	Stress time
t_{total}	Total thickness
u	Periodic potential
V	Voltage, Bias
v	Average thermal velocity
V(y)	Bias voltage at point y in the channel
V(z)	Potential energy
V_0	Neutral oxygen vacancy
V_o^{2+}	Ionized oxygen vacancy
V_B	Applied bias across MOS structure
V_{base}	Base Voltage

v_d	Drift velocity
V_{DD}	Supply voltage
V_{di}	Total potential across the gate stack
$V_{di,dipole}$	Potential of the interface dipole at the IL/high-k interface
$V_{di,gsc}$	Gate stack potential due to the gate stack charge alone
$V_{di,high\text{-}k}$	Potential across the high-k layer
$V_{di,IL}$	Potential across the intermediate layer (IL)
$V_{di,inv}$	Gate stack potential in strong inversion
$V_{di,sc}$	Gate stack potential due to the semiconductor space charge alone
$V_{di,sc,0}$	Gate stack potential due to the semiconductor space charge alone at the source
$V_{di,sc,inv,L}$	Gate stack potential due to the semiconductor space charge alone in strong inversion at the drain
$V_{di,sc,L}$	Gate stack potential due to the semiconductor space charge alone at the drain
$V_{discharge}$, $V_{Discharge}$	Discharge voltage
V_{DS}	Saturation drain voltage
V_{ds}, V_{DS}, V_D	Drain voltage
v_e	Average thermal velocity of electrons
V_{FB}^{∞}	Flat-band voltage for infinite bottom high-k layer thickness
V_{FB}, V_{fb}	Flat-band voltage
V_{FB}^{0}	Flat-band voltage for zero bottom high-k layer thickness
V_{FB}^{exp}	Flat-band voltage for a given bottom high-k layer thickness
V_{FB}^{Norm}	Normalized flat-band voltage
V_G, V_g	Gate voltage
v_h	Average thermal velocity of holes
V_m	Molar volume
V_{ox}	Oxide potential
V_{stress}	Stress voltage
V_{TH}, V_T, V_{th}, V_t	Threshold voltage
V_{thr}	Threshold voltage after relaxation
V_u	Volume of the structure containing a single oxygen atom
W	Weight related to FOM component
W	Channel width
W	Space charge width
W_{GATE}	Gate Width
x	Molar fraction
x_{max}	Position from the silicon surface of the most distant traps in the gate stack which communicate with the silicon surface
x_t	Location of the trap in the gate stack
y	Direction along the channel
y_p	Pinch-off point
z	Position vector
Z	Atomic number

ΔE_c	Conductance band offset
ΔE_v	Valence band offset
ΔEN	Electro-negativity difference
ΔV_t	Threshold voltage shift
δE_{FS}	Imref separation

Chapter 1
Introduction to High-k Gate Stacks

Samares Kar

Abstract The manifold aim of this chapter is: (1) to present a simple summary of the contents of the eleven other chapters of the book in a manner as continuous and cohesive as possible; (2) more importantly, to provide the basics, the definitions, simple explanations, and the missing links; (3) and to acclimatize the uninitiated reader. This chapter begins with an historical account going back some four decades when research was initiated into the Metal Oxide Semiconductor (MOS) tunnel diodes much ahead of its commercial adoption for nano-transistors. An introduction is then presented into the desirable characteristics of a current and future high-k gate stack followed by a discussion of the properties of the available and possible high-k candidates to replace the single SiO_2 gate dielectric. The subsequent sections of the chapter deal with: (a) the basics, theory, and characteristics of the MOS structures and the MOS field effect transistor including its energy band diagrams, equivalent circuits, admittance-voltage and current–voltage characteristics, and parameter extraction methodologies; (b) the physics of the Hf-based gate stacks, which is the current favorite; (c) the processing of the Hf-based gate stacks including process optimization and control; (d) metal gate electrodes, work-function tuning, and metal gate integration; (e) the flat-band and threshold voltage anomaly including the role of the bottom and the top interface dipoles in this anomaly; (f) the channel mobility including the scattering mechanisms, the factors behind its degradation, and the options for attaining high mobility; (g) the mechanisms of gate stack degradation including fast and slow charge trapping and the reliability issues; (h) physics and technology of Ln-based gate stacks—perhaps the next generation; (i) the ternary higher-k gate dielectric materials including the roles of the structure modifiers and the higher-k phase stabilizers; (j) the crystalline high-k oxides including their novel applications; and (k) high mobility channels including their surface passivation and electrical characterization. This chapter concludes with a model for the figure of merit for evaluating the high-k material for gate stack application.

S. Kar (✉)
Department of Electrical Engineering, Indian Institute of Technology, Kanpur 208016, India
e-mail: skar@iitk.ac.in

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5_1,

1.1 Thin Tunneling SiO_2 Single Gate Dielectric

The precursor to the current high-k gate stacks is the thin (say, <4 nm thick) dry thermal (i.e. grown by high temperature thermal oxidation of the Si substrate) SiO_2 single gate dielectric (The dry thermal SiO_2 is a true dielectric in the sense that there are no charges inside its bulk, and it can be represented by a dielectric capacitance alone.), and the MOS/MIS (Metal Oxide/Insulator Semiconductor) tunnel structure on non-degenerate silicon. Peter Gray [1] was one of the earliest investigators to look into these tunnel structures in 1965 (in keeping with his guide John Bardeen's keen interest in surface/interface states, Schottky barrier, and tunneling). Significant basic analysis of the MOS/MIS tunnel structures was carried out, among others, by Kar and Dahlke [2–4], Card and Rhoderick [5, 6], and Green and Shewchun [7, 8], some 40 years ago. (These basics would be applicable to all gate dielectrics, including the high-k materials.) However, the prevailing wisdom in the microelectronics community at that time, and for a long time to come, was that the leaky (say, current density $>1 \times 10^{-9}$ A cm^{-2}) tunnel oxides would never see the light of day, when it came to their industrial application as Complimentary Channel MOS Field Effect Transistor (CMOSFET) gate dielectric.

The experimental investigation of Si/2–4 nm SiO_2/Metal structures by Kar and Dahlke [2–4] made the following basic contribution:

1. Equivalent circuit representation (see Fig. 1.1) of the MOS/MIS tunnel structures were developed, based on which, an MOS admittance technique was outlined for extracting the electronic parameters of these tunnel structures, including all the interface trap parameters, from the small signal capacitance and conductance measurement. The experimental results of Kar and Dahlke [3] confirmed that the Direct Current (DC) conductance could be easily and correctly separated from the trap conductance, using the superposition principle; this was a crucial issue in the development of the MOS tunnel admittance technique. This contribution (the equivalent circuit and the tunnel admittance technique) has been the basis of all the past and the current parameter extraction methodologies (conductance technique, low–high capacitance technique, etc.), which make use of the measured small signal admittance characteristics (C–V–f, G–V–f, I–V) of the ultrathin gate stacks. (C is capacitance density, G is conductance density, I is current, V is applied bias, and f is frequency.)
2. Perhaps, for the first time, Kar and Dahlke [3] confirmed experimentally that impurity elements at the interface can generate characteristic traps (see Table 1.1).
3. The experimental results of Kar and Dahlke [3] indicated that, in the MOS configuration, if the SiO_2 layer thickness was less than about 3.3 nm, but thicker than 1.3 nm, no strong inversion layer formed, but the interface trap occupancy was still controlled by the bulk Si majority carrier Fermi level.

Perhaps, the most important contribution made by Card and Rhoderick [5, 6] to the basics of the Si/SiO_2/Metal tunnel structures was a simple and elegant

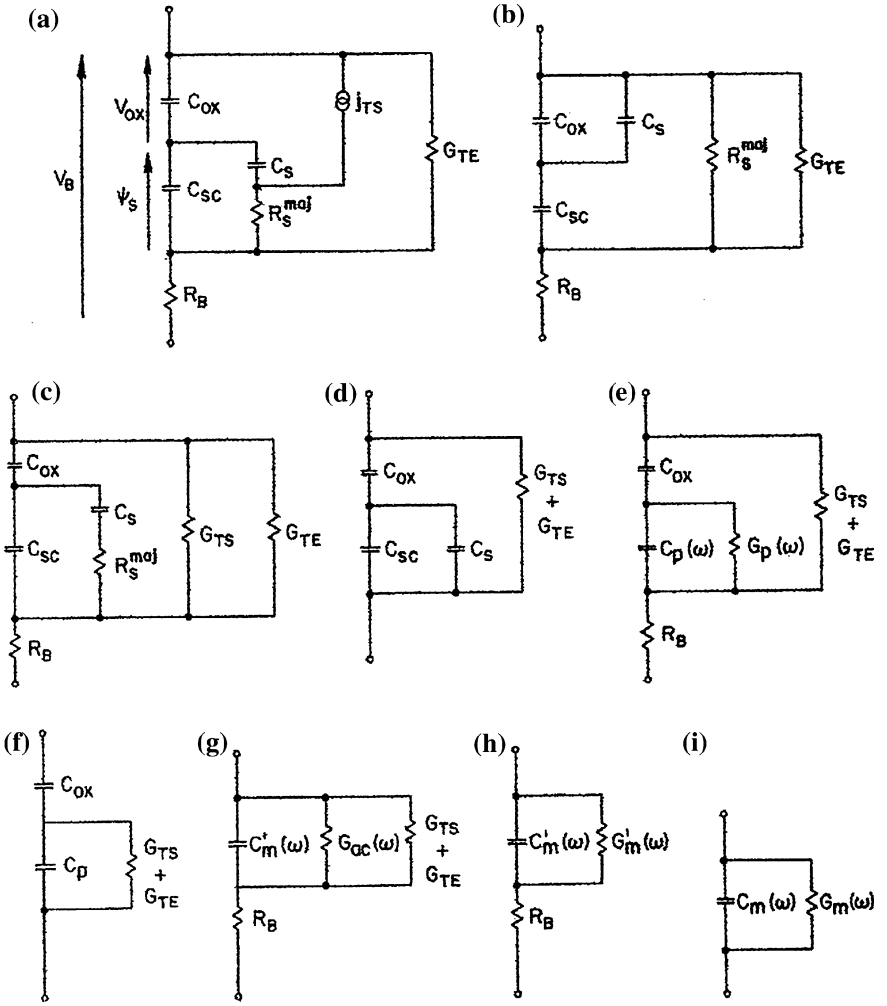


Fig. 1.1 Equivalent circuits for the intermediate MOS($20\ \text{Å} \leq t_{ox} \leq 40\ \text{Å}$) structure. **a** General circuit for a single interface state level: C_{ox} oxide capacitance, C_{sc} space charge capacitance, C_s interface state capacitance, R_s^{maj} interface state resistance for recombination of majority carriers, R_B bulk and back contact resistance, G_{TE} conductance due to thermionic field emission current, j_{TS} density of current tunneling from the metal through the oxide into the interface state, V_B applied bias, φ_s surface potential, V_{ox} oxide voltage. **b** Equivalent circuit **a** simplified for $\tau_T \ll \tau_R^{maj}$, i.e. the interface recombination controlled case. **c** Circuit **a** simplified for $\tau_T \gg \tau_R^{maj}$, i.e. the oxide tunneling controlled case: G_{TS} tunneling conductance due to j_{TS} **d** Circuit **c** at low, i.e. equilibrium, frequency. **e** Circuit **c** transformed, also valid for a continuum of interface states. **f** Circuit **e** at low, i.e. equilibrium, frequency. **g–i** Successive reductions of circuit **e**. τ_T is the tunneling time, τ_R^{maj} is the majority carrier interface recombination time, C_p, G_p are equivalent parallel capacitance and conductance, $C_m(\omega)$ and $G_m(\omega)$ are the measured device capacitance and conductance, and ω is the angular frequency. (After Kar and Dahlke [3])

Table 1.1 Properties of interface traps caused by impurity elements at or near the interface. (After Kar and Dahlke [3])

Element	Trap energy above the silicon valence band edge (eV)	Capture cross-section (cm^2)
Magnesium, Mg	0.54	1.1×10^{-18}
Chromium, Cr	0.20–0.21	5.2×10^{-17}–1.1×10^{-15}
	0.52	6.6×10^{-10}
Copper, Cu	0.20–0.26	1.1×10^{-17}–3.9×10^{-16}
Gold, Au	0.97	2.2×10^{-16}

formulation of the density of the direct tunneling current through the thin gate dielectric, J_{DT}, as represented in the following relation [5]:

$$J_{DT} = AT^2 \exp\left[-\sqrt{\phi m^*} t_{ox}\right] \exp(-q\phi_b/kT)[\exp(qV/nkT) - 1] \tag{1.1}$$

A is the Richardson constant, T is the absolute temperature, ϕ is the height of a rectangular potential barrier (in eV) equivalent to the actual barrier in SiO_2, m* is the effective carrier mass (in kg), t_{ox} is the SiO_2 thickness (in Å), q is the electron charge, ϕ_b is the Schottky barrier height, k is the Boltzmann constant, V is the applied bias across the MOS tunnel structure, and n is the diode quality factor. Card and Rhoderick correctly recognized that, in the device grade structures, the dominant carrier transport process across the combined potential barrier extending over the silicon space charge layer and the SiO_2 insulator, will be the thermionic (Schottky) emission over the silicon barrier, followed by direct tunneling through the SiO_2 band gap, i.e. the thermionic emission current could be multiplied by the tunneling transmission coefficient $[\exp - (\phi m^*)^{1/2} t_{ox}]$ to yield not just a closed-form, but also, a simple mathematical expression for the direct tunneling current density. In a device-grade high-k gate stack, particularly, in the low gate voltage regime, the dominant gate leakage current may also be a direct tunneling current. For the ultrathin high-k gate stacks, the formulation by Card Rhoderick has been the basis for all the closed-form (although far more complicated and far less elegant) as well as numerical expressions for the direct tunneling current across the gate stack.

The investigation by Green and Shewchun of MIS tunnel diodes [7, 8] made the following basic contribution. Generally, as mentioned above, the MOS/MIS tunnel structures are majority carrier devices, i.e. the dominant carrier transport mechanism is thermionic emission of the majority carriers over the Schottky barrier in silicon, followed by direct tunneling through the gate stack. Green and Shewchun [7, 8] demonstrated that, if the Schottky barrier in silicon is high enough, i.e. close to the Si band gap, such that the thermionic emission over the Schottky barrier is suppressed, then minority carrier injection from the metal electrode into the silicon minority carrier band is likely to be the dominant carrier transport mechanism, resulting in minority carrier MOS/MIS tunnel structures.

A large number of studies were undertaken in the academia following the above three pioneering investigations into the basics of the MOS/MIS tunnel structures/systems; finally, in 1996, Momose and co-workers from the Toshiba Corporation demonstrated that the leakage current through the ultra-thin tunneling gate stacks did not necessarily disable reasonable operation of nano-CMOS transistors, employing leaky SiO_2 gate dielectrics [9, 10].

1.2 The Need for High Permittivity Gate Stacks

The gate stack capacitance influences all the important performance parameters of the CMOSFET, as will be outlined in Sect. 1.3. To increase the gate stack capacitance density (to enhance the CMOSFET performance), $C_{di} = \varepsilon_{di}/t_{di}$ (ε_{di} is gate stack permittivity and t_{di} is gate stack thickness) and to continue miniaturization in accordance with Moore's law, two options are available: (a) Reduce the gate stack thickness; and/or (b) enhance the gate stack permittivity. The SiO_2 gate dielectric is universally considered a unique gift of the nature to the CMOSFET technology, being perfect or near-perfect in all the requirements for a gate dielectric, save its very low permittivity, as indicated below:

1. Dry thermal SiO_2 is a near-perfect dielectric, practically with no bulk charges and no space-charge capacitance.
2. Si–SiO_2 interface is a true and marvelous gift of the nature with an interface state density of the order of 10^{10} cm^{-2} V^{-1}.
3. When the gate electrode is poly-silicon, the Si/SiO_2/poly-Si (chemically) symmetric structure is practically immune to any thermal and chemical stability problems.
4. The dry thermal SiO_2 owes its unmatched physical, chemical, and most importantly, electronic property to its primarily covalent character, i.e. the majority chemical bond is covalent.
5. The covalent character of SiO_2 lends excellent matching with Si.
6. While it is not impossible, SiO_2 is difficult to crystallize; its crystallization temperature is far above any CMOSFET processing temperature including the implant activation temperature.
7. Dry thermal silicon dioxide has the highest band-gap (of about 9 eV) among the inorganic solids.
8. Dry thermal SiO_2 has the highest electrical resistivity (of the order of 10^{23} Ω cm) ever recorded (The resistivity is purely electronic; there is no ionic contribution.).
9. Dry thermal SiO_2 has perhaps the lowest dielectric constant (The electrical polarization is mainly electronic, with a very low ionic contribution.) among the inorganic oxides.

For the above obvious factors, the microelectronics/nanoelectronics industry was very reluctant to part with the SiO_2 gate dielectric, for as long as four decades,

till the thin SiO_2 leakage current density became untenable, when the SiO_2 thickness reached 1.3 nm or so. The place of the SiO_2 gate dielectric was taken over by silicon oxynitride, SiO_xN_y, in what we may call the transition period. SiO_xN_y has a higher permittivity than SiO_2, which allows a thicker SiO_xN_y layer, for the same equivalent dielectric capacitance, thereby resulting in a lower dielectric leakage current than that for SiO_2; this enabled scaling of the EOT (Equivalent Oxide Thickness) up to 1.0 nm or so. To achieve sub-nanometer EOT, a gate dielectric constant in excess of, say, 20 is required, which the current favorite Hf-based (Hf oxynitride, etc.) high-k gate dielectrics satisfy (see Chaps. 3, 4). Further scaling down of the EOT (beyond 0.7 nm or so) would perhaps require still higher-k materials such as La-based, ternary, or novel dielectric materials (see Chaps. 9, 10, 11).

1.3 Important Material Constants of the Gate Stack

Device grade high-k gate stacks consist of ultra-thin layers of different dielectric materials, separated by interfacial regions of mixed chemical composition. For the Si substrate, the two bulk dielectric layers are SiO_2/SiON and a high-k material (e.g. HfO_2, La_2O_3), while the three interfaces are Si/SiO_2, SiO_2/high-k, and high-k/metal interfaces.

The important material properties of the gate stack are decided by the main CMOSFET performance parameters. The physical properties of the gate stack influence many properties of the MOS structure, which in turn, determine the performance parameters of the CMOSFET. These inter-relationships are illustrated in Table 1.2 in a brief manner. The main CMOSFET performance parameters included in Table 1.2 are: the drain current, I_D, the trans-conductance, g_m, the channel conductance, g_D, the threshold voltage, V_T, the gate stack reliability, and the gate direct tunneling current density, J_{DT}. Most of these are directly influenced by the properties of the MOS structure, namely, gate dielectric capacitance, C_{di}, channel mobility, μ_{ch}, metal–semiconductor work function difference, ϕ_{MS}, gate stack charge density, Q_{gsc}, interface trap density, D_{it}, and bulk dielectric trap density, D_{bt}, cf. Table 1.2. Rigorous quantitative relations are not available in many cases to link the gate stack material properties to the CMOSFET performance parameters; however, it may be possible to suggest qualitative correlations in most cases. Reliability concepts for the ultra-thin high-k gate stacks are under development including new measurement approaches and analysis of the measured data (see Chap. 8); hence, the linkage between the gate stack degradation and the high-k material properties are not yet clearly understood.

The channel mobility is influenced by a host of high-k factors, including Coulomb scattering by charged defects (located at interfaces and in the bulk high-k layers), remote phonon scattering, and interface and remote interface roughness scattering (see Chap. 7). Electrically active defects at the interfaces, represented by D_{it} in Table 1.2, and inside the bulk layers, represented by D_{bt}, are crucially

Table 1.2 Important CMOSFET parameters and their dependence on the material properties of the gate stack. (Adapted from Kar and Singh [11])

CMOSFET performance parameter	Determined by MOS parameter	Influenced by the gate stack property
Drive current, $I_D = (W/L)C_{di}\mu_{ch}(V_G - V_T)V_D$	(1) Gate dielectric capacitance, $C_{di} = \varepsilon_0 k/t_{di}$	Dielectric constant, k
	(2) Gate stack charge density, Q_{gsc}	
	(3)Work-function difference, ϕ_{MS}	Dielectric thickness, t_{di}
	(4) Channel mobility, μ_{ch}	
Trans-conductance, $g_m = (W/L)C_{di}\mu_{ch}V_D$		Electro-negativity difference, ΔEN
		Density of oxygen vacancies, D_{Vo}
Channel conductance, $g_D = (W/L)C_{di}\mu_{ch}(V_G - V_T)$		Cation/anion coordination number, CN_c/CN_a
Threshold voltage, V_T	C_{di}	Ionicity, I
	Semiconductor doping density	Mismatch at the interfaces
	Work-function difference, ϕ_{MS}	Interface roughness
	Gate stack charge density, Q_{gsc}	Metal work function
	Interface trap density D_{it}	Metal/high-k Schottky barrier height
	Bulk dielectric trap density D_{bt}	Density of diffused impurities, e.g. Hf, Si
	Interface dipole	Electron/hole effective mass, m_e^*/m_h^*
Reliability (lifetime)	Pre-existing defect density	High-k band gap, E_G
	Degradation-induced defect density	High-k electron affinity, χ_{di}
Direct tunneling current density, J_{DT}	Conductance/valence band offset, $\phi_{b,c}/\phi_{b,v}$	
	Electron/hole effective tunneling mass, m_e^*/m_h^*	

W/L are channel width/length, V_G/V_D are gate/drain voltages, ε_0 is free space permittivity, and k is dielectric constant

important elements, as these influence almost all the CMOSFET parameters—channel mobility, μ_{ch}, threshold voltage, V_T, and gate stack reliability, cf. Table 1.2. The nature and characteristics of the high-k interface and bulk traps, and the mechanism of high-k gate stack degradation, and the mechanism of new defect creation are still to be well understood (see Chaps. 2, 8). However, a qualitative analysis of the basic factors behind the generation of defects is possible. The intrinsic defects in the high-k gate stack are likely to be bonding defects (dangling bonds, weak bonds, bond length/angle variation) associated with point defects (vacancies, interstitials); in other words, these intrinsic defects are influenced mainly by the arrangement and the network of atoms in the bulk layers and the mismatch at the interfaces. The atomic arrangement, the packing of atoms, the nature of chemical bonding, and the interface mismatch may depend upon the electro-negativity difference, the cation and anion coordination numbers, the ionicity of the mixed bond (In the high-k oxides, the chemical bond is mainly ionic and partly covalent.), and the cation–anion distance (ionic radii). Experimental evidence suggests oxygen vacancy to be a dominant intrinsic defect throughout the high-k gate stack; also, there is experimental evidence for interface dipole. Chemical interactions at the interfaces of the high-k gate stack and inter-layer diffusion during high-temperature post-deposition processing appear to generate extrinsic defects in the form of chemical impurities, e.g. Hf in the intermediate layer of SiO_2 and Si in the bulk HfO_2 layer in Hf-based gate stacks.

In a device-grade high-k gate stack, both the band offsets have to be high enough (say, $\phi_{b,c}/\phi_{b,v} \approx 2$ eV, after image force lowering), to suppress the thermionic emission over the barrier. Across such gate stacks, the dominant carrier transport mechanism is likely to be direct tunneling, as in SiO_2, for low and moderate voltage operation. In that case, the gate dielectric tunneling current, will have the following dependence on the high-k material constants [3, 5], cf. (1.1) and Table 1.2:

$$J_{DT} = C_1 \left[\exp C_2 (\phi_b m_t)^{\frac{1}{2}} t_{di} \right], \tag{1.2}$$

where C_1 and C_2 are constants, ϕ_b is the conduction or valence band offset, and m_t is the effective tunneling mass. A very important point to note is that the effective tunneling mass is as important as the band offset in determining the direct tunneling current. (A barrier height of 1.5 eV and a tunneling mass of 1.0 m is equivalent to a barrier height of 3 eV and a tunneling mass of 0.5 m.)

The interface trap density (particularly, the bonding defects) D_{it} may depend significantly on the change in the atomic arrangement at the interface, which can be represented by the difference in the coordination number, CN, the difference in the cation–anion distance, Δd_{c-a}, and the electro negativity difference ΔEN, cf. Table 1.2. The bulk trap density, D_{bt}, may be influenced by ΔEN, the ionicity of the mixed bond, I, and the coordination number CN, cf. Table 1.2. It has been established that the processes by which the gate stack is formed, including the pre-deposition interface preparation/annealing, bear upon the interface roughness, cf. Table 1.2.

Table 1.3 Important material constants of gate stack. (Adapted from Kar and Singh [11])

Material constant	CMOSFET parameters that the material constant is likely to influence
Clock-frequency dielectric constant, k ⇑	Gate dielectric capacitance, C_{di} ⇑:⇒ Drain current ⇑; Threshold voltage, V_T ⇓; Transconductance ⇑; Channel conductance ⇑
Dielectric band-gap, E_G ⇓ Dielectric electron affinity, χ ⇑ Effective tunneling mass, m_t ⇓	Barrier height ⇓:⇒ Dielectric leakage current ⇑⇒ EOT ⇑
Cation coordination number, CN ⇑ Ionicity of the mixed bond, I ⇑ Inter-ionic distance mismatch Δd_{c-a} ⇑ Electro negativity difference ΔEN ⇑	Interface trap density, D_{it} ⇑; Bulk dielectric trap density, D_{bt} ⇑; Oxygen vacancy density, D_{Vo} ⇑:⇒ Channel mobility μ_{ch} ⇓; Threshold voltage V_T ⇑; Gate stack reliability ⇓; Drain current I_D ⇓; Transconductance g_m ⇓; Channel conductance g_D ⇓

Table 1.3 lists the important gate dielectric constants, identified on the basis of the above analysis, and their influence on the CMOSFET performance parameters. Also indicated in this table is how an increase/decrease in the dielectric material constant affects the performance parameters.

1.4 Experimental Values of High-k Material Constants

Appendix V contains the experimental values of the material constants of a number of high-k gate dielectrics, SiO_2, and Si. In most cases, for each value, at least five reliable references were consulted. However, for some of the material constants, e.g. tunneling/effective mass, relaxation frequency, the data available are meager, and not five references could be found in a number of cases. Few salient features of the data in Appendix V may be noted:

1. Some of the high-k oxides can solidify in a number of crystal structures, e.g. HfO_2 can crystallize in monoclinic, tetragonal, and cubic structures.
2. In the case of the high-k oxides, with the odd cation coordination number, CN_c, of 7, for half the anions, the coordination number is 3, while it is 4 for the rest [12]. This leaves many vacant sites for the self-diffusion of the anions, as has been found in the cases of HfO_2 and ZrO_2 [13, 14].
3. Many of the high-k oxides have unequal cation–anion distances, d_{c-a}. In ZrO_2 and HfO_2, all the seven cation–anion distances are unequal [15], cf. Appendix V. In such cases, the concept of the coordination number, i.e. the number of nearest neighbors, itself is in question, when one d_{c-a} is 0.197 nm, while another is 0.260 nm, as in the case of Ta_2O_5 [16], cf. Appendix V–B.
4. There is considerable uncertainty in the values of the constants, indicated by the broad range in the values. Large variations occur in the values of the band-gap, the electron affinity, the effective mass, and the dielectric constant. The origins

of this uncertainty could partly be: variations in crystal structure, coordination number, phase (amorphous or poly-crystalline), anisotropy, processing conditions, and measurement frequency, across the data samples.

1.5 Correlation Between the High-k Material Constants

Inter-linkages may exist between the important material constants of the high-k oxides, identified in Table 1.3, as would be apparent from the data of Appendix V–A and V–B. As an example, the correlation, between the clock-frequency dielectric constant, k_{c-f}, and the other material constants, is illustrated in Table 1.4. The contribution of the electronic polarization to the dielectric constant, k_{el} should be n^2, where n is the refractive index, measured at optical frequencies. (In Si and diamond, $k_{static} = n^2$, cf. Appendix V, which confirms that there is no polarization in these solids other than electronic.) The electronic polarization and k_{el} may be expected to increase with the atomic number, i.e. with the number of electrons. There may be a correlation between the band gap and k_{el}, as the Moss rule $\{E_G(k_{el})^2 \approx 77\}$ suggests [17], although the quantitative relation of this empirical rule does not appear to apply universally, cf. Appendix V. Qualitatively, one could argue that large cations lead to large inter-ionic distances, to weaker bonds, and hence, to smaller band-gaps. Large band gaps are compatible with small electron affinities, as is larger ionicity of the mixed bond with a larger cation coordination number. It can be expected that larger ionicity and larger cation/anion coordination number will lead to higher ionic polarization and k_{ion}, cf. Appendix V.

Unfortunately, the nature of the correlation between the important material constants of the high-k oxides is such that, when some constants become more favorable for application as gate dielectrics, then others become less favorable. Desirable are high dielectric constant, high band-gap, low electron affinity, large effective mass, small electro-negativity difference, low ionicity (therefore,

Table 1.4 Correlation between important material constants of the high-k oxides (After Kar and Singh [11])

Material constant	Interlinked with	Nature of linkage
Clock-frequency dielectric constant, $k_{c\text{-}f} = k_{el} + k_{ion}$	Refractive index, n	$k_{el} = n^2$
	Bandgap, E_G	Atomic number of cation $\uparrow \Rightarrow k_{el} \uparrow \Rightarrow d_{c-a} \uparrow \Rightarrow E_G \downarrow$ Moss rule: $(k_{el})^2 E_G = 77$
	Electron affinity, χ_{di}	$E_G \uparrow \Rightarrow \chi_{di} \downarrow$
	Ionicity, I	$I \uparrow \Rightarrow CN_c/CN_a \uparrow$
	Cation/Anion coordination number, CN_c/CN_a	$CN_c/CN_a \uparrow$, $I \uparrow \Rightarrow k_{ion} \uparrow$

covalent bonding), and fourfold cation coordination. This will suggest that there is a set of optimal values for the important material constants of the gate dielectric.

1.6 MOSFET: Basics, Characteristics, and Characterization

A basic understanding of the Metal Oxide Semiconductor Field Effect Transistor (MOSFET)—its mechanisms, its performance, and its characteristics—is likely to enhance the readability of the book. The main objective of Chap. 2 is to provide a basis for understanding various MOSFET and related topics encountered in the chapters of the book. The aim also has been to provide the missing links in the realm of theory in order to enhance the continuity in the coverage of the subjects. The MOS structure being the control electrode, is the most important constituent of the MOSFET, and has been discussed first in Chap. 2 including its energy band diagrams, electrostatic analysis, flat-band and threshold voltages, circuit representation, and the capacitance–voltage (C–V) characteristics. The C–V characteristic permeates and appears throughout the book; hence it is useful to understand all its nuances, particularly, its unusual features in the case of non-Si channels.

In an approach which is gradual and easy to follow, the case of the single gate dielectric—non-leaky SiO_2—has been taken up first, subsequently graduating to leaky, ultrathin high-k gate stacks. As for the MOS structure, the MOSFET operation, including its channel characteristics (drain current, channel conductance, transconductance) has been analyzed first for the single SiO_2 gate dielectric. Significant progress has been made in understanding many aspects of the high-k gate stacks, particularly, its deposition including the precursors, post-deposition processing, passivation of the silicon channel surface, formation of the device quality intermediate layer, and thermal and chemical stability of the high-k gate stack. Unfortunately, the electrical characteristics of the high-k gate stack—electrically active defects and oxygen vacancies, fixed charges, interface traps, bulk traps, and interface dipoles, gate stack degradation—is currently among the less understood topics. While deriving the channel characteristics of the high-k gate stacks, the aim has been to highlight the degradation of the drain current, the channel conductance, and the transconductance by the non-ideal factors present in the high-k MOSFETS—namely the gate stack charges, the work function anomaly, and the non-saturating surface potential, cf. Chap. 2.

The energy band diagram across a high-k gate stack, the nature and origin of defects, traps, and charges in a high-k gate stack, the electrostatic (charges and corresponding potentials, trap capacitance and conductance) analysis, and circuit representation have been presented in Chap. 2; these subjects are not only complicated to deal with but also are hampered by our incomplete knowledge of the location, magnitude, and physical properties of the large variety of the high-k defects, traps, and charges. To analyze the contribution of the electrically active

defects and traps to the gate stack charges, potentials, and capacitance and conductance, one needs to know the Fermi occupancy of these defects, which required a treatment of the topics of electron wave function penetration into the gate stack, quantum–mechanical tunneling, and trap time constant, cf. Chap. 2. The related topic of carrier confinement in and quantization of the strong inversion and the accumulation layers have also been discussed.

One topic still under development and evolution is the flat-band voltage anomaly and MOSFET threshold voltage tuning. This topic has been analyzed in Chap. 2 including the flat-band voltage versus the EOT characteristic. As mentioned already, the C–V and the G–V (conductance–voltage) characteristics of the high-k gate stacks appear in many chapters and their features are debated, particularly, in the case of the MOSFETs with non-silicon channels. In the light of this, the basics of the impedance characteristics of high-k gate stacks have been presented in Chap. 2, and the unusual features of the G–V characteristics and the frequency dispersion of the accumulation and the strong inversion capacitance have been scrutinized.

One of the main factors behind the unmatched success of the MOSFET technology and the amazing quality of the Si/SiO_2 interface achieved was the development of the small signal MOS admittance techniques which were employed for the extraction of MOS and interface trap parameters. Chapter 2 concludes with a discussion of the most effective techniques available for the extraction of device and trap parameters of the high-k gate stacks; these include the Terman technique, the low/high frequency capacitance technique, the conductance technique, and the charge pumping technique. High-k MOS parameter extraction is many times more complicated due to the numerous interfaces, multiple dielectric layers, many different types of traps, very high trap density, and a high leakage current. Also presented in Chap. 2 are techniques for accurate determination of the gate stack capacitance and the surface potential.

1.7 Hafnium-Based Gate Dielectric Materials

Matsushita and Intel started mass producing 45 nm chips in late 2007, and AMD started production of 45 nm chips in late 2008, while IBM, Infineon, Samsung, and Chartered Semiconductor soon thereafter completed a common 45 nm process platform. By the end of 2008, SMIC was the first China-based semiconductor company to move to 45 nm, having licensed the bulk 45 nm process from IBM. Chipmakers initially voiced concerns about introducing new high-k materials into the gate stack, for the purpose of reducing the gate leakage current density. However, both IBM and Intel announced that they had high-k dielectric with metal gate solutions, which Intel considered to be a fundamental change in transistor design. In the case of the Intel 45 nm technology node (refers to the minimum feature length—generally the average half pitch of a memory cell), 1 nm was the equivalent oxide thickness (EOT) with a 0.7 nm transition layer, the gate length

was 35 nm, the gate-last process was used with $Si_{0.7}Ge_{0.3}$ stressors; 1.36 and 1.07 mA/μm were the nFET and pFET drive currents, respectively. The world's first high-k material in commercial production was a hafnium-based material—some form of HfSiON. The 32 nm technology was introduced in production by Intel and AMD in early 2010; while the 22 nm technology node has been introduced in 2012 in which the MOSFET gate length may be around 25 nm and the value of EOT may be around 0.5 nm.

The main attraction of hafnia lies in its combination of a high dielectric constant with high conduction and valence band offsets. The serious weaknesses of hafnia include its significant diffusion constant for oxygen as well as for dopant/impurity atoms, moderate crystallization temperature, and silicide formation. Oxygen diffusion through hafnia to the silicon surface leads to the thickening of the intermediate silica film with attendant increase in the EOT. Diffusion of dopant or impurity atoms through hafnia to the Si surface and the channel may destabilize the threshold voltage and/or create interface states. Poly-crystallization of hafnia during the dopant activation annealing will increase the gate leakage current. Silicide formation will also increase the gate leakage current. Nitridation of HfO_2 resulting in the formation of HfON can prevent oxygen, dopant, and impurity diffusion; however, the band-gap is significantly reduced. Solid solution of hafnia with silica or alumina, resulting in the formation of Hf silicate or Hf aluminate is observed to suppress crystallization, as Hf silicate or aluminate has a higher crystallization temperature. Also, silicide formation is suppressed. However, phase separation in HfSiO and HfAlO alloys has been observed (with the appearance of HfO_2 crystallites) to take place at high processing temperatures which can be prevented by the incorporation of nitrogen into HfSiO or HfAlO. N incorporation is also found to enhance the dielectric constant and effectively block the diffusion of oxygen, dopant, and impurity atoms.

The above issues have been examined in Chap. 3 by Nishiyama. Nishiyama begins with a discussion of the properties desired in a high-k gate dielectric followed by an outline of the important physical properties of HfO_2 including its permittivity and polarizability, its band-gap and conduction and valence band offsets, and its thermal stability during the high temperature processing steps. Next, nitridation of HfO_2 is discussed and the physical properties of HfON are outlined. Nishiyama then takes up the formation and processing of the pseudo-alloys of HfO_2 with SiO_2 and Al_2O_3 and the physical properties of these solid solutions, followed by the incorporation of N into these pseudo-alloys to prevent phase separation in the latter. The physical properties of Hf-silicates and Hf-aluminates containing N have been described in Chap. 3. Finally, Nishiyama describes how the dielectric constant of HfO_2 could be enhanced by transforming the monoclinic phase into the tetragonal or the cubic phase and how doping by rare earth elements (such as Y or La or Gd or Er or Dy) can promote this transformation by stabilizing the tetragonal or the cubic phases which normally obtain at high temperatures. The transformation of the monoclinic to the tetragonal/cubic phase can be promoted by a gate electrode cap also (see Chap. 3).

The microscopic picture of electrical polarization and the dielectric constant is important to the understanding of many topics presented in this book. In the microscopic view, electrical polarization arises due to charge asymmetry, i.e. the center of gravity of the positive charges differs from that of the negative charges, when an external electric field is present. The origin of electrical charges and dipoles can be electrons, ions, molecular dipoles, defect complexes, and/or space-charge. In the case of electronic polarization, the electron orbit is asymmetric with respect to the nuclei. In the case of ionic polarization, the cations are relatively displaced from the anions. In the case of dipolar or orientational polarization, the built-in dipoles of some molecules change their random orientation to align with the external electrical field. Charged defects and interfacial (space-charge) dipoles, i.e. defect and space charge polarization, may make additional contributions to the electrical permittivity.

The permittivity and the dielectric constant are very strong functions of the ac frequency, as is indicated in Chap. 3. An important parameter in connection with the electrical polarization is the dielectric relaxation frequency, which depends among other factors on the mass of the charge entity. Electrons have much smaller mass, hence, electronic polarization can follow very high frequencies; the relaxation frequency in this case is around the visible range, i.e. around 10^{15}–10^{16} Hz. The mass increases going from ions to molecular dipoles to defect complexes and to interfaces; hence the corresponding relaxation frequencies are: for ionic polarization around the infrared range—10^{11}–10^{13} Hz, for dipole polarization in the range of 10^{6}–10^{8} Hz, for defect polarization in the range of 10^{3}–10^{5} Hz, and for interfacial polarization in the range of 10^{-2}–10^{3} Hz. Generally, only a few of the possible polarizations will be present in a given material. In a material in which several polarizations are present, the permittivity will have its highest value at static, i.e. very low frequency (the static permittivity). At a given ac frequency, only those polarizations will contribute to the total dielectric constant, whose relaxation frequencies are higher than the ac frequency. Hence, the permittivity is zero for all materials above the visible range. At the operating frequency (i.e. some GHz), only the electronic and the ionic polarization are relevant for the MOSFETs. The electronic polarization, which generally increases with the atomic number, is small even for the heavy high-k oxides. Consequently, ionic polarization is the main source of permittivity for the high-k gate stacks. A related topic of importance is the dielectric loss. As is indicated in Chap. 3, the permittivity or the dielectric constant has a real and an imaginary component, representing the capacitive and the resistive or the loss parts. The imaginary component of the permittivity goes through a peak around each relaxation frequency; at other frequencies, its value is zero, cf. Chap. 3.

Unfortunately, there exist several misconceptions of the dielectric constant, which is the core topic of this book. Many confuse the dielectric constant, which is unit-less, with the permittivity (unit—F cm^{-1}), and use the phrase "relative dielectric constant"; dielectric constant is itself relative permittivity; hence the

word "relative" in the phrase "relative dielectric constant" is redundant. Sometimes a reference is made to the dielectric constant or the permittivity at infinity, perhaps, without realizing that both of these are zero above the optical frequencies. A third issue of concern is that what is relevant for the MOSFETs is the value of the dielectric constant at GHz frequencies. The experimental value of the dielectric constant is seldom extracted at GHz frequencies; almost always it is evaluated from measurements at or below 1 MHz, at which frequency several polarizations, e.g. dipolar and defect polarizations, may in some cases, inflate the value of the dielectric constant much above its value at the GHz frequencies.

As has been done in Chap. 3, often the band-gap of the high-k oxide E_G is linked to its dielectric constant in an inverse relationship, i.e. a higher value for k is assumed to imply a lower value for E_G. What, if any, could be the physical basis for this linkage? There exists till date no rigorous relation between E_G and k. The empirical relation—Moss rule—between E_G and k_{el} has been outlined in Sect. 1.5. However, k_{el} is a small component of the dielectric constant in the high-k oxides.

Nishiyama presents very interesting experimental and calculated data in Chap. 3 on the variation of the value of k with the nitrogen content in HfON; this variation is not monotonic, but undergoes maxima and minimum. Depending on N content, HFON can form several compounds: HfO_2, $Hf_7O_{11}N_2$, $Hf_7O_8N_4$, and Hf_2ON_2. The features (maxima, minimum) in the k versus the N content occur at the N concentration at which there is compound formation. These features could be explained by invoking the Clausius-Mossotti relation; that the maxima occur when the molar volume shrinks and the minimum occurs when the molar volume expands due to a structural change.

The issue of phase separation in the alloys of the high-k oxides has been examined by Nishiyama in Chap. 3. Phase stability is not a problem in the case of semiconductor solid solutions which are formed by mixing or alloying solids, which are eminently miscible and often completely soluble in each other, having the same crystal structures and nearly the same lattice constants, in addition to other identical or nearly same properties. In great contrast, in the case of high-k ternary/quaternary alloys, not only are the component structures dissimilar, the inter-atomic distances also differ significantly; consequently, the result is a thermodynamically unstable system with a high internal energy. The system is unstable with high internal energy, because it is amorphous and because the matrix contains atoms with different electro-negativity, different valence, and different coordination number. Such a system will look for opportunities to lower its internal energy, which the high temperature processing provides, by providing the thermal energy for atomic diffusion and nucleation and growth of new phases. Phase separation leads to several negative developments, which include enhanced diffusion, increase in leakage current, and mobility degradation. Chapter 3 presents interesting experimental data to show how incorporation of N into Hf silicates or aluminates suppresses phase separation, perhaps by blocking atomic diffusion (of Hf and/or Si), which is a precursor to any phase change.

1.8 Hafnium-Based Gate Stack Processing

Several issues have to be addressed while processing the high-k gate stack. Processing includes: (1) deposition or growth of the layers comprising the gate stack, i.e. the interfacial layer (IL), the high-k layer, the gate electrode, and the capping layer, if any, (2) post-deposition annealing (PDA), (3) post-metallization annealing (PMA), and implant activation annealing. Growth is an intrinsic process in which the substrate takes part in the chemical reaction and therefore is itself consumed, whereas, deposition is an extrinsic process, in which all the constituents of the layer are transported from external sources. The issues which concern the processing of the Hf-based gate stack include its thermodynamic stability and its ionic bonding. Thermodynamic stability reflects resistance to change at high temperatures (typically 1,000 °C in chip manufacturing); the change can involve phase transformation such as crystallization or phase separation, and/or chemical reaction between adjoining layers including inter-diffusion across them. At low temperatures (i.e. at typical deposition or growth temperatures), the high-k layers are amorphous to begin with. Any amorphous layer will tend to transform to a crystalline phase above a certain temperature which is a characteristic property of the material (crystallization temperature). After the transformation from the amorphous phase, the high-k dielectric is a polycrystalline material with grains and grain boundaries. The grain boundaries are highly undesirable as these are defect-rich regions and promote a high gate leakage current and diffusion of contaminants along the grain boundaries.

As discussed already, the high permittivity of the high-k materials originates from the contribution from the ionic polarization; hence, all the promising high-k materials are highly ionic. Almost all the deficiencies of and the problems related to the high-k gate dielectrics stem from their high ionicity if not also from their thermodynamic instability. A common deficiency of the ionic oxides is the predominance of oxygen vacancies, which in turn create several reliability problems like a very high trap density, flat-band voltage roll-off, and dielectric degradation. The processing of the Hf-based gate stack has to be optimized to: (1) raise the crystallization temperature, (2) limit the growth of the interfacial layers at both the interfaces of the high-k layer (i.e. between the high-k layer and the Si substrate and the high-k layer and the gate electrode), (3) reduce the diffusion of contaminants across the interfaces and through the bulk high-k layer, (4) reduce adverse interfacial chemical reactions, (5) promote EOT scaling, (6) enhance the permittivity of the high-k layer by doping and/or structural transformation, and (7) control the threshold (or the flat-band) voltage.

Doping of the high-k layer to enhance its dielectric constant has been discussed by Toriumi and Kita in Chap. 10. Flat-band voltage roll-off has been analyzed by Toriumi and Nabatame in Chap. 6. Metal work function control has been presented in detail in Chap. 5. Chapter 4 begins with a discussion of the important techniques for depositing the Hf-based dielectric layer, namely the Metal Organic Chemical Vapor Deposition (MOCVD), Atomic Layer Deposition (ALD), and Sputtering,

and their relative merits and demerits. The high-k deposition techniques currently in use can be classified into two broad categories: Chemical Vapor Deposition (CVD) which includes MOCVD and ALD, and Physical Vapor Deposition (PVD) which includes Sputtering. In both the CVD and the PVD processes, the constituents of the high-k layer are transported on to the heated substrate by the volatile precursors (gas phase transfer), and the film is formed by chemisorption or physisorption on the substrate surface.

The precursors in a CVD process are metal-organics and contain elements (e.g. C, H, Cl) which are not constituents of the high-k film; the CVD precursor therefore needs to decompose before the film formation, thereby creating by-products which can contaminate the high-k layer. In contrast, the precursors in a PVD process contain only the needed chemical components of the film to be deposited. Niwa outlines in Chap. 4 the relative advantages of MOCVD, ALD, and Sputtering. MOCVD handles a batch of wafers with the obvious advantage of a high throughput. ALD handles only a single wafer at a time; alternate monolayers of the high-k components are formed by cycling the complimentary precursors; and has the advantage of self-limiting monolayer growth. Sputtering involves transporting the film components on to the substrate surface by ion bombardment of a target; it does not suffer from by-products as in the MOCVD process but is affected by ion-induced surface damage and target impurity. PVD (sputtering) also enjoys the merits of low cost, good adhesion of the film to the substrate surface, and the ability to deposit high melting point materials. A host of precursors is available for the Hf-based dielectrics; Niwa analyzes in Chap. 4 their relative merits on the basis of ease of handling, by-products and contamination, and cost. Niwa presents SIMS data to show that a post-deposition ozone treatment is effective in reducing C and H contamination in the HfSiO film.

As outlined by Nishiyama in Chap. 3, also according to Niwa, nitrogen incorporation into the Hf-based gate stack appears to be a favorite manufacturing approach for resolving many of the problems which afflict it, and HfSiON with/without a nitride cap layer is a very popular high-k gate stack. It appears that N is useful both in the bulk of the gate stack and also in an interfacial cap layer. As already mentioned, nitrogen incorporation in the bulk HfSiON raises the crystallization temperature, reduces diffusion by decreasing the diffusion coefficient, by preventing crystallization, and the resultant formation of the grain boundaries, and also enhances the permittivity; however, the band-gap is reduced. Introduction of N into the top interface capping layer reduces inter-diffusion (e.g. of B from the poly-Si gate through the gate stack to the Si substrate) and interfacial chemical reaction (e.g. Hf reacting with Si in the poly-Si gate). The Si component in the HfSiON layer increases the band-gap, thereby compensating the band-gap reduction by the N component; the Si content also enhances the crystallization temperature, but reduces the dielectric constant. It is believed that the Si–N bond in the HfSiON material has a benign influence on thermodynamic stability. It is not clear what effect N has on the oxygen vacancies and the trap density. Experimental data on electron and hole trapping, showing hysteresis in the drain current versus gate voltage curves, have been presented in Chap. 4. The trapping phenomenon

has serious implications for gate stack instability (NBTI and PBTI) and reliability; this topic has been analyzed in detail in Chap. 8. Experimental data in Chap. 4 suggests that nitrogen close to the channel may degrade the mobility. According to Niwa, an N content of 3–6 % may be the optimal amount. A typical process for the formation of the HfSiON layer is plasma nitridation of a HfSiO film or its annealing in the NH_3 ambient. The doping of the Hf-based layer to improve its important properties (such as dielectric constant, thermodynamic stability, and electrical characteristics) has been outlined by Niwa in Chap. 4; this topic has been discussed in detail in Chap. 10.

In the processing of the high-k gate stack, a crucial step is the formation of the interfacial layer (IL) between the semiconductor substrate and the high-k bulk layer. One important function of the IL is to passivate the Si substrate; it is also a crucial factor in determining the gate stack reliability and the gate leakage current density. The interfacial layer is typically a SiO_2 layer; it could also be a SiON layer. The silica IL is grown thermally either by rapid thermal oxidation (RTO) in dry O_2, or by oxidation in in situ steam generation, or in ozone. Naturally, the IL formation is the first step in the gate stack processing. Unfortunately, as Niwa discusses in Chap. 4, subsequent processing steps for the high-k bulk layer, the capping layer, and the gate electrode modify the properties of the IL, in particular, its thickness and the resultant increase in the EOT of the gate stack, which nullifies the very advantage of employing a high permittivity gate stack. Diffusion into the interfacial layer and the Si/IL interface is the genesis of the IL growth during subsequent high temperature processing. Diffusion of oxygen into the IL and out-diffusion of Si from the substrate leads to a further growth of the SiO_2 layer. Niwa presents data in Chap. 4 to show that Hf can also diffuse from the Hf-based bulk dielectric layer into the IL leading to a growth at the IL/high-k interface and also a degradation of the gate stack such as interface roughness which has adverse effect on the gate stack reliability. Incorporation of N into the IL and/or into the Hf-based dielectric layer has been found to reduce this type of inter-diffusion, cf. Chap. 4. Inter-diffusion also affects the profile of the permittivity across the affected region. In principle, the dielectric constant reflects the chemical composition as well as the structural (atomic) arrangement in the lattice; inter-diffusion can lead to a significant change in both of these. Experimental data are presented in Chap. 4 indicating layers with transitional dielectric constant between the IL and the high-k layer, because of inter-diffusion.

Whereas the IL between the high-k layer and the Si substrate is a standard and necessary feature of the current high-k gate stack, an interfacial layer may form at the other (top) interface of the high-k layer, namely between the gate electrode and the high-k bulk. A capping layer may be intentionally introduced at the top interface. A common purpose of a capping layer (e.g. AlO for PMOS or LaO for NMOS) is the setting up of an interface dipole to adjust the threshold voltage; this topic is discussed in detail in Chap. 6. Another common purpose of a capping layer (e.g. SiN) could be to reduce inter-diffusion between the gate electrode layer and the high-k bulk. An unintentional layer may form at the top interface, for example, due to the formation of oxygen vacancies in the high-k bulk and the resultant

diffusion of oxygen to the gate electrode and oxidation of its surface. Niwa presents experimental data in Chap. 4 to demonstrate the efficacy of a thin SiN capping layer in reducing the degradation of the top (i.e. high-k/gate-electrode) interface between Ni-rich FUSI (Fully Silicided) gate electrode and HfO_2. It is believed that the interfacial chemical reaction, which takes place during a high temperature processing, generates oxygen vacancies in the bulk high-k layer, which in turn modulates the effective gate electrode work function.

Niwa presents in Chap. 4 a rather uncommon fabrication approach for processing the Hf-based layer and controlling the thickness of the IL and the resultant EOT of the gate stack; this involves deposition of a metallic Hf layer on the Si substrate and its subsequent oxidation. The metallic Hf layer could be deposited by DC sputtering and then could be oxidized by Rapid Thermal Oxidation (RTO) in an Ar/O_2 plasma during reactive sputtering of the HfO_2 film or by RPO (Remote Plasma Oxidation). The metallic Hf film reduces the growth of the IL during the high-k deposition. Niwa concludes RPO to be the best tool for controlling the IL and the EOT, and the experimental data he presents indicate RPO to result in lower CET (Capacitance Equivalent Thickness), lower gate leakage current, and lower flat-band voltage; also, the IL thickness could be effectively controlled by the RPO time.

Chapter 4 concludes with the discussion of the related topics of Fermi Level Pinning (FLP), work-function control, and gate-first versus the gate-last option. The gate-first option involves forming the gate stack first and then the source and the drain, using the gate stack as a mask for the source/drain implantation. This means that the gate stack has to be exposed to the high temperature implant activation annealing with all its negative consequences on the gate stack integrity. The gate-last approach involves forming the source/drain regions first using a sacrificial gate stack for the source/drain implantation, which then has to be etched out, and the real gate stack formed, in the last step. In the gate-first option the casualty is the gate stack integrity, whereas in the gate-last option, the casualty is the gate area. Fermi level pinning has been discussed at length and work-function control has been analyzed in detail in Chap. 5. The gate electrode materials and the gate-first and gate-last options have been analyzed in Chap. 5. As is outlined in Chap. 4, the gate-first approach is beset with the serious problems of crystallization of the high-k layer and the flat-band voltage roll-off, both of which have to be very carefully managed. Perhaps the most important factor to be managed in connection with the latter is the process-induced oxygen vacancies in the Hf-based bulk layer.

1.9 Metal Gate Electrodes

The metal gate electrode has been a very challenging problem in realizing high quality high-k gate stacks and the 45 nm and more advanced technology nodes. Poly-silicon, which has been the gate electrode material since the dawn of

integrated circuit manufacturing, had to be replaced by a metal, to eliminate two major undesirable features of the poly-silicon electrode on a high-k gate dielectric:

- Formation of a depletion layer in the poly-silicon subsurface, which enhances the EOT; and
- Pinning of the Fermi level at the high-k/poly-silicon interface, which disables adjustment of the threshold voltage of the transistor by the poly-silicon doping, perhaps due to generation of a very high density of interface states.

Schaeffer discusses these two handicaps in Chap. 5, and presents experimental results which show that just one monolayer of the high-k dielectric deposited between SiO_2 and the poly-silicon electrode makes the flat-band voltage invariant of the EOT.

Schaeffer outlines the important criteria for the selection of the metal for a gate electrode application; these criteria are the metal work function and its thermal and thermodynamic stability on the high-k layer, when subjected to high temperature processing (the most challenging being the high temperature implant activation anneal, in the case of gate-first integration). The ideal metal work function for NMOSFET is 4.1 eV (near the conduction band edge E_c of silicon); and is 5.2 eV for PMOSFET (near the valence band edge E_v of silicon). There appears to be a rough periodicity in the relation between the atomic number of the metal and its vacuum work function, with the vacuum work function increasing along a row in the periodic table from the left to the right, cf. Appx. VII. It may be useful to keep in mind that, although the vacuum work function is a basic physical property of the metal, there is no perfect method for its experimental determination, and there is always a spread in the values of the vacuum work function reported in the literature.

Even in the case of a metal electrode on a high-k gate dielectric, the Fermi level gets pinned, apparently due to the formation of an interfacial layer between the metal and the high-k layer and the attendant generation of the metal-induced gap states (MIGS) by the dangling bonds at the high-k/metal interface. Chapter 5 discusses the concept of the pinning parameter S, the value of which indicates the severity of the Fermi level pinning, and the charge neutrality level (CNL), which represents the pinning position of the Fermi level on the high-k layer surface. The issue of the Fermi level pinning and the related concepts of the charge neutrality level and the pinning parameter are also discussed in Chap. 6.

The remaining of Chap. 5 outlines two very practical and challenging issues in nano-CMOSFET manufacturing:

- Engineering of the effective metal work function by manipulation of the interface dipole (two layers of equal but opposite charges on the two sides of an interface), and/or oxygen vacancy concentration.
- Gate metal integration issue, namely gate first integration or gate last integration.

The basic idea behind the metal work function engineering (adjustment of the effective metal work function towards its desired value by processing means) is the

manipulation and control of the charge inside the high-k layer. As explained in Chap. 2, there is a nonlinear relation between the charges inside the high-k gate stack and all the various potentials across the gate stack layers, including the flat-band voltage and the threshold voltage of the transistor. Setting up an interface dipole at the SiO_2/high-k interface and controlling the magnitude and the direction of the dipole charge has been observed to be one effective method of realizing the desired threshold voltage. The dipole is set up at the SiO_2/high-k interface by diffusing certain metals to that interface from the high-k/metal interface. For NMOSFETs, examples of the doping metals are Mg, La, Gd, Dy, while for the PMOSFETs, an example of the doping metal is Al. The doping metal can be introduced at the high-k/metal interface as a capping oxide (1–10 Å thick oxide of Mg, La, Gd, Dy for NMOSFET, and Al_2O_3 for PMOSFET) or as a cap on the gate metal (e.g. TaMgC in place of TaC as gate metal for NMOSFET; MoAlN in place of MoN as gate metal for PMOSFET). Incorporation of the doping metal into the gate electrode may be the better option than the capping oxide, as the latter may enhance the value of EOT, which is not desired. In fact, it is possible to reduce the EOT, by having La, Mg, Al scavenge oxygen from the interfacial SiO_2 layer. One adverse effect of the threshold voltage control by the bottom interface (the SiO_2/high-k interface) dipole is a possible degradation of the channel mobility and the gate leakage current due to the presence of the scattering charges and the high density of traps, respectively.

In principle, an interface dipole can exist at both the high-k interfaces, one with the metal (the top interface), and the other with the intermediate SiO_2 layer (the bottom interface). Which interface dipole—the bottom or the top—dominates and controls the flat-band voltage and the threshold voltage has generated a large amount of debate and has been the topic of many investigations. Results of some of the most interesting and successful investigations have been outlined by Schaeffer in Chap. 5 and also by Toriumi in Chap. 6. Both Schaeffer and Toriumi present strong evidence for the dominance of the bottom interface dipole in controlling the flat-band voltage. Schaeffer presents capacitance–voltage (C–V) and SIMS (Secondary Ion Mass Spectroscopy) experimental data showing that, in the case of PMOSFET (NMOSFET), the presence of Al (Mg) at the top interface does not have any appreciable effect on the effective work function (EWF) of the MoN (TaC) metal electrode, but the diffusion of Al (Mg) to the bottom interface has a large effect on the EWF. Although Schaeffer (in Chap. 5) and Toriumi (in Chap. 6) present convincing proof for the bottom interface being the real location of the effect producing the flat-band voltage change, the mechanism behind the interface dipole formation has not yet been fully understood. Schaeffer presents data which support the electro-negativity difference between SiO_2 and the high-k dielectric being the source, while Toriumi presents data which support the difference in the areal density of oxygen between SiO_2 and the high-k material as being the primary factor behind the creation of the dipole layer at the SiO_2/high-k interface.

Just as ionic bonds (i.e. the majority bond being ionic) are an intrinsic feature of the high-k oxides, so also are the oxygen vacancies an inalienable component of its

characteristics. Oxygen vacancies, always present in the high-k oxides, will be discussed in Chap. 8. Theoretical calculations indicate that an oxygen vacancy in monoclinic hafnia may exist in any of the five charged states, from −2 to +2, and act as an electron, a hole trap, or as a fixed charge, depending upon the occupancy of its five states. Therefore, the flat band voltage can also be engineered by oxygen diffusion into the high-k oxide through the metal electrode (cf. Chap. 6). Annealing in oxygen removes oxygen vacancies in the high-k oxide, thereby changing the magnitude of the charge in the high-k layer, leading ultimately to V_{FB} change. The annealing temperature is critical, i.e. there is a threshold temperature, below which there is no change in V_{FB}, and at high temperatures, the change in V_{FB} reverses, perhaps due to the equilibrium vacancy concentration increasing faster than the vacancy removal (cf. Chap. 5).

Manufacturing an integrated circuit requires tens of process steps, generally at higher-than-ambient temperatures, and some of which involve high processing temperatures. These many process steps are carried out one after the other, but there is no unique sequencing of these steps. Any of the various possible options for the process sequence, or the integration scheme, always has some drawbacks. The main challenge while executing an integration scheme is to minimize the unintended changes in the composition and the atomic arrangement in the multitude of thin layers of diverse materials which constitute an integrated circuit. When integrating the gate metal, the main concerns are that the gate metal does not interact chemically with and/or diffuse into the surrounding environment, during the subsequent process steps. As illustrated in Chap. 5, there are two main options for carrying out the gate metal integration—namely the gate-first integration, and the gate-last (replacement-gate) integration. The concern for the thermodynamic stability of the gate metal promotes the choice of gate metals which resist high temperatures and atomic diffusion, such as W, Re, Ta, Mo, MoN, WN, NbN, TiAlN, TaN, TiN, TaSiN. The nitrides are very good diffusion barriers by virtue of the nitrogen occupying an interstitial lattice site.

In the gate-first integration, the gate metal is deposited first, and the source/drain implantation and the subsequent high temperature (1,000 °C or so) implant activation annealing are carried out later. As Schaeffer illustrates in Chap. 5, several alternatives exist and have been implemented to realize an effective high-k gate stack with a gate metal for the CMOSFET:

1. Dual metal gates (one metal gate for the NMOSFET, another for the PMOSFET, having different effective work functions) and dual high-k oxides (one set of high-k oxide and the doping element for the NMOSFET, another set of high-k oxide and the doping element for the PMOSFET, to allow better control of the threshold voltages) for realizing the CMOSFET.
2. Dual metal gates, dual high-k oxides, and dual channels (different semiconductor substrates, such as Si and SiGe, for threshold voltage adjustment; in addition for higher channel mobility) for realizing the CMOSFET.

3. Dual metal gates with alloying for realizing the CMOSFET. In this approach, an elemental metal is used for either the NOMSFET or the PMOSFET, and a metal alloy for the other transistor.
4. Single metal gate with dual high-k oxides for realizing the CMOSFET.

1.10 Flat-Band and Threshold Voltage Control

As has been explained already, the threshold voltage of the CMOSFET is one of the most important design and performance parameters, and the ability to control and tune it is therefore crucial. The threshold voltage was quite easy to control in the case of the single SiO_2 gate dielectric for two main reasons. The threshold voltage could be engineered accurately, simply by adjusting the doping of the poly-silicon gate electrode. Secondly, the interface between the SiO_2 gate dielectric and the poly-silicon electrode was chemically and thermodynamically very stable. The variation of the poly-silicon work-function by changing its doping from n^+ to p^+ resulted in exactly an equal amount of change in the potential across the SiO_2 gate dielectric. This can happen only when the gate dielectric or the gate stack is a perfect dielectric, devoid of any charges inside, as was true in the case of the SiO_2 gate dielectric. In principle, this can happen in the case of a gate stack with charges inside it, only if the charges are invariant of the gate electrode material and the gate electrode material in no way alters the gate stack charges.

One of the main challenges confronting the high-k gate stack technology today is the difficulty in controlling the threshold voltage and the anomalous relation between the work function of the gate metal, the gate stack EOT, and the threshold and the flat-band voltages. Experimental observations on the latter can be summarized in the following manner. The change in the threshold voltage of the high-k MOSFETs does not reflect the change in the metal work function in vacuum or its work function on the SiO_2 gate dielectric. The high-k threshold voltage (also flat-band voltage) is sensitive to both pre-metallization and post-metallization processing, i.e. does not enjoy chemical as well as thermodynamic stability. Chapter 6 is devoted to analyzing the flat-band/threshold-voltage anomaly and to examining the various hypotheses offered to explain this phenomenon. The hypotheses include the Fermi level pinning model, the metal/high-k interface dipole model, the oxygen vacancy model, and the high-k/SiO_2 interface dipole model. Toriumi examines all these hypotheses in Chap. 6 and the related experimental observations and concludes that the experiments support the bottom (high-k/SiO_2) interface dipole model to be valid in the case of the metal gate electrodes and the top (poly-Si/high-k) interface dipole model to be valid in the case of the poly-Si electrodes.

As has been outlined already, high-k gate stacks have proved to be incompatible with the poly-Si electrodes for a number of reasons. One of the reasons was the serious flat-band voltage anomaly observed in the case of high-k/poly-Si systems

(see Chap. 6). Ideally, and as observed on the single SiO_2 gate dielectric, the difference (ΔV_{FB}) between the flat-band voltages for p^+-poly-Si and n^+-poly-Si electrodes, respectively, should be close to the Si band-gap, i.e. about 1.0 V, irrespective of the EOT value. But on high-k gate stacks, this difference (ΔV_{FB}) is not only not 1.0 V, but varies with the high-k material and decreases with increasing high-k gate stack EOT. This anomaly signifies that not only the gate stack charge depends upon the high-k/poly-Si combination, but also depends upon the poly-Si doping or the Fermi level position in poly-Si.

Two models (pinning model and the oxygen vacancy model) have been suggested to explain this anomaly: (1) In the case of HfO_2/poly-Si system, the Fermi level pinning model rests on the assumption of pinning states located in the poly-Si upper band-gap due to Hf–Si bonding, whereas in the case of Al_2O_3/poly-Si system, the same rests on the assumption of pinning states located in the poly-Si lower band-gap due to Si–O–Al bonding. (2) The oxygen vacancy model invokes the reduction of the high-k layer (HfO_2) by the poly-Si (In a redox reaction, the reducer gets oxidized and the oxidizer gets reduced.), which involves transfer of oxygen atoms from the high-k layer to poly-Si, thereby creating oxygen vacancies V_0 in the high-k layer. The two electrons of the oxygen vacancy are assumed to occupy states below the conduction band of HfO_2. If the electrode is p^+-poly-Si, then it is energetically favorable for the these two electrons to transfer to poly-Si, leaving behind ionized oxygen vacancies V_0^{2+}, thereby setting up a top interface dipole layer to account for the V_{FB} anomaly.

The flat-band/threshold-voltage anomaly in the case of the metal gate electrodes is discussed in the later part of Chap. 6. Various hypotheses advanced to explain the V_{FB} anomaly in the case of the high-k/metal systems invoke the formation of a dipole either at the top (metal/high-k) or at the bottom (SiO_2/high-k) interface. Toriumi analyzes the strengths of the electro-negativity model (genesis: difference in electro-negativity of the two layers at the top interface), the oxygen vacancy model (genesis: redox reaction at the top interface), and the bottom interface dipole model (genesis: difference in areal oxygen density on the two sides of the bottom interface) in terms of support from the experimental observations. Interesting experimental data on bi-layer high-k gate stack (e.g. $Al_2O_3 + HfO_2$) suggest that the observed flat-band voltage is not determined by the top high-k/metal interface but by the bottom SiO_2/high-k interface and that a dipole is involved as the flat-band voltage is invariant of the thickness of the bottom high-k layer if that thickness is larger than a characteristic thickness, which is about 0.4 nm. Toriumi argues that the bottom interface dipole is the result of the difference in the areal oxygen density between the SiO_2 layer and the high-k layer. The magnitude and the direction of the dipole were found to vary with the high-k layer (HfO_2 and Al_2O_3 have negative part, whereas Y_2O_3 and La_2O_3 have positive part of the dipole.). Two layers forming an interface always have unequal electron energies; hence an electron transfer takes place across any interface to reach equilibrium in electron energy. However, Toriumi proposes oxygen transfer across the interface to balance the areal oxygen density difference; this transfer results in structural

relaxation. It is not clear, as the EOT is reduced, up to what value of the EOT, the interface dipole model can be expected to be valid.

1.11 Channel Mobility

The drift velocity v_d is proportional to the applied electric field $E_{applied}$; the proportionality constant is known as the electron/hole mobility μ: $v_d = \mu E_{applied}$. As they drift, electrons and holes are scattered by a variety of entities. Electrons drifting in a perfect crystal (represented by a periodic potential and a Bloch function) will encounter no scattering; crystal imperfections (lattice vibrations or phonons, alien atoms, surfaces and interfaces) and any deviation from or perturbation of a perfect periodic potential will constitute a scattering object. Any scattering phenomenon reduces the mobility. The mobility depends upon the effective mass m* and the mean free time τ between two consecutive scattering phenomena: $\mu = m^*/\tau$. The average time between two collisions τ will depend upon the density of the scattering objects. The effective mass has the unit of a mass, but has no direct physical meaning; it is a device for not having to invoke the internal forces acting on the electron/hole, while writing the force equation (Newton second law of motion). Hence, the effective mass represents the internal forces; it is expressed as: $m^* = \left[\hbar^2/\left(d^2E/dk^2\right)\right]$, and therefore, is given by the curvature of the electron/hole eigen-energy versus wave vector, E(k), diagrams. ($\hbar$ is Planck's constant/2π.)

Rigorous quantum-mechanical treatments exist for electron–phonon scattering and electron scattering by perturbations of the periodic potential. These treatments are a bit involved; however, simpler physical descriptions are possible. The E(k) diagrams are obtained from the solution of the Schroedinger wave equation, and therefore depend strongly upon the nearest neighbor distance, as the latter determines the potential energy term in this equation. Basically, lattice vibration results in the deviation of the nearest neighbor distance, hence in the E(k) diagrams as well. Similarly, any perturbation of the periodic potential will also lead to an alteration of the energy bands. Alien atom, charged center, and any other imperfection will perturb the lattice potential, and will affect the E(k) diagram and hence will reduce mobility.

Charge carriers traversing the channel are subjected to additional scattering mechanisms, which are absent in a bulk semiconductor. The MOSFET channel being at or in the immediate vicinity of the interface and the defect-rich layers of the high-k gate stack, represents a very imperfect region from the point of view of carrier mobility because of both intrinsic and extrinsic factors; resultantly, the channel mobility is only a fraction of what obtains in the bulk material. The additional mechanisms include: (i) carrier confinement scattering; (ii) interface roughness scattering; (iii) remote interface roughness scattering; (iv) remote phonon scattering; and (v) remote Coulomb scattering. One way of looking at the scattering process is to view these as causing loss of symmetry and uniformity and

introducing randomness. In the strong inversion layer, carriers are confined in a potential well in the direction perpendicular to the channel, and are scattered at the well boundary. Interface roughness affects the thickness of the inversion layer potential well and makes it non-uniform, hence also the E(k) diagrams of the energy sub-bands. Remote interface roughness, e.g. at the high-k/metal interface, makes the thickness of the gate stack random, which in turn makes the surface potential or the inversion layer band-bending, the potential well profile, and ultimately the inversion layer energy sub-bands random.

To keep the analysis manageable, it has been the practice to treat the different scattering mechanisms independent of each other. Under this assumption, the Matthiessen rule can be extended to include all types of scattering processes to yield the following relation:

$$\frac{1}{\mu_{ch}} = \frac{1}{\mu_{coul}} + \frac{1}{\mu_{phonon}} + \frac{1}{\mu_{potential-well}} + \frac{1}{\mu_{rough}} + \frac{1}{\mu_{remote-coul}} + \frac{1}{\mu_{remote-rough}} + \frac{1}{\mu_{remote-phonon}} \tag{1.3}$$

The main source of Coulomb scattering may be the ionized dopants in the channel; other sources may be the interface trap charges and the charges in the interfacial layer, if their densities are high. However, in device grade gate stacks, scattering by the latter may be a minor Coulomb scattering component. As the phonon density increases with the temperature, and as the operating temperature keeps rising with each new CMOS generation, phonon scattering occupies an important place. As the Coulomb potential decreases strongly with increasing distance, remote Coulomb scattering becomes significant only if the distance of the charge center is small as may be the case for sub-nanometer EOT; at the same time, as the Coulomb potential is inversely proportional also to the dielectric constant, it will still remain less significant than the dopant charges. These two factors (i.e. decreasing EOT and the high dielectric constant) will apply to remote roughness scattering in the same manner in the case of future CMOS generations. Decreasing EOT may not make remote roughness scattering more important in the future, as this will depend upon the ratio of the surface potential to the potential across the high-k/metal interface, which will actually increase instead of decreasing.

Experimental data indicate different scattering mechanisms to dominate in the different ranges of the perpendicular electric field, i.e. Coulomb, phonon, and roughness scattering dominate low, moderate, high electric field, respectively. Coulomb scattering is less important at high electric fields perpendicular (i.e. in the x direction) to the channel, because in that situation, the Coulomb potential of the charge centers becomes more localized, and will not reach majority of the channel electrons. In contrast, the vicinity interface roughness scattering (i.e. the semiconductor surface roughness scattering) becomes more important at high perpendicular fields, because the inversion layer becomes thinner; consequently the ratio of the roughness to the channel thickness increases. The inter-atomic

distance (or the lattice constant) is not primarily affected by the perpendicular field; consequently, phonon scattering is relatively field-invariant; hence, phonon scattering may dominate at moderate perpendicular electric field.

Different aspects of the channel mobility are discussed in Chap. 7 by Young. Chapter 7 begins with an introduction to the different electron/hole scattering mechanisms and their relative importance in the high-k gate stacks. Then different factors behind the channel mobility degradation in the case of high-k gate stacks (in particular, the Hf-based gate stacks) are outlined, including how the fast transient charge trapping, the interfacial layer quality, the interfacial layer thickness, the high-k layer thickness, and the Hf content influence the channel mobility.

Experimental data are presented in Chap. 7 which indicate the channel mobility to degrade as the interfacial layer thickness is reduced from 1.9 to 1.0 nm; Young has attributed this degradation to deterioration in the interface quality and to soft optical phonons in the high-k layer. Experiments suggested thicker high-k layers to reduce the mobility extracted by the split C–V technique. Flat-band voltage shifts indicated electron trapping to be more severe in the case of thicker high-k layers, leading to larger mobility degradation. According to the data presented in Chap. 7, higher Hf content in the HfSiO film translated to a larger electron trap density and an attendant sharp decrease in the channel mobility.

In the linear regime, fast transient charging effects can have a significant impact on the extracted mobility. In a conventional DC sweep of the I_D–V_G (drain current vs. gate voltage) measurement, the threshold voltage can shift as the measurement progresses, thereby decreasing the drive current at each bias sweep point and the value of the extracted mobility. Young outlines several fast transient charging correction techniques to correct for this effect; the correction techniques are (1) split C–V and conventional I_D–V_G correction technique, (2) direct measurement of inversion charge using charge pumping, and (3) pulsed I–V with model fitting and parameter extraction.

The later part of Chap. 7 is devoted to a variety of approaches for enhancing the effective mobility in the channel, such as surface orientation, strain-induced mobility increase, and high mobility semiconductors—Ge, GaAs, InGaAs, and other III–V compound semiconductors—as channel material. Young reports in Chap. 7 that NMOSFET on (100) Si has significantly higher electron mobility than the same on (110) Si, while the reverse is true in the case of the hole mobility in PMOSFET. This feature is taken advantage of in the hybrid orientation technology. Young presents data in Chap. 7 to show that the multi-gate transistors, such as the FinFETs enjoy significantly higher channel mobility. The employment of high mobility semiconductors as channel material is covered in detail in Chap. 12.

1.12 Reliability Issues

As outlined in Chap. 2, the variety of defects, traps, and charges in the high-k gate stack can relate to:

1. The states/traps at the channel/IL interface;
2. The defects/traps/charges inside the IL (Intermediate Layer);
3. The dipole layer or the states/traps at the IL/high-k interface;
4. The defects/traps/charges inside the high-k layer.

In operation, an MOSFET keeps continually degrading, culminating finally in an electrical breakdown, followed by a thermal (run-away) breakdown and burn-out. For an MOSFET, degradation mainly signifies a deterioration of the gate stack, which is characterized by a continual generation of defects in the gate stack and an attendant increase in the gate stack charges. The continually increasing additional gate stack charges manifest in an ever-increasing MOSFET threshold voltage and an ever-decreasing drain current, transconductance, and channel conductance, while the continually-increasing defects manifest in an ever-increasing leakage current through the gate stack. The chief objective of an MOSFET reliability study is to estimate the lifetime, which is defined as the operating period at the end of which the MOSFET will deliver at least 90 % of its rated performance; or to specify the maximum supply voltage which will yield a specified lifetime. The integrated circuits (ICs) are required to have a lifetime of 10 years. As testing an IC under the specified operating conditions for 10 years is not feasible, for examining the reliability, the devices undergo accelerated testing, which may involve subjecting the devices to higher-than-operating voltages (called stress voltage) at higher-than-operating temperatures for a practical duration of time (called stress time). The reliability of the estimated device lifetime and/or its maximum supply voltage permissible for yielding a 10 year lifetime, critically hinges upon the relation used between the accelerated device ageing (degradation) rate and the actual device ageing rate.

Bersuker points out in Chap. 8 that the reliability problem is more complicated in the high-k gate stacks than in the SiO_2 gate dielectric due to a number of reasons. One of these is the presence of a very high density of intrinsic, or pre-existing, defects in the gate stack. The d-shell bonding may lead to a large number of as-grown defects in the high-k layer, such as oxygen vacancies. Also, the intermediate layer (IL) may contain a high density of pre-existing defects such as oxygen vacancies and under-coordinated Si^{2+} and metal (e.g. Hf) ions. The author outlines how charging/discharging at these pre-existing traps may have significant bearing upon the reliability measurements and their interpretation, and may even dominate over the trapping process at the stress-generated defects; the latter is believed to make the main contribution to the device instability in the SiO_2 gate dielectric.

The reliability experiments outlined by Bersuker in Chap. 8 reveal that under a voltage stress, the threshold voltage increases not at a monotonous rate, but perhaps at three different rates; in other words, the incremental threshold voltage versus the stress time plot could be divided into three different regimes: a fast transition regime of the order some μs, an intermediate regime of the order of a second, and a slow regime of the order of tens of seconds. Bersuker argues that the threshold voltage increase in the fast transition regime is due to a reversible

process, i.e. the threshold voltage increase can be reversed by a reverse stress voltage or can relax to the original value by itself. The experimental data presented by the author suggests the origin of this reversible process to be trapping and de-trapping in the process-induced defects in the high-k (HfO_2) layer by tunneling of carriers through the gate stack, and as the process is reversible, it does not constitute any threat to the integrity of the gate stack, and therefore should play no role in the lifetime estimation. Bersuker presents a host of experimental data to demonstrate that the slow regime has its origin in the continual generation of additional states and traps in the precursor defects in the IL close to the high-k (HfO_2) interface; these he argues are the main source of degradation in the high-k gate stacks. The process responsible for the intermediate regime is also reversible and Bersuker suggests that the same type of traps in the high-k (HfO_2) layer are responsible as those that gave rise to the fast transient regime, and all effects of these traps on the reliability measurements should be eliminated while estimating the reliability of the high-k gate stacks.

Therefore the data illustrated in Chap. 8 suggest that only two of the four different defect/trap types, outlined in Chap. 2, are reflected in the incremental threshold voltage versus the stress time data, $\Delta V_T(t)$, under a constant voltage stress. Bersuker argues that $\Delta V_T(t)$ due to the bulk defects in the high-k layer should not be considered in the reliability estimation, but only that part of the $\Delta V_T(t)$ which is due to the bulk defects in the intermediate layer (IL).

Chapter 8 illustrates how trapping/de-trapping at the pre-existing defects greatly alters the nature of the reliability measurements and gives rise to some new reliability phenomena. A crucial parameter in this connection is the trap time constant or the relaxation time of the pre-existing trap and how this time compares to the other times, such as the sense measurement time and the stress-induced defect generation time. The following two situations are possible:

1. If the pre-existing trap's capture time is shorter (i.e. trapping is fast) than the sense time used to monitor the device parameter, then the fast trapping would erroneously reflect as a part of the time-zero or the intrinsic MOSFET characteristic. However, during the MOSFET operation, the frequency would be much higher (in the GHz range) than the test measurement frequency (in the kHz–MHz range), resulting in a very different time-zero MOSFET characteristic.
2. If the pre-existing trap's capture/emission time is longer than the sense time (i.e. trapping is slow) but is shorter than the stress-induced defect generation time, then the slow trapping at the pre-existing defects may dominate the time-dependent device instability, which would be erroneously read as being caused by the gate dielectric degradation due to the generation of new defects. This would lead to an incorrect interpretation of the device lifetime.

Chapter 8 analyzes where the most crucial traps, as far as gate stack degradation is concerned, are located—in the IL or in the high-k layer, and whether these traps are the as-grown ones or are the stress-induced ones. The author outlines a methodology for separating the fast transient trapping contribution from the total

device instability, so that a correction can be made for this effect. Bersuker concludes that at low voltages and moderate temperatures, defect generation in the IL is a major degradation factor, and the stress-induced leakage current reflects this phenomenon and could be an effective monitoring tool for it.

Chapter 8 highlights the following points:

1. In monoclinic hafnia, oxygen vacancies have been shown by ab initio calculations to exist in five charge states, from −2 to +2 and may function as electron/hole traps, as well as fixed charges. The IL (SiO_2) layer may be oxygen deficient at the Si/IL interface and also at the IL/Hafnia interface due the diffusion of oxygen vacancies from the hafnia layer. The IL layer may contain under-coordinated Si^{2+} and metal Hf ions.
2. At low operating voltages, the device instability in the HfO_2 gate stack is dominated by electron trapping in the shallow traps located in the high-k layer. These traps have activation energy of about 0.5 eV, a density of about 10^{14} cm^{-2}, a capture cross-section of ca. 10^{-13} cm^2, and a time constant of about 0.5 μs and this process is fully reversible in the NMOSFETs, indicating that only the as-grown defects are responsible for the fast transient charging, and that no new defects are generated. However, the instability in the PMOSFETs is not fully recoverable, indicating symptoms of some NBTI.
3. The reversible electron trapping into the precursor defects has a fast and a slow component, differing in response times by over six orders of magnitude, i.e. the fast process may have a characteristic time of μs, while the slow process may have characteristic times of s. While the fast reversible process is temperature independent, the slow process is quite temperature sensitive, suggesting a different mechanism involved in the latter trapping process. The experimental results yield a capture cross-section of 10^{-18}–10^{-19} cm^2 which the author argues is too small to have any physical meaning. Bersuker argues that the precursor defect in both the fast transient and the slow reversible charging effect is the same—namely a shallow defect about 0.5 eV below the conduction band edge, which could be a negatively charged oxygen vacancy in monoclinic HfO_2.
4. The device reliability, i.e. the device lifetime, is predicted on the basis of the so-called power law and the power law exponent (in other words the slope of the $\Delta V_T(t)$ characteristic). Therefore, correct estimation of the power law exponent is very crucial. Unfortunately, the fast transient charging effect can seriously affect the reliability of the $\Delta V_T(t)$ plot and its power exponent. Hence, it is necessary to eliminate the effect of the fast charging process in the $\Delta V_T(t)$ data.
5. Bersuker has outlined in Chap. 8 two approaches for correcting the effect of the fast charging process: the single pulse stress-sense-stress method and on-the-fly method, both of which are shown to yield nearly identical results. In contrast, the traditional DC I_D–V_G technique produces significantly different results. After corrections for the reversible fast charging effects, the long term $\Delta V_T(t)$

data indicate about the same rate of stress-induced interface state generation as in the SiO_2 gate dielectric with $\Delta V_{th}(t) = t^n$ behavior and n = 0.2.

6. In PMOSFETs, hole injection into the precursor defects in the IL generates additional defects at a faster rate than the so-called slow stress-induced defect generation. These precursor defects are likely to be Hf in the IL due to diffusion from the Hf-based bulk high-k layer. The density of these Hf atoms could be 10^{13} cm^{-2}.

1.13 Lanthanide Based High-k Gate Stack Materials

The lanthanide (also referred to as lanthanoid) series includes fifteen metallic elements from lanthanum through lutetium (atomic number: 57 to 71), which along with scandium and yttrium is popularly called the "rare earth elements", although some of these are not that rare. All except lutetium are f-block elements, many of these are trivalent, and the ionic radius, which decreases monotonically from lanthanum to lutetium, has a strong influence on their chemistry. Often the generic chemical symbol Ln is assigned to the lanthanides. Lanthanide based dielectrics are said to offer higher dielectric constant and higher crystallization temperatures than Hf based dielectrics, while offering about the same band offsets. Ternary compounds of lanthanides are more attractive than the binary compounds in terms of a high dielectric constant, a high band gap, and a higher crystallization temperature, i.e. higher amorphous phase stability with no crystallization or phase separation as happens in the case of the Hf based gate stacks. In addition, a thinning or even a complete removal of the intermediate layer (IL) by the ternary lanthanide oxide allows scaling the EOT well below 1.0 nm. A high dielectric constant cannot alone lead to an EOT lower than 1.0 nm; removal or at least a down-scaling of the lower-k IL is also necessary.

Chapter 9 begins with a review of the physical, structural, dielectric, and electrical properties of a host of important lanthanide based oxides having the binary Ln_2O_3 and ternary $LnMO_3$, $LnLnO_3$, $Ln_xM_yO_z$, $Ln_xSi_yO_z$ configurations. Among these oxides, La_2O_3, Gd_2O_3, $LaAlO_3$, $LaScO_3$, $LaLuO_3$ appear to be promising as sub-1-nm-EOT gate dielectrics on the basis of the experimental values of the dielectric constant (k = 24–32), band-gap (5.2–5.5 eV), and band offsets (ca. 2 eV), cf. Chap. 9. The lanthanide oxides, Ln_2O_3, crystallize in cubic, monoclinic, or hexagonal structure, depending upon the temperature. Unfortunately the binary oxides are sensitive to the atmosphere, in particular the water vapor; this weakness can be removed by forming ternary oxides, e.g. lanthanide silicates and aluminates, cf. Chap. 9. As Lichtenwalner outlines in Chap. 9, the ternary oxides—lanthanide scandates, hafnates, aluminates, and alloy oxides—offer in general higher amorphous phase stability and better lattice matching if epitaxial oxides are to be grown. The deposition techniques for the lanthanide based gate dielectrics include thermal and e-beam evaporation, pulsed laser

abalation and atomic layer deposition (ALD). The ALD is the preferred deposition technique for which the proper precursor is the main item of concern. The lanthanides being reactive and hygroscopic need special handling during the deposition process. Impressive EOT values (0.50–0.75 nm) have been reported for the lanthanide based gate stacks, cf. Chap. 9.

Lichtenwalner presents interesting experimental results in Chap. 9 which demonstrate how low temperature (400 °C) post-metallization RTA in nitrogen of a W/TaN/lanthania/silica/Si gate stack leads to a significant EOT reduction (from 1.57 to 0.69 nm) and yields a lanthanum silicate gate dielectric with a sub-nm EOT but a very high interfacial trap density; the latter is perhaps caused by a high density of dangling bonds due to incomplete chemical reaction (i.e. silicate formation). Further high temperature (1,000 °C) RTA in nitrogen reduces the trap density at the expense of an SiO_x layer growth with the attendant increase in the EOT; this phenomenon illustrates the current challenge in obtaining a sub-nm EOT with a low trap density. A final high temperature FGA reduces the trap density further. The PBTI and the NBTI of these gate stacks also improve after the high temperature RTA, in agreement with the reduction in the trap density. The experimental data appear to indicate the metal electrode to be the source of oxygen for the growth of the SiO_x layer during the high temperature RTA; hence elimination of excess O from the gate electrode may be a key factor in realizing Ln based gate stacks with EOT less than 1.0 nm and low interface trap density.

Chapter 9 contains a comparison of various Ln based ternary (an extra cation) oxides—lanthanide aluminates, scandates, hafnates, zirconates—on the basis of higher dielectric constant, band-gap, crystallization temperature, phase stability, cation diffusion, leakage current, channel mobility, epitaxial oxide on Si, and bias temperature instability (threshold voltage shift). Lanthanum scandate appears to have a dielectric constant higher than 30 and a conduction band offset of about 2.0 eV. LaAlON and LaSiO have the highest amorphous phase stability. Addition of La to HfSiON enhances amorphous phase stability, with an attendant decrease in the defect density and also in PBTI, perhaps because of a mixed cation coordination in LaHfSiON. The multi-component dielectrics offer (1) a higher-temperature amorphous phase-stability; (2) a potential for limiting IL SiO_2 formation; (3) an ability to control the device V_T; and (4) a means of reducing V_T shift during the PBTI stress. The later part of Chap. 9 presents an analysis of the topics of threshold voltage control and passivation of the III–V compound surfaces by lanthanide oxides.

1.14 Ternary High-k Gate Stack Materials

In device grade high-k MOSFETs, the gate stack is generally a bi-layer system, and the high-k layer is seldom a binary oxide. (Even the intermediate layer—IL—is often intentionally or unintentionally a ternary compound, e.g. SiON, Hf silicate.) The ternary high-k dielectrics can be called mixed oxides or oxide solid

solutions, or oxide alloys; often the name doped oxides is used perhaps to suggest the possibility of the physical effects far exceeding what normally would be expected from the mere concentration of the dopant—in analogy to the semiconductor properties, which are dictated by the dopants. The addition of a cation or an anion to the binary high-k material may be motivated by any or some or all of the following objectives: (a) to raise the crystallization temperature; (b) to inhibit diffusion of atoms and ions, i.e. attenuate the diffusion constant; (c) to increase the band-gap and decrease the electron affinity; (d) to bring about a phase transformation, i.e. modify the atomic arrangement, which can significantly enhance the dielectric constant; and (e) to stabilize a certain phase which is in the binary configuration meta-stable. Thus alloying or doping offers several possibilities to enhance the performance of the high-k gate stack, notwithstanding the fact that the system becomes more complex and prone to cross-contamination.

Mixing has been a common practice in the case of III–V compound semiconductor materials and devices; most often the physical properties (e.g. lattice constant, band-gap) change monotonically, i.e. almost linearly, with the chemical composition of the mixture or the mixing index, between those of the two components. Interesting is when the physical property (say, the dielectric constant) would change non-linearly with the mixing ratio, with features such as a peak or a spike. One main focus of Chap. 10 is to analyze how unusually large increases in the permittivity can be engineered by the addition of structure modifiers to the high-k material. Toriumi presents interesting experimental data (k versus composition, X-ray diffraction, and molar volume V_m versus composition) on several doped oxides (Y-doped hafnia—YDH, Si-doped hafnia—SDH) to demonstrate this mechanism. These structure modifiers change the atomic arrangement to reduce the molar volume V_m while keeping the molar polarizability α_m unaltered; consequently the ratio of α_m/V_m increases, to which, the dielectric constant k, according to the Clausius-Mossotti relation is very sensitive. Thus, Toriumi explains, it is possible to enhance the dielectric constant significantly beyond what the average composition alone would achieve, by decreasing the molar volume through a favorable rearrangement of the atom matrix.

As already analyzed in Sect. 1.7, ionic polarization is the main source of permittivity for the high-k gate stacks. Toriumi analyzes ionic polarization in Chap. 10 to suggest that low frequency phonons are responsible for the high ionic polarization of hafnia. Chapter 10 presents several examples of hafnia and lanthania based ternary materials to demonstrate how the dielectric properties can be greatly enhanced by the addition of suitable dopants. Toriumi outlines how doping of hafnia by elements such as Y (YDH) or Si (SDH) brings about a phase transformation (YDH: monoclinic to cubic, SDH: monoclinic to tetragonal) to result in a reduced molar volume with an attendant increase of the dielectric constant. Lanthanum doping is very effective in raising the crystallization temperature of HfO_2 with the result that HfLaO remains amorphous with a high dielectric constant even after high temperature processing, cf. Chap. 10. (Hf silicate and aluminate have high crystallization temperature but suffer from a greatly

reduced dielectric constant, making these unsuitable for the future high-k gate stack.) Toriumi suggests that the large difference in the ionic radius between Hf and La suppresses the long range order and thereby promotes the amorphous phase with its random close packed network. Lanthania is not a stable material in particular when exposed to air and moisture. Doping of La_2O_3 by Y is very effective in stabilizing the hexagonal phase which also is the higher-k phase of lanthania, cf. Chap. 10. Doping of La_2O_3 by Lu or Ta is very effective in raising both the crystallization temperature and the dielectric constant. The large difference between the ionic radius of La and Lu or La and Ta promotes the amorphous phase.

1.15 Crystalline Gate Oxides

As explained earlier, to reduce the interface trap density in high-k MOSFETs to the device quality level, the semiconductor surface has to be passivated by an intermediate layer (IL) such as SiO_2 or SiON. After a high temperature annealing process, this results in the growth of a silicate or oxynitride IL, which generally increases the value of the EOT. Values of EOT below 0.7 nm cannot be achieved unless this IL is done away with. One option of doing this is to grow a defect-free epitaxial gate dielectric directly upon the semiconductor surface, which is the focus of Chap. 11 by Osten. Epitaxy is the growth of a crystalline layer upon a crystalline substrate; it is called homo-epitaxy if the layer and the substrate are the same material, whereas hetero-epitaxy refers to the case when the substrate and the layer are different materials. The quality of the crystalline gate dielectric will depend upon how well the conditions of the hetero-epitaxy are satisfied. The conditions for hetero-epitaxy are that the substrate and the layer have identical crystal structure, nearly same lattice constant, identical valency, nearly same electro-negativity, and nearly same thermal expansion. These conditions basically mean that the substrate and the layer have nearly the same atomic arrangement. The amount of difference between the substrate and the layer on these counts will determine the nature and the magnitude of defects in the epitaxial layer, most importantly the bonding defects. It may be recalled from our earlier discussion that the direct semiconductor/high-k interface suffers from a very high density of electrically active defects because of a strong mismatch between the semiconductor and the high-k layer in terms of lattice constant, crystal structure, chemical bonding, electro-negativity, coordination number, etc. A good hetero-epitaxial system can be expected to have a lower mismatch with a lower number of bonding defects, while at the same time fulfilling the basic aim of eliminating the IL to achieve a lower EOT.

The semiconductors of interest—Si, GaAs, InAs, GaSb, GaN—have the cubic diamond or the zinc blende crystal structure, cf. Appendix III. However, there are no insulating oxides which solidify in these crystal structures. The closest oxide

crystal structures are those of the perovskite oxides and the lanthanide oxides. The focus in Chap. 11 is on the lanthanide oxides (LnO_x), in particular the gadolinium oxide, Gd_2O_3. The rare earth metals can have several oxidation states—2, 3, and 4; consequently the LnO's can have multiple stoichiometries and phases. The Ln_2O_3 oxides exhibit only one valence state, hence are most suitable. Gd_2O_3 has the bixbyite crystal structure. As illustrated in Chap. 11, the MBE grown Gd_2O_3 layer on (100) Si, exhibits two orthogonal (110) type domains. For a gate stack of EOT = 0.9 nm, the trap densities obtained from the conductance technique were 1.0, 2.4, and 9.0 times 10^{12} cm^{-2} V^{-1} for (100), (111), and (110) Si surface orientation, respectively. As Osten points out, oxygen partial pressure during the MBE growth of the Gd_2O_3 layer by evaporation of the Gd_2O_3 granules is a very crucial parameter in determining the electrical quality of the epitaxial layer. The optimal oxygen partial pressure was observed to be 5×10^{-7} mbar and the optimal substrate temperature was found to be 600 °C which yielded films with a dielectric constant of 20, EOT < 0.7 nm, hysteresis < 10 mV, and leakage current at $V_{FB-}1.0$ V of 5 mA cm^{-2}. Lower oxygen partial pressure leads to silicide formation while higher oxygen partial pressure results in the growth of silicon oxides. An intimate interface between Si and the crystalline oxide was observed from the TEM micrographs, cf. Chap. 11.

The growth of the epitaxial rare earth oxides with two (110) orthogonal domains on (100) Si can have negative consequences, cf. Chap. 11. Domains have grain boundaries and the mismatch at these boundaries manifest by broken bonds which generate a high density of traps and also act as a sink to getter chemical impurities. One attendant result is a large leakage current along the grain boundaries. Osten outlines in Chap. 11 how careful Si surface preparation and a 4° miscut substrate surface can produce a single domain epitaxial oxide layer eliminating the grain boundaries and achieving a reduced leakage current. Investigations outlined in Chap. 11 demonstrate the scope of interface engineering for improving the electrical properties. Control of oxygen partial pressure and temperature during the growth of the Gd_2O_3 layer can yield either an oxide-like or a silicate-like interface; the latter interface results in lower EOT and leakage current. Ge passivation of the Si surface results in much lower interface trap density and fixed charge density. The standard forming gas anneal is effective in eliminating the large hysteresis observed in the case of as-grown crystalline Gd_2O_3 gate dielectrics and in reducing the leakage current, apparently by saturating the dangling bonds. Rapid thermal annealing at temperatures higher than 800 °C degraded the Gd_2O_3 crystalline layer, cf. Chap. 11.

Many applications other than as a gate dielectric are possible for crystalline gadolinia; some of these applications have been illustrated in Chap. 11. These include a double barrier insulator/Si/insulator quantum-well structure with epitaxial gadolinia as the insulator and single crystal gadolinia layer with embedded Si nano-clusters for non-volatile flash memory application.

1.16 High Mobility Channels

Much like its natural oxide SiO_2, silicon is blessed with many unmatched assets, but for some applications, these assets are neutralized by one of its few weaknesses which include its low hole mobility and also by its modest electron mobility. The assets of silicon include the following. (1) Device grade silicon is by far the purest material obtained so far. In addition, much larger single crystals are obtained for silicon than for any other semiconductor; large crystal diameter translates into less expensive ICs. (2) The physics of silicon has been extensively studied and is much better understood than that of any other material. (3) The technology of silicon, promoted by the preceding two factors, is very advanced and is nowhere matched by that of any other semiconductor or any other material. This is the primary reason for using silicon as the substrate (base) even when employing other channel materials (Ge, SiGe, III–V compounds). (4) Silicon is the second most abundant element (The most abundant element is oxygen.) on the earth's surface and occurs in nature as silica sand. Very pure silica sand occurs in nature, which is one reason why such phenomenal purity is achieved for silicon.

Very sadly, much like the unique and near-perfect dielectric SiO_2 is undone as a gate dielectric for CMOSFETs with sub-nanometer EOT by its exceptionally low dielectric constant, Si as a channel material for the same devices is undone by its low hole and modest electron mobility. When the EOT is below 1 nm, experimental results indicate that a host of scattering entities, see Sect. 1.11, sink the hole and the electron mobility of silicon channels to such values that its drain current becomes unviable. In other words, the gain in the drain current due to the increase in the gate dielectric capacitance density C_{di} is more than neutralized by the attenuation of the effective channel mobility μ_{eff}. In order to overcome this challenge to the drain current, one option is to replace the silicon channel by a channel of much higher carrier mobility.

Appendix III presents a list of semiconductors with their important physical properties, including the lattice constant, the band gap, electron and hole mobility. Looking at the column of carrier mobility, two features stand out, namely the enormous spread in the values of the mobility, and wide disparity between the electron and the hole mobility. Some semiconductors stand out with respect to their properties of carrier mobility. Diamond and Ge, both having the diamond crystal structure like Si, are the only two semiconductors with high hole as well as high electron mobility. Diamond is a unique material: it is the hardest of materials with extremely high natural purity and extremely high melting point; it may also be one of the most perfect and covalent of crystals. Its nearly equal electron and hole mobility perhaps reflects its closeness to a perfect covalent material. Unfortunately, diamond is not yet available in a form which lends itself to use as a channel material. Till then, Ge remains our only option for realizing high mobility PMOSFETs. Some III–V compounds (InSb, InAs, GaAs) stand out on account of their extremely high electron mobility; unfortunately, the hole mobility is depressingly low in these semiconductors. As is outlined in Chap. 12, at this time,

NMOSFETs with Ge channels exhibit poor performance. Therefore, at the current time, the best option for realizing a high mobility CMOSFET is a hybrid with PMOSFET on Ge channel and NMOSFET on a III–V compound semiconductor, see Chap. 12.

As Houssa recalls in Chap. 12, the transistor was invented on Ge but was soon forgotten, because of silicon's overwhelming advantages over Ge, as outlined above, till its much higher hole mobility than that of Si generated the current interest. For decades, surface passivation has been the main challenge in realizing MOSFETs on any semiconductor other than Si, be it Ge or III–V compound semiconductors. On silicon, the dry thermal SiO_2 was always found to be orders of magnitude superior to the deposited (sputtered or CVD or PVD) SiO_2. The inherent advantages of a grown over a deposited oxide include the inward movement of the semiconductor/dielectric interface and much higher purity of the dielectric source components. Silicon oxidizes in O_2 or H_2O by the inward diffusion of the oxidant rather than the outward diffusion of Si resulting in the growth of the newest oxide monolayer at the Si/SiO_2 interface [18]. This helps in keeping the initial contamination on the Si surface away from the crucial Si/SiO_2 interface. The source components of the dielectric, i.e. Si substrate and the oxidant are much purer than the precursors for a deposited SiO_2.

In light of the experience with the passivation of Si, thermal oxidation of the Ge surface would seem to be the best option for the latter's passivation. As Houssa recounts in Chap. 12, this is what was tried for many decades with very poor results (unacceptably high trap density, etc.) for the Ge/GeOx interface quality. Fortunately, the situation has seen significant improvement as a result of the renewed interest and current research on Ge surface passivation. Houssa outlines three successful approaches for removal of the dangling bonds on the Ge surface and for obtaining an acceptable Ge/gate-dielectric interface. These are the use of a Si cap and its dry thermal oxidation, dry thermal oxidation of the Ge surface, and deposition of a rare-earth high-k dielectric, all of which yield 2–3 times higher hole mobility than the same in Si channels.

In the first approach, a few monolayers of Si are deposited on the Ge surface by hetero-epitaxy followed by thermal oxidation of a part of the Si epitaxial layer cap. Experiments revealed the quality of the Ge PMOSFET to depend on the number of remaining Si monolayers after oxidation; optimal values of the drain current and the threshold voltage were obtained for about 5 remaining monolayers of the Si cap, cf. Chap. 12. The physical reason lies in the following. For a higher number of Si monolayers, the Ge channel is less affected by the interface imperfection, but at the same time the field effect is diluted as the channel is capacitance-wise further removed from the gate electrode. As an example of the second approach, thermal oxidation of Ge in dry O_2 followed by in situ deposition of Al_2O_3 yielded high quality GeO_2 layers with sharp GeO_2/Al_2O_3 interfaces and Ge/GeO_2 interface trap densities of the order of 10^{11} cm^{-2} V^{-1} and low C–V hysteresis, see Chap. 12. Houssa describes as an example of the third approach, how deposition of rare-earth oxides like La_2O_3 has been successful in passivating the Ge surface perhaps because of the formation of La–O–Ge (germinate) bonds. While recent surface

passivation research has yielded promising PMOSFETs with Ge channels, NMOSFETs with Ge channels remain disappointing. Houssa reports that this perhaps could be related to the position of the charge neutrality level (CNL) being closer to the valence than the conduction band edge of Ge and a very high interface trap density near the latter.

As Ge NMOSFETs are not viable, at least at present, one needs to fall back upon NMOSFETs with III–V channels for realizing high mobility CMOSFETs. The III–V compound semiconductors have been beset with a very high interface trap density and the resulting pinning of its surface Fermi level. One of the lasting features of many decades of III–V compound semiconductor research has been the inability to unpin the surface Fermi level and to scan the band-gap by applying a reasonable range of applied voltage. Crucial to a successful passivation of GaAs and other III–V semiconductor surfaces may be the removal of the native oxide and surface contamination and obtaining a hydrophilic surface. (A hydrophilic surface is not wet by water at all, while a hydrophobic surface is completely wet by water. A hydrophilic surface represents the absence of any surface contamination, and any particulate matter present can be easily removed by a de-ionized water rinse.) An HF last clean renders a hydrophilic Si surface and ensures successful dry thermal oxidation to realize a near perfect Si/SiO_2 interface. Ye outlines in Chap. 12 the recent success in realizing a hydrophilic GaAs surface by wet chemical surface preparation by NH_4OH and $(NH_4)_2S$ treatments. He also describes how the ALD of the high-k layer (Al_2O_3) has a self-cleaning effect in the removal of the native oxide. Ye presents in Chap. 12 the characteristics of the depletion-mode and accumulation-mode GaAs/AlGaAs MOS transistors with $Al_2O_3/HfO_2/HfAlO$ gate dielectric and impressive drain current and transconductance.

Inspite of the significant success achieved recently in the passivation of the III–V semiconductor surfaces by surface pre-treatment and ALD high-k, the interface trap density is still high and makes scanning of the band-gap through the high density traps difficult for the of enhancement mode inversion type GaAs MOSFETs. Ye explains in Chap. 12 an approach in reducing this difficulty by decreasing the band-gap by the use of a GaAs solid-solution. InAs has a much smaller band-gap but a much higher electron mobility than GaAs, cf. Appendix III. A semiconductor binary solid solution (i.e. an alloy) generally has a band-gap varying monotonically between those of the two constituents. The mobility of the semiconductor alloy will also vary monotonically between those of the two constituents, but will have a minimum because of enhanced scattering in a solid solution. The InGaAs alloy will have a much higher electron mobility and a lower band-gap than those of GaAs, both of which are advantageous. The lower band-gap eases the attainment of inversion because a lower supply voltage is required to move the surface Fermi level through the high density interface traps. Chapter 12 contains details of In rich GaAs ($In_{0.65}Ga_{0.35}As$) NMOSFETs with drain currents of 630 $\mu A/\mu m$, transconductance of 350 mS/mm, effective electron mobility of 1,550 cm^2/V s, and a mid-gap interface trap density of 1.4×10^{12} cm^{-2} V^{-1}.

Electrical characterization of the III–V MOSFETs and MOS capacitors is more complicated because of a number of reasons. If the III–V band-gap is high, as is

the case for GaAs, then the minority carrier generation rate may be too low and the corresponding minority carrier and inversion layer response time too high for measuring the full capacitance in the weak and strong inversion regimes. In other words, the low frequency (i.e. the equilibrium) capacitance–voltage (C–V) characteristic, which is the backbone of the MOS admittance techniques for parameter extraction, cannot be obtained. The inversion layer response time is inversely proportional to the intrinsic carrier density. If one uses the MOS configuration for measuring the C–V curve, then the inversion layer can form only from the thermal minority carrier generation process and the generation rate will be low for a high band-gap semiconductor. Heyns outlines in Chap. 12 three options for resolving the inversion layer response problem. One of these involves using the MOSFET configuration with minority carrier injection into the channel from the source for building up the inversion layer quickly. The second option is measurement at elevated temperature with a greatly enhanced minority carrier generation rate. The third option is measurement under illumination, which greatly enhances the minority carrier generation rate. There is another serious problem in the electrical characterization of III–V semiconductor and also Ge MOS devices, which relate to the significant frequency dispersion of both the accumulation and the strong inversion capacitance. Normally, as in the Si MOSFETs, the accumulation or the strong inversion capacitance is frequency-invariant (i.e. for the usual ac frequency range employed in electrical characterization). This frequency dispersion is discussed in Chap. 2.

1.17 A Figure of Merit for a High-k Material as a Gate Dielectric

It perhaps is desirable to have a methodology for ranking, even if only qualitatively, the high-k materials either currently under investigation, application, or exploration for the future. In order to develop such a methodology or a figure of merit for the high-k material, we need to list the relevant material properties and their importance or weight to which the figure of merit (FOM) may be linked. Such a merit list has to be preceded by a consideration of the important properties we need for a device quality MOSFET gate stack. On the basis of the analysis we have in this chapter, and also in the later chapters, these considerations may include:

1. Gate stack capacitance C_{di} is directly proportional to the dielectric constant k; hence, k is a very important material constant and should be an FOM component.
2. The channel mobility is equally important, but unfortunately, is not directly linked to the high-k material constants. The channel mobility is affected by a host of scattering mechanisms or potential energy perturbations; even these scattering mechanisms are not directly linked to the high-k material properties. So, it is difficult to formulate the channel mobility in terms of the high-k material constants. Perhaps, the least difficult among these is the case of the

Coulomb scattering which may be linked to the trap density at the semiconductor/IL interface D_{it} and the net gate stack charge density $Q_{di,gsc}$. As has been discussed, the gate stack charges and the interface traps may be related to the ionicity or the electro-negativity difference and the coordination number in an indirect manner. Since D_{it} can be experimentally extracted (from the conductance technique), it may be prudent to link the channel mobility μ_{ch} to the interface trap density D_{it}; hence D_{it} may be included as an FOM component.

3. The mathematical relations derived in Chap. 2 for the channel parameters (drain current, channel conductance, and the transconductance) in the case of the high-k gate stacks indicate the classical or the ideal drain current to be degraded by three factors—namely the non-saturating inversion surface potential, the net gate stack charge density $Q_{di,gsc}$, and the work function difference ϕ_{MS}. It is not clear at the moment whether the first factor has any correlation with the high-k properties; hence we may include ϕ_{MS} as an FOM component; the parameter $Q_{di,gsc}$ may not be readily available from experiments.
4. The gate leakage current density J_{di} disables the MOSFET if it exceeds a certain limit. In device grade high-k MOSFETs this leakage current is generally due to direct tunneling or trap-assisted tunneling (Frenkel-Poole mechanism) and may be directly related to high-k material constants such as band-gap E_G, electron affinity χ_{di}, effective tunneling mass m_t*, and the gate stack property—the bulk trap density distribution $D_{bt}(E,x)$ in terms of both trap energy and its physical location inside the gate stack. There are several problems in linking J_{di} to these parameters. Firstly, the effective tunneling mass is generally unknown and is difficult to extract reliably from experiments. Secondly, the bulk trap distribution is equally unknown and to date there exists no reliable technique to extract it. Thirdly, closed-form reliable mathematical relations do not exist to represent the tunneling process inside a multi-layer high-k gate stack. Under these circumstances the best option may be to include J_{di} itself at a bias of 1.0 V and for a CET of 1.0 nm, as an FOM component.
5. The high-k gate stack degradation may be one of the most important criteria for the figure of merit; unfortunately, it is still not well understood for the high-k multi-layer system. There are slow and fast trapping processes in the case of the high-k gate stacks, as has been discussed in Chap. 8. The fast transient trapping/de-trapping is reversible and is believed to be due to traps in the high-k layer; hence it may be ignored. The slow trapping process and the resultant degradation have characteristics similar to that of the degradation process in the case of the single SiO_2 gate dielectric. Experiments suggest the slow degradation process to originate from the pre-existing traps in the intermediate layer and/or at the Si/IL interface. Therefore, one may include the interface trap density D_{it} to represent the high-k gate stack degradation in the relation for FOM.
6. The discussion in this chapter will suggest that, unless the gate dielectric is a single crystalline layer, the preferred phase for the high-k gate stack is the amorphous phase. Therefore, the amorphous to crystalline phase change

temperature T_{ther} (thermal stability) is an important consideration and component for the FOM.

7. Another important thermal/chemical stability issue is interlayer diffusion and subsequent chemical reaction. Therefore, diffusivity or diffusion constant may be chosen as an FOM component. However, the diffusion constant may vary a great deal from one diffusing species to another; further, the diffusion constants, for the high-k gate stack with a location-variable chemical composition, are difficult to extract and moreover may change during the device processing.
8. The band-gap E_G of the high-k layer and the electro-negativity difference ΔEN have so far not been linked to any of the MOSFET or high-k gate stack property, but, these may be included as FOM components, as the band-gap of the high-k layer will have some say on the gate leakage current and the gate stack degradation, whereas the electro-negativity difference will have influence on the gate stack charges and traps.
9. As the discussion in this chapter will suggest, an effective intermediate layer (IL) is an important component of the gate stack, particularly in the case of sub-nanometer values of EOT; the optimal value of the thickness (EOT) of the IL will depend upon its quality, the high-k layer, the semiconductor substrate, and the gate stack processing. The quality of the intermediate layer will determine D_{it}, the gate stack leakage current, the channel mobility, the gate stack degradation, etc., whereas its thickness will determine the EOT of the gate stack. As the thickness (EOT) of the IL is available from experiments, it may be included as an FOM component.

In the course of the above discussion, we have identified the following as the appropriate FOM components and the MOSFET properties each is linked to:

1. Dielectric constant k (total gate stack capacitance density C_{di});
2. Interface trap density D_{it} (channel mobility μ_{ch} and gate stack degradation);
3. EOT of the intermediate layer EOT_{IL} (gate stack EOT, D_{it}, μ_{ch}, gate stack degradation, J_{di});
4. Work function difference ϕ_{MS} (drain current I_D, channel conductance g_D);
5. Gate dielectric leakage current density J_{di} at 1.0 V for a CET = 1.0 nm (CET limit of MOSFET);
6. Amorphous to poly-crystalline phase transformation temperature T_{crys} (structural/phase stability);
7. Band-gap E_G (Dielectric leakage current, gate stack reliability);
8. Electro-negativity difference ΔEN (Gate stack defects, charges, traps).

We have tried to select those material constants/properties or parameters as FOM components whose values can be extracted reliably and avoided properties/constants such as the effective tunneling mass or the high-k bulk trap density whose values are unknown and also would be difficult to obtain. The dielectric constant k can be obtained from the literature or a plot of capacitance density of an MIM capacitor versus its physical thickness. The interface trap density D_{it} can be obtained from the conductance technique. EOT_{IL} can be obtained from a plot of

Table 1.5 Figure of merit (FOM) component, its normalization constant, and its relative weight

FOM component	Normalization value	Weight	Remarks
k	$k_{op} = 30$	0.18	At GHz frequency
D_{it}	$D_{it,op} = 1 \times 10^{11} cm^{-2}\ V^{-1}$	0.10	Around Si mid-gap
EOT_{IL}	$EOT_{IL,op} = 0.4$ nm	0.17	
ϕ_{MS}	$\phi_{MS,op} = 0.2$ V	0.10	
J_{di} at 1.0 V for CET = 1.0 nm	$J_{di,op} = 1\ A\ cm^{-2}$	0.15	Inversion bias
T_{ther}	$T_{ther,op} = 1{,}000$ °C	0.10	
E_G	$E_{G,op} = 6.0$ eV	0.10	
ΔEN	$\Delta EN_{op} = 2.14$	0.10	Value for HfO_2

C_{di} versus gate stack EOT. The metal work-function difference ϕ_{MS} can be obtained from the V_{FB} versus EOT measurement. J_{di} is a directly measured parameter. The crystallization temperature T_{crys} can be obtained from the high temperature processing and the physical characterization (TEM/X-ray) experiments. The band gap E_G and the electro-negativity difference can be obtained from the literature.

In order to link the above eight FOM factors to each other and finally to the figure of merit $FM_{high\text{-}k}$, we need to normalize each factor and assign a weight W to each of them; both of these have to be chosen carefully after due consideration. The weight W is meant to represent the relative importance of the parameter, whereas the normalization factor could be the optimal or the ideal value of the FOM component. Table 1.5 is a list of the FOM component, its normalization factor, and its weight.

As explained already and will be further discussed in the chapters to follow, there is a trade-off between a high k-value and its negative consequences—higher gate stack leakage current, defects, and charges. To realize an EOT of 0.5 nm, we may need a k-value of around 30, which may be achievable by intelligent doping (i.e. decreasing the molar volume) and/or lowering the phonon frequency. Strictly speaking, the value of k should be measured at GHz frequencies (operating frequency of MOSFETs). For a device grade single SiO_2 gate dielectric, the mid-gap interface state density D_{it} is of the order of $1 \times 10^{10} cm^{-2}\ V^{-1}$; such a low D_{it} may not be a realistic target for the high-k gate stack. Experimental data indicate D_{it} to increase with decreasing IL thickness (cf. Chap. 2), perhaps due to the proximity of the high-k layer to the Si/IL interface; hence a target of $1 \times 10^{11}\ cm^{-2}\ V^{-1}$ may be realistic. The value chosen for $J_{di,op}$ is the target set by ITPR for an EOT of 1.0 nm. The value of 1,000 °C chosen for $T_{ther,op}$ may be reasonable as the implant activation annealing is generally carried out at this temperature. As there is a trade off between the band-gap and dielectric constant, the value of 6.0 eV chosen for $E_{G,op}$ may be high enough. The value chosen for ΔEN_{op} is what obtains for HfO_2 on a Pauling scale.

The weights assigned in Table 1.5 are rather tentative; it may not be possible to have a rigorous basis for these. The values for the weights have been chosen on the basis of the importance and the number of MOSFET performance parameters the

FOM components influence (as analyzed above). In conclusion of the above discussion, the following relation is suggested for the figure of merit of the high-k gate stack:

$$FM_{high-k} = \frac{k}{k_{op}} W_k + \left(\ln e \frac{D_{it,op}}{D_{it}}\right) W_D + \left(\ln e \frac{EOT_{IL,op}}{EOT_{IL}}\right) W_{IL} + \left(\ln e \frac{\phi_{MS,op}}{\phi_{MS}}\right) W_\phi + \left(0.20 \log_{10} \frac{J_{di,op}}{J_{di}}\right) W_J + \frac{T_{ther}}{T_{ther,op}} W_T + \frac{E_G}{E_{G,op}} W_E + \frac{\Delta EN_{op}}{\Delta EN} W_{EN} \quad (1.4)$$

It may be noted that for a gate stack with the optimal values of all the eight FOM parameters, the value of the figure of merit is 1.0. Illustrative calculations of the FOM for 4 indicative gate stacks are presented below. The data for these calculations were chosen on the basis of the data presented for these gate stacks in the indicated chapters and the rest from the literature.

Amorphous La_2SiO_5 Chap. 9
FM = 0.01[(20/30) × 18 + (lne10^{11}/5 × 10^{11}) × 10 +
(lne0.4/0.7) × 17 + (lne × 0.2/0.2) × 10 + (0.20log10/10^{-2}) × 15 +
(900/1,000) × 10 + (6.5/6) × 10 + (2.14/
1.94) × 10] = 0.01[12 + 10 + 7.48 + 10 + 9 + 9 + 10.8 +11] = 0.79
Amorphous $LaLuO_3$ Chaps. 9 and 10
FM = 0.01[(32/30) × 18 + (lne10^{11}/10^{11}) × 10 +
(lne0.4/0.5) × 17 + (lne × 0.2/0.1) × 10 + (0.20lne10/1) × 15 +
(1,000/1,000) × 10 + (5.2/6) × 10 + (2.14/
2.26) × 10] = 0.01[19.2 + 10 + 13.2 + 16.9 + 3 + 10 + 8.7 + 9.5]= 0.91
Epitaxial Gd_2O_3 Chap. 11
FM = 0.01[(24/30) × 18 + (lne10^{11}/10^{11}) × 10 + (lne0.4/
0.3) × 17 + (lne × 0.2/0.1) × 10 + (0.20log10/10^{-4}) × 15 +
(1,000/1,000) × 10 + (6/6) × 10 + (2.14/
2.24) × 10] = 0.01[14.4 + 10 + 21.9 + 16.9 + 15 + 10 + 10 + 9.6] = 1.08
Amorphous HfSiON: Chap. 4
FM = 0.01[(15/30) × 18 + (lne10^{11}/10^{11}) × 10 +
(lne0.4/0.7) × 17 + (lne × 0.2/0.2) × 10 + (0.20log10/1) × 15 +
(1,000/1,000) × 10 + (6.1/6) × 10 + (2.14/
1.60) × 10] = 0.01[9.0 + 10 + 7.48 + 10 + 3 + 10 + 10.2 + 13.4] = 0.73

1.18 Summary

As a gate dielectric material, SiO_2 is almost perfect, i.e. it has all the properties desired in a gate dielectric for the CMOSFET, except its very low dielectric constant k. Its exceptionally high band-gap and its relatively large effective mass

allowed SiO_2 gate dielectric layers as thin as 1.3 nm to compensate for the low value of its k. That was as thin the SiO_2 layer could get as gate dielectric. A decade or so ago, the binary high-k oxides began their career with a number of weaknesses; these included a poor interface with silicon, crystallization temperature much lower than that for the implant activation, high leakage current density, threshold voltage anomaly, high gate stack charge density, easy diffusion of oxygen through the gate stack, and gate stack degradation. Many of these problems have since then been solved and today the high-k gate stacks with the metal electrodes are a much better lot having found employment in the 45, 32, and 22 nm node technologies. The interface problem with silicon has been solved by introducing an intermediate Si-based layer between Si and the high-k bulk layer. The phase stability and the low crystallization problem have been solved by alloying with suitable oxides, whereas the easy diffusion of oxygen and impurities has been reduced by introducing nitrogen. It is now possible to tune the threshold voltage by several options such as manipulation of the interface dipole or the oxygen vacancies; however, complete control of the threshold voltage is still eluding. The gate leakage current density has been brought down to satisfactory levels. The gate stack charge density is still very high and the gate stack degradation is still not well understood. Much remains to be done in the area of high mobility channels. Ge surface passivation has made impressive progress; the passivation of some III–V compound semiconductors has also made progress, but the problem is still not satisfactorily solved. An EOT of 0.5 nm looks a reachable goal for CMOSFETs with high-k/metal-gate gate stacks.

References

1. P.V. Gray, Tunneling from metal to semiconductors. Phys. Rev. **140**, A179–A186 (1965)
2. S. Kar, W.E. Dahlke, Appl. Phys. Lett. **18**, 401 (1971)
3. S. Kar, W.E. Dahlke, Interface states in MOS structures with 20-40 a thick SiO2 films on nondegenerate Si. Solid-State Electron. **15**, 221–232 (1972)
4. S. Kar, W.E. Dahlke, Potentials and direct current in Si-(20 to 40 A)-metal structures. Solid-State Electron. **15**, 869–875 (1972)
5. H.C. Card, E.H. Rhoderick, J. Phys. D: Appl. Phys. **4**, 1589–1601 (1971)
6. H.C. Card, E.H. Rhoderick, J. Phys. D: Appl. Phys. **4**, 1602–1611 (1971)
7. M.A. Green, J. Shewchun, Current multiplication in metal-insulator-semiconductor (MIS) tunnel diodes. Solid-State Electron. **17**, 349–365 (1974)
8. M.A. Green, J. Shewchun, Solid-State Electron. **17**, 941 (1974)
9. H.S. Momose, M. Ono, T. Yoshitomi, T. Ohguro, S. Nakamura, M. Saito, H. Iwai, 1.5 nm direct-tunneling gate oxide Si MOSFET's. IEEE Trans. Electron Devices **43**, 1233–1242 (1996)
10. H.S. Momose, S. Nakamura, T. Ohguro, T. Yoshitomi, E. Morifuji, T. Morimoto, Y. Katsumata, H. Iwai, Study of the manufacturing feasibility of 1.5-nm direct-tunneling gate oxide MOSFET's: uniformity, reliability, and dopant penetration of the gate oxide. IEEE Trans. Electron Devices **45**, 691–700 (1998)
11. S. Kar, R. Singh, Correlation between the material constants of and a figure of merit for the high-K gate dielectrics. in *Proceedings of ECS PV 2002–28*, pp. 13–24, 2002

12. E.V. Stefanovich, A.L. Shluger, C.R.A. Catlow, Phys. Rev. **B49**, 11560 (1994)
13. K.P. Bastos, J. Morais, L. Miotti, R.P. Pezzi, G.V. Soares, I.J.R. Baumvol, R.I. Hegde, H.H. Tseng, P.J. Tobin, Appl. Phys. Lett. **81**, 1669 (2002)
14. H.-J. Cho, C.S. Kang, K. Onishi, S. Gopalan, R. Nieh, R. Choi, S. Krishnan, J.C. Lee, IEEE Electron Device Lett. **23**, 249 (2002)
15. R. Ruh, P.W.R. Corfield, J. Am. Ceram. Soc. **53**, 126 (1970)
16. N.C. Stephenson, R.S. Roth, Acta. Cryst. **B27**, 1037 (1971)
17. J.I. Pankove, *Optical Processes in Semiconductors* (Prentice Hall, Englewood Cliffs, 1971)
18. B.E. Deal, A.S. Grove, J. Appl. Phys. **36**, 3770 (1965)

Chapter 2
MOSFET: Basics, Characteristics, and Characterization

Samares Kar

Abstract This chapter attempts to provide a theoretical basis for the Metal Oxide (Insulator) Semiconductor (MOS/MIS) Structure and the MOS/MIS Field Effect Transistor (MOSFET/MISFET), their characteristics, and their characterization (parameter extraction); the theoretical treatment starts from the first principles. While deriving the mathematical relations, assumptions have been avoided as far as possible. A comprehensive treatment is included which covers the important aspects of the function, mechanism, and operation of the MOS/MIS devices; in particular topics have been covered which are relevant to all the later chapters of the book and which will aid in reading the rest of this book. We begin this chapter with the theory of the classical MOS structure (non-leaky and SiO_2 single gate dielectric) and the classical MOSFET and then graduate to the MOS structure and the MOSFET with the high-k gate stack and the high mobility channels. Various aspects of the MOS/MOSFET devices analyzed in this chapter include the energy band profiles, circuit representations, electrostatic analysis (charge–voltage and capacitance–voltage relations), drain current versus drain voltage relation, quantum-mechanical phenomena (wave function penetration, tunneling, carrier confinement), nature of the high-k gate stack traps, and the pseudo-Fermi function inside the gate stack and the occupancy of the gate stack traps. Features such as capacitance–voltage (C–V) characteristics, flat-band and threshold voltages (V_{FB} and V_T), V_T versus EOT characteristics permeate the chapters; hence these features and characteristics such as conductance–voltage (G–V) characteristics have been discussed. A significant part of this chapter contains topics which are rarely seen in the literature and are yet to be well understood. As these topics (composition of the high-k gate stack, nature of the high-k gate stack charges, effects of the degradation factors) are of vital significance for the progress of the high-k gate stack technology, we have tried to analyze these issues. The final part of this chapter treats the various methods available for characterization of the high-k gate

S. Kar (✉)
Department of Electrical Engineering, Indian Institute of Technology, Kanpur 208016, India
e-mail: skar@iitk.ac.in

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5_2,

stacks, in particular, for the determination of the trap parameters—trap density, trap energy, trap capture cross-section, and the trap location inside the gate stack.

2.1 Introduction

This chapter will attempt to cover the basics of the electrical properties and the electrical characteristics, in particular, the concepts and the theory, which will help in reading the different chapters of the book. Our approach will be the following:

1. We will define any topic (e.g. energy band diagrams), technical term (e.g. frequency dispersion) or parameter (e.g. transconductance) used in this chapter at the place of its introduction in textual and/or mathematical form.
2. We will outline the importance and/or usefulness and/or application of the technical entity.
3. Set up the boundary conditions and outline all the assumptions made for obtaining any closed-form (analytical) equation, beginning from the first principles; outline all the important steps and the logic inputs in the solution process, and if necessary cite the source of a more detailed mathematical treatment.

Our focus will be on simplicity, continuity, and a complete coverage, and on avoiding all unnecessary complexity in algebra. The basics of the Metal Oxide Semiconductor (MOS) structure and the MOS Field Effect Transistor (MOSFET) and their important electrical characteristics and parameters will be covered. Our treatment will focus on the high-k (k is the dielectric constant) gate stacks, and will also include the high mobility channels.

The material presented in this chapter is linked to and overlaps with almost all the following chapters of the book. The following is an indicative outline of the linkage:

1. Chapter 3: Quantum mechanical tunneling;
2. Chapter 4: Quantum mechanical tunneling;
3. Chapter 5: Flat-band voltage, interface traps, Metal Induced Gap States (MIGS), charge neutrality level, Capacitance–Voltage (C–V) characteristics;
4. Chapter 6: Flat-band voltage, interface traps, MIGS, charge neutrality level, C–V characteristics;
5. Chapter 7: Carrier mobility, drain current in the linear regime, C–V characteristics;
6. Chapter 8: Traps, trap time constant, oxygen vacancies;
7. Chapter 9: C–V characteristics;
8. Chapter 10: Quantum mechanical tunneling, C–V characteristics, flat-band voltage;
9. Chapter 12: C–V and Conductance–Voltage (G–V) characteristics, parameter extraction.

References [1–3] are quality text/reference books on MOS structures and MOSFET containing the SiO_2 single gate dielectric. Reference [4] is a classical paper on many important topics of the MOS structures, particularly, its conductance originating from the interface traps. These references together contain most of the details of analysis of the classical devices with the SiO_2 single gate dielectric.

2.2 MIS/MOS Structure: Single SiO_2 Gate Dielectric

The Metal-Insulator/Oxide-Semiconductor (MIS/MOS) structure is the most important component of the MOSFET/MISFET. Its basic function is to modulate (and change the conductivity type—p to n or n to p) the channel conductivity, thereby modulating the drain current, which flows by drift, provided a channel exists from the source to the drain. In the high-k transistors, the semiconductor is silicon or a higher mobility material, e.g. Ge, SiGe, GaAs, InGaAs, InAs, InAlAs, InP, GaN, InSb, HgTe, PbTe, and the insulator is multi-layered, whose core is a high-k material, e.g. HfO_2, La_2O_3, HfSiO, HfAlO, HfNO, HfSiON, $ErTiO_5$, $SrTiO_3$, $LaScO_3$, $LaAlO_3$, $GdScO_3$, $LaLuO_3$, $La_2Hf_2O_7$, Gd_2O_3, La_2SiO_5, $SrHfO_3$. To increase the permittivity (the gate stack capacitance), and the channel mobility, the future trend is towards higher-k gate dielectrics, and higher-mobility substrates.

We begin our analysis of the MIS structure with the following assumptions:

1. The semiconductor is p-type silicon; i.e. the channel on this substrate will be an n-channel—NMOSFET.
2. The gate dielectric is a dry thermal SiO_2, as is the case when the Equivalent Oxide Thickness (EOT) is more than 1.3 nm or so.
3. The leakage current through the gate dielectric is insignificant.

Later, we will analyze, what complications the following features of a gate stack introduce into the treatment:

1. A leaky gate dielectric or gate stack.
2. A gate stack with multiple dielectrics (both covalent and ionic).
3. A high mobility channel material.

The most important aids in understanding the electrical characteristics of the MOS structure and the MOSFET are their energy band diagrams and the circuit representations.

2.2.1 Metal-Semiconductor Contact (Schottky Barrier)

One may consider the metal-semiconductor (MS) contact (or junction) to be a subclass of the MOS/MIS structure where the oxide/insulator thickness is nil. This device is popularly known as a Schottky barrier or diode or contact, named after

Walter Hermann Schottky, who made important contributions to the development of its theory, in particular the image force barrier lowering. Reference [5] is a good source for details on this topic. The Schottky diode itself is an important semiconductor device; in addition, it is an important component of the Schottky transistor (a bipolar junction transistor with a Schottky barrier between the base and the collector), Metal Semiconductor FET (MESFET in which the control electrode is the MS contact), High Electron Mobility Transistor (HEMT), and Schottky barrier carbon nano-tube (CNT) transistor. The MESFET and HEMT [6] have been the substitutes for the MOSFET/MISFET configuration on compound semiconductors, as earlier and still today it is difficult to realize high quality MOSFET/MISFET configurations on compound semiconductor substrates.

We have a more direct cause for looking at and analyzing the formation of the Schottky barrier and its basic nature. The MS contact is a limiting case of the MOS structure. The ultrathin MOS gate stack, with EOT less than 1.0 nm and scaling down to an EOT of 0.5 nm, is much closer to the MS contact than it is to the classical MOS structure. Therefore, an understanding of the MS contact is useful in analyzing the ultrathin MOSFET gate stack.

Many MS contacts may have an interfacial layer (say up to a nm thick) between the semiconductor and the metal [5]; however, it may be possible to avoid the interfacial layer as in a silicon/silicide MS contact [2]. When an intimate contact is established between the semiconductor and the metal, electron or hole exchange must take place between these two layers to equalize the energy and the Fermi level (What we call the Fermi level in semiconductor physics, is actually the chemical potential. At times, Fermi level is confused with the Fermi energy. The Fermi energy is defined only at absolute zero—0 K; it is the energy of the highest energy electron at 0 K). Figure 2.1a illustrates the energy band diagram across an MS contact before the intimate contact and Fig. 2.1b after the intimate contact. In the situation represented by Fig. 2.1a, the higher energy electrons in the conduction band are transferred from the n-type semiconductor to the metal, thereby setting up a dipole layer across the MS interface with the positive layer of ionized donors in the semiconductor sub-surface and the negative layer of transferred electrons on the metal surface.

The dipole layer sets up an electric field and a surface potential $\varphi_{s,0}$ in the semiconductor space charge layer. The charge on the metal side must remain on its surface, as no electric field can penetrate a conductor. The metal surface charge density Q_M has to be equal and opposite of the semiconductor space charge density Q_{sc}, if the interface trap charge density Q_{it} is nil. The penetration of the electric field into a semiconductor depends upon its free carrier density and the charge on the semiconductor side resides in its space charge layer, a detailed analysis of which will be taken up later in Sect. 2.2.4. For a Schottky barrier, the semiconductor space charge is dominated by the charge of the ionized dopants. Perhaps the most important MS contact parameter is the Schottky barrier height, which dictates almost all the other parameters. The Schottky barrier height is the potential energy barrier perceived by a metal electron when transported from the metal to the semiconductor. The barrier height depends critically on the nature of the MS

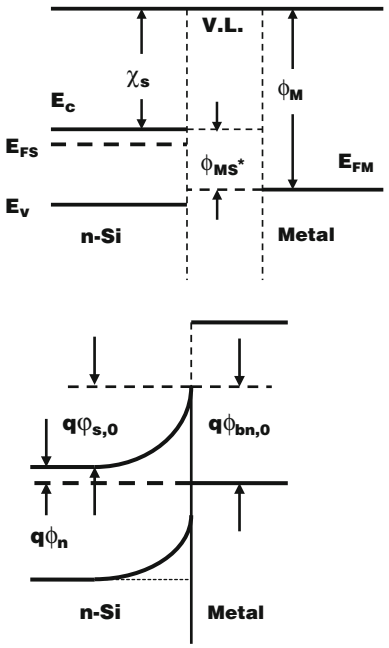


Fig. 2.1 Energy band diagram across an n-semiconductor/metal MS contact (*top*) before and (*bottom*) after an intimate contact. E_c, E_v are the conduction, valence band edges, E_{FS}, E_{FM} are the semiconductor, metal Fermi levels, q is the electron charge, ϕ_M is the metal work function, χ_s is the semiconductor electron affinity, $\phi_{bn,0}$ is the zero-bias Schottky barrier height on n-type semiconductor, and $\varphi_{s,0}$ is the zero-bias surface potential. V.L. is the vacuum level

interface—the interface trap density D_{it}. Two limits are invoked while discussing the Schottky barrier height—the Schottky–Mott limit in which case $D_{it} = 0$ and the Bardeen-limit in which case the high interface trap density pins the interface Fermi level. Let us take up the classical or the ideal case (Schottky–Mott limit) first. In this case, as represented by Fig. 2.1b, the barrier height for zero applied bias $\phi_{b,0}$ is given by:

$$\phi_{bn,0} = \frac{\phi_M - \chi_s}{q} = \varphi_{s,0} + \phi_n \tag{2.1}$$

ϕ_n is the Fermi potential in n-semiconductor. Generally an interface state represents an allowed state inside the band-gap at the interface. A band-gap normally is an energy interval where the density of states is zero, i.e. where any states are forbidden. An interface represents a severe discontinuity in the periodic lattice; hence the solutions of the Schrödinger equation for the bulk crystal (with a periodic atomic arrangement and a periodic potential energy) do not obtain at the interface. Therefore allowed states may exist at the interface in the energy range which is a band-gap for the bulk crystal; in fact the entire distribution of the states at the interface may be different in the other energy ranges also, from what obtains in the bulk of the crystal. In addition to the intrinsic interface states, which are present due to the lattice mismatch at the interface, extrinsic states may also be present due to lattice defects and alien atoms, for which the interface is a center of attraction and a sink. In spite of a large amount of research carried out, the theoretical basis of the interface states is still incomplete [7].

Several concepts have been introduced in this area, one of which is the Metal Induced Gap States (MIGS), and one which we will come across in several chapters of the book. MIGS, which are one type of the intrinsic interface states, are due to the metal wave functions penetrating and decaying inside the semiconductor. A question related to the MIGS is whether the metal and the semiconductor wave functions can mix at the interface region. Charge Neutrality Level (CNL) is another concept which will be encountered often in connection with the interface traps. The interface trap charge density is zero or the interface trap charge is neutral if the Fermi level is at the CNL. Another related concept is the pinning parameter S, defined as $= \partial\phi_b/\partial\phi_M$. The Schottky-Mott limit corresponds to $S = 1$, i.e. $D_{it} = 0$, whereas the Bardeen limit corresponds to $S = 0$, i.e. $D_{it} = \infty$. When the interface trap density is infinite, then the interface traps pin the interface Fermi level so strongly that the Schottky barrier height becomes invariant of the metal work function, and in such a case, the barrier height is given by:

$$\phi_{bn,0} = \phi_M - \phi_{CNL} \tag{2.2}$$

ϕ_{CNL} is the CNL measured from the vacuum level. In most cases, the pinning parameter will lie between 0 and 1; in such cases, the barrier height will be given by:

$$\phi_{bn,0} = (\phi_M - \phi_{CNL}) + S(\phi_{CNL} - \chi_s) \tag{2.3}$$

The Schottky barrier is a rectifying contact, i.e. a large current flows in the forward (a positive voltage applied on the metal with respect to an n-type semiconductor) direction, and a much smaller current in the reverse direction. Several mechanisms can operate simultaneously across an MS contact to transport carriers. In the case of a device quality Si/Metal Schottky barrier, the dominant carrier transport mechanism is likely to be thermionic emission over the potential energy barrier.

2.2.2 Energy Band Diagram

The profile of the energy bands along the axis of the applied electric field is very helpful in understanding the basic mechanisms of any semiconductor device. In these diagrams, the x axis represents most often the direction of the applied electric field, and the y axis the electron energy (generally electron energy increasing in the positive y direction and hole energy increasing in the negative y direction). Figure 2.2 illustrates the energy band profile of a p-Si/SiO_2/Metal structure across the x-axis, along which the gate voltage, V_G, is applied, for the equilibrium condition, i.e. the drain voltage $V_D = 0$. E_G is the semiconductor band gap, φ_s is the semiconductor band-bending (surface potential), and V_{di} is the potential across the gate dielectric. If we sum the energy parameters between the vacuum level (VL)

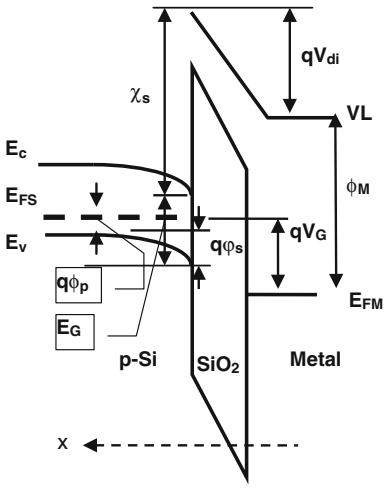


Fig. 2.2 Energy band profile across a p-Si/SiO_2/metal structure across the x direction, along which the gate voltage has been applied. VL is vacuum level, i.e. the energy an electron would have in absolute vacuum

and E_{FS} on both surfaces of the gate dielectric and equate, the following relation results:

$$\chi_s + E_G - q\varphi_s - q\phi_p = qV_{di} + \phi_M - qV_G$$

Rearranging the terms, we obtain a relation for the gate voltage in terms of the potentials across the semiconductor and the gate dielectric, φ_s and V_{di}, respectively, and the work-function difference between the metal and the semiconductor, ϕ_{MS}.

$$V_G = \varphi_s + V_{di} - \lfloor (\chi_s + E_G - \phi_M)/q - \phi_p \rfloor \tag{2.4}$$

$$\phi_{MS} = (\chi_s + E_G - \phi_M)/q - \phi_p \tag{2.5}$$

In the ideal case, the work function difference is zero, and this is what one tries to achieve by a suitable choice of the metal work function and the semiconductor doping density.

2.2.3 Equivalent Circuit Representation

The salient parts of a device are generally reflected as elements in its circuit representation, which also has to be consistent with its energy band profile, as well as with the theoretical current voltage relations of the device. Figure 2.3 illustrates a relatively simple (assuming single-level interface states, equipotential semiconductor surface, etc.) equivalent circuit for the p-Si/SiO_2/Metal structure at an intermediate frequency. In Fig. 2.3, the gate dielectric is represented by the dielectric capacitance density, C_{di}, the semiconductor space charge layer is

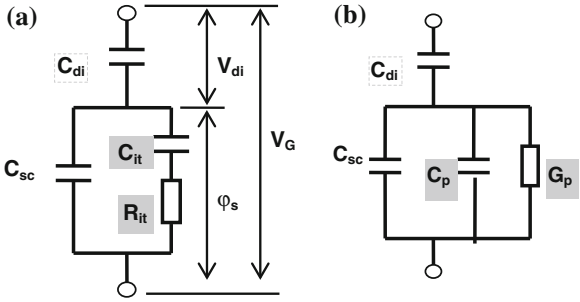


Fig. 2.3 **a** Circuit representation of an MIS structure with a single gate dielectric, at an intermediate frequency, i.e. neither low nor high; **b** transformation of the circuit in **a**

represented by the semiconductor space-charge capacitance density, C_{sc}, and the interface traps by the interface trap capacitance density, C_{it} and the interface trap charging resistance, R_{it}, or alternatively by the equivalent parallel trap conductance density, G_p, and the equivalent trap capacitance density, C_p. The theoretical treatment is greatly simpler when the gate dielectric is a dry thermal SiO_2 layer, which for all practical purposes, is a near-perfect dielectric, i.e. is free of any bulk charges—both fixed (which do not charge or discharge in the operating voltage range) as well as trap charges. A perfect dielectric has only a dielectric capacitance, cf. (2.6), as in a plane parallel capacitor, and the electric field is constant inside, as reflected in Fig. 2.2 by the linear energy band profiles across the SiO_2 layer. Presence of bulk charges—both fixed and trap—create an additional nonlinear potential and a varying electric field across the gate dielectric; additionally, the presence of bulk trap charges creates a trap capacitance in series with a trap resistance in parallel with the dielectric capacitance of the gate insulator. This issue will be further treated later, when we analyze the high-k gate stacks.

It may be noted that in Fig. 2.3a, the space charge layer in the semiconductor, which varies (charges or discharges) with the surface potential φ_s (i.e. the potential across this layer), is represented only by a capacitance (C_{sc}), while the interface traps (at the Si/SiO_2 interface) have been represented by a capacitance (C_{it}) in series with a charging resistor (R_{it}). In principle, any charging/discharging phenomenon is to be represented by a series RC branch, where C is the rate of charging/discharging ($\partial Q/\partial V$; Q is the charge and V is the potential.), R is the charging resistor, and τ (=RC) is the time constant. The series RC branch representation is the electrical engineering equivalent of the Kramers-Kronig relation in physics. In plain words, charging/discharging lags the application of a potential change by a typical time, called the relaxation time in physics and the time constant in electrical engineering. Often to simplify the analysis, we short-circuit the charging resistor, if the inverse of applied angular signal frequency, ω, is much larger than the relaxation time (i.e. the applied signal is easily followed by the charging/discharging process), and open-circuit it if the vice versa is the case. We have short-circuited the charging resistor for C_{sc} in Fig. 2.3, as the relaxation time (of the order of ps [3]) in this case is many orders of magnitude smaller than the inverse of the operating frequency (a few GHz).

The circuit representation in Fig. 2.3b results, when we transform the series $R_{it}C_{it}$ branch in Fig. 2.3a into a parallel G_p and C_p combination. We will see in later sections how the circuit of Fig. 2.3b is more appropriate for the analysis of the experimental interface state admittance data. The circuit representation of Fig. 2.3a simplifies to the one in Fig. 2.4a at high frequencies (f_h), and to the one in Fig. 2.4b at low frequencies (f_l). Any frequency much higher than the inverse of the inversion layer time constant τ_{inv} as well as the inverse of the minimum interface state time constant τ_{it} is a high frequency. Any frequency much smaller than the inverse of the inversion layer time constant τ_{inv} as well as the inverse of the maximum interface state time constant τ_{it} is a low (also called equilibrium) frequency.

The inversion layer time constant τ_{inv} and the interface state time constant τ_{it} are functions of the surface potential φ_s, and depend upon whether the device is an MOS capacitor or an MOSFET in operation. Both the low and the high frequency circuit representations of Fig. 2.4 are capacitive. At low frequencies, total MOS capacitance density C_{lf} is given by the relation:

$$\frac{1}{C_{lf}} = \frac{1}{C_{di}} + \frac{1}{C_{sc} + C_{it}}; \text{ where } C_{di} = \frac{\varepsilon_{di}}{t_{di}}; \text{ and } C_{it} = q \cdot D_{it} \tag{2.6}$$

ε_{di} is the dielectric permittivity, t_{di} is the dielectric thickness, and D_{it} is the interface state density. The simple relation for C_{it} in (2.6) follows from the definition: $C_{it} = \partial Q_{it}/\partial \varphi_s$ (Q_{it} is the interface trap charge density.), as will be shown later. At high frequencies, the total MOS capacitance density C_{hf} is given by the relation:

$$\frac{1}{C_{hf}} = \frac{1}{C_{di}} + \frac{1}{C_{sc,hf}} \tag{2.7}$$

The relation for the high frequency space charge capacitance density $C_{sc,hf}$ depends upon whether the device is an MOS capacitor or an MOSFET. For an MOS capacitor, $C_{sc,hf}$ will reduce to the depletion space charge capacitance density C_{dep}, whereas for the MOSFET $C_{sc,hf}$ will be the same as C_{sc}. This point will be elaborated later.

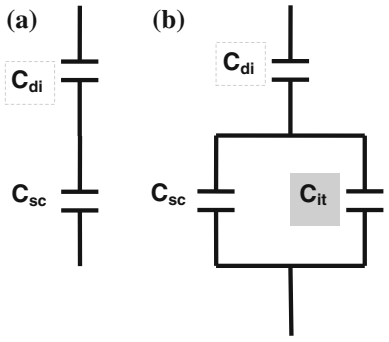


Fig. 2.4 Simplification of the equivalent circuit in Fig. 2.3a at (**a**) high frequencies; and at (**b**) low frequencies or equilibrium

2.2.4 Electrostatic Analysis

The response of an equivalent circuit to an applied signal can be expressed once the mathematical relations for the different elements of the circuit are known which we now proceed to develop from an electrostatic analysis.

The electrostatic relations between potentials, charges, capacitances, and resistances, are much simpler, when the gate dielectric is a single near-perfect insulator like SiO_2. To expand the relation in (2.4), we need to substitute the relations for V_{di} and ϕ_p. The latter has the well-known relation:

$$\phi_p = kT/q \ln N_v/N_A = V_T \ln N_v/N_A \tag{2.8}$$

T is the absolute temperature, N_v is the effective density of states in the semiconductor valence band, N_A is the acceptor density, and V_T is thermal voltage. For a near-perfect gate dielectric, i.e. SiO_2, the expression for V_{di} is much simpler, while for a high-k gate stack (with as many as two bulk layers and three interfacial layers, full of fixed charges and traps), it can be very complicated. For the dry thermal SiO_2 gate dielectric, which has, for all practical purposes, no traps or fixed charges inside, V_{di} can be given by:

$$V_{di} = Q_M/C_{di} = -\frac{Q_{sc} + Q_{it} + Q_F}{C_{di}} \tag{2.9}$$

Q_F is the fixed charge density in the vicinity of the interface. It may be noted that Q_M on the metal surface is balanced by the charges on the silicon surface: $Q_{sc} + Q_{it} + Q_F$ and therefore has the opposite sign. The interface trap charge density can be expressed as:

$$Q_{it} = q \int \left(D_{it}^D f_{FD}^h - D_{it}^A f_{FD}^e\right) dE \tag{2.10}$$

$D_{it}{}^D$ and $D_{it}{}^A$ are respectively the density of the donor-type/acceptor-type interface states, $f_{FD}{}^h$ and $f_{FD}{}^e$ are respectively the hole/electron Fermi occupancy (Fermi-Dirac distribution), and E is energy. It can be easily shown that if the expression for Q_{it} is differentiated with respect to φ_s, the expression for C_{it} in (2.6) will result, since $D_{it} = D_{it}^D + D_{it}^A$.

Unfortunately, even in the case of the non-leaky Si/SiO_2/Metal system, an expression for Q_{sc} and C_{sc} (the remaining elements of the circuit representation) cannot be found in closed form, because of two main reasons:

1. In accumulation and strong inversion, one cannot use the Boltzmann distribution, but, needs to use the Fermi occupancy.
2. Carrier confinement (rather restrictions) in the x-direction (perpendicular to the interface) in the accumulation or the strong inversion potential well leads to standing waves (not Bloch functions, which are traveling waves), resulting in energy sub-bands inside the valence and conduction bands. This phenomenon

requires simultaneous application of the Poisson equation and the Schrödinger equation to carry out the electrostatic analysis.

In the case of a high-k gate stack, further complications arise:

(a) Significant penetration of the semiconductor electron/hole wave-function occurs into the high-k gate stack and also transmission through the high-k gate stack to the metal takes place resulting in a high tunneling current, particularly for EOT in the 0.5–1.0 nm range.
(b) Intermixing and interference between the semiconductor and the metal electron/hole wave-functions is possible.

The above complications notwithstanding, a closed-form solution promotes much clearer physical understanding. This is possible only if we ignore the carrier confinement effect as well as assume a Boltzmann distribution. The mathematical formulation and treatment in the following is generally credited to Ref. [8]. The starting point for the electrostatic analysis is the solution of the Poisson equation, for which we need an expression for the charge density per unit volume at a point x in the space charge layer, cf. Fig. 2.2, $\rho(x)$. The net charge density is the sum of the positive charge densities [ionized donor density, N_D^+, and hole density, p(x)] minus the sum of the negative charge densities [ionized acceptor density, N_A^-, and electron density, n(x)]. We assume the doping density to be constant in the space charge layer.

$$\rho(x) = q\left[\sum N_D^+ - \sum N_A^- + p(x) - n(x)\right]$$

In the neutral regime, $\rho = 0$, which leads to the following simplification of the above relation:

$$\sum N_D^+ + p_0 = \sum N_A^- + n_0 \Rightarrow \rho(x) = q[(p - p_0) - (n - n_0)]$$

p_0/n_0 are the hole/electron concentrations in the neutral regime. The free carrier (electrons and holes) densities at a point x, p(x), n(x), will vary exponentially with the potential $\varphi(x)$ at point x, and therefore across the space charge layer, cf. Fig. 2.5.

$$p(x) = p_0 \exp\{-\beta\varphi(x)\}; \quad n(x) = n_0 \exp\{\beta\varphi(x)\};$$

β (=q/kT) is the inverse thermal voltage. The following steps follow from the application of the Poisson equation to the semiconductor space charge layer under the simplifications made:

$$d^2\varphi/dx^2 = 1/2\, d/d\varphi\left(d\varphi/dx\right)^2 = -q/\varepsilon_s\left[p_0\left(e^{-\beta\varphi} - 1\right) - n_0\left(e^{\beta\varphi} - 1\right)\right]$$

One integrates the Poisson equation (the left side with respect to $d\varphi/dx$, and the right side with respect to φ, after a slight rearrangement of the terms) to obtain the square of the electric field, $E(x) = -d\varphi/dx$, at point x in the space charge layer.

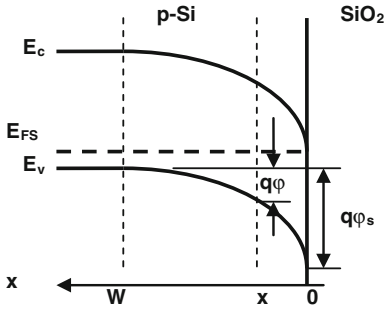


Fig. 2.5 Energy band profile across the space charge layer for a p-Si/SiO_2 system in depletion

$$-d\varphi/dx = E(x) = \pm\left[2qp_0/\beta\varepsilon_s\left\{\left(e^{-\beta\varphi} + \beta\varphi - 1\right) + n_0/p_0\left(e^{\beta\varphi} - \beta\varphi - 1\right)\right\}\right]^{1/2}$$

Subsequently one obtains the space charge density $Q_{sc} = -\varepsilon_s E_i$, where ε_s is the semiconductor permittivity, and E_i is the electric field at the interface (also the semiconductor surface, $E_i = E(x)$ at $x = 0$, cf. Fig. 2.5):

$$Q_{sc} = \int_0^W \rho(x)dx = -\varepsilon_s E_i$$

W is the space charge width, cf. Fig. 2.5. The negative sign in the above relation results from the fact that for a positive Q_{sc}, E_i is in the negative x direction, cf. Fig. 2.5. Since $E_i = E(\varphi = \varphi_s)$, the closed-form expression for Q_{sc} takes the form:

$$Q_{sc} = \mp\left[\frac{2q\varepsilon_s N_A}{\beta}\left\{\left(e^{-\beta\varphi_s} + \beta\varphi_s - 1\right) + n_0/p_0\left(e^{\beta\varphi_s} - \beta\varphi_s - 1\right)\right\}\right]^{1/2} \tag{2.11}$$

Differentiation of the space charge density with respect to the surface potential yields the closed-form mathematical relation for the space charge capacitance C_{sc}.

$$C_{sc} = \left|\frac{dQ_{sc}}{d\varphi_s}\right| = \left(\frac{q\varepsilon_s N_A\beta}{2}\right)^{1/2} \frac{\left|1 - e^{-\beta\varphi_s} + n_0/p_0\left(e^{\beta\varphi_s} - 1\right)\right|}{\left[\left(e^{-\beta\varphi_s} + \beta\varphi_s - 1\right) + n_0/p_0\left(e^{\beta\varphi_s} - \beta\varphi_s - 1\right)\right]^{1/2}} \tag{2.12}$$

Equations (2.11) and (2.12) are valid in depletion (the majority carrier density < doping density but > intrinsic carrier density n_i, i.e. for a p-type semiconductor, $N_A > p_s > n_i$) and weak inversion (the minority carrier density < doping density but > intrinsic carrier density n_i, i.e. for a p-type semiconductor, $N_A > n_s > n_i$), as in these regimes, there is no potential well formation at the semiconductor surface and the attendant carrier confinement, and the Boltzmann distribution is valid. Although (2.11) and (2.12) are not valid in accumulation (the majority carrier density > doping density, i.e. for a p-type semiconductor, $p_s > N_A$) and strong inversion (the minority carrier density > doping density, i.e.

for a p-type semiconductor, $n_s > N_A$), these still are useful in providing some insight into the physical picture (p_s, n_s are hole, electron densities at the semiconductor surface).

The issues of quasi-thermal equilibrium, Fermi-Dirac occupancy, validity of such thermal-equilibrium entities as the law of mass action, etc., are important in the operation of high-k MOSFET/MISFET and ultrathin leaky gate dielectrics, but a clear analysis of these questions may be elusive. It may be noted that while deriving (2.11) and (2.12), we have assumed the law of mass action to be valid in the x direction in the space charge layer, i.e. $p(x)n(x) = n_i^2$. This is a good assumption even in an MOSFET in operation, if the gate dielectric is non-leaky.

Equations (2.1)–(2.12), would enable determination of the gate voltage V_G and the total low frequency and the high frequency MOS capacitance densities, C_{lf} and C_{hf}, respectively, for any value of the surface potential φ_s in depletion and weak inversion regimes. Equations (2.11) and (2.12) can be significantly simplified in depletion and in weak inversion. For a p-type semiconductor both in depletion and in weak inversion, φ_s is > 0; hence $\exp(-\beta\varphi_s) \ll \beta\varphi_s \ll \exp(\beta\varphi_s)$, and $\beta\varphi_s \gg 1$. Moreover, $(n_0/p_0)\exp(\beta\varphi_s) = n_s/p_0 \ll 1$. Under these conditions, (2.11) and (2.12) approximate to:

$$Q_{sc} \approx -(2q\varepsilon_s N_A \varphi_s)^{1/2} \tag{2.13}$$

$$C_{sc} \approx \left(q\varepsilon_s N_A/2\varphi_s\right)^{1/2} \tag{2.14}$$

Equations (2.13) and (2.14) are good approximations in both depletion and in weak inversion. It is worth noting that under these conditions, the semiconductor space charge layer reduces to a dielectric capacitor, whose capacitance density is given by ε_s/W, as there are only ionized dopants (whose charge is fixed) in this layer, and the free carrier density is insignificant; so, there is no charging or discharging, when φ_s is varied.

2.3 Flat-Band Voltage and Threshold Voltage: Single SiO_2 Gate Dielectric

Two very important MOS and MOSFET parameters are the flat-band voltage V_{FB}, corresponding to the condition: $\varphi_s = 0$ (or $p_s = N_A$ for a p-type semiconductor), and the threshold voltage V_T, corresponding to the onset of the strong inversion regime ($n_s = N_A$, for a p-type semiconductor). The flat-band and the onset of strong inversion conditions are illustrated by the energy profiles in Figs. 2.6 and 2.7, respectively. Equations (2.4), (2.5), and (2.9) yield the following expression for V_{FB}: which will be valid only for a gate dielectric free of bulk charges, cf. Fig. 2.6.

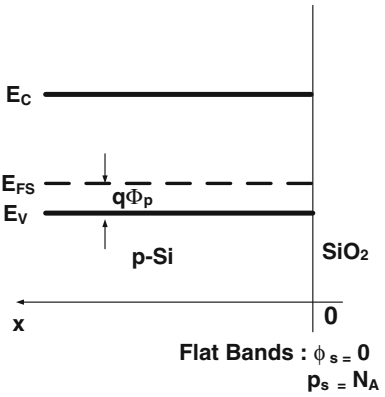


Fig. 2.6 Energy band profile across the space charge layer for a p-Si/SiO_2 system at flat band condition

$$V_{FB} = -\left[\left(Q_{it,fb} + Q_F\right)/C_{di}\right] - \phi_{MS} = -\left[\left(Q_{it,fb} + Q_F\right)\frac{t_{di}}{\varepsilon_{di}}\right] - \phi_{MS} \qquad (2.15)$$

$Q_{it,fb}$ is the interface trap charge density at flat-band (Often, $Q_{it,fb}$ is ignored, and only Q_F is considered in calculating V_{FB}.). Ideally, V_{FB} should be zero; in practice, one tries to obtain a value for V_{FB} as close as possible to zero. The flat-band voltage is for the MOSFET/MISFET a performance parameter, as it relates to the threshold voltage of the transistor, and also is a quality indicator, as it reflects the magnitude of the unwanted entities such as the trap density and the fixed charges. The device quality SiO_2 is a near-perfect gate dielectric; the experimental plot of the flat-band voltage V_{FB} versus the dielectric thickness t_{di} of an MOS structure with such a gate dielectric has been demonstrated to be a linear characteristic, as (2.15) indicates, yielding accurate values of the Si/metal work function difference and the interface charge density at flat-band, $(Q_{it,fb} + Q_F)$ [9, 10].

Equations (2.4) and (2.9) yield the following expression for the threshold voltage (MOSFET turn-on voltage) V_T: cf. Figs. 2.7 and 2.2:

$$\begin{aligned} V_T &= \varphi_{s,inv,th} + V_{di,inv,th} - \phi_{MS} \\ &= \left(\frac{E_G}{q} - 2\phi_p\right) - \left(\frac{Q_{sc,inv,th} + Q_{it,inv,th} + Q_F}{C_{di}}\right) - \phi_{MS} \end{aligned} \qquad (2.16)$$

$\varphi_{s,inv,th}$ is the surface potential, $V_{di,inv,th}$ is the potential across the gate stack, $Q_{sc,inv,th}$ is the semiconductor space charge density, and $Q_{it,inv,th}$ is the interface trap charge density, at the onset of strong inversion. At the onset of strong inversion, the Fermi level at the interface is $q\phi_p$ below the minority carrier band edge E_c, cf. Fig. 2.7. Hence the surface potential at the onset of strong inversion is given by: $\varphi_{s,inv,th} = (E_G/q) - 2\phi_p$.

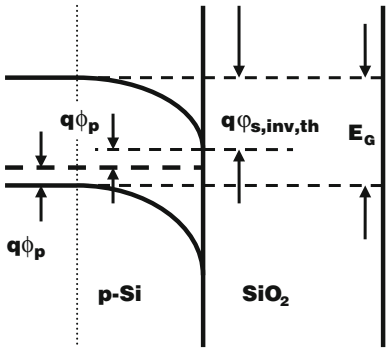


Fig. 2.7 Energy band profile across the space charge layer for a p-Si/SiO_2 system at the onset of strong inversion

2.4 Capacitance–Voltage (C–V) Characteristics of the Si/SiO_2/Metal Structures

The capacitance–voltage, the C–V, characteristic of the MOS structure is perhaps the most frequently used tool in the characterization of both the MOS and the MOSFET devices; perhaps, the most important reasons are:

1. The ease with which the C–V characteristic can be measured, even in the case of high gate stack leakage current.
2. The accuracy with which it can be measured.
3. The sensitivity with which it reflects the defects, traps, and other deficiencies of the MOS structure and the MOSFET. In other words, the C–V curve is the most direct image of the MOS quality.
4. The large number of the important device parameters which can be extracted from this characteristic.

The fact that the C–V characteristic has been referred to throughout this book underscores the need to understand this characteristic accurately, which we will try to do in this section. Equations (2.4–2.12) enable us to understand a measured high/low frequency C–V characteristic. Such characteristics are displayed in Fig. 2.9 for a p-Si/SiO_2/Al capacitor, illustrated by the schematic of Fig. 2.8. The non-leaky, 5.2 nm thick SiO_2 layer was grown in dry O_2 at 1,100 °C. The high frequency characteristic was measured at 300 kHz, while the low frequency C–V was measured at a voltage ramp rate of 0.01 V/s. In Fig. 2.9, the accumulation, depletion, weak inversion, and the strong inversion regimes have been indicated. The surface potential needed to delineate these regimes was obtained from the Berglund integral [4] of the low frequency C–V curve. Figure 2.9 demonstrates one significant change as the gate dielectric becomes thinner, and the gate dielectric capacitance becomes comparable to the space charge capacitance in accumulation. This important change is the dominance of the accumulation and the strong inversion regimes over the depletion and the weak inversion regimes in the C–V characteristics.

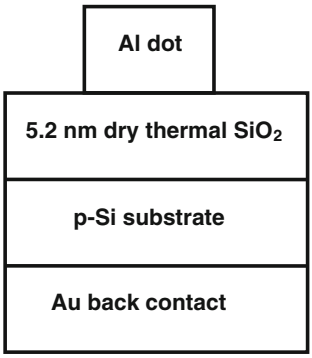


Fig. 2.8 A schematic of a p-Si//SiO_2/Al MOS capacitor, whose low and high frequency C–V characteristics are displayed in Fig. 2.9

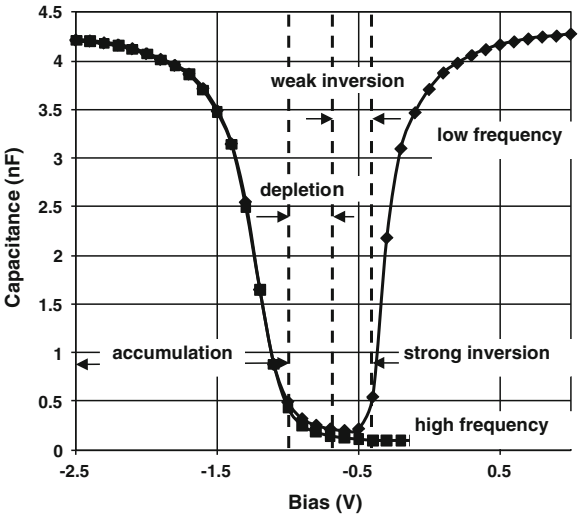


Fig. 2.9 *Low* and *high frequency* capacitance–voltage (C–V) characteristics of a p-Si/SiO_2/Al MOS capacitor. The dry thermal SiO_2 layer was 5.2 nm thick

Figure 2.9 indicates that the capacitance is a weaker function of the bias in depletion and in weak inversion, and is a strong function of the bias in the early part of the accumulation and the early part of the strong inversion regimes; whereas the capacitance tends to saturate in the later part of accumulation and strong inversion. We will now proceed to qualitatively explain these features of the C–V characteristics. In this context, it may be noted that the C–V characteristics in Fig. 2.9 reflect a very low interface trap density, which means a small influence of the parameters C_{it} and Q_{it} on both the total C and the total V, cf. Figs. 2.3 and 2.4 and (2.11) and (2.12). Equations (2.13) and (2.14) show that in depletion and in weak inversion, both the space charge capacitance density C_{sc} and the space charge density Q_{sc} are parabolic (i.e. weak) functions of the surface potential. This is the main reason behind the slow variation of C with V in depletion and weak inversion.

For a p-type semiconductor in accumulation, φ_s is <0; hence $\exp(-\beta\varphi_s) \gg |\beta\varphi_s| \gg 1 \gg \exp(\beta\varphi_s)$. Under these conditions, (2.11) and (2.12) approximate to:

$$Q_{sc} \approx \sqrt{\frac{2q\varepsilon_s N_A}{\beta}} \exp -\left(\frac{\beta\varphi_s}{2}\right); \quad C_{sc} \approx \sqrt{\frac{q\varepsilon_s N_A \beta}{2}} \exp -\left(\frac{\beta\varphi_s}{2}\right) \tag{2.17}$$

Equation (2.17) indicates the space charge density, and therefore, the space charge capacitance density to be a strong exponential function of the surface potential in accumulation (It may be noted that $\beta\varphi_s$ is $\gg 1$). Even when the Fermi occupancy is used instead of the Boltzmann distribution and carrier confinement is considered, Q_{sc} and C_{sc} are still strong functions of the surface potential in accumulation. The mathematical relations for and the functional form of Q_{sc} and C_{sc} in accumulation and strong inversion will be discussed in a later section. The strong increase of C_{sc} as an accumulation layer grows is the reason why the capacitance C rises rapidly after accumulation sets in. As C_{sc} is in series with C_{di}, cf. Figs. 2.3 and 2.4, the total MOS capacitance tends to saturate to C_{di}, once C_{sc} becomes $\gg C_{di}$.

Once strong inversion sets in, for a p-type semiconductor, φ_s is > 0 and also many times β^{-1}; hence $\exp(\beta\varphi_s) \gg |\beta\varphi_s| \gg 1 \gg \exp(-\beta\varphi_s)$. Under these conditions, (2.11) and (2.12) approximate to:

$$Q_{sc} \approx -\sqrt{\frac{2q\varepsilon_s n_0}{\beta}} \exp\left(\frac{\beta\varphi_s}{2}\right); \quad C_{sc} \approx \sqrt{\frac{q\varepsilon_s n_0 \beta}{2}} \exp\left(\frac{\beta\varphi_s}{2}\right) \tag{2.18}$$

Equation (2.18) also indicates the space charge density, and therefore, the space charge capacitance density to be a strong exponential function of the surface potential in strong inversion (It may be noted that $\beta\varphi_s$ is $\gg 1$). Even when the Fermi occupancy is not approximated by the Boltzmann distribution and the carrier confinement is considered, Q_{sc} and C_{sc} are still strong functions of the surface potential in strong inversion. The strong increase of C_{sc} as a strong inversion layer grows is the reason why the capacitance C rises rapidly, once strong inversion sets in. As C_{sc} is in series with C_{di}, cf. Figs. 2.3 and 2.4, the total MOS capacitance tends to saturate to C_{di}, once C_{sc} becomes $\gg C_{di}$. The saturation of the MOS capacitance to the total gate dielectric capacitance C_{di} allows it to be extracted, and therefore also an EOT or Capacitive Equivalent Thickness (CET), from the measured C–V characteristic.

The MOS capacitance rises rapidly not only in strong inversion but also just before and at the onset of strong inversion, cf. Fig. 2.9. It may be noted that at the onset of strong inversion, for a p-type semiconductor, φ_s is > 0 and also many times β^{-1}; hence $\exp(\beta\varphi_s) \gg |\beta\varphi_s| \gg 1 \gg \exp(-\beta\varphi_s)$, further, we have $n_0\exp(\beta\varphi_s)/p_0 = 1$, since $n_0\exp(\beta\varphi_s) = n_s = p_0$. Under these conditions, (2.11) and (2.12) approximate to:

$$Q_{sc} \approx -\sqrt{2q\varepsilon_s N_A \varphi_s}; \quad C_{sc} \approx \sqrt{\frac{2q\varepsilon_s N_A}{\varphi_s}} \tag{2.19}$$

It is important to note that (2.19) yields a value for C_{sc} that is twice of what the depletion approximation, cf. (2.14), would yield. Depletion approximation ignores the contribution by the minority carriers. As the onset of strong inversion approaches, the minority carrier density approaches the doping density, and therefore, C_{sc} begins to increase. It may be noted that the MOS C–V has an asymmetry, and this asymmetry signifies whether the semiconductor is p- or n-type.

It is observed in Fig. 2.9, that the high frequency MOS capacitance saturates to a minimum capacitance in strong inversion. We now proceed to understand this phenomenon. At a high frequency, by the virtue of its definition, there is no contribution from the minority carriers to the space charge capacitance density C_{sc}. Once, a strong inversion layer forms and keeps growing, the surface potential φ_s changes very slowly with the bias V, since the space charge density Q_{sc} becomes very large, making the potential across the gate dielectric $\delta V_{di} \gg \delta\varphi_s$, cf. (2.9). A near-constant φ_s makes C_{sc} saturate, cf. (2.14), which in series with C_{di}, renders a nearly constant minimum C in strong inversion. This minimum capacitance allows the doping density to be extracted from its measured value.

Perhaps, the most important parameter that the MOS C–V characteristics reflect is the interface trap density and its attendant effects. As Figs. 2.3 and 2.4 suggest, the difference between the low and the high frequency capacitance should directly yield the interface trap capacitance C_{it}, hence the interface trap density directly. Hence, the low frequency capacitance is taken to be a very useful and reliable indicator of the quality of the MOS capacitor and the MOSFET. The characteristics of Fig. 2.9 show no difference between C_{lf} and C_{hf} in accumulation and a only small difference in depletion and weak inversion, thereby demonstrating that this is indeed a high quality capacitor with a very low interface trap density. The mid-gap trap density obtained for this MOS structure from the MOS conductance data was about 2×10^{10} cm^{-2} V^{-1}.

2.5 Drain Current–Voltage Characteristics of MOSFET with SiO_2 Gate Dielectric

Our aim is to obtain a closed-form mathematical relation for the drain current I_D as a function of the drain voltage V_D with the gate voltage V_G as an important parameter, to gain an insight into the physical operation and behavior of the MOSFET. Application of the drain voltage along the channel, in the y direction, and of the gate voltage perpendicular to the channel and the semiconductor/insulator interface, in the x direction, cf. Fig. 2.10, requires a two-dimensional analysis, which cannot yield a closed-form solution.

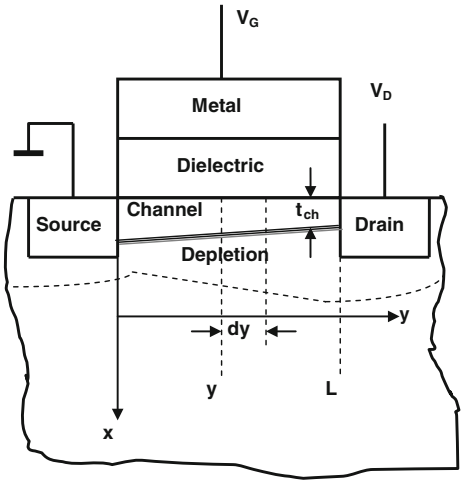


Fig. 2.10 Idealized representation of the basic features of an MOSFET structure. The physical dimensions are disproportionate to allow emphasis on such elements of the MOSFET as the channel and the gate dielectric, both of which to a large extent dictate the MOSFET performance. Also, the contours of such elements as the channel and the depletion layer are only indicative. This representation assumes that the drain voltage $V_D <$ the gate voltage V_G

2.5.1 Ideal MOSFET: Linear Regime

Let us begin with a very simple one-dimensional analysis, which will provide some basic but important understanding of the device. Let us also characterize an ideal MOSFET gate stack with which we could later compare an MOSFET in reality and analyze the degradation of the channel parameters by the non-ideal factors. The following is a list of the non-ideal factors and we assume these to be absent in the ideal MOSFET gate stack:

1. Zero charges of any kind in the bulk and the interfaces of the gate insulator;
2. Zero semiconductor/metal work function difference;
3. Saturating inversion surface potential, i.e. the surface potential remains frozen at its value at the onset of the strong inversion regime and does no longer change with the gate voltage;
4. Low drain voltage compared to the gate voltage;
5. The charge of ionized dopants is zero.

In other words, in an ideal MOSFET, the entire gate voltage (100 % of it) is utilized in generating the channel charge Q_{ch} (consisting only of electrons in an n channel) and none of it is wasted on the non-ideal factors. The drain current is directly proportional to the channel charge Q_{ch}; if parts of the gate voltage are wasted on the non-ideal factors and are therefore not available to generate Q_{ch}, then the drain current is reduced proportionately—this results in the degradation of the channel parameters by the non-ideal factors.

In an ideal MOSFET, under strong inversion, the source-channel-drain becomes a homogeneous semiconductor (with no junctions), allowing the majority carriers to flow by drift; hence the drain current is simply the drain voltage times the

channel conductance. For an n-channel MOSFET (i.e. p-type substrate), we could write, cf. Fig. 2.10:

$$\begin{aligned} I_D &= g_d V_D = \frac{Wt_{ch}}{L}\sigma_{ch}V_D = \frac{Wt_{ch}}{L}qn\mu_{ch}V_D = \frac{W}{L}|Q_{ch}|\mu_{ch}V_D \\ &= \frac{W}{L}C_{di}(V_G - V_T)\mu_{ch}V_D = \beta(V_G - V_T)V_D, \text{ where } \beta = \frac{\mathrm{W}}{\mathrm{L}}C_{di}\mu_{ch} \end{aligned} \tag{2.20}$$

g_d is the channel conductance, σ_{ch} is the channel conductivity, L is the channel length, cf. Fig. 2.10, W is the channel width, μ_{ch} is the channel mobility, Q_{ch} is the channel charge density, and β is the MOSFET quality factor. In deriving the above relation, we have assumed that the channel thickness t_{ch} is constant (due to low V_D) and is not a function of y or V_D, i.e. the channel conductance is a geometric one and is given by the channel area perpendicular to the y direction, Wt_{ch}, the channel length L, and the channel conductivity σ_{ch}. We have assumed that the channel charge is dominated by the electrons; the channel charge density $Q_{ch} = qnt_{ch}$.

Since we had assumed zero bulk and interface charges for the gate dielectric, $Q_{it} = 0$ and $Q_F = 0$. Hence, $V_G = Q_M/C_{di} = -(Q_{sc} + Q_{it} + Q_F)/C_{di} = -Q_{sc}/C_{di}$. We assume that at the onset of strong inversion, i.e. when $V_G = V_T$, the space charge (i.e. basically the depletion charge) is negligible [assumption (5) above]. Therefore, any additional gate voltage, $(V_G - V_T)$ results in the creation of the channel charge, since the surface potential is frozen at its value at the onset of strong inversion, $\varphi_{s,inv,th}$. Hence, $|Q_{ch}| = C_{di}(V_G - V_T)$. In strong inversion, the channel charge density Q_{ch} is predominantly that of the electrons. We must bear in mind that the relations in (2.20) are valid for very low drain voltages. The channel conductance g_d and the transconductance g_m can be expressed as:

$$g_D = \frac{\partial I_D}{\partial V_D} = \beta(V_G - V_T); \quad g_m = \frac{\partial I_D}{\partial V_G} = \beta V_D \tag{2.21}$$

There are some remarkable features of the relations in (2.20) and (2.21). These relations are very simple with a very simple derivation, still provide an insight into and understanding of the basic operation and performance of the MOSFET. The main significance of (2.20) could be summed as:

1. For low drain voltages (i.e. $V_D \ll V_G$), the $I_D(V_D)$ characteristic is linear. In other words, the ideal $I_D(V_D)$ relation is linear; any deviation from this linearity is a manifestation of the deviation from the ideal state and of degradation. We will see later how each of the five assumptions, made in deriving this relation, degrades and makes the drain current–voltage characteristic to droop and deviate from linearity.
2. Each of the five assumptions is an imperfection (a non-ideal factor) and causes the drain current to attenuate from its ideal value.
3. Ideally, the entire gate voltage, i.e. all of V_G, should be utilized for modulating the channel charge Q_{ch}. However, in a real MOSFET, each of the five non-ideal factors consumes a large part of the gate voltage.

4. Among the most important MOSFET performance parameters (Which directly influence the drain current, the switching speed, the channel conductance, and the transconductance.) are: (a) the channel mobility μ_{ch}; and (b) the gate dielectric capacitance density C_{di}.
5. The slope $(\partial I_D/\partial V_D)$ of the $I_D(V_D)$ characteristic can be used for the extraction of the important device parameters.

Equation (2.20) and the above remarks illustrate why in the case of the single gate dielectric (SiO_2 or SiNO), the technological trend was to reduce the gate dielectric thickness t_{di}, thereby increasing C_{di}. In the case of the high-k gate stack, the trend is both to increase the gate stack permittivity (by choosing insulators such as HfO_2, La_2O_3, thereby increasing C_{di}, and to enhance the channel mobility (by choosing semiconductors as Ge, GaAs, graphene).

2.5.2 Classical Model

We will now proceed to derive a more general and less ideal $I_D(V_D)$ relation, for which the first three assumptions of Sect. 2.5.1 will be retained, but the last two assumptions will be removed. The following mathematical treatment is generally credited to [1, 11, 12]. As soon as the drain voltage is applied, the uniformity of the channel disappears, i.e. the channel thickness t_{ch} becomes a function of the drain voltage, i.e. the channel can no longer be represented by a geometric conductance, but, becomes a V_D-dependent entity. This is what makes the $I_D(V_D)$ relation non-linear (As V_D increases in comparison to V_G, the $I_D(V_D)$ relation turns from linear to non-linear, and finally to saturation.), when V_D becomes significant. With the application of the drain voltage V_D, the voltage along the y direction, V(y), increases from zero at the source (y = 0) to V_D at the drain (y = L), cf. Fig. 2.10. Consequently, the channel thickness $t_{ch}(y)$ is maximum at the source and minimum at the drain. At any point y, the voltage across the gate insulator is $[V_G - \varphi_s(y)]$. When $V_D = 0$, φ_s is not a function of y. But, for a non-zero V_D, φ_s is a function of y; such that at any point y, $\varphi_s(y) = \varphi_s(y = 0) + V(y)$, cf. Fig. 2.10.

It may be noted that for a non-zero drain voltage, the channel is in thermal non-equilibrium; consequently, the electron imref (i.e. the quasi-Fermi level) separates from the hole imref, as illustrated in Fig. 2.11. After a drain voltage has been applied, the voltage V(y) equivalent energy, qV(y), at a point y in the channel, must appear between the majority carrier imref in the neutral substrate and the majority carrier imref in the channel, as illustrated in Fig. 2.11. To maintain the channel and the strong inversion condition at any point y, for a p-Si substrate, the conduction band edge E_c has to maintain a certain minimum energy from the electron imref. Also, the surface potential at y for the strong inversion regime, $\varphi_{s,inv}(y)$, will exceed the surface potential at the source, $\varphi_{s,inv}(y = 0)$, by V(y):

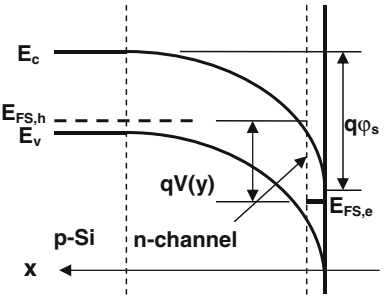


Fig. 2.11 Energy band profile of the semiconductor space charge region along the x-axis for the gate voltage $V_G >$ the threshold voltage V_T (a channel exists at the semiconductor-dielectric interface), at the point y along the y direction (i.e. the direction of V_D) for a non-zero V_D. $E_{FS,h}$ is the hole (majority carrier) imref in the p-Si neutral region, while $E_{FS,e}$ is the electron imref in the channel. Note that electrons are minority carriers in the neutral p-Si region, but are majority carriers in the channel, i.e. after the p-Si substrate has been inverted

$\varphi_s(y) = \varphi_s(y = 0) + V(y)$. Therefore, the voltage across the gate dielectric $V_{di} = [V_G - \varphi_s(y = 0) - V(y)] = -Q_{sc}/C_{di}$. In other words, the voltage across the gate dielectric is maximum at the source and minimum at the drain, leading to a maximum channel charge and channel width at the source and a minimum channel charge and channel width at the drain. This makes the channel a non-geometric conductor, requiring a formulation of the current–voltage relation, different from the simple approach of Sect. 2.5.1.

As the channel thickness t_{ch} is now a function of y, we express the channel conductance Δg_d of an infinitesimal channel length dy, cf. Fig. 2.10, in the following manner:

$$\Delta g_d = \left(\frac{W}{dy}\right) \int_0^{t_{ch}} \sigma(x) dx, \quad \text{where } \sigma(x) = qn(x)\mu_e(x)$$

It may be noted that the electron density in an n-channel is a function of x, because of the potential $\varphi(x)$; hence the channel conductivity is also a function of x. The above relation may be rearranged as:

$$\Delta g_d = \left(\frac{W}{dy}\right) \mu_{ch} \int_0^{t_c} qn(x) dx = \left(\frac{W}{dy}\right) \mu_{ch} |Q_{ch}|, \text{ where } |Q_{ch}| = q \int_0^{t_c} n(x) dx$$

It may be noted that, in principle, both the channel carrier density n as well as the channel carrier mobility μ are functions of both x and y. (As will be explained later, we assume an effective channel mobility, to simplify the treatment).

The voltage across the channel element of length dy, dV(y), can be represented as;

$$dV(y) = \frac{I_D}{\Delta g_d} = \frac{I_D}{W\mu_{ch}} \frac{dy}{|Q_{ch}|}$$

The above relation can be rearranged as:

$$\int_0^{V_D} |Q_{ch}| dV(y) = \left(\frac{I_D}{W\mu_{ch}}\right) \int_0^{L} dy = I_D L/(W\mu_{ch})$$

Hence, the drain current I_D can be expressed as:

$$I_D = \frac{W}{L}\mu_{ch} \int_0^{V_D} |Q_{ch}| dV(y)$$

In Sect. 2.5.1, the charge of the ionized dopants was ignored [i.e. assumption (5)]. We will now formulate the channel charge without neglecting the space charge due to the ionized dopants.

$$Since\, V_G = \varphi_s + V_{di} = \varphi_s - \frac{Q_{sc}}{C_{di}}, \text{ we have: } Q_{sc} = -C_{di}(V_G - \varphi_s)$$

The channel charge density Q_{ch} may be equated to the total space charge density minus the charge density of the ionized dopants, for which, we may use the notation Q_{dep}; to express the latter, we may use the depletion approximation, cf. Sect. 2.2.4 and (2.13).

$$Q_{ch} = Q_{sc} - Q_{dep} = -C_{di}(V_G - \varphi_s) + \sqrt{2q\varepsilon_s N_A \varphi_s} \tag{2.22}$$

It may be noted that for a p-type semiconductor, the depletion charge density Q_{dep} is negative, being due to the ionized acceptors. As the value of φ_s will be minimum at the source and maximum at the drain, the values of Q_{dep} will be the same. If we use the notation $\varphi_s(y = 0)$ for the surface potential at the source, and the relation $\varphi_s(y) = \varphi_s(y = 0) + V(y)$, then the drain current may be expressed as:

$$I_D = \frac{W}{L}\mu_{ch} \int_0^{V_D} \left[\{C_{di}(V_G - \varphi_s(y=0) - V(y)\} - \sqrt{2q\varepsilon_s N_A[\varphi_s(y=0) + V(y)]}\right] dV(y)$$

Upon integration, we obtain a closed-form relation for the drain current:

$$I_D = \frac{W}{L}\mu_{ch} C_{di}\left[\left(V_G - \varphi_s(y=0) - \frac{V_D}{2}\right)V_D - \frac{2}{3}\gamma\left[(V_D + \varphi_s(y=0))^{\frac{3}{2}} - (\varphi_s(y=0))^{\frac{3}{2}}\right]\right] \tag{2.23}$$

$$\gamma = \frac{\sqrt{2q\varepsilon_s N_A}}{C_{di}} = \frac{Q_{dep}}{C_{di}\varphi_s^{1/2}} \tag{2.24}$$

Since we have assumed (assumption 3 in Sect. 2.5.1) that the surface potential remains frozen at its value at the onset of strong inversion, $\varphi_s(y = 0) = \varphi_{s,inv,th} = (E_G/q) - 2\phi_p$.

As illustrated below, the relation in (2.23) reduces to the one in (2.20), for small values of V_D, i.e. $V_D \ll V_G$. It may be noted that the relation for the threshold voltage in (2.16) simplifies to what is written below for the assumptions made in Sect. 2.5.1, namely: $\phi_{MS} = 0$; $Q_{it} = 0$; $Q_F = 0$.

$$\begin{aligned} I_D &\approx \beta\left[\left(V_G - \varphi_{s,inv,th} - \frac{V_D}{2}\right)V_D - \frac{2}{3}\gamma\left(\varphi_{s,inv,th}^{3/2} + \frac{3}{2}\frac{V_D}{\varphi_{s,inv,th}}\varphi_{s,inv,th}^{3/2} - \varphi_{s,inv,th}^{3/2}\right)\right] \\ &\approx \beta\left(V_G - \varphi_{s,inv,th} - \gamma\sqrt{\varphi_{s,inv,th}}\right)V_D \approx \beta(V_G - V_T)V_D,\ V_T = \varphi_{s,inv,th} + \gamma\sqrt{\varphi_{s,inv,th}} \end{aligned}$$

When the drain voltage V_D is no longer small compared to the gate voltage V_G, the third term within the first parenthesis of (2.23), namely $V_D/2$, gains weight; consequently, the drain current does not increase rapidly with V_D, and begins to droop. The bracket in (2.23), which is multiplied by γ, is net positive and is to a lesser extent also responsible for the drain current deviating from a linear increase with the drain voltage and finally tending to saturate. It may be noted that the derivation of (2.23) is based upon the assumption that the channel exists in the entire region between the source and the drain, cf. Fig. 2.10, and that the carriers traverse the channel by drift. As the drain voltage increases, a situation comes about, when the channel disappears at the drain ($y = L$). This condition is known as pinch-off, which occurs first at drain, and with V_D increasing further, the pinch-off point y_p moves towards the source, cf. Fig. 2.12. The pinch-off condition can be defined as: $n_s = N_A$ (for p-type semiconductor) or $Q_{ch} = 0$. The saturation drain voltage, V_{DS}, is the drain voltage for which the pinch-off point $y_p = L$, and the corresponding drain current is the saturation drain current, I_{DS}. An expression for V_{DS} can be obtained by using (2.22) and the condition $Q_{ch}(y = L) = 0$:

$$\begin{aligned} Q_{ch} &= -\left(V_G - V_{DS} - \varphi_{s,inv,th}\right)C_{di} + \gamma C_{di}\sqrt{V_{DS} + \varphi_{s,inv,th}} = 0 \\ \Rightarrow V_{DS} &= V_G - \varphi_{s,inv,th} + \frac{\gamma^2}{2}\left[1 - \left(1 + \frac{4V_G}{\gamma^2}\right)^{\frac{1}{2}}\right] \end{aligned} \tag{2.25}$$

For ultrathin gate insulators, $\gamma \ll 1$, cf. (2.24). For example, for EOT = 1 nm and $N_A = 10^{17}$ cm^{-3}, $\gamma = 0.02\sqrt{\text{V}}$. Equation (2.25) approximates, for $\gamma \ll 1$, to:

$$V_{DS} \approx V_G - V_T', \quad \text{where } V_T' = \varphi_{s,inv,th} + \gamma\sqrt{V_G} \tag{2.26}$$

Substitution of the relation for V_{DS} in (2.26) in (2.23), yields a relation for the saturation drain current I_{DS} versus the saturation drain voltage V_{DS}, for $\gamma \ll 1$:

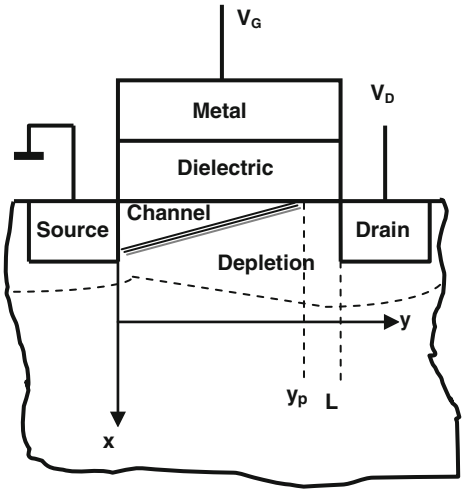


Fig. 2.12 Schematic representation of an MOSFET, illustrating the onset of the saturation regime ($V_D > V_{DS}$), the pinch-off point y_p, and the attendant disappearance of the channel in the region between y_p and L

$$I_{DS} \approx \frac{\beta}{2}\left(V_G - V_T'\right)^2 = \frac{\beta}{2}\left(V_{DS}\right)^2 \tag{2.27}$$

In the saturation regime, the transconductance is given by:

$$g_m = \frac{\partial I_{DS}}{\partial V_G} = \beta\left(V_G - V_T'\right) = \beta V_{DS} \tag{2.28}$$

Figure 2.13 presents the experimental drain current versus drain voltage characteristics of NMOSFETs: one with an ultrathin SiO_2 (EOT = 1.84 nm) and another with oxide-nitride (EOT = 1.93 nm) gate dielectric [13]. All the non-ideal

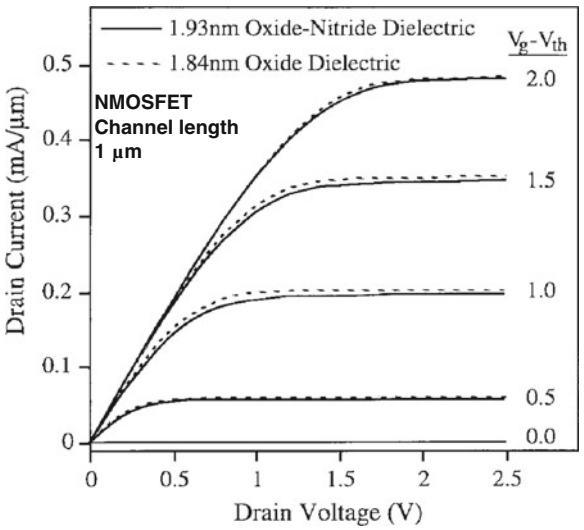


Fig. 2.13 Drain current versus drain voltage of n-channel MOSFET's with ultrathin oxide or oxide-nitride gate dielectric of EOT = 1.84 or 1.93 nm, respectively. The channel length was 1 μm. ($V_G - V_T$) varied between 0 and 2.0 V. Adapted from [13]

factors listed in Sect. 2.5.1 will be present in this MOSFET; however many of these factors will have much lower magnitudes than those in MOSFETs with high-k gate stacks. These measured characteristics of MOSFETs with silicon dioxide and silicon oxide-nitride dielectrics will be compared with those of MOSFETs with high-k gate stacks in Sect. 2.6.1.

2.6 High Dielectric Constant (k) Gate Stacks

In many ways (in physical, chemical, and electrical characteristics), the high permittivity gate dielectrics represent a drastic change, from the SiO_2 gate dielectric:

1. Dry thermal SiO_2 is a near-perfect dielectric, practically with no bulk charges and no space-charge capacitance; in that sense, high k materials are poor dielectrics with enormous charges ($>10^{13}/cm^2$); and high k gate stacks need to be represented by many (perhaps, as many as five) bulk trap and interface trap capacitances.
2. While the Si-SiO_2 interface is a true and marvelous gift of the nature (with an interface state density of the order of $10^{10}/cm^2/V$), the Si/high-k (and more so, Ge/high-k, GaAs/high-k, etc.) interface is a difficult one (with an interface state density $>10^{13}/cm^2/V$) with an uncertain chance of improvement.
3. With a poly-silicon gate electrode, the Si-SiO_2-poly-Si symmetric structure is practically immune to the kind of thermal and chemical stability problems encountered in the case of semiconductor/high-k-gate-stack/metal-electrode structures.
4. While the dry thermal SiO_2 owes its unmatched physical, chemical, and most importantly, electronic properties to its primarily covalent character, in sharp contrast to the high-k materials, which owe their many weaknesses to their primarily ionic character.
5. The covalent character of SiO_2 lends excellent matching with Si, while the ionic character of high-k is at the root of their poor matching with all semiconductors (including Si), which are all almost totally covalent (While Si and Ge are totally covalent, compound semiconductors are necessarily partly ionic).
6. SiO_2 is difficult to crystallize, and is a very stable material thermally. The opposite is true of the high-k materials, which crystallize easily much below the implant activation temperature.
7. Dry thermal silicon dioxide has the highest band-gap (of about 9 eV) among the inorganic solids, while the high-k materials have only moderate to high (say, 4–6 eV) band-gaps.
8. Dry thermal SiO_2 has the highest electrical resistivity (of the order of 10^{23} Ω cm) ever recorded (The resistivity is purely electronic; there is no ionic contribution); the ionicity of the high-k materials lends them a conductivity, much higher than what their electronic or band-gap alone would suggest.

9. Dry thermal SiO_2 has perhaps the lowest dielectric constant (The electrical polarization is mainly electronic, with a very low ionic contribution.) among the inorganic oxides, whereas that of the high-k materials can be many times higher.

Succinctly expressed, everything would seem to be wonderful of the dry thermal SiO_2, except its abysmally low dielectric constant, whereas it is difficult to find a near-perfect property of the high-k gate dielectrics, except their high permittivity. It is hard to imagine how just one handicap undoes the dry thermal SiO_2, in spite of its truly amazing qualities, when gate dielectrics with sub-1-nm EOT are required. Fortunately, the high permittivity of the high-k gate stacks mitigates some of their problems:

1. Perhaps, the most effective among these relate to the huge reduction in the electric field across the different layers of the gate stack due to the high permittivity. It may be noted that a much lower electric field will induce the same inversion layer charge, as $Q_{ch} = \varepsilon_{di} E_i$.
2. This, in turn, is responsible for the moderate potentials across the high-k gate stack layers, although, the high-k bulk and interface trap charge densities may be very large. Consequently, it has become possible to maintain the trend of reducing the threshold and the supply voltages.
3. The low electric fields may also promote gate stack reliability, even in the presence of huge charges inside.

We will discuss various aspects of the high-k gate stacks (physical structure and representation of the various layers, chemical nature, energy band profiles, nature and representation of the various bulk and interface traps, electrostatic analysis, circuit representations) in the later sections. First, we examine the implications of the high-k gate stack properties on the channel's electrical (drain current–voltage, etc.) characteristics.

2.6.1 Drain Current–Voltage Characteristics of MOSFETs with High-k Gate Stacks

We will now reexamine the validity of all the assumptions, made while deriving the classical $I_D(V_D)$ relation of (2.23), in the case of high-k gate stacks, and also examine how the drain current–voltage relation can be obtained in the closed form for MOSFETs with high-k gate stacks. The assumptions of Sect. 2.5.2 for the classical model were:

1. No semiconductor/metal work function difference, i.e. $\phi_{MS} = 0$.
2. No charges in the bulk of the gate stack.
3. No fixed oxide or interface state charges, i.e. $Q_{it} = 0$ and $Q_F = 0$.

4. Once strong inversion sets in, the inversion surface potential at the source remains frozen at the value $\varphi_{s,inv,th}$ and no longer increases with the gate voltage.

None of the above assumptions is tenable in the case of MOSFETs with high-k gate stacks, which are beset with enormous bulk charges and traps and work-function anomaly [14, 15]. After strong inversion sets in, the surface potential far from saturating has been observed to keep increasing continuously over the entire strong inversion regime [16, 17]. Hence the above assumptions taken together represent a serious contradiction to the reality that obtains in the case of the current high-k gate stacks.

Numerical models have been developed to make more realistic estimates of the drain current and the related parameters of MOSFETs. However, closed-form, text-book type equations, derived from the first principles, are needed to provide a sound physical insight into the critical parameters. Needed are mathematical relations in closed form for the drain current I_D, the channel conductance g_D, and the transconductance g_m which transparently illustrate how much the high-k gate stack charges, the non-saturating inversion surface potential, and the work function anomaly degrade the channel parameters. Even the recent semiconductor device and gate stack reference books still present the classical closed-form equations for the MOSFET channel parameters [18–20]. Our aim in this section will be four-fold, namely to: (1) have direct incorporation of the threshold-excess inversion surface potential ($\Delta\varphi_{s,inv} = \varphi_{s,inv} - \varphi_{s,inv,th}$), the total gate-stack charge density $Q_{di,gsc}$, and the semiconductor-metal work-function difference ϕ_{MS} in the equations for the drain current I_D, the channel conductance g_D, and the transconductance g_m; (2) illustrate the scale of degradation of I_D, g_D, and g_m by the non-ideal factors of $\Delta\varphi_{s,inv}$, $Q_{di,gsc}$, and ϕ_{MS}, by quantitative analysis and estimation, using the available experimental data; (3) present text-book-appropriate relations in such a form that, how much each of the adverse factors degrades each of the channel parameters, becomes visible in a glance.

To derive the mathematical relations for the channel parameters, we will use the following general formulation for the drain current arrived at in Sect. 2.5.2:

$$I_D = \frac{W}{L}\mu_{ch}\int_0^{V_D} |Q_{ch}|dV(y) \tag{2.29}$$

The challenge is to find an expression for the channel charge $Q_{ch}(y)$ which would be valid in the case of the high-k gate stacks; in particular, the main problem is finding a realistic mathematical representation for the gate stack charge $Q_{di,gsc}(y)$, and the corresponding gate stack potential $V_{di}(y)$ and the channel charge $Q_{ch}(y)$, which will allow integration of (2.29) into a closed-form.

2.6.1.1 Gate Stack Potential $V_{di}(y)$

The gate stack potential V_{di} originates from the semiconductor space charge density Q_{sc} and all the charges in the gate stack. As Fig. 2.14 illustrates, the gate stack potential has many components (The energy band diagram of Fig. 2.14 will be discussed in detail in Sect. 2.6.3. It reflects experimental values of the surface potential and gate stack potentials and gate stack trap densities obtained from admittance-voltage-frequency and flat-band voltage versus EOT measurements over many MOS structures with varying EOT [17]. Figure 2.14 illustrates the complicated nature of the charge-potential relation inside a high-k gate stack). Unfortunately, at present, there is scarce information on the nature, location, and distribution of the charges in the high-k gate stacks [14–17]. Therefore, it is not possible to express realistically the potentials across the different layers and interfaces of the high-k gate stack, e.g. V_{diIL}, $V_{di,dipole}$, in mathematical form as a function of y. We outline below an option for tackling this obstacle. We write down the total gate stack voltage as a sum of two components, one which is a strong function of y and the other which is relatively invariant of y.

$$V_{di}(y) = V_{di,sc}(y) + V_{di,gsc} \tag{2.30}$$

$V_{di,sc}$ is the total potential across the entire gate stack due to the semiconductor space charge density Q_{sc}, whereas $V_{di,gsc}$ is the total potential across the entire gate stack due to all the charges in the layers and the interfaces of the high-k gate stack. $V_{di,sc}(y)$ is a strong function of y, because of the strong variation in the surface potential $\varphi_s(y)$ in the y-direction, cf. Sect. 2.5.2 and (2.31), causing the space charge density $Q_{sc}(y)$, which is a function of φ_s, to vary strongly.

$$\varphi_{s,inv}(y) = \varphi_{s,inv}(y=0) + V(y) = \varphi_{s,inv,0,th} + \Delta\varphi_{s,inv,0}(V_G) + V(y) \tag{2.31}$$

On the other hand, the bulk and the interface trap charges in the gate stack may not be a significant variant in the y-direction. The charge in traps inside the gate stack is decided by the occupancy of these traps; and the trap occupancy is decided by the pseudo-Fermi level inside the gate stack, cf. Fig. 2.14. The pseudo-Fermi level at a plane x inside the gate stack indicates the occupancy of traps at that plane and whether the traps at that plane is communicating more readily with the semiconductor surface, i.e. is controlled and given by $E_{FS,h}$, or more readily with the metal surface, i.e. is controlled and given by E_{FM}, or is communicating with neither, cf. Fig. 2.14 [12, 16]. To analyze their variation with y, the gate stack traps can be classified into three groups:

Group 1 The charge in traps at the metal surface and the charge in the traps inside the gate stack which exchange electrons more readily with the metal surface will not be a function of y, since this charge is controlled by the metal Fermi level E_{FM}, which is not a function of y, the potential on the metal surface being the same everywhere, cf. Fig. 2.14.

Group 2 There may be traps deep inside the gate stack which are unable to exchange electrons/holes either with the Si or the metal surface. Moreover, there

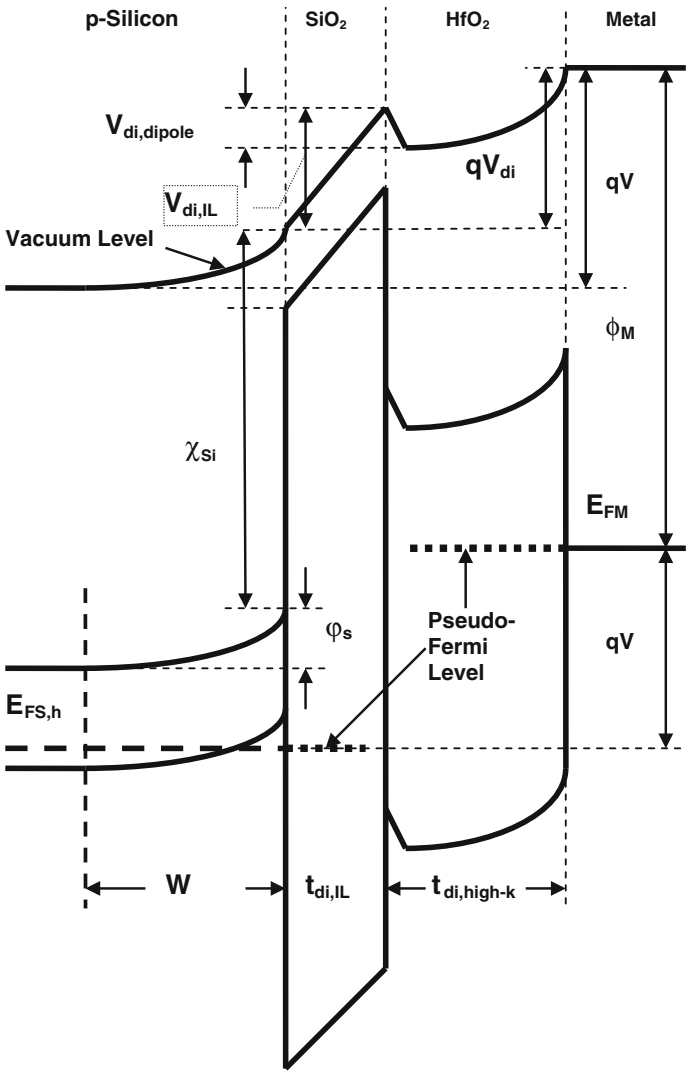


Fig. 2.14 A schematic representation of the energy profiles across a p-silicon/SiO_2/.HfO_2/TaN MOS capacitor in strong accumulation, under a bias of −1.82 V. The SiO_2 layer was about 1 nm and the HfO_2 layer about 2 nm thick, while the EOT was about 1.9 nm. $V_{di,dipole}$ is the potential drop across the dipole at the IL/high-k interface. $E_{FS,h}$ is the hole (majority carrier) imref. $t_{di,IL}$, $t_{di,high-k}$ are the thickness of the IL and the high-k layer, respectively. The pseudo-Fermi function is supposed to indicate the occupancy of the traps in the gate stack

may be traps in the gate stack whose energy levels remain much higher or much lower than the Si quasi-Fermi level $E_{FS,h}$, cf. Fig. 2.14. Effectively, these traps will act as fixed charges and will remain invariant of potential and therefore also of y.

Group 3 The charge in traps at the Si surface and the charge in traps inside the gate stack which exchange electrons more readily with the Si surface will be controlled by the Si quasi-Fermi level $E_{FS,h}$ the variation of this charge with y will depend upon the variation of $E_{FS,h}$ with y, cf. Figs. 2.11 and 2.14. The variation in the surface potential $\varphi_{s,inv}$ will be $= V_D$ between the source and the drain according to (2.31). However, the variation of $E_{FS,h}$ with y will be a small fraction of eV_D, as for a channel (i.e. strong inversion) to exist, $E_{FS,h}$ has to remain close to E_c, cf. Fig. 2.11. Two other factors need to be considered to analyze how much the charge in group 3 traps will change with y. All experimental evidence consistently suggests that the trap density is lowest at the Si/IL interface, and the S/IL interface trap charge is a very small fraction of the total gate stack charges [14, 15, 17]. The magnitude of charge exchange between the Si surface and the traps inside the gate stack will be determined by the electron density at the Si surface, the wave function attenuation (depends on the conduction band offset, the tunneling electron mass, and the distance between the trap location and the Si surface), the trap capture/emission cross-section, and the frequency, cf. Fig. 2.14 [17], and may be a small fraction of the total gate stack charge.

Hence the charge of only the group 3 gate stack traps may vary with y and the error may be small if we ignore the variation with y of the charge in the group 3 traps, and consider $V_{di,gsc}$ to be invariant of y to enable its integration with respect to V(y) into a closed-form expression. Combination of (2.4), (2.30), and (2.31) would then yield:

$$V_{di,sc}(y) = V_G - \varphi_{s,inv}(y=0) - V(y) - V_{di,gsc} - \phi_{MS,p} = -\frac{Q_{sc}(y)}{C_{di}} = -\frac{Q_{ch}(y) + Q_{dep}(y)}{C_{di}} \tag{2.32}$$

or:

$$Q_{ch}(y) = -C_{di}\left[V_G - \varphi_{s,inv}(y=0) - V(y) - V_{di,gsc} - \phi_{MS,p}\right] - Q_{dep}(y) \tag{2.33}$$

In (2.33), only V(y) and $Q_{dep}(y)$ are functions of y. The work function difference (anomaly) may be considered invariant of y.

2.6.1.2 Drain Current Versus Drain Voltage Characteristic

Substitution of the expression for the channel charge density in (2.33) into the equation for the drain current in (2.29) yields:

$$I_D = \frac{W}{L}\mu_{ch}\int_0^{V_D}\left[\left\{C_{di}\left[V_G - \varphi_{s,inv,0} - V(y) - V_{di,gsc} - \phi_{MS,p}\right]\right\} + Q_{dep}(y)\right]dV(y)$$
$$= \frac{W}{L}\mu_{ch}\int_0^{V_D}\left[\left\{C_{di}\left[V_G - \varphi_{s,inv,0} - V(y) - V_{di,gsc} - \phi_{MS,p}\right]\right\} - \sqrt{2q\varepsilon_s N_A \varphi_s(y)}\right]dV(y)$$
$$= \frac{W}{L}\mu_{ch}\int_0^{V_D}\left[\left\{C_{di}\left[V_G - \varphi_{s,inv,0} - V(y) - V_{di,gsc} - \phi_{MS,p}\right]\right\} - \sqrt{2q\varepsilon_s N_A\left(\varphi_{s,inv,0} + V(y)\right)}\right]dV(y) \tag{2.34}$$

$\varphi_{s,inv,0}$ is the inversion surface potential at y = 0, $\varphi_{s,inv}(y = 0)$. In (2.34), the classical depletion approximation relation has been used for Q_{dep} [3]. Even though all the non-ideal factors present in the high-k gate stack have been considered in (2.34), we have been able to represent these factors in such a manner that (2.34) can be integrated into a closed form to yield the following mathematical relation for the drain current:

$$I_D = \frac{W}{L}\mu_{ch}C_{di}\left[\begin{array}{c}\left(V_G - \varphi_{s,inv,th,0} - \Delta\varphi_{s,inv,0} - V_{di,gsc} - \phi_{MS,p} - \frac{V_D}{2}\right)V_D \\ -\frac{2}{3}\frac{\sqrt{2q\varepsilon_s N_A}}{C_{di}}\left\{\begin{array}{c}\left(\varphi_{s,inv,th,0} + \Delta\varphi_{s,inv,0} + V_D\right)^{3/2} \\ -\left(\varphi_{s,inv,th,0} + \Delta\varphi_{s,inv,0}\right)^{3/2}\end{array}\right\}\end{array}\right] \tag{2.35}$$

The classical drain current versus the drain voltage relation, cf. (2.23) which was derived in Sect. 2.5.2 could be expressed as:

$$I_D = \frac{W}{L}\mu_{ch}C_{di}\left[\begin{array}{c}\left(V_G - \varphi_{s,inv,th,0} - \frac{V_D}{2}\right)V_D \\ -\frac{2}{3}\frac{\sqrt{2q\varepsilon_s N_A}}{C_{di}}\left\{\left(\varphi_{s,inv,th,0} + V_D\right)^{3/2} - \left(\varphi_{s,inv,th,0}\right)^{3/2}\right\}\end{array}\right] \tag{2.36}$$

Normalization of (2.35) by (2.36) yields a relation which directly illustrates the effects of the gate stack charges $V_{di,gsc}$, the non-saturating inversion surface potential $\Delta\varphi_{s,inv,0}$, and the work function difference $\phi_{MS,p}$ on the drain current:

$$\frac{I_{D,high-k}}{I_{D,ideal}} = \frac{\mu_{ch,high-k}}{\mu_{ch,ideal}} \times \frac{\left(V_G - \varphi_{s,inv,th,0} - \Delta\varphi_{s,inv,0} - V_{di,gsc} - \phi_{MS,p} - \frac{V_D}{2}\right)V_D - \frac{2}{3}\gamma\left[\left(\varphi_{s,inv,th,0} + \Delta\varphi_{s,inv,0} + V_D\right)^{\frac{3}{2}} - \left(\varphi_{s,inv,th,0} + \Delta\varphi_{s,inv,0}\right)^{\frac{3}{2}}\right]}{\left(V_G - \varphi_{s,inv,th,0} - \frac{V_D}{2}\right)V_D - \frac{2}{3}\gamma\left[\left(\varphi_{s,inv,th,0} + V_D\right)^{\frac{3}{2}} - \left(\varphi_{s,inv,th,0}\right)^{\frac{3}{2}}\right]} \tag{2.37}$$

The following are the important features of the normalized drain current as represented by (2.37). As the channel conductance g_D and the transconductance g_m are directly linked to the drain current I_D, many of these features belong to the latter as well.

Non-Saturating Surface Potential—The surface potential does not saturate in the case of the ultrathin high-k gate stacks because of a number of reasons: (1) As the EOT is decreased, the gate stack dielectric capacitance density C_{di} increases and becomes comparable to the semiconductor space charge capacitance density

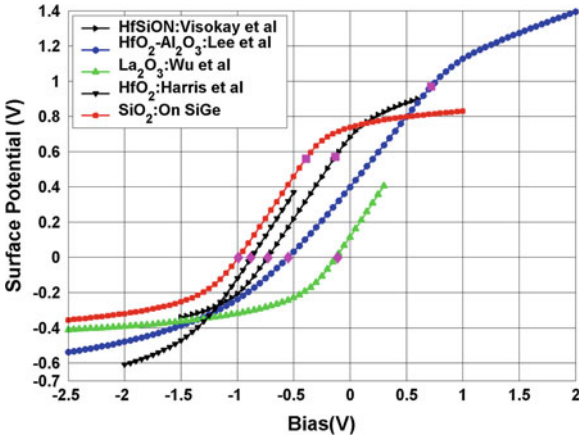


Fig. 2.15 Experimental surface potential versus applied gate bias plot extracted from the Berglund integral [60] of the measured equilibrium capacitance–voltage (C–V) characteristics of MOS structures on p-type silicon with five different gate stacks: HfSiON (*right faced triangle*, *black*), HfO_2-Al_2O_3 (*circle*, *blue*), La_2O_3 (*triangle*, *green*), HfO_2 (*inverted triangle*, *black*), and SiO_2 (*square*, *red*), and EOT values of 2.0, 1.7, 1.4, 0.5, and 3.9 nm respectively. Surface potential in strong inversion could be extracted only for three of the five MOS structures. The C–V data for four of the characteristics were taken from the literature: HfSiON [56], HfO_2-Al_2O_3 [57], La_2O_3 [58], HfO_2 [59]. It maybe noted that the *diamond marker* indicates the flat-band point, while the *square marker* indicates the onset of strong inversion

C_{sc}. Consequently, the surface potential change given by: $\Delta\varphi_{s,inv} = \Delta V_{di}C_{di}/C_{sc}$ becomes larger in comparison to the gate stack potential change ΔV_{di}. (2) Because of quantum-mechanical carrier confinement in the strong inversion layer, C_{sc} is significantly less than its classical value. (3) The very high trap density of high-k gate stacks makes ΔV_{di} larger. The degrading feature of the non-saturating surface potential has not been widely recognized in the literature. $\varphi_{s,inv,th,0}$ is the surface potential at the onset of strong inversion at the source: $\varphi_{s,inv,th,0} = E_G/q - 2\phi_p$ for a p-type semiconductor. In the classical formulation [1, 11, 12], the surface potential at the source was assumed to remain frozen at the value $\varphi_{s,inv,th,0}$ once the channel was formed. In the current high-k transistors, the inversion surface potential $\varphi_{s,inv,0}(V_G)$ has been observed to increase significantly with the gate voltage V_G; experimental data indicate that the surface potential may increase by as much as 0.4 V after strong inversion has set in, i.e. $\Delta\varphi_{s,inv,0} = (\varphi_{s,inv,0} - \varphi_{s,inv,th,0}) = 0.4$ V [21, 22]. Figure 2.15 illustrates the strong variation of the surface potential in both accumulation and in strong inversion for four different high-k gate stacks in comparison with a control SiO_2 gate dielectric. The excess inversion surface potential $\Delta\varphi_{s,inv,0}$ appears twice in the numerator of (2.37), once inside the parenthesis, then within the bracket; consequently $\Delta\varphi_{s,inv,0}$ degrades (i.e. reduces) the drain current twice.

Gate Stack Charges and Traps—Flat-band voltage measurements and results from the conductance technique [17] indicate the charges and the traps in the current device-quality high-k gate stacks to be orders of magnitude higher (net gate stack charge density exceeding $q \times 10^{13}$ cm^{-2}) than what obtained in the case of the dry thermal SiO_2 gate dielectrics. Experimental data [17] indicate the high-k gate stack potential due to the gate stack charges $V_{di,gsc}(V_G)$ to be significant (hundreds of mV). $V_{di,gsc}(V_G)$ will reduce the drain current below its ideal value, cf. (2.37), and this reduction will increase if the gate stack charge increases, e.g. due to gate stack degradation [23, 24].

Work Function Difference Anomaly—The work-function difference term is absent in the classical drain current relation, cf. (2.36), (2.37). Unfortunately, the high-k gate stacks suffer from a serious work-function anomaly and lack of control [25, 26]; consequently the desired work function cannot be realized, and the work-function difference remains not only significant (a few hundred mV) but also changes with subsequent processing. A significant ϕ_{MS} will decrease the drain current below its ideal value and its lack of control introduces drain current instability, cf. (2.35), (2.37).

2.6.1.3 Estimate of Drain Current Degradation

All the three factors discussed above degrade the high-k gate stack drain current $I_{D,high\text{-}k}$ to a fraction of its ideal value, $I_{D,ideal}$, cf. (2.37). It may be noted that the values of the parameters V_G, $\varphi_{s,inv,th,0}$, V_D, and γ in (2.37) and (2.24) may be obtained from the device design, whereas the values of the parameters $\Delta\varphi_{s,inv,0} = (\varphi_{s,inv,0} - \varphi_{s,inv,th,0})$, $V_{di,gsc}$, and ϕ_{MS} may be extracted from carefully planned experiments on the MOS device. It may be instructive to make a rough estimate of the normalized drain current in (2.37), and obtain a feel for the importance of the various degrading factors. The body effect parameter γ is proportional to EOT and to $(N_A)^{1/2}$; choosing EOT = 1.0 nm and $N_A = 3.55 \times 10^{17}$ cm^{-3} yields a round figure of 0.1 for γ. We may choose: a gate voltage $V_G = 2.2$ V; a drain voltage $V_D = 0.4$ V and the surface potential at the source at the onset of strong inversion $\varphi_{s,inv,th,0} = 0.9$ V. For the estimation, it may be assumed that the increase in the surface potential after the onset of strong inversion $\Delta\varphi_{s,inv,0} = (\varphi_{s,inv,0} - \varphi_{s,inv,th,0}) = 0.2$ V [21, 22]. Flat-band voltage versus EOT measurements indicate that the total gate stack potential due to all its fixed and trap charges may be as much as 0.3 V [17]; we may choose a gate stack potential due to the net gate stack charges $V_{di,gsc} = 0.2$ V. The work function anomaly varies a great deal depending upon the metal, the gate stack high-k layer, and the processing. At the current state of the work function control, it may be reasonable to assume a work-function difference $\phi_{MS,p} = 0.1$ V. Hence, a numerical estimate of the drain current according to (2.37) would be:

$$\frac{I_{D,high-k}}{I_{D,ideal}} = \frac{\mu_{ch,high-k}}{\mu_{ch,ideal}} \times \frac{(2.2-0.9-0.2-0.2-0.1-0.2)0.4 - \frac{2}{3}\cdot 0.1\left[(0.9+0.2+0.4)^{\frac{3}{2}} - (0.9+0.2)^{\frac{3}{2}}\right]}{(2.2-0.9-0.2)0.4 - \frac{2}{3}\cdot 0.1\left[(0.9+0.4)^{\frac{3}{2}} - (0.9)^{\frac{3}{2}}\right]}$$
$$= \frac{\mu_{ch,high-k}}{\mu_{ch,ideal}} \times \frac{0.6\cdot 0.4 - 0.067\left((1.5)^{3/2} - (1.1)^{3/2}\right)}{1.1\cdot 0.4 - 0.067\left[(1.3)^{\frac{3}{2}} - (0.9)^{\frac{3}{2}}\right]} = \frac{\mu_{ch,high-k}}{\mu_{ch,ideal}} \times \frac{0.24 - 0.067\cdot 0.69}{0.44 - 0.067\cdot 0.63}$$
$$= \frac{\mu_{ch,high-k}}{\mu_{ch,ideal}} \times \frac{0.24-0.04}{0.44-0.04} = \frac{0.20}{0.40}\frac{\mu_{ch,high-k}}{\mu_{ch,ideal}} = 0.50\frac{\mu_{ch,high-k}}{\mu_{ch,ideal}} \tag{2.38}$$

These calculations suggest that the part of the relation containing the bracket in (2.37) or (2.38), is not significant for small drain voltages; the part containing the bracket represents the difference between the depletion charge at the drain and at the source and assumes importance for large drain voltages. Therefore, for the triode regime, we may focus on the part within the parentheses. The calculation in (2.38) suggests that the non-saturating surface potential, the gate stack charges, and the work-function difference may have the same order of weight; in the present example, the three factors degrading the drain current roughly 20, 20, and 10 %, respectively, such that the total degradation is 50 %, i.e. the drain current is reduced to 50 % of its ideal value by these three factors in addition to the reduction by the degraded channel mobility. The calculations illustrated by (2.38) show that even for moderate drain voltages (such as $V_D = 0.4$ V), the drain current versus the drain voltage relation becomes significantly more non-linear for the high-k gate stacks, because the term V_D inside the parenthesis in (2.37) becomes a more significant fraction of the rest of the terms inside the parenthesis.

There are some experimental results [27–31] on the degradation of the channel parameters I_D, g_D, and g_m in high-k MOSFETs by the non-ideal factors of $\Delta\varphi_{s,inv}$, $Q_{di,gsc}$, and ϕ_{MS} which support the basic suggestions of (2.37) and (2.38). High pressure annealing of Si/Hf-silicate/HfAlO/TiN gate stacks in pure H_2 was reported to bring down the interface state density by a factor of 2 (from 12 to 6×10^{10} cm^{-2}) and enhance the drain current and the transconductance by 10–15 % [27]. Fluorine treatment and GeO_2 passivation of HfO_2/Ge gate stacks was observed to reduce the interface trap density from 4 to 1×10^{12} cm^{-2} V^{-1}, while enhancing the drain current by 18 % [28]. Fluorine incorporation into HfO_2/InP and $HfO_2/In_{0.53}Ga_{0.47}As$ gate stacks was reported to improve the drain current and the transconductance and these improvements were attributed to a reduction in the fixed charge and the interface trap density [29]. Passivation of HfO_2/InP gate stacks by an intermediate Al_2O_3 layer was observed to enhance the drain current 2.5 times and the transconductance 4 times while reducing the interface trap density [30]. Fluorine treatment of $InP/Al_2O_3/HfO_2$/TaN gate stacks was found to increase the drain current 100 %, the transconductance 50 %, and the channel mobility 56 %; this enhancement was attributed to a reduction in the gate stack charge [31].

2.6.1.4 Channel Conductance and Trans-conductance

The channel conductance may be expressed as, cf. (2.35):

$$g_D = \frac{\partial I_D}{\partial V_D} = \frac{W}{L}\mu_{ch}C_{di}\begin{bmatrix} \left(V_G - \varphi_{s,inv,0} - V_{di,gsc} - \phi_{MS,p}\right) \\ -2/3\frac{\sqrt{2q\varepsilon_s N_A}}{C_{di}}3/2\sqrt{\varphi_{s,inv,0} + V_D} - V_D \end{bmatrix}$$

or,

$$g_D = \frac{W}{L}\mu_{ch}C_{di}\left[\left(V_G - \varphi_{s,inv,th,0} - \Delta\varphi_{s,inv,0} - V_{di,gsc} - \phi_{MS,p}\right) - V_D - V_{di,sc,L}\right] \tag{2.39}$$

The term $V_{di,sc,L}$ represents the gate stack potential at the drain (y = L) due to the ionized dopant charge; it is a function of both the gate and the drain voltages. Equation (2.39) may be compared with its classical formulation, cf. (2.15):

$$g_D = \frac{W}{L}\mu_{ch}C_{di}\left[\left(V_G - \varphi_{s,inv,th,0}\right) - V_D - V_{di,sc,inv,L}\right] \tag{2.40}$$

The quantity $\varphi_{s,inv,th,0}$ is a constant. The term $V_{di,sc,inv,L}$ represents the gate stack potential due to the ionized dopant charge at the drain (y = L) at the onset of strong inversion; it is a function of the drain voltage but not of the gate voltage. The channel conductance may be normalized by its ideal value and expressed as, cf. (2.39), (2.40):

$$\begin{aligned} g_{D,norm} &= \frac{g_{D,high-k}}{g_{D,ideal}} \\ &= \frac{\mu_{ch,high-k}}{\mu_{ch,ideal}}\frac{V_G - \varphi_{s,inv,th,0} - \Delta\varphi_{s,inv,0} - V_{di,gsc} - \phi_{MS,p} - V_D - V_{di,sc,L}}{V_G - \varphi_{s,inv,th,0} - V_D - V_{di,sc,inv,L}} \end{aligned} \tag{2.41}$$

The normalized channel conductance has many parameters in common with the normalized drain current, cf. (2.37) and (2.41); however, g_D is degraded more than I_D, as may be illustrated by making a rough numerical estimate of $g_{D,norm}$ assuming the same values of the parameters, selected for estimating the normalized drain current, cf. (2.38):

$$\begin{aligned} g_{D,norm} &= \frac{\mu_{ch,high-k}}{\mu_{ch,ideal}}\frac{2.2 - 0.9 - 0.2 - 0.2 - 0.1 - 0.4 - 0.12}{2.2 - 0.9 - 0.4 - 0.11} \\ &= \frac{0.28}{0.79}\frac{\mu_{ch,high-k}}{\mu_{ch,ideal}} = 0.35\frac{\mu_{ch,high-k}}{\mu_{ch,ideal}} \end{aligned} \tag{2.42}$$

As (2.38) and (2.42) indicate, the estimated degradation in channel conductance is nearly 50 % higher than that in the drain current.

The transconductance may be expressed as, cf. (2.35):

$$g_m = \frac{\partial I_D}{\partial V_G} = \frac{W}{L}\mu_{ch}C_{di}\left[\begin{array}{l}\left(1 - \frac{\partial \varphi_{s,inv,0}}{\partial V_G} - \frac{\partial V_{di,gsc}}{\partial V_G}\right)V_D \\ -2/3\frac{\sqrt{2q\varepsilon_s N_A}}{C_{di}}3/2\left(\sqrt{\varphi_{s,inv,0}+V_D} - \sqrt{\varphi_{s,inv,0}}\right)\frac{\partial \varphi_{s,inv,0}}{\partial V_G}\end{array}\right] \tag{2.43}$$

or:

$$g_m = \frac{W}{L}\mu_{ch}C_{di}\left[\left(1 - \frac{\partial \varphi_{s,inv,0}}{\partial V_G} - \frac{\partial V_{di,gsc}}{\partial V_G}\right)V_D - \left(V_{di,sc,L} - V_{di,sc,0}\right)\frac{\partial \varphi_{s,inv,0}}{\partial V_G}\right] \tag{2.44}$$

$V_{di,sc,0}$ is the gate stack potential at the source (y = 0) due to the ionized dopant charge. Equation (2.44) may be compared with its classical counterpart as expressed below, cf. (2.36):

$$g_m = \frac{W}{L}\mu_{ch}C_{di}V_D \tag{2.45}$$

The transconductance normalized by its ideal value may be expressed as, cf. (2.44) and (2.45):

$$\begin{aligned} g_{m,norm} &= \frac{g_{m,high-k}}{g_{m,ideal}} \\ &= \frac{\mu_{ch,high-k}}{\mu_{ch,ideal}}\left(1 - \frac{\partial \varphi_{s,inv,0}}{\partial V_G} - \frac{\partial V_{di,gsc}}{\partial V_G} - \frac{\left(V_{di,sc,L} - V_{di,sc,0}\right)}{V_D}\frac{\partial \varphi_{s,inv,0}}{\partial V_G}\right) \end{aligned} \tag{2.46}$$

The equation for the transconductance is clearly different in nature from those for the drain current and the channel conductance, cf. (2.46), (2.41), (2.37); namely, (2.46) contains differentials (all of which are fractions) while (2.41) and (2.37) do not. Moreover, the transconductance is not degraded by the work function difference anomaly. It is important to note the strong dependence of the transconductance degradation on the gate voltage. The rate of change of the surface potential with respect to the gate voltage, $\partial\varphi_{s,inv,0}/\partial V_G$, can be obtained from an experimental $\varphi_s(V_G)$ plot, cf. Fig. 2.15. In Fig. 2.15, $\partial\varphi_{s,inv,0}/\partial V_G$ for both the high-k gate stacks is about 0.25 in strong inversion, while it is only 0.08 for the single SiO_2 gate dielectric. This strongly illustrates how important the non-saturating inversion surface potential is in the case of the high-k gate stacks. Similar to the exercise in the case of the drain current and the channel conductance, cf. (2.38) and (2.42), a numerical estimate could be made for the normalized transconductance:

$$
\begin{aligned}
g_{m,norm} &= \frac{\mu_{ch,high-k}}{\mu_{ch,ideal}}\left(1-\frac{\partial\varphi_{s,inv,0}}{\partial V_G}-\frac{\partial V_{di,gsc}}{\partial V_G}-\frac{0.01}{0.40}\frac{\partial\varphi_{s,inv,0}}{\partial V_G}\right)\\
&= \frac{\mu_{ch,high-k}}{\mu_{ch,ideal}}\left(1-1.025\frac{\partial\varphi_{s,inv,0}}{\partial V_G}-\frac{\partial V_{di,gsc}}{\partial V_G}\right)\\
&= \frac{\mu_{ch,high-k}}{\mu_{ch,ideal}}\left(1-1.025\times 0.25-\frac{\partial V_{di,gsc}}{\partial V_G}\right)\\
&= \frac{\mu_{ch,high-k}}{\mu_{ch,ideal}}\left(1-0.26-\frac{\partial V_{di,gsc}}{\partial V_G}\right)=\frac{\mu_{ch,high-k}}{\mu_{ch,ideal}}\left(0.74-\frac{\partial V_{di,gsc}}{\partial V_G}\right)
\end{aligned}
\tag{2.47}
$$

The rate of change in the gate stack potential due to the gate stack charges, $V_{di,gsc}$, with respect to the gate voltage V_G will depend upon how the trap charge inside the gate stack will change and this will be determined by the trap energy levels and the wave function penetration into the gate stack.

Equations (2.38), (2.42), and (2.47) suggest that the degradation is most severe in the case of the channel conductance, followed by the drain current, and then the transconductance.

2.6.1.5 Factors Attenuating Channel Parameters

We may recapitulate our analysis in Sect. 2.5 and Sect. 2.6.1 to list all the factors which force the drain current to attenuate. Ideally, as already stated in Sect. 2.5.1, the entire gate voltage should be spent in enhancing the channel conductivity equally everywhere in the channel irrespective of the y position. The factors which each consumes and wastes a part of the gate voltage are:

1. The drain voltage V_D. The drain voltage causes the voltage across the gate stack to reduce; the amount of reduction depends upon the position along the direction y—at the drain, the amount of reduction is $= V_D$.
2. The ionized dopant charge Q_{dep}. Ideally, the semiconductor surface charge layer should have only electrons or holes. A part of the gate voltage is wasted in sustaining the dopant charge; this effect is more significant for large drain voltages.
3. Work function difference ϕ_{MS}. Ideally, the semiconductor–metal work function difference ϕ_{MS} should be zero. A part of the gate voltage is directly wasted in neutralizing the non-zero work function difference.
4. Gate stack charge $Q_{di,gsc}$. The gate dielectric should be an ideal one; in other words, the gate stack should have only ideal gate dielectrics as constituents, in which the charge density should be zero. A significant part of the gate voltage is wasted in supporting the gate stack charge $Q_{di,gsc}$.
5. Non-saturating inversion surface potential $\Delta\varphi_{s,inv}$. Ideally, the inversion surface potential should remain frozen at its value at the onset of strong inversion

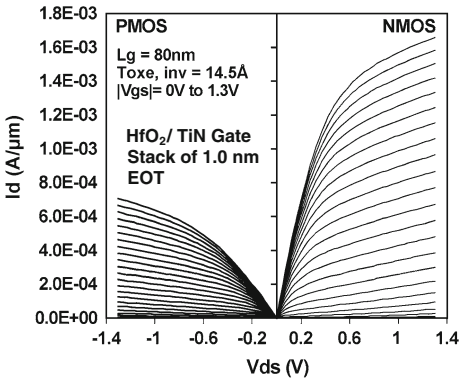


Fig. 2.16 Drain current versus drain voltage characteristics of n-channel and p-channel MOSFETs with HfO_2/TiN gate stack. The EOT was 1.0 nm, whereas the CET was 1.45 nm. The channel length was 80 nm. The gate voltage varied between 0 and 1.3 V. I_D for the n-channel was 1.66 mA/μm and for the p-channel was 0.71 mA/μm at a V_D of 1.3 V. Adapted from [32]

$\varphi_{s,inv,th}$. The excess inversion surface potential $\Delta\varphi_{s,inv}$ reflects several factors and consumes a significant fraction of the gate voltage.

Figure 2.16 illustrates the drain current versus the drain voltage characteristics measured for both NMOSFETs and PMOSFETs with HfO_2/TiN high-k gate stacks [32]: The channel length was 80 nm, EOT was 1.0 nm, CET was 1.45 nm [32]. It should be of interest to compare the characteristics of Fig. 2.16 representing high-k gate stacks with those of Fig. 2.13 representing SiO_2 single gate dielectric; the two sets of characteristics differ in several respects both in the triode as well as in the saturation regimes. In the triode regime, the high-k characteristics deviate more from a linear form than do the single gate SiO_2 characteristics, whereas in the so-called saturation regime, the former are non-saturating. The deviation from linearity manifested in Fig. 2.16 supports the conclusions of Sect. 2.6.1 that the three non-ideal factors—of $\Delta\varphi_{s,inv}$, $Q_{di,gsc}$, and ϕ_{MS}—make the $I_D - V_D$ characteristics more non-linear and degrade the channel parameters of high-k gate stacks significantly, particularly when the supply voltage is reduced, as is the case with the decreasing EOT trend.

2.6.2 Composition of the High-k Gate Stack

The high-k gate stack presents a very formidable challenge to effective (in the sense of being accurate and yet practical in use) modeling and representation. Our aim is to evolve realistic energy band diagrams and circuit representations of the high-k gate stack, and modeling of the various high permittivity layers, and traps and charges, which abound in these systems, and the potentials across the various

dielectric layers. Among the important basic issues, which will be discussed in this section, are:

1. Which constituents/components of the gate stack are to be recognized for representation? Each layer with a different permittivity should in principle be represented by a capacitor in series with those of the adjacent layers.
2. Are there chemically graded layers, intentionally or unintentionally in the gate stack? It has been established [33, 34], that across a Si/SiO_2 interface, there exists a 0.3–0.5 nm thick chemically graded (hence, band-gap and permittivity graded) layer. It is possible that chemically graded layers exist also across the intermediate-layer/high-κ-layer interface and even the high-κ-layer/metal interface (cf. Fig. 2.14). How does one represent a graded permittivity layer in the equivalent circuit, which in principle would be an infinite set of capacitors in series?
3. Is there a metal oxide layer formed between the high-k bulk and the metal electrode?
4. What are the tangible consequences of the very high density of traps in the gate stack layers? The SiO_2 gate dielectric could be represented by a simple dielectric capacitance (plane-parallel capacitor), because there were no significant traps inside, i.e. it could be represented as an ideal dielectric layer. In great contrast, the charging and discharging in the myriad traps, inside the high-κ gate dielectric stack, in principle, need to be represented by a large number of R_tC_t combinations.
5. This brings us to the issue of the Fermi occupancy of the traps and the profile (x-direction-wise) of the pseudo-Fermi function inside the gate dielectric stack. One could argue that in strong accumulation and in strong inversion, it may be possible for many traps inside the high-κ gate dielectric stack, if not all of them, to follow the applied signal, and exchange electrons/holes, either with the silicon bands or with the metal, if the EOT < 1.0 nm.
6. Is there a Schottky barrier at the high-k/metal interface?

The basic to any treatment of the gate stack, be it the energy band diagram, the charge analysis, the potential profile, or the circuit representation, is the identification or selection of the constituents of the gate stack. As already discussed, no unique identification is possible, and one has to make a choice or compromise between complexity, accuracy, and practicality. An important point in this context is the strong variation in the approach and practice for an effective passivation of the different semiconductor (Si, Ge, GaAs, InGaAs) surfaces. This may result in significantly different gate stacks and metal electrodes on different semiconductor substrates. Our analysis in this section will be based upon the silicon substrate. We may have a composition of the gate stack, as complicated as illustrated in Fig. 2.17. On the silicon substrate, the intermediate layer is typically SiO_2 or a silicon-oxynitride, whereas the most common high-k layer, currently, is a Hf-based oxide or oxynitride or silicate or aluminate with or without a dopant (e.g. Y, La). The grading of the interfacial layer between the intermediate layer and the high-k

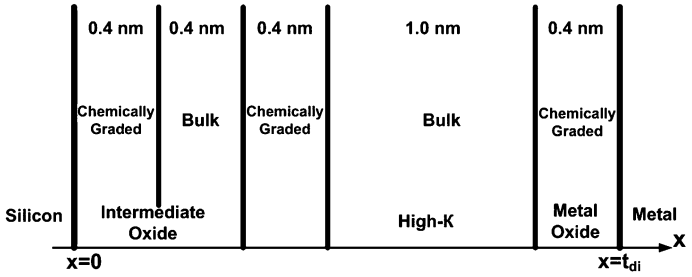


Fig. 2.17 Schematic representation of the five different layers, each with a characteristic permittivity, constituting the high-k gate stack

layer and that between the high-k layer and the gate metal has not been investigated as much as has been the Si-SiO_2 graded interface [33, 34], but graded layers are possible, when one considers atomic, ionic, and inter-layer diffusions, that are likely to occur in the gate stack, due to high concentration gradients [25, 26]. The thickness indicated in Fig. 2.17 for the graded layer at the Si-SiO_2 interface has been established [33, 34], but for the other two graded layers is only indicative.

2.6.3 Energy Profile of the High-k Gate Stack

The representation of Fig. 2.17 would be too complicated to deal with even in terms of the energy band profile, let alone the circuit representation, the representation of the traps, and the electrostatic relations. To arrive at a practical solution, we may consider the following simplifications:

1. The possibility of a metal oxide between the high-k layer and the metal electrode may be ignored. Investigations suggest oxidation of the metal electrode surface and the resultant presence of an oxide to be likely if the metal electrode (e.g. AlN) contains a reactive metal such as Al.
2. Representation of the graded layers may be eliminated to avoid complications too difficult to deal with.
3. The above two simplifications will leave us with two bulk layers—the intermediate layer (e.g. SiO_2 or SiON or a silicate) and a high-k layer (e.g. Hf- or La-silicate or aluminate or oxynitride)—and their three interfaces—Si/IL, IL/high-k, and high-k/metal.
4. The traps and charges at or near the Si/IL interface are comparatively small, as conductance measurements indicate these traps in the device quality MOSFETs to be of the order of 10^{11} cm^{-2}, i.e. orders of magnitude less than the high-k charges and traps. Experimental results will be presented in the later sections.
5. The traps and charges inside the IL are also smaller, as evidence to be presented in the later sections will indicate these to be an order of magnitude smaller than the high-k charges. Experiments on devices with a graded IL indicate these traps and charges to cause voltage drops not exceeding a few tens of mV.

6. The above simplifications will leave us with traps and charges at the IL/high-k interface, the same in the high-k bulk, and in the semiconductor space charge as the dominating ones; these are the traps and charges whose influence on the energy band profile, we will consider.

Figure 2.18 represents a schematic of the simplified high-k gate stack, consisting of two dielectric layers, namely the intermediate layer (IL) and the high-k layer, and three interfaces, namely the silicon/IL interface, the IL/high-k interface, and the high-k/metal interface. Figure 2.18 illustrates the location of the dominant gate stack traps and charges, and identifies the three most important charge locations, which are likely to dominate the electrostatic relations and the carrier transport through the gate stack. It may be borne in mind that the gate stack reliability may not necessarily be dominated by these charge centers. The dominant charges suggested by Fig. 2.18 are:

1. The semiconductor space charge of density Q_{sc} spread over the space charge layer of width W, which has been analyzed in Sect. 2.2.4. This charge will cause significant voltage drop across the gate stack.
2. The traps and charges inside the high-k layer. This charge density is of the order of 10^{13} cm^{-2}, and causes voltage drop of the order of hundreds of mV. The profiles (x-direction variation) of these traps and charges have not been firmly established.
3. The traps and charges at the IL/high-k interface may consist of two entities—an interface dipole and/or interface traps. A dipole, i.e. a positive and a negative layer of equal but opposite charges, of large magnitude at the IL/high-k has been well established by the experimental results [25, 26]. Traps of significant

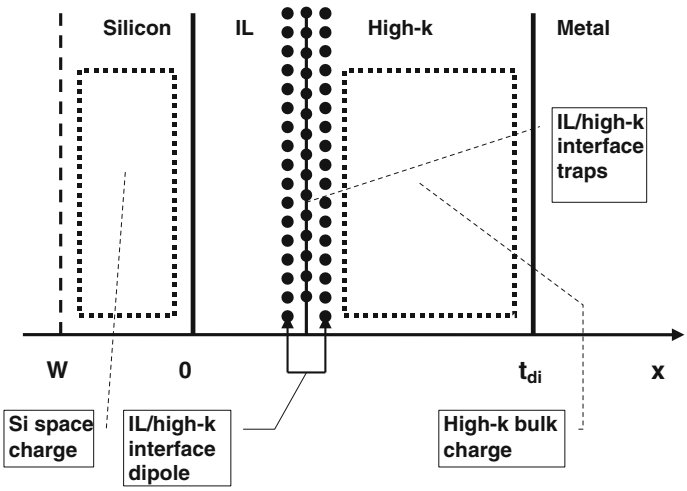


Fig. 2.18 Simplified schematic representation of a high-k gate stack. W is the semiconductor space charge width; t_{di} is the total gate stack physical thickness

density are likely to be present at the IL/high-k interface due to a large mismatch in the chemical bonding.

Several issues need to be considered to render an energy profile representation of even this simplified gate stack.

Image Force Barrier Lowering—When an electron/hole leaves the emitting surface, it experiences an attractive force due to its image charge in the emitter. An attractive force lowers the potential barrier to the electron/hole. Hence, the band offsets and the electron/hole energy barriers are reduced. As the attractive force depends upon the distance between the free carrier and its image charge, the entire gate stack energy profile is modified and the potential energy barrier maximum is shifted from the interface. The image force barrier lowering $\Delta\phi_b$ may be expressed as [2]:

$$\Delta\phi_b = \sqrt{\frac{qE_{\max}}{4\pi\varepsilon_{di}}}; \quad x_{\max} = \sqrt{\frac{q}{16\pi\varepsilon_{di}E_{\max}}} \tag{2.48}$$

E_{max} is the maximum electric field, and x_{max} is the shift in the position of the maximum barrier energy. Equation (2.48) shows that the image force barrier lowering becomes less important in the case of the high-k dielectrics; hence, it may be ignored in the interest of simplification.

Schottky Barrier Formation at the High-k/Metal Interface—As discussed in Sect. 2.2.1, a free carrier exchange across a metal/semiconductor interface leads to the formation of a Schottky barrier; the dipole consists of a charge layer on the metal surface and the space charge layer in the semiconductor sub-surface. A dipole and therefore a Schottky barrier, in principle, cannot form at the metal/ideal-dielectric interface (e.g. at the SiO_2/metal interface), because an ideal dielectric, or even a near-ideal dielectric, such as the dry thermal SiO_2, does not have any free carriers to exchange with the metal. A high-k dielectric layer, likewise the dry thermal SiO_2, has no free carriers in the conduction/valence band, but in strong contrast to an ideal dielectric, has a high density of traps, with which an exchange of free carriers can, in principle, take place with the metal. The effect of such a Schottky barrier on the electrostatic relations and the carrier transport is likely to be muted on account of its proximity to the metal electrode and the high band gap of the high-k layer. Therefore, in the interest of practicality, we will not consider this feature.

Figure 2.14 represents the energy profiles across a p-silicon/SiO_2/HfO_2/TaN MOS capacitor with an intermediate layer of about 1 nm thick SiO_2, a high-k layer of about 2 nm HfO_2, and a total EOT of about 1.9 nm. The profile of the vacuum level has been represented. The vacuum level is the hypothetical energy level of an electron in absolute vacuum (Where no other entity exists with which the electron could interact.). It is the highest potential energy an electron can have, and is used as an electron energy reference, i.e. against which other electron energies are compared. All the electrostatic potentials, i.e. the surface potential φ_s and the components of the gate stack potential V_{di}, have been marked. It may be noted that the internal vacuum level, being a potential energy, has necessarily to follow the

same profile as the electrostatic potential. The total potential across the gate stack can have different components, depending upon which physical components of the gate stack have been recognized. In the representation of Fig. 2.14, we have recognized three components, namely, the intermediate layer (the dry thermal SiO_2 layer), the high-k layer (the HfO_2 layer), and the dipole at the SiO_2/high-k interface.

An important point that is illustrated in Fig. 2.14 is that the net electric field and the electrostatic potential across any component of the gate stack do not have to be of the same direction or polarity as those obtaining across any other component of the gate stack. In turn, the net electric field and the potential across any gate stack component can result from a number of diverse charges not having the same polarities. The charges, which we have considered in constructing the potential profile of Fig. 2.14, are: the semiconductor space charge Q_{sc}, the bulk high-k charge, the dipole charge, and the charge of traps at the IL/high-k interface; these charges have been illustrated in Fig. 2.18; each of these charges contributes to the net electric field and the net potential across the high-k (HfO_2) layer. As the p-type semiconductor sub-surface is in accumulation, cf. Fig. 2.14, the space charge Q_{sc} consists of holes and is positive; hence the electric field is positive and the surface potential φ_s is negative (has value of -0.42 V at a bias of -1.82 V [16], as obtained from the Berglund integral). As we have ignored traps at the Si/SiO_2 interface as well as in the bulk SiO_2 traps, the charge contributing to the field across the SiO_2 layer is only Q_{sc}; hence, the electric field across the SiO_2 layer is positive and the potential $V_{di,IL}$ is negative (estimated to be about -0.44 V).

The electrostatic picture for the HfO_2 layer or for that matter any other high-k layer, is complicated because of the diverse charges present in its bulk and at its interfaces. The experimental data [16] on the MOS capacitor of Fig. 2.14 (to be presented in more detail later) indicate a total gate stack potential V_{di} of about -1.38 V corresponding to an applied bias of -1.82 V. For the approximations made on the gate stack charges, the total gate stack potential may be expressed as the net sum of the following components:

$$V_{di} = V_{di,IL} + V_{di,dipole} + V_{di,IL/high-k} + V_{di,sc} + V_{di,high-k} \tag{2.49}$$

$V_{di,IL/high-h}$ and $V_{di,high-k}$ are potentials across the HfO_2 layer due to the charge density of the traps at the IL/high-k interface and the high-k bulk charge density, respectively. The potential across SiO_2 layer $V_{di,IL}$ has been estimated to be about -0.44 V. The potential $V_{di,sc}$ across the HfO_2 layer due to the semiconductor space charge Q_{sc} is estimated to be about -0.17 V. This will suggest that the net sum of ($V_{di,dipole} + V_{di,IL/high-k} + V_{di,high-k}$) is about -0.77 V. From the experimental data [16], under the flat-band condition, the potential across the HfO_2 layer due to its bulk charges $V_{di,high-k}$ is estimated to be about +0.10 V, and the sum of ($V_{di,dipole} + V_{di,IL/high-k}$) is estimated to be +0.15 V. Hence, at flat-band, the sum of ($V_{di,dipole} + V_{di,IL/high-k} + V_{di,high-k}$) is +0.25 V, whereas in strong accumulation at an applied bias of -1.82 V, this sum is -0.77 V. This discrepancy can be

reconciled, if all or some of the high-k related charges considered by us vary with the applied bias.

2.6.4 Occupancy of Interface Traps and Bulk Traps in the High-k Gate Stack

The Fermi-Dirac distribution function, in which the Fermi level is the critical parameter, determines the occupancy of an eigenstate. The Fermi level (popular name for the chemical potential, and sometimes mixed up with the Fermi energy, which is defined only at 0 K.) and the law of mass action ($pn = n_i^2$) are thermal equilibrium concepts. The law of mass action is synonymous with a common Fermi level for both electrons and holes. Strictly speaking, thermal equilibrium no longer holds once a bias is applied across the MOS capacitor.

The concept of the quasi-Fermi level (imref—Fermi written in reverse) has no rigorous basis—it is only a practical tool. The electron imref at any location enables us to determine the free electron density at that point, while the hole imref enables us to determine the hole concentration. The deviation of the hole imref from the electron imref represents the scale of the thermal non-equilibrium and the magnitude of the direct current flowing at that point. Notwithstanding its empirical basis, the imref tool has been extensively used inside the semiconductor space charge layer. The occupancy of traps in the dielectric/insulator bulk is a rarely discussed subject; and the extension of the quasi-Fermi level concept to the inner region of a dielectric or insulator is questionable, as there are no free carriers in its conduction/valence band, except in transparent conductors such as SnO_2/In_2O_3.

As already mentioned, the high-k dielectric (its bulk and its interfaces) may contain a multitude of traps and charges possibly of different origin and a multitude of electron and hole trap levels; hence it is necessary to know the occupancy of the diverse traps, and the variation of the trap occupancy with the applied bias.

The concept of pseudo-Fermi function and pseudo-Fermi level has been invoked a long time ago to represent the occupancy of a trap inside a dielectric [12], but has rarely been used and developed further. This concept would approximately (i.e. the 0 K approximation) mean that the gate stack traps below the pseudo-Fermi level are occupied by electrons, and are empty if above. In Fig. 2.14, the hole (i.e. the majority carrier) imref in the semiconductor space charge region, and the pseudo-Fermi level in the gate stack have been illustrated. It would be meaningful to apply the concept of the pseudo-Fermi level inside the gate stack, only if the gate stack is thin enough for significant wave function penetration and tunneling. This context leads us to discuss the topic of quantum-mechanical tunneling, which has an important bearing upon a variety of phenomena in and around the high-k gate stack.

2.6.5 Potential Well and Quantum-Mechanical Phenomena

In quantum mechanics, all properties of the electron—its eigen (discrete) energy, the probability of finding it in a given volume of space, its energy bands, its effective mass—are obtained by solving the Schrödinger equation: $H\psi = E\psi$, where the operator H is the Hamiltonian (sum of the potential energy V and the kinetic energy of the electron), E is the eigen-energy, and ψ is the electron wave function. The electron wave function ψ has no direct physical meaning, however, the entity $\psi\psi^*dV$ represents the probability of finding the electron in the infinitesimal volume dV. For no situation in a solid, there is a closed-form solution for the Schrödinger equation, because even in a perfect periodic crystal, the potential energy term is complicated even in its approximated expression. We need to invoke some quantum-mechanical concepts in order to gain some insight into the following phenomena in and around the gate dielectric stack:

Carrier confinement in the potential well of the strong inversion layer (*or of the accumulation layer*)—Strong inversion or accumulation leads to the formation of a potential well in the semiconductor sub-surface; the profile of the potential well is defined by the (conduction or valence) band bending in the semiconductor, $\varphi(x)$, and the band offset at the semiconductor/gate-stack interface, $\phi_{b,c}$ or $\phi_{b,v}$, see Fig. 2.19. An electron or a hole, having a kinetic energy less than the barrier energy $q\varphi_s$, is confined in the x-direction to the perimeters of the potential well, and can no longer be represented by a Bloch wave function, i.e. a travelling wave having the periodicity of the lattice, see (2.50):

$$\Psi_{nk}(\mathbf{r}) = u_{nk}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) \tag{2.50}$$

However, an electron with a kinetic energy exceeding the barrier energy $q\varphi_s$ may be represented by a travelling wave, while an electron with a kinetic energy less than $q\varphi_s$ will be represented by a standing (or stationary) wave, cf. Fig. 2.19. The latter electrons are localized in the potential well and will be characterized by bound states in the energy sub-bands, cf. Fig. 2.19.

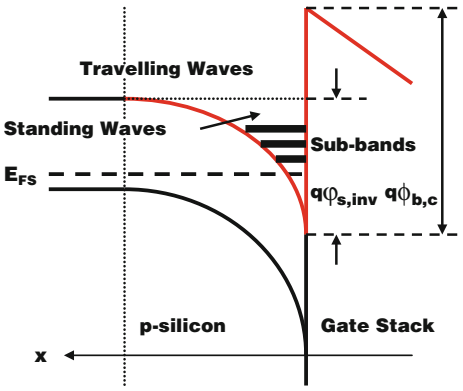


Fig. 2.19 Energy band profile across a p-Si/gate-stack system in strong inversion, illustrating the effects of carrier confinement in the potential well (barrier energy highlighted in *red*) ≫ energy sub-bands, standing waves, etc

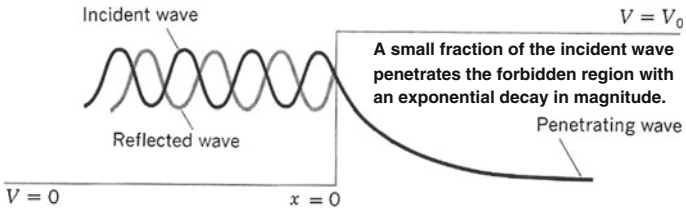


Fig. 2.20 Penetration of a wave, incident at a step potential barrier, into a classically forbidden region. The incident electron energy E is < the barrier energy V_0. A small fraction of the incident wave penetrates; the rest is reflected. Adapted from [35]

Carrier wave function penetration into the forbidden potential energy barrier of the gate stack—As the dielectric layers of the gate stack have high band gaps, the gate stack presents a potential energy barrier to the free carriers in the semiconductor and the metal, see Fig. 2.19. In classical physics, an electron of energy E incident at a potential barrier V (V > E) will be turned back, i.e. totally reflected, at the so-called classical turning point, cf. Fig. 2.20. In the quantum-mechanical description, unless the potential barrier V is infinite, which is never the case in a gate stack or in any reality, there will exist a finite, howsoever small, probability of finding the electron inside the potential barrier. In other words, the electron wave function penetrates the potential barrier with an exponentially decaying amplitude, and since a wave function exists, the probability $\psi\psi^*dV$ of finding it at a point inside the barrier is finite, cf. Fig. 2.20. To be of some consequence, this point inside the potential barrier, as represented by the gate stack, has to be within a nm or so from the semiconductor or the metal surface. It may be noted that wave function penetration can occur from both the semiconductor as well as from the metal surface.

Tunneling of the free carriers from the semiconductor to the metal through the composite potential barrier of the gate stack and vice versa—Free carrier transport by tunneling across a potential barrier occupies a special place in quantum mechanics, as its quantum mechanical formulation is relatively simple, and, more importantly, it was perhaps the first demonstration of the validity of the quantum mechanical description of matter including the wave-particle duality of an electron. The name of the process under discussion (i.e. tunneling) derives itself from an analogy to a tunnel through a mountain barrier (standing for the forbidden potential barrier). According to the concept of elastic tunneling, a free carrier of energy E incident at a potential barrier of height V (V > E) and thickness t, can be transmitted through the potential barrier to an empty or partially empty eigenstate of the same energy on the other side of the barrier, cf. Fig. 2.21. The tunneling transmission coefficient T or the tunneling probability is exponentially dependent upon the product of the barrier thickness and the square root of the excess barrier energy (motive energy) [V(x) − E]. Due to its stronger dependence on the barrier thickness, the tunneling probability becomes insignificant for barriers much thicker than a nm. To gain a rough estimate of the transmission coefficient, one could use a rule of thumb: For a motive energy of an eV and a tunneling electron/hole mass of

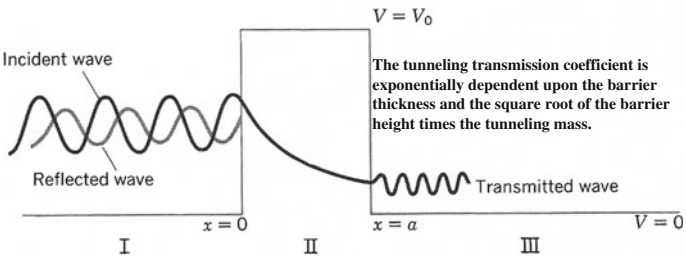


Fig. 2.21 Wave penetration and quantum-mechanical tunneling through a *rectangular barrier* of height V_0 and thickness a, of an electron of energy $E < V_0$, incident at the barrier at $x = 0$ from the *left*. The incident wave is partly transmitted and partly reflected. Adapted from [35]

1.0 m, the transmission coefficient reduces by 1/e per each 0.1 nm, such that for 1 nm thick barrier, $T \cong e^{-10}$ if $(V - E)$ is 1 eV, and $T \cong e^{-20}$ if $(V - E)$ is 4 eV.

2.6.5.1 Tunneling Through the Gate Stack

Types of tunneling, relevant for the high-k gate stack, include: direct elastic tunneling (incident and transmitted states are of the same energy), inelastic tunneling (incident and transmitted states are of different energy), trap-assisted tunneling (could be a chain process mediated by several traps), and Fowler-Nordheim tunneling (tunneling into a conduction/valence band in the gate stack). Theoretical treatments exist for direct tunneling through a rectangular potential barrier and other simple barriers in the text books on quantum mechanics [35, 36]. Simplest is the case for a rectangular potential barrier (of height V_0), see Fig. 2.21, for which solution of the one-dimensional, time-independent Schrödinger equation:

$$-\frac{\hbar^2}{2m^*}\Psi'' + (V_0 - E)\Psi = 0, \tag{2.51}$$

yields the following closed-form wave functions for an electron of energy E incident on the barrier from the left:

$$\Psi_I = Ae^{ikx} + Be^{-ikx}, \quad x<0 \tag{2.52}$$

$$\Psi_{II} = Ce^{\kappa x} + De^{-\kappa x}, \quad 0<x<a \tag{2.53}$$

$$\Psi_{III} = Fe^{ikx}, \quad x > a \tag{2.54}$$

$$k = \sqrt{\frac{2m^*E}{\hbar^2}}, \quad \kappa = \sqrt{\frac{2m^*}{\hbar^2}(V_0 - E)} \tag{2.55}$$

ħ is Planck's constant/2π, m* is the tunneling effective mass, A, B, C, D and F are constants, k is the electron wave vector, κ is the wave attenuation constant, and a is the barrier thickness. It is not clear what the tunneling effective mass should be.

First of all, the tunneling process does not involve any electron motion in the usual sense; so Newton's second law of motion and the usual effective mass concept may not apply. Secondly, it is also not clear whether any motion at all is involved in the tunneling process.

Application of the boundary conditions (continuity of the wave function and its derivative at the boundaries x = 0 and at x = a) yields the following relations among the constants:

$$\begin{aligned} C &= \frac{1}{2\kappa}\{(\kappa + ik)A + (\kappa - ik)B\} = \frac{\kappa + ik}{2\kappa} e^{-(\kappa - ik)a} F \\ D &= \frac{1}{2\kappa}\{(\kappa - ik)A + (\kappa + ik)B\} = \frac{\kappa - ik}{2\kappa} e^{-(\kappa + ik)a} F \end{aligned} \tag{2.56}$$

The profile of the potential barrier presented by the gate stack, V(x), is very different from a rectangular shape. Nevertheless, the relations in (2.52–2.56) could still be applied, if we replace V_0 in (2.55) by the potential energy profile V(x) across the gate stack. For EOT of 1 nm or less, the gate leakage current is very significant and is one of the most important issues for the MOSFET. The dominant mechanism of carrier transport through the gate stack is likely to be some form of tunneling. Hence, the ability to reliably estimate the tunneling current would be very useful. Unfortunately, this ability is seriously compromised by the following factors, among others:

1. The profile of the potential barrier across the gate stack, V(x), cannot be known, because the composition of the layers of the gate stack is complicated by interlayer diffusion and chemical reaction, which in turn determine the energy bands and the tunneling mass.
2. It is not clear what mass one is to use for the free carriers in the tunneling equations. Even if the usual effective mass is applied, this entity is not accurately known for most of the high-k dielectrics.

2.6.5.2 Carrier Confinement in the Strong Inversion and Accumulation Layers

The strong inversion layer, i.e. the MOSFET channel, could be a few nm thick; hence for an electron (hole) inside this potential well, i.e. for $E < V_0 = q\varphi_s$, the movement in the x-direction (i.e. perpendicular to the interface) is restricted, although parallel to the surface, the electron movement is free and the 2D (2-dimensional) periodicity of the semiconductor crystal is retained. This restriction in the x direction leads to a quantization of the eigenstates of the conduction or the valence band, ultimately resulting in sub-bands inside the potential well, separated by regions of forbidden energy; cf. Fig. 2.19; however, for electron energies higher than $V_0 = q\varphi_s$, the usual distribution of states in the band remains basically

unaltered. In other words, an electron ($E > V_0 = q\varphi_s$) inside the potential well is less delocalized than the channel electrons with $E > V_0 = q\varphi_s$, which are still treated as a free electron gas. The Schrödinger equation for an electron with $E < V_0 = q\varphi_s$ could be expressed as:

$$\left[-\frac{\hbar^2}{2}\left(\frac{1}{m_x^*}\frac{\partial^2}{\partial x^2}+\frac{1}{m_y^*}\frac{\partial^2}{\partial y^2}+\frac{1}{m_z^*}\frac{\partial^2}{\partial z^2}\right)+V(x)\right]\Psi(\mathbf{r})=E\Psi(\mathbf{r}) \tag{2.57}$$

where V(x) is the electrostatic potential energy in the strong inversion or the accumulation layer and is a function of x only. A solution of (2.57) could be expressed as:

$$\Psi(\mathbf{r})=\phi_i(x)\exp(ik_y y+ik_z z) \tag{2.58}$$

The x component of the bound electron is obtained by solving the equation:

$$\left(-\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2}-V(x)\right)\phi_i(x)=\varepsilon\phi_i(x) \tag{2.59}$$

For the electrostatic potential energy V(x), the boundary conditions are: $V(x = 0) = -q\varphi_s$ and $V(x = \infty) = 0$. The electrostatic potential energy $V(x) = -q\varphi(x)$, and as demonstrated in Sect. 2.2.4, the electrostatic potential $\varphi(x)$ (i.e. the band bending) are obtained by solving the Poisson equation, in which the net charge density $\rho(x)$ will take the expression:

$$\rho(x)=q\left(-\sum_i n_i|\phi_i(x)|^2+\sum N_D^+-\sum N_A^-\right) \tag{2.60}$$

where n_i is the electron density having the i-th eigenenergy ε_i. The two relations (2.59) and (2.60) are coupled requiring a self-consistent solution of the Schrödinger and the Poisson equations [37, 38]. The treatment of the quantization of the accumulation layer is similar to that of the strong inversion layer, except the following item. A weak inversion layer and a depletion layer separate the neutral region from the strong inversion layer, which is not the case in accumulation; consequently, the free electrons (or holes) have also to be included in the expression for the space charge density $\rho(x)$ in addition to those in the sub-bands.

Following are among the significant consequences of the quantization of the strong inversion and the accumulation layers:

1. Quantization significantly alters the electron (hole) density profile n(x)/p(x) at the semiconductor surface in strong inversion and accumulation, as illustrated in Fig. 2.22. In the case of the 3-dimensional (3D) Bloch wave representation, i.e. in the classical analysis, the electron (hole) density is determined solely by the energy separation between the band-edge (E_c or E_v) and the semiconductor Fermi level E_{FS} at the interface (x = 0); hence, the electron (hole) density peaks at the interface, cf. Fig. 2.22. However, in the 2D representation, the

wave function is required to vanish at its nodes, see Fig. 2.23, which includes the interface and the other perimeter of the potential well. This approach is similar to the text-book treatment of a particle in a box in one dimension [39], where infinite potential barriers are assumed at the boundaries of the box, cf. Fig. 2.24. The same assumptions of infinite potential barriers have been made in most of the treatments of carrier confinement in strong inversion and accumulation layers [37, 40]. In the 2D representation, the electron density does not peak at the interface, but away from the interface and inside the potential well, cf. Fig. 2.22.

2. The free carrier charge density of the MOSFET channel (strong inversion layer) or the accumulation layer is reduced, resulting in a lower space charge density Q_{sc} and therefore a lower space charge capacitance density C_{sc} for the same value of the surface potential (or band-bending). The gate dielectric capacitance density C_{di} is also said to effectively reduce on the basis of the following argument: The gate dielectric separates equal and opposite charges on its two sides. Electric field lines emanate from the metal surface and terminate in the semiconductor space charge layer or the vice versa. The effective separation between the metal surface charges and the semiconductor space charge layer charges increase, see Fig. 2.22, leading to an effective increase in the capacitive dielectric thickness. As a result, the total saturated capacitance of the MOS structure, C, is reduced.
3. The effective semiconductor band-gap is increased in strong inversion and in accumulation, as the first sub-band ε_0 lies above/below the band edge E_c/E_v by a significant amount, cf. $\Delta\varepsilon$ in Fig. 2.23.

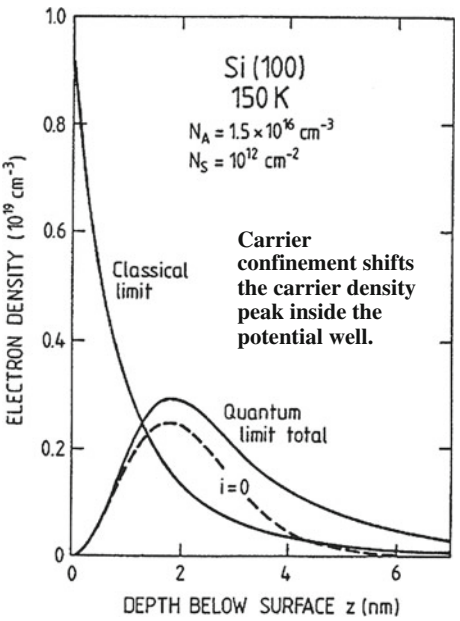


Fig. 2.22 Electron density profile n(x) in the strong inversion layer for a silicon substrate at 150 K with (100) orientation, acceptor density of 1.5×10^{16} cm^{-3} and a total electron density of 10^{12} cm^{-2}. The *broken line* represents the contribution to n(x) of the lowest sub-band alone. Adapted from [37]

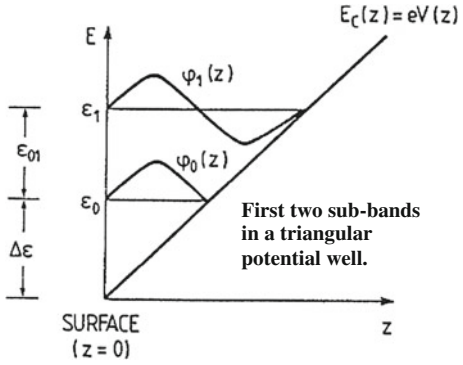


Fig. 2.23 Schematic illustration of the electron wave functions ϕ_0 and ϕ_1 for the lowest two sub-bands ε_0 and ε_1, for a triangular potential well approximation of the strong inversion or the accumulation layer. Adapted from [38]

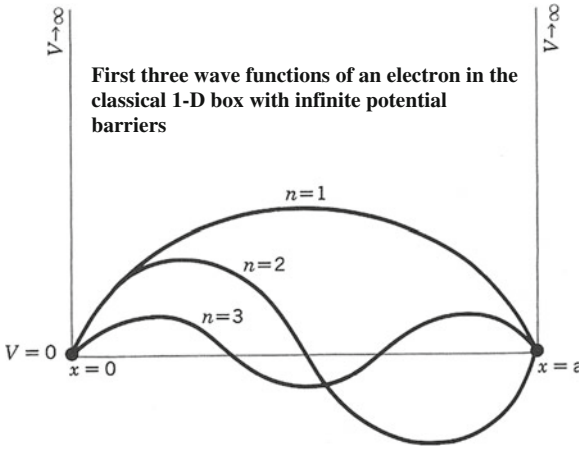


Fig. 2.24 Schematic representation of the first three wave functions for an electron confined to a one-dimensional box. Adapted from [35]

There exists experimental evidence for the validity of the 2D representation of the strong inversion and the accumulation layers and the resultant sub-bands, from infrared absorption and magneto-conductance experiments. However, there are good reasons to believe that many of the quantization and 2D treatments [37, 41], which appear to be very popular, overestimate the effects of the carrier confinement. One reason for the popularity could be that these estimates or calculations predict an EOT, which is significantly lower than what the physical dimensions would allow; *this—much lower estimate of the EOT*—would appear to be a success and a progress for realizing the target set by the ITRS roadmap. The carrier confinement effects may be overestimated for the following reasons:

1. *Electron wave function penetration*—As mentioned already, the potential barrier was assumed to be infinite at the boundaries of the potential well, see Fig. 2.24, to simplify the mathematical treatment. For the high-k gate stack with much lower band offsets than SiO_2, and, generally, with much smaller effective mass, this assumption is hard to justify [It may be noted that the wave

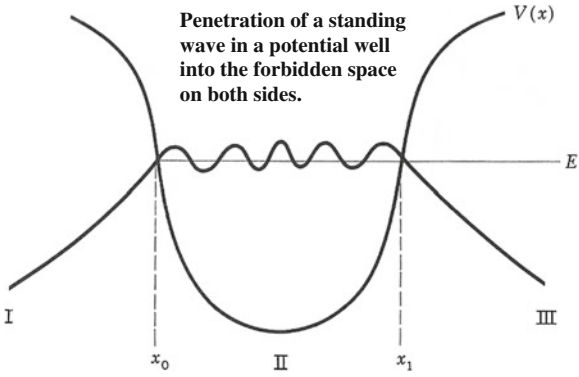


Fig. 2.25 Penetration of the bound state E (standing wave) in a potential well into the classically forbidden regions outside the potential well (exponentially attenuating wave). Adapted from [35]

attenuation constant depends upon the product of the effective mass and the barrier height, cf. (2.55)]. Penetration of the bound states (in classically allowed regions) into both sides of the classically forbidden regions is a text book problem in quantum mechanics, see Fig. 2.25. However, it is only recently, that penetration of the standing waves in the strong inversion or the accumulation layer into the gate stack has been looked into [41]. Calculations indicate significant effect of the wave function penetration in enhancing the electrical oxide thickness: by as much as 0.33 nm or more [42]. Major problems in correctly estimating the extent of wave function penetration into the gate stack include the unknown potential barrier profile in the gate stack and the corresponding tunneling mass, as mentioned in Sect. 2.6.5.1.

2. *Metal induced states*—Another factor which may dilute the carrier confinement effect is the possibility of metal induced states in the semiconductor sub-surface, when EOT is <1 nm. A metal-induced gap state is an old concept [43], which was invoked in the case of the metal–semiconductor interface. In fact, when EOT is very small, interference between wave functions of carriers at the semiconductor surface and at the metal surface may be a significant possibility; such an interaction may significantly alter the properties of the gate stack.
3. *Very large tunneling currents*—For EOT, say = 0.5 nm, the tunneling current may be as large as a significant fraction of the drain current, and the tunneling time (If that concept is valid) may be of the order of the channel traversing (transit) time and also the semiconductor relaxation time. In such a case, it is not known what the Fermi occupancy of the sub-bands would be and which Fermi level—the semiconductor or the metal—would apply to these sub-bands. The large tunneling current itself will dilute the carrier confinement phenomenon.

2.6.5.3 Wave Function Penetration into the Gate Stack

Penetration of an incident electron wave with energy E into a region of higher potential energy V_0 is classically forbidden, but is quantum-mechanically allowed, as illustrated in Fig. 2.20. It does not matter quantum-mechanically, if the barrier is infinite in thickness, as Fig. 2.20 would suggest. However, there would be no penetration if the barrier energy is infinite. As already outlined, the concept of wave-function penetration is almost as old as quantum mechanics itself. If the wave function ψ exists inside the potential barrier, $\psi\psi^*dV$ would be finite, howsoever small it might be; therefore there would be a finite probability of finding the electron inside the barrier. However, it does not appear that electron/hole traps were contemplated at that time to exist inside the forbidden potential barrier; for, such an entity would also require the corresponding trap energy level (i.e. an eigenstate) to exist. If there are no allowed energies inside the potential barrier, it is not clear in the classical treatment of wave function penetration, how one would find the electron inside the barrier. The issue of traps, the occupancy of these traps, and the related pseudo-Fermi function was investigated many years later [12], but this lone treatment of these issues was not followed subsequently. For a potential barrier profile, where the potential energy is a function of x, it would follow from (2.55), that the electron density attenuation $A = |\psi|^2$ at a distance x from the point of incidence would be given by:

$$A = \exp -2\sqrt{\frac{2m^*}{\hbar^2}\int_0^x \sqrt{\phi(x)}dx} \tag{2.61}$$

where ϕ is the barrier height.

2.6.6 Trap Time Constant

As already outlined, the high-k gate stack is beset with a high density of traps inside its bulk layers and at the interfaces between the gate stack layers, cf. Fig. 2.26. Figure 2.26 represents the energy band diagram of a high-k gate stack on a p-Si at the onset of strong inversion. This diagram represents schematically the traps at various locations inside the gate stack—at the Si-IL interface, inside the IL layer, at the IL/high-k interface, inside the high-k layer, and at the high-k/metal interface. Indicated in this diagram are the physical thicknesses of the various layers of the gate stack and the different components of the gate stack potentials. The charges in the gate stack, particularly when these are high as inside the high-k layer, will make the potential profile non-linear; for the sake of simplicity, this point has been ignored in Fig. 2.26, and the indicated potential profiles are linear. An important feature illustrated in Fig. 2.26 is the pseudo-Fermi level, cf. Fig. 2.14; the pseudo-Fermi level is meant to indicate the trap occupancy function. As indicated in Fig. 2.26, the pseudo-Fermi function and the trap

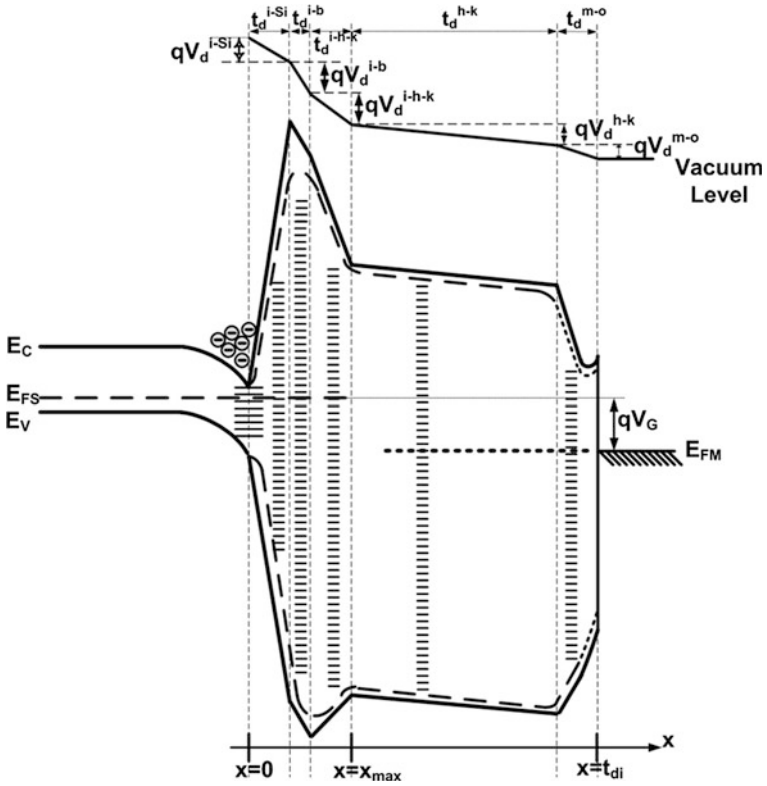


Fig. 2.26 Energy band diagram of an MOS structure with a high-k gate stack, consisting of bulk intermediate oxide and bulk high-κ layers and chemically graded layers at the three interfaces, illustrating the effects of the graded band-gaps and the electric field. i-Si and i-h-k are respectively the transition layers between Si and IL and IL and high-k layers. i-b and h-k are respectively the bulk IL and high-k layers. m-o is the transition metal oxide layer between the high-k layer and metal electrode. The effect of the charges in the gate dielectric layers, on the potential profile (i.e. variation of the potential along direction x), has not been represented (The actual potential profile will be non-linear). The broken profile schematically represents the effect of the image force on the potential and the energy barrier profile. Figure 2.26 illustrates the formation of a Schottky barrier at the high-k/metal interface

occupancy are given by the semiconductor majority carrier imref $E_{FS,h}$ up to a distance x_{max} into the gate stack, whereas in the rest of the gate stack it is given by the metal Fermi level E_{FM}.

The gate stack traps are likely to capture and emit electrons or holes, depending upon the bias conditions and the signal frequency. How deep into the gate stack and how many of these traps exchange free carriers with either the semiconductor substrate or the metal electrode, will depend upon the potential barrier profile of the gate stack, including the barrier energy and the physical thickness. For a low EOT, such as 0.5 nm, it is in principle possible for traps at all locations inside the gate stack to communicate either with the semiconductor energy bands or with the

metal energy bands, cf. Fig. 2.26. There may be traps of different physical and chemical nature with their characteristic capture and emission probabilities and relaxation times. The capture or emission time constant of a trap may carry its signature or reflect its identity.

The electron or the hole capture (electron/hole emission is a different process from electron/hole capture) is a complex process, and there is no comprehensive theory for the capture probability of a trap inside a forbidden region (dielectric or an insulator). In a simple formulation, the capture probability is equated to $v_e\sigma_e$ or $v_h\sigma_h$ (v_e/v_h is thermal velocity, and σ_e/σ_h is capture cross-section for electron/hole), which has the unit of cm^3/s. The capture cross-section remains an ambiguous concept; it hides our inability to formulate a clear set of relations for the capture process. Experimental results [44] suggest that the capture cross section can vary over many orders of magnitude, such as from 10^{-12} to 10^{-18} cm^2. Can one explain such an enormous variation of the capture cross-section? The usual explanation offered is that scattering by Coulomb attraction entails the largest and by Coulomb repulsion the smallest capture cross-sections.

The capture cross-section of a trap would likely depend upon its neighborhood, which may change strongly along the x direction inside the gate stack. So, the capture cross-section may change strongly with the trap location inside the gate stack, x_t; it may also be strongly dependent upon the trap energy (Modeling [45] indicates, even for one kind of defect, for example for oxygen vacancies, trap levels at different energies with different effective charges.).

In the simplest formulation, one can argue that the capture time, for a trap at the location x_t inside the gate stack, will be inversely proportional both to the density of the free carriers (electron or hole) at x_t and also to the capture probability, cf. Fig. 2.27. The time for a hole capture by the trap located at x_t is then given by:

$$\tau_t^h(x_t) = \frac{1}{v_h \sigma_h p(x_t)} \tag{2.62}$$

where $p(x_t)$ is the hole density at the trap's location.

Since the wave-function of an electron or a hole at the silicon surface (i.e. $x = 0$) can penetrate the potential barrier presented to it by the gate stack, there is a non-zero probability of finding the electron/hole at any trap location x_t, cf. Fig. 2.27. The hole density at x_t is determined by the electron wave function attenuation constant κ_t [12]:

$$p(x_t) = p_s e^{-2\kappa_h x_t}; \quad \kappa_h x_t = \sqrt{\frac{2m_h}{\hbar}\int_0^{x_t}\sqrt{\phi_h(x)}dx} \tag{2.63}$$

where p_s is the hole density at the Si surface, m_h^* is the effective tunneling mass of holes, and $\phi_h(x)$ is the potential energy barrier profile for holes in the gate stack, as defined by the valence band edge, the electric field, and perhaps the image force barrier lowering.

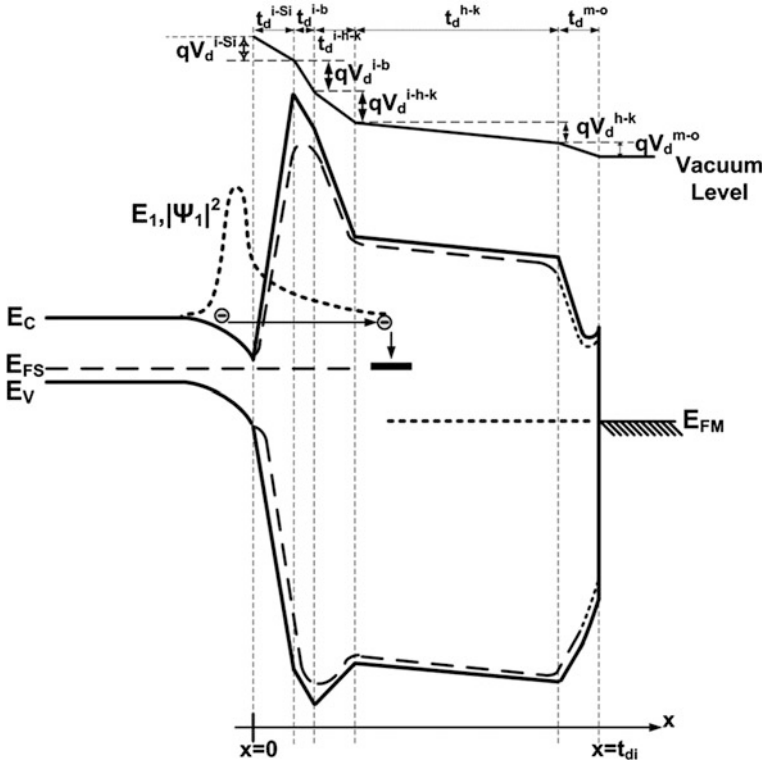


Fig. 2.27 Energy band diagram of an MOS structure with a high-k gate stack, consisting of bulk intermediate oxide and bulk high-κ layers and chemically graded layers at the three interfaces, illustrating the attenuation of the silicon-surface electron wave-function, and the electron capture at a trap located in the gate stack. The *long dashed line* is an indicative representation of the image force barrier lowering. The *short dashed line* is an indicative representation of the electron density of the first sub-band electrons in the strong inversion layer and inside the gate stack

It may be useful to have a quantitative feel for the carrier density attenuation in the gate stack; for example, for a rectangular barrier with $\phi_h = 2\,\text{eV}$ and $m_h^* = 0.18$ m, $(1/2\kappa_h) = 0.16$ nm, cf. (2.63), which will mean that for each 0.16 nm of barrier thickness, the hole density will be attenuated by e^{-1}. If one considers the combined effect of the electric field and image-force barrier-lowering, the value of $(1/2\kappa_h)$ will be higher.

The occupancy of a trap inside the gate stack under a dc bias will depend upon the bias, the potential barrier profile between the trap location and the free carrier source, i.e. the semiconductor or the metal surface, cf. Fig. 2.27, the tunneling mass, and the trap capture or emission cross-section. In the presence of an ac signal, charging and discharging would occur, giving rise to a trap capacitance. For the charging or the discharging of the trap to take place, the trap time constant τ_t has to be much smaller than the inverse angular frequency ω^{-1} of the ac signal. For a certain surface carrier density p_s/n_s, there would be a maximum penetration

depth, x_{max}, such that all traps in the gate stack in the range of 0 and x_{max} would follow the applied small signal. This maximum penetration depth x_{max} would be higher for a lower signal frequency. Further, as the semiconductor surface carrier density increases with increasing accumulation or strong inversion, x_{max} would be strongly bias dependant.

To get a feel for this maximum penetration depth x_{max}, let us consider a 100 kHz signal and a surface hole density of 10^{20} cm^{-3} (This magnitude of hole density would represent a p-type semiconductor in deep accumulation or a n-type semiconductor in deep inversion.). For the traps to follow the applied signal, the condition ω (=$2\pi f$) $\ll$ $(\tau_t^h)^{-1}$ has to be fulfilled. Assumption of a hole capture cross-section of 10^{-15} cm^2, and a hole density of 10^{16} cm^{-3} at the trap location x_t, and a hole thermal velocity of about 10^7 cm/s at 300 K, leads to a trap time-constant of 10 ns, cf. (2.62). Hence, the traps should be able to follow the 100 kHz signal under these conditions. So, for a 100 kHz signal, and a rectangular barrier of 2 eV and $m_h^* = 0.18$ m, according to (2.63)

$$x_{\max} = \frac{\ln \frac{p_s}{p}}{2\kappa_h} = 0.16 \ln \frac{10^{20}}{10^{16}} \text{ nm} = 1.47 \text{ nm}$$

Similarly, for a 10 kHz signal, x_{max} would be 1.84 nm, assuming $(2\kappa_h)^{-1} = 0.16$ nm. An x_{max} of 1.47 or 1.84 nm would mean that, in the case of high-k gate stacks with EOT = 1.0 nm or less, all traps in the gate stack may follow the 100/10 kHz signal, either through communication with the semiconductor energy bands or through the same with the metal electrode energy bands, cf. Figs. 2.14 and 2.19.

In the case of the ultrathin gate stacks, it is not necessary that the traps inside the gate stack exchange free carriers only with the semiconductor surface. Free carriers are available on the metal surface as well; hence free carriers on the metal surface will compete with the free carriers on the semiconductor surface to fill/empty (i.e. charge/discharge) the gate stack traps. Which exchange will dominate will depend upon the carrier density at the two surfaces and the rate of tunneling. What this means is that the quasi-Fermi occupancy of the traps between $x = 0$ and $x = x_{max}$ will be given by the semiconductor quasi-Fermi level E_{FS}, while the quasi-Fermi occupancy of traps in the range of $x = x_{max}$ and $x = t_{di}$ would be given by the metal E_{FM}, as illustrated in Fig. 2.26. In other words, depending on the bias, a part of the gate stack is in quasi-equilibrium with the semiconductor surface, and the rest of the gate stack is in quasi-equilibrium with the metal surface. As the semiconductor surface carrier density is strongly bias-dependent, while the metal surface carrier density is constant (depends on the metal), x_{max} will be bias-dependent. In deep accumulation or in deep inversion, x_{max} is likely to be in the vicinity of the plane interfacing with the intermediate layer and the high-k layer, as has been indicated in Fig. 2.26. Consequently, the bulk semiconductor imref E_{FS} and the pseudo-Fermi function will extend into the high-k gate stack up to a distance x_{max} from the silicon surface, while in the rest of the gate stack, the Fermi occupancy will be given by the metal Fermi level E_{FM} (cf. Figs. 2.26, 2.27).

2.7 Nature of Traps and Charges in the High-k Gate Stack

Gate stack traps may be defined as electrically active defects in the gate stack which manifest themselves by capturing or emitting electrons or holes. The traps may be characterized by the localized trap level (an allowed state), its capture and emission cross-section, and its charged state. It is not necessary for the trap energy level to be located inside the band-gap; the trap levels could be located inside the allowed energy bands of the host dielectric. It is not uncommon to come across a perception that the trap states are required to be inside the band-gap. The only real difference is that if the trap state is inside an allowed band, then it supplements the allowed states of the host, whereas if it is inside the band-gap, then it exists where the host has a zero density of states. The charge of a trap may or may not vary with the applied gate voltage; if it does not, then it will act as a fixed charge, and will not contribute to any trap capacitance; if it charges or discharges with the gate voltage, then a trap capacitance will result. As analyzed in Sect. 2.6.4, whether charging or discharging happens at the trap will depend upon whether the trap occupancy changes with the gate voltage.

An unusually high trap density inside the gate stack, as currently obtains even in the device quality high-k gate stacks, has serious implications which have not received much attention so far. We have already made a clear distinction between a dielectric or a dielectric stack, which is free of charges inside, and the high-k gate stack in reality, whose capacitance is not a dielectric capacitance, but is a space charge capacitance. We may recapitulate that a near-ideal dielectric like the SiO_2 has no significant charges inside it with the result that the electric field is constant and the potential is a linear function of distance inside it. As discussed in Sect. 2.6.6, it is possible for all the traps inside the gate stack to communicate with the semiconductor or the metal surface and exchange electrons or holes (i.e. charge or discharge), if the EOT is small (say <1.0 nm), the frequency is 100 kHz or even 1 MHz, and the device is biased in strong accumulation or inversion, cf. Figs. 2.26 and 2.27. In such a situation, as we will see in detail in later sections, the trap capacitance (due to trap charging or discharging) will add to the dielectric capacitance of the gate stack, yielding a value of C_{di} higher than what would obtain in the GHz range—the operating frequency of the MOSFET, cf. Fig. 2.28. At GHz frequencies, the traps inside the high-k gate stack are unlikely to exchange electrons or holes with either the semiconductor or the metal surface. In other words, the parameter extraction technique (normally carried out at 1 MHz or below) would yield a higher value of the gate stack capacitance C_{di} than what would obtain during the MOSFET operation and would overestimate its performance. It would also give an erroneously lower value of EOT or CET and induce a false sense of progress on downscaling of the gate stack thickness.

As already mentioned, the topic of surface states and interface states have engaged many researchers over several decades, beginning with William Shockley, Igor Tamm, John Bardeen, Walter Schottky, and Volker Heine. Igor Tamm was perhaps the first to realize the possibility of localized states existing at the

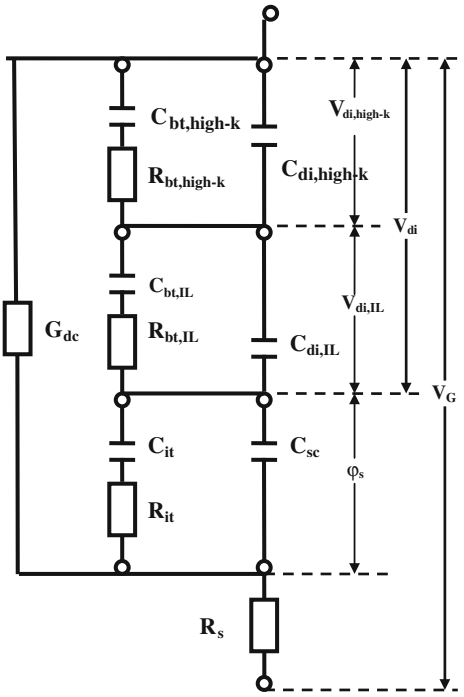


Fig. 2.28 Equivalent circuit diagram for the MOS structure in accumulation and strong inversion, at an intermediate frequency, illustrating the effect of charging/discharging in traps located inside and at the interfaces of the two layers constituting the high-κ gate stack

surface. The classical concepts of Tamm states [46] and Shockley states [47] are generally not invoked in the case of the high-k gate stacks. Theoretical and even experimental information is incomplete on the nature of traps inside the high-k gate stack. From common sense one may expect to find different types of intrinsic traps in the different regions of the high-k gate stack, because the electronic properties of the trap would primarily be decided by its nearest neighborhood:

1. *Si-IL* (*SiO_2-like*) *interface*—Transition region from a covalent semiconductor to a primarily covalent insulator, cf. Fig. 2.26. The interface traps, investigated most, experimentally, are those existing at the Si-SiO_2 interface. The experimental information on the characteristics of these traps may be considered both reliable and complete. However, the theoretical understanding and identification of the origin of these traps are incomplete notwithstanding several hypotheses which exist on their origin. From the experimental results, the device grade Si–SiO_2 interface appears to exhibit the following interface state distribution across the Si band-gap: two peaked profiles overlying a U-shaped background. There are good reasons to believe that some of the intrinsic traps at the Si-IL interface may be the P_b center with the structure of $Si \equiv Si_3$—amphoteric trap which is a threefold-coordinated Si with a dangling bond [48, 49]. An amphoteric trap can be positively charged, when it has no electron, and is a donor state, neutral, when it has one electron, and negatively charged, when it has two electrons, and is an acceptor state. Experiments suggest that the peak

in the lower Si band-gap (around 0.30–0.35 eV above E_v) represents the donor P_b center, while the peak in the upper Si band-gap (around 0.80–0.85 eV above E_v) represents the acceptor P_b center. Less certain is the origin of the U-shaped background $D_{it}(E)$ distribution. Possible origin could be the tail states of the conduction band and the valence band, like in amorphous silicon [50], and Si-IL strain-induced states. In addition to the intrinsic traps, chemical impurities can generate characteristic traps. Experiments by Kar and Dahlke [51] have convincingly demonstrated that metal impurities from the gate electrode can generate metal-specific trap levels with characteristic capture cross-section. Results presented in many chapters suggest the possibility of high-k cation being present at the Si-IL interface.

2. *IL bulk*—When the semiconductor is Si, the intermediate layer (IL) is SiO_2 or SiON. The dry thermal SiO_2 layer normally should be defect-free. But, according to Bersuker, gate stack degradation experiments reveal the precursor defects located in the IL layer, cf. Fig. 2.26, to be the main concern for high-k gate stack reliability, cf. Chap. 8. These may be oxygen vacancies, with trap levels located 2.2–3.5 eV below the silica conduction band [52], induced by the interaction of the intermediate layer with the high-k/metal layers. Stress-induced traps are found to be generated in these precursor defects.
3. *IL/high-k interface*—Transition region from a primarily covalent insulator to a highly ionic insulator with strong mismatch in terms of ionicity (electro-negativity), coordination number, and lattice constant, cf. Fig. 2.26. This multi-faceted mismatch can in principle generate a high density of interface traps of various kinds. Unfortunately, it is difficult to access this interface by the interface trap extraction techniques such as the MOS conductance technique and also the charge-pumping technique. However, the limited experimental results suggest a high density of traps at the IL/high-k interface and also indicate that the trap density increases with distance from the Si surface. Experimental results presented in Chaps. 6 and 5 confirm the presence of a strong interface dipole, cf. Fig. 2.14. The origin of this dipole could be the areal oxygen density difference [53] or the electro-negativity difference [54] across the IL/high-k interface.
4. *High-k bulk*—Ionic oxides are always rich in oxygen vacancies; intrinsic defects are likely to be dominated by oxygen vacancies. Extrinsic defects due to diffusion from the semiconductor (e.g. Si) or from the metal are possible. As has been discussed in Chap. 8, in monoclinic hafnia, oxygen vacancies have been shown by ab initio calculations to exist in five charge states, from -2 to +2 and may function as electron/hole traps, as well as fixed charges. The IL (SiO_2) layer may be oxygen deficient at the Si/IL interface and also at the IL/Hafnia interface due to the diffusion of oxygen vacancies from the hafnia layer. The results presented in Chap. 8 indicate the Vo^{++} traps to have an energy level about 0.5 eV below the hafnia conduction band, a capture cross-section of 10^{-13} cm^2, a time constant of 0.5 μs, and an extremely high areal density of 10^{14} cm^{-2}.

5. *High-k/metal interface*—As already mentioned, one does not talk of interface traps at any SiO_2/metal interface. But a high-k material is not a near-perfect dielectric as SiO_2 is. The concepts of Fermi level pinning and charge neutrality layer have often been invoked recently to analyze the anomalous behavior at the high-k/metal interface. Both of these concepts will not apply unless the interface trap density is extremely high. The moot question remains what could generate these traps. The most unfortunate part is that normally a high trap density should be easier to measure directly. But there is no trap extraction technique available which can be applied on a metal surface. So, all the evidence on these traps is indirect. One possible origin could be tails of the metal wave functions which decay into the high-k layer. In the case of the metal-semiconductor contacts, as these mid-gap states penetrate some distance into the semiconductor, one treats these states as mixtures of the valence and conduction band wave functions with the metal wave functions [55].

2.8 Potentials and Circuit Representations of the High-k Gate Stack

The energy band diagram of Fig. 2.26 illustrates five components of the gate stack potential V_{di}. Three of these potential components are across the three interface regions—Si/IL, IL/high-k, and high-k/metal—in which the chemical composition has been considered to be graded. This may be a more realistic representation, but, the mathematical representation of the electrostatic potential across a layer with a graded composition is too complicated; in particular, the permittivity of such a graded layer will be graded also, which will be difficult to handle. To simplify the equations, we will merge the three interfaces into the two bulk layers—IL and high-k—and consider constant permittivity in each of these layers. Such a less complex situation is illustrated in the energy band profile in Fig. 2.14. With this simplification, the total potential across the gate stack, V_{di}, will have two components, across the IL, $V_{di,IL}$, and across the high-k layer, $V_{di,high\text{-}k}$:

$$V_{di} = V_{di,IL} + V_{di,high-k} \tag{2.64}$$

Each of these potentials will depend upon the silicon space charge density Q_{sc}, the charge density in the interface states at the Si-SiO_x interface Q_{it}, the fixed charge density in the proximity of the silicon surface Q_F, the total charges in each of the gate stack layers preceding the layer in question, the charges in the layer itself, the dielectric capacitance of the layer and its permittivity, and the physical thickness of the layer. The potential across the bulk intermediate-oxide layer, $V_{di,IL}$, may be expressed as:

$$V_{di,IL} = -\frac{Q_{sc} + Q_{it} + Q_F}{C_{di,IL}} - \int_0^{t_{di,IL}} \frac{\rho_{IL}(x)dx}{\varepsilon_{di,IL}}\left(t_{di,IL} - x\right) \tag{2.65}$$

where $C_{di,IL}$ is the dielectric capacitance, ρ_{IL} is the volume charge density, $\varepsilon_{di,IL}$ is the permittivity, and $t_{di,IL}$ is the thickness of this layer. The potential across the bulk high-k layer, $V_{di,high\text{-}k}$, may be expressed as:

$$V_{di,high-k} = -\frac{Q_{sc} + Q_{it} + Q_F + \int_0^{t_{di,IL}} \rho_{IL}(x)dx}{C_{di,high-k}} - \int_{t_{di,IL}}^{t_{di}} \frac{\rho_{high-k}(x)dx}{\varepsilon_{di,high-k}}(t_{di} - x) \tag{2.66}$$

where $C_{di,high\text{-}k}$ is the dielectric capacitance, $\rho_{high\text{-}k}$ is the volume charge density, and $\varepsilon_{di,high\text{-}k}$ is the permittivity of this layer. The dielectric capacitance of a layer is its capacitance if it were charge-free, i.e. an ideal dielectric. This is the same as the plane parallel capacitance. The dielectric capacitances of the two layers of the gate stack may be expressed as:

$$C_{di,IL} = \frac{\varepsilon_{di,IL}}{t_{di,IL}}; \quad C_{di,high-k} = \frac{\varepsilon_{di,high-k}}{t_{di,high-k}}; \tag{2.67}$$

Just as the expressions for the potentials across the gate stack layers, a valid representation of the equivalent circuit of the gate stack is enormously more complicated than that in the case of the SiO_2 gate dielectric, as illustrated by Fig. 2.28. In this representation, R_s is the series resistance representing the entire device region outside of the silicon space charge layer and the gate stack, and is an important element in the case of large gate leakage currents. C_{sc} is the traditional silicon space charge layer capacitance, C_{it} is the traditional capacitance of the traps at the Si-SiO_x interface, R_{it} is the traditional Si-SiO_x interface trap resistance, and φ_s is the surface potential. G_{dc} represents the flow of the direct current through the gate stack [51]. Charging and discharging in traps in each of the two bulk layers of the gate stack and at two of the three interfaces has been represented by an R_tC_t combination. It may be noted that there will be no charging or discharging of the traps on the metal surface (i.e. at the high-k/metal interface), as the metal Fermi level is bias invariant. An important issue in the case of charging/discharging of a trap inside the gate stack is where the free carrier is being supplied from, i.e. the silicon surface or the metal surface. If the free carrier exchange of the trap is with the silicon surface, then the R_tC_t branch is connected to the silicon majority carrier band; if the free carrier exchange of the trap is with the metal surface, then the R_tC_t branch is connected to the metal, cf. Fig. 2.28.

The circuit representation of Fig. 2.28 can be simplified, depending upon the bias and the small signal frequency. At a signal frequency too high for charging/discharging at any trap in the gate stack to follow, f_h, the circuit representation of the gate stack can be simplified to what is illustrated in Fig. 2.29. The circuit representation on the right side in Fig. 2.29b is close to the traditional

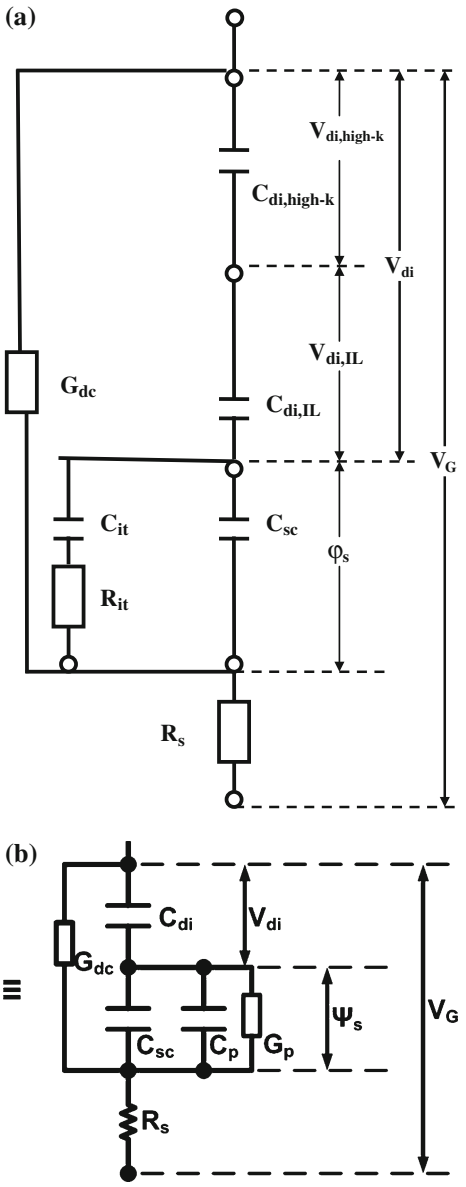


Fig. 2.29 **a** Reduction of the circuit representation in Fig. 2.28 at a high frequency, which no trap in the gate stack can follow. **b** Transformation of **a**

representation, except that the total gate stack capacitance C_{di} is now a series sum of two plane-parallel capacitors:

$$\frac{1}{C_{di,hf}} = \frac{1}{C_{di,IL}} + \frac{1}{C_{di,high-k}} \tag{2.68}$$

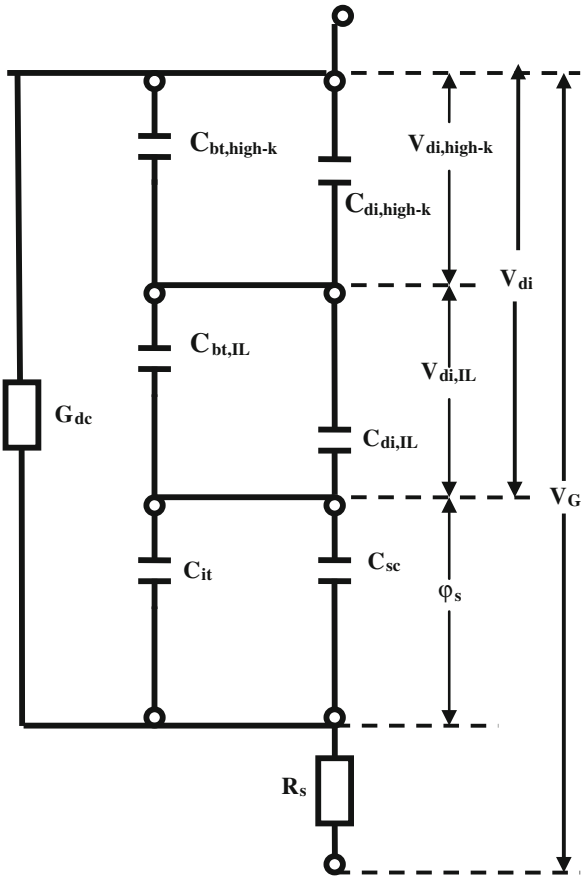


Fig. 2.30 Reduction of the circuit representation in Fig. 2.28 at a low frequency, which all traps in the gate stack and the interface traps can follow

This is likely to be case, for all bias values, at the MOSFET clock frequency of a few GHz. At a signal frequency low enough for charging/discharging at every trap in the gate stack to follow, f_l, the circuit representation of the gate stack can be simplified to what is illustrated in Fig. 2.30, in which case the total MOS capacitance would be given by:

$$\frac{1}{C_{lf}} = \frac{1}{C_{sc} + C_{it}} + \frac{1}{C_{di,IL} + C_{bt,IL}} + \frac{1}{C_{di,high-k} + C_{bt,high-k}} \tag{2.69}$$

2.8.1 *Flat-Band Voltage Characteristics of High-k Gate Stack*

SiO_2 Single Gate Gate Dielectric.—As already outlined, the flat-band voltage V_{FB} is a very frequently used indicator of the MOS device quality; the use of V_{FB}

permeates most of this book. The flat-band voltage characteristics, in particular, its variation with the gate dielectric thickness or the EOT, is also very useful for parameter extraction. In the case of the SiO_2 single gate dielectric, the V_{FB} versus the t_{di} characteristic, as represented by (2.15) and reproduced below:

$$V_{FB} = -\frac{Q_{it,fb} + Q_F}{\varepsilon_{di}} t_{di} - \phi_{MS} \tag{2.15}$$

has been used very effectively to extract the silicon-metal work function difference ϕ_{MS} and the interface charge density at flat-band ($Q_{it,fb} + Q_F$) [9, 10]. The $V_{FB}(t_{di})$ plots were found to be straight lines over the entire dielectric thickness range, the intercept of which with the $t_{di} = 0$ line yielded the work-function difference ϕ_{MS} [10] and the slope of which yielded the interface charge at flat-band [9]. Wafers with graded SiO_2 thickness was used in the study [9, 10], which perhaps was instrumental in producing high quality straight-line characteristics. Some points are worth noting in this context. Often, the interface charge density at flat-band is mistakenly taken as the fixed charge density Q_F; this is erroneous, as for the device quality SiO_2 layers, $Q_{it,fb}$ and Q_F are of the same order of magnitude. The other point is that if Q_F and the interface traps are invariant of the silicon doping, which is generally assumed to be true, then, $Q_{it,fb}$ for n-Si will be different from $Q_{it,fb}$ for p-Si, i.e. $Q_{it,fb}$ for n-Si will be more negative than $Q_{it,fb}$ for p-Si. Experimental results clearly revealed the $V_{FB}(t_{di})$ straight line for n-Si to have a smaller slope than the same for p-Si [9, 10]. Although Q_F has never been determined, as there exists no technique to do so, the common wisdom considers the fixed charge density Q_F to be positive in the case of the SiO_2 gate dielectric [3]; this will support the above experimental result.

High-k Gate Stack.—In contrast to the experience with the SiO_2 gate dielectric, in the case of even the device quality gate stacks, the experimental flat-band voltage V_{FB} versus the EOT characteristic is non-linear, and it becomes increasingly non-linear as the value of EOT becomes small; for EOT < 1.0 nm, V_{FB} rolls off, leading to the emergence of the phrase: "V_{FB} roll-off". Making the assumptions used for obtaining (2.65) and (2.66), we may express the flat-band voltage for a high-k gate stack as, cf. (2.15), (2.65), and (2.66):

$$V_{FB} = -\phi_{MS} - \frac{Q_{it,fb} + Q_F}{\varepsilon_{SiO_2}} EOT - \frac{\int_0^{t_{di,IL}} \rho_{IL}(t_{di,IL} - x)dx}{\varepsilon_{di,IL}} - \frac{\int_0^{t_{di,IL}} \rho_{IL} dx}{\varepsilon_{di,high-k}} t_{di,high-k} - \frac{\int_{t_{di,IL}}^{t_{di}} \rho_{high-k}(t_{di} - x)dx}{\varepsilon_{di,high-k}} \tag{2.70}$$

Even a casual look at (2.70) will suggest that, unless many parameters in (2.70) are insignificant, any plot of V_{FB} versus any dielectric thickness will not yield a linear characteristic, which can be used to extract ϕ_{MS} or any of the gate stack charge densities. Unfortunately, in the relation of (2.70), the largest magnitudes have those charges, which induce the largest non-linearity, namely, $\rho_{high\text{-}k}$ and ρ_{IL}.

We may recall that in our formulation, both $\rho_{high\text{-}k}$ and ρ_{IL} include both the bulk and the interface trap charges. In general, the charge density as well as the permittivity increase moving from the Si/SiO_2 interface to the SiO_2 bulk to the SiO_2/high-k interface and finally to the high-k bulk. Consequently the potentials induced by the higher charges get partly neutralized by the corresponding higher permittivity, cf. (2.70).

In many experiments with high-k gate stacks, the high-k layer thickness is kept constant, while a graded SiO_2 intermediate layer is used to vary the IL thickness and thereby the EOT. In such a case, one would expect the V_{FB}(EOT) characteristic to be less non-linear, since the Si/IL interface is Si/SiO_2 and only the thickness of the SiO_2 layer is being varied. Even in such a situation, a linear V_{FB}(EOT) characteristic is not obtained, as is illustrated in Fig. 2.31. The wafers of Fig 2.31 had a graded SiO_2 layer (1–6 nm) grown in dry O_2 at 900 °C. Sample D-04 had no HfO_2 layer; samples D-06 and D-12 had 2 nm and samples D-10 and D-14 had 3 nm thick HfO_2 layer. Samples D-06 and D-10 had a post-deposition annealing (PDA) in O_2 at 500 °C for 1 min.

The results of Fig. 2.31 will be discussed in more detail later, see Sect. 2.10.3.2. Some relevant points worth mentioning here are:

1. As there are no data in the EOT range of 0–2 nm, the V_{FB} roll-off feature cannot be observed.
2. The inaccuracy in the flat-band voltage V_{FB} can be of the order of a few mV. This could have partly contributed to the scatter and the non-linearity of the V_{FB}(EOT) characteristics, as the change in V_{FB} is in the range of 10–20 mV/nm.
3. For the samples of Fig. 2.31, the SiO_2 trap density obtained from the conductance technique at about 0.25–0.30 eV above E_v was rather low and varied between 0.6 and 2.6 × 10^{11} cm^{-2} V^{-1}, depending upon the SiO_2 layer thickness, see Sect. 2.10.3.2. The trap density increased with decreasing SiO_2 layer

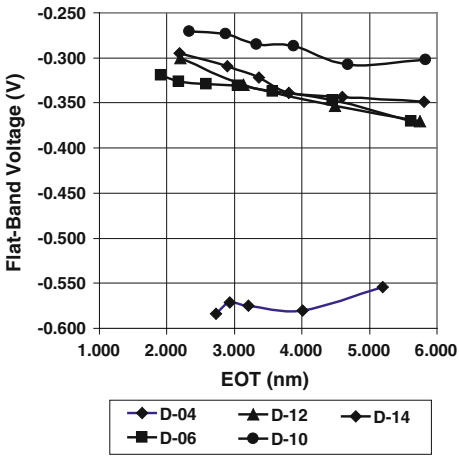


Fig. 2.31 Flat-band voltage versus EOT for p-Si/SiO_2/HfO_2/TaN MOS capacitors on different graded-SiO_2 wafers with different sets of PDA and HfO_2 thickness. The *lower most curve* belongs to wafer D-04 with no hafnia layer. The *upper four curves* belong to wafers with the hafnia layer [17]

thickness. As will be discussed in a later section, these traps may be located 0.4 nm inside the SiO_2 layer from the Si surface; this suggests that the traps at the Si surface and in the SiO_2 layer may depend upon the SiO_2 layer thickness. This would contribute to the non-linearity of the V_{FB}(EOT) characteristics in Fig. 2.31.

5. Wafer D-04 has a positive gradient, while the other wafers have a negative gradient, suggesting that the flat-band interface charge is positive in the presence of the hafnia layer, while it is negative without.
6. The approximate flat-band interface charge density obtained from the slope of the V_{FB}(EOT) characteristic in Fig. 2.31 was -2.1×10^{11} charges.cm^{-2} for wafer D-04, 3.2×10^{11} charges.cm^{-2} for wafer D-06, 4.1×10^{11} charges.cm^{-2} for wafer D-12, 2.1×10^{11} charges.cm^{-2} for wafer D-10, and 3.3×10^{11} charges.cm^{-2} for wafer D-14. As the HfO_2 layer thickness was not varied, these charges at flat-band are likely to include only the flat-band charges at the Si-SiO_2 interface and in the SiO_2 layer, if we assume that the charges in the HfO_2 layer and at its two interfaces are not affected by the SiO_2 layer thickness, cf. Fig. 2.31.
7. There is a huge amount of charge in the hafnia layer and/or its two interfacial regions: one with the silica layer, and another with the TaN metal electrode; the difference between the flat-band voltage of wafer D-04 (no HfO_2 layer) and the other wafers is an indication of this charge, cf. Fig. 2.31. The total charge at flat-band in the HfO_2 layer and its two interfaces is net negative and its magnitude is roughly around $q \times 1.5 \times 10^{13}$/cm^2.

2.9 Impedance Characteristics of Leaky High-k MOS Structures

There are characteristic differences between the admittance characteristics of MOS devices with ultra-thin (say EOT < 1.5 nm) high-k gate stacks and those of MOS devices with non-leaky thicker (say EOT > 6 nm) gate dielectrics. The admittance characteristics undergo alterations as the gate dielectric thickness is reduced, say, from 6.0 to 1.5 nm. Some changes in the characteristics occur gradually, whereas some changes occur more drastically as a threshold thickness is crossed. One observes several unusual features in the impedance characteristics of ultrathin gate stacks on silicon channels, e.g. large accumulation and strong inversion regimes, flat-band voltage in a much less steep region, and narrow depletion and weak inversion regimes. In the case of ultrathin gate stacks on high mobility channels—such as Ge, GaAs, and InGaAs channels—several anomalous features, such as frequency dispersion of the accumulation and strong inversion capacitance, are observed, in addition to the salient features seen in the case of the silicon channels.

The classical electrical characterization tools were developed, for the MOS devices with the single SiO_2 gate dielectric, more than 3–4 decades ago in the

golden age of the MOS and the MOSFET research [1–4]. These tools are being employed for the leaky ultrathin high-k gate stacks without any significant change or adaptation. Many of these tools, such as the quasi-static C–V technique, simply do not function in the case of the leaky ultrathin gate stacks. Some other tools need to be modified to yield reliable values of the parameters. The electrical characterization tools need to be tuned and matched to the nature of the electrical characteristics. As there are features in the electrical characteristics, which are new, an opportunity and a scope exist for the development of new techniques for measurements and for analysis. Following is an analysis of the new features in the characteristics of the high-k gate stacks.

2.9.1 Si Channels

We will first analyze the admittance characteristics of the high-k gate stacks on the silicon substrate. The MOS characteristics on the high mobility substrates have always been a challenge to interpret and understand; these will be taken up in a later section.

2.9.1.1 Capacitance–Voltage (C–V) Characteristics

Figure 2.32—normalized capacitance, C/C_{di}, versus voltage V—and Fig. 2.15—the corresponding surface potential φ_s versus voltage V—illustrate some of the salient features of the characteristics of the MOS devices with ultra-thin high-k gate dielectrics. Figures 2.32 and 2.15 represent four carefully chosen devices in MOSFET and MOS capacitor configuration with different ultrathin (EOT = 0.46–1.94 nm) high-k materials (HfO_2, $HfAl_2O_5$, La_2O_3, HfSiON) as gate dielectrics, different high-k deposition methods (ALD, e-gun evaporation, oxidation), and gate electrode materials (poly-Si, Ti, Al, TiN), offering a wide variation in band offsets ($\phi_b = 2.00 - 4.19\,eV$), effective tunneling mass ($m^* = 0.22$–0.46 m), and dielectric constant (k = 14–33); the measured C–V data of these devices were taken from the literature [56–59] as indicated in the caption of Fig. 2.15. The C–V characteristic of the MOS capacitor with the SiO_2 gate dielectric has been included for the sake of comparison. Indicated, in Fig. 2.32, are the flat-band voltage, and also the voltage for the onset of strong inversion; these voltages were obtained from the $\varphi_s(V)$ characteristics of Fig. 2.15, which was extracted by the integration of the measured C–V characteristic, in accordance with the Berglund relation [60], cf. Sect. 2.10.2.

One salient feature that one can observe in Fig. 2.32 is that, most of the C–V characteristic is in the accumulation regime and the strong inversion regime. For example, in the case of the MOS transistor with the HfO_2–Al_2O_3 gate dielectric, the accumulation and the strong inversion regime cover a voltage range of nearly 3.62 V, while the depletion and the weak inversion regimes cover only a fraction

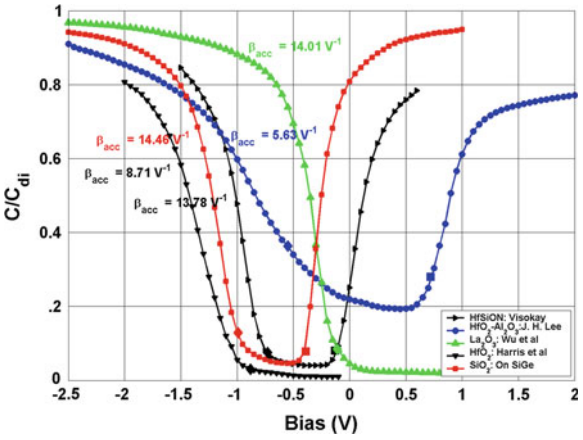


Fig. 2.32 High frequency normalized capacitance (C/C_{di}) versus bias for four MOS capacitors or transistors on p-type silicon with four different high-k gate stacks as indicated in the legend, and for the control sample with the SiO_2 gate dielectric: HfSiON (*right faced triangle*, *black*), HfO_2-Al_2O_3 (*circle*, *blue*), La_2O_3 (*triangle*, *green*), HfO_2 (*inverted triangle*, *black*), and SiO_2 (*square*, *red*) with EOT values of 2.0, 1.7, 1.4, 0.5, and 3.9 nm respectively. It maybe noted that the *diamond marker* indicates the flat-band point, while the *square marker* indicates the onset of strong inversion. The C–V data for four of the characteristics were taken from the literature: HfSiON [56], HfO_2-Al_2O_3 [57], La_2O_3 [58], HfO_2 [59]

of it, namely, about 1.38 V. Secondly, a lower slope in the falling or the rising parts of the C–V curve does not reflect or represent a higher density of interface states in the silicon band-gap, as it used to be the case for the classical MOS devices. It is not clear yet what this slope represents or reflects; a lower slope of C–V characteristic in accumulation or in strong inversion may reflect a higher density of states inside the conduction or the valence band and/or a low direct tunneling current index [22]. Thirdly, certain traditional MOS parameter extraction techniques, if applied, would lead to less reliable results. For example, as the quasi-static C–V technique is not available, often, one finds recourse to the Terman technique [61] in the current literature for extracting the interface state density D_{it}. For the ultra-thin gate dielectrics, it can be shown that the Terman technique is very unreliable, even when D_{it} is high, because of a small potential across the gate stack and a strong and uncertain doping density profile.

In the case of the classical MOS devices, the C–V characteristics were dominated by the depletion and the weak inversion regimes; these regimes yielded almost all the information that could be extracted. The situation is almost reversed in the case of the high-k gate dielectrics of Fig. 2.32, with the accumulation and the strong inversion regimes dominating the C–V characteristics and offering a huge window, which permits a deep look into the accumulation and the strong inversion layers. This opens up the scope of extracting significantly more information (than allowed before) from these two regimes and also for the development of new parameter extraction techniques to make use of this opportunity.

Significance of the parameter—accumulation surface potential index β_{acc}, values of which have been indicated in Fig. 2.32, will be discussed in Sect. 2.10.2. This index was found to vary with the high-k material and its processing and may represent the quality of the high-k gate stack [22].

Noteworthy in Fig. 2.15 are the large surface potential ranges in the accumulation and the strong inversion regimes, particularly when the accumulation surface potential index is small, e.g. see the enormous accumulation surface potential range of 0.62 V for the MOS transistor with the HfO_2 high-k layer. It may be seen that in the depletion and the weak inversion regimes, φ_s is linear with V, with a slope very close to unity, except for the MOSFET with the HfO_2–Al_2O_3 gate dielectric. The parallel displacement of one φ_s(V) from the other reflects the difference in the flat-band voltage. As already discussed in Sect. 2.6.1, the non-saturating surface potential in both strong inversion and accumulation regimes is a strong departure from the classical MOS behavior, cf. Fig. 2.15.

2.9.1.2 Conductance–Voltage (G–V) Characteristics

Figure 2.33 presents the capacitance–voltage (C–V) characteristics of a p-Si/SiO_2/HfO_2/TaN MOS capacitor, measured at 10 and 100 kHz, respectively, and Fig. 2.34 represents the corresponding G–V characteristics. The physical thickness of the HfO_2 layer was 2 nm, and the EOT of the gate stack was 1.9 nm. In Fig. 2.33, the slightly higher plot in the accumulation regime, represents the lower frequency (i.e. 10 kHz), while in Fig. 2.34, the higher plot represents the higher frequency (i.e. 100 kHz). Indicated in both Figs. 2.33 and 2.34 are the flat-band point V_{FB} and the accumulation and the depletion regimes; the surface potential was extracted from the Berglund integral [60]. It may be noted that the conductance observed in Fig. 2.34 in accumulation and also the slight frequency dispersion of the accumulation capacitance observed in Fig. 2.33 are due to the series resistance. The observed right peak in conductance, in the weak inversion regime,

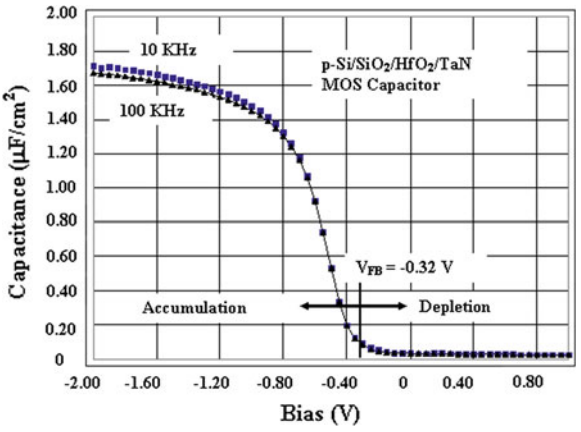


Fig. 2.33 Capacitance–voltage (C–V) characteristics of a p-Si/SiO_2/HfO_2/TaN MOS capacitor, measured at 10 and 100 kHz, respectively [72]

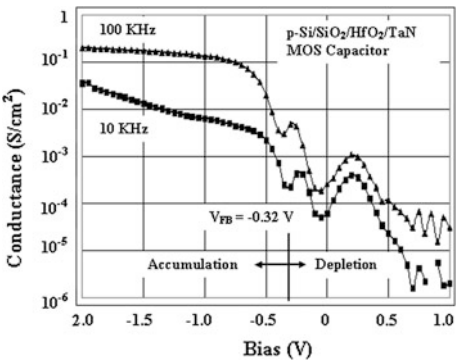


Fig. 2.34 Conductance–voltage (G–V) characteristics of a p-Si/SiO_2/HfO_2/TaN MOS capacitor (the same as in Fig. 2.33), measured at 10 and 100 kHz, respectively [72]

is due to the partial (lossy) response of the inversion layer. In the depletion regime, the conductance peak (the left peak) observed is a manifestation of the loss involved in the capture of holes by the traps at some location x_t in the gate stack. These points will be discussed in more detail in Sect. 2.10. The conductance-voltage (G–V) characteristics of ultrathin leaky gate stacks are complicated by the effect of the series resistance in the accumulation regime and the lossy response of the inversion layer in the weak inversion regime. These features are generally not observed in the case of device grade non-leaky classical gate dielectrics.

2.9.2 Ge and III–V Compound Semiconductor Channels

As indicated already, the admittance-voltage-frequency characteristics of ultrathin high-k gate stacks on high mobility substrates—such as Ge, GaAs, InGaAs—exhibit several unusual features [62, 63]. One of these is the strong frequency dispersion of the accumulation and also the strong inversion capacitance; it may be added that this dispersion has been observed for gate dielectrics—such as 1-octadecene monolayers—on silicon substrates [64]. The capacitance–voltage characteristics of Fig. 2.35 illustrate the huge frequency dispersion of the accumulation as well as the depletion capacitance. The latter is most likely caused by the very high density of the interface traps—exceeding 10^{13} cm^{-2} V^{-1}. Several models [62, 63] have been proposed to explain the frequency dispersion of the accumulation and the strong inversion capacitance; however, this phenomenon is not yet adequately explained. The following is a possible interpretation of this unusual feature.

In the case of the classical gate dielectrics and MOS devices, the accumulation or the strong inversion capacitance approached the capacitance of the gate dielectric, which was assumed to be frequency-independent, as the capacitance of an ideal dielectric should be. The leaky high-k gate stack is a strong departure from an ideal gate dielectric—it has a very high density of traps and as the gate stack is ultrathin and leaky, many of the traps inside the gate stack can exchange

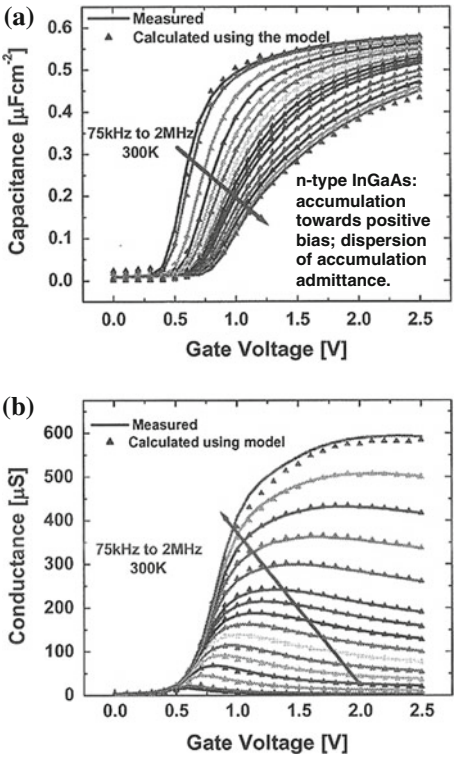


Fig. 2.35 C–V and G–V Characteristics of $In_{0.53}Ga_{0.47}As/LaAlO_3/TaN$ gate stacks measured at different frequencies in the range of 75 kHz and 2 MHz, illustrating frequency dispersion of the accumulation capacitance and conductance. Adapted from [62]

electrons or holes with the semiconductor or the metal surface, thereby contributing and adding trap capacitance to the dielectric capacitance of the gate stack, cf. Sect. 2.8, Figs. 2.28, 2.29 and 2.30. When the ac signal frequency is high enough, the total capacitance of the gate stack reduces to its high frequency value, see Fig. 2.29 and (2.68). As the frequency is reduced, the gate stack capacitance increases because its dielectric capacitance is augmented by the charging capacitance of the traps, see Fig. 2.30 and (2.69). The frequency dispersion of the accumulation conductance, cf. Fig. 2.33, may be caused partly by the charging or discharging loss at the gate stack traps; other sources of this dispersion may be the series resistance and the leakage current.

2.10 Parameter Extraction Techniques

As already outlined, extraction of high-k gate stack parameters is confronted with several difficult roadblocks: (1) the high-k gate stack is a much more complicated system than the SiO_2 gate dielectric; (2) many of the classical MOS parameter extraction techniques are disabled by the nature of the leaky ultrathin high-k gate

stack; (3) no new parameter extraction technique has so far been successfully developed for the high-k gate stack; (4) the frequency range of measurements should extend up to GHz. Ideally, we should have techniques which cover very low to the microwave range: (1) static (10^{-5} to 10^{-2} Hz); (2) quasi-static (10^{-3} to 10^{-1} Hz); (3) LF (Low Frequency −10 Hz to 100 kHz); (4) HF (High Frequency −1 to 50 MHz); (5) Ultra High Frequency (UHF −100 to 900 MHz); (6) Microwave (1 to 20 GHz). We need the static and the quasi-static frequencies to access traps deep inside the high-k gate stack and also to get full response of the inversion layer. We need the HF, UHF, and microwave frequencies to examine the dependence of the dielectric constant on frequency and also to be able to measure the leaky capacitance more reliably. We need the LF, HF, and the UHF frequencies to obtain the full range of the G–V-f characteristics and the full range of the G_p/ω peaks. Because of the gate leakage current, measurements at static, quasi-static, and even most of the LF ranges are not possible. These acute constraints make parameter extraction a difficult exercise in the case of the high-k gate stacks.

2.10.1 Determination of the High-k Gate Stack Capacitance C_{di}

The gate stack capacitance, C_{di}, is a crucial parameter, as it is directly related to the drive current, the trans-conductance, the channel conductance, and the threshold voltage. Furthermore, it reflects somewhat the physical thickness of the gate stack, which is the most important parameter in deciding all phenomena related to quantum-mechanical tunneling (wave-function penetration, carrier confinement, tunneling current, wave-function mixing). Therefore, accurate determination of C_{di} is a crucial exercise. This exercise used to be rather simple in the case of thick gate dielectrics, as the MIS capacitance in very strong accumulation, C_{acc}, saturated nicely to the gate dielectric capacitance C_{di}. In the case of ultrathin, say EOT <3 nm, gate dielectrics, extraction of C_{di} is beset with several problems. The accumulation capacitance does not saturate to C_{di}, because C_{di} is no longer insignificant in comparison to the space charge capacitance in accumulation, and also due to the carrier confinement in the accumulation layer. An ever-increasing gate dielectric leakage current density and the parasitic impedance of the substrate, also make the accumulation capacitance C_{acc} differ from its ideal value. Other factors which may cause C_{acc} to deviate significantly from the gate stack capacitance C_{di} include: (a) capacitance of traps in the gate stack; and (b) low density of states in the conduction band, in the case of compound semiconductors.

Several capacitance techniques [65–69] exist, which were originally developed to extract the capacitance of SiO_2 gate dielectrics; three of these extract the dielectric capacitance directly, while the fourth involves modelling and curve-fitting. These techniques operate under various assumptions, all of which are

invalid in the case of the high-k gate dielectrics. Three of the main differences between SiO_2 and high-k gate dielectrics, which make the above assumptions invalid, relate to the Si/IL interface trap density, D_{it}, which is higher in the latter case, the charge densities in the high-k gate stack, which are orders of magnitude higher in the latter case, and the conductance or the valence band offsets, $\phi_{b,c}$ or $\phi_{b,c}$, which are much lower in the latter case. In these techniques, the interface trap charge, Q_{it}, is neglected in accumulation and around flat-band. Even, in the case of the SiO_2 gate dielectric, D_{it} can be much higher near the band edges; in the case of the high-k gate dielectrics, D_{it} and the corresponding Q_{it} can be too large in accumulation to neglect. The interface trap capacitance, C_{it}, is neglected in the 100 kHz or the 1 MHz (taken as high frequency) accumulation capacitance. It can be easily shown that the interface traps are likely to follow the 1 MHz ac signal in accumulation. The charges in the high-k gate stack are simply too large to ignore. The above-mentioned techniques briefly are:

1. *McNutt and Sah Technique* [65]—One plots $|dC/dV|^{1/2}$ versus C in the accumulation regime. If a linear fit is obtained, then its intercept with the C axis yields C_{di}.
2. *Maserjian Technique* [66, 67]—One plots $|dC^{-2}/dV|^{1/4}$ versus 1/C in the accumulation regime. If a linear fit is obtained, then its intercept with the C^{-1} axis yields $1/C_{di}$.
3. *Ricco Technique* [68]—In this technique, a factor, F_{Ricco} is calculated around the flat-band condition. The flat-band voltage, V_{FB} is determined from the condition that for $V = V_{FB}$, $F_{Ricco} = 0$. Subsequently, the gate dielectric capacitance is calculated from the MIS flat-band capacitance and the flat-band space-charge capacitance, calculated using an approximate relation, obtained by Ricco et al. [68].
4. *Curve-Fitting Technique* [69]—In this technique, a calculated capacitance–voltage, C–V, characteristic is matched to the measured 100 kHz or 1 MHz C–V, adjusting a large number of fitting parameters, to yield the values of the physical (i.e. the fitting) parameters. There can be justifiable concerns regarding all the assumptions of this technique, as well as the technique itself of adjusting many fitting parameters, the number of which will be higher, and their independent measurements more difficult in the case of the high-k gate dielectrics. Particularly, difficult is the determination of the physical parameters (dielectric constants and thicknesses) of gate dielectric stacks, where significant interlayer diffusion and interlayer chemical reactions have taken place, by design or otherwise. Assumption of an infinite potential barrier at the interface is questionable, as the conductance and valence band offsets are in many cases around 2 eV or even less, cf. Appendix V. At a signal frequency of 100 kHz or 1 MHz, there may be significant contribution to the accumulation gate stack capacitance from the traps inside the gate stack, as has been discussed in Sect. 2.8, which the curve-fitting model does not take into account.

2.10.1.1 A New Direct Capacitance Extraction Technique

More recently, a capacitance technique has been proposed to better address the current problems; this technique yields the gate stack capacitance directly, and makes none of the above assumptions. In this technique, two simple options are available for the determination of the capacitance of a leaky ultrathin gate dielectric, using mathematical relations in closed form [21]. These relations [21], presented below, are valid in strong accumulation, if the corresponding parallel capacitance, $C_{p,acc}$, which is the sum of the space charge capacitance, $C_{sc,acc}$, and the interface trap capacitance, $C_{it,acc}$, is an exponential function of the surface potential φ_s, i.e. $C_{p,acc} \propto \exp(\beta_{acc}\varphi_s)$, which was confirmed convincingly by experimental results on a variety of MOS structures with high-k gate stacks, cf. Fig. 2.36.

$$\left|\frac{1}{C}\frac{dC}{dV}\right|^{\frac{1}{2}} = \frac{|\beta_{acc}|^{\frac{1}{2}}}{C_{di}}(C_{di} - C) = \frac{-\sqrt{|\beta_{acc}|}}{C_{di}}C + \sqrt{|\beta_{acc}|}. \tag{2.71}$$

$$\frac{1}{C} = \frac{1}{C_{di}} + \left|\frac{1}{2\beta_{acc}}\frac{d}{dV}\frac{1}{C^2}\right|^{\frac{1}{2}} \Rightarrow \left|\frac{dC^{-2}}{dV}\right|^{\frac{1}{2}} = \frac{\sqrt{2|\beta_{acc}|}}{C} - \frac{\sqrt{2|\beta_{acc}|}}{C_{di}}. \tag{2.72}$$

It can be seen from (2.71) that, a plot of $|C^{-1}dC/dV|^{1/2}$ versus C, in the accumulation regime, should result in a straight line, whose x-intercept would yield the gate dielectric capacitance C_{di}, whose y-intercept would yield $|\beta_{acc}|^{1/2}$, and whose slope (dy/dx) would yield $(-|\beta_{acc}|^{1/2}/C_{di})$. Linear plots (cf. Fig. 2.37) were obtained for all the high-k gate stacks presented in Table 2.1, from which values of C_{di} and β_{acc} were obtained from the x- and y-intercepts, respectively, cf. Table 2.1. Similarly, a plot (cf. Fig. 2.38) of $|dC^{-2}/dV|^{1/2}$ versus C^{-1}, in the accumulation regime, also resulted in a straight line, for the wide variety of gate stacks of Table 2.1. According to (2.72), the x-intercept of this line would yield C_{di}^{-1}, the y-intercept would yield $(-|2\beta_{acc}|^{1/2}/C_{di})$, and whose slope would yield $(2|\beta_{acc}|)^{1/2}$. Values of C_{di} and β_{acc} were obtained from the x-intercept and the

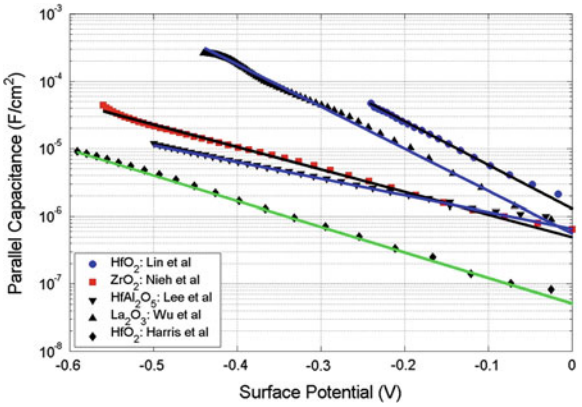


Fig. 2.36 Experimental parallel capacitance, in the strong accumulation regime, $C_{p,acc} = C_{sc,acc} + C_{it,acc}$, versus the surface potential, φ_s, for five MOS devices (on p-type silicon) containing different high-k gate stacks, cf. Table 2.1 [22]

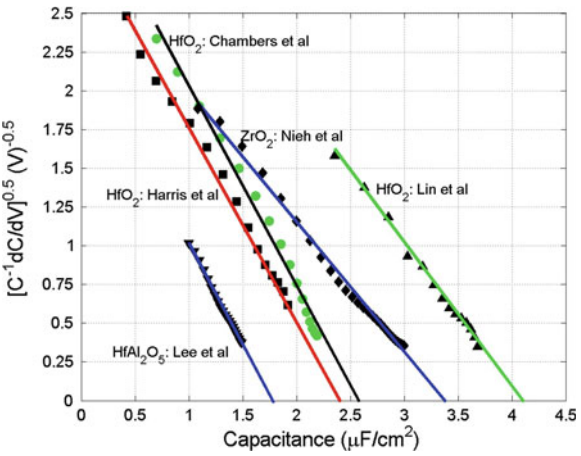


Fig. 2.37 $|C^{-1}dC/dV|^{1/2}$ versus capacitance C, in strong accumulation, for different high-k gate stacks, cf. Table 2.1. The intercept, of the linear fit to the data points, with the x-axis, yielded experimental values of C_{di} and its slope yielded experimental values of β_{acc} [21]

slope, respectively, of the straight lines in Fig. 2.38, cf. Table 2.1. It may be noted that experimental values of C_{di} or β_{acc} may be obtained from the slope of Fig. 2.37 or the y-intercept of Fig. 2.38; however, these may have lower accuracy.

Unsatisfactory results were obtained from the application of three of the existing direct capacitance techniques [65–68] to the high-k gate stacks. The McNutt and the Maserjian plots were very non-linear, thereby suggesting invalidity of the assumptions under which the approximate mathematical relations were derived, in the cases of high-k gate stacks. Different linear fits could be made to different parts of the plots, i.e. in different parts of the accumulation regime, but, this led to ambiguity and improbable and multiple values for the dielectric capacitance. Figure 2.39a and b illustrate some of the deficiencies and present some comparison of the efficacy of the McNutt and the Maserjian Techniques with the technique proposed by Kar [21].

Table 2.1 Reference [21]: experimental values of the gate stack capacitance C_{di} and the exponential index of the surface potential φ_s of MOS structures with a wide variety of high-k gate stacks, with different composition, and/or band offsets, and/or deposition conditions, fabricated by various research groups [57–59, 82–84]

Substrate/High-K/Gate electrode	From Fig. 2.37		From Fig. 2.38		From Fig. 2.36
	C_{di} ($\mu F/cm^2$)	β_{acc} (V^{-1})	C_{di} ($\mu F/cm^2$)	β_{acc} (V^{-1})	β_{acc} (V^{-1})
p-Si/$HfAl_2O_5$/poly-Si [57]	1.76	−5.54	1.77	−5.37	−5.63
p-Si/ZrO_2/TaN [82]	3.43	−7.60	3.41	−7.73	−7.81
p-Si/HfO_2/Al [59]	4.08	−14.47	4.09	−14.47	−14.49
n-Si/HfO_2/TaN [83]	2.91	8.50	2.98	7.75	9.06
p-Si/La_2O_3/Al [58]	7.36	−13.38	7.37	−13.14	−14.01
p-Si/HfO_2/Ti [84]	2.47	−8.76	2.48	−8.66	−8.71

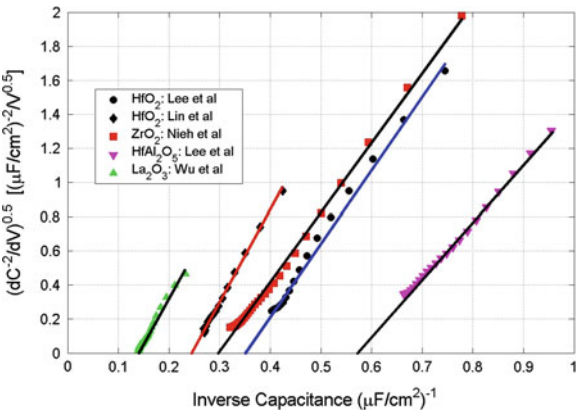


Fig. 2.38 $|dC^{-2}/dV|^{1/2}$ versus inverse capacitance C^{-1}, in strong accumulation, for different high-k gate stacks, cf. Table 2.1. The intercept, of the linear fit to the data points, with the x-axis, yielded experimental values of C_{di}^{-1} and its slope yielded experimental values of β_{acc} [21]

Figure 2.39a illustrates the comparison of the quality of the plots and the attendant results between the Kar technique and the McNutt and Sah technique in the case of a lanthanum oxide gate dielectric with an EOT of 0.48 nm. (p-Si/La_2O_3/Al structure by oxidation of lanthanum on silicon; 3.3 nm La_2O_3; gate area of 1×10^{-4} cm^2). It can be seen in Fig. 2.39a that all the data points from the Kar technique, i.e. almost the entire accumulation regime, fall on a straight line. The McNutt data points lie on a very non-linear curve, which is likely to reflect the neglect of trap charges and states. The value of the dielectric capacitance, obtained from the intercept of the linear fit to the data points of the Kar technique is 722.22 pF. The highest accumulation capacitance measured was 708.61 pF at − 3.50 V. A value of 0.48 nm is obtained for Capacitive Equivalent Thickness (CET), corresponding to a C_{di} of 722.22 pF. The value of β_{acc} obtained from the slope of the linear fit is 15.61 V^{-1}. If one makes a linear fit to the McNutt data points in very strong accumulation, i.e. to only a part of the curve, one obtains a value of 734.21 pF for C_{di}.

Figure 2.39b illustrates the quality of the plots and the attendant results from the Kar technique in comparison to the Maserjian technique for the same sample as in Fig. 2.39a. It can be seen that the data points from the Kar technique fit a straight line even better, over the entire accumulation regime, than in Fig. 2.39a. The value of the dielectric capacitance obtained from the intercept of the linear fit to the data points of the Kar technique is 718.82 pF, while a value of 16.93 V^{-1} is obtained for β_{acc} from its slope. The Maserjian data points lie on a very non-linear curve. Absurd values of C_{di} result from linear fits to parts of the curve in accumulation or moderate accumulation. The best values are obtained from linear fits to the curve in very strong accumulation, which are 945.63 pF from the quantum-mechanical Maserjian technique, and 845.67 pF from the classical Maserjian technique.

The Kar gate stack capacitance extraction technique [21] has since been applied to a very large number of high-k MOS structures [15]; this technique has worked well in all the cases applied, irrespective of the gate stack composition, the

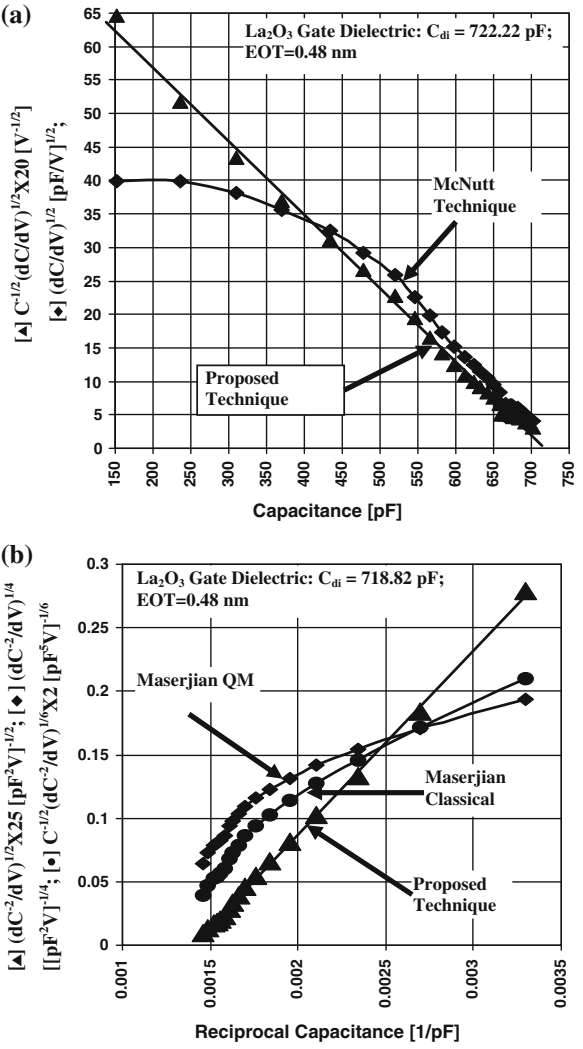


Fig. 2.39 (**a**, *top*) Comparision of the quality of plots obtained from the Kar (proposed) technique versus the McNutt technique for a p-Si/La_2O_3/Al structure (3.3 nm La_2O_3). The intercept from the Kar technique yields a C_{di} of 722.22 pF. The slope yielded a value of 15.61 V^{-1} for the exponential constant β_{acc}. (**b**, *bottom*) Comparision of the quality of plots obtained from the Kar (proposed) technique versus the classical and the quantum-mechanical (QM) Maserjian techniques for the same gate dielectric as in Fig. 2.39a. The intercept from the proposed technique yields a C_{di} of 718.82 pF. The slope yielded a value of 16.93 V^{-1} for the exponential constant β_{acc} [21]

deposition/fabrication conditions, and the value of EOT, consistently producing linear plots over the entire accumulation region, in accordance with (2.71) or (2.72); also, the extracted accumulation capacitance was without any exception observed to be an exponential function of the surface potential.

2.10.1.2 The Curve-Fitting Capacitance Extraction Technique

Perhaps, the most frequently applied and the most popular gate stack capacitance extraction technique, currently, is the curve-fitting technique [40, 69] or some variation of the same. It is not clear what the reason for this popularity and what

the attraction of this technique could be. At the present time, no option exists to ascertain the reliability or the veracity of the values of C_{di} and EOT obtained from the curve-fitting technique, or for that matter, from any other C_{di} extraction technique. Physical thicknesses of the different layers of the gate stack could be obtained from the cross-sectional transmission electron micro-graphs, perhaps, not too inaccurately. Also, a value of the EOT can be calculated from the extracted gate stack capacitance density C_{di}, but the connection of the EOT to the physical thicknesses of the gate stack, cf. (2.73), is not possible, as it is beyond our reach at the moment to determine the values of the dielectric constant of the individual gate stack layers.

$$EOT = \frac{\varepsilon_{SiO_2}}{C_{di}} = t_{di,IL}\frac{k_{SiO_2}}{k_{IL}} + t_{di,HfO_2}\frac{k_{SiO_2}}{k_{HfO_2}} + t_{di,cap}\frac{k_{SiO_2}}{k_{cap}} \tag{2.73}$$

In (2.73), ε_{SiO_2} represents the electrical permittivity of the SiO_2, the respective t's represents the physical thickness, and the respective k's represents the dielectric constant of the different gate stack layers: IL, HfO_2, and the cap layer, if a cap layer is present. Therefore, as there is no independent verification of the accuracy of the generic curve-fitting technique, it could not be that the curve-fitting technique is popular, because it has been proved to be a reliable technique.

One possibility for the popularity of the generic curve-fitting technique could be that it is easy to foresee that the technique could yield a significantly lower value for the EOT than what the C–V characteristic would realistically suggest and what the real value could be. The curve-fitting technique may be overestimating the carrier confinement effects, overestimating the effect of the parasitic resistance, and may be overestimating the effect of the gate leakage current, on the gate stack capacitance. All of these three factors, if present, make the measured accumulation capacitance lower than the gate stack capacitance.

1. As already analyzed, carrier confinement is significantly diluted by a moderate band offset (in place of an infinite barrier, generally assumed in the curve-fitting technique), is significantly diluted by significant wave-function penetration in deep accumulation and deep inversion, is significantly diluted by the tunneling current, and is significantly diluted by the mixing with the metal wave function.
2. A series resistance is a very poor representation of the parasitic impedance of the passive elements of the CMOSFET, which are vast in size and in number. Perhaps, more importantly, it is impossible to measure it. Even in a non-leaky MOS structure (say, with an EOT of 4.0 nm) with a SiO_2 gate dielectric, one does not measure a frequency-independent value of the series resistance in deep accumulation. Resultantly, the modelling of the series resistance effect in the curve-fitting technique is rather baseless.
3. Likewise the carrier confinement effect, both the series resistance and the gate stack leakage current make the measured accumulation capacitance to be lower than C_{di} in the deep accumulation regime. Similar is the effect of inductance at high frequencies (1 MHz). The modelling of the leakage current effect in the curve-fitting technique is too empirical and has no sound basis.

4. There has been no experimental verification of the models of any of the above three effects on the C–V characteristic, as it is not possible to separate these effects from one another or from that of the parasitic inductance. A drooping C–V characteristic in the deep accumulation regime could be an indication of the series resistance, and/or, the leakage current, and/or the parasitic inductance effect.
5. As already discussed, the deep accumulation capacitance may include significant contribution from the charging capacitance due to trapping or de-trapping inside the gate stack. This also will give a higher gate stack dielectric capacitance density and a lower EOT than what the reality is. It is important to note that, at the operating frequency (GHz range), there will be no trapping or de-trapping inside the gate stack; so, what will determine the drain current, transconductance, switching time, etc., will be the dielectric capacitance of the gate stack, and without any contribution from the gate stack trap capacitance.

2.10.2 Extraction of the Surface Potential φ_s

The surface potential (same as the semiconductor band bending or the interface potential) φ_s is one of the most useful parameters in analyzing the MOSFET function.

1. Knowledge of the surface potential enables us to know whether the MOSFET is operating in the accumulation, in the depletion, in the weak inversion, or in the strong inversion regime.
2. Knowledge of the surface potential is necessary for determining the interface trap energy.
3. Knowledge of the surface potential is indispensible for using the MOS conductance technique, and for determining the interface state time constant and the interface state capture cross-section.
4. An accurate value of the surface potential enables us to calculate the free carrier density at the semiconductor surface, and hence the free carrier density at the site of the gate stack trap.
5. Knowledge of the surface potential enables us to know the profile of the carrier confinement potential well.

There is only one way for extracting the surface potential accurately, i.e. by an integration of the equilibrium (or the low frequency) capacitance–voltage (C–V) characteristic. An equilibrium C–V is obtained when both the traps and the minority carriers can follow the applied small signal. As the minority carriers contribute to the capacitance only in strong and weak inversion, the signal needs to be followed by the minority carriers only in strong and weak inversion, but, by the traps for all the bias values. A typical frequency for measuring the equilibrium C–V is of the order of 10^{-3} Hz. Although, sinusoidal frequencies as low as 10^{-5} Hz

are available, small signal C–V measurement is difficult below 100 Hz, because of noise and also because of signal source instability. Therefore the equilibrium C–V is generally obtained from the quasi-static or the static measurement. In the quasi-static measurement, a ramp voltage, V = at, where a is a constant, is applied and the charging current is measured, which yields the quasi-static capacitance:

$$C = \frac{\partial Q}{\partial V} = \frac{\partial Q}{\partial t}\frac{\partial t}{\partial V} = \frac{I_{charging}}{\partial V/\partial t} = \frac{I_{charging}}{a}$$

A typical ramp rate is 1 mV/s. For the measurement samples (i.e. the test structures), the quasi-static charging current is typically in the pA range. The gate stack leakage current, if significant, adds to the charging current; unless the gate leakage current is significantly smaller than a pA, the quasi-static capacitance measurement becomes faulty. In the static C–V measurement, a voltage step (say, 10 mV) is applied; the corresponding incremental charge is measured after a time delay (say, 100 s), thereby yielding the static capacitance. For a reliable quasi-static or a reliable static measurement, it is essential that the gate leakage current is insignificant. This means that both the quasi-static and the static C–V measurements are not possible in the case of the ultrathin, leaky gate stacks. For the leaky gate stacks, the equilibrium C–V can be obtained only for the accumulation regime, for which the only option that is available is to measure the C–V at 1 kHz or so.

As the minority carrier contribution to the accumulation capacitance can be easily ignored, only the response of the majority carriers and the interface traps to the applied small signal, will decide the equilibrium frequency. The majority carrier response time in the silicon substrate is of the order of a ps or less; so, the majority carriers would have absolutely no problem with a 1 kHz or even a 100 kHz signal. The interface trap time-constant for hole capture may be expressed as:

$$\tau_h = \frac{1}{v_h \sigma_h p_s} \tag{2.74}$$

Assuming a hole capture cross-section of 10^{-15} cm^2, and a hole density of 10^{16} cm^{-3} at the silicon surface, the interface trap time-constant estimates to 10 ns. For the interface traps at the Fermi level to follow the signal, the condition ω (=2πf) $\ll (\tau_h)^{-1}$ has to be fulfilled; this is the case for f = 100 kHz. Hence, the 100 kHz or a lower signal frequency can be considered to be an equilibrium frequency in accumulation.

The equilibrium C–V is to be integrated to obtain the surface potential φ_s, according to the relation [60]:

$$\varphi_s - \varphi_{s,0} = \int_0^V \left(1 - \frac{C}{C_{di}}\right) dV \tag{2.75}$$

where $\varphi_{s,0}$ is the surface potential at zero bias. The relation (2.75) can be derived from the circuit representation of Fig. 2.4b and the incremental voltage division defined in Fig. 2.3a:

$$\frac{d\varphi_s}{dV} = \frac{dV - dV_{di}}{dV} = 1 - \frac{dV_{di}}{dV} = 1 - \frac{dQ_M dV_{di}}{dQ_M dV} = 1 - \frac{C}{C_{di}} \tag{2.76}$$

It is apparent from (2.75) and (2.76) that the area above the C–V curve represents the surface potential φ_s while the area below represents the gate stack potential V_{di}.

2.10.2.1 Integration Constant $\varphi_{s,0}$

There are traditional ways of extracting the integration constant in (2.76), namely the zero-bias surface potential $\varphi_{s,0}$ [3]. We outline here two new methodologies for obtaining this integration constant.

New Integration Constant Extraction Method 1.—The 100 kHz capacitance–voltage curve of Fig. 2.33 was integrated, according to (2.75), to obtain the surface potential. First the parallel capacitance, C_p (=C_{sc} + C_{it}), was calculated, using the relation:

$$\frac{1}{C_p} = \frac{1}{C} - \frac{1}{C_{di}} \tag{2.77}$$

Subsequently, C_p is plotted as a function of $\varphi_s - \varphi_{s,0}$, as illustrated by Fig. 2.40. It can be observed in Fig. 2.40, that the parallel capacitance C_p is very much an exponential function of the surface potential in the accumulation regime, and can be expressed as:

$$C_{p,acc} = \alpha_{acc} \exp(\beta_{acc} \varphi_s) \tag{2.78}$$

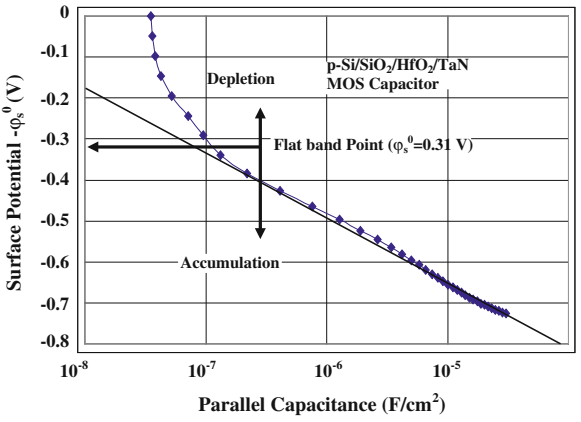


Fig. 2.40 Plot of the experimental parallel capacitance C_p (=C_{sc} + C_{it}) versus the surface potential $\varphi_s - \varphi_{s,0}$ (where $\varphi_{s,0}$ is the surface potential at zero bias). This plot is obtained using the 10 kHz C–V of Fig. 2.33 [72]

In (2.78), the pre-exponential constant α_{acc} is the value of $C_{p,acc}$ for $\varphi_s = 0$, and β_{acc} is the exponential constant of the surface potential φ_s.

In the next step, the flat-band space-charge capacitance, $C_{sc,fb}$, is calculated, using the bulk value of the acceptor density in the following relation, cf. (12):

$$C_{sc,fb} = \sqrt{q\varepsilon_s N_A \beta} \tag{2.79}$$

Under the flat-band condition, $\varphi_s = 0$, the mathematical relation (2.12) simplifies exactly to (2.79). In Fig. 2.40, where $C_p = C_{sc,fb}$, there, the corresponding value of $\varphi_s - \varphi_{s,0}$ is $-\varphi_{s,0}$. As illustrated in Fig. 2.40, the value of the zero-bias surface potential is 0.31 V.

New Integration Constant Extraction Method 2.—The above methodology 1 is susceptible to errors if the surface doping density deviates significantly from the initial substrate value, and/or there is significant contribution from the interface traps to the flat-band parallel capacitance at the measurement frequency. A second option for calculating $\varphi_{s,0}$ would be to make use of the following empirical relation:

$$\frac{C_p}{\alpha_{acc}} = \sqrt{2} \tag{2.80}$$

In Fig. 2.40, where the above empirical relation holds, the corresponding value of $\varphi_s - \varphi s^0$ is $-\varphi_s^0$. The genesis of (2.80) could be outlined in the following manner: If one compares (2.78) to (2.17), then, it would emerge that:

$$\alpha_{acc} = \sqrt{\frac{q\varepsilon_s N_A \beta}{2}} \tag{2.81}$$

Comparison of (2.81) with (2.79) would yield the relation (2.80), if one neglects the contribution of the interface traps to the parallel capacitance density at flat-band:

$$C_{p,fb} = C_{sc,fb} + C_{it,fb} \cong C_{sc,fb}$$

Since the relation (2.80) is a ratio, any variation in the doping density and/or any contribution by the interface traps to the flat-band capacitance would affect both C_p and α_{acc} in the same manner; hence, methodology 2 is more immune to a variation in the value of the doping density in the semiconductor sub-surface or to a contribution to $C_{p,fb}$ from D_{it}.

Figure 2.41 presents the experimental surface potential versus the bias relation. The flat-band voltage, V_{FB}, is the value of the bias, corresponding to $\varphi_s = 0$, which came out to be -0.32 V, cf. Fig. 2.40. In Fig. 2.40, both the options for the extraction of $\varphi_{s,0}$ yielded nearly the same result, perhaps because there was no additional doping of the silicon sub-surface, and the interface trap density near the flat-band point was below 10^{11} cm^{-2} V^{-1}, cf. Sect. 2.10.3.2. The surface potential plot φ_s(V) of Fig. 2.41 can be considered to be reliable and accurate inside the accumulation regime, which covers much of the plot in Fig. 2.33. A small part of

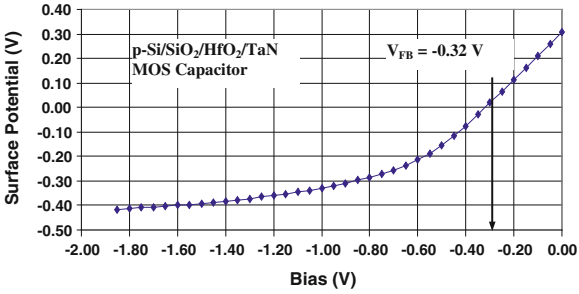


Fig. 2.41 The Experimental surface potential versus the bias plot for the MOS capacitor of Figs. 2.40 and 2.33 [72]

the plot, i.e. in the voltage range of −0.32 to 0 V, in Fig. 2.41 is in the depletion regime. The accuracy of the surface potential in the depletion regime will depend upon the magnitude of the interface trap density in this range; the interface trap density extracted from the conductance technique was observed to be below 10^{11} cm^{-2} V^{-1} in this range, cf. Sect. 2.10.3.2.

The gate stack capacitance C_{di} used to be considered a constant and invariant of the bias and the signal frequency. This assumption is valid if the gate stack is a perfect dielectric and is devoid of any traps and therefore is devoid of any trap capacitance. This was the case for the single SiO_2 gate dielectric. However, as has been pointed out in detail, the high-k gate stack capacitance may contain a significant contribution from the gate stack traps in both deep accumulation and in deep inversion; in that case C_{di} would be a function of the bias V and the frequency f. The relation (2.75) would still be valid, as there is no assumption made in (2.76) that C_{di} is not a function of the bias V. However, there would be serious implementation problems. Since, the gate stack capacitance density C_{di} would keep increasing with the intensity of accumulation because of contribution from additional gate stack traps, it is not clear how C_{di} could be determined. Secondly, it is not clear how calculations would be made using the relation of (2.75) with a gate-voltage variant C_{di}.

2.10.3 Different Techniques for Trap Parameter Extraction

As mentioned in Chap. 1, the classical MOS trap parameter extraction techniques were developed for and applied to (i.e. were suitable for) non-leaky SiO_2 gate dielectrics. The mainstream techniques, responsible for the remarkable success of the SiO_2 gate technology, were the Low–High Frequency Capacitance Technique and the Conductance Technique [3, 4], both of which required the ability to obtain the quasi-static or the static C–V characteristic and reliable measurement of the G–V characteristic over a wide frequency range. Other (significantly less reliable and/or versatile) techniques included the Terman Technique and the Charge Pumping Technique. As mentioned already, leaky high-k gate stacks enormously complicate trap parameter extraction because the quasi-static, leave alone the static, C–V characteristic cannot be obtained under the normal conditions and because the

high-k gate stacks host a very complicated and still-unknown network of traps. As is outlined below, the low–high frequency technique and the conductance technique can still be used in a significantly truncated form and the information these techniques yield is also greatly reduced; moreover, the classical procedure for data analysis has to be modified (as detailed in Sects. 2.10.3.1 and 2.10.3.2) to make both these techniques usable in the case of the high-k gate stacks.

Inversion Capacitance–The main problems with the high gate stack leakage current are two-fold, namely, (a) the gate leakage drains the inversion layer, preventing its formation at the semiconductor surface and the measurement of the inversion capacitance, whereas (b) the high direct conductance submerges the alternating conductance measured at low and even moderate frequencies, thereby significantly restricting the scope of the conductance technique. There exist in principle three options for obtaining the inversion capacitance–voltage characteristic when this cannot be obtained under the normal conditions. Each of these three options enhances the minority carrier generation rate by orders of magnitude over its thermal value in the dark. An inversion layer will form at the semiconductor surface if the minority carrier supply rate exceeds its drainage from the surface by the gate stack leakage [51]. The rate of supply of the minority carriers to the surface will be proportional to its generation rate.

1. Option 1 involves measuring the admittance-voltage-frequency characteristics under illumination; the photo-generation enhances the electron–hole pair generation rate. This option is described in Sect. 2.10.3.3.
2. Option 2 involves the measurement of the admittance-voltage-frequency characteristics at elevated temperatures; this methodology may be referred to as the High Temperature Admittance Technique. The genesis of this option is the enhanced minority carrier generation rate at higher temperatures. The minority carrier generation rate is proportional to the intrinsic carrier density $n_i = (N_c N_v)^{1/2} \exp(-E_G/2kT)$; therefore, higher temperatures translate into higher n_i and therefore higher minority carrier generation rate (N_c, N_v are the effective density of states in the conductance, valence band, respectively). The High Temperature Admittance Technique has been rarely used in practice. Consequently, it has not developed into a mature trap extraction technique. Moreover, it is not clear how much the enhanced minority carrier generation rate will mitigate the inversion layer drainage, as higher temperature can translate into a higher gate leakage rate.
3. Option 3 involves measurement of the MOS admittance-voltage-frequency characteristics using the MOSFET configuration and formation of the inversion layer by minority carrier injection from the source, to compensate for the gate stack leakage. Much like the option 2, this technique remains to be developed into a mature one with all the issues carefully analyzed and a comprehensive methodology is available for data analysis.

The problem of the high direct conductance submerging the alternating conductance can perhaps be solved if the small signal conductance could be measured

at frequencies higher than what has been used so far, namely frequencies higher than 1 MHz—say in the range of 1 MHz–10 GHz—to compensate for the loss of the range of, say, 1 Hz–10 kHz (The small signal conductance increases with the signal frequency). Attempts have been made to measure the conductance in this range (HF, UHF, Microwave), but serious measurement problems including those of the appropriate sample configuration, lead impedance, etc. are still to be overcome.

Terman Technique—Once the Low–High Frequency technique matured, the Terman Technique [61] (also referred to as the High Frequency Capacitance Technique) was seldom used in the case of the SiO_2 gate dielectrics on account of its unreliability and inaccuracy. In brief, it consists of the following steps: (1) The high frequency C–V characteristic is measured. (2) The ideal high frequency C–V characteristic is calculated. (3) The voltage shift between the measured and the calculated C–V curves, δV, is calculated as a function of the gate bias V_G and subsequently transformed into a function of the surface potential φ_s, using the calculated $\varphi_s(V)$ relation. (4) The interface trap density is calculated from the differential $d(\delta V)/d\varphi_s$, using the relation $D_{it} = (C_{di}/q)\ d(\delta V)/d\varphi_s$. (5) The trap energy is calculated from the calculated value of φ_s. The Terman Technique has been applied to high-k gate stacks in some investigations, but it is highly unreliable because of the following reasons: (a) The potential δV, which is a part of the potential across the high-k gate stack, is small, when EOT is small, and is therefore vulnerable to error. (b) A significant inaccuracy is possible in the calculated high frequency C–V characteristic because of uncertainty in the value of the doping density. (c) As explained in Sect. 2.9.1.1, the depletion regime where the Terman Technique applies is tiny, in the case of the ultrathin gate stacks, with a flat profile which will promote errors in the value of δV.

Charge Pumping Technique—All the techniques discussed above are steady state small signal techniques. In contrast, the Charge Pumping Technique [70, 71] is a non-steady-state technique, in which while applying a reverse bias to the source/substrate and drain/substrate diodes, the gate is periodically pulsed between accumulation and strong inversion. The substrate direct current, which is called the charge pumping current, and is monitored over the duration of the pulse, originates from the exchange (emission and capture) of electrons and holes between the traps and the conduction band and the valence band of the semiconductor. All the non-steady-state and the transient techniques involve measurements during the period of trap relaxation in response to a change in the trap occupancy caused by a scanning of the imref through the trap levels; a transient capacitance technique measures the transient capacitance, whereas the Charge Pumping Technique measures the transient charging current flowing to adjust the trap charge due to a change in its occupancy.

The charge pumping model [71] indicates the current to depend upon a host of parameters including the trap density and the trap capture/emission cross-section. The charge pumping current I_{CP} has been observed to be proportional to the pulse frequency and the gate area, but its mathematical relation to many other parameters of the model has not been established. The usual variable measurement

parameters are the pulse frequency, the pulse width, and the pulse rise/fall times, the most potent among which is the pulse frequency. As the trap response time has to match the pulse frequency, in principle, it may be possible to access traps at different locations inside the gate stack and traps with widely varying capture/emission cross sections. However, in practice, there are several issues connected to the application of this technique to the high-k gate stacks:

1. In reality, the Charge Pumping Technique involves a complicated process, particularly in the case of the high-k gate stacks. The connection of the charge pumping current to the trap density (D_{it}), the trap capture cross-section (σ_t), and the trap location (x_t) in the gate stack is not clearly established yet; there are several ambiguities in the charge pumping models developed so far.
2. Any of the three trap parameters—D_{it}, σ_t, or x_t—cannot be determined reliably, unless the other two parameters are known from independent measurements. Since the latter is seldom the case, the values of the other two trap parameters are assumed. For example, often, the trap capture/emission cross-section (σ_t) is assumed to be of the order of 10^{-15} cm^{-2}; but, experiments indicate that the capture cross-section of traps in the high-k gate stacks can vary by orders of magnitude [17], also see Chap. 8.
3. It is not clear what exactly the parameter D_{it} in the model for the charge pumping current I_{CP} represents.
4. The trap energy E_t cannot be extracted from this technique.
5. It is unclear what the effect is of the large gate leakage current on the reliability of this technique.
6. The charge pumping current has a clearer interpretation when the trap density is not a function of energy, the trap capture/emission cross-sections for electrons/holes are same for all the traps which respond to the pulse frequency, and all the traps which emit or capture electrons/holes are located on the same plane in the gate stack. In a high-k gate stack, the traps are located at various planes throughout the gate stack with capture/emission cross-sections varying over several orders of magnitude. So, it is possible that traps at different locations, with varying cross-sections and varying densities respond to a particular pulse frequency. What is the expression for the charge pumping current in such a case; can any trap parameter be extracted from this current?

2.10.3.1 Low-High Frequency Capacitance Technique

The high frequency C–V and the low frequency C–V are used in this technique. The low–high frequency capacitance technique consists of the following steps [3].

1. For any value of the bias in the depletion regime, the space charge capacitance density C_{sc} is calculated from the high frequency capacitance density C_{hf}, cf. (2.7) and Fig. 2.4a, whereas the parallel capacitance density $C_p = C_{sc} + C_{it}$ is

calculated from the low frequency capacitance density C_{lf}, cf. (2.6) and Fig. 2.4b, using the extracted value of the gate stack capacitance density C_{di}, cf. Sect. 2.10.1.1.
2. The experimental C_{sc} is subtracted from the experimental C_p to obtain the interface trap capacitance density C_{it}, from which the interface trap density D_{it} is extracted according to (2.6).
3. The surface potential is obtained from the bias value as outlined in Sect. 2.10.2, from which the trap energy E_{it} is extracted according to the following relation, for p-type semiconductor:

$$E_{it} = E_v + q\left(\phi_p + \varphi_s\right) \tag{2.82}$$

4. The low–high frequency capacitance technique yields only the interface trap density distribution $D_{it}(E_{it})$; no trap time constant or capture cross-section data are obtained.
5. To obtain the $D_{it}(E_{it})$ distribution in the weak inversion regime, one has to use a space charge capacitance density calculated using a value for the doping density; this procedure is vulnerable to the uncertainty in the value of the doping density.

The low–high frequency capacitance technique is difficult to implement in the case of the leaky ultrathin gate stacks because of the following reasons. Generally the lowest frequency at which the C–V can be reliably measured is 1 kHz or so, which is far above the equilibrium frequency, and the highest frequency at which the C–V can be reliably measured is 100 kHz or so, which is below what can qualify as the high frequency. Consequently, even in the depletion regime, the low–high frequency capacitance technique can yield only a fraction of the total trap density at and around the Si/SiO_2 interface. However, it may be still be used only for an indicative comparison between samples.

2.10.3.2 Conductance Technique

As in the case of the leaky ultrathin gate stacks, the quasi-static C–V cannot be measured, the only reliable techniques, that are available for obtaining the trap parameters (trap density D_t, trap energy E_t, trap capture cross-section σ_t, and trap location x_t inside the gate stack), are the conductance technique [3, 4] and the charge-pumping technique [70, 71]. The conductance technique remains the most reliable technique for yielding D_t, E_t, and σ_t of traps in the majority carrier band-gap half, i.e. under the depletion condition [3]. To use the conductance technique, one needs to obtain the surface potential φ_s fairly accurately and the parallel conductance G_p, see Figs. 2.29 and 1.1. In principle, the measured total conductance G_m needs to be corrected for the series resistance R_s and the direct conductance G_{dc}. The direct conductance G_{dc} can be obtained by differentiating the measured I–V (direct current–voltage) characteristic [51]. However, no

satisfactory technique exists for obtaining the series resistance R_s in the case of a leaky high-k gate stack. The main reason is that the passive regions of the device cannot be represented by a lumped resistance, but, needs to be represented by impedances at several locations. In the case of non-leaky MOS devices with a single SiO_2 gate dielectric, the series resistance could be obtained from the real part of the impedance measured around 1 MHz in strong accumulation [3, 51].

After the measured total parallel conductance G_m has been corrected for the series resistance R_s and the direct conductance G_{dc}, see Fig. 2.29, to obtain the total parallel conductance G_{ac}, the parallel conductance G_p is obtained using the following relation:

$$\frac{G_p}{\omega} = \frac{\omega(C_{di})^2 G_{ac}}{\left[(G_{ac})^2 + \omega^2(C_{di} - C)^2\right]} \tag{2.83}$$

Experimental G_p/ω is then plotted as a function of the surface potential φ_s, as depicted in Fig. 2.42, for 10 and 100 kHz, respectively, for the sample of Figs. 2.41 and 2.34. It may be noted that the curve, closer to the accumulation regime, represents the higher frequency (i.e. 100 kHz). As illustrated in Fig. 2.42, G_p/ω undergoes a peak, as φ_s is changed; the peak value of G_p/ω, $(G_p/\omega)_p$, is a measure of the trap density; while the corresponding value of φ_s, $(\varphi_s)_p$, represents the trap's capture probability [3].

As already mentioned, when SiO_2 (grown by dry thermal oxidation) is the gate dielectric, it is safe to assume that traps exist mainly at or in the vicinity of the Si-SiO_2 interface. But, in the case of the high-k gate stacks, traps are likely to exist throughout the gate stack. Moreover, the trap density is perhaps lowest at the Si-SiO_2 (or Si-SiON) interface, and is higher or much higher elsewhere in the high-k gate stack [16]. So it is not necessary at all that only traps at the Si-SiO_2 interface will contribute to the observed conductance peaks in Fig. 2.42; the G_p/ω peaks in Fig. 2.42 will in general represent traps at some location x_t in the high-k gate stack. The trap density D_{it} can be calculated using the relation [3]:

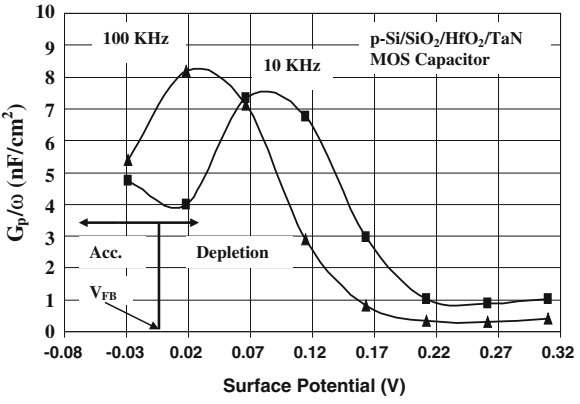


Fig. 2.42 The experimental G_p/ω versus the surface potential under the depletion condition for the MOS capacitor of Figs. 2.31 and 2.32, for 10 and 100 kHz, respectively [72]

$$D_{it} = \frac{1}{f_D q}\left(\frac{G_p}{\omega}\right)_p \tag{2.84}$$

The parameter f_D represents the effect of the statistical fluctuation of the surface potential on G_p/ω [3]. The trap energy E_t can be calculated (for p-type silicon) according to (2.82). The trap's hole-capture cross-section can be extracted using the relation [3], cf. (2.62):

$$\sigma_t^h(x_t)p(x_t) = \frac{\omega}{f_\sigma v} \tag{2.85}$$

$\sigma_t^h(x_t)$ is the hole-capture cross-section of the trap at location x_t from the silicon surface, $p(x_t)$ is the hole density at x_t, v is the average thermal velocity of holes, and f_σ represents the effect of the statistical fluctuation of the surface potential on the trap time constant. In the case of single level interface traps at the Si/SiO_2 interface, $f_D = 0.5$ and f_σ is 1.0; otherwise, for a trap eigenenergy continuum at the Si/SiO_2 interface and statistical fluctuation of φ_s, the value of f_D is <0.5, and can be as low as 0.15, while the value of f_σ is >1.0 and can be as high as 2.6 [3]. There is no analysis of how, in the case of high-k gate stacks, the statistical fluctuation of traps throughout the gate stack will affect the parameters f_D and f_σ. These two parameters are complex functions of the standard deviation, σ_s, of the surface potential [3].

A practical approach is outlined in the following for calculating the statistical parameters f_D and f_σ, which differs from that outlined by Nicollian and Brews [3]. Normally, the ratio $(G_p/\omega)/(G_p/\omega)_p$ is calculated from the plots of G_p/ω versus the small-signal frequency f, at 5 or 0.2 times the peak frequency f_p, at which the peak of G_p/ω occurs. Since in the case of the ultrathin high-k gate dielectric stacks, the frequency range for the conductance measurements is severely limited by the high gate dielectric leakage, one does not have enough frequency variation to obtain the peak and/or enough of the profile in the G_p/ω versus f plot. A solution for this problem is to use the value of G_p/ω in Fig. 2.42 at a surface potential, either $\Delta\varphi_s$ higher or lower than $(\varphi_s)_p$; $\Delta\varphi_s$ can be calculated using the relation:

$$\Delta\varphi_s = \frac{\ln 5}{\beta} \tag{2.86}$$

Once the ratio of $(G_p/\omega)/(G_p/\omega)_p$, is obtained at $(\varphi_s)_p \pm \Delta\varphi_s$, the standard deviation σ_s can be obtained from a numerical plot of $(G_p/\omega)/(G_p/\omega)_p$ versus σ_s, contained in [3]. Subsequently, f_D and f_σ can be obtained from numerical plots of f_D versus σ_s and f_σ versus σ_s, respectively, also provided in [3]. Tables 2.2 and 2.3 present the experimental values of σ_s, f_D, f_σ, D_t, E_t, and the trap's hole-capture probability, $\sigma_t p(x_t)$.

As (2.85) suggests, it is not possible to extract either the trap's hole-capture cross-section or the trap's location x_t in the gate stack, unless the other parameter is determined independently, e.g. from the charge-pumping technique. If the trap is located at the silicon surface ($x_t = 0$), i.e. it is an Si/SiO_2 interface trap, then the

Table 2.2 Experimental trap parameters

f (kHz)	$(G_p/\omega)/(G_p/\omega)_p$	σ_s $(1/\beta)$	f_D	f_σ	D_t (cm^{-2} V^{-1})
10	0.803	1.86	0.250	2.400	1.75×10^{11}
100	0.829	2.05	0.238	2.425	2.00×10^{11}

Table 2.3 Experimental trap parameters

f (kHz)	$E_t - E_v$ (eV)	$\sigma_t^h p(x_t)$ (cm^{-1})	$\sigma_{it}^h(x_t = 0)$ (cm^2)	$\sigma_{it}^h(x_t = 0.4$ nm) (cm^2)
10	0.31	2.62×10^{-3}	4.21×10^{-17}	1.05×10^{-15}
100	0.26	2.59×10^{-2}	4.95×10^{-17}	1.24×10^{-15}

values of the trap's hole-capture cross-section come out to be very small (of the order of 10^{-17} cm^2), as indicated in Table 2.2. In principle, a range as large as 10^{-12}–10^{-18} cm^2 is possible for the trap capture cross-section [44], depending upon whether the trap is charge-wise neutral, or Coulomb-attractive (very large σ_t) or Coulomb-repulsive (very small σ_t). If on the other hand, the trap is located inside the gate stack, its hole-capture cross-section would be larger. For example, if we assume an x_t of 0.4 nm, and $(2\kappa_h)^{-1} = 0.16$ nm, then the hole capture cross-sections are of the order of 10^{-15} cm^2 (typical values), as Table 2.3 indicates.

As Table 2.2 amply illustrates, if the effects of the statistical fluctuation of the surface potential are ignored, large errors will result in the values of the trap density and the trap's capture cross-section. In the literature, almost always, not only has this phenomenon been ignored in extracting the trap parameters from the conductance data, but single level traps have been assumed without any validation. Also, in most cases, only the trap density has been estimated from the conductance data, but not the trap energy or the trap capture cross-section. The main reason behind this deficiency has perhaps been the inability to extract the surface potential, in the absence of the quasi-static C–V data.

This brings us to the issue of the error in the surface potential data of Fig. 2.41 and the methodology outlined in Sect. 2.10.2 for extracting the surface potential in the absence of the quasi-static C–V characteristic. The extracted values of the surface potential, as explained earlier, are accurate in the accumulation regime. The magnitude of error in the surface potential in the depletion regime would increase as the surface potential moves away from the flat-band point and towards the weak inversion condition; the error would also depend on the magnitude of the trap density. The values of the trap density, as reflected in Table 2.2, are relatively low for a high-k gate stack, particularly, when one considers that these traps are close to the valence band edge, an energy range, where the trap density is likely to be significantly higher than the mid-gap trap density. One may bear in mind that, these values, of the trap density in Table 2.2, may not represent the total trap density, which decides the magnitude of the total potential across the gate stack. It is useful to note that, in a situation, where traps are present throughout the gate stack, the trap capacitance at the equilibrium frequency will reflect the total trap density, but the conductance at any frequency will reflect the trap density only at a certain x_t [72].

The experimental data of Figs. 2.33, 2.34, 2.40, 2.41 and 2.42 belong to the same sample; this sample belonged to a group of samples on wafer D-06 which had a graded SiO_2 layer (1–6 nm) grown in dry O_2 at 900 °C. Figures 2.43 and 2.44 illustrate the variation in the experimental interface trap density and the hole capture cross-section, respectively, with EOT, for different sets of such wafers (All these wafers had graded SiO_2 layer). Wafer D-04 had no HfO_2 layer; wafer D-06 and D-12 had 2 nm and wafer D-10 and D-14 had 3 nm thick HfO_2 layer. Wafers D-06 and D-10 had a post-deposition annealing (PDA) in O_2 at 500 °C for 1 min. The wafers of Fig. 2.31 are the same as those of Figs. 2.43 and 2.44. The 100 kHz conductance yielded the parameters of traps located in the range of 0.24–0.27 eV above the valence band edge E_v, while the 10 kHz conductance yielded the parameters of traps located in the range of 0.30–0.32 eV above the valence band edge E_v (see Table 2.3). The trap energy E_t and the hole capture cross-section σ_h did not vary significantly with the SiO_2 layer thickness or the HfO_2 layer thickness, suggesting that in the case of all the MOS capacitors (high-k gate stacks), we are

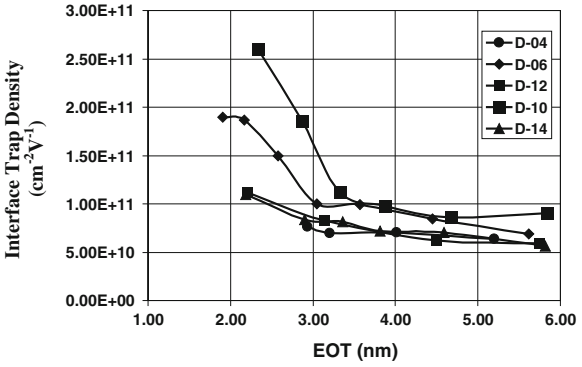


Fig. 2.43 Interface trap density versus EOT for p-Si/SiO_2/HfO_2/TaN MOS capacitors on different graded-SiO_2 wafers with different sets of PDA and HfO_2 thickness [17]

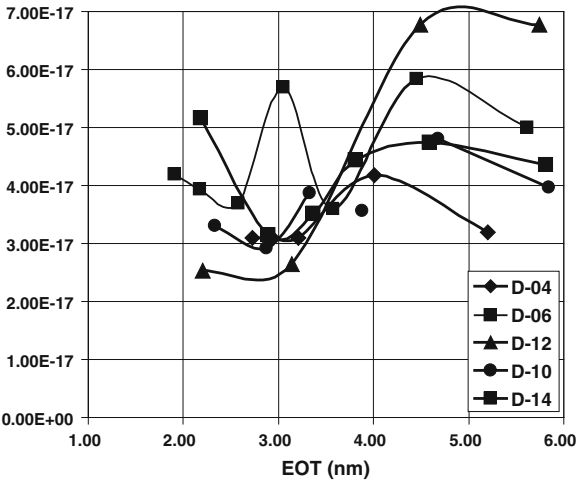


Fig. 2.44 Hole capture cross-section versus EOT for p-Si/SiO_2/HfO_2/TaN MOS capacitors on different graded-SiO_2 wafers with different sets of PDA and HfO_2 thickness [17]

dealing with the same nature of traps. Figures 2.43 and 2.44 represent the interface state density at $E_v + 0.26 - 0.28$ eV.

The experimental data on the trap density, as illustrated by Fig. 2.43, may be summarized in the following:

1. The trap density increases for all wafers with decreasing SiO_2 thickness, except for wafer D-04 (SiO_2 gate dielectric).
2. The increase in the trap density for the thinner SiO_2 layer is significantly higher for wafers subjected to Post Deposition Annealing (PDA in oxygen at 500 °C for 1 min.).
3. The increase in the trap density is higher for the thicker HfO_2 layer, in the case of wafers undergoing PDA.

Following interpretations are possible from the results on the trap parameters in Figs. 2.43 and 2.44 and on the flat-band voltage profile as a function of EOT in Fig. 2.31:

1. That the trap density does not change significantly with decreasing SiO_2 thickness in the case of wafer D-04 (no HfO_2 layer) suggests that the increase in D_{it} in the case of wafers with the HfO_2 layer has something to do with the HfO_2 layer itself; also it has nothing to do with the TaN gate electrode.
2. This conclusion is reinforced by the fact that the flat-band interface charge density changes from a net negative value for the wafer D-04 (no HfO_2 layer) to a net positive value for the other wafers all of which have HfO_2 layer, cf. Fig. 2.31 and Sect. 2.8.1. The change in the sign of the flat-band interface charge with the induction of the HfO_2 layer suggests introduction of new defects into the intermediate SiO_2 layer and the Si-SiO_2 interface by the HfO_2 layer.
3. That the trap density in Fig. 2.43 and its increase for the thinner SiO_2 are higher for the thicker HfO_2 layer is another fact indicating the crucial role of the hafnia layer. Chapter 7 presents experimental data showing higher trap density in the intermediate SiO_2 layer as the HfO_2 layer thickness is increased.
4. The HfO_2 layer and its two interfaces with a huge flat-band charge density of $q \times 1.5 \times 10^{13}$ cm^{-2} (see Fig. 2.31 and Sect. 2.8.1) represent an inexhaustible store of defects and source of contamination to diffuse into the neighbor SiO_2 layer and the Si-SiO_2 interface with a two orders of magnitude lower trap density. The defect concentration gradient across the HfO_2-SiO_2 layers is very steep; what exactly diffuses is not clear.
5. The increase in D_{it} for thinner SiO_2 can be explained by enhanced diffusion of impurity ions or atoms from the HfO_2 layer through the thinner SiO_2. It can also be explained by the influence of the SiO_2/HfO_2 interface on the generation of traps in the SiO_2 layer—due to mismatch in chemical bonding, in coordination, and in impurity or defect solid solubility; such a trap is likely to have a profile with a peak near the SiO_2/HfO_2 interface; so, D_{it} at the Si/SiO_2 interface will be higher for a thinner SiO_2 layer.

6. The formation of an interface dipole at the SiO_2/HfO_2 interface has been analyzed in depth in Chap. 6 and an interface dipole model has been presented based upon the difference in the areal density of oxygen between SiO_2 and HfO_2, which results in the exchange of oxygen across the SiO_2/HfO_2 interface, culminating in an interface dipole. The traps of Fig. 2.43 may be the tail of this interface dipole.
7. Oxygen vacancies are generated in the hafnia layer; the released oxygen atoms may diffuse through the thinner silica layer into the silicon/silica interface [73] and induce new traps.
8. As oxygen diffuses through a thinner silica layer to the Si-SiO_2 interface; a new oxide layer grows at lower temperature (500–600 °C during HfO_2 deposition and/or PDA), resulting in a higher trap density. The source of oxygen could be the HfO_2 layer (oxygen released by generation of oxygen vacancies), the gas phase during the HfO_2 deposition, and/or the gas phase during the PDA.
9. The degradation of the HfO_2 gate stack has been analyzed in detail in Chap. 8 with the conclusion that the main seat of the stress-induced traps, which are the main agents of degradation, is inside the intermediate SiO_2 layer. The experimental data presented in Chap. 8 suggest that these traps/defects are positively-charged oxygen vacancies induced in the SiO_2 layer by the hafnia layer.

2.10.3.3 Photo-Admittance Technique

The photo-admittance technique was proposed by Poon and Card [74]; subsequently it was developed by Kar and Varma [75] with a complete methodology for parameter extraction. This technique involves measuring the MOS admittance (capacitance and conductance) as a function of bias at different frequencies and at different illumination levels, see the C–V characteristics in Fig. 2.45. In principle, this technique enables measurement of the equilibrium C–V characteristic including the inversion capacitance in the case of the leaky gate stacks and/or channel (semiconductor substrate) materials with a low minority carrier generation rate. Under illumination, the minority carrier generation rate is enhanced by many orders of magnitude over its thermal generation rate in the dark; this increased minority carrier generation rate enables build-up of the inversion layer in leaky gate stacks and/or in the case of channels with a low thermal generation rate (as in GaAs). This technique was applied by Kar and Varma [75] to non-leaky SiO_2 gate dielectrics on Si to extract the trap density and the capture cross-section in both halves of the band-gap, i.e. under both the depletion and the weak inversion conditions. An important point to note is that this is the only technique which allows determination of both the electron and the hole capture cross-sections and the trap density in both halves of the band-gap from the measured MOS conductance irrespective of whether the gate stack is leaky or not.

As has been discussed in Chap. 12 and also in the other chapters, high mobility channels have to be employed to enhance the drain current. Many of these high

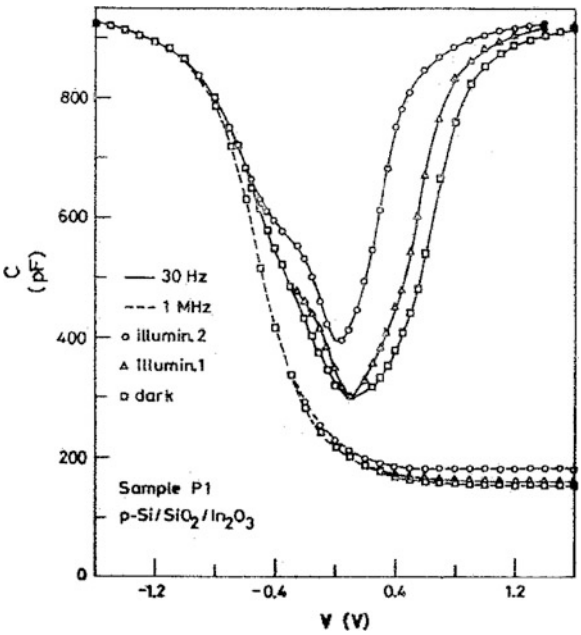


Fig. 2.45 Capacitance–voltage characteristics of a p-$Si/SiO_2/In_2O_3$ MOS structure (sample P1) measured at 30 Hz and 1 MHz in the dark and under two different levels of illumination [81]

mobility channels have low minority carrier generation rate in the dark and consequently suffer from the inability to measure the inversion capacitance and the C–V characteristic in the weak and the strong inversion regimes. In such situations, the photo-admittance technique may be an effective aid. As will be explained later, the photo-capacitance or the photo-conductance can alone yield the trap density in both halves of the band-gap, but to extract the trap energy reliably, both the photo-capacitance and the photo-conductance have to be measured and analyzed [75].

Under illumination, the law of mass action will no longer hold and the magnitude of the imref (i.e. the quasi-Fermi level) separation will increase with the illumination level. Under the low-level injection, the majority carrier imref will remain the same as in the dark, while the minority carrier imref will move towards the minority carrier band edge, see Fig. 2.46. Consequently, under low-level injection, while the majority carrier density will remain the same as in the dark, the minority carrier density will increase by many orders of magnitude. Hence, the space charge capacitance under illumination $C_{sc,photo}$ will change from its value in the dark C_{sc}, and the mathematical relation of (2.12) has to be modified as indicated below, in which δE_{FS} is the imref separation [75]:

$$C_{sc,photo} = \left(\frac{q\varepsilon_s N_A \beta}{2}\right)^{1/2} \frac{\left|1 - e^{-\beta\varphi_s} + {}^{n_0}/_{p_0}\left(e^{\beta\varphi_s} - 1\right)\exp\left(\frac{\delta E_{FS}}{kT}\right)\right|}{\left[\left(e^{-\beta\varphi_s} + \beta\varphi_s - 1\right) + {}^{n_0}/_{p_0}\left(e^{\beta\varphi_s} - \beta\varphi_s - 1\right)\exp\left(\frac{\delta E_{FS}}{kT}\right)\right]^{1/2}} \tag{2.87}$$

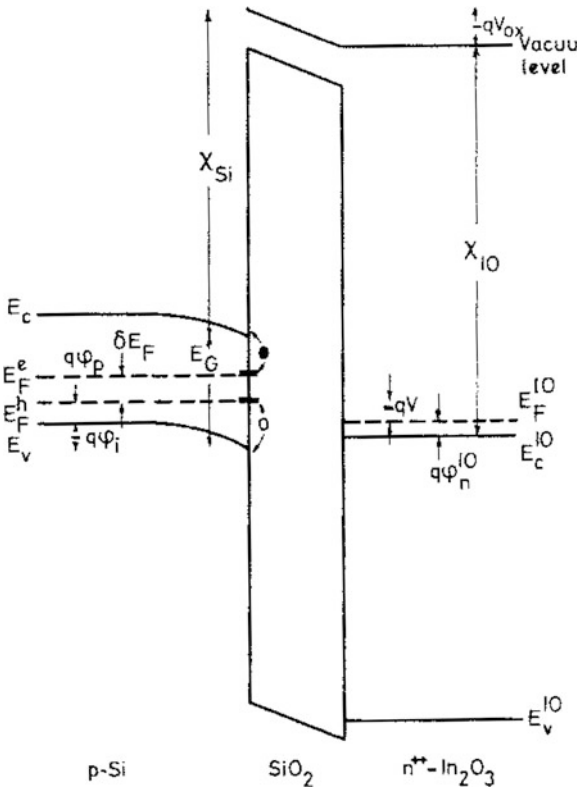


Fig. 2.46 Energy band diagram of a p-Si/SiO_2/In_2O_3 MOS structure at a certain applied bias V. E_F^h and E_F^e are hole and electron imref; φ_i is the surface (interface) potential and φ_p is the Fermi potential; E_F^{IO} is the indium oxide Fermi level and E_v^{IO} and E_c^{IO} are the indium oxide valence and conduction band edge; φ_n^{IO} is the Fermi potential in the degenerate indium oxide [75]

Extraction of the imref separation δE_{FS} is an important exercise in the photo-admittance technique. There are three options for determining the imref separation:

1. From the measured high frequency photo-capacitance in strong inversion. Under illumination, strong inversion will set in at a surface potential $\delta E_{FS}/q$ less than its value in the dark: $\varphi_{s,inv,th,photo} = (\varphi_{s,inv,th} - \delta E_{FS}/q)$. The surface potential at the onset of strong inversion under illumination $\varphi_{s,inv,th,photo}$ can be determined from the high frequency photo-capacitance minimum in strong inversion; its deviation from its dark counterpart will yield the imref separation [75].
2. From the Berglund integral of the equilibrium photo-capacitance–voltage characteristic over strong accumulation to strong inversion, see (2.75), which will yield a surface potential $\Delta\varphi_s$, the deviation of which from the band-gap equivalent voltage will yield the imref separation: $E_G - q\Delta\varphi_s, = \delta E_{FS}$ [75].
3. From the parallel shift of the parallel capacitance C_p (or the space charge capacitance C_{sc}) in strong inversion between its dark value $C_{p,inv}$ and that under illumination $C_{p,inv,photo}$. The experimental parallel capacitance C_p is extracted using (2.77) and the experimental surface potential φ_s is determined as outlined

in Sect. 2.10.2. The parallel capacitance C_p in dark as well as under illumination is plotted as a function of the surface potential φ_s, as is illustrated in Fig. 2.47. In strong inversion, the parallel capacitance, $C_p = C_{sc} + C_{it}$, generally reduces to the space charge capacitance C_{sc}. Figure 2.47 demonstrates the experimental C_p to be an exponential function of φ_s, in both dark and under illumination, as could be expected from (2.12), (2.17), and (2.87). As is indicated in Fig. 2.47, the parallel shift along the φ_s axis between the illuminated and the dark characteristic yields the imref separation δE_{FS}.

The trap parameters—the trap density D_{it}, trap energy E_t, and the capture cross-section of the trap σ_h/σ_e—are determined using the procedure outlined in Sects. 2.10.3.1 (Low–High Capacitance Technique) and 2.10.3.2 (Conductance Technique), with the following modifications. In the dark, interface trap recombination is dominated by exchange of carriers by traps at the Fermi level with the majority carrier band, even in the weak inversion regime [3]. However, under illumination as illustrated in Fig. 2.46, carrier exchange between interface traps and both the conduction and the valence bands are possible. Before the trap energy or the trap cross-section can be determined, one needs to establish whether the exchange of holes between the traps at the hole imref $E_{FS,h}$ and the valence band or the exchange of electrons between the traps at the electron imref $E_{FS,e}$ and the conduction band is dominating the trap recombination process, cf. Fig. 2.46. If $p_s\sigma_h$ is $\gg n_s\sigma_e$, then the interface trap recombination process is dominated by the hole capture by traps at $E_{FS,h}$; in this case, the trap energy is given by (2.82). On the other hand, if $p_s\sigma_h$ is $\ll n_s\sigma_e$, then the interface trap recombination process is

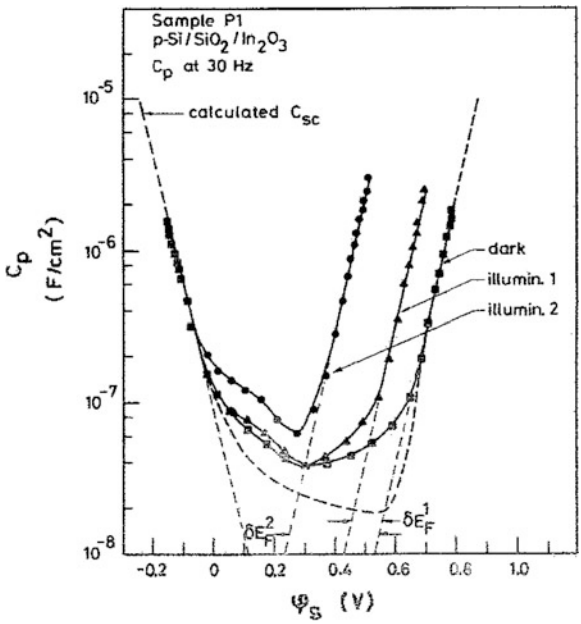


Fig. 2.47 Experimental In (C_p) as a function of the surface potential of the same device P1 as in Fig. 2.45 in *dark* and two different levels of illumination. The *broken line* represents the calculated low frequency space charge capacitance density as a function of the surface potential in the *dark condition* [81]

dominated by the electron capture by traps at $E_{FS,e}$; in that case the trap energy is given by:

$$E_t = E_v + q(\varphi_s + \phi_p) + \delta E_{FS} \tag{2.88}$$

To determine the capture cross-section from the conductance data by the procedure outlined in Sect. 2.10.3.2, one has to confirm whether a particular G_p/ω versus the φ_s characteristic represents carrier exchange with the valence band or with the conductance band. Unless the hole and the electron capture cross-sections are very unequal, generally, for p-type semiconductor, carrier exchange with the valence band will prevail in the depletion regime, and with the conduction band in the weak inversion regime, and vice versa for the n-type semiconductor. For a p-type semiconductor, if the hole exchange with the valence band dominates, then the position of the G_p/ω peak—$(\varphi_s)_p$—would move towards more positive values of φ_s with decreasing frequency, and towards more negative values if electron exchange with the conduction band dominates.

2.11 A Fundamental Basis for the Ultimate EOT

The issue of the lowest value of the EOT possible for a gate stack of an MOSFET has engaged our attention over 4 decades or longer. In olden times the question asked was how small the thickness of the SiO_2 gate dielectric could be in principle? The prevailing wisdom for a long time was that the gate dielectric had to be thick enough to prevent a direct tunneling current flowing through it. Mead, in spite of being a keen reader of the times to come, could only predict an ultimate thickness of 5 nm for the SiO_2 layer in 1972 [76]. Nicollian and Brews, notwithstanding the knowledge of the MOS basics they had, could not contemplate even in 1982 that tunneling SiO_2 layers would ever see the light of day as gate dielectrics in MOSFETs and considered the thin tunnel SiO_2 layers, leave alone the ultrathin ones, to be useless [3]. The current wisdom envisages the ultimate EOT to be around 0.5 nm; this prediction is based upon the gate leakage current and not so much on a more fundamental point of physics. It is interesting to note that the two most frequent parameters on which the ultimate EOT has been based are the gate leakage current and the sensitivity of lithography. Both these parameters have witnessed several generations; each generation has experienced a scaling down. Is there a more basic or fundamental feature which has an important bearing on the lowest EOT possible for the gate stack? In the following we will attempt to suggest that the EOT threshold for the conversion of the tunnel MOS structure to a Schottky barrier is such a feature.

Kar and Dahlke [51] and Kar [77] had classified the MOS Tunnel Diodes into two groups: (1) *Intermediate Tunnel MOS Structure* in which the minority carrier imref at the semiconductor surface was pinned to the metal Fermi level, but the majority carrier imref remained pinned to the majority carrier imref in the

semiconductor bulk; (2) *Schottky Tunnel MOS Structure* in which both the majority carrier as well as the minority carrier imref at the semiconductor surface are pinned to the metal Fermi level. In the intermediate tunnel MOS structures, the occupancy of the states at the Si-SiO_2 interface is determined by the majority carrier imref in the bulk semiconductor, whereas in the Schottky tunnel MOS structures, the occupancy of the states at the Si-SiO_2 interface is determined by the metal Fermi level. There exist two threshold values of the EOT at which the transformation from the thick (non-leaky) MOS structure to intermediate tunnel MOS structure and from the intermediate tunnel MOS structure to the Schottky tunnel MOS structure, respectively, occur [77].

For an EOT lower than the $EOT_{threshold,min}$, the time the minority carriers take to reach the semiconductor surface from the semiconductor bulk ($=\tau_{generation}+\tau_{drift}$) exceeds their time of tunneling from the semiconductor surface to the metal ($\tau_{tunneling}$), as a result of which the minority carriers are drained away to the metal, no inversion layer forms at the semiconductor surface, and the minority carrier imref at the semiconductor surface gets pinned to the metal Fermi level, as illustrated in Fig. 2.48 [78]. The minority carrier generation time depends on the semiconductor substrate and its quality; in the case of device quality silicon, a typical value of $\tau_{generation}$ could be 10^{-5} s. After generation in the semiconductor bulk (neutral region), which is the main source of the supply of the minority carriers to the semiconductor surface, the minority carriers traverse the space-charge region by drift to reach the semiconductor surface; an estimate of this drift time could be 10^{-12} s, as suggested in Fig. 2.48. Hence, as indicated in Fig. 2.48, the conversion of the thick MOS to the intermediate tunnel MOS will occur at the threshold EOT of $EOT_{threshold,min}$, for which the minority carrier tunneling time $\tau_{tunneling}$ becomes much smaller than their generation time. Naturally the value of the $EOT_{threshold,min}$ will depend upon the composition of the gate stack, the

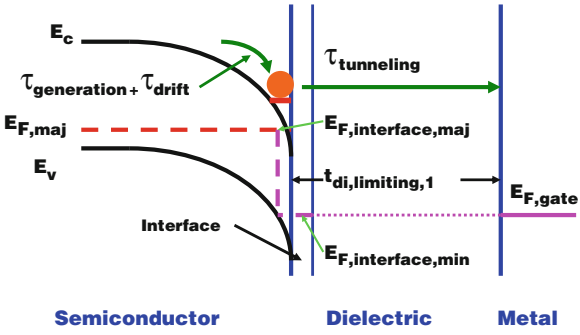


Fig. 2.48 Energy band diagram across semiconductor/dielectric/metal structure illustrating minority carrier interface imref pinning and the corresponding EOT threshold: $EOT_{threshold,min}$

$\tau_{drift} = 10^{-5}/10^{7}$ s $= 10^{-12}$ s $\tau_{generation} + \tau_{drift} \approx \tau_{generation} = 10^{-5}$ s $\tau_{tunneling} << 10^{-5}$ s $\Rightarrow$ No strong inversion layer forms in an MOS structure ; $E_{F,interface,min}$ pinned to $E_{F,gate}$; Onset of direct tunneling.

semiconductor substrate and its quality, etc.; in the case of the SiO_2 gate dielectric, it was observed to be around 3.3 nm for a device quality silicon [51, 77].

For an EOT lower than the $EOT_{threshold,maj}$, the time the majority carriers take to reach the semiconductor surface from the semiconductor bulk ($=\tau_{relaxation} + \tau_{drift}$) exceeds their time of tunneling from the semiconductor surface to the metal ($\tau_{tunneling}$), as a result of which the states on the semiconductor surface exchange carriers faster with the metal than with the semiconductor bulk, the occupancy of states at the semiconductor surface is determined by the metal Fermi level, and the majority carrier imref at the semiconductor surface gets pinned to the metal Fermi level, as illustrated in Fig. 2.49. When, this happens, the isolation of the gate electrode from the semiconductor surface and the channel (the insulated gate character) completely disappears and no surface inversion or accumulation are possible. Consequently, with such a gate stack (i.e. with a Schottky tunnel MOS), the device will cease to function as an MOSFET and will be transformed to a Schottky barrier gate FET. In other words, $EOT_{threshold,maj}$ is the lowest value the EOT of an MOSFET gate stack can have. In a Schottky barrier gate FET, no accumulation or inversion layer is possible; the semiconductor surface can only be under the depletion condition.

As indicated in Fig. 2.49, the source of supply for the majority carriers is the bulk semiconductor relaxation (given by $\tau_{relaxation}$); to reach the semiconductor surface, the majority carriers traverse the space-charge layer by drift (given by τ_{drift}). Estimates of the different components of the total time $-\tau_{relaxation} + \tau_{drift}$—it takes the states inside the conduction/valence band at the semiconductor surface to exchange carriers with the semiconductor bulk are indicated in Fig. 2.49 for silicon MOSFETs. For a semiconductor surface state inside the band-gap, an additional time—the recombination time $\tau_{R,maj}$—has to be added to $-\tau_{relaxation+\tau drift}$—to obtain the total exchange time. To estimate the threshold EOT—$EOT_{threshold,maj}$—we need to consider the smallest time the state at the semiconductor surface needs to communicate with the semiconductor bulk.

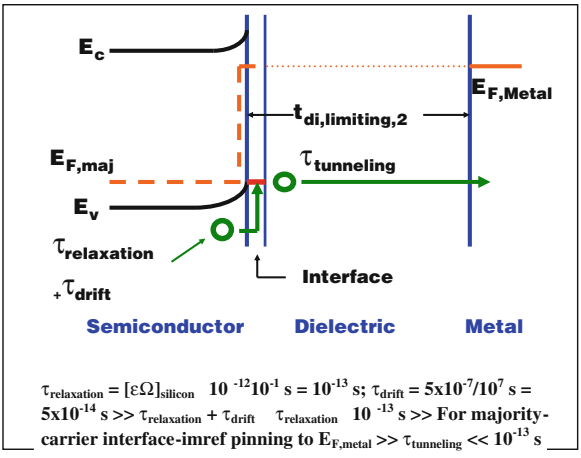


Fig. 2.49 Energy band diagram across semiconductor/dielectric/metal structure illustrating majority carrier interface imref pinning and the corresponding EOT threshold: $EOT_{threshold,maj}$

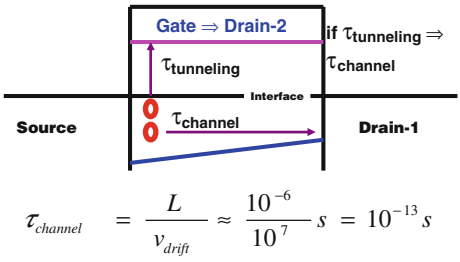


Fig. 2.50 A schematic of the MOSFET illustrating the competition between the gate metal and the regular drain to capture (by tunneling and by drift, respectively) the carrier injected by the source. The time to traverse the channel, $\tau_{channel}$, is estimated assuming a channel length of 10 nm

The estimated times in Fig. 2.49 suggest that we basically need to compare the semiconductor relaxation time with the majority carrier tunneling time. The semiconductor relaxation time will depend upon the semiconductor permittivity and resistivity; the majority carrier tunneling time will depend upon the composition of the gate stack and the EOT. When the majority carrier tunneling time becomes smaller than the semiconductor relaxation time, then the intermediate tunnel MOS transforms into a Schottky tunnel MOS, the majority carrier surface imref gets pinned to the metal Fermi level, and the gate stack is no longer functional for the MOSFET operation. The threshold EOT for which this occurs, i.e. $EOT_{threshold,maj}$, is the lowest value of EOT possible for the gate stack for MOSFET operation, unless some other fundamental consideration yields a higher value for the minimum EOT. An experimental value for the $EOT_{threshold,maj}$ is not available at this time; in the case of the SiO_2 single gate dielectric, $EOT_{threshold,maj}$ should be less than 0.8 nm, as Intel reports MOSFET operation for an SiO_2 gate dielectric of 0.8 nm [79]. Muller reports that some 0.7 nm of SiO_2 is necessary for a meaningful tunnel barrier to take shape [80].

A basic consideration which may assign a higher value for the ultimate EOT than the $EOT_{threshold,maj}$ is illustrated in Fig. 2.50. If the time to traverse the channel exceeds the time of tunneling through the gate stack, then the carriers injected by the source would rather end up in the metal electrode than in the customary drain; in other words, the gate metal will become a secondary or even primary drain, and the MOSFET will cease to function. The time estimated for the channel traverse time, $\tau_{channel}$, i.e. 10^{-13} s, in Fig. 2.50 is about the same as the semiconductor relaxation time estimated in Fig. 2.49. The channel traverse time would be determined by the channel length, the electric field along the y axis, and the channel mobility.

2.12 Summary

The material presented in this chapter can be classified into two groups—one set for the MOS/MOSFET devices with the SiO_2 single gate dielectric and the other set for the same with the high-k gate stacks. What comes out clearly and strongly is the large difference in the nature, behavior, and the characteristics of the two sets;

almost everything deviates strongly—be it the energy band profile, the circuit representation, the traps, the parameter extraction, and the mathematical relations for the channel parameters, for the gate stack charges, and for the gate stack potentials. In the case of the high-k gate stacks, the channel parameters—the drain current, the channel conductance, and the transconductance—are severely degraded by the non-ideal factors of the work-function anomaly, the high-k gate stack charges, and the non-saturating inversion surface potential; in addition, the drain current versus the drain voltage relation becomes more non-linear even in the triode regime. The energy band profile across the high-k gate stack is enormously complicated by its composition—two bulk dielectrics (intermediate layer—IL—and the high-k layer) and three transition layers with varying composition, varying band-gap, and varying permittivity respectively at the semiconductor/IL, IL/high-k, and high-k/metal interfaces. The corresponding circuit representations and the charge-potential relations are equally complicated; in fact, it is not possible to represent the transition region of varying permittivity with a finite number of circuit elements. The theoretical treatment and the mathematical relations are complicated also by the quantum-mechanical phenomena. The analysis presented in this chapter on the latter suggest a strong dilution of the carrier confinement in the strong inversion and the accumulation layers by wave function penetration into the gate stack, tunneling through the gate stack, and the metal/semiconductor wave function mixing. The concept of the pseudo-Fermi potential and the Fermi occupancy of traps inside the gate stack have been analyzed; this analysis suggests that it is likely that all the traps communicate and exchange carriers with either the semiconductor or the metal surface in accumulation and in strong inversion if the EOT is <1.0 nm and the ac signal frequency is <1 MHz. In such a case, for a measurement frequency of 1 MHz or less, the MOS capacitance in accumulation or in strong inversion will exceed the dielectric capacitance of the gate stack, since it will be augmented by the charging capacitance of the gate stack traps. An erroneous and false value of EOT will result if the same is extracted from this capacitance. Moreover, the augmentation of the dielectric capacitance by the charging capacitance of the gate stack traps can explain the huge frequency dispersion of the accumulation or the strong inversion capacitance observed particularly in the case of the high-mobility channels. The C–V characteristic of the ultrathin gate stack deviates strongly from the classical form. The classical MOS parameter extraction techniques do not apply in the case of the ultrathin gate stacks and need to be modified; one reason for this is the inability to apply the quasi-static C–V technique on account of the high gate leakage current.

References

1. A.S. Grove, *Physics and Technology of Semiconductor Devices* (Wiley, New York, 1967)
2. S.M. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981)
3. E.H. Nicollian, J.R. Brews, *MOS Physics and Technology* (Wiley, New York, 1982)

4. E.H. Nicollian, A. Goetzberger, The Si-SiO_2 interface—electrical properties as determined by the MIS conductance technique. Bell Syst. Tech. J. **46**, 1055 (1967)
5. E.H. Rhoderic, R.H. Williams, *Metal-Semiconductor Contacts* (Clarendon Press, Oxford, 1988)
6. S.M. Sze (ed.), *Modern Semiconductor Devices* (Wiley, New York, 1998)
7. H. Lueth, *Surfaces and Interfaces of Solid Materials* (Springer, Berlin, 1995)
8. C.G.B. Garrett, W.H. Brattain, Physical theory of semiconductor surfaces. Phys. Rev. **99**, 376 (1955)
9. S. Kar, Interface charge characteristics of MOS structures with different metals on steam grown oxides. Solid-St. Electron. **18**, 723–732 (1975)
10. S. Kar, Determination of Si-metal work function differences by MOS capacitance technique. Solid-St. Electron. **18**, 169–181 (1975)
11. H.K.J. Ihantola, J.L. Moll, Solid-State Electron. **7**, 423 (1964)
12. S.R. Hofstein, F.P. Heiman, Proc. IEEE **51**, 1190 (1963)
13. C.G. Parker, G. Lucovsky, J.R. Hauser, Ultrathin oxide–nitride gate dielectric MOSFET's. IEEE Electron Device Lett. **19**(4), 106 (1998)
14. S. Kar, M. Houssa, S. Van Elshocht, D. Misra, K. Kita (eds.), Physics and technology of high-K materials IX. ECS Trans. **41**(3) (2011), ch. 8, ch. 7
15. S. Kar, S. Van Elshocht, D. Misra, K. Kita (eds.), Physics and technology of high-K materials X. ECS Trans. **41**(3) 2012
16. S. Kar, S. Rawat, ECS Trans. **16**(5), 443 (2008)
17. S. Kar, ECS Trans. **25**(8), 399 (2009)
18. T. Hori, *Gate Dielectrics and MOS ULSIs* (Springer, Berlin, 1997). (ch. 3)
19. S.M. Sze, *Modern Semiconductor Device Physics* (Wiley, New York, 1998). (ch. 3)
20. C.C. Hu, *Modern Semiconductor Devices for Integrated Circuits* (Prentice Hall, Upper Saddle River, 2009). (ch. 6)
21. S. Kar, IEEE Trans. Electron Devices **50**, 2112 (2003)
22. S. Kar, S. Rawat, S. Rakheja, D. Reddy, IEEE Trans. Electron Devices **52**, 1187 (2005)
23. R. Choi, S.J. Rhee, J.C. Lee, B.H. Lee, G. Bersuker, IEEE Electron Device Lett. **26**, 197 (2005)
24. G. Bersuker et al., IEEE Trans. Device Mater. Reliab. **7**, 138 (2007)
25. A. Toriuma, K. Kita, ECS. Trans. **19**(1), 243 (2009)
26. J.K. Schaeffer et al., Appl. Phys. Lett. **85**, 1826 (2004)
27. H. Park et al., IEEE Electron Device Lett. **26**, 725 (2005)
28. R. Xie, T.H. Phung, W. He, M. Yu, C. Zhu, IEEE Trans. Electron Devices **ED-56**, 1330 (2009)
29. Y.-T. Chen et al., Appl. Phys. Lett. **96**, 253502 (2010)
30. Y. Wang et al., ECS Trans. **33**, 487 (2010)
31. Y. Wang et al., ECS Trans. **41**, 243 (2011)
32. R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, M. Metz, High-k/Metal–gate stack and its MOSFET characteristics. IEEE Electron Device Lett. **25**(6), 408 (2004)
33. N. Nakagawa, H. Y. Hwang, and D. A. Muller, Nat. Mater. **5**, 204 (2006)
34. H. Watanabe, D. Matsushita, K. Muraoka, IEEE Trans. Electron Devices **53**, 1323 (2006)
35. S. Borowitz, *Fundamentals of Quantum Mechanics* (W. A. Benjamin, New York, 1967)
36. E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1961)
37. T. Ando, A. B. Fowler, F. Stern, Rev. Mod. Phys. **54**, 437 (1982)
38. H. Lueth, *Surfaces and Interfaces in Solid Materials* (Springer, Berlin, 1995)
39. C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1967)
40. J.R. Hauser, K. Ahmed, Characterization of ultra-thin oxides using electrical CV and IV measurements, in *International Conference on Characterization and Metrology for ULSI Technology Proceedings* (1998) pp. 235–239
41. S. Krishnamurthy, S. Jallepalli, C.-F. Yeap, K. Hasnat, A.F. Tasch, C.M. Maziar, A computationally efficient model for inversion layer quantization effects in deep submicron N-channel MOSFETs. IEEE Trans. Electron Devices **43**(1), 90–96 (1996)

42. L.F. Register, A.F. Tasch, S.K. Banerjee, Understanding the effects of wave function penetration on the inversion layer capacitance of NMOSFETs. Electron Device Lett. **22**(3), 145–147 (2001)
43. V. Heine, Theory of surface state. Phys. Rev. **138**, A1689–A1696 (1965). MIGS
44. A. Rose, *Concepts in Photoconductivity and Allied Problems* (Wiley Interscience, New York, 1963)
45. D. Muñoz Ramo, J.L. Gavartin, A.L. Shluger, G. Bersuker, Phys. Rev. B **75**, 205336 (2007)
46. I.E. Tamm, Z. Physik **76**, 849 (1932)
47. W. Shockley, Phys. Rev. **56**, 317 (1939)
48. G.J. Gerardi, E.H. Poindexter, P.J. Caplan, N.M. Johnson, Interface traps and Pb centers in oxidized (100) silicon wafers. Appl. Phys. Lett. **49**, 348 (1986)
49. A.H. Edwards, Theory of the P_b center at the <111> Si/SiO_2 interface. Phys. Rev. B **36**, 9638 (1987)
50. J. Dong, D.A. Drabold, Atomistic structure of band-tail states in amorphous silicon. Phys. Rev. Lett. **80**(9), 1928–1931 (1998)
51. S. Kar, W.E. Dahlke, Interface states in MOS structures with 20–40 Åthick SiO2 films on non-degenerate Si. Solid-State Electron. **15**, 221–232 (1972)
52. A.V. Kimmel, P.V. Sushko, A.L. Shluger, G. Bersuker, ECS Trans. **19**(2), 3 (2009)
53. A. Toriumi, K. Kita, ECS Trans. **19**(1),243 (2009)
54. H. Jagannathan, V. Narayanan, S. Brown, ECS-Trans. **16**, 19 (2008)
55. J. Tersoff, Schottky barrier heights and the continuum of gap states. Phys. Rev. Lett. **52**, 465–468 (1984)
56. M.R. Visokay, J.J. Chambers, A.L.P. Rotondaro, A. Shanware, L. Colombo, Appl. Phys. Lett. **80**, 3183 (2002)
57. J.-H. Lee et al., in *2002 Symposium on VLSI Technology Digest of Technical Papers*, 2002
58. Y.H. Wu, M.Y. Yang, A. Chin, W.J. Chen, C.M. Kwei, IEEE Electron Device Lett. **21**, 341 (2000)
59. H. Harris, K. Choi, N. Mehta, A. Chandolu, N. Biswas, G. Kipshidze, S. Nikishin, S. Gangopapadhyay, H. Temkin, Appl. Phys. Lett. **81**, 1065 (2002)
60. C.N. Berglund, IEEE Trans. Electron Devices **13**, 701 (1966)
61. L.M. Terman, Solid-State Electron. **5**, 285 (1962)
62. A. Ali, H. Madan, S. Koveshnikov, S. Datta, ECS Trans. **25**(6), 271 (2009)
63. M. Heyns et al., ECS Trans. **25**(6), 51 (2009)
64. S. Kar, C. Miramond, D. Vuillaume, Properties of electronic traps at silicon/1-octadecene interfaces. Appl. Phys. Lett. **78**, 1288 (2001)
65. J. McNutt, C.T. Sah, J. Appl. Phys. **46**, 3909 (1975)
66. J. Maserjian, G. Petersson, C. Svensson, Solid-State Electron. **17**, 335 (1974)
67. J. Maserjian, in *The Physics and Chemistry of SiO2 and the Si/SiO2 Interface*, ed. by C.R. Helms, B.E. Deal (Plenum Press, New York, 1988)
68. B. Ricco, P. Olivo, T.N. Nguyen, T.-S. Kuan, G. Ferriani, IEEE Trans. Electron Devices **35**, 432 (1988)
69. K. Ahmad, E. Ibok, G. Bains, D. Chi, B. Ogle, J.J. Wortman, J.R. Hauser, IEEE Trans. Electron Devices **47**, 1349 (2000)
70. J.S. Brugler, P.G.A. Jespers, IEEE Trans. Electron Devices **16**, 207 (1969)
71. G. Groeseneken, H.E. Maes, N. Beltran, R.F. deKeersmaecker, IEEE Trans. Electron Devices **31**, 42 (1984)
72. S. Kar, S. Rawat, ECS Trans. **16**(5), 111 (2008)
73. K. Shiraishi, K. Yamada, K. Torii, Y. Akasaka, K. Nakajima, M. Konno, T. Chikyo, H. Kitajima, T. Arikado, Y. Nara, Thin Solid Films **508**, 305 (2006)
74. T.C. Poon, H.C. Card, J. Appl. Phys. **51**, 6273 (1980)
75. S. Kar, S. Varma, J. Appl. Phys. **58**, 4256 (1985)
76. B. Hoeneisen, C.A. Mead, Solid-State Electron. **15**, 819 (1972)
77. S. Kar, Characterization of silicon MOS tunnel diodes. IEDM Tech. Dig. **79** (1976)

78. S. Kar, Two limiting thinnesses of the ultrathin gate oxides, in *Silicon Nitride and Silicon Dioxide Thin Insulating Films*, ed. by K.B. Sundaram, M.J. Deen, D. Landheer, W.D. Brown, D. Misra, M.D. Allendorf, R.E. Sah, Electrochem. Soc. Proc., vol PV-2001-7, **60** (2001)
79. R. Chau, B. Boyanov, B. Doyle, M. Doczy, S. Datta, S. Hareland, B. Jin, J. Kavalieros, M. Metz, Silicon nano-transistors for logic applications. Physica E **19**, 1 (2003)
80. D.A. Muller, T. Sorsch, S. Moccio, F.H. Baumann, K. Evans-Lutterodt, G. Timp, The electronic structure at the atomic scale of ultrathin gate oxides. Nature **399**, 758 (1999). 24 June
81. S. Kar, S. Varma, J. Appl. Phys. **54**, 1988 (1983)
82. R. Nieh, R. Choi, S. Gopalan, K. Onishi, C.S. Kang, H.-J. Cho, S. Krishnan, J.C. Lee, Evaluation of silicon surface nitridation effects on ultra-thin ZrO_2 gate dielectrics. Appl. Phys. Lett. **81**, 1663–1665 (2002)
83. C.H. Lee, J.J. Lee, W.P. Bai, S.H. Bae, J.H. Sim, X. Lei, R.D. Clark, Y. Harada, M. Niwa, D.L. Kwong, Self-aligned ultra thin HfO2 CMOS transistors with high quality CVD TaN gate electrode, in *2002 Symposium on VLSI Technology Digest of Technical Papers*
84. Y.-S. Lin, R. Puthenkovilakam, J.P. Chang, Dielectric property and thermal stability of HfO_2 on silicon. Appl. Phys. Lett. **81**, 2041–2043 (2002)

Chapter 3
Hafnium-Based Gate Dielectric Materials

Akira Nishiyama

Abstract In this chapter, we focus on hafnium-based gate dielectrics. HfO_2 is regarded as the most promising material for the high–k gate dielectrics owing to its large dielectric constant and large band-gap energy. In the first part of this chapter, these characteristics are addressed in a comparison with SiO_2 and other high-k materials. Thermal stability is a major issue for the application of high-k materials to MOSFETs in LSIs. Although HfO_2 satisfies this requirement, severer process conditions may cause problems even with this material. Suppression of these issues is also addressed in the second part of this chapter. In order to enhance the characteristics of HfO_2, transformation of the monoclinic phase to the tetragonal and the cubic phases with larger dielectric constant has been pursued recently. We mention the issue in the last part of this chapter.

3.1 Introductory Remarks and Brief Outline of the Chapter

Since the proposal of the materials in the early 2000s [1–3], hafnium-based gate dielectrics have been regarded as the most promising materials for the application to large scale integrations (LSIs). In 2007, Intel announced the start of manufacturing of 45 nm-node LSIs with high-k/metal gate stack, with the remark that the world's first high-k material in commercial production was hafnium-based [4].

There are several characteristics that the high-k materials should meet for the application to metal oxide semiconductor field effect transistors (MOSFETs) in LSIs. First, we clarify the properties so that readers can understand their

A. Nishiyama (✉)
Corporate R&D Center, Toshiba Corporation, 1 Komukai-toshiba-cho, Saiwai-ku, Kawasaki 212-8582, Japan
e-mail: akira.nishiyama@toshiba.co.jp

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5_3,

importance in the selection process of high-k materials. Next, we present concrete data on HfO_2, and identify the reasons why this material is considered the most promising candidate. This is followed by data on its alloy with SiO_2 or Al_2O_3. Along with pure HfO_2, these two alloys have been investigated intensively by many companies, institutes and universities world-wide in order to enhance thermal stability and compatibility with the LSI processes. Since HfO_2 takes several phases and they influence the high-k properties, this topic is also addressed in this chapter. Taking the higher-k HfO_2 phase is regarded as one of the measures to extend the usability of this material through several LSI generations.

3.2 Properties Required for Gate Dielectrics

Listed in Fig. 3.1 are the properties required for high-k materials as alternative gate dielectrics of advanced ultra large scale integrations (ULSIs). They are large dielectric constant, large band-gap energy, large effective mass for tunneling carriers, good integrity of high-k/substrate interface, high reliability against electrical stresses, and compatibility with ULSI processes. Obviously, good uniformity of properties throughout the wafer is essential.

The probabilities of the quantum mechanical tunneling of electrons and holes are governed by the integral of the barrier height that carriers encounter in their attempt to tunnel through the insulators [5]. The simplified equation of the direct tunneling J_g through a gate dielectric is presented in Ref. [6] as follows:

$$J_g \propto \exp\left(-\frac{4\pi\sqrt{2qm^*\phi_b}}{h}\frac{k}{3.9}t_{EOT}\right), \tag{3.1}$$

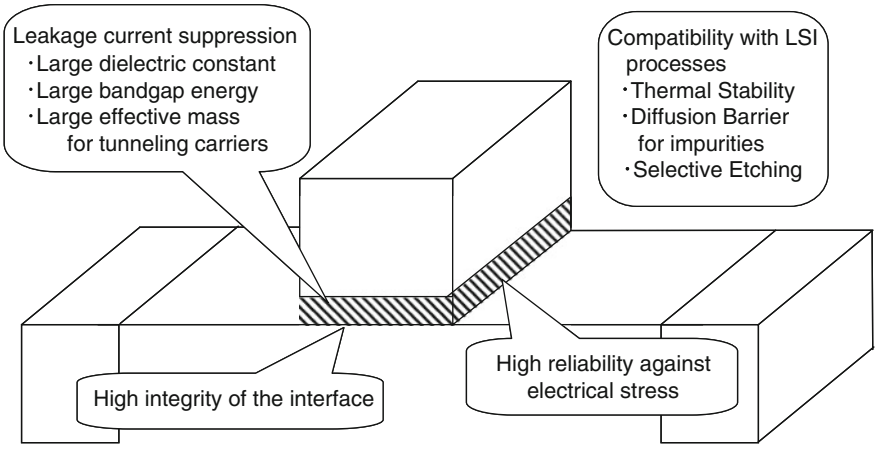


Fig. 3.1 Properties required for high-k materials applied as alternative gate dielectrics of advanced MOS transistors

where t_{EOT} and ϕ_b, k and 3.9 correspond to the electrical thickness of the gate dielectric, the energy barrier height formed at the gate dielectric/substrate interface, dielectric constant of the gate insulator and that of SiO_2, respectively. m*, q, and h correspond to the effective tunneling mass of the carrier, elementary charge and the Planck constant. The electrical thickness is called the 'equivalent oxide thickness (EOT)', which is the SiO_2 thickness equivalent to the high-k gate dielectric inducing the same amount of carrier density in the channel region with the same gate voltage.

According to the International Technology Roadmap for Semiconductors (ITRS), the EOT decreases as the device scaling proceeds in order to maintain a high sheet carrier concentration with lower voltages [7]. However, this leads to a drastic increase in power consumption due to the severe gate leakage current from the smaller t_{EOT} in (3.1) [7]. Therefore, material with a larger dielectric constant is used. The physical thickness of high-k gate dielectric t_{gate} with the dielectric constant k is calculated using the following equation:

$$t_{gate} = \frac{k}{3.9} t_{EOT} \tag{3.2}$$

3.9 is the dielectric constant of SiO_2. Equation (3.2) indicates the thicker physical thickness is obtained with larger dielectric constant and this leads to a decrease in J_g.

In addition to the physical thickness, the energy barrier formed between the high-k material and the substrate should be large. Since silicon has a conduction band edge and a valence band edge at about 4.0 and 5.1 eV from the vacuum level, the conduction band edge of high-k material should be much nearer the vacuum level, whereas the valence band edge should be further from the vacuum level, in order to diminish the tunneling probability of electron transport from the conduction band of silicon in n-channel MOSFETs (NMOS) and hole transport from the valence band of silicon in p-channel MOSFETs (PMOS). A schematic view of the band alignment concept for NMOS is shown in Fig. 3.2.

The tunneling probability is also governed by the effective mass of the carriers: a heavier mass leads to lower probability. Therefore, the high-k materials through which carriers tunnel with heavier mass are desirable for the gate insulator. The effective mass of electrons in HfO_2 was extracted with the fit between the experimental tunneling leakage current data and theoretical calculations. Those studies have provided figures such as 0.22 m_0 [6], 0.15 m_0 [8], or 0.1 m_0 [9], where m_0 represents the free electron mass. It should be noted that the tunneling mass of electrons through SiO_2 is as large as 0.4 m_0 [6].

Using the function inside the brackets in (3.1), we can define the figure of merit (F.O.M.) of the high-k dielectrics as follows [6]:

$$F.O.M. \equiv k\sqrt{m^* \phi_b} \tag{3.3}$$

A large value of F.O.M. of a high-k material indicates its high potential to diminish the amount of J_g with the fixed t_{EOT} value.

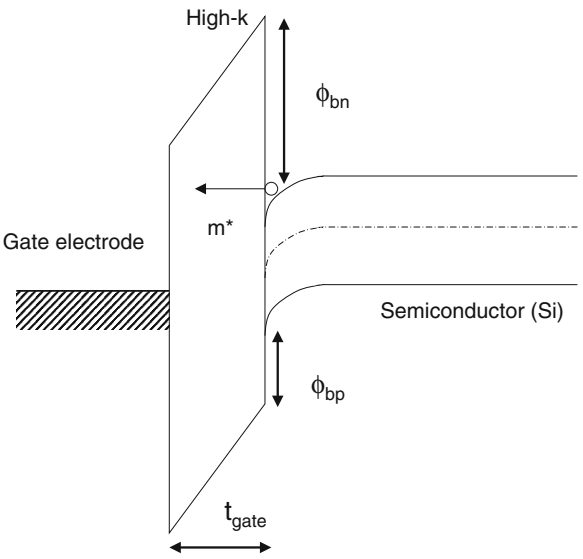


Fig. 3.2 Schematic view of the concept of the band alignment among gate electrode, high-k material, and semiconductor substrate. ϕ_{bn}, ϕ_{bp} and m* represent energy barrier for electrons, energy barrier for holes and effective mass for tunneling carriers, respectively

Since the integrity of high-k/semiconductor substrate interface and film reliability issues against electrical stresses are discussed in other chapters, the thermal stability issue is addressed in this chapter. The LSI processes consist of several types of high-temperature annealing such as oxidation annealing, nitrogen annealing, and chemical vapor deposition. When the high-k dielectrics are applied in the conventional LSI processes, a few high-temperature processes including the source/drain impurity activation annealing are performed after the deposition process. Currently, several methods such as spike rapid thermal annealing, flash lamp annealing, and laser thermal annealing are proposed in order to limit the thermal budget of the impurity activation process. The concept of these processes is that thermal energy is provided to incorporate impurity atoms in the substrate crystalline network, limiting the diffusion of the atoms in the device structures. Since the typical temperature does not go below 900 °C, high-k materials on semiconductor substrate should maintain their above-mentioned properties through the high-temperature annealing. It is particularly important to avoid a thermal reaction between high-k materials and the substrate, since such a reaction ruins the integrity of the high-k properties, in many cases.

3.3 Physical and Electrical Properties of Hafnium Oxide

Hafnium is a transition metal with a steel grey color. The atomic number of this element is 72. With its very high neutron-capture cross-section, it is regarded as a good material for the control rods of nuclear reactors [3]. Like other elements in group IV (titanium and zirconium), it is a very reactive material in air and forms

several phases of oxides. Among several oxide phases of hafnium, HfO_2, called 'hafnia', is the only stable compound showing the insulating behavior.

3.3.1 Dielectric Constant and Microscopic Polarization

In LSI processes, HfO_2 is deposited either by physical vapor deposition (PVD) such as sputtering [2], chemical vapor deposition (CVD) [1], oxidation of Hf metals [10], or by atomic layer deposition (ALD) [11–14]. In many cases, HfO_2 forms monoclinic phase just after the deposition process. Even in the case in which it has amorphous phase after deposition, low-temperature annealing transforms the material to polycrystalline monoclinic HfO_2 [10].

In the monoclinic crystal structure, the coordination number of Hf is 7, whereas that of oxygen is 3 or 4 [15]. The static dielectric constant of monoclinic HfO_2 ranges from 17 [1, 2] to 22 [16]. Throughout this chapter, 'dielectric constant' means relative permittivity k. The absolute permittivity ε is the vacuum permittivity ε_0 multiplied by this number. The dielectric constant of materials comes from a few components as shown in Fig. 3.3.

For high-k gate dielectrics, the ionic and the electronic polarization components are important, because they follow a high clock frequency (about GHz order) used in advanced ULSIs. The electronic polarization component of the dielectric constant can be extracted using the following equation:

$$k_e = n^2, \tag{3.4}$$

where n represents reflectivity. This component is called the 'the dynamic dielectric constant'. This is also called the 'optical dielectric constant' because it is the dielectric constant in the optical frequency region. Knowing that the reflectivity of HfO_2 is about 2.1 [17], the electronic polarization component of this material is estimated to be around 4.4. Since this value is small compared to the static dielectric constant value [1, 2, 14], one can assume that the reason for the high static dielectric constant of this material resides in a large ionic polarization.

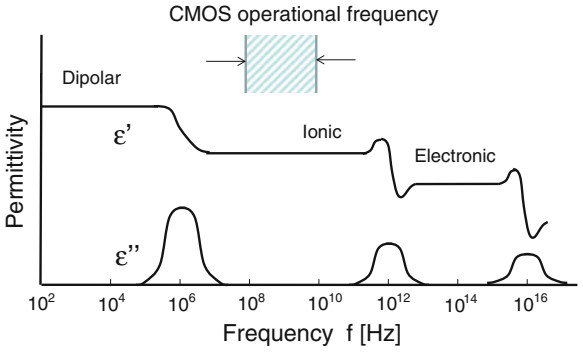


Fig. 3.3 The frequency dependence of permittivity. ε' represents the real part, while ε'' represents the imaginary part of the permittivity. Since advanced ULSI operation clock frequency is in the GHz domain, ionic and electronic polarization contributes to the permittivity of high-k materials at the frequencies

The Clausius–Mossotti theory correlates the static dielectric constant k and the structural parameters as follows:

$$\frac{k-1}{k+2} = \frac{4\pi\alpha}{3V_m}, \tag{3.5}$$

where V_m and α are the molar volume in $Å^3$ and the polarizability of each molecule, respectively. This equation can be rewritten as follows:

$$k = \frac{1 + \frac{8\pi\alpha}{3V_m}}{1 - \frac{4\pi\alpha}{3V_m}} \tag{3.6}$$

This clearly states that the increase in α and the decrease in V_m lead to the increase in the dielectric constant of the material. In the case of HfO_2, the polarizability of each molecule and molar volume is about 7.3 and 35 $Å^3$, and its combination results in large dielectric constant among many metal oxides [18]. The polarizability of each molecule and molar volume of SiO_2 is 4.8 and 38 $Å^3$, for comparison [18]. Although there is little difference between the molecular volumes of the two materials, the polarizability is much larger for HfO_2 than for SiO_2. This large polarization can be ascribed to the large difference between the electro-negativity of Hf and O: Hf electronegativity is 1.3 (Pauling), whereas that of oxygen is 3.44 [19]. It should be noted that the electronegativity of Si is as large as 1.9, which is larger than that of Hf and nearer to the value of oxygen (Table 3.1).

3.3.2 Bandgap and Band Alignment with Silicon

As mentioned in Sect. 3.2, band-gap energy of a high-k material is the basic parameter influencing the F.O.M. of the material. Figure 3.4 shows the relationship of band-gap energy and dielectric constant of metal oxides.

The general trend of these two parameters indicates that as the dielectric constant of material increases, the bandgap energy decreases. Considering that the transition probability of electrons from the stable states to the excited states increases owing to the decrease in the bandgap, it can be understood that the

Table 3.1 Pauling's electronegativity for typical cation and anion elements [19]

Cation							
Al	Si	Ti	Y	Zr	La	Hf	Ta
1.6	1.9	1.5	1.2	1.3	1.1	1.3	1.5
Anion							
N				O			
3				3.4			

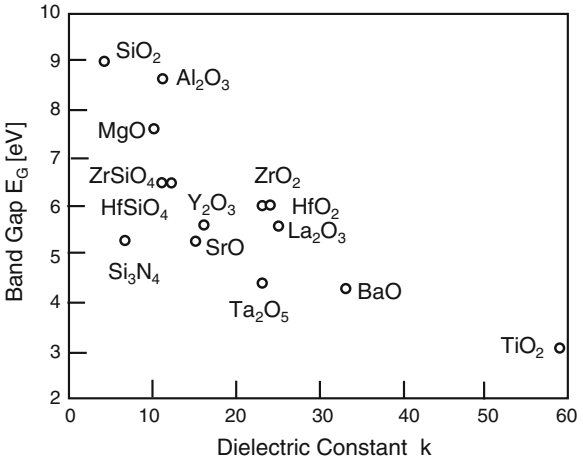


Fig. 3.4 Correlation of static dielectric constant with band gap of metal oxides. (Reprinted with permission from [26]. Copyright 2002, American Institute of Physics)

dynamic dielectric constants increase for high-k materials with narrower bandgap energies [20].

Bandgap energy of materials can be measured with a few analytical techniques such as XPS (X-ray photoelectron spectroscopy) [21], ellipsometry [22], and REELS (reflection electron energy loss spectroscopy) [23]. The measured bandgap of the monoclinic HfO_2 is 5.7 eV [22] or 5.8 eV [23, 24]. Although the bandgap is smaller than that of SiO_2 (about 9 eV) [23, 25], the value is still large among high-k materials. Theoretical calculation of this material elucidates that its conduction band originates from 5d band of Hf, whereas the valence band originates from 2p states of oxygen [26].

XPS measurement is also a very useful technique for the extraction of the band alignment of high-k materials to the underlying substrate. It is well known that work functions of pure metals are not realized when they contact semiconductors and form metal/semiconductor systems. This is certainly the case for a semiconductor with a relatively small bandgap such as Si (about 1.1 eV) [27]. For example, aluminum, which has been used as the interconnection metal in LSIs for a long time, has bulk work function of about 4.0 eV; however, its effective work function in Al/Si system is about 4.7 eV [27]. This discrepancy is explained using the CNL (Charge Neutrality Level) of semiconductor and the formation of dipole at the metal/semiconductor interface [28]. A similar theory is applicable to the insulator/semiconductor systems using the CNL of insulator as an alternative to Fermi level of metal in metal/semiconductor systems [26].

Normally, valence band spectra from XPS measurements are used to extract the band alignment of the high-k/semiconductor systems [29]. This measurement reveals binding energies of the tops of the valence bands for high-k materials and the underlying semiconductor substrate. The important point is that those two binding energy values come from one sample. By calculating the energy difference of the two, one can extract valence band offset between them. In addition, if the band-gap energies of two materials are known by means of the other methods

mentioned above, the conduction band offset can be readily obtained. See Fig. 3.2 for the reference of this procedure. Another common method of extracting the band offset is IPES (internal photo-emission spectroscopy) measurement: When electrons are provided with sufficient energy to overcome the band offset, they are detected as the current [30]. The extracted conduction band offset of HfO_2 was 1.2 eV [31] or 1.9 eV [32]. This relatively large value along with the large dielectric constant promises a dramatic decrease in leakage current by 4 or 5 orders of magnitude through HfO_2, compared to that through SiO_2 with the same EOT [1, 2].

3.3.3 Compatibility with LSI Processes

Another important factor considered in high-k material selection is the thermal stability in the ULSI fabrication processes. 'Thermal stability' has several aspects as follows:

1. The melting point of the material must be higher than the temperature in ULSI fabrication process
2. The metal oxide should not be reduced by the neighboring materials including the substrate semiconductor during ULSI fabrication processes
3. Reaction with neighboring materials other than reduction should not occur in ULSI fabrication process.

Hubbard et al. [33] used these criteria to select possible high-k insulators among many metal oxides. Regarding the second and the third criteria, they supposed the following reactions and tried to find metal oxides (MOx) whose reaction Gibbs free energies, ΔG, become positive. With positive ΔG value, these reactions become endothermic, indicating that the metal oxide is stable in contact with the substrate silicon. Through this procedure, they selected several metal oxide candidates, including HfO_2.

$$MO_x + Si \rightarrow M + SiO_2 + \Delta G \quad (R1)$$

$$MO_x + Si \rightarrow MSi_x + SiO_2 + \Delta G \quad (R2)$$

With a melting point of 2,758 °C and a boiling point of 5,400 °C, HfO_2 is expected to be a very stable material. Moreover, Hubbard et al. calculated the Gibbs free energy of R1 at 1,000 K and concluded that R1 reactions are not expected for HfO_2 at least at this temperature. They could not calculate Gibbs free energy of R2 reactions for HfO_2, because of the lack of thermodynamic data for $HfSi_2$. However, given that R1 and R2 reactions are both endothermic for ZrO_2 at 1,000 K, one can expect the stability of HfO_2 against R2 reactions, considering the similarity of the two materials. This high thermal stability along with the drastic

reduction of the gate leakage current makes HfO_2, as well as ZrO_2, a promising material for the high-k gate dielectrics in advanced ULSIs.

However, in ULSI processes, severer process condition may cause problems even for these materials: zirconium silicide formation was reported when high-temperature annealing (around 950–1,000 °C) was performed under high vacuum or in an oxygen-deficient ambient [34–36] as shown in Fig. 3.5 [37]. Although silicide formation temperature is higher for HfO_2 than for ZrO_2 [38], a similar reaction was detected [39, 40]. A different reaction model that includes reduction of the interfacial SiO_2 layer was proposed as follows (R3) [41] and this may be the probable reason for the increase of the leakage current of HfO_2/SiO_2 gate stack after 950 °C annealing in N_2 ambient [42]:

$$HfO_2 + SiO_2 + Si \rightarrow HfSi_x + 2SiO(g) + Si \quad (R3)$$

Another issue concerning the compatibility of HfO_2 with ULSI process is the severe diffusion of oxygen through the film [43, 44]. High-temperature annealing in oxygen-containing atmosphere is commonly practiced to improve the leakage characteristics of the gate stack. Moreover, the residual atmosphere in the production furnaces contains oxygen. The oxygen diffusion during high-temperature annealing in oxygen-containing ambient leads to EOT increase due to the interfacial layer growth and ruins the leakage suppression characteristics. The experimental result shows that the activation energy of this phenomenon is low, 0.6 eV [45]. This fast interfacial layer growth is attributed to the fast oxygen diffusion in HfO_2 and the enhanced oxidation at the interface owing to the catalytic function of the HfO_2 layer [40, 46, 47].

The diffusion of common impurities such as arsenic and boron would pose a problem when poly-Si gate electrode is used with the high-k material [48]. When poly-silicon is used as the gate electrode, it must be highly doped with phosphorous and arsenic for NMOSFETs and with boron for PMOSFETs to reduce the sheet resistance of the electrode and adjust the threshold voltage. Thin HfO_2 cannot work as the diffusion barrier for arsenic [49] or boron [50, 51] and this diffusion leads to unintentional decrease in the threshold voltage or even to severe leakage current between source and drain (punch-through phenomenon).

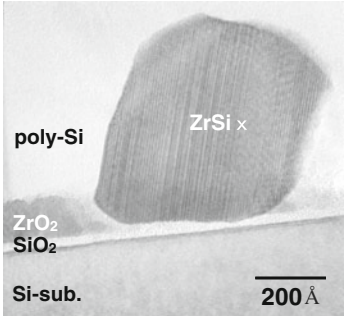


Fig. 3.5 ZrO_2/SiO_2 gate stack reacts with Si substrate when annealed in a N_2 ambient at 1,000 °C, resulting in the $ZrSi_x$ formation [37]

Therefore, suppression of the above-mentioned issues is the incentive for incorporation of other elements such as nitrogen, aluminum and silicon into HfO_2. The change in the material's properties as a result of their incorporation is also addressed in the following sections.

3.4 Hafnium–Nitrogen-Based Gate Dielectrics

Nitrogen has been introduced into HfO_2 in order to enhance the thermal stability of the material. HfON is deposited by N_2/O_2 reactive sputtering using Hf target [52]. HfON can also be formed by N_2 reactive sputtering from Hf target followed by rapid thermal oxidation [42]. CVD deposition of this material has also been reported [53]. Kang et al. [42] claimed that nitrogen incorporation at the gate dielectric/Si interface region and SiN formation may lead to the suppression of SiO formation in the reaction R3 mentioned above, resulting in the suppression of the leakage current increase. The suppression of diffusion of oxygen [42, 43] as well as impurities such as boron [48, 53, 54] through the film has also been reported to be a beneficial influence of the nitrogen incorporation.

As in the case of SiON [55], nitrogen incorporation reduces the bandgap of HfO_2. This is ascribed to the replacement of the valence band edge from O2p orbital to N2p orbital [56]. Although it leads to the decrease of its valence-band barrier offset to Si substrate by about 1.1 eV, a sufficiently large energy barrier (about 2.2 eV) is retained to suppress the hole injection from the substrate in PMOSFETs [56] (Fig. 3.6).

Pure nitrides of fourth-column transition metals (Ti, Zr and Hf) share similar characteristics. All form MN-type nitrides with NaCl structure, and with more nitrogen in the structures, they form M_3N_4-type nitrides. Hafnium pure nitrides

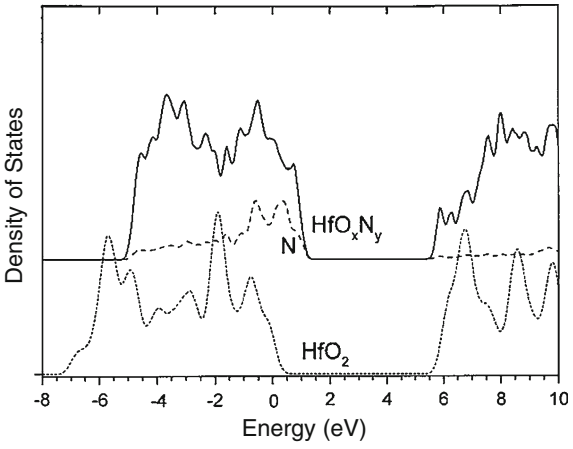


Fig. 3.6 Calculated densities of states for HfO_2 and HfON. Band-gap modification due to the nitrogen incorporation into HfO_2 is shown. (Reprinted with permission from [56]. Copyright 2004, American Institute of Physics)

such as HfN [57, 58] and Hf_3N_4 [59, 60] have been reported. Among them, Hf_3N_4 shows insulating behavior. It is reported that Hf_3N_4 may solidify in different crystalline structures according to temperature and pressure around the material [61, 62]. First-principle calculation predicted the band-gap of Hf_3N_4 to be as large as 1.8 eV [62] and electrical measurement of CVD Hf_3N_4 revealed the dielectric constant to be about 30 [60]. Since HfON can be regarded as an alloy of HfO_2 and Hf_3N_4 [63], intermediate characteristics are anticipated for HfON. Having thoroughly investigated the physical properties of HfON with various nitrogen content, Ino et al. concluded that the above-mentioned Clausius–Mossotti equation (3.6) describes the dielectric constants of the alloys well [64]. They calculated the static polarizabilities α_{total} using the additive rule of the static polarizability of each constituent element [18] as follows:

$$\alpha_{total} = C_{Hf}\alpha_{Hf^{4+}} + C_O\alpha_{O^{2-}} + C_N\alpha_{N^{3-}}, \tag{3.7}$$

where C_{Hf}, C_O, and C_N represent content of hafnium, oxygen, and nitrogen in the alloy, respectively. Given the polarizability of the oxygen ion [18], they estimated the polarizability of hafnium ion by using the material data of HfO_2 [65]. They also estimated that of the nitrogen ion by using data for some metal nitrides [64]. Figure 3.7 shows dielectric constants calculated using (3.7) and estimated molar volumes for α (m-HfO_2), γ_2 ($Hf_7O_{11}N_2$), γ_3 ($Hf_7O_8N_4$), and γ_4 (Hf_2ON_2). The calculated values fit experimental data very well, especially for N content of less than 18 %. Ino et al. also extracted the band alignment of the alloys using REELS and valence spectra of XPS and concluded that this material is one of the better candidates for high-k gate dielectrics.

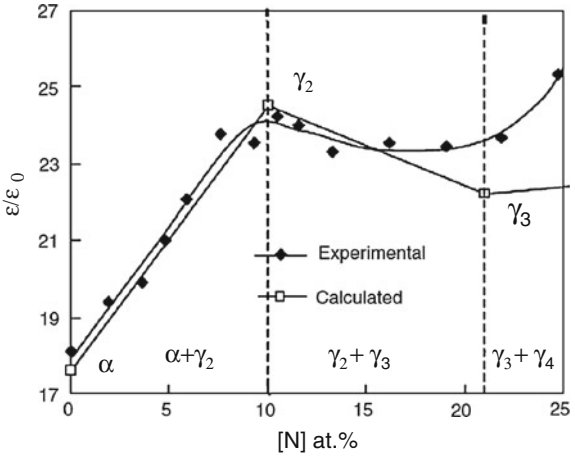


Fig. 3.7 Dielectric constant of HfON versus N content. Calculation results based on the Clausius-Mossotti theory are compared to the experimental values [64]. In this figure, α, γ_2, γ_3, and γ_4 represent m-HfO_2, $Hf_7O_{11}N_2$, $Hf_7O_8N_4$ and Hf_2ON_2, respectively

3.5 Hafnium–Silicon-Based Gate Dielectrics

HfO_2–SiO_2 alloy is formed using various techniques such as chemical vapor deposition (CVD) [66, 67], plasma-enhanced CVD [25], and atomic layer deposition (ALD) [68, 69]. For silicon deposition by the CVD method, organic precursors such as $SiH[N(C_2H_5)_2]_3$, tetraethoxy-silane (TEOS:$Si(OC_2H_5)_4$), and tetrakis-dimethylamino-silane (TDMAS:$Si[N(CH_3)_2]_4$) were normally used. By changing the ratio of precursor flow, Hf content and Si content were varied in the material. Physical vapor deposition methods are also used such as pulsed-laser deposition (PLD) [70], sputtering from composite target [71], from HfO_2 target with SiH_4 gas [72], reactive sputtering from Hf and Si targets [73], and electron beam evaporation [74].

Since Hf is an element in group IV, tetravalent coordination of Hf in the silicate is expected for material with low Hf concentration, in which the SiO_2 tetrahedron is the main component and a few Hf cations are expected to replace some Si in the network [75]. On the other hand, as mentioned above, the coordination number of Hf is 7 and that of oxygen is 3 or 4 in the monoclinic HfO_2 crystal structure. Therefore, modification of the coordination number is expected when Hf content increases in the pseudo-alloy materials [76]. Some microscopic inhomogeneity originating from the coordination variability has been observed as the broadening of Si2p and Hf4f peaks in the XPS spectra from the alloy [77].

Unfavorable Hf silicide formation observed in the case of HfO_2 is suppressed even at high temperature by the inclusion of SiO_2 [74]. In addition, the oxygen diffusion in HfSiO is retarded, leading to the suppression of interfacial layer growth in an oxygen-containing atmosphere at high temperature [78]. SiO_2 incorporation is also effective for suppressing the impurity penetration phenomena, which are problematic when poly-Si gate electrode is used. Boron diffusion through the HfSiO film with the Hf/Hf+Si ratio of 0.52 (atomic concentration of Hf is about 10–12 %) is suppressed up to 950 °C [79]. However, considering that the process temperature for the activation of impurities in the source/drain of MOSFET is as high as 1,000 °C, the capability of Hf silicate as a diffusion barrier is insufficient. Moreover, it is reported that phosphorous and arsenic diffused through the HfSiO with the same Hf concentration during 1,000 °C rapid thermal annealing (RTA) [80]. Such impurity diffusion is ascribed to the enhanced diffusion at the grain boundaries formed in the resultant ‘phase separation phenomena’ in HfSiO at high temperatures [79, 80] as described below.

When the pseudo-alloy is annealed at high temperature, it is observed that the material loses its uniformity and HfO_2 precipitates in the remaining silica-rich matrix [16, 81]. Transmission electron microscope (TEM) photographs clearly show the crystallization of HfO_2, which results in the inhomogeneity in the materials after 800 °C–1,000 °C high-temperature annealing [82]. Even in the alloys with relatively low content such as Hf/Hf+Si = 0.1, small HfO_2 crystals emerge after 1,000 °C annealing [83]. Figure 3.8 shows the planar TEM photographs of the HfSiO alloy after high-temperature annealing. Elements with larger

mass in the film have larger capability for the diffraction of incident electrons. Therefore, the dark area corresponds to those containing Hf in the film. Actually, the enlargement of the portion suggests that this area consists of orthorhombic HfO_2 crystal as shown in the inset of the figure [83].

These phase separation phenomena are ascribed either to the spinodal decomposition or to the nucleation and growth [84]. Cahn treated alloys as a supercooled liquid solution and reported that the system's total Gibbs free energy change becomes negative owing to the compositional fluctuation for a certain initial compositional region. This is where the following equation is satisfied, and the decomposition takes place spontaneously in this region [85]:

$$\frac{\partial^2 G}{\partial C^2} < 0, \tag{3.8}$$

where C represents the initial uniform composition of the alloy. This compositional region and its boundary are referred to as the spinodal region and the spinodal boundary.

Figure 3.9 shows the phase diagram of HfO_2–SiO_2 system, in which the spinodal boundary is indicated as the inner dotted line [16]. The alloys whose compositions are located inside the spinodal boundary are unstable and tend to suffer from phase separation.

The compositions of the resultant two different phases approach those indicated by the miscibility gap curve (the outer dotted line in Fig. 3.9) when the system reaches equilibrium [16].

Even alloys whose compositions are outside the spinodal boundary may also suffer from phase separation by the nucleation and growth mechanism: once nuclei

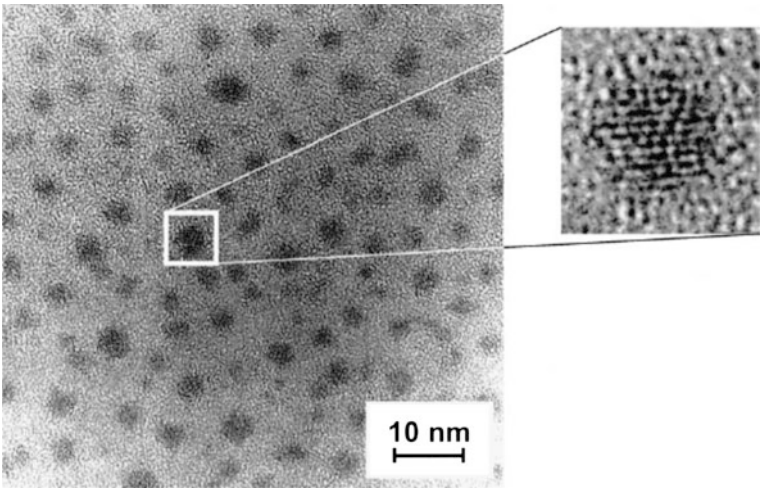


Fig. 3.8 HfO_2 precipitation after 1,000 °C annealing of the HfSiO alloy with Hf relative content (Hf/Hf+Si) of 10 % [83]

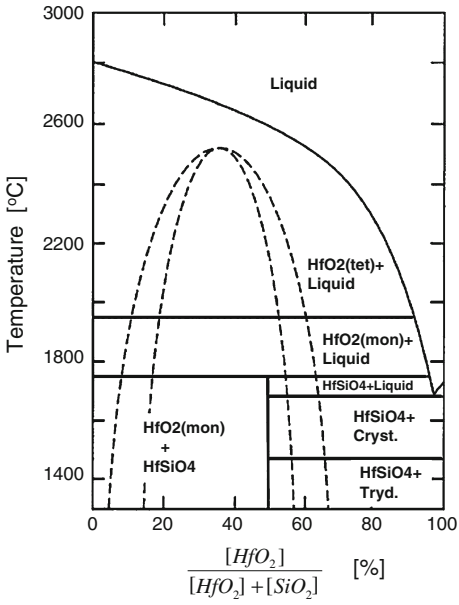


Fig. 3.9 Phase diagram of HfO_2–SiO_2 compositional systems. Inner dotted line and outer dotted line represent the spinodal boundary and the miscibility gap, respectively. (Reprinted with permission from [16]. Copyright 2002, Materials Research Society)

of a certain phase with a diameter larger than the critical size are formed, the growth of the nuclei is favorable for reducing the free energy of the system [86]. The phase separation observed for an alloy with low HfO_2 contents (Hf/Hf+Si = 0.1) as shown in Fig. 3.8 [83] may be caused by the nucleation and growth mechanism. This assumption is supported by the fact that the HfO_2 crystals precipitated in the alloy are round, which is typical for this mechanism [86]. On the other hand, HfO_2 crystal forms rather dendritic shapes in samples with larger HfO_2 content when the phase separation takes place. This phenomenon is shown in Fig. 3.11a for the material with Hf/Hf+Si = 26 % and in Ref. [87] for that with Hf/Hf+Si = 0.4. The kinetics of these phenomena is considered to be limited by the diffusion of the constituent material, i.e. Hf, Si or both in the alloy system [84].

The phase separation has several adverse effects in addition to the above-mentioned enhancement of the impurity diffusion through the alloy. One of the influences is the degradation of the dielectric constant: Ino et al. measured the dielectric constant of HfSiO alloy with HfO_2 content of 0.1 and 0.2 and concluded that the resultant non-uniformity in the film after the high-temperature annealing leads to a severe decrease in the average dielectric constant as shown in Fig. 3.10 [88]. Ono et al. [83] modeled the film structure due to the phase separation and calculated the change of the film's dielectric constant. They obtained a reasonable match to the experimental results.

It has also been reported that additional scattering of carriers moving through the channel occurs owing to the non-uniformity of the electric field in the gate insulator and this degrades the electron mobility of transistors with the alloy [89].

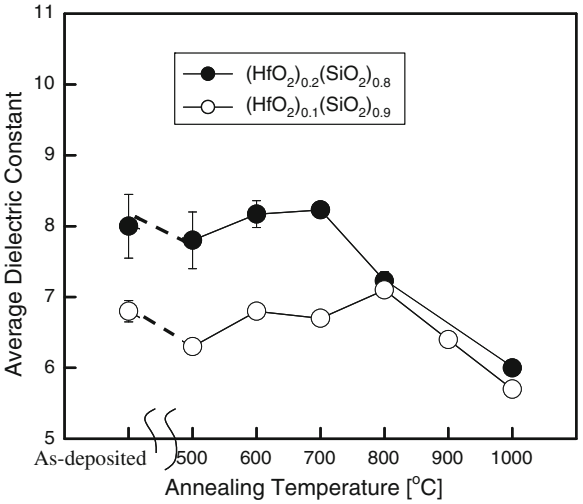


Fig. 3.10 Average dielectric constant of the film just after deposition and after the phase separation at high-temperature annealing [88]

Nitrogen is incorporated in order to suppress the phase separation in HfSiO alloy [90, 91]. Nitrogen is incorporated by the introduction of nitrogen-containing gas in reactive sputtering [90]. Oxidation of HfSiN was also tried in order to obtain the HfSiON structure [92]. There is another method in which HfSiO is firstly deposited and nitrogen is introduced by subsequent NH_3 annealing [93, 94] or nitrogen plasma treatment [94, 95]. Sekine et al. compared the applicability of the two techniques to short-channel MOSFETs and concluded that the plasma treatment is favorable for keeping high carrier mobility in the transistor. As in the case of SiON, nitrogen at the interface between the gate dielectric and the substrate silicon induces charges that scatter carriers in the channel region. Since the plasma treatment is effective for obtaining larger concentration at the surface of HfSiO while maintaining low concentration in the vicinity of the interface, this method is effective for keeping larger carrier mobility of MOSFETs [94].

The effect of the nitrogen incorporation on the phase separation phenomenon is investigated by changing the amount of nitrogen in the material. Figures 3.11 show the TEM planar views of the Hf-silicate films (a) without nitrogen, (b) with a nitrogen atomic concentration [N] of 5 at.%, and (c) 30 at.% after the 1,000 °C annealing [73]. The relative concentration of Hf to Hf+Si (Hf/Hf+Si) was about 25 % for all samples. Microcrystals with a diameter of several nanometers formed in the film without nitrogen during the annealing as shown in Fig. 3.11a. The diameter decreased in the film with the nitrogen incorporation of 5 at.% and uniform amorphous film was obtained for the sample with the 30 at.% nitrogen concentration.

The mechanism of this phase separation suppression due to the nitrogen incorporation has not been investigated in detail yet. However, considering that the phase separation speed is limited by the diffusion of the constituent material, i.e. Hf, Si or both, in the alloy system [84], it is reasonable to think that nitrogen incorporation in the amorphous network retards such diffusion in the materials [96].

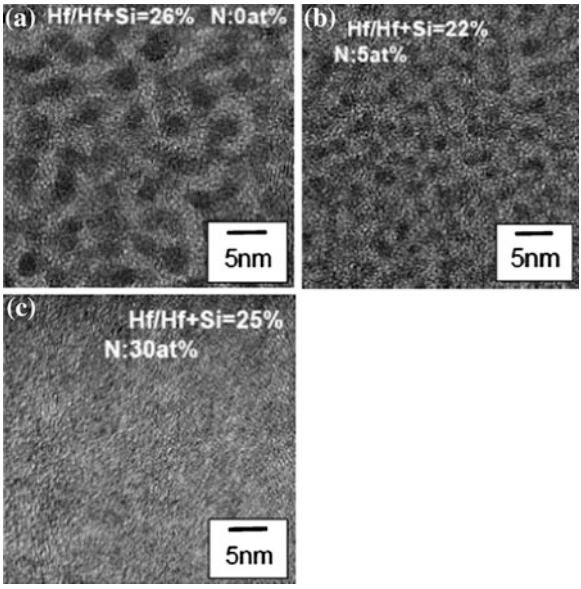


Fig. 3.11 TEM photographs of HfSiON with different nitrogen content after 1,000 °C annealing in nitrogen ambient. Hf ratio (Hf/Hf+Si) is around 22 %. Nitrogen atomic concentration is **a** 0 at.%, **b** 5 at.% and **c** 30 at.%. Annealing was performed in N_2 ambient at 1,000 °C for 30 s. [73]

The nitrogen concentration criterion for keeping the thermal stability of HfSiO is definitely dependent on the hafnium content in the film. The lower limit of nitrogen concentration that can sustain the amorphous structure, avoiding the phase separation and crystallization, was investigated extensively for the case of 1,065 °C spike annealing in nitrogen ambient [97]. The lower limit increases gradually when Hf content increases in the film, however, for films with Hf/Hf+Si larger than 80 %, the lower limit increases dramatically.

Another beneficial effect of the nitrogen incorporation is the suppression of the diffusion of the impurities such as boron, phosphorous, and arsenic through the film [73, 98, 99]. As mentioned above, Quevedo-Lopez et al. [99] ascribed high diffusivity of the impurities in HfSiO to the formation of the grain boundary as a result of the phase separation and high diffusion coefficient at the grain boundary. They claim the reason for the lower diffusion coefficient of impurities through HfSiON resides in the film uniformity as well as the fact that the film retains its amorphous structure. Koyama et al. [100] also investigated the correlation between the diffusion coefficient and the film structure, concluding that not only film uniformity but also the film microscopic structure changes the coefficient: larger nitrogen concentration and lower Hf concentration lead to the lower diffusion coefficient of boron at 1,000 °C even among amorphous films.

Nitrogen incorporation in the HfSiO alloy increases the film dielectric constant as shown in Fig. 3.12 [97]. Hf/Hf+Si ratio is taken as a parameter in this figure. The measurement is performed for films without any high-temperature annealing, meaning that there is no influence of the phase separation in this case. As expected, increase in Hf content leads to an increase in the dielectric constant from the SiO_2 value to the HfO_2 value. For any relative Hf concentration, larger nitrogen

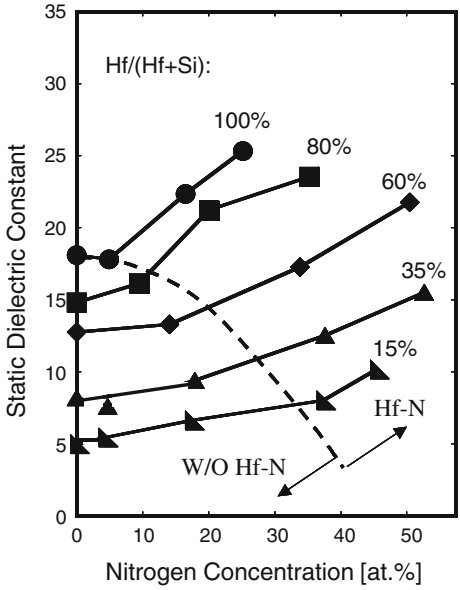


Fig. 3.12 Dielectric constant of HfSiON as a function of the nitrogen incorporation, evaluated from CV measurement for thick (about 100 nm) films with various concentrations. Hf/Hf+Si is taken as a parameter. Nitrogen atomic concentration [N] region where clear Hf–N bonds are observed in XPS is also shown [97]

incorporation generally increases the dielectric constant. Considering that Si_3N_4 has a larger k-value (about 7.8) than that of SiO_2, nitrogen incorporation is expected to enhance the dielectric constant; however, Koike et al. claimed that the dielectric constant increases drastically with Hf–N bond formation for a certain [N] level depending on Hf content in the film [97]. The dotted line in Fig. 3.12 indicates the critical [N] concentration. It is considered that HfSiON in the high dielectric constant region can be regarded as a pseudo-quaternary alloy consisting of four insulating components: SiO_2, HfO_2, Si_3N_4, and Hf_3N_4 [17]. Since Hf_3N_4 is an insulating material with a large dielectric constant as mentioned in Sect. 3.4 [60, 62], it is plausible to think that this component contributes to the abrupt increase in the dielectric constant of the alloy.

Band-gap energies of the HfSiO with various concentrations have been investigated with XPS, Ellipsometry, and REELS techniques [23, 25]. Kato et al., used XPS and ellipsometry to extract band-gap energy of HfSiO alloy and concluded that Hf incorporation into SiO_2 network decreases the band-gap energy of the materials from that of SiO_2 to the HfO_2 value as in Fig. 3.13. An interesting point is that the value shows steep decrease at low Hf content and saturates at HfO_2 value already at Hf/Hf+Si of about 0.6. This may be due to the difference in the microscopic structure between SiO_2 and HfO_2. Coordination of Hf atoms in SiO_2 crystal network may be different from that of Hf atoms in HfO_2 ionic crystal structure, leading to the different conduction band energies [76].

Nitrogen incorporation further decreases the band-gap [23, 101]. Figure 3.14 shows band-gap energy of HfSiON with Hf/Hf+Si value of 0.6 and 0.8. Although nitrogen incorporation gradually decreases the band-gap energy at low nitrogen concentration [N], it leads to abrupt decreases at critical [N]s. It is ascribed to the

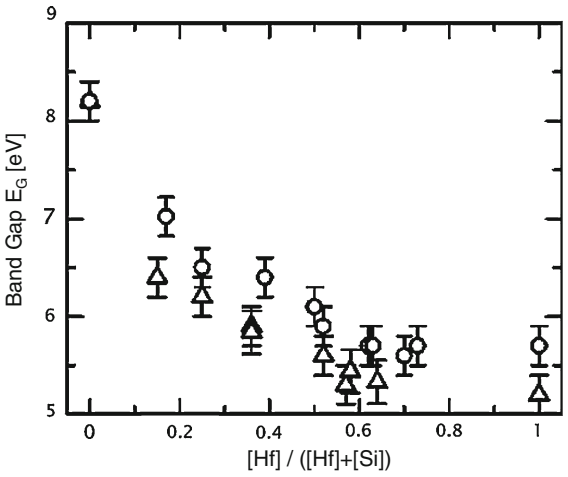


Fig. 3.13 Band-gap energy estimated from the XPS analyses as a function of the composition x(=Hf/Hf+Si). Circles and triangles represent data for HfSiO and ZrSiO, respectively. (Reprinted with permission from [25]. Copyright 2002, American Institute of Physics)

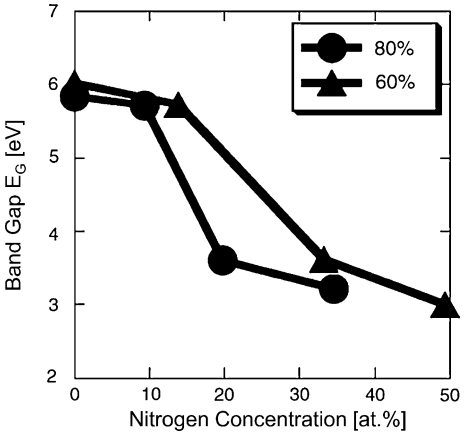


Fig. 3.14 Band-gap energy of HfSiON with Hf/Hf+Si of 0.6 and 0.8. Nitrogen incorporation gradually decreases the band-gap energy at low nitrogen concentration [N]; however, it leads to abrupt decreases at critical [N]s, depending on the Hf contents in the alloy. It is ascribed to the Hf–N bond formation [23]

Hf-N bond formation [23]. As mentioned in Sect. 3.4, the nitrogen incorporation moves the valence band edge nearer to the vacuum level by the replacement of the edge from O_2p band to N_2p band [56], leading to the valence band offset decrease.

Contrary to the prediction by the theoretical calculation [56], the experimental result shows that nitrogen incorporation in HfO_2 as well as HfSiO network also leads to the decrease in the conduction band offset [23]. It has been reported that a large amount of N in HfO_2 network could induce a large amount of oxygen vacancy, Vo, and lead to Vo–Vo interactions, resulting in the modification of the conduction band structure [102].

As mentioned in Sect. 3.2, the potential of high-k material as a gate leakage suppressor is measured conveniently with the figure of merit (F.O.M.),

$$F.O.M. \equiv k\sqrt{m^*\phi_b}, \tag{3.3}$$

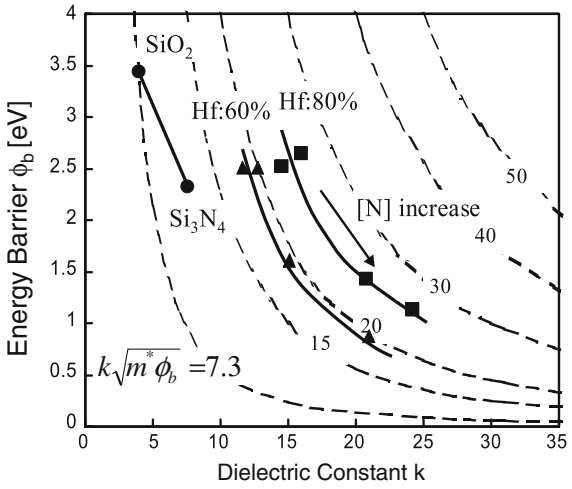


Fig. 3.15 Contour plots of the F.O.M. for electrons tunneling from Si conduction band. Values for HfSiON films with various Hf and N concentrations are plotted. The effective mass of carriers m* was set to be 1 for all concentrations for the purpose of relative comparison [103]

where k, m* and ϕ_b are the film dielectric constant, effective mass of the carrier and the barrier height, respectively. The F.O.M. for HfSiON films with various Hf and N concentrations are plotted in Fig. 3.15 [103]. Since the effective masses both for electron and hole are not clear yet for HfSiON, m* was set to be 1 for all concentrations for the purpose of relative comparison in this figure. Figure 3.15 shows that the F.O.M. becomes larger as Hf concentration increases, indicating the larger potential as a suppressor of gate leakage current. This figure also shows that, contrary to the F.O.M. enhancement due to the Hf relative concentration increase, [N] hardly changes the F.O.M. value. This is because the dielectric constant increase leads to the band-gap and barrier height decrease in HfSiON [23]. A similar figure concerning the hole tunneling is shown in Ref. [103]. Considering that larger [N] leads to higher thermal stability, HfSiON with high Hf and N concentrations seem to be more desirable. It should be mentioned, however that the discussion above is valid only when modulation of the Hf and N concentration does not change the effective mass of the carriers much. As mentioned in Sect. 3.2, Hf incorporation decreases the effective mass, while nitrogen incorporation into SiO_2 network increases it [104, 105]. The effective mass should be evaluated for HfSiON before concluding the F.O.M. dependence on the concentration of each constituent material.

3.6 Hafnium–Aluminum-Based Gate Dielectrics

HfO_2–Al_2O_3 alloy is formed using chemical vapor deposition (CVD) [106], atomic layer deposition (ALD) [107], pulsed laser deposition (PLD) [108], jet vapor deposition (JVD) [109], and sputtering from composite target [110]. Normally, trimethyl-aluminum (TMA: $Al(CH)_3$) is used for the Al_2O_3 deposition in the CVD/

ALD process, whereas organic materials such as tetrakis-dimethylamino-hafnium (TDMAH:($Hf[N(CH_3)_2]_4$) or tetrakis-diethylamino-hafnium TDEAH ($Hf[N(C_2H_5)_2]_4$) are used for HfO_2 deposition. Therefore, a certain amount of carbon is expected to be present in the film. The intermittent high-temperature annealing between ALD stages [111] or intermittent NH_3/Ar plasma treatment [112] is effective for reducing the carbon content in the material and this leads to the decrease in the leakage current through the film [112].

As in the case of Hf silicate, Hf silicide formation that leads to the increase in leakage current through the film is suppressed even at high temperature with Al_2O_3 incorporation [108, 113], except for the case in which oxygen deficiency occurs in the deposition process [108]. Oxygen diffusion in HfO_2 is also retarded by the presence of Al_2O_3 component, leading to the suppression of interfacial layer growth in an oxygen-containing atmosphere at high temperature [113]. This phenomenon can be attributed to the very small diffusivity of oxygen in Al_2O_3 compared to that in HfO_2 [114].

However, impurity penetration phenomena such as boron diffusion, which is problematic for poly/high-k gate stack, are hardly suppressed in HfAlO system [115] because of the limited impurity diffusion barrier characteristics of Al_2O_3 [116]. Moreover, phase separation between HfO_2 and Al_2O_3 accompanied by HfO_2 crystallization occurs as in the case of HfSiO, when HfO_2 content is large relative to Al_2O_3 [111] and this partial crystallization may cause the inhomogeneity and the enlargement of the leakage current through HfAlO after the high-temperature annealing [117].

As expected from the dielectric constant values for HfO_2 and Al_2O_3, the dielectric constants of HfAlO are intermediate between the two. They have been investigated thoroughly by the conventional C–V measurement with a number of samples with various Hf content and Al content as shown in Fig. 3.16 [118]. It should be noted that the measurements were performed with HfAlO layer deposited by metal organic chemical vapor deposition (MOCVD) in which Hf and Al precursors were introduced in the chamber at the same time with various flow ratios. The deposition process was performed at 500 °C, followed by post-metallization annealing at 400 °C. Therefore, homogeneous alloy structure was expected without the phase separation in this case. The film dielectric constant decreases from the value of HfO_2 to that of Al_2O_3 almost linearly as Al atomic concentration increases. Note that Al atomic concentration in pure Al_2O_3 is 40 at.%.

The band-gap energy also takes intermediate values between those of HfO_2 and Al_2O_3 [32, 107]. Ohta et al. [119] used XPS measurement in order to extract the band-gap energy and the band alignment to Si using HfAlO with the cation ratio Hf/Hf+Al of 0.3. This film was deposited by ALD using $HfCl_4$, $Al(CH_3)_3$, and H_2O at 300 °C. Although the film was subjected to 1,050 °C RTA, it was confirmed that this film was not crystallized through the annealing [119]. The extracted band-gap was as large as 6.5 eV, which is an intermediate value between 5.8 eV (HfO_2) and 8.8 eV (Al_2O_3). From the valence band spectra of XPS, it was confirmed that the valence band and conduction band offsets of this HfAlO are as

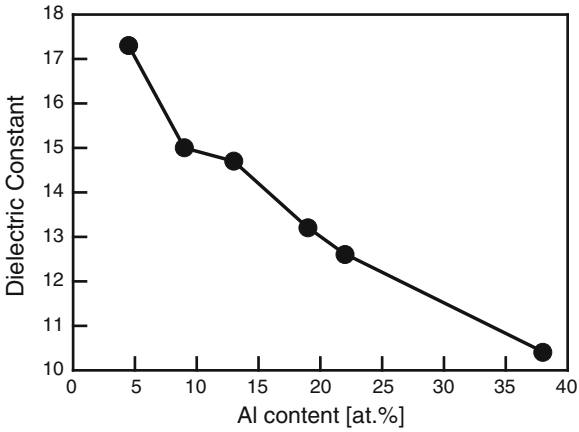


Fig. 3.16 Dielectric constant modification by the increase of Al content in HfAlO alloy [133]

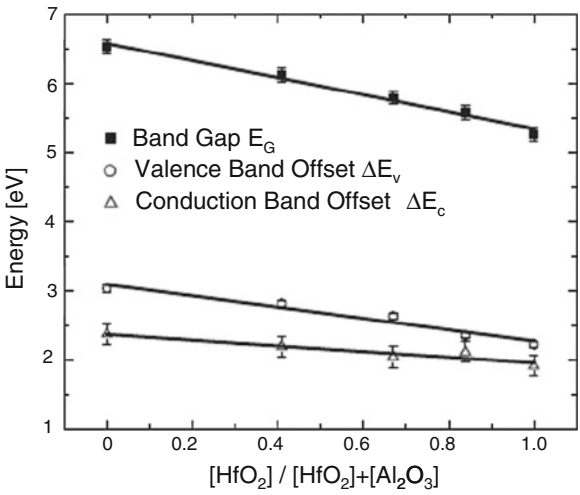


Fig. 3.17 Dependence of band-gap E_G, the valence band offset ΔE_v, and the conduction band offset ΔE_c for $(HfO_2)_x(Al_2O_3)_{1-x}$. The E_G and ΔE_v data were obtained by XPS measurements. The ΔE_c data are calculated by the $\Delta E_c = E_G - \Delta E_v - E_{G,\,Si}$. The solid lines are linear least square fits of the data points. (Reprinted with permission from [32]. Copyright 2002, American Institute of Physics)

large as 1.78 and 3.6 eV, respectively [119]. Similar result is reported from another group as shown in Fig. 3.17 [32]. These values are large enough to suppress the leakage current through the high-k film.

Nitrogen is added to the material in order to suppress the thermal stability issues such as impurity penetration as well as phase separation and crystallization. Nitrogen incorporation is realized by co-sputtering of HfO_2 and Al_2O_3 in $Ar/O_2/N_2$ ambient [115], or NH_3 annealing during or after the HfAlO deposition process [111]. Nitrogen incorporation is effective for suppressing the boron diffusion [115]. It is also effective for enhancing the crystallization temperature [111]. This practice even enhances the dielectric constant of the material [120] as in the case of HfSiO, which is beneficial for the realization of smaller EOT gate stack.

3.7 Doped Hafnium-Based Gate Dielectrics

For the further increase in figure of merit (F.O.M.) of the gate insulator, high-k material with a dielectric constant of more than 25 has been searched for, recently. As La has very large ionic polarizability among cation candidates [18], La-based high-k material exemplified by La_2O_3 has been investigated intensively. Since this topic is discussed in detail in Chap. 9, research for obtaining a large dielectric constant of more than 25 using Hf-based material is discussed in this section.

A reasonable way of enlarging the dielectric constant is to mix HfO_2 with other MOx that has larger dielectric constant. Since TiO_2 has a very large dielectric constant of more than 50 [26], this material may be a good candidate for the mixed oxide. HfTiAlO was deposited by co-sputtering for that purpose and it realized a high dielectric constant of 36 [121]. The drawbacks of TiO_2 are its small band-gap energy (about 3 eV) and almost negligible barrier height from Si conduction band edge [122]. However, this HfTiAlO showed a barrier height of 1.3 eV from the Si conduction band edge, which is not much smaller than that of HfO_2. Mixing TiO_2 with HfO_2 may change the coordination of Ti atoms inside the alloy, resulting in the realization of a larger barrier height.

Since theoretical study predicted that the cubic HfO_2 and the tetragonal HfO_2 could have large dielectric constants of 29 and 70 [123] respectively, much work has been done to transform the material to such higher dielectric constant phases. It is well known that the monoclinic phase is the most stable one below 1,750 °C [10, 124]. Therefore, the research interest resides in how to stabilize the cubic and tetragonal phases at temperatures as low as 1,000 °C. Small amount of yttrium doping into HfO_2 network (about 5 at.%) enhanced the dielectric constant value to 27, as a result of the formation of cubic HfO_2 phase [125]. The process temperature for the transformation was as low as 800 °C in this case. The transformation was confirmed by the X-ray diffraction (XRD) spectrum, which indicated the molar volume reduction of the material from that of the monoclinic phase [125]. The doping species is not limited to yttrium but other rare earth materials such as La [126], Gd [127], Er [127], Dy [127, 128], Sc [128] or Ce [129] show a similar effect. Figure 3.18 shows the dielectric constant variation of Gd-, Er-, and Dy-doped HfO_2 as a function of the doping concentration [127]. In the case of Gd, the larger doping with 20 at.% shows the largest dielectric constant, 27, whereas in the case of Er and Dy, doping level of 10 at.% gives the largest value, 28, as a result of the tetragonal HfO_2 formation. It has also been reported that Dy doping transforms HfO_2 to its cubic phase and enhances the dielectric constant to 32 [128]. It should be noted that these enhancements cannot be ascribed to the rare earth oxide formation in the materials, because those oxides have smaller dielectric constant than monoclinic HfO_2. Actually, larger doping of the rare earth elements resulted in the decrease of the dielectric constant of the materials, as shown in Fig. 3.18. It has also been reported that even a small amount of Si doping (about Si/Hf+Si ratio of 0.05) could lead to the tetragonal HfO_2 formation, which enhances the dielectric constant to 27 [130].

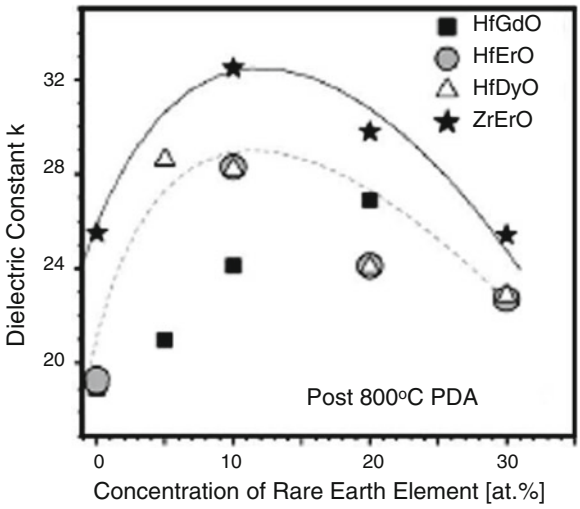


Fig. 3.18 Dielectric constant variation of Gd-, Er-, and Dy-doped HfO_2 as a function of the doping concentration. Data for Er-doped ZrO_2 is also shown. (Reprinted with permission from [127]. Copyright 2007, American Institute of Physics)

The doping was realized by using the rare earth oxide or SiO_2 target simultaneously when HfO_2 is sputter-deposited. Doping is also performed by adding Dy or Sc containing precursors in HfO_2 MOCVD. Although the doping effect on the HfO_2 phase transformation is not clear yet, some authors claim that the larger ionic radii of rare earth elements raise internal compressive stress in the HfO_2 network, leading to the realization of the smaller molar volume (tetragonal or cubic) phases [127]. Theoretical study using first-principles calculations of doped HfO_2 indicated that the doping modifies the energy of the monoclinic and the tetragonal phases and concluded that Si and Ge are favorable elements for the modification. However, it could not elucidate any clear dependence of the energy modification on the ionic radii of the dopants [131].

There is another method in which the stress is induced not by the doping but by the annealing with a cap layer. Migita et al. indicated that the low-temperature post-deposition annealing (PDA) of HfO_2 with the cap gate electrode (TiN or TaN) enhances the dielectric constant up to about 50 [132]. Figure 3.19 shows the XRD spectra and indicates that the crystal phase after PDA with the cap electrode is that of cubic HfO_2, whereas the monoclinic phase is obtained when PDA is performed without the cap. The advantage of this method is that there is no need to use elements other than Hf and oxygen. This could circumvent the problems associated with the use of other elements in the LSI process such as cross-contamination.

These larger dielectric constant phases have band-gap energy similar to that of the monoclinic HfO_2 [128, 132]. Even with the doping of rare earth elements, the band-gap energy does not decrease [24, 128]. Therefore, the gate stacks with these materials show drastic reduction of leakage current by about three orders of magnitude, compared with monoclinic HfO_2 with similar EOT [127, 128].

A practical issue concerning the usage of tetragonal and cubic phase HfO_2 materials as gate insulators is that subjecting them to the higher annealing

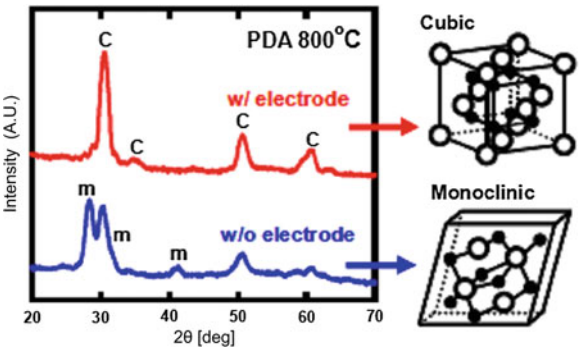


Fig. 3.19 XRD spectra of HfO_2 films after PDA at 800 °C with and without electrode. 'C' represent peaks from cubic HfO_2, whereas 'm' represents those of monoclinic phase. The cubic HfO_2 is successfully formed by PDA only with the electrode cap. (©2008 IEEE, Reprinted, with permission, from [132])

temperature typical of LSI processes diminishes their dielectric constant. For all the cases mentioned above, low-temperature annealing (typically at 800 °C) was performed for the realization of the large dielectric constant phases. However, it was reported that annealing at higher temperatures such as 1,000 °C drastically reduced the dielectric constant of tetragonal HfO_2 from 50 to 25 [132]. Yttrium-doped HfO_2 also showed the decrease after 1,000 °C annealing [125]. This degradation phenomenon seems reasonable because the stable phase of HfO_2 is monoclinic below 1,750 °C [124]. This tetragonal or cubic HfO_2 is thought to be unstable at 1,000 °C. Therefore, higher-temperature annealing and longer annealing yield the stable monoclinic phase, leading to the reduction of the dielectric constant. Some measures for keeping the high dielectric constant phase even at 1,000 °C annealing should be developed for the application of this measure to advanced LSIs. The subject of phase modifiers in promoting the higher-k phases of the high-k oxides has been analyzed in detail in Chap. 10.

3.8 Summary

HfO_2 is regarded as one of the most promising materials for the high-k gate dielectrics. This is mainly because this material possesses a large dielectric constant originating from a large polarizability of the Hf–O molecule, and large band-gap energy. These two characteristics lead to the dramatic reduction of the leakage current through the thin film of this material, compared to the conventional SiO_2 gate dielectrics. Thermal stability is also required for the high-k materials in order to maintain the integrity of the MOSFET through the LSI processes. HfO_2 satisfies this requirement, however severer process condition may cause some problems even with this material: anomalous silicide formation in the film may enhance the leakage current, enhanced interfacial layer growth may increase the equivalent oxide thickness, and dopant-impurity penetration through the film may cause threshold voltage fluctuation. Suppression of those issues is the incentive for the incorporation of other components such as SiO_2 and Al_2O_3 into HfO_2.

The pseudo-alloy shows good characteristics to overcome most of those issues. Nitrogen incorporation is very effective for keeping homogeneous structure avoiding the phase separation of the pseudo-alloy. As HfO_2 can have several phases, transformation of the monoclinic phase to the tetragonal and the cubic phases with larger dielectric constant has been pursued recently in order to further enhance the figure of merit of HfO_2 as the gate dielectric.

References

1. S.H. Lee, H.F. Luan, W.P. Bai, C.H. Lee, T.S. Jeon, Y. Senaki, D. Roberts, D.L. Kwong, IEDM Tech. Dig., p. 31, San Francisco, (December 2000)
2. L. Kang, K. Onishi, Y. Jeon, B.H. Lee, C. Kang, W.-J. Qi, R. Nieh, S. Gopalan, R. Choi, J.C. Lee, IEDM Tech. Dig., p.35, San Francisco, (December 2000)
3. G.D. Wilk, R.M. Wallace, J.M. Anthony, J. Appl. Phys. **89**(10), 5243 (2001)
4. M.T. Bohr, R.S. Chau, T. Ghani, K. Mistry, IEEE Spectrum, p. 23 (October 2007)
5. J.G. Simmons, J. Appl. Phys. **34**(6), 1793 (1963)
6. Y.-C. Yeo, T.-J. King, C. Hu, IEEE Trans. Elect. Devices **50**(4), 1027 (2003)
7. International Technology Roadmap for Semiconductors (2012), Edition http://www.itrs.net/Links/2012ITRS/Home2012.htm
8. C.L. Hinkle, C. Fulton, R.J. Nemanich, G. Lucovsky, Microelectron. Eng. **72**, 257 (2004)
9. W.J. Zhu, T.-P. Ma, T. Tamagawa, J. Kim, Y. Di, IEEE Electron. Device Lett. **23**(2), 97 (2002)
10. G. He, Q. Fang, M. Liu, L.Q. Zhu, L.D. Zhang, J. Cryst. Growth **268**, 155 (2004)
11. H.B. Park, M. Cho, J. Park, S.W. Lee, C.S. Hwang, J.-P. Kim, J.-H. Lee, N.-I. Lee, H.-K. Kang, J.-C. Lee, S.-J. Oh, J. Appl. Phys. **94**(5), 3641 (2003)
12. G.D. Wilk, D.A. Muller, Appl. Phys. Lett. **83**(19), 3984 (2003)
13. R. Chen, H. Kim, P.C. McIntyre, S. Bent, Appl. Phys. Lett. **84**(20), 4017 (2004)
14. M.S. Akbar, L.C. Lee, N. Moumen, J. Peterson, Appl. Phys. Lett. **88**(8), 082901 (2006)
15. A.S. Foster, F. Lopez Gejo, A.L. Shluger, R.M. Nieminen, Phys. Rev. B **65**, 174117 (2002)
16. J.-P. Maria, D. Wickaksana, J. Parrette, A.I. Kingon, J. Mater. Res. **17**(7), 1571 (2002)
17. M. Koike, T. Ino, Y. Kamimuta, M. Koyama, Y. Kamata, M. Suzuki, Y. Mitani, A. Nishiyama, Phys. Rev. B Condens. Matter. Mater. Phys **73**(12), 125123 (2006)
18. R.D. Shannon, J. Appl. Phys. **73**(1), 348 (1993). (In this article, values for ZrO_2 and SiO_2 are mentioned. We assume the values for HfO_2 are similar to those for ZrO_2)
19. E.J. Little, M.M. Jones, J. Chem. Educ. **37**, 231 (1960)
20. J.M. Ziman, *Principles of the Theory of Solids*, 2nd edn. (Cambridge University Press, Cambridge, 1972)
21. S. Miyazaki, M. Narasaki, M. Ogasawara, M. Hirose, Microelectron. Eng. **59**(1–4), 373 (2001)
22. H. Takeuchi, D. Ha, T.-J. King, J. Vac. Sci. Technol., A **22**(4), 1337 (2004)
23. Y. Kamimuta, M. Koike, T. Ino, M. Suzuki, M. Koyama, Y. Tsunashima, A. Nishiyama, Jpn. J. Appl. Phys. **1**, Regul. Pap. **44**(3), 1301 (2005)
24. S.-G. Lim, S. Kriventsov, T.N. Jackson, J.H. Haeni, D.G. Schlom, A.M. Balbashov, R. Uecker, P. Reiche, J.L. Freeouf, G. Lucovsky, J. Appl. Phys. **91**(7), 4500 (2002)
25. H. Kato, T. Nango, T. Miyagawa, T. Katagiri, K.S. Seol, Y. Ohki, J. Appl. Phys. **92**(2), 1106 (2002)
26. P.W. Peacock, J. Robertson, J. Appl. Phys. **92**(8), 4712 (2002)
27. E.H. Rhoderick, R.H. Williams, *Metal Semiconductor Contacts* (Oxford University Press, Oxford, 1988)
28. W. Mönch, Phys. Rev. Lett. **58**(12), 1260 (1987)
29. S. Miyazaki, J. Vac. Sci. Technol., B **19**(6), 2212 (2001)

30. V.V. Afanas'ev, M. Houssa, A. Stesmans, M.M. Heyns, Appl. Phys. Lett. **78**(20), 3073 (2001)
31. S. Sayan, E. Garfunkel, S. Suzer, Appl. Phys. Lett. **80**(12), 2135 (2002)
32. H.Y. Yu, M.F. Li, B.J. Cho, C.C. Yeo, M.S. Joo, D.L. Kwang, J.S. Pan, C.H. Ang, J.Z. Zheng, S. Ramanathan, Appl. Phys. Lett. **81**(2), 376 (2002)
33. K.J. Hubbard, D.G. Schlom, J. Mater. Res. **11**(11), 2757 (1996)
34. T.S. Jeon, J.M. White, D.L. Kwong, Appl. Phys. Lett. **78**(3), 368 (2001)
35. H. Watanabe, N. Ikarashi, Appl. Phys. Lett. **80**(4), 559 (2002)
36. K. Muraoka, Appl. Phys. Lett. **80**(24), 4516 (2002)
37. M. Koyama, K. Suguro, M. Yoshiki, Y. Kamimuta, M. Koike, M. Ohse, C. Hongo, A. Nishiyama, IEDM Tech. Dig., p. 459, Washington, D.C., (December 2001)
38. Y.-S. Lin, R. Puthenkovilakam, J.P. Chang, Appl. Phys. Lett. **81**(11), 2041 (2002)
39. M.-H. Cho, Y.S. Roh, C.N. Whang, K. Jeong, S.W. Nahm, D.-H. Ko, J.H. Lee, N.I. Lee, K. Fujihara, Appl. Phys. Lett. **81**(3), 472 (2002)
40. N. Miyata, M. Ichikawa, T. Nabatame, T. Horikawa, A. Toriumi, Jpn. J. Appl. Phys. **42**(Part2(2B)), L138 (2003)
41. S. Sayan, E. Garfunkel, T. Nishimura, W.H. Schulte, T. Gustafsson, G.D. Wilk, J. Appl. Phys. **94**(2), 928 (2003)
42. C.S. Kang, H.-J. Cho, K. Ohnishi, R. Nieh, R. Choi, S. Gopalan, S. Krishnan, J.H. Han, L.C. Lee, Appl. Phys. Lett. **81**(4), 2593 (2002)
43. H.-J. Cho, C.S. Kang, K. Onishi, S. Gopalan, R. Nieh, R. Choi, E. Dharmarajan, J.C. Lee, IEDM Tech. Dig. **655**, (December 2001)
44. K.P. Bastos, J. Morais, L. Miotti, R.P. Pezzi, G.V. Soares, I.J.R. Baumvol, R.I. Hedge, H.H. Tseng, P.T. Tobin, Appl. Phys. Lett. **81**(9), 1669 (2002)
45. M. Zhao, K. Nakajima, M. Suzuki, K. Kimura, M. Uematsu, K. Torii, S. Kamiyama, Y. Nara, H. Watanabe, K. Shiraishi, T. Chikyow, K. Yamada, Appl. Phys. Lett. **90**, 133510 (2007)
46. B.W. Busch, W.H. Schulte, E. Garfunkel, T. Gustafsson, W. Qi, R. Nieh, J. Lee, Phys. Rev. **B62**(20), R13290 (2000)
47. H. Watanabe, Appl. Phys. Lett. **78**(24), 3803 (2001)
48. K. Onishi, L. Kang, R. Choi, E. Dharmarajan, S. Gopalan, Y.J. Jeon, C.S. Kang, B.H. Lee, R. Nieh, J.C. Lee, in *Symposium on VLSI Technology,* Digest of Technical Papers, p.131, Kyoto, June 2001
49. C. Lee, J. Choi, M. Cho, J. Park, C.S. Hwang, H.J. Kim, J. Jeong, W. Lee, Appl. Phys. Lett. **83**(7), 1403 (2003)
50. K. Suzuki, H. Tashiro, Y. Morisaki, Y. Sugita, IEEE Trans. Electron Devices **50**(6), 1550 (2003)
51. C.-L. Liu, Z.X. Jiang, R.I. Hedge, D.D. Sieloff, R.S. Rai, D.C. Gilmer, C.C. Hobbs, S. Lu, Appl. Phys. Lett. **81**(8), 1441 (2002)
52. D. Gui, J. Kang, H. Yu, H.F. Lim, Appl. Surf. Sci. **231–232**, 590 (2004)
53. C.H. Choi, T.S. Jeon, R. Clark, D.L. Kwong, IEEE Electron Device Lett. **24**(4), 215 (2003)
54. T. Nabatame, K. Iwamoto, K. Yamamoto, K. Tominaga, H. Hisamitsu, M. Ohni, K. Akiyama, M. Ikeda, T. Nishimura, H. Ota, T. Horikawa, A. Toriumi, J. Vac. Sci. Technol. B **22**(4), 2128 (2004)
55. K. Muraoka, K. Kurihara, N. Yasuda, H. Satake, J. Appl. Phys. **94**(3), 2038 (2003)
56. G. Shang, P.W. Peacock, J. Robertson, Appl. Phys. Lett. **84**(1), 106 (2004)
57. J. Lindström, L.I. Johansson, P.E.S. Persson, A. Callenås, Phys. Rev. B **39**(6), 3599 (1989)
58. A.F. Guillermet, J. Haeglund, G. Grimvall, Phys. Rev. B **48**(16), 11673 (1993)
59. A. Zerr, G. Miehe, R. Riedel, Nat. Mater. **2**(3), 185 (2003)
60. J.S. Becker, E. Kim, R.G. Gordon, Chem. Mater. **16**, 3497 (2004)
61. P. Kroll, Phys. Rev. Lett. **90**(12), 125501 (2003)
62. P. Kroll, J. Phys.: Condens. Matter **16**, S1235 (2004)
63. S.J. Clarke, C.W. Michie, M.J. Rosseinsky, J. Solid State Chem. **146**, 399 (1999)
64. T. Ino, Y. Kamimuta, M. Suzuki, M. Koyama, A. Nishiyama, Jpn. J. Appl. Phys. **45**(6B), 2908 (2006)
65. R.E. Hann, P.R. Suitch, J.L. Pentecost, J. Am. Ceram. Soc. **68**, C285 (1985)

66. B.C. Hendrix, A.S. Borovik, C. Xu, J.F. Roeder, T.H. Baum, M.J. Bevan, M.R. Visokay, J.J. Chambers, A.L.P. Rotondaro, H. Bu, L. Colombo, Appl. Phys. Lett. **80**(13), 2362 (2002)
67. Y. Ohshita, A. Ogura, M. Ishikawa, T. Kada, H. Machida, Jpn. J. Appl. Phys. **42**, Part2(6A), L578 (2003)
68. Y. Xuan, D. Hojo, T. Yasuda, Appl. Phys. Lett. **84**(25), 5097 (2004)
69. Y. Senzaki, S. Park, H. Chatham, L. Bartholomew, W. Nieveen, J. Vac. Sci. Technol., A **22**(4), 1175 (2004)
70. M. Filipescu, N. Scarisoreanu, V. Craciun, B. Mitu, A. Purice, A. Moldovan, V. Iona, O. Toma, M. Dinescu, Appl. Surf. Sci. **253**(19), 8184 (2007)
71. G.D. Wilk, R.M. Wallace, Appl. Phys. Lett. **74**(19), 2854 (1999)
72. A. Callegari, E. Cartier, M. Gribelyuk, H.F. Okum-schmidt, T. Zabel, J. Appl. Phys. **90**(12), 6466 (2001)
73. M. Koyama, A. Kaneko, T. Ino, M. Koike, Y. Kamata, R. Iijima, Y. Kamimuta, A. Takashima, M. Suzuki, C. Hongo, S. Inumiya, M. Takayanagi, A. Nishiyama, IEDM Tech. Dig., p. 849, San Francisco, (December 2002)
74. G.D. Wilk, R.M. Wallace, J.M. Anthony, J. Appl. Phys. **87**(1), 484 (2000)
75. G. Lucovsky, G.B. Rayner Jr, Appl. Phys. Lett. **77**(18), 2912 (2000)
76. P. Broqvist, A. Pasquarello, Microelectron. Eng. **84**, 2416 (2007)
77. G.B. Rayner Jr, D. Kang, G. Lucovsky, J. Vac. Sci. Technol. **B21**(4), 1783 (2003)
78. L.V. Goncharova, A.M. Dalponte, D.G. Starodub, T. Gustafsson, E. Garfunkel, P.S. Lysaght, B. Foran, J. Barnett, G. Bersuker, Appl. Phys. Lett. **89**, 044108 (2006)
79. M.A. Quevedo-Lopez, M. El-Bouanani, M.J. Kim, B.E. Gnade, R.M. Wallace, M.R. Visokay, A. LiFatou, M.J. Bevan, L. Colombo, Appl. Phys. Lett. **81**(6), 1074 (2002)
80. M.A. Quevedo-Lopez, M. El-Bouanani, M.J. Kim, B.E. Gnade, R.M. Wallace, M.R. Visokay, A. LiFatou, M.J. Bevan, L. Colombo, Appl. Phys. Lett. **81**(9), 1609 (2002)
81. A.I. Kingon, J.-P. Maria, S.K. Streiffer. Nature **406**(6799), 1032 (2000)
82. S. Stemmer, Z. Chen, C.G. Levi, P.S. Lysaght, B. Foran, J.A. Gisby, and J.R. Taylor, Jpn. J. Appl. Phys. Part 1, **42**(6A), 3593 (2003)
83. M. Ono, T. Ino, M. Koyama, A. Takashima, A. Nishiyama, Solid-State Electron **48**(12), 2191 (2004)
84. H. Kim, P. McIntyre, J. Appl. Phys. **92**(9), 5094 (2002)
85. J.W. Cahn, Acta Metall. **9**, 795 (1961)
86. J.W. Cahn, R.J. Charles, Phys. Chem. Glasses **6**(5), 181 (1965)
87. S. Stemmer, Y. Li, B. Foran, P.S. Lysaght, S.K. Streiffer, P. Fuoss, S. Seifert, Appl. Phys. Lett. **83**(15), 3141 (2003)
88. T. Ino, M. Koyama, M. Ono, A. Takashima, Y. Kamimuta, M. Suzuki, A. Nishiyama, in *Proceedings of the 2nd ECS International Semiconductor Technology Conference 2002*, Tokyo, Japan
89. T. Yamaguchi, R. Iijima, T. Ino, A. Nishiyama, H. Satake, N. Fukushima, IEDM Tech. Dig., p. 621 San Francisco, (December 2002)
90. M.R. Visokay, J.J. Chambers, A.L.P. Rotondaro, A. Shanware, L. Colombo, Appl. Phys. Lett. **80**(17), 3183 (2002)
91. A.L.P. Rotondaro, M.R. Visokay, J.J. Chambers, A. Shanware, R. Khamankar, H. Bu, R.T. Laaksonen, L Tsung, M. Douglas, R. Kuan, M.J. Bevan, T. Grider, J. McPherson and L. Colombo, in *Symposium on VLSI Technology*, Digest of Technical Papers, Honolulu, June 2002
92. G. Pant, P. Punchaipetch, M.J. Kim, R.M. Wallace, B.E. Gnade, Thin Solid Films **460**, 242 (2004)
93. M.S. Akbar, S. Gopalan, H.-J. Cho, K. Ohnishi, R. Choi, R. Nieh, C.S. Kang, Y.H. Kim, J. Han, S. Krishnan, J.C. Lee, Appl. Phys. Lett. **82**(11), 1757 (2003)
94. K. Sekine, S. Inumiya, M. Sato, A. Kaneko, K. Eguchi, and Y. Tsunashima, IEDM Tech. Dig. p. 102, Washington, D.C., (December 2003)

95. S. Inumiya, K. Sekine, S. Niwa, A. Kaneko, M. Sato, T. Watanabe, H. Fukui, Y. Kamata, M. Koyama, A. Nishiyama, M. Takayanagi, K. Eguchi, Y. Tsunashima, in *Symposium on VLSI Technology Digital Technical Papers*, p. 17, Kyoto, June 2003
96. S. Sayan, N.V. Nguyen, J. Ehrstein, J.J. Chambers, M.R. Visokay, M.A. Quevedo-Lopez, L. Colombo, D. Yoder, I. Levin, D.A. Fischer, M. Paunescu, O. Celik, E. Garfunkel, Appl. Phys. Lett. **87**(21), 212905 (2005)
97. M. Koike, T. Ino, Y. Kamimuta, M. Koyama, Y. Kamata, M. Suzuki, Y. Mitani, A. Nishiyama, Y. Tsunashima, IEDM Tech. Dig. p. 107 Washington, D.C., (December 2003)
98. M.A. Quevedo-Lopez, M. El-Bouanani, M.J. Kim, B.E. Gnade, R.M. Wallace, M.R. Visokay, A. LiFatou, J.J. Chambers, L. Colombo, Appl. Phys. Lett. **82**(26), 4669 (2003)
99. M.A. Quevedo-Lopez, M.R. Visokay, J.J. Chambers, M.J. Bevan, A. LiFatou, L. Colombo, M.J. Kim, B.E. Gnade, R.M. Wallace, J. Appl. Phys. **97**, 043508 (2005)
100. M. Koyama, T. Ino, Y. Kamimuta, M. Suzuki, C. Hongo, A. Nishiyama, Integration of advanced micro- and nanoelectronic devices-critical issues and solutions . Mater. Res. Soc. **811**, 247 (2004)
101. N. Ikarashi, M. Miyamura, K. Masuzaki, T. Tatsumi, Appl. Phys. Lett. **84**(18), 3672 (2004)
102. T. Schimizu, M. Koyama, Appl. Surf. Sci. **254**, 6109 (2008)
103. A. Nishiyama, M. Koike, M. Suzuki, Y. Kamimuta, M. Koyama, ECS transactions, 1(5, Physics and Technology of High-k Gate Dielectrics III), 541 (2006)
104. C.Y. Ng, T.P. Chen, Y. Liu, C.Q. Sun, S. Fung, J. Appl. Phys. **96**, 5912 (2004)
105. J. Suné, M. Lanzoni, P. Olivo, IEEE Trans. Electron Devices **40**, 1017 (1993)
106. R.S. Johnson, J.G. Hong, C. Hinkle, G. Lucovsky, J. Vac. Sci. Technol. **B20**(3), 1126 (2002)
107. H.Y. Yu, M.F. Li, D.-L. Kwong, Thin Solid Films **462–463**, 110 (2004)
108. P.F. Lee, J.Y. Dai, K.H. Wong, H.L.W. Chan, C.L. Choy, J. Appl. Phys. **93**(6), 3665 (2003)
109. W.J. Zhu, T. Tamagawa, M. Gibson, T. Furukawa, T.P. Ma, IEEE Electron Device Lett. **23**(11), 649 (2002)
110. P. Sivasubramani, J. Kim, M.J. Kim, B.E. Gnade, R.M. Wallace, J. Appl. Phys. **101**, 114108 (2007)
111. A. Toriumi, K. Iwamoto, H. Ota, M. Kadoshima, W. Mizubayashi, T. Nabatame, A. Ogawa, K. Tominaga, T. Horikawa, H. Satake, Microelectron. Eng. **80**(6), 190 (2005)
112. T. Kawahara, K. Torii, R. Mitsuhashi, A. Muto, A. Horiuchi, H. Ito, H. Kitajima, Jpn. J. Appl. Phys. **43**(7A), 4129 (2004)
113. H.Y. Yu, N. Wu, M.F. Li, C. Zhu, B.J. Cho, D.-L. Kwong, C.H. Tung, J.S. Pan, J.W. Chai, W.D. Wang, D.Z. Chi, C.H. Ang, J.Z. Zheng, S. Ramanathan, Appl. Phys. Lett. **81**(19), 3618 (2002)
114. M.L. Green, E.P. Gusev, R. Dedraeve, E.L. Garfunkel, J. Appl. Phys. **90**(5), 2057 (2001)
115. M. Koyama, Y. Kamimuta, M. Koike, M. Suzuki, A. Nishiyama, Jpn. J. Appl. Phys. **43**(4B), 1788 (2004)
116. D.-G. Park, H.J. Cho, I.S. Yeo, J.S. Roh, J.M. Hwang, Appl. Phys. Lett. **77**(14), 2207 (2000)
117. X. Blasco, J. Pétry, M. Nafría, X. Aymerich, O. Richard, W. Vandervorst, Microelectron. Eng. **72**, 191 (2004)
118. O. Buiu, Y. Lu, S. Hall, I.Z. Mitrovic, R.J. Potter, P.R. Chalker, Thin Solid Films **515**, 3772 (2007)
119. A. Ohta, H. Nakagawa, H. Murakami, S. Higashi, T. Kawahara, K. Torii, S. Miyazaki, Jpn. J. Appl. Phys. **1**, Regul. Pap. **43**(11B), 7831 (2004)
120. M. Koyama, Y. Kamimuta, M. Koike, T. Ino, A. Nishiyama, Jpn. J. Appl. Phys. **44**(4B), 2311, (2005)
121. N. Lu, H.-J. Li, J.J. Peterson, D.-L. Kwong, Appl. Phys. Lett. **90**, 082911 (2007)
122. J. Robertson, J. Vac. Sci. Technol. **B18**(3), 1785 (2000)
123. X. Zhao, D. Vanderbilt, Phys. Rev. B **65**, 233106 (2002)
124. D.W. Stacy, D.R. Wilder, J. Am. Ceram. Soc. **58**, 285 (1975)
125. K. Kita, K. Kyuno, A. Toriumi, Appl. Phys. Lett. **86**, 102906 (2005)
126. Y. Yamamoto, K. Kita, K. Kyuno, A. Toriumi, Appl. Phys. Lett. **89**, 032903 (2006)

127. S. Govindarajan, T.S. Böscke, P. Sivasubramani, P.D. Kirsch, B.H. Lee, H–.H. Tseng, R. Jammy, U. Schröder, S. Ramanathan, B.E. Gnade, Appl. Phys. Lett. **91**, 062906 (2007)
128. C. Adelmann, V. Sriramkumar, S. Van Elshocht, P. Lehnen, T. Conard, S. De Gendt, Appl. Phys. Lett. **91**, 162902 (2007)
129. P.R. Chalker, M. Werner, S. Romani, R.J. Potter, K. Black, H.C. Aspinall, A.C. Jones, C.Z. Zhao, S. Taylor, and P.N. Heys, Appl. Phys. Lett. **93**, 182911 (2008)
130. K. Tomida, K. Kita, A. Toriumi, Appl. Phys. Lett. **89**, 142902 (2006)
131. D. Fischer, A. Kersch, Appl. Phys. Lett. **92**, 012908 (2008)
132. S. Migita, Y. Watanabe, H. Ohta, H. Ito, Y. Kamimuta, T. Nabatame, A. Toriumi, in *Symposium on VLSI Technology*, Digest of Technical Papers, p.152, Honolulu, June 2008
133. This figure was formed using data in Ref. [78]

Chapter 4
Hf-Based High-k Gate Dielectric Processing

Masaaki Niwa

Abstract This chapter focuses on the processing of Hf-based high-k gate dielectric film and its device fabrication to improve its electrical properties. First, the formation process of Hf-based high-k dielectric thin film is introduced followed by detail study of these films from the materials science point of view, such as its crystallization and its control, carrier trapping and doping effect on the bulk high-k film. Finally, the device processing of the Field Effect Transistor including the CMOSFET with Hf-based high-k gate dielectric and metal gate electrode is discussed.

4.1 Introductory Remarks

In accordance with the requirement of CMOS (Complimentary Metal Oxide Semiconductor) scaling, replacement of thermally grown SiO_2 gate dielectric with other dielectric materials of high dielectric constant (k) has become imperative. However, replacing of the SiO_2 conduces to a variety of other disadvantages and is not as simple as it may seem. Many challenges caused by different dielectric properties are to be conquered before the high-k dielectrics find practical use in industry. In spite of such difficult situations, thanks to intensive efforts, these insuperable difficulties have been solved gradually and finally CMOS with high-k/metal gate stack for high performance application has been in commercial production in 2007 for the first time [1]. However, still many issues were left unsolved and have been addressed continuously to overcome them. Most of them come from the fact that Hf-based high-k dielectric materials are thermodynamically unstable and the nature of the electrovalent (ionic)-bonded HfO_2 is quite

M. Niwa (✉)
Center for Innovative Integrated Electronic Systems, Tohoku University,
Sendai 980-0077, Japan
e-mail: niwa.masaaki@cies.tohoku.ac.jp

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5_4,

different from that of the covalently-bonded SiO_2. Although oxygen vacancy is a common problem for the oxides, behavior of each system is completely different. For SiO_2, the remaining unpaired electrons in Si tend to form Si–Si bond, whereas Hf is produced by releasing two electrons into the valence band of the HfO_2 and the amount of oxygen vacancy is rather large compared to SiO_2. Key issues to improve the high-k gate stack characteristics are (1) dielectric material has high permittivity, (2) gate electrode is metal to obtain the desired work function without a depletion layer, (3) process control is easy for interfaces between high-k gate dielectric, Si substrate, and gate electrode, and (4) their module/integration process is not complicated and cost effective, and (5) the gate stack has high reliability.

In this chapter, issues of material and processing of Hf-based high-k gate dielectrics are described. First, formation process and material properties of Hf-based high-k gate dielectric are discussed. Then its influence on electrical characteristics including reliability based on the intrinsic solid state properties are taken into consideration. Several intrinsic key issues such as interfacial control, optimization of the Hf-based high-k gate dielectric formation, breakdown issues, work function control, bulk trapping phenomena as well as process integration are discussed from the point of materials science and device process technology. Issues of metal gate electrode and gate stack reliability as well as its process integration including the comparison of gate first and gate last processes are relegated to other chapters and discussed in detail.

4.2 Hf-Based High-k Gate Dielectric Formation Process

Among the current high-k gate dielectrics, the Hf-based material is the most promising, in particular HfO_2 and its silicate ($HfSiO_x$) have been studied intensively for practical use. As a result HfO_2 has been introduced into industry for high performance application in 2007 [1]. Generally, the requirements for the gate dielectric are: (1) High k value (but $k < 50$ because of short channel effect of FET—Field Effect Transistor); (2) High barrier property with respect to N-, PMOS gate electrodes; (3) Robust against high thermal budget; (4) Good process controllability; and (5) Capability of thin film thickness controllability. In the following sections, typical formation methods are addressed.

4.2.1 Metal Organic Chemical Vapor Deposition

Chemical vapor deposition (CVD) is a chemical process used to produce high quality thin films, in which process the Si substrate is exposed to volatile precursors, which react and/or decompose on the substrate surface to obtain the desired composition. During the reaction, volatile byproducts are often produced, which are removed by gas flow in the reaction chamber. Among the variety of

CVD depositions, metal-organic chemical vapor deposition (MOCVD) based on metal-organic precursors is the principal method of depositing high-k gate dielectric layer. To form Hf-based high-k dielectric films, it is necessary to use metal-halide or metal-organic as a basic ingredient. Different from the Atomic Layer Deposition (ALD) method with a single wafer treatment, MOCVD has the advantage of its high throughput because of its batch processing.

HfSiON deposition is taken up as an example. For the high-k part, Si_2H_6 and HBT ($Hf(OtBu)_4$)) are used to form $HfSiO_x$. HBT has high vapor pressure among liquid Hf precursors and is generally moderate in price compared with other precursors. NH_3 is often used to form HfSiON, which leads to suppression of large gate leakage by keeping up the amorphous state in the $HfSiO_x$ film during high temperature annealing. The effect of nitrogen doping is considered in detail in Sect. 4.3. In addition, an interfacial oxide layer (SiO_2) which is necessary to obtain high quality interface between high-k and Si substrate is formed by conventional oxidation such as Rapid Thermal Oxidation (RTO), pyrogenic oxidation or chemical oxidation.

The HfSiON thickness is selected depending on the device application. Concentration of Hf, [Hf] and Si, [Si] in $HfSiO_x$ can be controlled by the polymerization reaction by desorption of C_4H_{10} from Si_2H_6 and HBT; and the intermediates such as $Si_2H_3[O–Hf(OC_4H_9)_3]_3$ ([Hf] = 60 %) and $Si_2H_2[OHf(OC_4H_9)_3]_4$ ([Hf] = 67 %) are decomposed on the Si substrate. However, it is interesting to note that the Hf concentration in the HfSiOx remained constant ([Hf] = approx. 62 %) regardless of the deposition condition. It may be considered that the decomposed intermediates contribute to form $HfSiO_x$ film under stable composition, i.e., [Hf] = approx. 60 %.

In order to reduce contaminants such as H and C in the high-k film, which are introduced from the basic ingredients, O_3 is often introduced to cause outward diffusion by forming H_2O and CO or CO_2. Figure 4.1 shows carbon and hydrogen concentration in the $HfSiO_x$ film observed by SIMS (Secondary Ion Mass Spectroscopy) indicating that these impurities have been effectively reduced by O_3 treatment (250 °C for 3 min.) [2]. In order to improve the film quality, MOCVD is often combined with other techniques such as Remote-Plasma Oxidation.

HfO_2 films were fabricated by MOCVD through an alternating supply process, which consists of deposition using tetrakis-diethylamino-hafnium (TDEAH_ $Hf(N(C_2H_5)_2)_4$), followed by oxidation using remote-plasma oxygen (RPO) [3]. Deposition rates are found to depend on temperature (above 350 °C) and the supply duration of TDEAH, which indicates a non-self-limiting growth mode. From X-ray photoelectron spectroscopy (XPS) analysis, it is found that as-deposited films formed only by the TDEAH are unstable due to presence of carbon and nitrogen impurities in the film. These impurities can be removed by the RPO treatment. When the supply duration of RPO is increased, the high-k film tends to be stoichiometric HfO_2 with low gate leakage current; however, excess RPO supply results in an increase of the equivalent oxide thickness (EOT) due to a decrease in the permittivity of the interfacial layer.

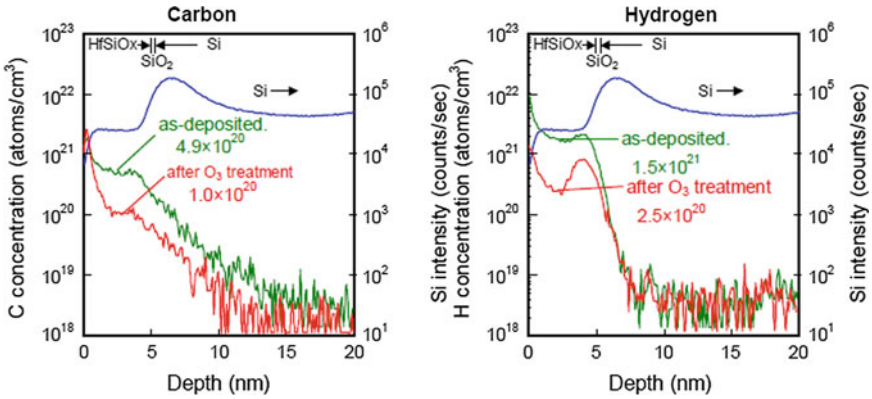


Fig. 4.1 Carbon and hydrogen concentrations in the $HfSiO_x$ film [2]

This implies that optimization of the supply durations of TDEAH and RPO is necessary. By this optimization, a minimal EOT of 1.6 nm and a leakage current of 1×10^{-4} A/cm^2 at -1 V relative to the flat-band voltage are obtained without any problem.

Effect of the RPO treatment is also confirmed by other metal-organic (MO) precursors. In case HfO_2 film is formed from $Hf(OC(CH_3)_2CH_2OCH_3)_4$, deposition rate of HfO_2 becomes higher when oxygen gas is supplied together with the precursor, however, the deposited film contains large amount of H_2O due to the oxidation of Hf precursor. Excess oxygen process tends to degrade the HfO_2 film properties. As a countermeasure, in situ RPO is effective in reducing the contaminants in the HfO_2. As a result, leakage current of HfO_2/Si capacitors with TiN gate electrode formed without excess oxygen flow and subsequent in situ RPO can be effectively reduced [4].

In order to obtain higher deposition rate with component flexibility than the conventional ALD, precise atomic growth at high deposition rates of advanced MOCVD is developed by means of accurate injection of small quantities of liquid precursors into a heated vaporizer volume using its short pulses for the delivery of the liquid [5]. This enables a precise control of the nucleation layer growth and in situ grading of composition during film formation.

4.2.2 Atomic Layer Deposition

Atomic layer Deposition (ALD) was, at the beginning, invented as Atomic Layer Epitaxy (ALE) which is a specialized form of epitaxy by depositing alternating monolayers of two elements onto a substrate and makes use of the fact that the occurring chemisorption will continue until all sites are occupied due to self-limiting surface reaction. ALD should be taken up as a special CVD, i.e. Atomic

Layer CVD as it is similar in chemistry to CVD, except that more than one ingredient gas does not exist in the reacting chamber in the ALD reaction. In ALD reaction, two complementary precursors (e.g. $Al(CH_3)_3$ and H_2O) are pulsed alternatively and then chemisorbed onto the surface of the substrate to form monolayer. More precisely, one of the precursors adsorbs onto the substrate surface until it saturates the surface sites and extra precursor is carried away by an inert gas such as nitrogen, purging the chamber, and further growth cannot occur until the second precursor is introduced. And this cycle is repeated alternatively until the film thickness reaches its desired value. Although the film thickness is controlled by deposition time for the conventional CVD processes, it is controlled by the number of precursor cycles for the ALD process rather than the deposition time. Thus, the ALD film growth is due to self-limited surface chemical reactions and in theory it allows pin-hole free film with excellent conformity and uniformity even against bumpy surfaces.

For the HfO_2 deposition, at least two types of precursors are used in the ALD process at around 300 °C. The typical precursors are $HfCl_4$ and H_2O. The key issue is to reduce the Cl contamination in the formed HfO_2 to obtain high quality film. The steady growth-per-cycle, defined as amount of HfO_2 deposited in one reaction cycle, is 1.3×10^{14} Hf/cm^2/cycle. As starting surface, a chemical oxide is usually introduced to allow good nucleation of the ingredient of precursors, resulting in continuous HfO_2 film growth. In order to obtain ultra-thin high-k gate stack with EOT (Equivalent Oxide Thickness) < 1 nm with minimum gate leakage as well as high channel mobility, HfAlON film was investigated by precise layer-by-layer deposition and annealing where annealing was carried out for each ALD layer [6].

4.2.3 Precursors

Precursors used for MOCVD and ALD are described in this section. Most precursors used in MOCVD and ALD are in the liquid form (liquid or solid dissolved in a convenient solvent). Liquid precursor is advantageous to limit particle formation during the deposition. Liquid solutions are injected in vaporization chamber towards injectors and vapor of the precursor is transported to the substrate. This technique is commonly used for solid precursors as well as liquid precursors and high growth rates can be achieved using this technique. Hafnium silicate layer, for example, is often deposited using the liquid precursors such as tetrakis-diethylamino-Hafnium (TDEAH) and tetrakis-dimethylamino-Silicon (TDMAS). Typical deposition temperature is around 600 °C. By changing the pressure and the gas flow rate, the produced composition can be controlled from Hafnium-rich (e.g. [Hf] = 65 %) to Silicon-rich (e.g. [Hf] = 23 %) phase. Typical precursors are listed in the Table 4.1. Order of the vapor pressure is TDEAH > HTTB > TDMAH > TEMAZ > TEMAH for all temperature ranges.

In addition to the above precursors, solid phase precursor in which the solid is easily sublimated has also been investigated. Attempt to introduce $Hf(NO_3)_4$ as

Table 4.1 Typical Hf-based precursors

Hf series chemical formula	Name of the precursor	Acronym name	State at room temperature
$Hf[N(CH_3)_2]_4$	Tetrakis-Di-Methyl-Amino-Hafnium	TDMAH	Liquid
$Hf[N(C_2H_5)_2]_4$	Tetrakis-Di-Ethyl-Amino-Hafnium	TDEAH	Liquid
$Hf[N(C_2H_5)\ (CH_3)]_4$	Tetrakis-Eethyl-Methyl-Amino-Hafnium	TEMAH	Liquid
$Hf(OtBu)_4$)	Hafnium-Tetra-Tertiary-Butoxy	HTTB HBT	Liquid
$Hf[OC(CH_3)_2CH_2OCH_3)]_4$	Tetrakis(1Methoxy-2Methyl-2Propoxy) Hafnium		Liquid
HfAlO	Tri-Methyl-Alminum+TDMAH	TMA+TDMAH	Liquid
$Hf(NO_3)_4$	Hafnium-Nitrate		Solid
$HfCl_4$	Hafnium-Tetra-Chloride		Solid

oxidizing agent as well as C free metal source was reported [7]. After RCA (Radio Corporation of America) cleaning followed by diluted HF cleaning, ultra-thin Si_3N_4 layer (approx. 0.7 nm) was formed as an interfacial layer by RTN (Rapid Thermal Nitridation). The RTN treatment was performed in 500 Torr $-NH_3$ at 700 °C. Then HfO_2 film was deposited on the RTN base layer. HfO_2 film formation process was as follows. By introducing Ar carrier gas, HfO_2 film was formed by sublimated Hf molecular gas which was generated from Hf-nitrato, i.e., $Hf(NO_3)_4$ solid phase precursor under the appropriate temperature (25–70 °C). Then the Hf gas was mixed with oxygen gas in the chamber to generate HfO_2 followed by the PDA (Post Deposition Anneal) at 800 °C for 30 s in N_2. This CVD-HfO_2 is called as "nitrato". During this process, substrate temperature was kept in the temperature range from 170 to 350 °C. The reason of using the $Hf(NO_3)_4$, rather than separate oxidizing agent such as H_2O, is to minimize the oxidation of the interfacial layer between high-k dielectric and Si substrate. Also, in contrast to other widely used precursors, $Hf(NO_3)_4$ has high reactivity with other materials, which initiates uniform deposition on H-terminated Si. However, because of its high reactivity, the interfacial layer tends to be relatively thick. Thick interfacial layer can avoid mobility degradation by remote coulomb scattering; in contrast, it leads to an increase of EOT. Furthermore, vaporization from solid phase precursor is generally unstable and thus the resulting deposition rate is unstable. Comparison with respect to different precursors is discussed in Sect. 4.3.2.

4.2.4 Physical Vapor Deposition

Different from MOCVD and ALD, physical vapor deposition (PVD) provides different advantages for thin film deposition. This deposition is purely physical deposition process rather than chemical reaction at surface. In the PVD, a wide

variety of vacuum deposition exists such as vacuum evaporation, ion plating, molecular beam epitaxy, sputtering, and laser ablation. Among them, sputtering is the most promising method from the point of commercial viability for high-k dielectric and metal gate deposition, where ions are generated by discharge and directed toward a target material and then these ions sputter atoms from the target material. The generated sputtered atoms are neutral and transported to substrate surface in vacuum and finally condense on the substrate, forming a thin film.

Sputtering is usually a simple system; therefore it provides low cost, contamination free deposition and capability of deposition of high-melting point metal as well as high adhesion to the substrate. Usually, the deposition rate is low; however, this is advantageous for ultra-thin film formation such as capping layer deposition (see Sect. 4.4) for the metal gate/high-k dielectric system. On the other hand, in contrast to ALD, it is difficult to achieve conformal coating on the bumpy structure because of the translatory movement of the sputtered atoms in the vacuum and also care should be taken for the damage generated by collision on the substrate surface. Nowadays, the PVD tool is capable of damage-free deposition and precisely controlled film composition through ultrahigh vacuum co-sputtering with excellent uniformity across the 300 mm wafer ($1\sigma < 1$ %).

Comparison of the film formation method is important because often material density and its microstructure influence more seriously than the expected chemical composition. An interfacial layer growth by unintentional oxidation during high-k deposition is a matter of concern. CVD and ALD methods suffer from contaminants such as carbon, hydrogen and chlorine that need to be removed by high temperature deposition or post deposition annealing (PDA), which brings excess oxygen into high-k dielectric as well as increase of interfacial layer. In addition, the PDA has a role to densify the high-k film. This additional thermal treatment often leads to undesirable phenomena such as composition change from as deposited composition, EOT increase as well as work function degradation as described in Sect. 4.4.6. Crystallization of high-k dielectric material due to thermal treatment is also an issue as described in Sect. 4.3.2. For this, nitrogen is often introduced to avoid crystallization of Hf-based film. Nitridation after high-k deposition is conventionally used for all types of deposition method. However, by means of PVD, the nitridation can easily be performed during the deposition. The deposited PVD-HfSiON gate dielectric with metal gate, for example, shows lower gate leakage because of its amorphous phase even after higher temperature (>1,000 °C) annealing [8].

Similar to the discussion of precursor in CVD and ALD, purity of the target material is important in PVD as well. Impurity level in gate dielectric required for Front End of Line (FEOL) is less than 1×10^{10} atoms/cm^2. This indicates that the desired impurity level for the target material of PVD depends on film thickness. For high-k gate dielectric formation, there seems to be no appreciable problem from contamination point of view for Hf target because HfO_2 thickness is thin enough (approx. 1.3 nm) compared to that of metal gate electrode and purity of Hf (6N) is higher than that of the target used for metal gate (5N).

For conventional gate first process, purity of the ultra-thin capping layer on high-k dielectric for work function tuning to be discussed in Sect. 4.4.6 is also important. Generally, Al_2O_3 and La_2O_3 are intensively investigated as capping material for P- and NMOS respectively. Al or AlO target is used for Al_2O_3 capping for PMOS since its thickness is thin enough and its purity is sufficient. In contrast, La or LaO target is used for La_2O_3 capping for NMOS; however, it has low purity (e.g. 2N5 i.e. 99.5 %) in origin and has problematic deliquescence nature. The purity of La target material has been improved recently to 4N level. By improving the purity, its deliquescence could also be suppressed recently.

Regarding the metal gate electrode, the purity of the metal target is normally 5N. In case of Ti target, when Fe content is 9 wt ppm (Fe in the target material should be suppressed to less than 9 wt ppm), Fe concentration in the target turns out to be 4.4×10^{10} atoms/cm^2 for 1 nm-thick Ti. When 20 nm-thick TiN (Ti/N = 1/1) is deposited as gate electrode, contamination concentration comes to 4.4×10^{11} atoms/cm^2 which is not acceptable in the clean room. In this respect, by means of PVD, low purity (4–5N) target such as Ti is difficult to be introduced for fabrication. In contrast, Mo and/or Ta target whose purity is better than Ti, can be applicable.

In addition to pure metal and metal oxide target, alloy target has been intensively developed. In case of TiSiN, it can stay amorphous after annealing and in contrast, TiN tends to conduce to crystalline phase. Owing to its amorphous state, the obtained work function can be uniform across the wafer, however, in case of crystalline materials, distribution of work function has some variation depending on each individual grain orientation.

From purity point of view, however, it is very difficult to obtain high purity alloy target, such as TiSi and/or TaC used for metal gate electrodes because high purity alloying of Si and/or C is difficult to achieve. Similarly, from alloy target, it is difficult to obtain high purity alloy insulators such as HfLaO, AlLaO and HfTiO for future gate dielectrics. Especially, it is extremely difficult to apply them as thick insulator for flash application. For this reason, co-sputtering method is considered to be useful and practical method.

For commercial production with wafer of large diameter, PVD for gate dielectric deposition has challenge to reduce plasma damage and improve the uniformity and reducing of particles across the wafer. On the other hand, MOCVD or ALD are easily used for mass-production, however, they suffer from hydrocarbon or halogen contaminants. In order to avoid such disadvantages, post deposition annealing is necessary to suppress plasma damage or impurity incorporation.

4.2.5 Interfacial Oxide Layer

In the research of high-k gate stack, control of the interfacial reaction with the silicon substrate is one of the key issues to maintain the property of the high-k gate dielectric with suppressed interfacial defects, in particular, how we build-in an interfacial oxide. In this respect, an interfacial layer (IL) between Si substrate and high-k gate

dielectric plays an important role and usually silicon oxide layer is introduced to obtain a high quality SiO_2/Si interface. Generally, conventional high quality ultra-thin oxidation techniques are used such as RCA-type SC-1 and/or SC-2 clean, thermally grown dry oxide by RTO (Rapid Thermal Oxidation), thermally grown oxide by ISSG (In-Situ Steam Generation), ozonated wafer dispense, ozone chemical oxidation followed by HF treatment. In regard to ultra-thin RTO and ISSG oxides, the ISSG film is confirmed to have lower defect concentration at interface than in the RTO film [9]. By means of leakage-free charge pumping, trap behavior with respect to high-k and interfacial layer has been possible to be observed [10].

Effective k value is obtained from the observed EOTs by changing the physical thickness (T_{phys}) of the high-k layer as shown in Fig. 4.2. Corresponding EOT is obtained by changing the T_{phys} of high-k on the interfacial oxide layer which has constant thickness (see Fig. 4.2a). The EOT changes linearly with the high-k thickness as shown in Fig. 4.2b and thus the corresponding quantitative equation is obtained as indicated in (4.1). The first term in this equation is known value and the k value of the high-k film is obtained from the slope in the EOT–T_{phys} relation. From the EOT at $T_{phys} = 0$ nm (intercept of y axis), k value of the interfacial layer is obtained experimentally. In this case, k value of the high-k layer and the interfacial oxide layer underneath turned out to be approx. 12 and 8, respectively.

$$\mathrm{EOT} = \underbrace{\frac{k(SiO_2)}{k(IL)} \cdot T_{phys}(IL)}_{\text{Constant}} + \frac{k(SiO_2)}{k(\text{high}-k)} \cdot T_{phys}(\text{high}-k) \tag{4.1}$$

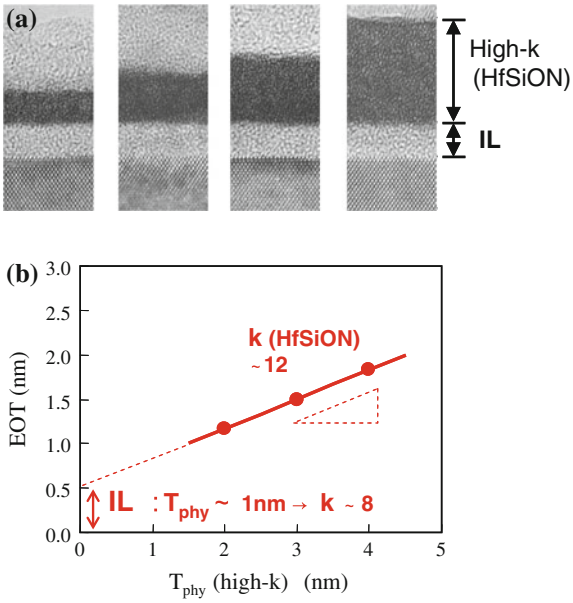


Fig. 4.2 Relation between EOT of the gate stack and the physical thickness of the high-k film. **a** Cross-sectional TEM (XTEM) of the high-k/IL. **b** Linear relation of EOT-physical thickness of the high-k film

Table 4.2 Typical high-k dielectric material properties

Material	k_d	Heat of formation (Kcal/mol)	Melting point (°C)	Structure
SiO_2	3.9	217	1,730	Amorphous
Si_3N_4	7.8		~1,900	Amorphous
Al_2O_3	8–10	267	2,050	Amorphous, α phase
ZrO_2	12–25	262	2,677	Monoclinic → Tetraganal(1,127 °C) Tetraganal → Cubic(2,377 °C) Amorphous up to 800 °C
Ta_2O_5	25–65	90	1,870	β phase → α phase (1,320 °C)
HfO_2	~30	266	2,900	Monoclinic → Tetraganal(1,700 °C) Tetraganal → Cubic (2,700 °C) Amorphous up to 700 °C
TiO_2	30–90	226	1,855	β phase → α phase (1,320 °C)
$BaSrTiO_3$	100–300			Monoclinic → Tetraganal (1,700 °C) Tetraganal → Cubic (2,700 °C)

4.3 Hf-Based Gate Dielectric Material and Its Intrinsic Phenomena

In addition to the deposition methods, many attempts to improve Hf-based high-k dielectric properties have been studied involving related process such as robustness against high thermal budget, permittivity, structure and chemistry with metal gate electrode etc. In this section, Hf-based gate dielectric material and its intrinsic phenomena are considered.

4.3.1 Hf-Based Dielectric Material

In Table 4.2, physical properties of typical high-k materials are listed. $BaSrTiO_3$ has large permittivity; however, it has huge hysteresis due to its ferroelectricity. This table indicates HfO_2 as well as Al_2O_3 and ZrO_2 are easier to be formed under exothermic reaction with relatively higher melting point compared to other high-k materials. And HfO_2 and ZrO_2 have relatively higher dielectric constant (k_d) among them and higher melting point than Al_2O_3. It should be noted that the lattice constant values of these materials are similar to Si lattice constant.

Also, as is shown in Fig. 4.3, band gap (E_g) varies inversely proportional to dielectric constant (k_d) and these three materials reveal higher k_d with larger E_g, especially HfO_2 and ZrO_2. In this respect, Hf-based high-k materials including Hf-Aluminate and Hf-Zirconate as well as Hf-silicate have been intensively investigated. Incorporation of doped oxide into HfO_2 is discussed in Sect. 4.3.5.

The EOT dependence of the gate leakage current density is indicated in Fig. 4.4. As is shown in this figure, HfO_2 and ZrO_2 dielectrics reveal lower leakage

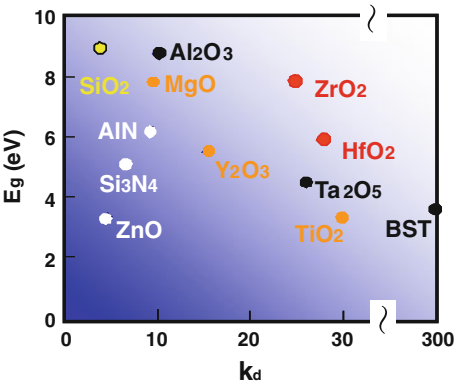


Fig. 4.3 Relation between band gap (E_g) varies inversely proportional to dielectric constant (k_d)

by six orders of magnitude compared with SiO_2 [11]. Generally, a conventional high-k gate stack has a bi-layer system with high-k dielectric and interfacial SiO_2 to obtain the high quality of interfacial properties. In this case, interpretation of the electrical characteristic is rather complicated, which is discussed in Sect. 4.4.1.

4.3.2 Crystallization and Related Issues

4.3.2.1 Crystallization

An important issue to note is the crystallization of HfO_2 which induces large gate leakage current due to the leakage path along with grain boundary when it is crystallized after the thermal anneal. The ZrO_2–SiO_2 phase diagram [12] shown in Fig. 4.5 is well known and characterized in detail compared to the case for HfO_2.

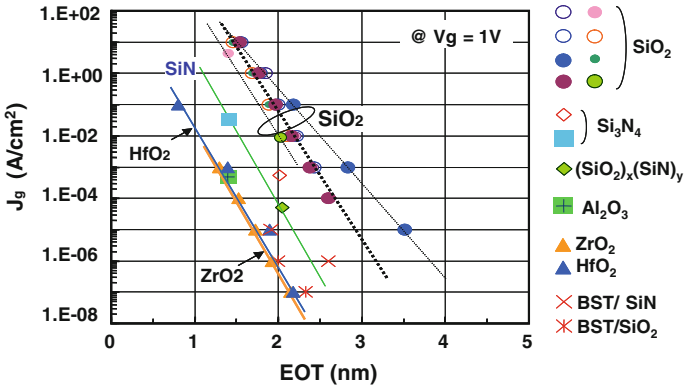


Fig. 4.4 EOT dependence of the gate leakage for various gate dielectrics

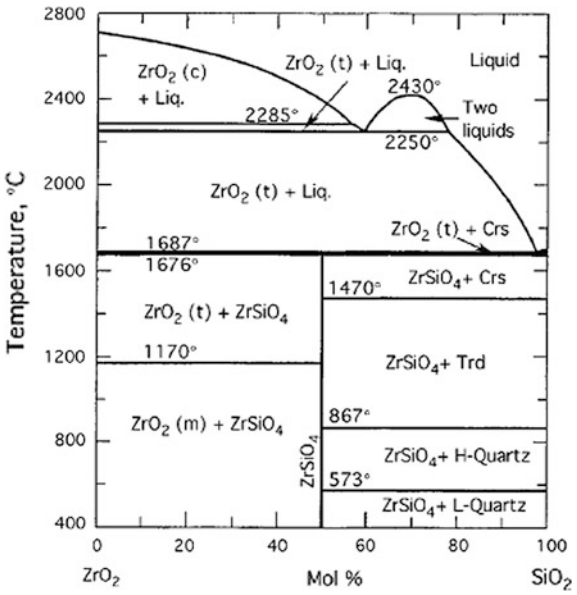


Fig. 4.5 Phase diagram of SiO_2-ZrO_2 system [12]

This is because very few reports are available on the phase diagram of HfO_2–SiO_2 at high temperature region. However, the ZrO_2–SiO_2 phase diagram can be used as a reliable phase diagram for HfO_2–SiO_2 system since Zr and Hf have very similar chemical behavior. It should be noted that the behavior indicated in the phase diagram is correct only in the closed chemical system and additional reaction have to be taken into consideration for the practical case which is an open chemical system, such as oxide transformation as well as competitive reaction between oxidation and silicidation. Using the metastable phase diagrams, the microstructure evolution during annealing of amorphous HfO_2–SiO_2 thin films for gate dielectric applications is studied and this evolution during annealing of amorphous ZrO_2–SiO_2 and/or HfO_2–SiO_2 thin films can be characterized by a low solid solubility, a liquid miscibility gap, and a kinetic barrier to the formation of the complex, crystalline silicate [13].

Nitrogen is often introduced to keep the amorphous state by preventing the crystallization. To avoid large gate leakage, nitridation is important to keep up the amorphous state in the $HfSiO_x$ film against high temperature annealing. In case of poly Si electrode, this treatment is also effective for preventing boron in the electrode from diffusing into the Si substrate. Introduction of nitrogen is often accomplished by forming Hf-silicate. By optimizing the nitrogen content ([N]) in Hf-silicate, crystallization can be suppressed effectively.

Figure 4.6 indicates the cross-sectional TEM (XTEM) image of poly Si/HfSiON/interfacial SiO_2/Si after activation annealing at 1,000 °C with respect to different [N]. At [N] = 5 at %, crystalline region is easily observed in the HfSiON which cannot be seen when [N] is increased to 18 at % due to an increase of Si–N

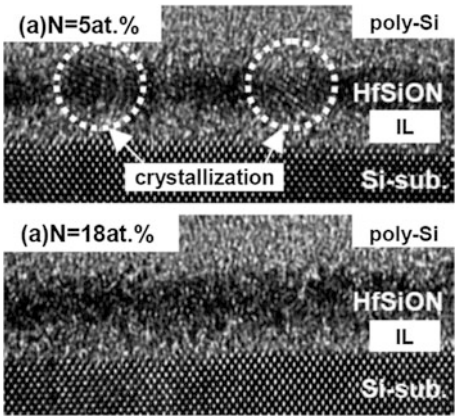


Fig. 4.6 XTEM image of the poly Si/HfSiON/interfacial-SiO_2/Si after activation annealing at 1,000 °C with respect to different nitrogen concentrations [14]

bond in the film. By increasing the [N], Weibull slope was confirmed to increase as well as the improvement of fluctuation of post-breakdown conductance [14].

Surface morphology by observing the crystalline domain is useful to understand crystallographic signature of HfO_2. "Nitrato" mentioned in Sect. 4.2.3, as an example, is CVD-HfO_2 and is deposited using the $Hf(NO_3)_4$ solid phase precursor on Si_3N_4 interfacial layer which is formed by a Rapid Thermal Nitridation (RTN). From X-Ray Diffraction (XRD) spectroscopy, the as-deposited 4-nm-thick HfO_2 film reveals amorphous phase and no crystal grains are observed by plan view TEM.

The grains begin to appear from elevated temperature at around 600 °C. Figure 4.7 indicates the in situ plan view TEM of "nitrato" observed after PDA (800 °C for 30 s in N_2) of the sample at 500 and 900 °C in Ultra High Vacuum (UHV), revealing approximate grain sizes of 8 nm for the sample annealed at 900 °C. The film thickness observed by ellipsometry before and after the PDA, indicates clear densification of the film as shown in Fig. 4.8.

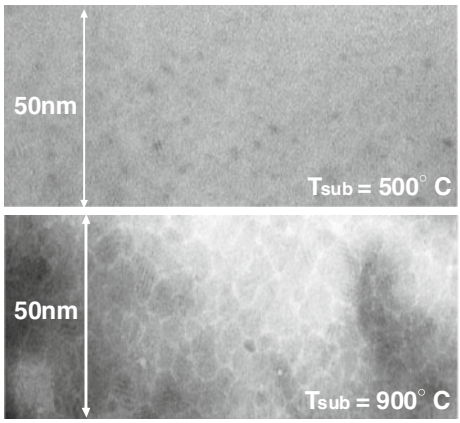


Fig. 4.7 In-situ plan view TEM of the CVD-HfO_2 ("nitrato"), deposited on Si_3N_4 interfacial layer by $Hf(NO_3)_4$ solid phase precursor, observed after PDA (800 °C for 30 s in N_2) at 500 and 900 °C in UHV

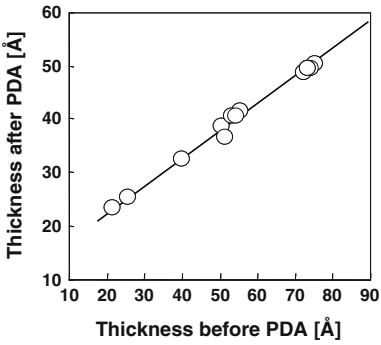


Fig. 4.8 Thickness change of the CVD-HfO_2 ("nitrato") before and after the PDA at an annealing temperature of 900 °C

4.3.2.2 Role of Nitrogen

Nitrogen incorporation is absolutely imperative from the points of not only preventing the crystallization of the bulk HfO_2 but also preventing various reactions at Hf-based high-k gate dielectric film as well as phase separation of the Hf-based bulk material. The amorphous state in $HfSiO_x$ film against high temperature annealing can be kept up by plasma nitridation and annealing under the NH_3 ambient, i.e., $HfSiO_x$ with PDA (NH_3 treatment under 600 °C).

A nitride layer between the poly silicon gate and HfSiON is well known to be capable of preventing Hf from reacting with Si and Boron in the p+ poly Si from diffusing into Si substrate. This is going to be discussed in detail in Sect. 4.4.4. For this purpose, plasma nitridation and/or annealing under the NH_3 ambient are common techniques.

In case that the nitrogen is incorporated into the bulk $HfSiO_x$, the dielectric constant is effectively increased and even after high temperature annealing at 1,100 °C, the high permittivity value is maintained as shown in Fig. 4.9. Also boron penetration from poly Si is effectively suppressed in the HfSiON. As illustrated in Fig. 4.10, these properties are ascribed to the homogeneity of the bond structure in the film [15]. As described in the previous section, in contrast to the HfO_2, the HfSiON can avoid silicidation with Si substrate and retain a higher k value than $HfSiO_x$. The interface property of the HfSiON can be improved by plasma oxidation, which leads to higher drivability [16].

Incorporation of nitrogen also induces a benefit for metal gate processing. It is confirmed, by a comparison with La capped HfSiON and HfSiO, that nitrogen has a role to prevent La from diffusion into the high-k whereby an interaction of La with $HfSiON/SiO_2$ is less than that with $HfSiO/SiO_2$ [17]. Thus, nitrogen incorporation is believed to improve the thermodynamic stability [18].

However, care should be taken for the nitrogen incorporation from a device characteristics viewpoint. Nitrogen incorporation influences the channel mobility, which is remarkable for PMOS due to remote coulomb scattering. Therefore, nitrogen should be located away from the channel and the content should be optimized in order not to degrade hole mobility as well as NBTI (Negative Bias

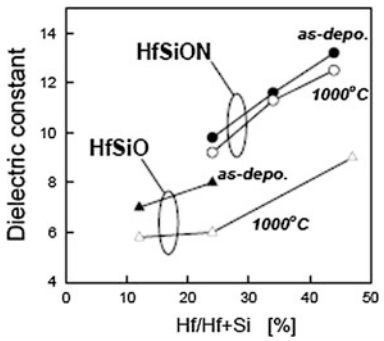


Fig. 4.9 Dielectric constant after annealing at 1,000 °C

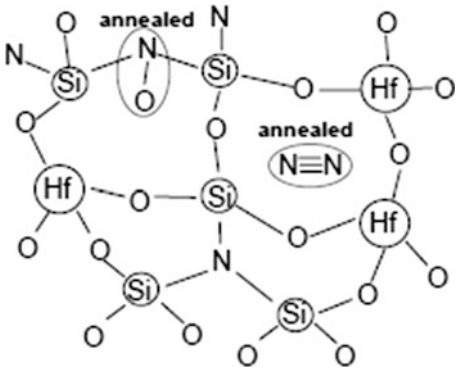


Fig. 4.10 Schematic of bond configuration before thermal annealing

Temperature Instability). Also, the gate leakage tends to increase due to the smaller band-gap when nitrogen is incorporated in large quantity. Nitrogen with an optimum content (3–6 at %) in the MOCVD-HfSiON was found to be a key parameter to obtain thin EOT of 0.8 nm for CMOS with Ni-FUSI (Fully Silicided) gate [19]. How to incorporate nitrogen into the Hf-based high-k gate dielectric is a key issue and a variety of approach has been reported in addition to plasma nitridation and NH_3 annealing. These include direct formation of HfON, deposition of a nitride layer such as SiON and/or SiN onto HfO_2, and adding other element such as Al, Si, Zr to form HfAlON, HfSiON, and HfZrON or HfZrO [20]. The Zr doping technique is taken up in Sect. 4.3.5.

4.3.3 Thermal Treatment and Interfacial Reaction

In relation to crystallization of the bulk of Hf-based high-k material, it is important to control the high-k gate stack interfaces. Not only prevention of crystallization but also its phase separation and inter-diffusion during thermal processing are critical issues. Reaction at the bottom interface, i.e., interface between high-k and interfacial layer on Si substrate (IL), is discussed as well as the control of the top

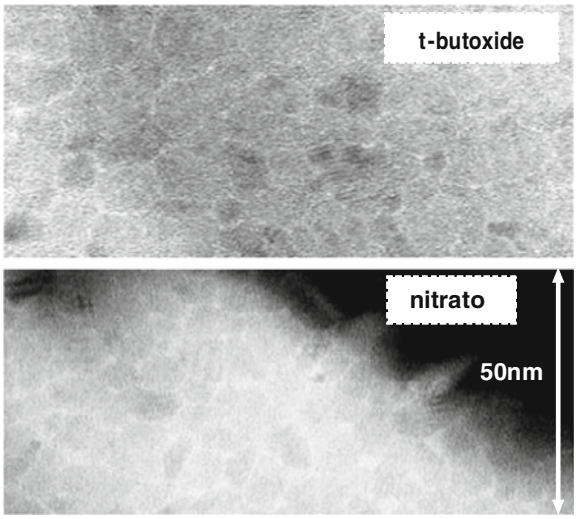


Fig. 4.11 In-situ plan view TEM image of the CVD-HfO_2 prepared from different precursors (Hf-t-butoxide and Hf-nitrato) after annealing at 900 °C

interface reaction. Difference in the solid state properties due to different precursors (Hf-t-butoxide vs. Hf-nitrato) is discussed here [21].

Hf-nitrato, a CVD-HfO_2 produced from the $Hf(NO_3)_4$ solid phase precursor as described in Sect. 4.2.3 is denoted as "nitrato". In contrast, a CVD-HfO_2 film denoted as "t-butoxide" is formed as follows. After RCA cleaning and diluted HF dipping, a HfO_2 film is deposited using the $Hf(OC(CH_3)_3)_4$ [Hf "t-butoxide"] precursor at 500 °C with O_2. This film was treated under PDA at 800 °C for 30 s in N_2, followed by poly-Si deposition with phosphorus implantation and activation annealing at 900 °C in N_2. All CVD-HfO_2 films are deposited on the Si_3N_4 interfacial layer prepared by RTN on Si substrate. As is seen in the Fig. 4.11, similar grains are observed in these samples after the 900 °C anneal and no specific difference is observed. However, the solid state properties of these films seem to have different features. The HfO_2 is prepared by "t-butoxide" (4.1 nm) and "nitrato" (3.2 nm). Figure 4.12 illustrates the Medium Energy Ion Scattering (MEIS) spectra of these HfO_2 after PDA (800 °C in N_2). From these figures, thickness of the HfO_2 from "t-butoxide" is found to be thicker than that from the "nitrato". And near the top surface, a gentle slope is observed in the Hf spectra for the "t-butoxide" indicating out-diffusion of Hf and, from the small surface peak in the Si spectra, out-diffusion of Si is observed for the "nitrato".

In-situ cross-sectional TEM (XTEM) images of these HfO_2 films are investigated as a function of the annealing temperature from room temperature to 860 °C in UHV. The diffusion of Hf into the interfacial layer is clearly observed by the in situ XTEM for the "t-butoxide" as shown in Fig. 4.13, where a new inter-diffusion layer is generated as a result of inter-diffusion of Hf and Si. The hydrogen is supplied from the CVD source of $Hf(OC(CH_3)_3)_4$ and the desorption of hydrogen from the film occurs at $\sim$700 °C resulting in the inter-diffusion of Hf and Si. And inter-mixing of these elements is enhanced by the vacancy evolution

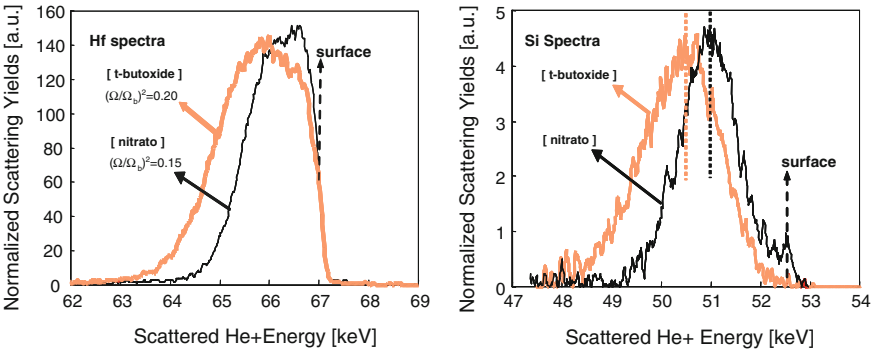


Fig. 4.12 MEIS spectra of the HfO_2 after PDA (800 °C in N_2)

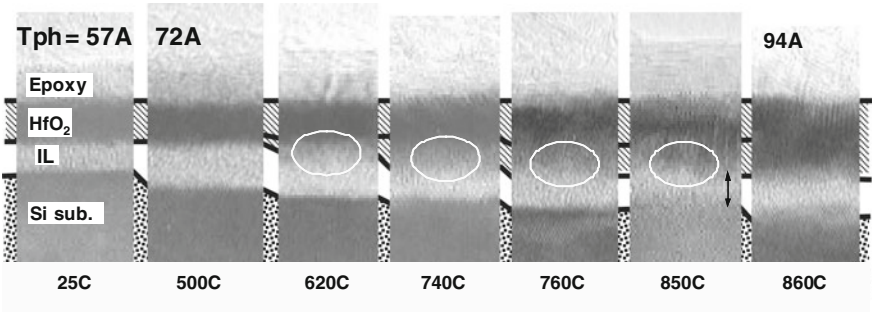


Fig. 4.13 In-situ XTEM images of CVD-HfO_2 by Hf-t-butoxide on 0.8 nm-Si_3N_4/Si substrate as a function of annealing temperature from room temperature to 860 °C in UHV

during the PDA. The oxidation of the Si substrate is observed at even low temperature regime (<500 °C). Gray layer indicated by circles appears in between the HfO_2 and the interfacial layer (IL) above 620 °C. This layer results from Hf diffusion toward the substrate. With elevating temperature, the thickness of the upper dark layer is increased with decreasing interfacial layer (IL) thickness. These diffusion reactions lead to a thinner EOT of the high-k stack. In case of "nitrato", as is shown in Fig. 4.14, the inter-diffusion of Hf and Si at temperature range above 620 °C is small and the thickness of the IL at higher temperature increases apparently compared to the "t-butoxide". This thicker IL of the "nitrato" due to less inter-diffusion leads to a large EOT.

Due to the inter-diffusion, in general, it is difficult to decrease the EOT. The inter-diffusion observed for the sample in Fig. 4.13 is caused by oxygen diffusion and additional oxidation during the conventional thermal process. The minimum EOT can be evaluated by the theoretical equations including densification of the film and the oxidation of the Si substrate as well as Hf diffusion into the interfacial layer [22]. Along with experimental results with poly-Si, the calculated results indicate that the minimum EOT of 1.1–1.2 nm exists at 3.0 nm of physical

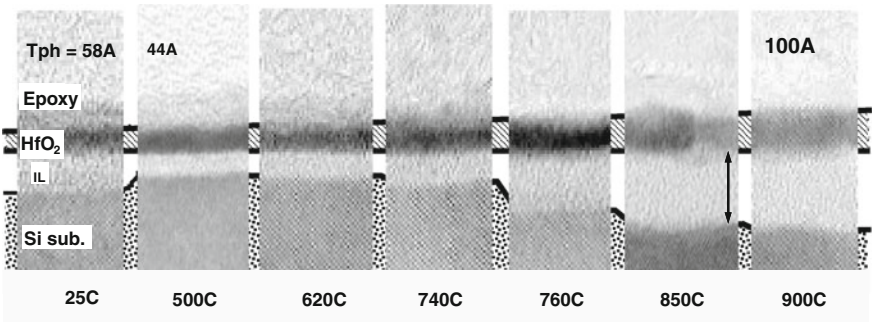


Fig. 4.14 In-situ XTEM images of CVD-HfO_2 by Hf-nitrato on 0.8 nm-Si_3N_4/Si substrate as a function of annealing temperature from room temperature to 900 °C in UHV

thickness of CVD-HfO_2 ("t-butoxide") as shown in Fig. 4.15. This implies that even though the physical thickness is reduced, it is difficult to reduce EOT due to the oxygen diffusion and the additional oxidation during the conventional process unless effective countermeasure to prevent such oxygen behavior is implemented. The partial pressure of the residual oxygen should be controlled rigidly to obtain smaller minimum EOT during fabrication.

An Important issue to note is the influence of interfacial degradation on the electrical characteristics. To understand this, "t-butoxide" and "nitrato" were considered. Figure 4.16 indicates the leakage current characteristics as a function of EOT with respect to three types of capacitor structure; A: TiN/HfO_2/Si_3N_4/Si, B: Poly Si,TiN/HfO_2/SiO_2,Si_3N_4/Si, C: Poly Si/HfO_2/SiO_2/Si. It is noteworthy that the leakage current increases with decreasing EOT. And they have minimum EOT depending on the structure.

This phenomenon is roughly interpreted as an interfacial degradation. For the type A sample, no minimum EOT is observed indicating no interfacial reaction. Meanwhile for the type B, an interfacial reaction at least one side of the interfaces seems to occur, which induces large gate leakage. This interfacial reaction happens for both top and bottom interfaces of the HfO_2 [23]. Each sample configuration is illustrated in the right hand side where wavy lines stand for interface reaction. The

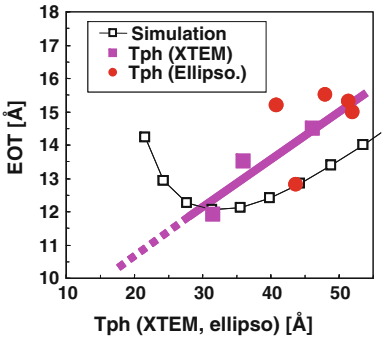


Fig. 4.15 Relationship of EOT and physical thickness for CVD-HfO_2 ("t-butoxide"). According to simulation, minimum EOT was found to exists as ~1.2 nm at around 3.0 nm of physical thickness

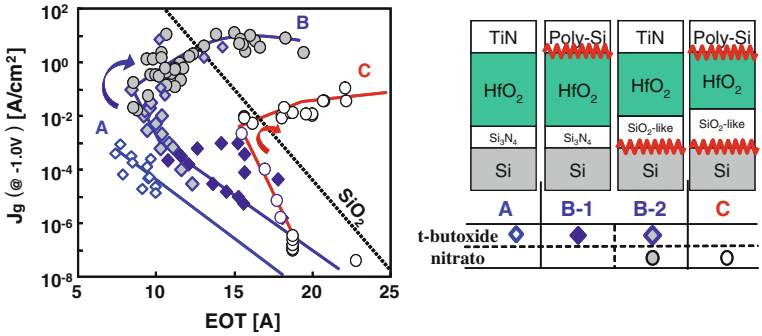


Fig. 4.16 *Left* Leakage current density (J_g) characteristics as a function of equivalent oxide thickness (*EOT*). *Right* Corresponding sample configurations

corresponding sample with respect to each CVD-HfO_2 film ("t-butoxide" and "nitrato") is denoted at the bottom of the right hand side. In type B, two types of reactions (B-1, 2) are possible.

Verification of the interfacial reaction is performed by cross-sectional TEM (XTEM). Figure 4.17 indicates cross-sectional view of poly-Si/CVD-HfO_2 ("t-butoxide")/IL/Si after annealing in N_2 at 900 and 1,150 °C respectively. The interface annealed at 900 °C apparently revealed smooth interfaces compared to the case at 1,150 °C annealing. The HfO_2 film became discontinuous by diffusion of Si atoms from poly-Si, possibly caused by the grain boundaries in the HfO_2. This Si diffusion from poly-Si through the grain boundaries results in an epitaxial growth of Si. This discontinuity of the HfO_2 causes the electrical breakdown. According to detailed analysis, at least two types of different extrinsic defects in the HfO_2 are found to exist. The first one is due to the defects which cannot be seen in the image generated by the impurity in the bulk HfO_2 and roughness at both the interfaces or indirectly related to the grain boundary. The defect density is roughly estimated as 1.5×10^3 cm^{-2} which may induce gate leakage. The second one shown is directly related to the grain boundary. The defect density for the latter case is around $2–8 \times 10^3$ cm^{-2}, which is the direct evidence of the electrical breakdown [24].

Also, specific structural factor which influences reliability of the CVD-HfO_2 ("t-butoxide") is considered as an example [24]. By means of theoretical calculation on the diffusion phenomena in the film, behavior of the out-diffused Si concentration into the HfO_2 is studied based on the reactions of oxidation and diffusion of Si. The oxidation happens at the Si substrate surface and Si within the oxidized layer can out-diffuse by a fundamental diffusion process even though the physical thickness is thin enough. In Fig. 4.18, the Si concentration of the calculated results and the experimental results observed by XPS are plotted. Small deviation near the interface implies the presence of SiO_x and a gentle slope for the HfO_2 film with Si nitride underneath indicates suppressed diffusion of Si compared to the case without the SiN layer.

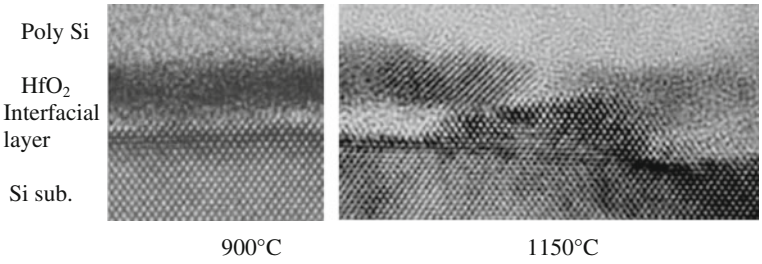


Fig. 4.17 Cross-sectional view of poly Si/CVD-HfO_2 ("t-butoxide")/IL/Si after annealing in N_2 at 900 and 1,150 °C respectively

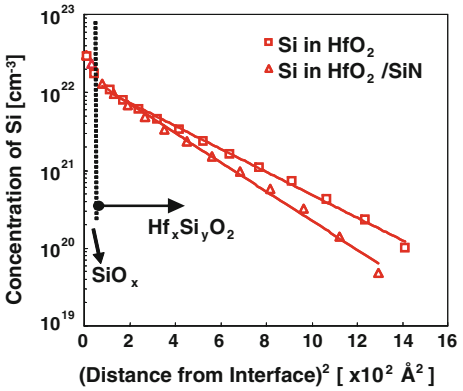


Fig. 4.18 Experimental results on Si concentration in the CVD-HfO_2 ("t-butoxide") with and without SiN layer underneath. *Solid lines* indicate calculated results from diffusion equation [24]

The dielectric constant k of Hf silicate as a function of the Hf ratio [Hf/(Si+Hf)] is a key issue for the electrical property of high-k gate stack. Relationship between the dielectric constant k and the Hf ratio [Hf/(Si+Hf)] as obtained by XPS and observed for Hf silicate is shown in Fig. 4.19. The dielectric constant k is obtained by EOT divided by the total physical Thickness (T_{ph}), EOT/T_{ph}. The k value increases rapidly at lower Hf concentration and stays constant afterwards, and then it increases again for higher Hf concentration.

According to the results in Figs. 4.18 and 4.19, depth profiles for Hf ratio and dielectric constant are estimated as shown in Fig. 4.20a. From the fact that good agreement is found between the experimental data and estimated result by adding "0.4 nm" to the EOT thickness estimated from the k values, interfacial layer with EOT of 0.4 nm is considered to exist in between the high-k dielectric and Si substrate. From Fig. 4.20a, it is clarified that the stoichiometrical composition, Hf/(Hf+Si) is found to be approximately 63 % and the center value of the dielectric constant denoted as "k-transition" indicates approximately as $k = 17$. Although it is difficult to determine the location of the interface from the contrast in the XTEM images in Fig. 4.20b, the stoichiometric interface exists at approximately 1.2 nm and the k-transition locates at approximately 2.0 nm from Si substrate surface [24].

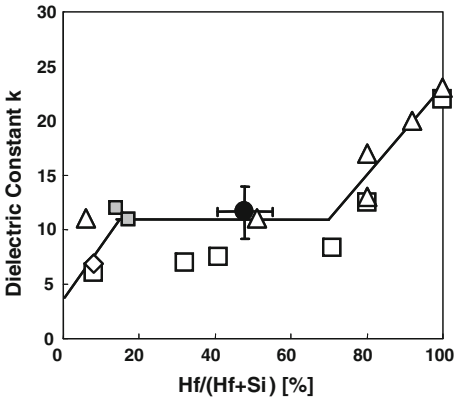


Fig. 4.19 Dielectric constant k for Hf silicate as a function of the Hf ratio [Hf/(Si+Hf)] as obtained by XPS. The k value is obtained by T_{ph}/EOT [24]

The interfacial layer between high-k and Si substrate where the electric field is high enough is an important region because, in the high-k/SiO_2 bi-layer system, it determines the breakdown and reliability factors such as time to breakdown (T_{bd}), Weibull slope (β). This lower k layer is defined at the k-transition interface as shown in Fig. 4.20a. Different from the homogeneous mono-layer gate dielectric like SiO_2 where the breakdown is explained by the percolation model [25], the high-k bi-layer structure reveals complicated behavior. As an example, a physical thickness dependence of the Weibull slope (β) with the same EOT of the high-k dielectric stacked structure is considered. The high-k dielectric is assumed to consist of three layers, i.e., SiO_2 (IL), Hf silicate, and HfO_2 on the Hf silicate. And

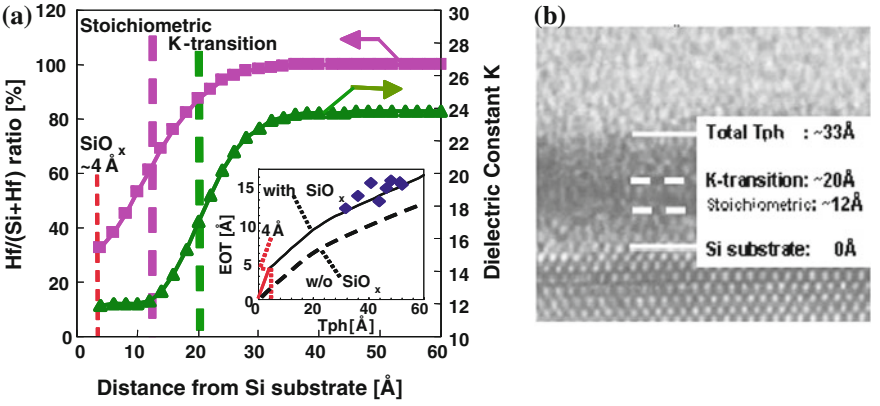


Fig. 4.20 Structural transition region at the *bottom* interface. **a** Hf ratio and k as a function of distance from Si substrate estimated from Figs. 4.18 and 4.19. By shifting the estimated EOT (*solid curve*) to fit to the experimental EOT (*diamonds*), a 0.4 nm thick-interfacial layer seems to exist at the interface. **b** XTEM image of CVD-HfO_2 ($\sim$3.3 nm). The interfaces of k-transition and stoichiometric SiO_x are located at $\sim$2 nm and $\sim$1.2 nm away from the Si substrate, respectively. The k-transition interface influences the reliability and is difficult to detect from the XTEM image

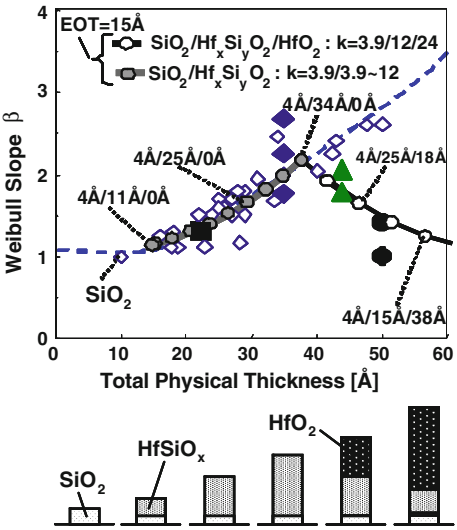


Fig. 4.21 Calculated relation between Weibull slope β and total physical thickness under the same EOT (1.5 nm) for a variety of gate stuck combination of 0.4 nm-SiO_2, Hf silicate, and HfO_2, based on the data in Figs. 4.19 and 4.20. The Weibull slope has its peak for a structure consisting of 0.4/3.4/0 nm thick SiO_2/$HfSiO_x$/HfO_2 stack

dielectric constant in each layer is assumed to be 3.9, 12 and 24, respectively. As indicated in Fig. 4.21, calculated physical thickness dependence of β under the same EOT (1.5 nm) for each structure is plotted together with other reported results. Here, it is assumed that β is dominantly determined by the distance between the k-transition interface and the Si substrate due to higher electrical field in the interfacial layer. Starting from SiO_2, the β increases with increasing physical thickness. Thickness configuration of each structure is changing from 0.4 nm (IL)/1.1 nm (Hf-silicate)/0 nm with $\beta \sim 1.1$–0.4 nm/3.4 nm/0 nm where the β has the maximum value $\beta \sim 2.2$ at around total physical thickness (T_{ph}) of ~ 3.8 nm. The film at this point consists of a single layer silicate on SiO_2 as shown below. Then, it decreases with decreasing physical thickness of Hf-silicate due to crystallization of the HfO_2 layer. This phenomenon implies that the β can be improved by increasing the Hf silicate thickness without the HfO_2 layer and for thicker physical thickness regime, crystallization tends to occur with increasing HfO_2 thickness. From this analysis, it is clarified that the β can be improved by increasing the Hf silicate thickness without any HfO_2 layer, indicating that a single layer silicate is the best structure from the reliability standpoint.

4.3.4 Trapping Property

Although reliability is taken up in detail in Chap. 8, difference in trapping with respect to high-k formation process as well as its material (HfO_2 or Hf-silicate) is considered in this section since carrier trap is an intrinsic phenomenon for Hf-based high-k bulk material.

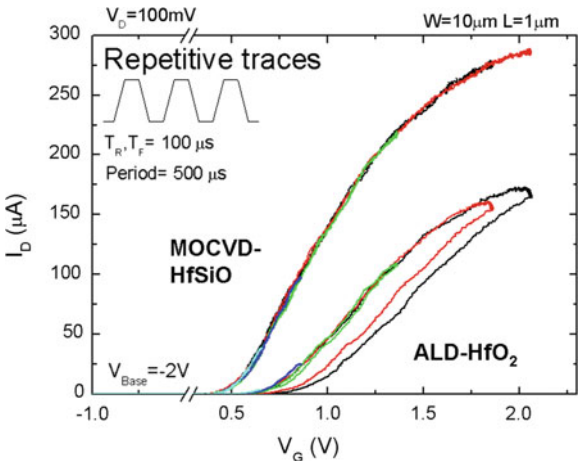


Fig. 4.22 Pulsed $I_d - V_g$ characteristics for NMOS (poly Si/HfSiO and HfO_2)

To understand the trapping behavior, a pulse I–V measurement as well as charge pumping (ICP) are useful techniques [26]. For the FET with the SiO_2/Si system, NBTI (Negative Bias Temperature Instability) for PMOS is more relevant than PBTI (Positive Bias Temperature Instability) for NMOS. This is because PBTI is considered as electron trapping in high-k gate dielectric and thus insensitive to the SiO_2/Si system [27]. In comparison to HfSiON, HfO_2 reveals larger hysteresis at positive gate bias regime, indicating presence of larger number of electron traps in HfO_2 than HfSiON. This becomes serious with less N content as well as increasing of Hf content.

Figure 4.22 presents pulsed $I_d - V_g$ (drain current—gate voltage) characteristics for NMOS with poly Si and HfSiO or HfO_2, respectively, indicating significant large hysteresis for the HfO_2 and strongly reduced electron trapping in the HfSiO layer compared to the ALD HfO_2 in the $V_g > 0$ region.

Precisely, since HfO_2 is formed by ALD and HfSiON is by MOCVD, it is uncertain whether the electron trap is caused by the defects generated in HfO_2 material or by the ALD process. Figure 4.23 illustrates the $I_d - V_g$ characteristics for NMOS with HfO_2 formed by PVD (NH_3 pre-anneal/2 nm Hf(PVD)/Oxidation anneal/SiN cap/poly Si gate), indicating behavior similar to the ALD-HfO_2 baseline and no significant instability is measured for the p-FET ($V_g < 0$ region). From this result, it is conceivable that the origin of the huge hysteresis, i.e., electron trap, is caused by the Hf induced defect in the HfO_2 and not in the Hf-silicate and hole trapping is insignificant in the Hf-based high-k dielectric material. In addition, this trapping becomes remarkable with poly Si gate rather than metal gates.

It is interesting to see the chemistry between the gate electrode and the high-k dielectric material. Conventional technique to measure the charge trapping for high-k material is to measure the hysteresis by stressing and sensing the sample. However, in the presence of the transient charging effect (cf. Chap. 8), pulsed $I - V$ measurement which provides alternative fast measurement is necessary.

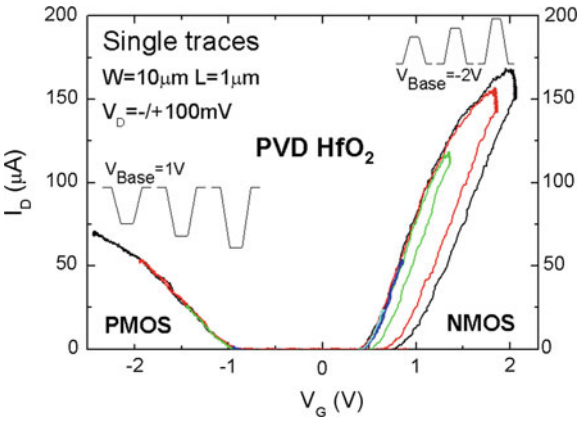


Fig. 4.23 Pulsed $I_d - V_g$ characteristics for NMOS with HfO_2 formed by PVD (NH_3 pre anneal/2 nm Hf (PVD)/Oxidation anneal/SiN cap/poly Si gate)

Figure 4.24 shows the pulsed $I_d - V_g$ characteristics for N- and PMOS with ALD-HfO_2 with TiN gate instead of poly Si. Electron trapping is negligible at small V_g and tends to be evident for $V_g > 1.5$ V. Compared to the result of ALD-HfO_2 with poly Si gate in Fig. 4.22, the drain current as well as the hysteresis are improved. This is due to the elimination of the depletion layer caused by the poly Si and less negative charge trapping in the bulk HfO_2. On the other hand, when compared to the PMOS results in Fig. 4.23, strong positive charge trapping is observed for negative V_g. This behavior depends on the HfO_2 thickness indicating hole trapping in the bulk HfO_2 when TiN is formed on the HfO_2 gate dielectric.

Nitrogen effect is confirmed by the pulsed $I_d - V_g$ characteristics in Fig. 4.25. The capping SiN layer is formed under the condition of 600 °C, 15 s. Compared with ALD-HfO_2 result in Fig. 4.22, this indicates a certain level of reduction of the traps for NMOS. Considering the sweeping direction, the PMOS behavior is more complicated than NMOS, where trapping and detrapping co-exist indicating that positive and negative charge trappings are occurring.

4.3.5 Doping Effect

Attempt to dope some elements into HfO_2 to improve the device performance is a promising approach. Al doped into HfO_2 and ZrO_2 [28] was reported to improve its thermal stability as well as the Zr-Al mixtured dielectric [29]. In addition, Si doping into HfO_2, i.e., hafnium silicate ($HfSi_xO_y$) gate dielectric films, has indicated better electrical characteristics such as lower leakage, relatively thin EOT with no significant hysteresis, low mid-gap interface state density and high breakdown voltage. Cross-sectional TEM showed no signs of reaction or crystallization in the films being annealed at 800 °C for 30 min [30]. Beside this, fluorine is found to be effective to obtain higher work function (lower V_{th}) by inducing negative charge in the film for PMOS [31].

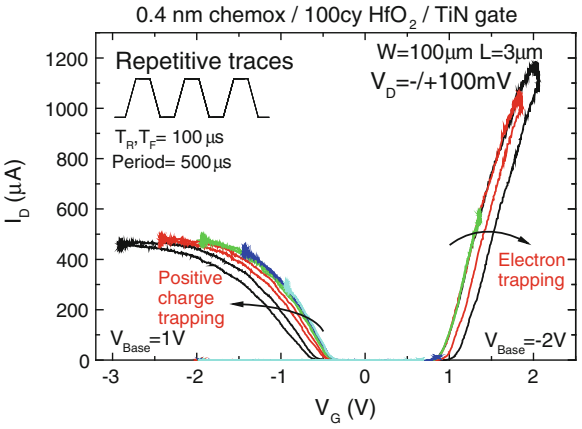


Fig. 4.24 Pulsed $I_d - V_g$ characteristics for N- and PMOS with ALD-HfO_2 and TiN gate

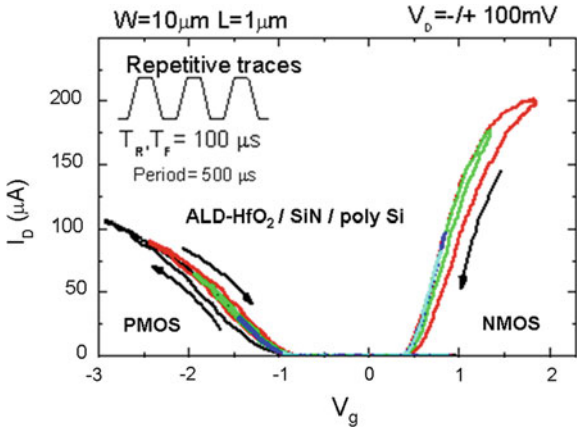


Fig. 4.25 Pulsed $I_d - V_g$ characteristics for FETs with 0.7 nm chemox/40 cy HfO_2/SiN (600 °C, 15 s)/poly Si structure

As an example, Zr doping is discussed here, in the quest to improve the reliability as well as its performance. A hafnium zirconate ($HfZrO_4$) alloy gate dielectric formed by ALD with tantalum carbide (Ta_xC_y) metal gate has been reported for advanced gate stack applications focusing on its dielectric properties, device performance, and reliability (Fig. 4.26) [32]. The $HfZrO_4$ dielectric film exhibits narrower band gap, smaller and more uniform grains, less charge traps than HfO_2 as well as good thermal stability with silicon. The $HfZrO_4$ gate dielectric is found to have smaller capacitance equivalent thickness (CET) owing to its thinner IL and higher k value as shown in Fig. 4.27, lower threshold voltage (V_{th}) and better reliability (with respect to charge trapping, C–V hysteresis, interface state density, and positive bias temperature instability) compared to HfO_2.

In this manner, the incorporation of zirconium dioxide (ZrO_2) into HfO_2 brings in a variety of advantages by enhancing the dielectric constant (k) of the resulting $HfZrO_4$ which is associated with structural phase transformation from mainly monoclinic to tetragonal (Fig. 4.28). The tetragonal phase has a higher k value as

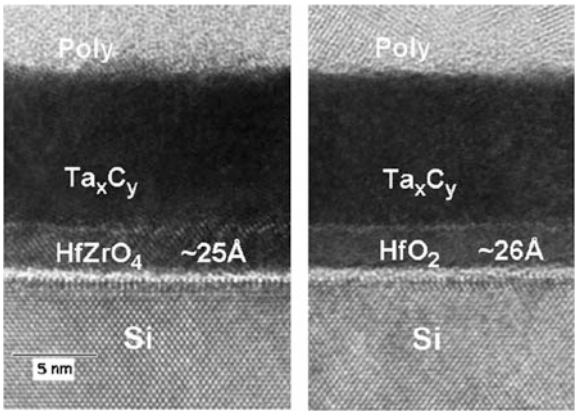


Fig. 4.26 XTEM of hafnium zirconate ($HfZrO_4$) alloy and HfO_2 with Ta_xC_y gate dielectrics

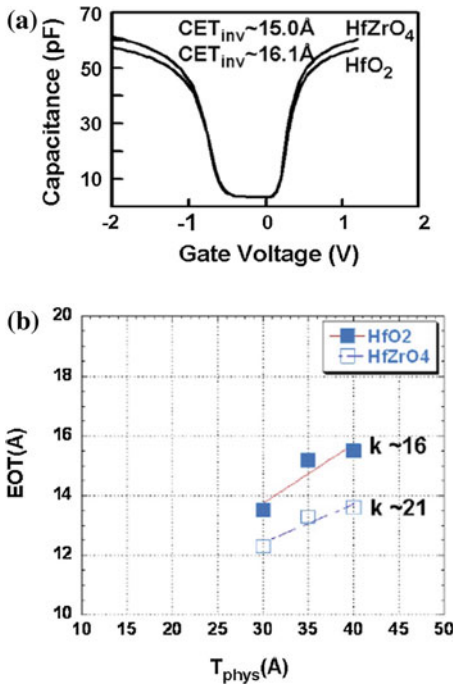


Fig. 4.27 CV curves and physical thickness dependence of EOT for $HfZrO_4$ alloy and HfO_2 gate stacks

predicted. The improved device characteristics are attributed to less oxygen vacancy in the fine-grained microstructure of the $HfZrO_4$ films. On the other hand, an issue of concern is the possibility of the presence of micro-crystals in the $HfZrO_4$ film, which will induce TDDB (Time Dependent Dielectric Breakdown) degradation. Also the electrical property of the HfZrO film as well as its microstructure depends on the thermal treatment method. The surface of the RTP (Rapid Thermal Processing) treated sample (1,000 °C, 5 s) has rough morphology, whereas, laser annealed

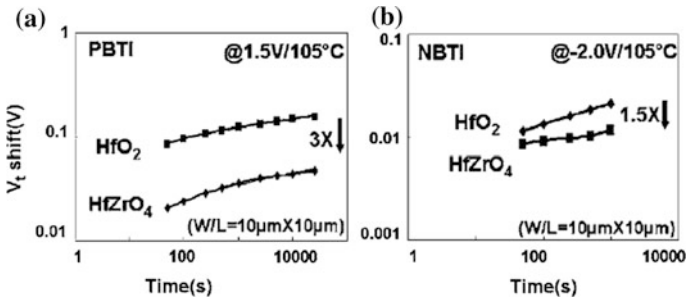


Fig. 4.28 BTI characteristics with respect to $HfZrO_4$ alloy and HfO_2 gate stack

(1,325 °C, spike) sample has smother surface and improved electrical characteristics [33]. Further analysis will be necessary for this issue.

4.4 Device Processing of Hf-Based High-k FET/CMOS

In this section, based on the previous discussions, characteristic features caused by the bulk material property and the interface control, which are closely related to the reliability, are discussed. Reliability assessment in terms of device degradation including measurement techniques, lifetime prediction and other trap related issues is important. Pre-existing traps which are the seeds of generating traps in the HfSiON is found to affect strongly the negative bias temperature instability (NBTI) [34].

4.4.1 Bi-Layer System

It is important to understand the bi-layer system conventionally used for the high-k/metal gate stack. Gate leakage characteristics of the high-k dielectrics were introduced in Sect. 4.3.1. Generally, the actual high-k gate stack has the interfacial layer (IL) to realize the high quality interface. Figure 4.29 illustrates the relation between the gate leakage current density and EOT for single layer systems with respect to SiO_2, high-k dielectric, bi-layer structure of high-k and interfacial SiO_2. In case of the single dielectric system, when the physical thickness of the dielectric is rather thin, the gate leakage indicates a direct tunneling (DT) behavior across its thickness and for larger physical thickness, the tunneling current behaves as Fowler-Nordheim tunneling (FNT) which has a rather gentler slope than DT in the J_g−EOT relation. Since the band gap is narrower for high-k dielectric than SiO_2, the tunneling current shows trap assisted tunneling (TAT) by means of hopping conduction through the shallow traps in the high-k dielectric. In this way, the gate leakage behavior has a

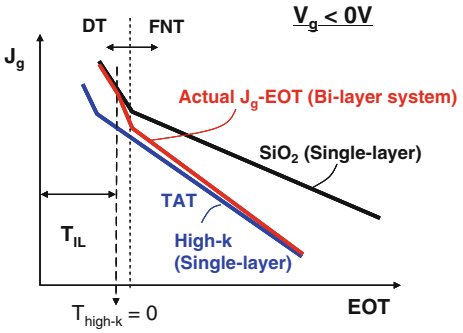


Fig. 4.29 Gate leakage characteristics with respect to single layer and bi-layer structure with high-k and interfacial SiO_2 under $V_g < 0$ V

kink at a turning point of EOT between FNT and DT in the J_g−EOT relation and this feature is remarkable for the gate injection mode ($V_g < 0$ V). Black line indicates the case for SiO_2 and blue line indicates for high-k dielectric under single layer structure. In case for the bi-layer stack system with high-k and SiO_2, the leakage characteristic is rather complicated and may be interpreted as follows. The FNT is dominant at thicker EOT regime and in contrast, DT becomes dominant at around the turning point and tends to approach to the DT value of SiO_2 asymptotically at ultra-thin regime and finally it reaches to the DT leakage value of SiO_2 when the high-k gate dielectric thickness becomes zero ($T_{high-k} = 0$). This, as is discussed in Sect. 4.2.5, is due to the presence of an interfacial oxide layer.

In case of thin interfacial SiO_2 layer, the electrons are injected into the bulk HfO_2 creating defects [35]. Concerning the overall degradation mechanism for high-k/metal gate stack, it has been reported that SiO_2 interfacial layer is considered to control the degradation and breakdown under the inversion mode [36].

In addition to the leakage characteristics, combination of the HfO_2 and the interfacial oxide can also modulate threshold voltage of the transistor. The conventional hysteresis measurement presents a fast screening method to study high-k gate dielectrics. Figure 4.30 shows stress induced V_{th} shifts versus maximum Si

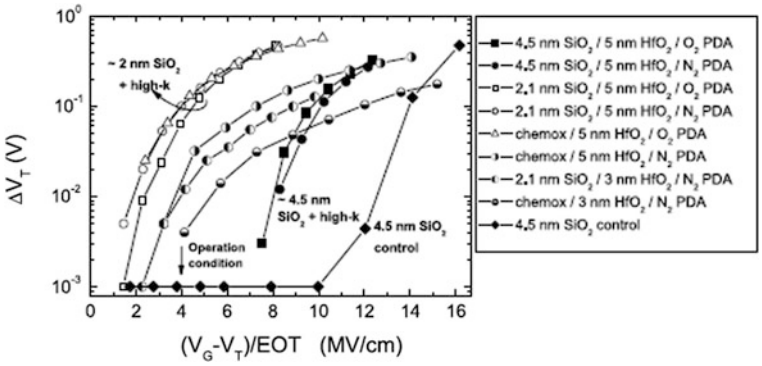


Fig. 4.30 Stress induced V_{th} shifts versus maximum Si field for various HfO_2 gate stacks with different interfacial SiO_2 thickness

field for various HfO_2 gate stacks with different interfacial SiO_2 thickness. The results show the shift in the V_{th} strongly depends on the thickness of the interfacial SiO_2 as well as the HfO_2 layer. The V_{th} shift increases with increasing HfO_2 and decreasing interfacial SiO_2 thickness. The former is due to the increase of Hf content in the high-k. Regarding the interfacial SiO_2 thickness effect, it is explained by the injection of carriers by tunneling through the interfacial SiO_2 layer. For such a bi-layer system, both TDDB and mobility increase with increasing high-k thickness and interfacial oxide thickness. TDDB improves with increasing the high-k thickness under the same interfacial oxide.

4.4.2 Oxidation of Metallic-Hf

High-k deposition technique by oxidizing the metallic-Hf (mHf) is considered here. Combination of a metal sputtering and successive oxidation is a reasonable approach for fabricating the high-k gate stack with controlled interface and thickness uniformity across the wafer. The important advantages of the PVD method are its higher degree of purity and less water content in the film compared to the other methods.

In this section HfO_2/Si interface stability is studied for various oxidation processes [37]. HfO_2 films are prepared on p-type Si(100) substrates by two different methods, i.e., (1) 3.2 nm-thick Hf metal layer is deposited by DC sputtering using Hf target in Ar (0.5 Torr), followed by Rapid Thermal Annealing (RTA) for 60 s in O_2 (760 Torr) and (2) Hf metal layer of 0–2.6 nm is pre-deposited and 1.3 nm-thick HfO_2 is successively deposited by reactive sputtering in Ar/O_2(=50/50) as illustrated in Fig. 4.31. The pre-deposited Hf was oxidized by the exposure to the Ar/O_2 plasma. The HfO_2/Si capacitors are fabricated using TiN or TaN gate electrodes [38]. Generally, thermal oxidation is achieved by oxygen diffusion at high temperature and facilitates crystallization. In contrast, plasma oxidation is performed by implantation of oxygen ion at low temperature, keeping an amorphous state. The Hf metal pre-deposition technique was studied to optimize the plasma oxidation conditions and to minimize CET [39]. The impact of Hf metal pre-deposition in CVD- and PVD-HfO_2 dielectrics was investigated [40]. As is illustrated in the figure, a deposition by reactive sputtering without the Hf metal layer on the Si substrate, oxygen species in plasma oxidize Si directly or through

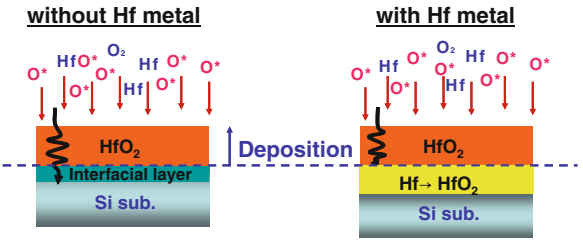


Fig. 4.31 Comparison of the deposition method by PVD with and without pre-deposited Hf metal layer

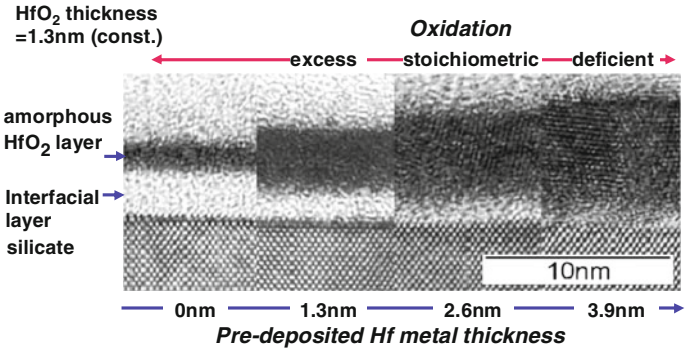


Fig. 4.32 XTEM images of the interface prepared by Hf/HfO_2 deposition

the HfO_2 layer during deposition, leading to a thick interfacial growth. In contrast, if Si substrate surface is coated with Hf metal, the oxygen species are consumed first to react with the pre-deposited Hf metal, therefore oxidation of Si is effectively suppressed. This indicates that Hf metal pre-deposition is an effective method to suppress the interfacial layer growth.

Figure 4.32 shows XTEM images of the interface prepared by the Hf/HfO_2 deposition. Target thickness of the HfO_2 is constantly 1.3 nm and pre-deposited Hf thickness is changed. At 0 nm when HfO_2 is directly deposited, a thick interfacial layer is formed. This thick layer is caused by an excess oxidation of the Si substrate. In case the Hf metal is pre-deposited, it is also oxidized in the amorphous phase and consequently the interfacial oxide layer growth is suppressed with increasing the Hf metal thickness. Thus, Hf metal pre-deposition is capable of minimizing the interfacial layer formation by limiting the amount of oxygen species available to interact with Si. In case of 3.9 nm Hf pre-deposited, the oxidation of Hf is deficient and metallic Hf remains. These observations are also confirmed by XPS analysis. From the image contrast and compositional data, the HfO_2 layer is considered to be amorphous and the interfacial layer is silicate.

Hf metal thickness dependence of electrical properties and thickness of each layer is shown in Fig. 4.33. The CET is found to have the largest value without the Hf metal pre-deposition, due to the thick interfacial layer. With increasing Hf metal thickness, the CET decreases dramatically, corresponding to the thickness reduction of the interfacial layer. For a further increase of the Hf metal thickness over 1.5 nm, the CET increases again, due to an increase of the total thickness of HfO_2. Thus, minimum CET is obtained at an optimum Hf metal thickness using plasma oxidation of Hf metal at low temperatures with amorphous HfO_2. Concurrently, the leakage current J_g is reduced for thin and thick Hf metal where the CET obtained is thick. It is plausible that the interfacial layer is a silicate ($k = 8$), which is conduced by the non-equilibrium nature and/or the mixing effect during plasma oxidation.

It is interesting to compare the electrical properties of the PVD and with the CVD HfO_2. Figure 4.34 indicates the relation between leakage current density J_g

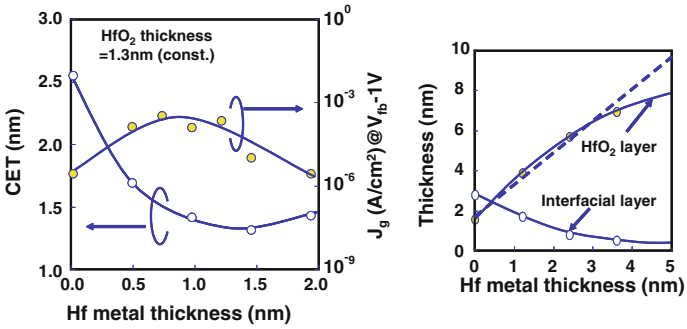


Fig. 4.33 Hf metal thickness dependences of CET, J_g, and thickness of each layer

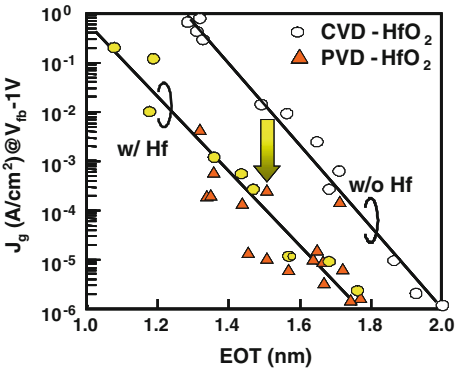


Fig. 4.34 Relation between leakage current and EOT value with respect to each type of HfO_2

and EOT value with respect to each type of HfO_2. Effective reduction of J_g or EOT is achieved by the Hf metal pre-deposition method.

Figure 4.35 indicates the flat band voltage, V_{fb}, (left hand side) observed from the high frequency CV (capacitance–voltage) curve (right hand side). Compared to the CVD HfO_2 without the Hf metal layer, the PVD-HfO_2 has a larger flat band voltage and a larger shift. Also V_{fb} of the CVD-HfO_2 on Hf metal (denoted as filled circle) has values in between them. Large humps observed for the PVD-HfO_2 with Hf metal in the CV curve imply that this flat band voltage shift originated from the sputtering damage. Longer sputtering time for thicker HfO_2 films is likely to be more damaging. However, as shown in the V_{fb} characteristics (left hand side), this damage can be recovered by post metallization annealing (PMA; 650 °C for 30 min in N_2) to the similar level as that of CVD-HfO_2/Hf.

Although the above discussion relates to the case of HfO_2 deposition on Hf metal layer, a simpler approach is the direct oxidation of a pre-deposited metal Hf layer. Figure 4.36 illustrates a specific PVD formation method which consists of two steps [40]. First, by sputtering, pure Hf-metal is deposited onto Si substrate prepared by HF last treatment. As described above, the pure Hf-metal is deposited uniformly onto any surfaces with high density, and less water content. By using a

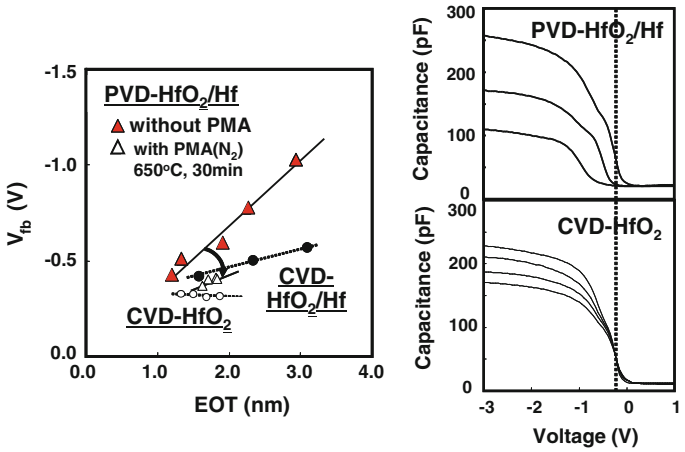


Fig. 4.35 V_{fb} and CV characteristics with respect to PVD- and CVD HfO_2 with and without Hf metal

pure Hf or HfSi target (5N grade), Hf or HfSi is deposited by co-sputtering. i.e., DC (Direct Current) sputtering in Ar under 400 Pa at room temperature. The second step is its oxidation to form HfO_2 by rapid thermal oxidation (RTO) or remote-plasma oxidation (RPO). For thick HfO_2, RPO is often used because of its higher deposition rate by a radical active oxygen [41].

Hf-metal can be oxidized uniformly by the oxygen radicals where Hf acts as a blocking layer to suppress the Si oxidation at low temperature. The HfO_2 thickness is determined by the pre-deposited Hf metal thickness and the interfacial layer can be adjusted by the RPO treatment time. This enables the gate stack structure to be optimized to achieve the best performance. As indicated in Fig. 4.37, the advantage of this method is that Hf can be easily and selectively oxidized when the oxidation temperature is chosen properly ($T = T_2$) where the oxidation rate is stable for both Hf and Si. In addition, the interfacial oxide layer thickness between Si substrate and HfO_2 layer can also be easily controlled by changing the RPO treatment time. Then, PDA is carried out at 650 °C for 15 min. As a gate

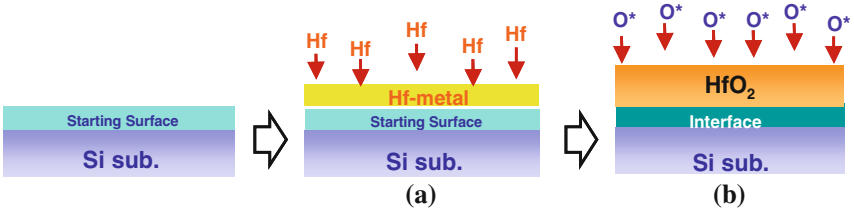


Fig. 4.36 PVD HfO_2 formation which consist of Hf-metal sputter (**a**) and remote plasma oxidation (**b**)

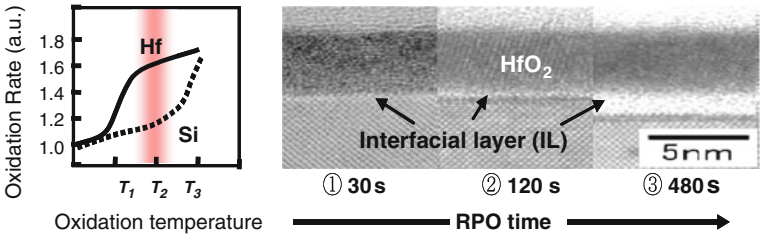


Fig. 4.37 Oxidation of pre-deposited Hf metal and precise control of the interfacial layer thickness by RPO

electrode, CVD-TiN is introduced. Similar electrical characteristics are obtained by this method. This systematic approach leads to a wider process window.

4.4.3 EOT Scaling

The gate stack scalability by the "Hf-metal sputter+RPO" technique is studied. As is explained previously, the interface thickness is controlled by the extended RPO time (Fig. 4.38). The RPO time dependence of EOT and leakage current density J_{g} for 3 nm-thick Hf metal are plotted in this figure. The pre-deposited Hf metal is selectively oxidized and then starts to grow during the extended RPO treatment time after complete oxidation of the Hf metal. Transistors using this technique are fabricated with a conventional self-aligned process flow up to Ni silicidation. Hf metal ranging from 0.5 to 3 nm is deposited on several kinds of starting surfaces pre-cleaned by HF-last followed by rapid thermal nitridation (RTN) and oxidation (RTO).

The deposited Hf metal was successively oxidized to form the HfO_2 layer increasing its thickness by a factor of 1.6 by RPO where oxygen radicals selectively oxidized the Hf at low temperature (400 °C). Poly-Si and source/drain activation treatment were performed under 1,000 °C for 3 s. The interface is controlled by the extended RPO time. Figure 4.39 presents plots of EOT versus the

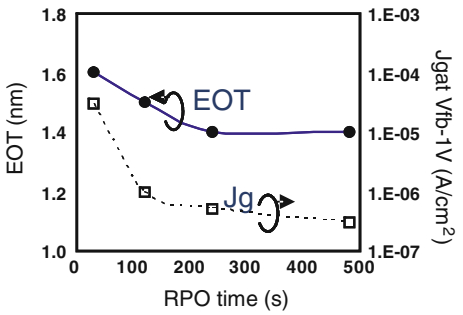


Fig. 4.38 RPO time dependence of EOT and corresponding J_g for 3 nm-Hf metal

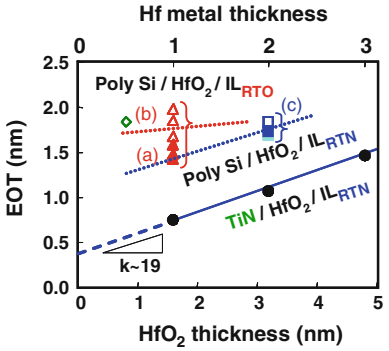


Fig. 4.39 Gate stack scaling by Hf-metal thickness change with respect to various combination of the interfacial layer (RTO and RTN) and the gate electrode (Poly Si and TiN)

physical thickness of HfO_2 with TiN gate, quoted from previous work [42]. From these results the k-value of HfO_2 is found to be k $\sim$19 and the contribution of the interface turned out to be approx. 0.4 nm in EOT and from (4.1) in Sect. 4.2.5, the ratio of physical thickness of nitride interface and the interfacial k value (T_{phys}(IL)/k(IL)) turn out to be approximately 0.1.

However, under the poly-Si gate process, overall increase in EOT by $\sim$0.5 nm occurred for the RTN interface, compared to the TiN gate, mainly due to an additional interface growth during the Poly-Si deposition or/and other possible reasons as discussed in Sect. 4.4.5. On the other hand, for the RTO interface, less EOT dependence on the HfO_2 thickness and a relatively thick EOT was observed in the thin Hf metal thickness region. This implies that an additional bottom interface growth and/or transformation into the silicate phase is enhanced in the thinner Hf metal regime for the RTO interface, compared to that of the RTN. Relations of EOT versus J_g (leakage current density) based on these scaling methods are plotted in Fig. 4.40. Apart from reducing of the bulk HfO_2 thickness, the EOT scaling has also been achieved by reducing the RPO time and/or introducing different types of the interface.

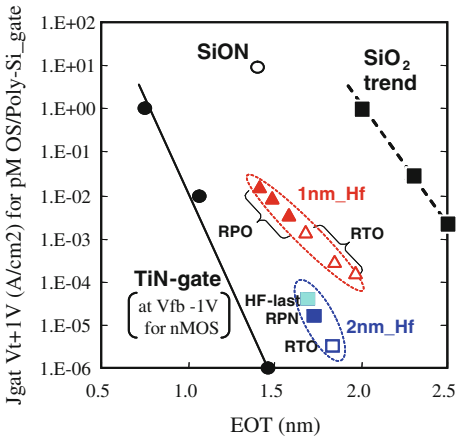


Fig. 4.40 EOT dependence of J_g for the scaled gate stacks with SiON reference and SiO_2 trend. The EOT scaling is performed by changing the RPO treatment time and/or different types of interfacial layer apart from thinning the bulk HfO_2

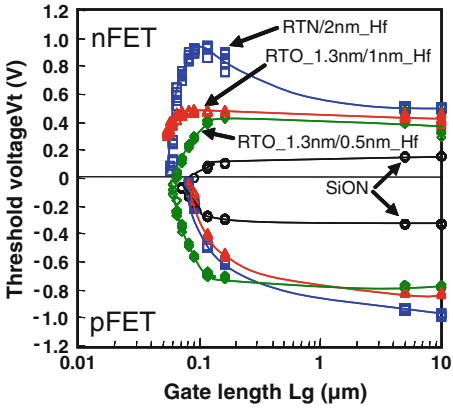


Fig. 4.41 Gate length dependence of the threshold voltage with respect to N- and P-MOS for various combinations of Hf-metal thickness and interfacial layer

In Fig. 4.41, V_{th} for the poly-Si gated FETs with a variety of gate dielectrics is plotted as a function of the physical gate length (L_g). As is seen in this figure, an anomalous V_{th} shift appears at around $L_g \sim 0.1$ μm in addition to the normal short channel effect (SCE). Also, an anomalous V_{th} lowering appears for p-FET. These V_{th} roll-off for pFET and inverse roll-off for nFET tend to be pronounced with increasing the HfO_2 thickness. Assuming that negative fixed charges remain in the HfO_2, these undesirable phenomena can be explained without contradiction and this is considered to be a main reason. Mechanical stress and/or Si-recess introduced during the gate fabrication may also influence these phenomena. The thick RTO interface is found to improve the V_{th} hysteresis to less than 10 mV and the $I_{on} - I_{off}$ characteristics can be optimized by tuning the RTO and the Hf-metal thickness to 1.3 and 1 nm, respectively [43].

4.4.4 Process Control

4.4.4.1 Bulk Process [44]

Suppression of the top interface degradation caused by the interfacial reaction between gate electrode and high-k gate dielectric is also a key issue, especially for the case when the metal gate is introduced. Improvement of the electrical properties, degraded by the interfacial reaction between HfO_2 and Ni-rich silicide for the metal gate/SiN/PVD-HfO_2 gate stack, is discussed in this section by introducing the ultra-thin nitrogen layer followed by optimization of the PDA condition. The high-k film is formed by sputtering of 1 nm-Hf metal on the 1.3 nm-RTO. The Hf metal is successively oxidized to form HfO_2 by RPO at 400 °C. A SiN cap is deposited after the PDA to suppress the interfacial reaction between the HfO_2 and the gate electrode. In this case, Ni-rich fully silicided (FUSI) gate is introduced as a metal gate electrode.

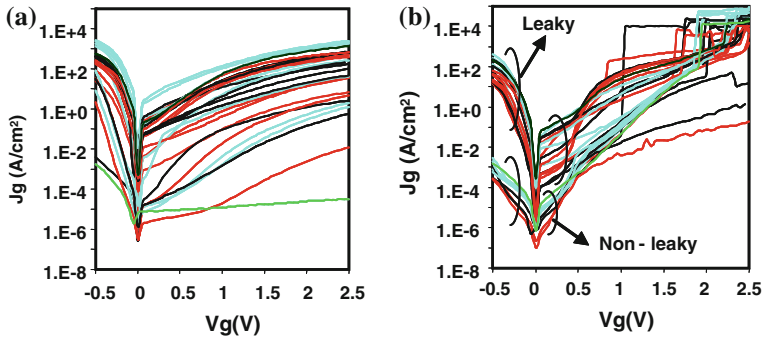


Fig. 4.42 $J_g - V_g$ of n-MOS capacitors for Ni-FUSI/(SiN)/HfO_2 gate stacks without SiN cap between FUSI and HfO_2 (**a**) and with SiN capping (**b**), after 700 °C PDA

Figure 4.42 shows the $J_g - V_g$ characteristics for the Ni-FUSI/(SiN)/HfO_2 gate stack after 700 °C PDA treatment. In the case without SiN capping, J_g is scattered over a wide range [(a) in this figure]; in contrast, two types of leakage modes are observed with the SiN cap: leaky part and non-leaky part as indicated in (b) of this figure. Similar dispersion is observed for the poly-Si/HfO_2 gate stacks without the SiN cap. On the other hand, the $J_g - V_g$ curves for poly-Si/SiN/HfO_2 gate stacks did not disperse so much (not shown here). From these results, SiN is found to have a role in suppressing the top interfacial reaction to some extent.

The PDA treatment strongly influences the WF by modulating the oxygen vacancy [45]. Comparison of $I_{on} - I_{off}$ behavior with respect to poly-Si or Ni-FUSI/SiN/HfO_2 gate stacks and with PDA at 700 °C is shown in Fig. 4.43. The poly-Si/SiN/HfO_2 results do not show data scatter; the $J_g - V_g$ curves had similar trend as well; in contrast, the data for the NiSi-FUSI gated transistors with the SiN cap revealed wide spread and in addition, some FETs do not work at all.

Possible explanation for this extraordinary spread in the $I_{on} - I_{off}$ characteristics, could be some micro holes in the flimsy SiN capping layer and/or roughness between the SiN and the HfO_2 layers formed during the processing. Some areas, which contain such holes or irregular thickness distribution due to roughness result in high leakage current whereas the other area without any holes nor roughness exhibit less leaky behavior (Fig. 4.44). Such undesirable issues can be suppressed by modifying the PDA condition which may change the surface roughness as well as the film quality. This can be performed by elevating the PDA condition and increasing the nitrogen pressure during the PDA. In order to confirm this hypothesis, structure of the interface is investigated with respect to conventional and new PDA treated samples. Since the SiN capping film is extremely thin (<0.5 nm), the film property is supposed to depend on the surface condition/roughness of the HfO_2.

The film quality was examined by elevating the PDA temperature from 700 to 1,000 °C. Figure 4.45 shows the XTEM images with respect to the conventional PDA condition at 700 °C and the new condition of 1,000 °C under rather high

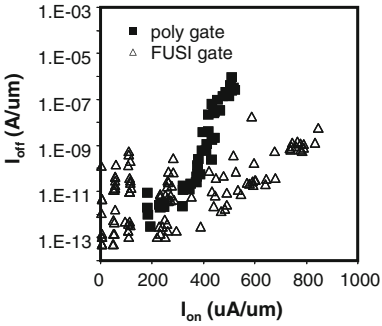


Fig. 4.43 $I_{off} - I_{on}$ characteristics for n-FET with poly and or FUSI gate, and with SiN cap after PDA treatment at 700 °C

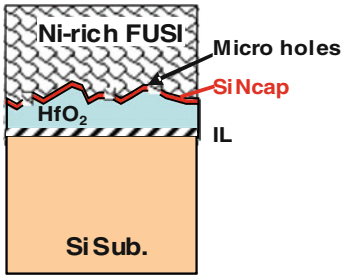


Fig. 4.44 A schematic representation of the flimsy SiN capping layer or roughness between the SiN layer and the HfO_2 to explain the spread behavior in the $I_{on} - I_{off}$ characteristics

nitrogen pressure. Although it is difficult to observe the direct evidence of such holes from the XTEM, some differences can be observed. As for the sample treated at higher PDA temperature (1,000 °C), an ultra thin layer indicated by the bright contrast is observed at upper interface, while this layer is not observable in the conventional PDA treated sample. The ultra-thin layer is considered to be a nitrided layer and the above dispersion in the $I_{on} - I_{off}$ can be explained. From these results, it is confirmed that the SiN capping at elevated temperature (1,000 °C) PDA condition is an important issue (Fig. 4.46). It should be noted that this argument is valid only when the FUSI gate is used. In the case of a pure metal such as TiN, since the interaction between gate and high-k layer does not happen, the SiN capping barrier is not necessary (see Fig. 4.16 in Sect. 4.3.3)..

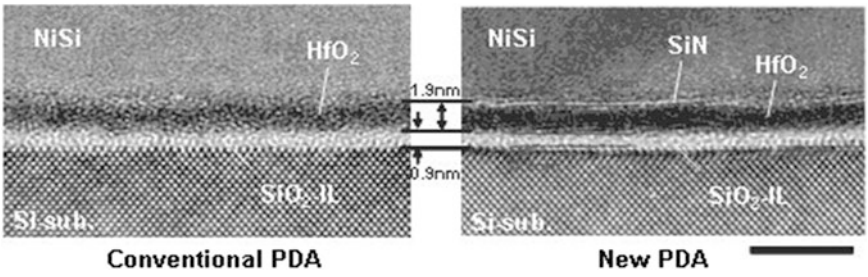


Fig. 4.45 XTEM image with respect to conventional PDA condition (700 °C in N_2) and new condition (1,000 °C in N_2)

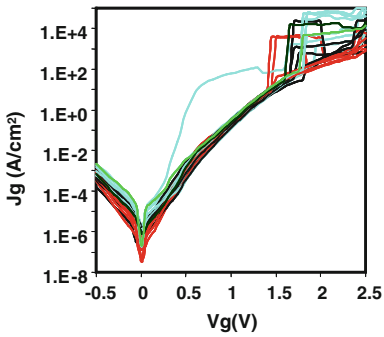


Fig. 4.46 The $J_g - V_g$ characteristics of n-MOS capacitors with Ni-FUSI/SiN/HfO_2 gate stacks after 1,000 °C PDA

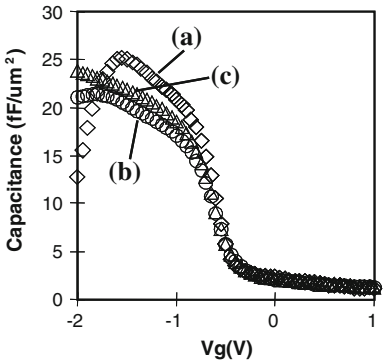


Fig. 4.47 CV of nMOS capacitors with NiSi-FUSI/HfO_2 gate stack: **a** w/o cap (EOT = 1.0 nm), **b** with SiN cap (EOT = 1.3 nm), PDA at 700 °C, while **c** with SiN cap (EOT = 1.2 nm), PDA at 1,000 °C

The EOT calculated from the CV curve of the Ni-FUSI/HfO_2 without the SiN cap is 1.0 nm and as is shown in Fig. 4.47; the EOT of Ni-FUSI/SiN/HfO_2 gate stacks after PDA is 1.3 nm for 700 °C treatment and 1.2 nm for 1,000 °C condition. As can be easily expected, the calculated EOT increases from 1.0 to 1.3 nm due to the SiN capping. The $I_{on} - I_{off}$ plots for the transistors with this elevated temperature PDA condition do not scatter as shown in Fig. 4.48 and the V_{th} has appropriate values. Thus, the $I_{on} - I_{off}$ characteristics of n-MOS transistors with Ni-FUSI/SiN/HfO_2 exhibit dramatic improvement due to the PDA optimization and SiN capping.

The carrier mobility is found to decrease a little by this treatment, however, the EOT is reduced by 0.1 nm (EOT = 1.2 nm). This EOT reduction can compensate the mobility degradation.

As discussed above, owing to these treatments, the gate leakage current is reduced and any splits in the $J_g - V_g$ curves are not observed. One more thing to note is that this wide dispersion can be seen for nFET rather than pFET. This implies that the dispersion is caused not only by the simple interfacial reaction discussed above but also a trap related issue which has polarity dependence. A different behavior might be observed for the Hf-silicate, due to a different charge trapping process [46].

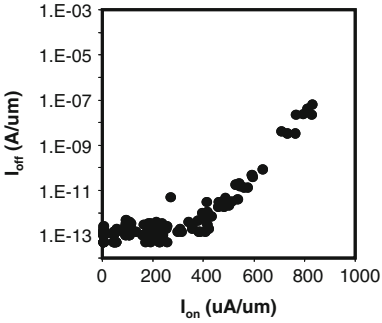


Fig. 4.48 $I_{on} - I_{off}$ characteristics of nMOS transistors (NiSi-FUSI/SiN/HfO_2) treated by the new PDA condition (1,000 °C). Owing to SiN cap and PDA optimizations, higher drivability is obtained

4.4.4.2 Capping Process

Fully silicided (FUSI) gate is one of the promising candidates as a metal gate since this process is cost worthy because it can make use of the poly silicon facility and the work function can be well tuned depending on the degree of silicidation such as NiSi for NMOS and Ni-rich phase for the PMOS [47]. Furthermore, it may not necessarily suffer from the "V_{th} roll-off" [48] seen for the gate first process Sect. 4.4.6.1. However, the shift of flat band voltage is not always enough for a band edge characteristic. In this respect, for the gate first process, a capping process is often introduced by slight doping of the material into the high-k dielectric to obtain the band edge work function with respect to N- and PMOSFET. Mechanism of this work function modulation has been discussed and considered as an electric dipole formation caused by the newly formed bond between Hf and the doping materials such as La for NMOS and Al for PMOS. This electric dipole moment produces a potential to modulate the initial work function. Figure 4.49 indicates the trends of V_{th} shift and areal density of oxygen atoms in various doped oxides [49]. The observed V_{th} shift for each oxide resulted in exactly the same trend as its areal oxygen density. This indicates that the oxygen density determines its direction and strength of the produced dipole as shown in Fig. 4.50. The oxygen moves toward lower density area leaving oxygen vacancy and this movement causes the electrical dipole. (The work function tuning is also discussed in Sect. 4.4.6.1.)

Figure 4.51 indicates a band diagram of the gate electrode (poly Si or metal)/high-k/IL/Si system for the case of the doped capping element. There are several factors to modulate the potential across the gate stack. Oxygen tends to move from high-k to poly Si, leaving the oxygen vacancy together with the electrical dipole at the lower interface which causes roll-off and due to the Fermi level pinning, the potential is modulated at the upper interface.

Also fixed charges in the high-k dielectric modulate the potential in the bulk according to Poisson's equation. In case of the sheet charge distribution, the potential shape has a polygonal line (black solid line) and in case of uniform distribution, the band bends in a quadratic curve as indicated by the red line. Direction of the potential modulation is determined by the polarity of the charge in the high-k. This actual retrograded work function is called "Effective Work

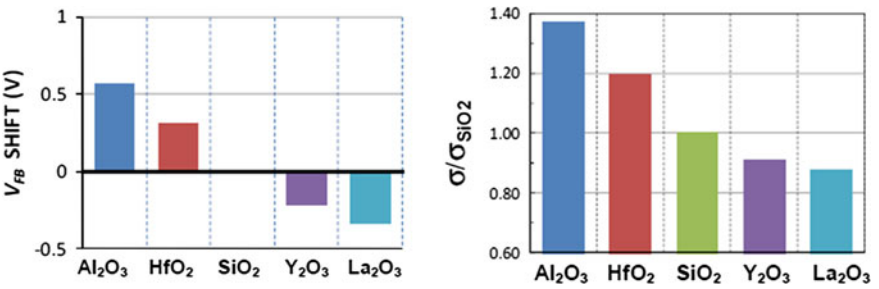


Fig. 4.49 Observed V_{th} shift (*Left*) and areal density of oxygen atoms (*right*) with respect to various oxide

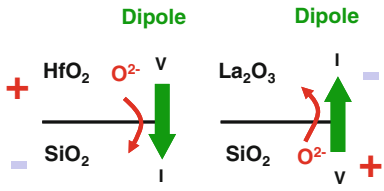


Fig. 4.50 Formation of the electric dipole caused by oxygen transfer

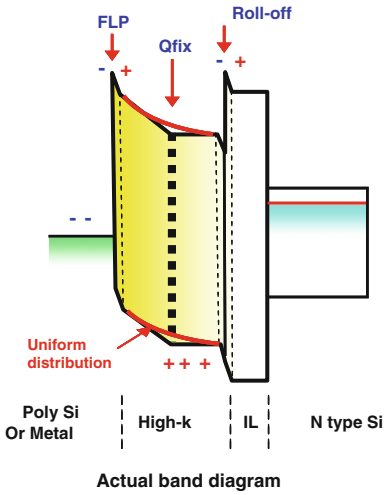


Fig. 4.51 Band diagram of gate/high-k/IL/Si system with capping material

Function (EWF)". How to locate the capping elements and retain them in the high-k dielectric against thermal treatment is a key issue for the "gate first process" and never for the "gate last process". The application of the capping layer for tuning the metal work function has been discussed in Chap. 5; the topic of the interface dipole has been analyzed in Chap. 6.

4.4.5 EOT Dependence on Gate Electrode

Along with the high-k dielectric, the gate electrode material is also a key issue for the high-k gate stack. Thanks to the intensive process development, in order to eliminate the depletion layer, pure metal gate is being introduced as the mainstream gate electrode by solving the various process problems. Study of the pure metal gate material is discussed in Chap. 5. In this section, gate electrode dependence on EOT is covered focusing on the interfacial reaction. And the fully silicided (FUSI) gate is considered as a metal gate electrode.

As is well known, when poly Si is used as gate electrode for Hf-based high-k gate dielectrics, Fermi level of both N- and P-type poly Si is pinned to a lower energy level approx. 0.2 eV below the conduction band edge and approx. 0.55 eV above the valence band edge, respectively, due to the Fermi level pinning (FLP) [50]. As is shown in Fig. 4.52, the pinning site is located more towards the CB side for both types.

For low power CMOS application, the effect of FLP at high-k/poly-Si interface strongly influences the device performance. The increase of threshold voltage due to the FLP leads to lower channel impurity concentration, which brings us better performance such as suppressed GIDL (gate induced drain leakage) as well as higher mobility. In addition, due to the mid-gap Fermi level (in other words, relatively higher threshold voltage) caused by the FLP, the electric field across the gate dielectric is effectively suppressed, hence the gate leakage current may be reduced [51]. By means of progressive application of FLP, the poly-Si/HfSiON gate stack is capable integration into the CMOSFET. This technique leads to suppression of the reverse narrow channel as well as superior cost performance [52].

It is worthwhile to understand the intrinsic phenomenon related to the interface between the gate electrode and the gate dielectric from the EOT-change point of view for metal gate and poly Si gate on the high-k gate dielectric. Figure 4.53 shows the CV curves for various combinations of gate electrode NiSi-FUSI [mono Ni silicide] indicated as "FUSI", poly Si/nitride capping (denoted as "PS/SiN") and gate dielectric (PVD-HfO_2 (denoted as "HfO_2") and SiO_2) for both p- and nMOSFETs. Comparison of poly Si vs NiSi-FUSI/SiO_2 gate dielectric for pMOSFET is considered first. When the poly Si is replaced with NiSi-FUSI, the

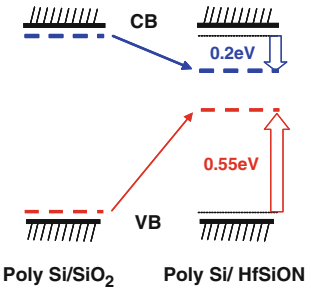


Fig. 4.52 Fermi level pinning occurring at Poly Si/HfSiON

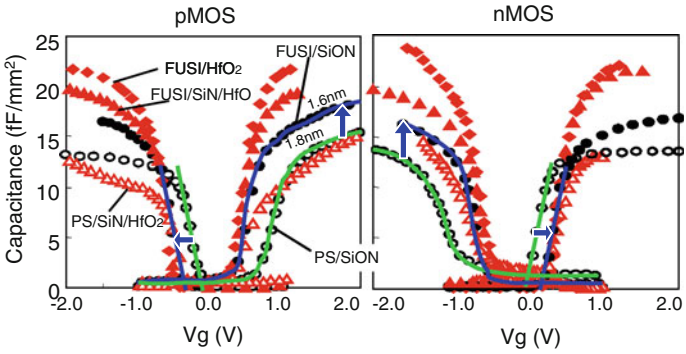


Fig. 4.53 Comparison between Poly Si and NiSi-FUSI on SiON among CV curves for a variety of gate stack: Ni-FUSI, Poly Si/(SiN)/SiON, and HfO_2 gate stack

EOT (under accumulation condition) obtained by Hauser fitting [53] decreases by 0.2 nm (from 1.8 to 1.6 nm) and CET (under inversion condition) by 0.5 nm. The reason is not yet well understood for this slight reduction in EOT which should not happen for both the gate electrodes. Concerning the CET reduction, considering the fact that the normal depletion layer thickness is approximately 0.3 nm, an additional unknown reduction by 0.2 nm seemed to exist under the inversion condition in addition to the depletion. Furthermore, the V_{th} is found to increase for FUSI/SiN/SiON. These results give an indication of the physical thickness reduction or changing of the k value of the SiON dielectric when the FUSI electrode is introduced. Further consideration is necessary to understand this mysterious phenomenon occurring in the PS, FUSI/SiN/SiON system.

On the other hand, changes in EOT and CET for the FUSI/SiN/HfO_2 seem to depend on the Hf concentration, i.e., density of Si–Hf bond at FUSI/SiN/HfO_2 interface. Figure 4.54 is the same as Fig. 4.53, but is focused on the HfO_2 dielectric with/without SiN capping (i.e., FUSI/(SiN)/HfO_2). In the case of SiN capped HfO_2, the EOT of PMOS decreases as well but by 0.5 nm (from 1.8 to 1.3 nm); in contrast, the V_{th} is found to be unchanged, which implies that it is not enough to unpin the Fermi Level Pinning (FLP) by the NiSi-FUSI [54]. For the case without the SiN cap, the EOT is decreased further by 0.7 nm (from 1.8 to 1.1 nm) and the corresponding V_{th} reveals unchanged value indicating NiSi-FUSI is not capable of unpinning the FLP regardless of the SiN capping on top of HfO_2. In case of Ni-rich FUSI electrode, the V_{th} can possibly be changed due to different pinning strength. Further reduction of the EOT by 0.2 nm (from 1.3 to 1.1 nm) is simply due to the absence of the SiN layer. Figure 4.55 indicates the possible phenomena for this "unexpected EOT reduction" [55]. Figure 4.55a shows the case when the interfacial reaction occurred at the upper interface for NiSi-FUSI gate stack. Due to a reaction at upper interface, EOT can be decreased. Figure 4.55b explains the case where an extra capacitance appeared due to a steep band bending caused by the FLP at upper interface for the case of poly Si gate

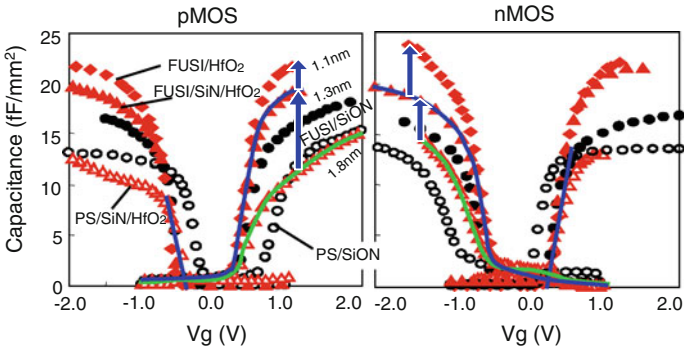


Fig. 4.54 Comparison between Poly Si and NiSi-FUSI on HfO_2 among CV curves for a variety of gate stack: Ni-FUSI, Poly Si/(SiN)/SiON, and HfO_2 gate stack

stack. Due to this additional capacitance, total C_{inv} is reduced, hence the CET increases. This hypothesis is contradictory to the CET decrease. However, this band bending caused by the FLP is likely to happen. These FLP related explanations can explain the unknown small EOT reduction for the NiSi-FUSI/HfO_2. However, these cannot explain the PS,FUSI/SiON stack. The third possible interpretation is the interface roughness as illustrated in Fig. 4.55c. Due to the roughness at the upper interface introduced during FUSI processing, the capacitance is increased, hence the EOT decreases. However, it is still open to question whether the total capacitance will increase by just increasing the roughness of one side of the electrode of the capacitor [56]. In any case, it should be noted that in addition to the poly depletion, an unknown additional reduction in EOT happens when the poly Si is replaced by NiSi-FUSI gate on SiON. Also the EOT reduction is more remarkable for HfO_2 gate dielectric than for SiON [54]. Although, these discussions related to the PMOS, similar trends are found for NMOS.

Apart from the complicated phenomena described above, accurate capacitance measurement including interfaces of the gate stack is necessary to obtain reliable EOT extraction and flat-band voltage shift [57]. Also, care should be taken for the flat-band voltage shift due to ionized dopant charge in the quantization layer [58].

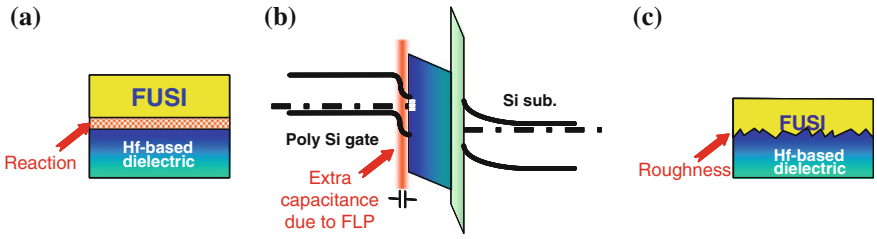


Fig. 4.55 Possible reasons for the unexpected EOT reduction

4.4.6 Influence of Processing on Device Performance

How will the intrinsic solid state property of Hf-based gate dielectric material such as crystallization and interfacial reaction influence the device performance? With aggressive scaling, the average channel electron energy increases and the electric field along the channel becomes stronger with the decreasing channel length. This leads to an increase of thermal noise of the carriers and flicker noise is generated by the trapping/detrapping behavior in the gate dielectric. The trap could be characterized by bias temperature instability (BTI) as well as 1/f and charge pumping (CP) measurement. The number of traps can be suppressed by introducing nitrogen into the gate oxide [59]. Furthermore, the carrier trapping becomes more serious when the high-k gate dielectric material is introduced. Number of the traps in the HfSiON for example is estimated to increase by three- to tenfold compared to those in SiO_2, due to the increase of carrier trap centers introduced by Hf, which is enhanced by its relatively large physical thickness. Besides the process treatment, it is necessary to take measures to suppress the 1/f noise by appropriate layout modification for CMOS with high-k gate dielectric.

Nitrogen in the HfO_2 has a role to prevent its crystallization which leads to large gate leakage and Boron penetration into the channel as well as reliability degradation. It is demonstrated that when nitrogen is introduced into HfO_2, the HfSiON (N = 20 %) obtained stays amorphous and keeps the high permittivity value even after a 1,000 °C anneal [15]. Also, by introducing the HfSiON, device function has been demonstrated with gate length of 50 nm and EOT of 1.2 nm by preventing silicidation at the bottom interface. Thanks to its higher k value with nitrogen, low gate leakage was achieved [60].

4.4.6.1 Work Function Control

From the device performance point of view, the threshold voltage (V_{th}) control is a crucial issue. For the gate first process, it is extremely important to control the oxygen vacancy behavior in the bulk HfO_2. This is because the amount of oxygen (or oxygen vacancy) in the bulk HfO_2 modulates the work function and the oxygen amount is strongly influenced by the CMOS integration process. In order to control the V_{th}, optimization of the channel implantation condition as well as well implantation is a common maneuver. When the channel impurity concentration (N_{ch}) becomes higher, the V_{th} becomes larger; in contrast, short channel effect (SCE) will be a problem when N_{ch} is too low. In case of high-k gate stack, the V_{th} is fixed at a rather higher value due to the "Fermi level pinning" (FLP) [50]. In order to suppress the V_{th} values against the FLP, the N_{ch} should be as low as possible. This can also be achieved by non-doping or counter doping. This treatment, however, requires a reformation of the halo/extension implantation condition which leads to transistor performance degradation. And to make the

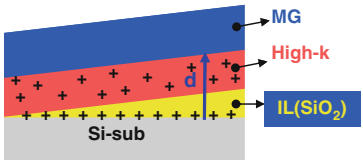


Fig. 4.56 Sample structure by the slant etch technique

matter worse, an immature dry etching during the gate patterning of high-k/metal gate stack makes this problem more complicated.

As discussed in Sect. 4.4.5, the "FLP" leads to flat band voltage shift. Due to the interfacial states at the PS/HfO_2 interface caused by the Hf–Si bond, an electrical dipole is formed at the interface which modulates the V_{fb}. And the strength of the dipole (dipole moment) is expressed by the fixed cation (+Q), anion (−Q) charge and the distance (d) between them. The dipole moment is expressed by the product of each value, "Qd" [61]. As shown in Fig. 4.56, the EOT can be changed intentionally by changing the interfacial oxide layer (IL) thickness, which is often called the slant etch method.

The measured flat band voltage V_{fb} could be expressed as an intrinsic flat band voltage and an additional potential ΔV_{fb} caused by the fixed charges density Q_{fix} in the high-k dielectric, which varies quadratically with the distance from the substrate surface, d as expressed in the following equations where S indicates the area of the corresponding electrode and κ is the permittivity of the dielectric.

$$V_{fb}(\text{measured}) = V_{fb}(\text{intrinsic}) + \Delta V_{fb}(Q_{fix}) \tag{4.2}$$

$$\Delta V_{fb} = dQ_{fix}/S\kappa + (1/2)d^2Q_{fix}/S\kappa \tag{4.3}$$

To obtain the intrinsic V_{fb} in the real device, the slant etch experiment is often introduced. Figure 4.57 shows the experimentally obtained EOT dependence of the measured flat band voltage V_{fb} (measured) with respect to each high-k material. Here, the sheet charge and the uniform charge distribution are assumed to be located at interface and in bulk high-k, respectively. The EOT is changed by changing the IL thickness and the intercept at y axes (EOT = 0 nm) indicates the intrinsic flat band voltage, V_{fb} (intrinsic) [62].

However, the actual V_{fb} obtained is found to deviate from the expected value at extremely small values of EOT which unfortunately corresponds to the EOT range of the gate dielectric of the real device. This undesirable degradation of V_{fb} is called "V_{fb} roll-off". In order to obtain the band edge work function, this roll-off should be suppressed as much as possible. The reason of this V_{fb} roll-off is considered to be due to the fix charges in the high-k bulk and the FLP.

Since the work function (WF) is closely related to V_{fb}, it behaves similarly as V_{fb} in addition to the influence of the FLP. Figure 4.58 illustrates how the WF behaves as EOT changes. The right hand axis corresponds to the empirical EOT dependence of V_{th} with respect to N and PMOS, indicating that −0.2 V corresponds to the VB (Valence Band) edge and +0.2 V corresponds to the CB (Conduction Band) edge [62].

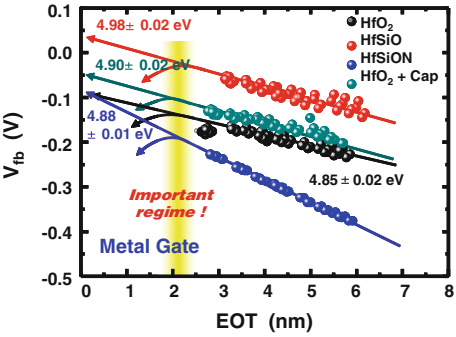


Fig. 4.57 V_{fb}−EOT relation for various high-k gate dielectrics (experimental results)

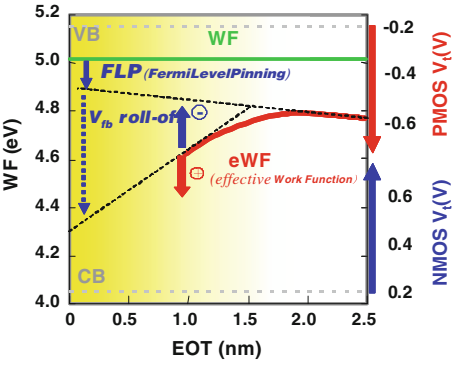


Fig. 4.58 Work function-EOT relation for high-k gate dielectric and the corresponding roll-off behavior together with the V_t trend [62]. "Gate last process" and "Gate first process" are indicated by *green* and *red curves*, respectively

Suppose that the green line indicates the net WF of the MG material, which is maintained as constant if it does not suffer from high temperature thermal treatment as is the case for the "Gate last process" which does not need high temperature thermal treatment. In contrast, for the conventional "Gate first process", due to the interfacial reaction, the Fermi level is pinned and the WF varies linearly as indicated by the upward dotted line, with decreasing EOT. In addition, due to the fixed charge in the high-k dielectric, the WF obtained deteriorates as indicated by the downward dotted line, with decreasing EOT. Consequently, the effective WF obtained results in the V_{fb}-roll-off behavior. When (−) charge is increased in the high-k dielectric, the WF obtained tends to roll-up; in contrast, it tends to roll-off for more (+) charge such as oxygen vacancy, V_o^{2+} in the high-k dielectric. Fermi level pinning and V_{fb} roll-off have been analyzed in Chap. 5. Chapter 2 contains a theoretical treatment of the flat-band and threshold voltages in the case of the high-k gate stack.

In this way, compared to the "Gate last process", the "Gate first process" often brings about undesirable work function degradation and improvement of the work function is extremely important for CMOS integration. The ideal band edge work function seems to be obtained by dual metal and dual dielectric for the "Gate first process"; however, these make the process extremely complicated for CMOS

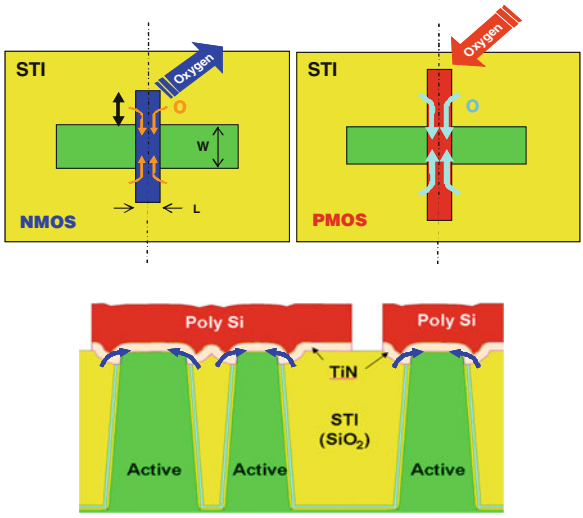


Fig. 4.59 Plan view (*upper side*) of N- and PMOSFETs formed by the "Gate first process" and cross-section (*lower side*) of respective MOSFETs surrounded by STI oxide for Poly Si/TiN/high-k gate stack structure

integration and leads to cost overrun. Thus, integration and V_{th} lowering are completely an antithetical. However, recently, by means of a variety of process ingenuities, the roll-off has been improved for the "Gate first process". For example, thick (thin) TiN electrode provides higher (lower) work function, indicating work function can be tuned to some extent by changing the thickness of a single metal gate material. This type of additional tuning brings about significant advantage to achieve a target band edge work function.

On the basis of the above discussion, the CMOS integration issue is now discussed. It has been reported experimentally that the defect concentration in the gate stack could be modulated by the oxygenation over the gate sidewall after the Source/Drain activation. By introducing a lot of oxygen in the gate dielectric, the work function could be shifted towards the band edge, thus lowering V_{th} for PMOS [63]. This is consistent with the trend described in Fig. 4.58. When the density of oxygen vacancy in the PMOS gate dielectric was decreased, the V_{th} revealed a higher value and vice versa for NMOS. Figure 4.59 shows the plan view (upper part) of the respective N- and PMOS and the cross-section (lower part) of the MOSFETs. Yellow area indicates surrounding STI (Shallow Trench Isolation) which is made from SiO_2. As is discussed above, a lot of oxygen is necessary for PMOS to keep the high WF and less oxygen is necessary for NMOS. From the process point of view, there is a concern that oxygen can invade into the gate from STI through the gate electrode and modulate the WF, hence V_{th}. To maintain the high functionality, effective process countermeasures to prevent invasion of oxygen into the NMOS gate stack and storing of the oxygen inside the PMOS gate stack are imperative for the "Gate first process".

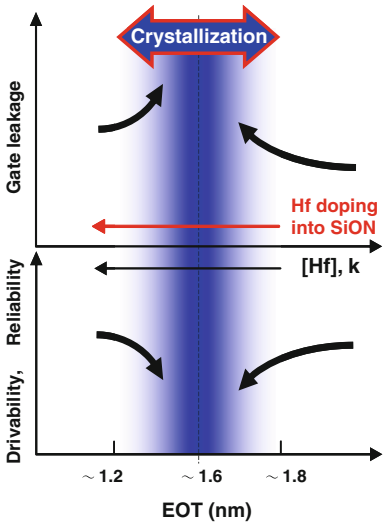


Fig. 4.60 EOT dependence of drivability, reliability, and gate leakage trend

4.4.6.2 Process Window

A more fundamental issue than the CMOS process lies in the realization of the MOSFET with the high-k/metal gate stack. That is the "crystallization" of the high-k layer, without which an MOSFET itself does not work and even more so for the CMOS realization. Rough prognostication based on the above discussion may give an insight into the guidelines for the high-k gate stack device development. Qualitative relation between EOT and key factors of FET such as I_{on}, reliability (Trapping, V_{th} shift, BTI, 1/f), and gate leakage is roughly illustrated in Fig. 4.60 for the gate dielectric system consisting of Hf-based high-k. According to the k-[Hf] trend of Fig. 4.19 in Sect. 4.3.3, EOT increases with decreasing [Hf] and k value. When [Hf] is small, which indicates the case for thick EOT regime, the dielectric property approaches to that of SiON. Each respective phase can be realized by increasing/decreasing the HfO_2 thickness and/or increasing/decreasing the [Hf] in the SiON. By small amount of Hf doping into SiON, i.e., "thinner HfO_2+thicker SiON", the film nature approximates that of SiON; on the other hand, in case of large [Hf], i.e., thinner EOT, the dielectric property comes close to that of HfO_2, which is formed by "thicker HfO_2+thinner SiON". And the regime in between these antithetical phases may induce crystallization. Empirically the crystallization appears at around EOT = 1.6 nm. This crystallization may cause the deterioration of electrical characteristics and this may also induce reliability degradation [64]. On the other hand, contrary to one's expectations, from a precise analysis by EXAFS (Extended X-ray Absorption Fine Structure Spectroscopy), it is suggested that for a 1.4 nm-thick HfO_2 it is possible to have crystalline feature with polymorph transition [65].

As discussed above, the "V_{fb} roll-off" and "Crystallization" are the serious intrinsic problems for the "gate first process" and these temperature process

related problems never happen in the "gate last process". However, the electrical instability of "as deposited high-k dielectric" as well as the profile degradation of extension doping together with the complicated, high cost process may be a serious issues for the "gate last process". Further study is necessary for these pros and cons.

4.5 Summary

Issues of material and processing of Hf-based high-k gate dielectric have been discussed. Different from the covalently-bonded SiO_2 gate dielectric, HfO_2 reveals a complex behavior. The root cause for this is based on its intrinsic solid state property. Due to the ionic-bond in HfO_2, a large number of oxygen vacancies conduce unexpected phenomena and HfO_2 is thermodynamically unstable from the crystallographic point of view such as crystallization and phase separation.

To control these unfavorable phenomena and maintain homogeneity, the material properties of Hf-based high-k gate dielectric and its influence on the electrical characteristics including reliability are analyzed. Several intrinsic key issues such as interfacial control, optimization of the Hf-based high-k gate dielectric formation, breakdown issues as well as bulk trapping phenomena are discussed on the basis of the experimental results.

Crystallization of the Hf-based gate dielectric material, thermal treatment, and interfacial reaction are most important issues in the device fabrication process. Among them, nitrogen has an important role to prevent chemical reaction in the film. For the process control of Hf-based high-k gate dielectrics, oxidation of the Hf-metal has potential for controlling the interfacial layer thickness and EOT scaling. When FUSI gate is introduced, chemical reaction control by nitrogen is effective at the top interface as well as in the bulk. Gate electrode influence on EOT is also an issue analyzed.

From the device performance point of view, these processing/material considerations mentioned above can be interpreted as follows.

For the high-k gate stack with a metal gate electrode including the capping technique and its CMOS processing by the "gate first" process, achievement of the band edge work function rests critically on controlling the number of fixed charges, in other words, control of the oxygen vacancy in the high-k gate dielectric; this will determine the drivability of the high-k transistor and crystallization which will seriously influence the gate leakage performance.

Based on the discussions above, as the guideline for an optimum FET with Hf-based gate dielectric, the medium EOT range where the crystallization tends to happen should be avoided in order to avoid the crystallization of the bulk Hf-based high-k gate stack.

The pros and cons between the "gate first" and the "gate last" process are key issues and are relegated to Chap. 5, where process complexity and cost in addition to performance improvement are considered in detail.

References

1. K. Mistry et al, in *Technical Digest of International Electron Devices Meeting (IEDM)*, Washington, 247, 2007
2. T. Aoyama, *Kougyou Zairyou*, **52**(3), 30 (2004)
3. S. Horii, K. Yamamoto, M. Asai, H. Miya, I. Kaneko, T. Ishihara, S. Hayashi, M. Niwa, Jpn. J. Appl. Phys. **43**, 6963 (2004)
4. K. Yamamoto, M. Asai, S. Horii, H. Miya, M. Niwa, J. Vac. Sci. Technol. **A21**(4), 1033 (2003)
5. M. Lukosius, Ch. Wenger, T. Schroeder, J. Dabrowski, R. Sorge, I. Costina, H.-J. Müssig, S. Pasko, Ch. Lohe, Microelectron. Eng. **84**(2165), 9–10 (2007)
6. A. Toriumi, K. Iwamoto, H. Ota, M. Kadoshima, W. Mizubayashi, T. Nabatame, A. Ogawa, K. Tominaga, T. Horikawa, H. Satake, Microelectron. Eng. **80**, 190 (2005)
7. Y. Harada, M. Niwa, J. Park, D-L. Kwong, in *Extended Abstracts of the 7th Workshop on Formation, Characterization and Reliability of Ultra-thin Silicon Oxides*, Shizuoka, p. 53, 2002
8. K. Yamamoto, S. Kubicek, A. Rothschild, R. Mitsuhashi, W. Deweerd, A. Veloso, M. Jurczak, S. Biesemans, S. De Gendt, S. Wickramanayaka, S. Hayashi, M. Niwa, Microelectron. Eng. **80**, 198 (2005)
9. N. Nagai, H. Hashimoto, T. Matsunobe, Y. Muraji, Y. Ohtsuka, IEICE SDM **101**(107), 1 (2001)
10. M. B. Zahid et al., in *Proceedings of Symposium on VLSI Technology (VLSI)*, 3B.1, Kyoto, 2007
11. M. Niwa, in *Proceedings of Symposium on VLSI Technology (VLSI),* Short course, Key Technologies for 130-100 nm Generations, Honolulu, 28, 2000
12. W.C. Butterman, W.R. Foster, Am. Mineral. **52**, 880 (1967)
13. S. Stemmer, Z. Chen, C.G. Levi, P.S. Lysaght, B. Foran, J.A. Gisby, J.R. Taylor, Jpn. J. Appl. Phys. **42**, 3593 (2003)
14. M. Koyama et al., in *Technical Digest of International Electron Devices Meeting (IEDM)*, Washington, p. 931, 2003
15. M. Koyama et al., in *Technical Digest of International Electron Devices Meeting (IEDM)*, San Francisco, p. 849, 2002
16. S. Inumiya et al., in *Proceedings of Symposium on VLSI Technology (VLSI)*, p. 17, Kyoto, 2003
17. B.O'Sullivan et al., in *Extended Abstracts of the 2007 International Conference on Solid State Devices and Materials,* (SSDM), 372, Tsukuba, 2007
18. J. Morais, L. Miotti, K.P. Bastos, S.R. Teixeira, I.J.R. Baumvol, A.L.P. Rotondaro, J.J. Chambers, M.R. Visokay, L. Colombo, M.C. Matrins Alves, Appl. Phys. Lett. **86**, 212906 (2005)
19. A. Rothchild et al., in *Proceedings of Sympsium on VLSI Technology (VLSI)*, Kyoto, p. 198, 2007
20. D.H. Triyoso, R.I. Hedge, J. Jiang, J.K. Schaefer, M.V. Raymond, IEEE Electron. Device Lett. **29**, 57 (2008)
21. M. Niwa et al., in *Proceeding of the 201st Meeting, MOS Gate Stacks: High-K Dielectrics, Rapid Thermal and other Short Time Processing Technologies III*, Abs. # 713, May, 2002, The Electrochemical Society
22. Y. Harada et al., in *Extended Abstracts of the 8th Workshop on Formation, Characterization and Reliability of Ultrathin Silicon Oxides*, Shizuoka, p. 23, 2003
23. Y. Harada et al., in *Extended Abstracts of the 8th Workshop on Formation, Characterization and Reliability of Ultrathin Silicon Oxides*, Shizuoka, p. 53, 2002
24. Y. Harada et al., in *Proceedings of Symposium on VLSI Technology (VLSI)*, Honolulu, p. 26, 2002

25. R. Degreave, G. Groeseneken, R. Bellens, J.L. Ogier, M. Depas, P. Roussel, H.E. Maes, IEEE Trans. Electron Devices **45**(4), 904 (1998)
26. A. Kerber et al., in *Proceedings of Symposium on VLSI Technology (VLSI)*, Kyoto, p. 159, 2003
27. R. Degraeve et al., in *Proceedings of IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits, IPFA*, 239, 2008
28. L. Manchanda et al., in *Technical Digest of International Electron Devices Meeting (IEDM)*, San Francisco, p. 23, 2000
29. P. J. Chen et al., in *Proceedings of Symposium on VLSI Technology (VLSI)*, Honolulu, p. 192, 2002
30. G.D. Wilk, R.M. Wallace, Appl. Phys. Lett. **74**, 2854 (1999)
31. M. Inoue et al., in *Technical Digest of International Electron Devices Meeting (IEDM)*, Washington, p. 425, 2005
32. R.I. Hegde, D.H. Triyoso, S.B. Samavedam, B.E. White Jr, J. Appl. Phys. **101**, 74113 (2007)
33. D.H. Triyoso, G. Spencer, R.I. Hedge, R. Gregory, X.D. Wang, Appl. Phys. Lett. **92**, 113501 (2008)
34. I. Hirano et al., in *Proceedings of the 46th Annual International Reliability Physics Symposium*, Phoenix, p. 659, 2008
35. A. Kerber, Doctral thesis, Technischen Universität Darmstadt, Fig. 6.20, p. 70, 2003
36. G. Bersuker et al., in *Technical Digest of International Electron Devices Meeting (IEDM)*, San Francisco, p. 791, 2008
37. K. Yamamoto, S. Hayashi, M. Kubota, M. Niwa, Appl. Phys. Lett. **81**, 2053 (2002)
38. S. Hayashi, K. Yamamoto, Y. Harada, K. Eriguchi, M. Kubota, M. Niwa, Appl. Surf. Sci. **216**, 228 (2003)
39. K. Yamamoto et al., in *Extended. Abstract on 2002 International Conference on Solid State Devices and Materials (SSDM'02)*, Nagoya, B-1-3 #20, 2002
40. K. Yamamoto, S. Hayashi, M. Niwa, M. Asai, S. Horii, H. Miya, Appl. Phys. Lett. **83**(11), 2229 (2003)
41. S. Horii et al., in *Extended. Abstract on 2002 International Conference on Solid State Devices and Materials (SSDM'02)*, Nagoya, p. 172, 2002
42. M. Hiratani et al., in *Extended. Abstract on 2003 International Conference on Solid State Devices and Materials (SSDM'03)*, Tokyo, p. 700, 2003
43. S. Hayashi et al., in *Proceedings of the 68th Symposium on Semiconductors and Integrated Circuits Technology*, Kyoto, p. 29, 2005
44. R. Mitsuhashi, K. Yamamoto, S. Hayashi, A. Rothschild, S. Kubicek, A. Veloso, S. Van Elshocht, M. Jurczak, S. De Gendt, S. Biesemans, M. Niwa, Microelectron. Eng. **80**, 7 (2005)
45. E. Cartier et al., in *Proceedings of Symposium on VLSI Technology (VLSI)*, Kyoto, p. 230, 2005
46. E. P. Gusev et al., in *Technical Digest of International Electron Devices Meeting (IEDM)*, San Francisco, p. 79, 2004
47. T. Hoffman et al., in *Technical Digest of International Electron Devices Meeting (IEDM)*, San Francisco, p. 269, 2006
48. B.H. Lee, J. Oh, H.H. Tseng, R. Jammy, H. Huff, Mater. Today, **9**, 32–40 (2006)
49. K. Kita et al., in *Technical Digest of International Electron Devices Meeting (IEDM)*, San Francisco, p. 29, 2008
50. C. Hobbs et al., in *Proceedings of Symposium on VLSI Technology (VLSI)*, Kyoto, p. 9, 2003
51. Y. Shimamoto et al., in *Proceedings of Symposium on VLSI Technology (VLSI)*, Kyoto, p. 132, 2005
52. G. Tsutsui et al., in *Proceedings of Symposium on VLSI Technology (VLSI)*, Kyoto, p. 176, 2007
53. J.R. Hauser, K. Ahmed, in *Characterization and Metrology for VLSI Tech*nology, pp. 235–239, 1998
54. M. Niwa, R. Mitsuhashi, K. Yamamoto, S. Hayashi, A. Rothchild, S. Kubicek, S. De Gendt, S. Biesemans, in *NATO Science Series II: Mathematics, Physics and Chemistry, Defects in*

High-k Gate Dielectric Stacks, Nano-Electronic Semiconductor Devices, ed. by E. Gusev, vol. 220 (Springer Netherlands), pp. 1–15, 2005
55. M. Niwa, R. Mitsuhashi, K. Yamamoto, S. Hayashi, Y. Harada, M. Kubota, A. Rothchild, T. Hoffmann, S. Kubicek, S. De Gendt, M. Heyns, S. Biesemans, Prospect of Hf-based gate dielectrics by PVD with FUSI gate for LSTP Application. in *Physics and Technology of High-k Gate Dielectrics III*, ed. by S. Kar, S. De Gendt, M. Houssa, D. Landheer, D. Misra, W. Tsai, ECS Transactions, vol. 1, no. 5 (The Electrochemical Society, Pennington, New Jersey, 2006)
56. Private communication with Prof. T. P. Ma, July, 2005
57. C.A. Richter, A.R. Hefner, E.M. Vogel, EDL **22**(1), 35 (2001)
58. N. Yasuda et al., in *Extended. Abstract on 2003 International Conference on Solid State Devices and Materials (SSDM'05)*, Kobe, A-5-4 p. 250, 2005
59. T. Ohguro et al., in *Proceedings of Symposium on VLSI Technology (VLSI)*, Kyoto, p. 91, 2001
60. M. Koike et al., in *Technical Digest of International Electron Devices Meeting (IEDM)*, Washington, p. 103, 2003
61. P. Sivasubramani et al., in *Proceedings of Symposium on VLSI Technology (VLSI)*, Kyoto, p. 68, 2007
62. M. Niwa, Short Course lecture, "Gate Stack Technology –From thin gate oxide to High-k gate dielectric", From Basic Theory to Newest Application in MOS devices, 2009 International Conference on Solid State Devices and Materials (SSDM'09), Sendai, Oct. 6, 2009
63. E. Cartier et al., in *Proceedings of Symposium on VLSI Technology (VLSI)*, Kyoto, p. 42, 2009
64. Private communication with Y. Harada, Sept. 2007
65. P.S. Lysaght, J.C. Woicik, M.A. Sahiner, B.-H. Lee, R. Jammy, Appl. Phys. Lett. **91**, 122910 (2007)

Chapter 5
Metal Gate Electrodes

Jamie K. Schaeffer

Abstract Metal gate electrodes are a vital enabler for the use of high-k gate dielectrics in advanced complimentary metal oxide semiconductor (CMOS) technology. This chapter will detail how metal gate electrodes were selected over poly-Si electrodes to be used in conjunction with high-k gate dielectrics, how metal gate electrodes are an important component of device scaling, and how the of the metal gate can be tuned to optimize the device performance. In this chapter you will see how the properties of the metal gate are inseparable from those of the high-k gate dielectric and how the metal gate can be manipulated to intentionally influence the properties of the dielectric material to produce a desired device response. This chapter will discuss the different ways in which metal gates can be used to change the effective work function of a MOS transistor, including the vacuum work function, dielectric/metal capping layers, and oxygen vacancy manipulation. In each of these cases there is a clear interplay between the metal gate and the gate dielectric and it is important to understand both materials systems in order to understand the device response. Finally, this chapter will detail the common approaches of integrating metal gate electrodes in a complimentary metal oxide semiconductor. Because there are various methods of modulating the effective work functions by using metal gate electrodes, there are multiple integration schemes that have been devised in order to produce optimized nMOS and pMOS transistors.

5.1 Reasons for Using Metal Gate Electrodes

In almost all instances metal gate electrodes are being introduced simultaneously with high-k gate dielectrics. There are two primary reasons for this. The first is that metal gate electrodes eliminate poly-silicon depletion and help with MOSFET

J. K. Schaeffer (✉)
GLOBALFOUNDRIES, Dresden 01109, Germany
e-mail: jamie.schaeffer@globalfoundries.com

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5_5,

(Metal Oxide Semiconductor Field Effect Transistor) scaling. The second became apparent when researches first attempted to introduce high-k gate dielectrics with poly-Si electrodes, and noticed an incompatibility between high-k based gate dielectrics and poly-silicon electrodes that results in the pinning of the work functions of n+ and p+ poly-silicon to undesirable values.

5.1.1 Elimination of Gate Depletion

Since increasing the gate capacitance is required for increasing MOSFET drive current, device scaling mandates a consideration of factors other than the gate oxide that contribute to the inversion capacitance. When a MOSFET is operated in inversion there are actually two additional capacitances in series with the oxide capacitance. Taking these factors into consideration the total capacitance in inversion, C_{inv}, is given by

$$\frac{1}{C_{inv}} = \frac{1}{C_{sub}} + \frac{1}{C_i} + \frac{1}{C_{poly}}$$

where C_i is the aforementioned oxide capacitance, C_{sub} is a capacitance due to quantum mechanical effects which force the centroid of inversion charge in the substrate to be a few Ångstroms away from the Si/SiO_2 interface, and C_{poly} is due to a gradual potential drop, or band bending, [1] in the poly-Si gate electrode (Fig. 5.1).

The C_{poly} contribution, which is frequently referred to as poly-silicon depletion, is a result of the sum of inversion and depletion charges in the substrate being greater than the impurity density (N_{poly}) near the poly-Si—oxide interface. Since charge neutrality requires that the field lines for every carrier in the substrate are screened by the ionized impurities in the electrode, the field lines from the substrate penetrate a finite distance into the poly-Si electrode before encountering enough charges to cancel the sum of their electric fields. The screening of charge over a finite distance into the poly-Si is the origin of the gradual potential drop and band bending in the poly-Si electrode.

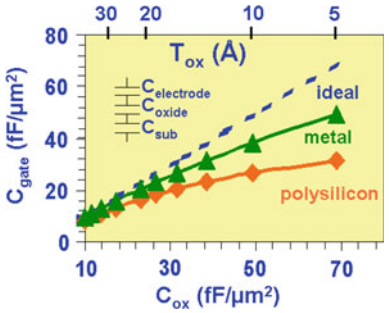


Fig. 5.1 The total gate inversion capacitance (C_{gate}) versus the oxide capacitance (C_{ox}) for metal gate electrodes and for poly-silicon electrodes. Higher C_{gate} with metal gate electrodes is due to the absence of gate-depletion effects

To overcome the additional inversion capacitance, poly-Si electrodes can be replaced with metal electrodes. Since metals have a shorter Debye length than poly-silicon, meaning a metal has a higher carrier concentration and is more effective at screening charge, no band bending occurs with metal electrodes. This eliminates the depletion capacitance and assists device performance.

5.1.2 Incompatibility of High-k Materials with Poly-Si Gates

Although there were many attempts to first introduce high-k with poly-Si electrodes, this encountered multiple challenges. These challenges include nFET and pFET threshold voltages that were higher than were expected based on n+ and p+ doped poly with SiO_2 electrodes and the observance of some physical reactions that occurred at the high-k/Si interface.

An experiment by Hobbs et al. [2], was one of the first experiments that investigated the high threshold voltages observed in poly-Si gated HfO_2 gate dielectrics. In this experiment, Hobbs et al. deposited sub-monolayers of HfO_2 on the surface of a 23 Å thermally grown SiO_2 gate dielectric (Fig. 5.2).

Atomic layer deposition (ALD) was used to deposit the sub-monolayers of increasing surface coverage of HfO_2. When analyzing poly-Si/HfO_2/n-type Si capacitors Hobbs et al. reported monotonic shifts in the flat band voltage electrodes that encouraged more researchers to evaluate flat band voltage of poly-Si with increasing HfO_2 surface coverage (Fig. 5.3). Compared to a SiO_2/Poly-Si system, flat band voltage shifts are observed with only one cycle of ALD HfO_2, and the shifts saturate at approximately 20 cycles, or when complete surface coverage is obtained.

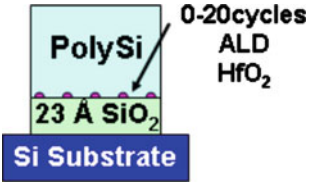


Fig. 5.2 Schematic showing the experimental technique used by Hobbs et al. to study the Fermi level pinning of poly-silicon electrodes on HfO_2

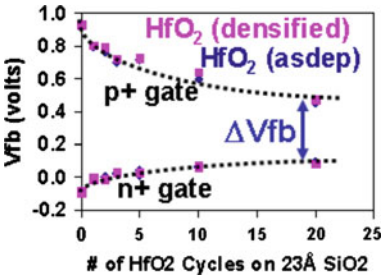


Fig. 5.3 V_{fb} shifts as a function of ALD precursor cycles for HfO_2 growth. The shifts tend to saturate when complete HfO_2 surface is obtained [2]

High-k dielectrics also had many undesirable physical interactions with poly-Si electrodes that encouraged more researchers to evaluate metal gate electrodes. The problems were first noticed when ZrO_2 was studied in conjunction with poly-Si electrodes. Large $ZrSi_x$ nodules would form at the ZrO_2/poly-Si interface demonstrating an incompatibility at the ZrO_2/poly-Si interface [3]. The gross $ZrSi_x$ formation at this interface is the primary reason HfO_2 became the preferred gate dielectric material even though ZrO_2 and HfO_2 are chemically very similar materials.

5.2 Work Function Considerations for Metal Gate Electrodes

This section focuses on the desired gate work function characteristics for metal gate electrodes and various methods of achieving the desired effective work function.

5.2.1 Work Function Requirements of Devices

For bulk CMOS devices, simulations indicate that the desired work function for NMOS (PMOS) electrodes is near the conduction (valence) band edge of silicon [4, 5]. These simulation studies imposed the constraint of a constant off current (I_{off}) for a sub-nominal length device. This constraint means that if the metal work function is shifted towards the middle of the Si band gap, the substrate doping has to be reduced to maintain a constant I_{off}. Reducing the substrate doping can benefit performance by improving carrier mobility. However, reducing the substrate doping too far also degrades the short channel performance of a MOSFET by lowering the potential barrier and increasing the depletion width between the source and the channel regions. As the gate length is reduced electric fields from the drain contribute to lowering this potential barrier. This phenomenon is referred to as drain induced barrier lowering (DIBL). DIBL manifests itself as a degraded sub-threshold swing, a measure of how rapidly the drain current responds to the gate bias. A degraded sub-threshold swing results in a lower inversion charge density for the same gate voltage which degrades the current drive of the MOSFET in saturation ($I_{d,sat}$). This suggests that work functions closer to $\sim$4.1 eV ($\sim$5.2 eV) near the conduction (valence) band edge of silicon are required to optimize NMOS (PMOS) device performance for short channel MOSFET devices.

It should also be mentioned that work function engineering provides an alternative means of achieving multiple threshold voltages on a semiconductor chip. Traditionally with poly-Si, the effective work function could not be engineered and therefore the only way to achieve multiple threshold voltages V_t was to change the

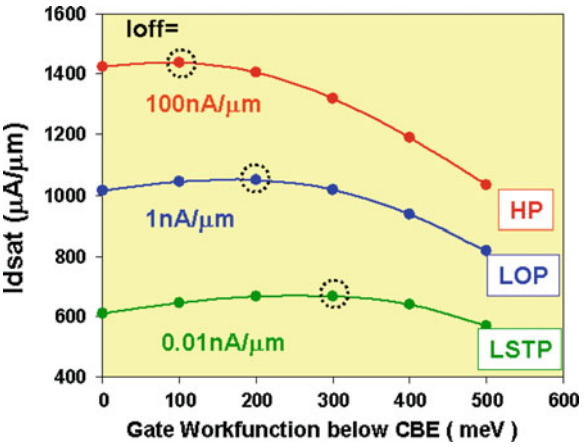


Fig. 5.4 Simulation of I_{dsat} as a function of nMOS gate work function relative to the silicon conduction band for different I_{off} targets. A maximum drive-current is achieved by using an off-band-edge metal gate to increase device V_t, and then lowering the substrate doping to increase mobility [5]

gate length or to increase the substrate well doping. One benefit of using metal gate electrodes is that a high-V_t device can be achieved by moving the metal work function towards mid-gap without increasing the substrate doping. This results in an inherent mobility and ultimately performance advantage for the devices in the circuit that utilize higher threshold voltages (Fig. 5.4).

5.2.2 *Methods of Achieving Desired Effective Work Function*

To achieve the appropriate threshold voltages in the devices, researchers are exploiting various methods broadly termed "work function engineering". Along with manipulating the metal work function to shift threshold voltages, there has also been a significant amount of work dedicated to modulating threshold voltages via interface dipoles and dielectric fixed charges. This section will review the factors that impact effective work function and also discuss the various methods that are currently being utilized to achieve this.

Accurate extraction of effective work function and dielectric fixed charge is the subject of multiple papers [6, 7]. For conventional SiO_2 gate dielectrics, the effective work function can be extracted from the y-intercept of plots of flat band voltage (V_{fb}) versus effective oxide thickness (EOT). The basic equation governing the electrostatics in a MOS capacitor is:

$$V_{fb} = \phi_{m,eff} - \phi_s - \frac{\int_0^{EOT} x\rho dx}{\varepsilon_{SiO_2}}$$

for a typical SiO_2-based MOSFET with negligible contributions from the bulk charges this is equivalent to:

$$V_{fb} = \phi_{m,eff} - \phi_s - \frac{Q_2 EOT}{\varepsilon_{SiO_2}} \tag{5.1}$$

where $\phi_{m,eff}$ = the effective metal work function, ϕ_s is the Fermi level of the silicon, $\rho(x)$ is the bulk charge distribution in the dielectric, Q_2 is the fixed charge at the Si/dielectric interface, and ε_{siO_2} is the permittivity of SiO_2.

For HfO_2 films deposited on silicon a ~1 nm SiO_2-like interface layer typically exists between the silicon substrate and the HfO_2. Therefore, (5.1) does not adequately account for the contributions of the fixed charges at the SiO_2/HfO_2 interface, nor does it capture the fact that HfO_2 can have a rather significant bulk fixed charge value, something that is often neglected for SiO_2 dielectrics. A more rigorous equation for the extraction of fixed charges in a high-k gate stack is given by

$$V_{fb} = \phi_{m,eff} - \phi_s - \frac{Q_1 EOT_1}{\varepsilon_{SiO_2}} - \frac{Q_2 EOT}{\varepsilon_{SiO_2}} - 1/2 \frac{\varepsilon_{HfO_2}}{\varepsilon_{SiO_2}} \frac{\rho_i (EOT_1)^2}{\varepsilon_{SiO_2}} \tag{5.2}$$

where Q_1 is the fixed charge at the SiO_2/HfO_2 interface, EOT_1 is the effective oxide thickness of the HfO_2 layer, EOT_2 is the effective oxide thickness of the SiO_2 layer, ρ_1 is the HfO_2 bulk charge density, ε_{HfO_2} is the permittivity of HfO_2, and $EOT = EOT_1 + EOT_2$ (Fig. 5.5).

This equation now illustrates some of the various layers in the gate stack that can be manipulated in order to shift threshold voltages to the desired device requirement. The next sub-sections will provide more detail into ways the flat-band voltage, or threshold voltage, can be controlled via the effective work function of the gate material or the fixed charge at various layers in the dielectric gate stack.

5.2.2.1 Vacuum Work Function

The discussion of effective work function begins with a quick review of the work functions of the elements [8]. Vacuum work functions, $\phi_{m,vac}$, show a periodicity between work function and atomic number. Work functions increase from left to right across a row on the periodic table (Fig. 5.6). This indicates that metals with work functions above the conduction band edge of Si (<4.1 eV) are typically in the first three columns of the periodic table. Alternatively, metals with work functions below the valence band edge of Si (>5.2 eV) tend to be late transition

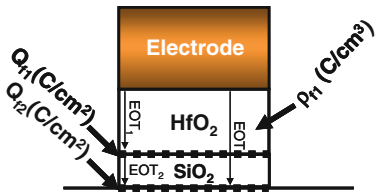


Fig. 5.5 Image showing interface fixed charges at the Si/SiO_2 and SiO_2/HfO_2 interfaces and bulk fixed charges inside the HfO_2 gate dielectric

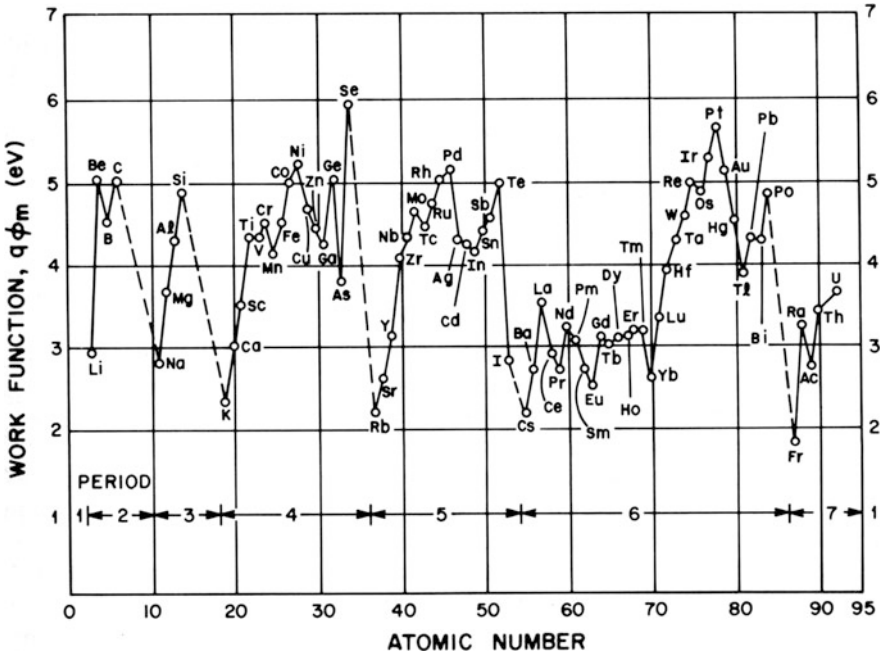


Fig. 5.6 Plot of work function versus atomic number. Work functions are generally observed to increase moving across a row of transition metal elements [20]

metals. In particular, the platinum group metals such as Pt, Ir, Os, Au, Ni, Ru, Pd, and Rh have high work functions. However, the vast majority of the transition metals have work functions that exist within the band edges of silicon (4.1 eV $< \phi_m <$ 5.2 eV) making it difficult to find metals that can easily replace p+ and n+ poly-silicon gates.

The limited number of metal gate electrode candidates with vacuum work functions between the conduction band edges of silicon is restricted even further by Fermi level pinning. Fermi level pinning is a consequence of forming an interface between a metal and a dielectric (or semiconductor). When an interface is formed, the effective metal work function becomes "pinned" at a different energy than its vacuum work function. This results from interfacial charge exchange between the metal Fermi level and gap states at the metal-dielectric interface causing it to shift with respect to its unpinned location. A recent comprehensive review of Schottky barrier concepts has been published [9] and Fermi level pinning models have been extensively tested on a broad class of interfaces [10–13], including for metal gate electrodes on HfO_2 [14, 15].

The most widely accepted of the Fermi level pinning models is the metal induced gap states (MIGS) model [10]. The origin of the gap states in the MIGS model is from the dangling bonds of under-coordinated surface atoms. These dangling bonds produce surface states that are dispersed in continuum at energies

throughout the band gap of the dielectric. The electron wave function for surface states is given by $\psi = u(r)exp(ikr)exp(-ik\perp z)$[16] (k is the wave vector, r, z are position vectors, and u is a periodic function in r). The two components of this equation are related to electron propagation in the plane of the surface and electron propagation normal to the surface. For electrons traveling parallel to the surface the wave vector $k_{\|}$ is real, and $\psi = u(r)exp\left(ik_{\|}r\right)$. However, for electron propagation normal to the surface $k_{\perp}$ is complex. This is because the $E - V(z)$ term under the square root in the expression $k_{\perp} = \left(2m\left[E - V(z)/\hbar^2\right)^{1/2}\right.$ is negative. This term is negative due to the potential energy $V(z)$ being greater than the electron energy E when an electron tunnels outside the crystal or when an electron tunnels into the crystal bulk. Therefore, the $exp(-ik\perp z)$ term results in an exponential decay of the electron wave function for directions normal to the surface. These are often referred to as evanescent states. Similar to a surface state on a dielectric, a cleaved metal surface also has a surface wave function that decays exponentially into vacuum. In the MIGS model, when a metal is placed in contact with a dielectric, the metal surface states induce the gap states in the dielectric. This results in interfacial charge exchange between the metal and the dielectric gap states causing the metal Fermi level to shift with respect to its unpinned location towards a characteristic energy level in the semiconductor. In the MIGS model this characteristic energy is referred to as the charge neutrality level $(\phi_{CNL,d})$. The charge neutrality level is the location of the highest occupied surface state in the dielectric band gap. The charge neutrality level has been described as the position where the surface states change from acceptor-like to donor-like character. Therefore, when the two surfaces are brought into contact, charge is exchanged between the metal and the dielectric surface states resulting in the formation of an interfacial dipole. The magnitude of the interfacial dipole depends on the pinning strength of the semiconductor which is defined by the value of the Schottky pinning parameter (S). The barrier height between a metal Fermi level and the dielectric conduction band depends on the pinning parameter and is given by $\Phi_b = S\left(\phi_{m,vac} - \Phi_{CNL,d}\right) + \left(\Phi_{CNL,d} - \chi_d\right)$ [17] (χ_d is the electron affinity of the dielectric). In the Schottky, or weak pinning, limit $S = 1$, while in the Bardeen, or strong pinning, limit $S = 0$ (Fig. 5.7). The magnitude of Fermi level pinning in a MOS capacitor can also be measured in terms of an effective metal work function $\left(\phi_{m,eff}\right)$, as opposed to barrier height shifts. The effective work function is related to the vacuum work function $\left(\phi_{m,vac}\right)$ by $\phi_{m,eff} = \phi_{CNL,d} + S\left(\phi_{m,vac} - \phi_{CNL,d}\right)$ [14]. It should be noted that the gap states in the MIGS model are intrinsic to any metal-dielectric interface.

The pinning parameter (S) is the slope obtained on plots of barrier height (ϕ_b) versus the vacuum work function $(\phi_{m,vac})$ for a given dielectric or semiconductor, $S = d\phi_b/d\phi_{m,vac}$. This relationship was developed for Schottky contacts, but in MOS capacitor structures effective work function $(\phi_{m,eff})$ should be substituted for ϕ_b since the metal Fermi level is being referenced with respect to the silicon Fermi level instead of the semiconductor (or dielectric) conduction band as in studies of

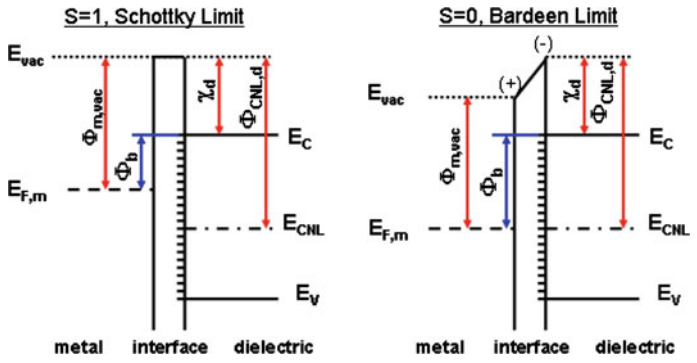


Fig. 5.7 Schematic of Fermi level pinning in the Schottky or weak-pinning limit and in the Bardeen or strong-pinning limit

Schottky contacts. The pinning parameter $S{=}d\phi_{m,eff}/d\phi_{m,vac}$ is identical to $S{=}d\phi_b/d\phi_{m,vac}$ if it is assumed that any internal dipoles that might exists at the SiO_2/HfO_2 or the Si/SiO_2 interfaces as well as any internal fields in the dielectric are independent of the electrode work function. Plots of $\phi_{m,eff}$ and ϕ_b versus $\phi_{m,vac}$ will therefore only differ by a parallel shift of the data, but the slope, or pinning parameter, remains unchanged.

An alternative approach is to express the slope parameter in terms of metal electronegativity (X) instead of vacuum work function, where $S' = d\phi_{m,eff}/dX$ [18, 19]. This variation on the pinning parameter arises because the Pauling electronegativity has been correlated to the vacuum work function of elemental metals by the expression $\phi_{m,vac} = 2.27X + 0.34$ [20, 21]. When this relationship is applied to elemental metals, the Schottky limit occurs on plots of X versus $\phi_{m,eff}$ when $S' \sim 2.27S$ [22, 23].

The relationship between group electronegativity and barrier height or effective work function has been extended to multi-element metal gate electrodes in contact with HfO_2. A plot of effective work function versus the geometric mean of electronegativity for a diverse set of electrode materials on HfO_2 produces a reasonable linear fit (see Fig. 5.8). This relationship provides a useful means of predicting effective work functions of either elemental or alloyed metals.

It should be noted that the sub-lattice elements in certain metal gate compounds can be a large factor in the calculation of the mean electronegativity. A rather sizeable electronegativity difference of $\Delta X_{\text{B to O}} = 1.5$ exists across four common sub-lattice elements (B, C, N, and O) that appear sequentially in the periodic table. For comparison, the electronegativity range across 30 transition metals is only $\Delta X_{\text{La to Au}} = 1.44$. Because of this fact, it is not surprising that low effective work functions have been obtained with LaB_6 and TaC, that interstitial carbide materials have consistently lower work functions than their analogous interstitial nitrides, and that conductive oxides, such as iridium oxide (IrO_2) and ruthenium dioxide

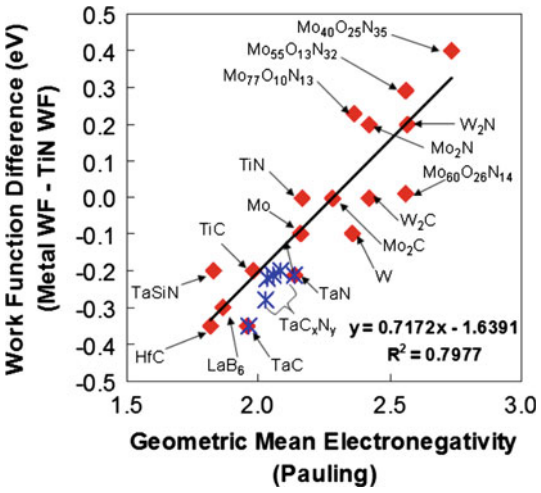


Fig. 5.8 Plot of the mean electronegativity of the constituent elements in metal gate electrodes versus the effective work function relative to TiN on HfO_2 gate dielectrics. All films were exposed to >900 °C anneal temperatures [65]

(RuO_2) have higher effective work functions than Ir or Ru. It should be noted that the correlation is not perfect, indicating that factors other than the geometric mean electronegativity likely contribute to the effective work function of metals on HfO_2. Some of these factors include the fact that the volume fraction of elements in the electrode may not be representative of the actual interface bonding, and differences in crystal phase and preferred orientation may contribute to some of the variability as well.

5.2.2.2 Manipulation of Fixed Charges/Interface Dipoles

Dielectric Capping Layers

Some promising techniques for engineering the metal work function towards band-edge involve the use of dielectric capping layers such as MgO or La_2O_3 for nFETs [24] and Al_2O_3 for pFETs [25], or by incorporating the La, Mg or Al into the metal electrode, instead of as an oxide capping layer. Both approaches work well for modulating the effective work function or threshold voltages (V_t) of MOS devices, but alloying the metal directly into the gate electrode can be advantageous in terms of effective oxide thickness scaling.

Capping Layers for NFET Electrodes

There has been a significant amount of work discussing the use of column IIA, IIIB, and rare earth elements for "tuning" the nFET threshold voltages [24, 26, 27]. The typical approach is to deposit oxides of these materials 1–10 Å thick between the high-k gate dielectric and the metal gate electrode in a gate first integration scheme. These capping layers produce negative V_t shifts, and reduce the threshold voltage of nFET device by up to 400 mV (Fig. 5.9). In some instances dielectric capping layers can be used for V_t shifts with slight reduction or

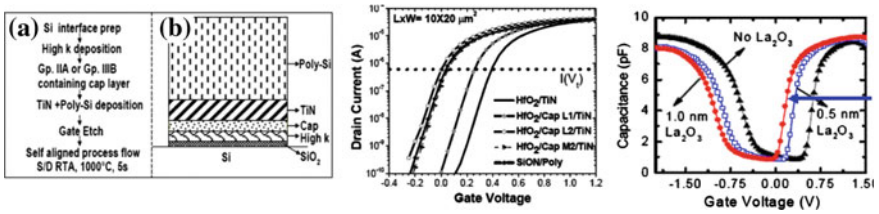


Fig. 5.9 Dielectric capping layer approach showing well behaved $I_d - V_g$ [24] and $C_g - V_g$ [75] curves

minimal increases in the effective oxide thickness of the gate dielectric. However, increasing the dielectric cap thickness too much will eventually result in increases in the effective oxide thickness. Well behaved $I_d - V_g$ (drain current–gate voltage) and $C_g - V_g$ (gate capacitance–gate voltage) curves showing monotonic shifts in the threshold voltage and no degradation in the sub-threshold slope are observed with this approach (Fig. 5.9).

To avoid the effective oxide thickness penalty associated with using an oxide capping layer, the V_t shift element can be incorporated directly into the electrode instead. One example of this is an experiment where Mg was incorporated directly into a TaMgC inter-layer between the gate dielectric and a TaC capping layer. To simultaneously decrease the NMOS threshold voltage and the effective oxide thickness, the thickness of the TaMgC inter-layer can be gradually increased or the Mg concentration in the TaMgC inter-layer can be gradually increased (Fig 5.10). Both approaches show V_{fb} shifts comparable to those achieved with MgO capping layers [28].

As will be explained in more detail later, the V_t shift mechanism for both the oxide capping and the alloyed electrode approach depends on the controlled diffusion of the capping elements to the SiO_2/HfO_2 interface where they react with SiO_2 at this interface to produce an interface dipole or positive fixed charge which results in the favorable reduction of the nFET threshold voltage. One potential negative side-effect of achieving threshold voltage reductions in this manner is that the creation of dipoles or fixed charges at this interface can result in the degradation of the mobility, sub-threshold slope, or gate leakage if the dielectric capping layer is too thick [24]. For device performance optimization, careful consideration needs to be given to the trade-offs between mobility, inversion capacitance equivalent thickness T_{inv}, and short channel control [29].

Capping Layers for PFET Electrodes

For pFETs, additional V_t shifts to help satisfy the requirements of high performance devices can be achieved by using Al_2O_3 capping layers or by alloying Al into the metal gate electrode. The impact of Al_2O_3 capping layers has been well documented [30–33] and shown to produce positive V_t shifts of up to $\sim$200 mV that are desirable for reducing the pFET threshold voltage (Fig. 5.11). The PMOS work function tuning situation is analogous to the NMOS one in that V_t lowering

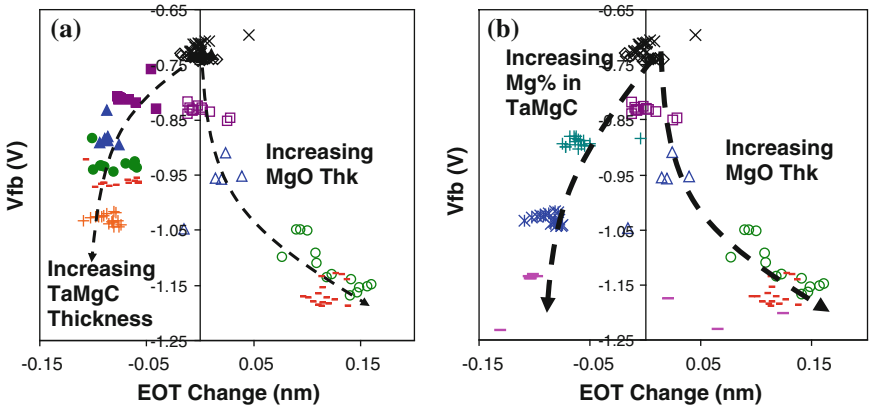


Fig. 5.10 V_{fb}/EOT shifts for increasing TaMgC thickness (**a**) and for increasing at percentage Mg in TaMgC (**b**) with V_{fb}/EOT for increasing MgO cap layer thickness as a reference

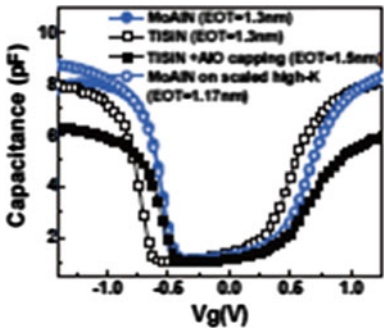


Fig. 5.11 Capacitance–Gate Voltage curves showing *positive* V_t shifts compared to reference devices by using MoAlN. MoAlN does not have any T_{inv} increase which is an added benefit over AlO capping layers [34]

can be achieved by either using Al_2O_3 capping or by alloying Al into a metal gate electrode.

For the alloyed metal electrode approach, MoAlN has been regularly evaluated [28]. Similar to the results achieved with nFET dopant species, comparisons of the Mo_2N/Al_2O_3 and MoAlN approaches indicate there is less effective oxide thickness increase with the alloyed electrode approach [34]. Unfortunately, increasing the total dose of Al in the MoAlN layer, either by increasing the MoAlN layer thickness or by increasing the Al concentration can result in a degradation of the device mobility. Tuning the threshold voltage of MOSFET devices with a MoAlN electrode requires careful engineering control to avoid some of the device degradations that occur when the Al dose is too large.

Mechanism for Threshold Voltage Shifts

Although there remains some debate as to whether the mechanism for the V_t, EOT, mobility, sub-threshold slope, and leakage changes is caused by internal dipoles, fixed charges, or modulation of dielectric oxygen vacancy concentrations the experimental results do suggest that the V_t shift elements must be located at the

SiO_2/high-k interface in order to produce the desired threshold voltage shifts. Some of the most compelling data elucidating this phenomenon was from the work that placed an Al_2O_3 inter-layer either at the SiO_2/HfO_2 interface or at the HfO_2/electrode interface [25]. This work indicated that large V_t shifts are achieved even with no thermal budget applied to the gate stack when the Al_2O_3 layer is placed at the SiO_2/high-k interface. On the other hand, when the Al_2O_3 is placed at the high-k/electrode interface, high temperature annealing is required in order to diffuse the Al towards the SiO_2/high-k interface where it produces the desired threshold voltage shift. This is a very critical result that suggests that without a significant thermal budget applied to the gate stack, the capping oxides must be placed at the SiO_2/HfO_2 interface in order to produce the desired V_t shifts. However, V_t shifts can also be produced if a thermal budget (>900 °C) is applied to "drive-in" the dopant species through the high-k to the bottom interface.

This has been extensively studied with various depth profiling techniques such as EELS (Electron Energy Loss Spectroscopy) and SIMS (Secondary Ion Mass Spectroscopy) to verify that the V_t shift elements are diffusing down to the SiO_2/high-k interface after annealing. In one experiment on MoAlN metal gate electrodes, back-side SIMS confirmed that there is indeed an inward diffusion of the Al from the electrode to the SiO_2-based interface layer with thermal budget and that this coincides with increases in the effective work function of the MoAlN metal electrode with increasing anneal temperature.

SIMS has also been performed on gate stacks with MgO caps or TaMgC electrodes before and after annealing. This analysis clearly shows that after a high temperature anneal there is a pile-up, or segregation, of Mg into the SiO_2-like interface layer. This is consistent with the fact that many of these materials are strong silicate forming elements [35]. It should be noted that the dopant species diffuse to the HfO_2/SiO_2 interface regardless of whether it is incorporated in the metal gate electrode or applied as an oxide dielectric cap layer.

There are multiple mechanisms that have been proposed to explain the observed V_t shifts. The first possible mechanism is that the effective work function of the electrode is actually changed by the electropositive (nFET) or electronegative (pFET) elements used as capping layers. This result is consistent with the effective work function and electronegativity correlation. However, it is not consistent with the electrical results in Fig. 5.12 that suggest the dopant species needs to be located at the SiO_2/HfO_2 interface to produce the desired V_t shifts. It is also possible that as the elements diffuse through the gate dielectric the HfO_2 defect chemistry and equilibrium oxygen vacancy concentration is being modulated [36]. In this argument one would expect Al to produce similar V_t shifts to La since they both have a valence of +3 and would favor the creation of positively charged oxygen vacancies in HfO_2. Since the opposite is observed this is also not the likely V_t shift mechanism. Finally, as the elements diffuse down further toward the SiO_2/high-k interface, they can react to form a silicate and create fixed charges or a dipole layer at the SiO_2/high-k interface layer. There is much supporting evidence for this model. First, electrical results suggest that for capping layers a high thermal budget needs to be applied to produce V_t shifts that are similar to those

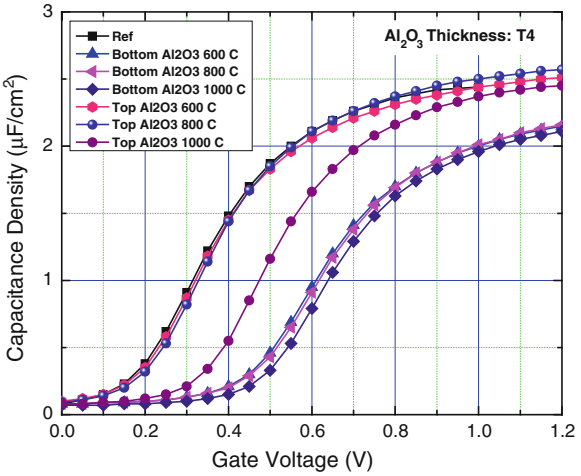


Fig. 5.12 Capacitance–Gate Voltage curves showing V_t shifts for *top*- and *bottom*-Al_2O_3 interface layers with different anneal temperatures. V_t shifts are achieved when the Al_2O_3 is at the bottom interface, or when the thermal budget diffuses the Al_2O_3 from the *top* to the *bottom* interface [25]

achieved at low thermal budgets when the cap layer is placed directly at the SiO_2/HfO_2 interface. SIMS and EELS depth profiling indicates that the capping elements diffuse down to the HfO_2/SiO_2 interface with anneal. The pile-up of dopant species at the SiO_2/HfO_2 interface is consistent with the fact that many of these dopant species are known to be strong silicate forming elements [35], and it has been theoretically demonstrated that these elements energetically prefer to segregate to the HfO_2/SiO_2 interface or the SiO_2 intermediate layer [37, 38]. Theoretical studies have evaluated a set of dopant species with different valences and work-functions/electro-negativities. Finally, the direction of the V_t shifts is consistent with the group electro-negativities of the doping oxide relative to HfO_2. The majority of evidence so far supports a model that the V_t shifts are caused by interface dipoles at the SiO_2/HfO_2 interface (Table 5.1, Figs. 5.13, 5.14).

Mechanism for Effective Oxide Thickness Reduction

The mechanism for EOT scaling that is achieved when the La, Mg or Al is incorporated into the gate electrode instead of as a dielectric capping layer is related to the thermodynamic competition for oxygen between the layers in the

Table 5.1 Comparison of the mean electro-negativities of some dopant oxides commonly employed to shift the threshold voltages in Hf-based gate dielectrics

Capping dielectrics	Geometric mean electronegativity (Pauling)
MgO	2.11
La_2O_3	2.18
Gd_2O_3	2.26
Y_2O_3	2.27
HfO_2	2.49
Al_2O_3	2.54
TiO_2	2.63
Ta_2O_5	2.71
SiO_2	2.82

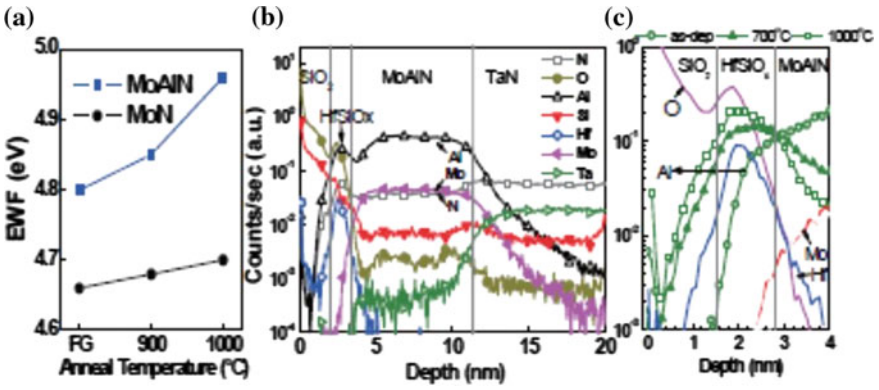


Fig. 5.13 Effective work function shift and aluminum diffusion into $HfSiO_x$ gate dielectric as a function of applied thermal budget. Backside SIMs profile of the10-nm MoAlN/2 nm $HfSiO_x$/2 nm terraced oxide stack showing **a** MAIN EWF as a function of annealing temperature **b** uniform MoAlN film even at 1,000 °C **c** local diffusion of Al through $HfSiO_x$ to the SiO_2 interface after annealing (a possible cause for EWF modulation) [88]

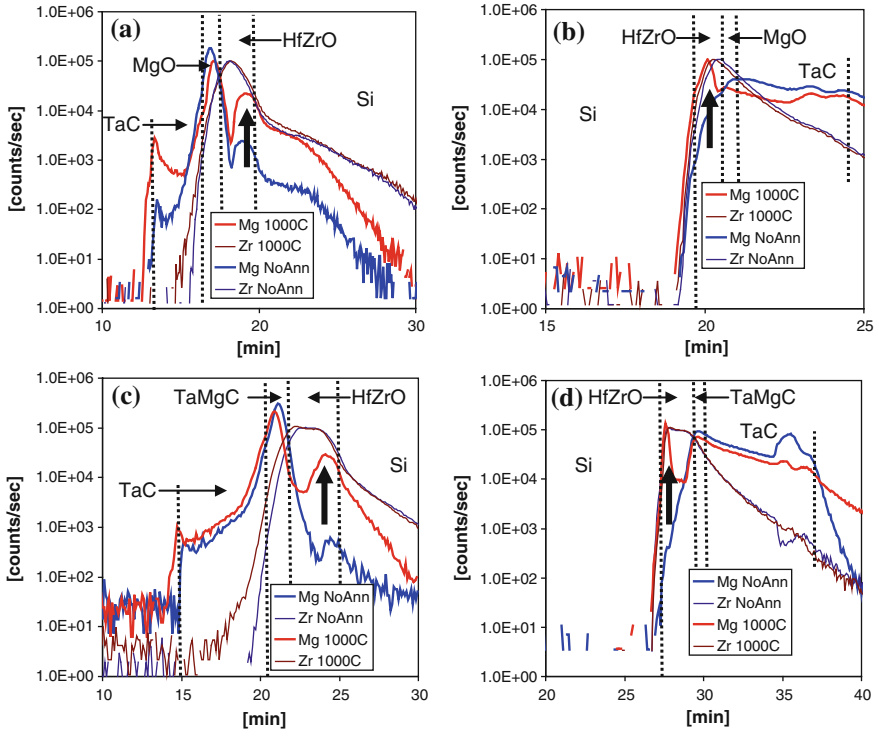


Fig. 5.14 *Front-* (**a**) and *back-* (**b**) side SIMS with MgO capping. *Front-* (**c**) and *back-* (**d**) side SIMS for TaMgC. The dielectric is $Hf_xZr_{1-x}O_2$ and only the Zr and Mg profiles are shown for clarity [28]

gate stack. This is summarized by the following reaction which illustrates how SiO_2 is converted into Si and a metal oxide with a higher dielectric constant if there is a large positive Gibbs free energy change at 1,000 K (ΔG°_{1000K}) [39].

$$Si + 2/yM_xO_y \rightarrow (2x)/yM + SiO_2$$

While using metallic capping layers to scavenge oxygen away from the bottom interface for EOT scaling, a commensurate degradation of mobility is also observed. A careful study comparing the use of La capping layers versus a remote interface scavenging technique (interface scavenging achieved without diffusing dopants into the high-k that would cause V_t shifts) suggests that the diffusion of La to the bottom interface does not result in additional mobility degradation beyond what is expected due to the thinning of the bottom SiO_2 interface layer. However, one should use caution because an extrinsic mobility degradation and device scatter can be observed depending on whether the dopant species or even with oxide cap layers if the cap layer is too thick. It should be noted that excellent process control is required for the deposition of these thin capping layers since the total dose of these layers is closely related to important device parameters such as mobility, EOT, and V_t.

Oxygen Vacancies

The threshold voltage shift mechanisms so far have focused on effective work function and charges at the SiO_2/high-k interface, but many experiments have shown that the bulk charge term in (5.2): $\Delta V_{fb} = -1/2\left(\varepsilon_{HfO_2}/\varepsilon_{SiO_2}\right)\rho_1 EOT_1^2/\varepsilon_{SiO_2}$ can be manipulated by controlling the oxygen vacancy concentration in the high-k dielectric. By annealing high-k devices in oxygen at an appropriate temperature and pressure positive shifts of the threshold voltage of up to 450 mV can be achieved by annihilating positively charged oxygen vacancies that exist in the high-k material.

Most of the studies on this topic focused on the top-down diffusion of oxygen through a metal gate electrode such as Pt [40], Re [41, 42], and TiN [43] as well as the lateral diffusion of oxygen from the sides of the gate [44]. The key results indicate that there are three key temperature ranges. At low temperatures no threshold voltage shifts are observed when the samples are annealed in oxygen. This is a diffusion limited temperature range where oxygen can not penetrate the metal gate electrode in the top–down experiments or diffusion through the high-k is sufficiently low such that only the shortest gate lengths show threshold voltage shifts in the lateral diffusion experiments. Above the critical temperature, a thermodynamic equilibrium exists between the oxygen partial pressure and the oxygen vacancy concentration in the HfO_2 film. As the oxygen annihilates positively charged oxygen vacancies large positive shifts of the threshold voltage can be produced. As the temperature continues to increase the maximum threshold voltage shift produced by the oxygen anneal reduces with increasing temperature.

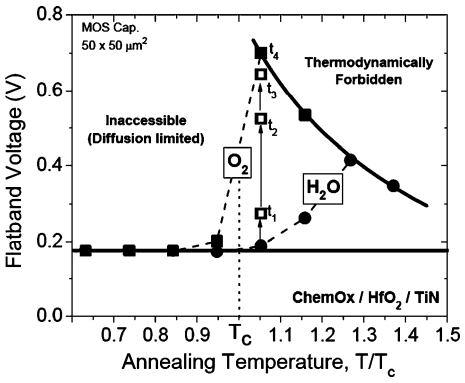


Fig. 5.15 Flatband voltage shifts in HfO_2 gate dielectrics as a function of oxygen annealing temperature [43]

This is because the equilibrium oxygen vacancy of the high-k film increases as temperature increases and because the equilibrium oxygen vacancy of the high-k film is balanced by the reaction of silicon and oxygen to form SiO_2 at the silicon substrate. When this occurs, an effective oxide thickness increase or gate leakage decrease accompanies the positive threshold voltage shifts due to the growth of additional SiO_2 at the silicon substrate (Fig. 5.15).

The studies of lateral oxidation are consistent with the top-down oxidation studies [44]. In general, at low temperatures only the devices with the shortest gate lengths begin to show threshold voltage shifts. With either increasing oxygen anneal time or increasing oxygen temperature, the threshold voltage shifts of the short devices become larger and the longer devices begin to show threshold voltage shifts also. This result indicates that a concentration gradient exists for the oxygen vacancy concentration with the highest concentration of vacancies in the center of the gate and the lowest concentration at the edge of the gate.

The positive threshold voltage shifts caused by the oxygenation of a high-k gate stack can be reversed by annealing in vacuum [41, 44], reducing ambients [40], or as previously discussed, at elevated temperatures where there is a higher equilibrium oxygen vacancy in the high-k material. This is the fundamental reason why it is difficult to achieve a low threshold voltage pFET device in a gate first integration scheme. This has been discussed in the context of a threshold voltage "roll-off" where low pFET threshold voltage devices can be achieved if the gate stack is never exposed to elevated temperatures, but at elevated temperatures there is an increase in the pFET threshold voltage when the interfacial layer separating the Si substrate from the high-k film is thin (Fig 5.16). It should be noted that the proximity of the high-k to the substrate appears to facilitate the formation of oxygen vacancies in the high-k film at elevated temperatures. When the high-k is near the substrate, oxygen exchange between the high-k and the silicon substrate readily occurs and the oxygen chemical potential is shifted. This results in a dipole offset in the gate stack that increases the pFET threshold voltage after high temperature processing [45]. The impact of this high temperature processing on the

pFET threshold voltage is an important consideration for integrating high-k materials into a CMOS-compatible process flow.

5.3 CMOS Metal Gate Integrations

This section deals with the integration of metal gate electrodes with different work functions to build both NMOS and PMOS devices. CMOS integration is easily done with poly-silicon electrodes by ion implanting n-type or p-type dopants into NMOS or PMOS electrodes to get appropriate work functions for both device polarities. For metal gate electrodes the integration scheme is more complex. Both conventional gate-first integrations [46, 47] and replacement gate [48–51] integrations have been proposed. Gate-first integrations have the advantage of being structurally similar to a poly/SiO_2-based CMOS process. In replacement gate integrations, the highest processing temperatures are performed prior to depositing the gate materials. Replacement gate integrations can therefore avoid the increase in the pV_t due to the generation of positively charged oxygen vacancies in the high-k.

5.3.1 Materials Considerations for Metal Gate Electrodes

When used in a gate-first integration scheme, metal gate electrode candidates must possess thermal stability up to the dopant activation temperatures of approximately 1,000 °C. Thermal stability includes the absence of gross reactions between the dielectric and other surrounding materials, no inter-diffusions with surrounding materials, sufficient bulk phase stability, and smooth interfaces with the dielectric [52]. Due to these stringent thermal requirements refractory metals such as W, Re, Ta, and Mo have attracted interest.

Metal nitrides and carbides have also gathered attention as metal gate electrodes. These materials are renowned for their use as diffusion barriers in the semiconductor industry. Transition metal nitrides and carbides are comprised of a

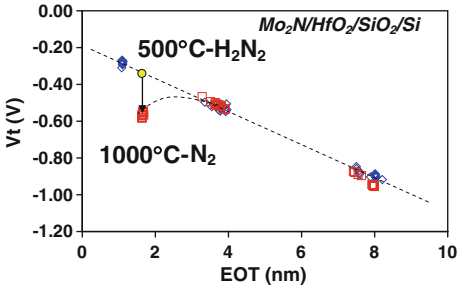


Fig. 5.16 Threshold voltage versus EOT for a constant thickness of HfO_2 on varying thicknesses of underlying SiO_2. Note that with high temperature processing the threshold voltage only shifts when the interface is thin

face-centered-cubic metal structure with nitrogen or carbon atoms occupying the octahedral interstices. This results in the rock-salt structure (NaCl). Since N and C occupy interstices the lattice parameter is typically only ~5 % larger than that of the pure metal compound. Stuffing the interstices with N or C helps give these materials their excellent diffusion barrier properties. These materials exhibit metal-like conduction, but at the same time are highly refractory compared to the pure metal constituent. It is believed that a mixture of metallic, covalent, and ionic bonding character is responsible for this behavior [53]. There have been numerous studies on the work function of interstitial nitrides for electrode and emitter applications. This includes MoN [54], WN [52], NbN [55], TaN [52, 56], TiN [52, 57–59], ZrN [55], and HfN [60], among others. TiAlN has also been widely studied as a metal gate electrode [61]. Although it is a ternary metal nitride, it is more closely related to the binary nitrides than the amorphous ternary metals because this compound exists in the rock salt structure for Al:Ti ratios <0.40 and in a mixture of rock salt and wurtzite structures for >0.40 Al:Ti ratio [62]. TiAlN loses its conductive properties with increasing presence of the AlN wutzite phase. Metal carbide materials have also shown promise as thermally stable metal gate electrodes. [27, 63, 64]. Solid-solutions of metal carbides and nitrides have been compared and it has been shown that carbide materials have lower effective work functions than their analogous nitride making them good nFET electrode candidates [65, 66].

Ternary alloys of a transition metal (TM), silicon, and nitrogen have also been widely investigated for their properties as diffusion barriers. These films are found to exist in a highly metastable amorphous structure [67]. This property makes them excellent diffusion barriers to elevated temperatures. Ta–Si–N has been studied as a stable NMOS electrode candidate [52, 68–70].

Even if a thermally stable metal gate electrode is selected there are more subtle effects detected with electrical measurements that are related to the thermo-chemistry of the gate stack. As previously discussed in the section on alloyed metal gate electrodes for work function tuning, if a metal/high-k/SiO_2/Silicon stack is annealed at high temperatures there is a thermo-chemical competition for oxygen in the gate stack that takes place between the silicon substrate and the metal gate material where oxygen transport across the HfO_2 layer is too rapid to limit the reactions [71]. Thermo-chemical calculations of a simple TiN_x/HfO_2/SiO_2/Si system were performed to illustrate this effect. If the composition is such that $x < 1$, then the oxygen in the SiO_2 interface layer can be dissolved into metal gate electrode and the Si returns to the substrate. However, if $x > 1$ then the calculations indicate that the system drives towards the creation of extra interfacial SiO_2. As was observed in the examples of alloyed metal gate electrodes these thermo-chemical interactions can be manipulated to achieve a desired electrical response (Fig. 5.17).

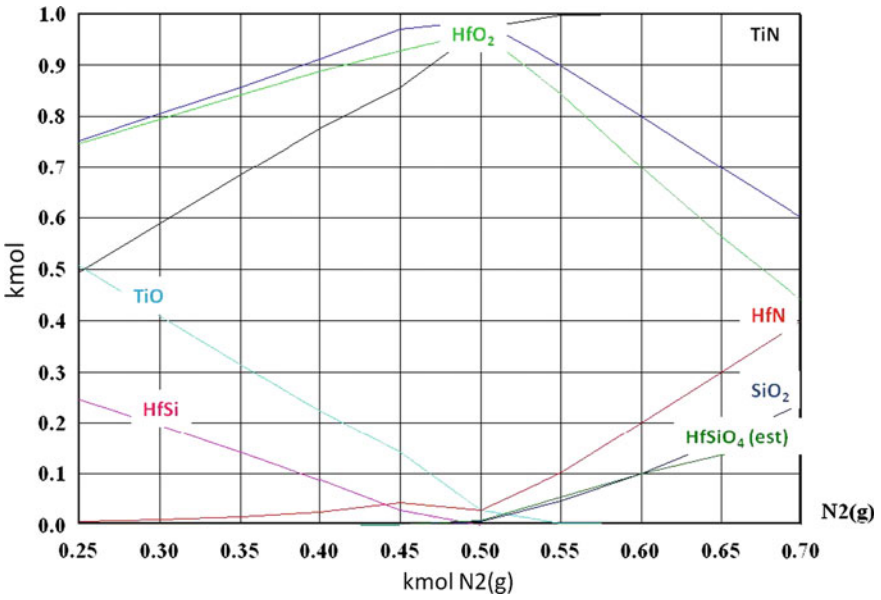


Fig. 5.17 Illustrates the fraction of each species present (Si, SiO_2, HfO_2, HfSi, TiN, TiO, HfN, and $HfSiO_4$) by varying the composition of a TiN_x metal gate electrode. For N-deficient TiN, there is a tendency to form TiO and HfSi. For N-rich TiN there is a tendency to form HfN and extra SiO_2 interface layer

5.3.2 Gate First Integration

A gate first integration is defined as one in which the high-k material and the metal gate electrodes are deposited prior to implanting the extension, halo, and source and drain dopants into the transistor. The gate first integration is desirable because it is the most similar to the conventional poly/SiON transistor integration and the least disruptive to the other modules that need to be integrated with the gate stack. Typically, in a gate first integration, thin metal gate layers are capped with poly-Si allowing for gate silicidation to be performed in the same fashion as it was in a poly/SiON integration. In a gate-first integration, the high-k metal gate materials must withstand significant process temperatures, typically in excess of 1,000 °C, which are required to activate the implanted dopants in the silicon substrate. Exposing the gate stack to such temperatures can present some challenges. It limits the number of gate stack metals to those that are thermally stable at these temperatures. In addition, as detailed in the section on the role of oxygen on the effective work function of the transistor, high processing temperature can result in higher than desired pFET threshold voltages.

5.3.2.1 Dual Metal Gate Electrodes

There are two main approaches to dual-metal gate electrode integration. The first is to use two separate metal gate electrodes with different effective work functions. To get different work functions, two separate metals can be used for nFET and pFET electrodes or one of the metals can be alloyed in a manner that shifts its effective work function. A second approach is to combine the dual-metal gate electrodes with some of the other threshold voltage engineering techniques that manipulate fixed charges or dipoles inside the gate dielectric layer. In this situation, the final structure is one in which the gate dielectric and the metal gates are different between nFET and pFET electrodes.

Dual-Metal Gate, Single-Gate Dielectric

The integration approach that accommodates vastly different metals for NMOS and PMOS electrodes is the stacked dual metal gate integration (Fig. 5.18). This integration involves depositing one metal over both the NMOS and the PMOS regions, then patterning one of the two regions and removing the electrode using a wet etch chemistry that is selective to the underlying gate dielectric. A second metal can then deposited over the first metal. The most difficult aspect of this integration is the plasma etching of two different gate stacks of different heights. This integration has been demonstrated on Si_3N_4 gate dielectric [72] and on HfO_2 gate dielectrics [73, 74].

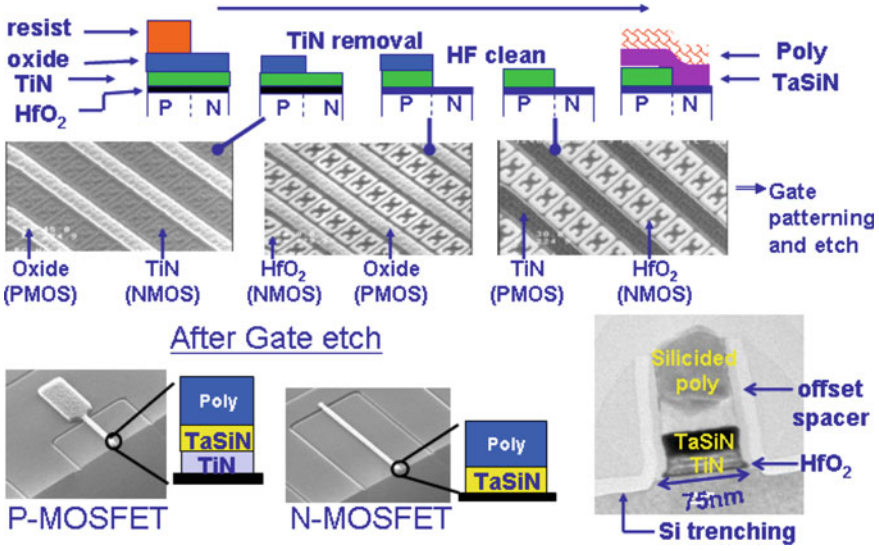


Fig. 5.18 Process flow for the stacked dual metal gate integration. The difficulty in this approach is that two gate stacks of different heights need to be etched simultaneously

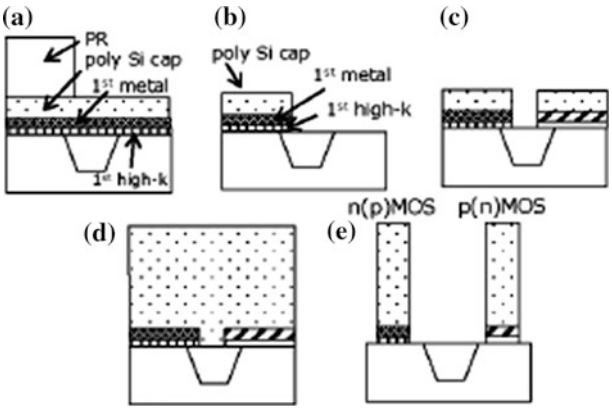


Fig. 5.19 Dual Metal Dual Dielectric (DMDD) CMOS integration scheme. [75]

The basic dual-metal gate integration scheme has been expanded to include additional work function tuning elements into the gate stack. One of these integration approaches uses two separate gate dielectrics, referred to as a dual-metal, dual-dielectric integration scheme. This allows for the use of different capping oxides on both the nMOS and pMOS devices as well as two separate work function metal gate electrodes to provide additional flexibility in threshold voltage tuning [75, 76] (Fig. 5.19).

An even more complicated approach employed dual-dielectrics, dual-metals, and dual-channels [77]. In this integration scheme the valance band-offset between SiGe and Si results in an additional reduction of the pFET threshold voltage.

Alloying for Dual-Metal Gate Work Functions

There have been numerous attempts to simplify the dual-metal gate integration to avoid having to selectively remove one metal from the gate dielectric without having any deleterious impact on the device. This can be achieved by alloying or tuning the work function of the metal in either the nMOS or pMOS region without actually depositing a second metal.

In the metal alloy approach, two metals are initially deposited on top of each other. Either the NMOS or PMOS region is then patterned so that the top metal can be selectively removed. A high temperature annealing process is then used to inter-diffuse the two metals. This leaves an inter-diffused metal for one of the electrodes and an elemental metal for the other electrode. This approach has been performed using Ru-Ta [78, 79], Ti-Ni [80], and Pt–Ta [81] alloys. The alloyed electrode approach has two foreseeable problems. First, it may be difficult to get one alloy system to span the entire 1.1 eV range to meet the NMOS and PMOS work function requirements. Second, some of the elements used in the alloy may not be thermally stable in contact with a high permittivity dielectric.

The nitrogen modulation integration has been reported in two forms. The first method is to ion implant nitrogen into a metal to shift its work function. This has been demonstrated for Mo-based electrodes [82, 83], and for TiN-based electrodes where the N concentration is modulated via implantation of N [84]. Adjusting the nitrogen concentration has also been accomplished by solid-state diffusion of nitrogen from a nitrogen-rich film into a nitrogen-deficient film [85]. Like the alloy approach, it may be difficult to span the entire work function range to meet the NMOS and PMOS electrode requirements by implanting nitrogen.

One material system that has been well suited for the nitrogen modulation technique is TaC. TaC has been shown to have a lower effective work function than its analogous nitride and it has been demonstrated that the work function of TaC can be controllably increased by alloying in nitrogen [63]. To exploit this property of the TaC_xN_y system in a metal gate integration, there has been some work to nitride TaC either by annealing or with plasma nitridation (Fig. 5.20).

5.3.2.2 Single Metal Gate, Dual-Dielectric

One of the primary manufacturing challenges for a dual-metal gate integration is the patterning and removal of one metal from the surface of the high-k film and the need to perform the simultaneous etch of two gate stacks with varying heights and/or materials. A simplification for manufacturing would be to integrate a gate stack that has a single metal gate electrode, but still achieves threshold voltages ideal for NMOS and PMOS devices by using separate dielectric capping layers for nMOS and pMOS. In this situation the challenge becomes the selective patterning of the dielectric capping layers. Depending on the integration approach, the dielectric capping layers can be placed at the SiO_2/high-k interface or on top of the high-k film and selective patterning with good resist adhesion is achieved by using a developable bottom anti-reflective coating (BARC) layer and wet chemistry to pattern the dielectric capping layers [86] (Fig. 5.21).

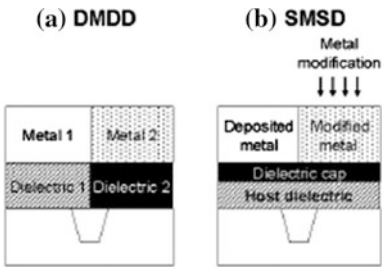


Fig. 5.20 Schematics of high-k/metal gate CMOS with deposited **a** dual metal-dual-dielectric (DMDD) and **b** single-metal-single-dielectric (SMSD) gate stacks [27]

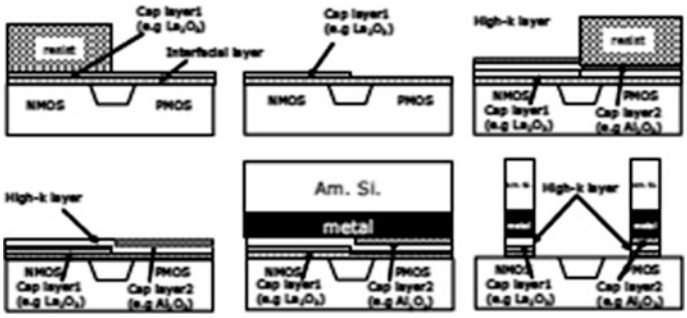


Fig. 5.21 Single-metal, dual-dielectric integration scheme (SMDD). CMOS process flow, with the nMOS cap located below and pMOS cap located above the bulk high-k dielectric [86]

5.3.3 Gate Last or Replacement Gate Integration

An alternative approach is the gate-last or replacement-gate integration. The main concept in the gate-last integration is that there is a "dummy gate" in place during the high temperature processing steps that are required to activate the dopants. This "dummy gate" can then be removed and replaced with the work function setting high-k/metal gate materials which is then never exposed to extreme subsequent processing temperatures. The replacement gate integration was the first high-k metal gate integration to be successfully implemented into manufacturing.

There are variations of the replacement metal gate integration scheme. One such integration that has been demonstrated is a hybrid gate-first/gate-last integration. In this integration the high-k layer, a thin portion of metal gate, and poly-Si capping layer are deposited. After encapsulation with spacers and implantation of the extensions/halos and source/drain, the poly-Si is eventually removed from both the nFET and pFET devices. At this stage the poly-Si voids are filled back in with a first work function setting metal. This will need to be patterned and removed in a manner that is similar to those used in the gate-first dual-metal gate integration. A second gate metal is then deposited for the opposite polarity transistor. The remainder of the gate is then filled back in with a low resistivity metal [48].

Another approach is a full replacement gate integration. In this approach, a poly/SiON gate is used for all the implantation and anneal steps. At a later stage of processing, the poly-Si is then removed and replaced with the high-k gate dielectric and the work function setting metal gate electrodes. In this scheme even the high-k film is not exposed to the source-drain activation temperatures [49].

There are a couple of benefits to the replacement gate integration that should be discussed. First, the replacement gate integration avoids many of the complications associated with the exposure of high-k/metal gate materials to elevated temperatures. For replacement gate integration, the gate stack materials are likely not exposed to temperatures greater than $\sim$500 °C. In this temperature range, there

are more options for engineering a pFET transistor by modulating the oxygen vacancy concentration. Another widely touted benefit of the replacement gate integration is that the removal of the "dummy gate" can enhance the strain applied to the channel region from some of the other commonly used stress enhancement techniques like embedded SiGe source-drain regions and stress liners. The stress of the work function setting metal gates and the fill metal can also be tailored to directly couple to the channel strain [87].

5.4 Conclusions

In conclusion, metal gate electrodes are now being used in production by multiple companies. Some companies have opted for gate-last integration approaches and others have opted for gate-first integration approaches. In both cases, a significant number of materials, device, and integration challenges had to be overcome. As CMOS technologies continue to scale to the next technology nodes, metal gate electrodes will be used together with entirely new device architectures, channel materials, and gate dielectrics. Much of the learning from the initial development of metal gate electrodes will continue to apply in these situations, but there will also be many exciting new challenges and opportunities to build upon the years of research that were required to ultimately implement metal gate electrodes into advanced planar CMOS technology nodes.

References

1. T. Hori, *Gate Dielectrics and MOS ULSIs* (Springer, Berlin, 24 1997)
2. C. Hobbs, L. Fonseca, V. Dhandapani, S. Samavedam, B. Taylor, J. Grant, L. Dip, D. Triyoso, R. Hegde, D. Gilmer, R. Garcia, D. Roan, L. Lovejoy, R. Rai, L. Hebert, H. Tseng, B. White and P. Tobin. in *Symposium on VLSI Tech. Digest*, pp. 9–10, 2003
3. M. Gutowski, J.E. Jaffe, C.L. Liu, M. Stoker, R.I. Hegde, R.S. Rai, P.J. Tobin, Appl. Phys. Lett. **80**(11), 1897–1899 (2002)
4. I. De, D. Johri, A. Srivastava, C.M. Osburn, Solid State Electron. **44**, 1077 (2000)
5. B. Cheng, B. Maiti, S. Samavedam, J. Grant, B. Taylor, P. Tobin, J. Mogab, 2001 IEEE International SOI Conference, pp. 91–92, 2001
6. H.-C. Wen, et al, IEEE Elec. Dev. Lett. **27**(7), 598–601 (2006)
7. R. Jha, J. Gurganos, Y.H. Kim, R. Choi, J. Lee, V. Misra, IEEE Elec. Dev. Lett. **25**(6), 420 (2004)
8. S.M. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981)
9. R.T. Tung, Mat. Sci. Engr. R, **35,** 1–138 (2001)
10. J. Tersoff, Phys. Rev. Lett. **52**, 465 (1984)
11. W.E. Spicer, P.W. Chye, P.R. Skeath, C.Y. Su, I. Lindau, J. Vac. Sci. Technol. **16**, 1422 (1979)
12. H. Hasegawa, H. Ohno, J. Vac. Sci. Technol. B **4**, 1130 (1986)
13. J.L. Freeouf, J.M. Woodall, Appl. Phys. Lett. **39**, 727 (1981)
14. Y. C. Yeo, P. Ranade, T.-J. King, C. Hu, IEEE Elec. Dev. Lett. **23**(6), 342 (2002)

15. Y.C. Yeo, T.-J. King, C. Hu, J. Appl. Phys. **92**(12), 7266 (2002)
16. J.A. Venables, *Introduction to Surface and Thin Film Processes* (Cambridge University Press, Cambridge, 2000), p. 31
17. J. Robertson, Appl. Surf. Sci. **190**, 2 (2002)
18. S. Kurtin, T.C. McGill, C.A. Mead, Phys. Rev. Lett. **22**(26), 1433–1436 (1969)
19. S.G. Louie, J.R. Chelikowsky, M.L. Cohen, J. Vac. Sci. Technol. **13**, 790 (1976)
20. H.B. Michaelson, IBM J. Res. Develop. **22**(1), 72–80 (1978)
21. W. Gordy, W.J.O. Thomas, Phs. Rev. **24**, 439 (1956)
22. S.G. Louie, J.R. Chelikowsky, M.L. Cohen, J. Vac. Sci. Technol. **13**(4), 790–797 (1976)
23. M. Schluter, J. Vac. Sci. Technol. **15**(4), 1347–1376 (1978)
24. V. Narayanan, et al., in *Digest of Technical Papers of Symposium on VLSI Technology*, pp. 178–179 (2006)
25. H. Jagannathan, V. Narayanan, S. Brown, Abstract ECS Transactions, Meet, **16**(5), 19–26 (2008)
26. H.R. Harris, et al., IEEE International Electron Devices Meeting IEDM, pp. 1–4 (2006)
27. V.S. Chang, et al., IEEE International Electron Devices Meeting IEDM 2007, **10–12**, 535–538 (2007)
28. J. Schaeffer, M. Raymond, D. Gilmer, R. Gregory, B. Taylor, J. Jiang, D. Triyoso, R. Hegde, S. Samavedam, ECS Trans. **16**(5), 3–10 (2008)
29. M. Goto, S. Kawanaka, S. Inumiya, N. Kusunoki, M. Saitoh, K. Tatsumura, A. Kinoshita, S. Inaba, Y. Toyoshima, in *Symposium on VLSI Technology*, pp. 214–215 (2009)
30. H.N. Alshareef, H.F. Luan, K. Choi, H.R. Harris, H.C. Wen, M.A. Quevedo-Lopez, P. Majhi, B.H. Lee, Appl. Phys. Lett. **88**, 112–114 (2006)
31. B. P. Linder, V. Narayanan, V. K. Paruchuri, E. Cartier, S. Kanakasabapathy, Technical Digest Solid State Devices and Materials Conference Japanese Applied Physical Society, Tokyo, 16 (2007)
32. C.C. Hobbs, L.R.C. Fonseca, A. Knizhnik, V. Dhandapani, S.B. Samavedam, W.J. Taylor, J.M. Grant, L. Dip, D.H. Triyoso, R.I. Hegde, D.C. Gilmer, H.H. Tseng, B.E. White, P.J. Tobin, IEEE Trans. Electron. Devices **51**, 971 (2004)
33. S.B. Samavedam, et al., IEDM Tech. Dig., 307 (2003)
34. H.-H. Tseng, P. Kirsch, C.S. Park, G. Bersuker, P. Majhi, M. Hussain, R. Jammy, Microelectron. Eng. **86**(7–9), 1722–1727 (2009)
35. D.G. Schlom, J.H. Haeni, MRS Bull. **27**(03), 198–204 (2002)
36. S. Guha, V.K. Paruchuri, M. Copel, V. Narayanan, Y.Y. Wang, P.E. Batson, N.A. Bojarczuk, B. Linder, B. Doris, Appl. Phys. Lett. **90**, 092902 (2007)
37. M. Copel, S. Guha, N. Bojarczuk, E. Cartier, V. Narayanan, V. Paruchuri, Appl. Phys. Lett. **95**(21), 212903 (2009)
38. J. Robertson, L. Lin, Appl. Phys. Lett. **95**, 012906 (2009)
39. T. Ando, M.M. Frank, K. Choi, C. Choi, J. Bruley, M. Hopstaken, M. Copel, E. Cartier, A. Kerber, A. Callegari, D. Lacey, S. Brown, Q. Yang, V. Narayanan, IEEE International Electron Devices Meeting (IEDM), pp. 1–4 (2009)
40. J.K. Schaeffer, Appl. Phys. Lett., **85**(10), 1826 (2004)
41. E. Cartier, et al., in *Proceedings of VLSI Symposium*, p. 230 (2005)
42. S. Guha, V. Narayanan, Phys. Rev. Lett. **98**, 196101 (2007)
43. E. Cartier, M. Hopstaken, M. Copel, Appl. Phys. Lett. **95**, 042901 (2009)
44. E. Cartier, M. Steen, B.P. Linder, T. Ando, R. Iijima, M. Frank, J.S. Newbury, Y.H. Kim, F.R. McFeely, M. Copel, R. Haight, C. Choi, A. Callegari, V.K. Paruchuri, V. Narayanan, in *Symposium on VLSI Technology*, pp. 42–43 (2009)
45. K. Shiraishi, K. Yamada, K. Torri, Y. Akasak, K. Nakajima, M. Konno, T. Chikyow, H. Kitajima, T. Arikado, Jap. J. Appl. Phys. **52**, L1413 (2004)
46. M. Chudzik, et al., in *IEEE Symposium on VLSI Technology*, pp. 194–195 (2007)
47. K. Henson, H. Bu, M.H. Na, Y. Liang, U. Kwon, S. Krishnan, J. Schaeffer, R. Jha, N. Moumen, R. Carter, C. DeWan, R. Donaton, D. Guo, M. Hargrove, W. He, R. Mo, R. Ramachandran, K. Ramani, K. Schonenberg, Y. Tsang, X. Wang, M. Gribelyuk, W. Yan,

J. Shepard, E. Cartier, M. Frank, E. Harley, R. Arndt, R. Knarr, T. Bailey, B. Zhang, K. Wong, T. Graves-Abe, E. Luckowski, D.-G. Park, V. Narayanan, M. Chudzik, M. Khare, IEEE International Electron Devices Meeting 2008 (IEDM 2008), pp. 1–4 (2008)
48. K. Mistry et al., in *IEEE International Electron Devices Meeting 2007 (IEDM 2007)*, pp. 247–250, 2007
49. S. Natarajan, M. Armstrong, M. Bost, R. Brain, M. Brazier, C.-H. Chang, V. Chikarmane, M. Childs, H. Deshpande, K. Dev, G. Ding, T. Ghani, O. Golonzka, W. Han, J. He, R. Heussner, R. James, I. Jin, C. Kenyon, S. Klopcic, S.-H. Lee, M. Liu, S. Lodha, B. McFadden, A. Murthy, L. Neiberg, J. Neirynck, P. Packan, S. Pae, C. Parker, C. Pelto, L. Pipes, J. Sebastian, J. Seiple, B. Sell, S. Sivakumar, B. Song, K. Tone, T. Troeger, C. Weber, M. Yang, A. Yeoh, K. Zhang, in *IEEE International Electron Devices Meeting 2008 (IEDM 2008)*, pp. 1–3, 2008
50. F. Ducroquet, H. Achard, F. Coudert, B. Previtali, J-F., Lugand, L. Ulmer, T. Farjot, Y. Gobil, M. Heitzmann, S. Tedesco, M-E. Nier, S. Deleonibus, Trans. Electron Devices **48**, 8, (2001)
51. R. Li, Q. Xu, Trans. Electron Devices **49**, 11 (2002)
52. J. Schaeffer et al., J. Vac. Sci. Tech. B **21**(1), 11 (2003)
53. H. Pierson, Handbook of Refractory Carbides and Nitrides (Noyes, 1996)
54. P. Ranade, H. Takeuchi, T-J. King, C. Hu, Electrochem. Sol. State Lett., 4 G85 (2001)
55. Y. Saito, S. Kawata, H. Nakane, H. Adachi, Appl. Surf. Sci. **146**, 177 (1999)
56. C.S. Kang, H.-J. Cho, Y.H. Kim, R. Choi, K. Onishi, A. Shahriar, J.C. Lee, J. Vac. Sci. Technol. **B 21**, 5 (2003)
57. M. Moriwaki, T. Yamada, Jpn. J. Appl. Phys., 40 Part 1, **4**B 2679 (2001)
58. D-G. Park, H-J. Cho, K-Y. Lim, T-H. Cha, I-S. Yeo, J-W. Park, J. Electrochem. Soc., **148**(9), F189 (2001)
59. D-G. Park, K-Y. Lim, H-J. Cho, T-H. Cha, J–J. Kim, J-K. Ko, I-S. Yeo, J-W. Park, in *Symposium VLSI Technology Digest*, pp. 65–66 (2001)
60. H.Y. Yu, H.F. Lim, J.H. Chen, M.F. Li, C. Zhu, C.H. Tung, A.Y. Du, W.D. Wang, D.Z. Chi, D.-L. Kwong, IEEE Elec. Dev. Lett **24**, 4 (2003)
61. T-H. Cha, D-G. Park, T-K. Kim, S-A Jang, I-S. Yeo, Appl. Phys. Lett., **81**(22), 4192 (2002)
62. F. Esaka, K. Furuya, H. Shimada, M. Imamura, N. Matsubayashi, T. Kikuchi, H. Ichimura, A. Kawana, Surf. Interface Anal. **27**, 1098 (1999)
63. J.K. Schaeffer, C. Capasso, L.R.C. Fonseca, S. Samavedam, D.C. Gilmer, Y. Liang, S. Kalpat, B. Adetutu, H.-H. Tseng, Y. Shiho, A. Demkov, R. Hegde, W.J. Taylor, R. Gregory, J. Jiang, E. Luckowski, M.V. Raymond, K. Moore, D. Triyoso, D. Roan, B.E. White Jr, P.J. Tobin, IEDM Tech. Dig., p. 287 (2004)
64. H.-H. Tseng, C.C. Capasso, J.K. Schaeffer, E.A. Hebert, P.J. Tobin, D.C. Gilmer, D. Triyoso, M.E. Ramon, S. Kalpat, E. Luckowski, W.J. Taylor, Y. Jeon, O. Adetutu, R.I. Hegde, R. Noble, M. Jahanbani, C. El Chemali, B.E. White, IEEE International Electron Devices Meeting (IEDM) Technical Digest, pp. 821–824 (2004)
65. J.K. Schaeffer, D.C. Gilmer, C. Capasso, S. Kalpat, B. Taylor, M.V. Raymond, D. Triyoso, R. Hegde, S.B. Samavedam, B.E. White Jr, Microelectron. Eng. **84**(9–10), 2196–2200 (2007)
66. J.K. Schaeffer, C. Capasso, R. Gregory, D. Gilmer, L.R.C. Fonseca, M. Raymond, C. Happ, M. Kottke, S.B. Samavedam, P.J. Tobin, B.E. White Jr, J. Appl. Phys. **101**, 014503 (2007)
67. M.-A. Nicolet, P.H. Giauque, Microelectron. Eng. **55**, 357 (2001)
68. Y-S. Suh, G. Heuss, V. Misra, J. Electrochem. Soc. **150**(5), F79–F82 (2003)
69. Y.-S. Suh, G.P. Heuss, J.-H. Lee, V. Misra, Electron. Dev. Lett **24**(7), 439–441 (2003)
70. D-G. Park, T-H. Cha, K-Y. Lim, H-J. Cho, T-K. Kim, S-A. Jang, Y-S. Suh, V. Misra, I-S. Yeo, J-S. Roh, J-W. Park, H-K. Yoon, IEDM Technical Digest, 671 (2001)
71. M.J. Hopstaken, J. Bruley, D. Pfeiffer, M. Copel, M.M. Frank, E. Cartier, T. Ando, V. Narayanan, ECS Trans. **28**(1), 105–113 (2010)
72. Y.-C. Yeo, Q. Lu, P. Ranade, H. Takeuchi, K.J. Yang, I. Polishchuk, T.-J. King, C. Hu, S.C. Song, H.F. Luan, D.-L. Kwong, IEEE Electron. Dev. Lett **22**(5), 227–229 (2001)

73. S. B. Samavedam, L. B. La, J. Smith, S. Daksha-Murthy, E. Luckowski, J. Schaeffer, M. Zavala, R. Martin, V. Dhandapani, D. Triyoso, H. H. Tseng, P. J. Tobin, D. C. Gilmer, C. Hobbs, W. J. Taylor, J. M. Grant, R. Hegde, J. Mogab, C. Thomas, P. Abramowitz, M. Moosa, J. Conner, J. Jiang, V. Aruachalam, M. Sadd, B-Y. Nguyen, B. White, IEDM Technical Digest, pp. 433–436 (2002)
74. S.C. Song, Z.B. Zhang, M.M. Hussain, C. Huffman, J. Barnett, S.H. Bae, H.J. Li, P. Majhi, C.S. Park, B.S. Ju, H.K. Park, C.Y. Kang, R. Choi, P. Zeitzoff, H.H. Tseng, B.H. Lee, R. Jammy, in *Symposium on VLSI Technology Digest of Technical Papers*, p. 13–14 (2006)
75. H.N. Alshareef, et al., in *Symposium on Digest of Technical Papers VLSI Technology*, 7–8 (2006)
76. S. Kubicek, T. Schram, E. Rohr, V. Paraschiv, R. Vos, M. Demand, C. Adelmann, T. Witters, L. Nyns, A. Delabie, L.-Å. Ragnarsson, T. Chiarella, C. Kerner, A. Mercha, B. Parvais, M. Aoulaiche†, C. Ortolland, H. Yu, A. Veloso, L. Witters, R. Singanamalla†, T. Kauerauf†, S. Brus, C. Vrancken, V.S. Chang1, S-Z. Chang1, R. Mitsuhashi2, Y. Okuno2, A. Akheyar3, H.-J. Cho4, J. Hooker5, B. J. O'Sullivan, S. Van Elshocht, K. De Meyer†, M. Jurczak, P. Absil, S. Biesemans, T. Hoffmann, in *Symposium on VLSI Technology Digest of Technical Papers*, pp. 130–131 (2008)
77. H.R. Harris, P. Kalra, P. Majhi, M. Hussain, D. Kelly, O. Jungwoo, D. He, C. Smith, J. Barnett, P.D. Kirsch, G. Gebara, J. Jur, D. Lichtenwalner, A. Lubow, T.P. Ma, S. Guangyu, S. Thompson, B.H. Lee, H.-H. Tseng, R. Jammy, in *IEEE Symposium on VLSI Technology*, pp. 154–155 (2007)
78. J-H. Lee, H. Zhong, Y.-S. Suh, G. Heuss, J. Gurganus, B. Chen, V. Misra, IEDM Technical Digest, pp. 359–362 (2002)
79. V. Misra, H. Zhong, H. Lazar, IEEE Electron. Dev. Lett. **23**(6), 354–356 (2002)
80. I. Polishchuk, P. Ranade, T-J. King, C. Hu, IEEE Electron Dev. Lett., **22**(9), 444–446 (2001)
81. B.-Y. Tsui, C.-F. Huang, IEEE Electron. Dev. Lett. **24**(3), 153–155 (2003)
82. Q. Lu, R. Lin, P. Ranade, T-J. King, C. Hu, in *Symposium on IEEE VLSI*, pp. 45–46 (2001)
83. P. Ranade, H. Takeuchi, T.-J. King, C. Hu, Electrochem. Sol. State Lett. **4**, G85 (2001)
84. H. Wakbayashi, Y. Saito, K. Takeuchi, T. Mogami, T. Kunio, IEEE Trans. Electron Dev. **48** 2363 (2001)
85. R.J.P. Lander, J.C. Hooker, J.P. van Zijl, F. Roozeboom, M.P.M. Maas, Y. Tamminga, R.A.M. Wolters, Mater. Res. Soc. Proc. V **716**, 253 (2002)
86. T. Schram, S. Kubicek, E. Rohr, S. Brus, C. Vrancken, S.-Z. Chang, V.S. Chang, R. Mitsuhashi, Y. Okuno, A. Akheyar, H.-J. Cho, J.C. Hooker, V. Paraschiv, R. Vos, F. Sebai, M. Ercken, P. Kelkar, A. Delabie, C. Adelmann, T. Witters, L.-A. Ragnarsson, C. Kerner, T. Chiarella, M. Aoulaiche, M.-J. Cho, T. Kauerauf, K. De Meyer, A. Lauwers, T. Hoffmann, P.P. Absil, S. Biesemans, in *Symposium on VLSI Technology*, pp. 44–45 (2008)
87. C. Auth et al., in *Symposium on VLSI Technology*, 128 (2008)
88. H.-C. Wen, S.C. Song, C.S. Park, C. Burham, G. Bersuker, K. Choi, M.A. Quevedo-Lopez, B.S. Ju, H.N. Alshareef, H. Niimi, H.B. Park, P.S. Lysaght, P. Majhi, B.H. Lee, in *IEEE Symposium on VLSI Technology*, pp. 160–161 (2007)

Chapter 6
V_{FB}/V_{TH} Anomaly in High-k Gate Stacks

Akira Toriumi and Toshihide Nabatame

Abstract One of the biggest challenges of metal/high-k gate stack technology is controlling the threshold voltage (V_{TH}) because achieving a high performance CMOS is almost impossible without it. We discuss anomalous V_{TH} in poly-Si/high-k gate stacks and in metal/high-k systems. The possible origin for anomalous behavior is also discussed, focusing on the dipole formation difference between the top and bottom interfaces.

6.1 Introduction

A central problem of high-k CMOS technology is that the threshold voltage (V_{TH}) of high-k MOSFETs cannot be described by the conventional formula. An unknown factor must always be taken into account to describe V_{TH}:

$$V_{TH} = \phi_{MS} + \delta\phi_X + 2\phi_F + \frac{\sqrt{2k_{Si}\varepsilon_0 q N_{sub}(2\phi_F)}}{C_{ox}}, \tag{6.1}$$

where ϕ_{MS} and ϕ_F are the work function difference between the metal and the semiconductor, the Fermi-potential in Si, respectively. $\delta\phi_X$ is an unknown term for describing V_{TH} obtained experimentally, and $\phi_{MS} + \delta\phi_X$ is generally called the effective work function, but it has not been clarified whether $\delta\phi_X$ is really derived from the work function modulation or not. k_{Si} is the dielectric constant of the semiconductor, ε_0 is the permittivity of vacuum, q is the magnitude of electron

A. Toriumi (✉)
The University of Tokyo, Tokyo 113-8656, Japan
e-mail: toriumi@material.t.u-tokyo.ac.jp

T. Nabatame
National Institute for Materials Science, Tsukuba 305-0044, Japan
e-mail: NABATAME.toshihide@nims.go.jp

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5_6,

charge, N_{sub} is the doping concentration in the substrate, and C_{ox} is the oxide capacitance density. Thus, the question is "what is the origin of $\delta\phi_X$ in Eq. (6.1)?" Since the flat-band voltage (V_{FB}) experimentally shows the same anomalous behavior as V_{TH}, the origin should be clarified by investigating V_{FB} instead of V_{TH}. A detailed electrostatic analysis and theoretical treatment of the interface and bulk trap charges, gate stack potentials, and flat-band and threshold voltages have been presented in Chap. 2.

A metal/high-k interface looks similar to a metal/wide band gap semiconductor interface. Thus, the Fermi-level pinning at the gate electrode/high-k interface was regarded as the origin of the anomalous V_{FB} shift [1, 2]. In particular, a poly-Si/HfO_2 gate stack was intensively investigated from the viewpoint of the formation of extrinsic charge neutrality level at the poly-Si/HfO_2 interface [3, 4]. The same kind of explanation was given to metal/high-k gate stacks [5]. Furthermore, the dipole formation at the metal/high-k interface due to the group electronegativity difference at the interface was also proposed as a reasonable extension of the above-mentioned model for anomalous V_{FB} [6]. For poly-Si/high-k [7] and metal/high-k gate stacks [8], charged oxygen vacancy formation in HfO_2 was also proposed as the origin of anomalous V_{FB}. The topics of threshold voltage tuning, charge neutrality level, oxygen vacancy model, interface dipole, and effective metal work function have also been analyzed in Chaps. 5 and 13.

Recently, a bottom interface model for anomalous V_{FB} based on the dipole formation at the high-k/SiO_2 interface was proposed [9, 10]. The validity of this model for various high-k materials has been demonstrated experimentally [11, 12]. Furthermore, plausible models for the bottom interface dipole formation have also been proposed very recently [13–15]. The oxygen vacancy/interstitial formation and electronegativity difference at the interface are considered to be the plausible origin of the bottom interface dipole formation.

What is the difference among the models proposed for the anomalous V_{FB} shift? And is the dipole located at the top or bottom interface? This chapter briefly reviews the various results so far reported for anomalous V_{FB} shift and discusses the models to consistently understand them.

6.2 Anomalous V_{TH} in Si-Gate/High-k MOSFETs

First, note that we should tolerate more bulk traps and interface states in high-k gate stacks than in SiO_2/Si systems. Some reports have not taken account of those effects in discussing the intrinsic V_{FB} anomaly. Those results may mislead us in understanding what happens in high-k gate stacks. In the first-order approximation, we can estimate intrinsic V_{FB} values from the linear extrapolation in the relationship between V_{FB} (experimental) and high-k thickness. Except in thick high-k films with a huge amount of bulk traps, the linear extrapolation method seems to be valid.

6.2.1 Poly-Si/High-k MOSFETs

In the early stage of high-k gate stack research and development, poly-Si gate electrode was considered because introducing two kinds of new materials to advanced CMOS seemed very dangerous. In addition, metal gate etching has been a serious concern in the fine gate patterning. However, the anomalous nature of V_{TH} in poly-Si/HfO_2 MOSFETs was unexpected. "Anomalous" refers to two characteristics of V_{TH}: one is the unexpected absolute value of V_{TH}, and the other is the asymmetric V_{TH} behavior between n^+poly and p^+poly-Si gates. A systematic study is shown in Fig. 6.1 [3, 4], where V_{FB} values in n^+ and p^+poly-Si gate stacks are plotted as a function of HfO_2 thickness on SiO_2 grown by atomic layer deposition (ALD). It is indicated that V_{FB} for p^+poly-Si gate shifts toward negative direction with HfO_2 thickness increase, while V_{FB} for n^+poly-Si gate is little changed.

The concept of the Fermi-level pinning in the metal/semiconductor and compound semiconductor interfaces is well known, though it is still a matter of controversy. Thus, the same view with the pinning parameter, S, and charge neutrality level, ϕ_{CNL}, was taken to understand the anomalous V_{FB}. From a phenomenological viewpoint, if ϕ_{CNL} is located near the conduction band edge with a strong pinning condition, the results in poly-Si/HfO_2 gate stack is understandable, although the physical origin of the ϕ_{CNL} is unclear. In contrast, it was also found that the n^+poly-Si gate showed the positive V_{FB} shifts with increase of Al_2O_3 thickness while the V_{FB} was almost same value for p^+poly-Si gate, as shown Fig. 6.2 [3]. In both cases, the top interface at poly-Si/high-k films seems to play a significant role in V_{FB} anomaly.

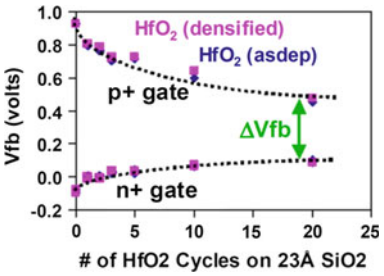


Fig. 6.1 V_{FB} values as a function of ALD-grown HfO_2 thickness on interfacial SiO_2 layer. The V_{FB} difference between n^+poly-Si and p^+poly-Si gate stacks sharply becomes small; particularly more significant change in the p^+poly-Si case is observed [3]

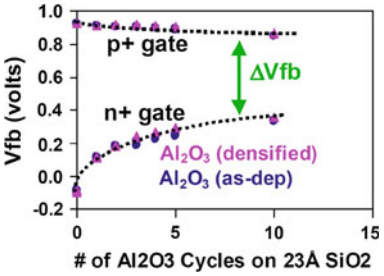


Fig. 6.2 V_{FB} values as a function of ALD-grown Al_2O_3 thickness on interfacial SiO_2 layer. The V_{FB} difference between n^+poly-Si and p^+poly-Si gate stacks sharply becomes small; particularly more significant change in n^+poly-Si case is observed [3]

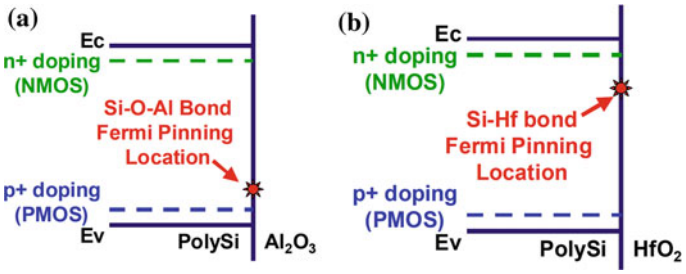


Fig. 6.3 Schematic view of the Fermi-level pinning position at **a** HfO_2/poly-Si and **b** Al_2O_3/poly-Si interfaces [3]

The Fermi-level pinning model at the poly-Si/high-k top interface was proposed in terms of the interfacial Si–Hf and Si–O–Al bonds for HfO_2 and Al_2O_3, respectively, as shown in Fig. 6.3 [3]. Especially, for Hf-based gate dielectrics, the presence of interfacial Si-Hf bonds or oxygen vacancies (V_o) (both are characteristic of oxygen-poor interface) strongly pins the Fermi-level just below the Si conduction band. On the other hand, for Al_2O_3 gate dielectrics, the pinning takes place just above the Si valence band and is not as strong as that for HfO_2.

Since the anomalous V_{FB} was severer in p^+poly-Si/HfO_2 than in n^+poly-Si/Al_2O_3, Al_2O_3 was introduced to HfO_2 so as to shift ϕ_{CNL} and then to tune V_{TH} to make it symmetric [16]. The $\phi_{m,eff}$ (effective metal work function) for n^+ and p^+poly-Si gates on four kinds of $HfAlO_x(N)$ are compared to those on SiO_2 as shown in Fig. 6.4. The charge neutrality level, ϕ_{CNL}, for four kinds of Al concentration in $HfAlO_x(N)$ is expected at the star symbol position, which is located both on the ideal line (broken line) and on the line representing the $\phi_{m,eff}$ on $HfAlO_x(N)$ versus the same on SiO_2 relationship between n^+poly-Si and p^+poly-Si. It is clearly seen that the charge neutrality level shifts toward higher $\phi_{m,eff}$ with increasing Al content.

It was also found that the V_{FB} difference between n^+ and p^+ poly-Si/HfSiON gate stacks increased as the Si content of Hf-silicates increases, as shown in

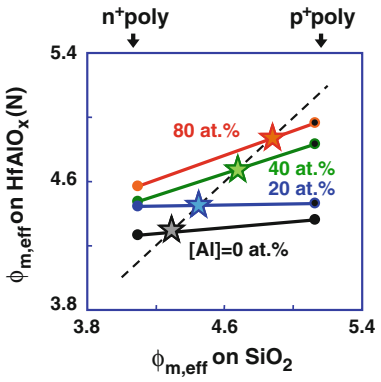


Fig. 6.4 Relationship between n^+ and p^+ poly-Si effective work function on $HfAlO_x$ and on SiO_2. Al concentration was changed from 0 to 80 at. %. A *star* symbol indicates the charge neutrality level for poly-Si/$HfAlO_x(N)$ gate stacks

Fig. 6.5 [17]. This suggests that the Fermi-level pinning effect may be reduced in Si-rich Hf-silicates by means of decreasing the number of Hf–Si bonds at poly-Si/Hf-silicates interface.

6.2.2 FUSI High-k MOSFETs

Fully silicided (FUSI) gates attracted much attention in terms of their process compatibility with conventional poly-Si technology in scaled CMOSFETs [18]. It was reported that another advantage of the FUSI process was V_{TH} controllability using the impurity segregation near the surface of the SiO_2 dielectric in FUSI/SiO_2 MOSFETs. To control V_{TH}, elements P, As and Sb were used for n-type, while B was used for p-type impurity doping. However, a structural/electrical instability was caused by a residual Si region resulting from the fluctuation of impurity doping during FUSI reaction, gate length dependent phase change of silicides, and Ni diffusion between n^+ and p^+poly-Si. Thus, we have to develop a pattern-dependent FUSI process, which is difficult to do for actual circuits with various gate lengths.

Furthermore, it was reported that the effect of the impurity segregation on V_{TH} control disappeared in MOSFETs with HfO_2 [19], as shown in Fig. 6.6. It seems to be due to the Fermi-level pinning at the FUSI/HfO_2 interface, which might affect V_{TH} more strongly than the impurity segregation, and the V_{TH} value becomes insensitive to the impurity.

It was found that Pt_3Si and Ni_3Si fabricated by changing the ratio of Pt/Si and Ni/Si, respectively, brought about a more significant recovery of V_{TH} compared to that for PtSi and NiSi gates [19]. Figure 6.7 shows that V_{FB} recovers from a strongly pinned to partially pinned state by increasing the Pt/Si ratio. The microscopic mechanism is not fully understood yet, but it is considered that the Fermi-level pinning strength would be reduced by decreasing the Si atom content at the interface, as schematically shown in Fig. 6.8a. The effect of SiN capping layer on V_{TH} in FUSI gates was also slightly observed and a possible mechanism is shown in Fig. 6.8b.

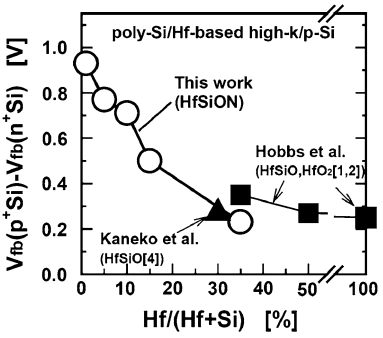


Fig. 6.5 Dependence of the V_{FB} difference between n^+poly-Si and p^+poly-Si/HfSiON gate stacks on Si concentration in Hf-silicate [17]

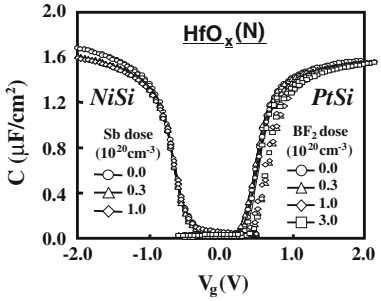


Fig. 6.6 Comparison of V_{FB} shift by introducing dopant impurities in NiSi and PtSi stacks on HfO_xN films. C is the MOS capacitance density and V_g is the gate voltage [19]

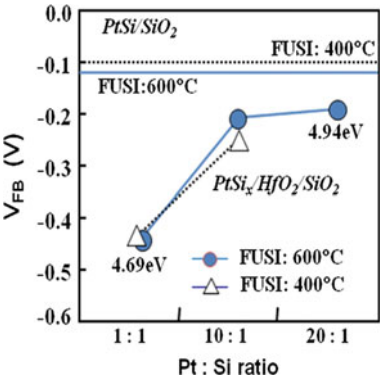


Fig. 6.7 V_{FB} shift of $PtSi_x/HfO_2/SiO_2$ gate stacks as function of Pt:Si ratio in $PtSi_x$ layer. Two kinds of FUSI formation temperature cases are shown [19]. The Pt:Si ratio is the nominal one in the sputtering, and the physical analysis (*TEM/EDX* Transmission Electron Microscope/Energy Dispersive X-ray Spectroscopy) indicated that PtSi was formed in the 1:1 Pt:Si ratio, while Pt_3Si was grown at the HfO_2 interface in the 10:1 and 20:1 cases. These are thermodynamically stable phases of Pt:Si system

It was also reported that $\phi_{m,eff}$ was controllable by changing the $NiSi_x$ stoichiometry such as Ni_3Si, NiSi and $NiSi_2$. It was called as the phase-controlled fully silicided (PC-FUSI) method, as shown in Fig. 6.9 [20]. The $\phi_{m,eff}$ for HfSiON gate stacks depends on the silicide phase: for Ni_3Si electrode, $\phi_{m,eff}$ increases about 0.4 eV compared to that for $NiSi_2$ electrode. It is likely that the $\phi_{m,eff}$ difference between Ni_3Si and $NiSi_2$ electrodes may be both due to the Fermi-level pinning relaxation by decreasing Si atoms at Ni-silicide/HfSiON interface and due to the change of work function of $NiSi_x$.

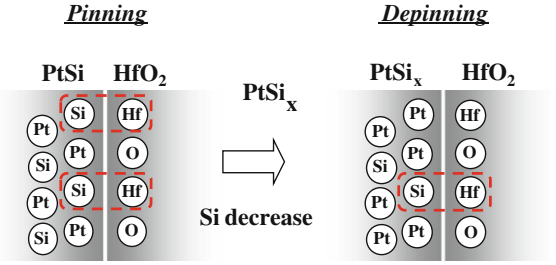


Fig. 6.8 Schematic explanation of the depinning mechanism in case of **a** the decrease of Si content in $PtSi_x$ and **b** insertion of SiN cap layer between PtSi and HfO_2. It seems that the Fermi-level pinning can be relaxed by reducing the bonding number of Hf–Si [19]

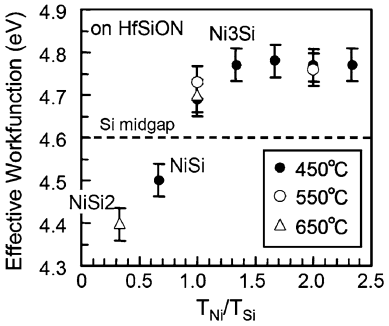


Fig. 6.9 Dependence of effective work function in $NiSi_x$/HfSiON gate stacks on Si concentration in $NiSi_x$. The effective work function seems to be determined by the Ni silicide phase rather than just the ratio of Ni to Si [20]

6.2.3 Oxygen Vacancy

As mentioned, the charged oxygen vacancy was reported as a plausible origin for anomalous V_{TH} shift [6]. So, the oxygen vacancy in HfO_2 is discussed in the following. The point is that higher formation energy of an oxide does not necessarily mean less oxygen vacancy formation. This is actually the case for HfO_2 as follows.

The theory related to the V_o formation in HfO_2 was proposed to understand anomalous V_{FB} shift in Hf-based high-k gate stacks with p^+poly-Si gates [7]. The mechanism of V_o formation in HfO_2 and subsequent electron transfer across the poly-Si/HfO_2 interface is schematically illustrated in Fig. 6.10. The poly-Si gate is partially oxidized by O atoms moving from HfO_2, which results in both SiO_2 growth at poly-Si/HfO_2 interface and V_o formation in HfO_2. Here, the electron entropy effect caused by the transition from neutral oxygen vacancy to charged one ($V_o \rightarrow V_o^{2+}$) should be taken into consideration. Two electrons occupy the V_o level which might be located around the bottom of Si conduction band. Indeed the V_o formation energy in HfO_2 is larger than that of SiO_2 by about 1.2 eV according to the first principles calculations [21], but in the case that V_o is ionized to V_o^{2+} and two electrons are released to the HfO_2 conduction band, the V_o formation energy

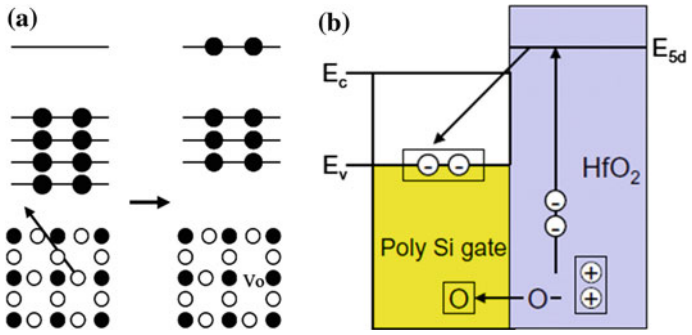


Fig. 6.10 Schematic model for the Fermi-level pinning at poly-Si/HfO_2 interface [7]. **a** The oxygen vacancy formation in HfO_2 moves up the electron energy level and **b** the charge transfer from HfO_2 to Poly-Si occurs. E_c is the conduction band edge and E_v is the valence band edge

becomes quite small thanks to the electron entropy effect. When HfO_2 is directly in contact with poly-Si gates, electrons in HfO_2 will favorably jump into p^+poly-Si gate with a certain energy gain than into n^+poly-Si gate, by assuming that V_o level is positioned around the conduction band of Si. As a result, the position of Fermi-level of p^+poly-Si gate will go up due to the interface dipole formation at p^+poly-Si/HfO_2 top interface and will result in the decrease of V_{FB}.

Namely, in this model, both the electron transfer from HfO_2 to poly-Si and the oxidation at HfO_2/poly-Si interface take place simultaneously at the p^+poly-Si/HfO_2 interface to minimize the total energy, resulting in charged oxygen vacancies forming in HfO_2.

After this view was proposed, significant roles of the oxygen vacancy in high-k films have frequently been discussed in many aspects. In that sense, it has quite a big impact on high-k gate stack technology not only for the V_{TH} issues but also the reliability properties and others. The experimental difficulty, however, lies in directly identifying and characterizing the oxygen vacancy quantitatively.

Let us revisit the experimental results in poly-Si/HfO_2 gate stacks in Fig. 6.1. With increasing HfO_2 thickness on SiO_2, the V_{FB} difference between n^+ and p^+poly-Si/HfO_2 gate stacks significantly decreases in spite of their 1 eV work function difference. If we do not assume bulk traps or interface charges differently depending on the doping property in poly-Si gates, it seems reasonable to consider the anomalous V_{FB} due to the above model based on the oxygen vacancy-induced positive charge formation in HfO_2. This implies the dipole formation at the top interface. In this model, it is interesting that the energy level of oxygen vacancy in high-k films plays an important role. On the other hand, it should be noted that results so far reported for poly-Si (FUSI)/high-k gate stacks include the bottom interface dipole contribution to the V_{FB} anomaly as will be discussed later. Thus, it is critical to subtract the bottom dipole contribution from the experimentally obtained anomalous V_{FB} for discussing the top Fermi-level pinning quantitatively.

6.3 Anomalous V_{TH} in Metal/High-k MOSFETs

6.3.1 Metal/High-k MOSFETs

Soon after the surprising report of the Fermi-level pinning at the poly-Si/HfO_2 interface, the top interface dipole formation model was also applied to metal/high-k MOSFETs [5], where the electronegativity difference between gate metal and Hf was a key to understanding anomalous V_{FB} behaviors, as schematically shown in Fig. 6.11. This view has been taken as one of the models that may explain the V_{TH} shift in metal/high-k gate stacks [6]. If this is the origin of anomalous V_{FB} in metal/high-k gate stacks, V_{FB} can be tuned by changing the top surface high-k material from HfO_2 to something else. On the basis of this viewpoint, a capping high-k technique to tune V_{FB} seems to have been successfully demonstrated [22]. However, the viewpoint of anomalous V_{FB} has been extending as discussed next.

Furthermore, the oxygen vacancy model to explain anomalous V_{FB} in metal/high-k gate stacks was proposed [8] (Fig. 6.12). The picture for this model is basically same as that shown in Fig. 6.10.

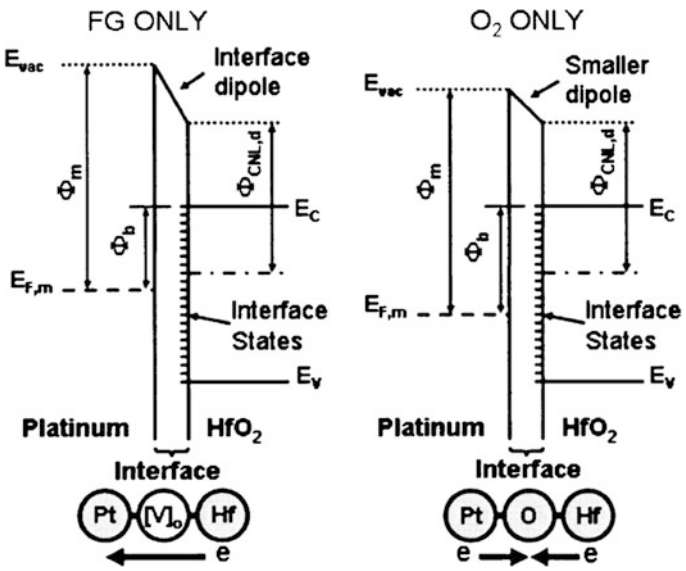


Fig. 6.11 Schematic views for interface dipole formation at a couple of metal/HfO_2 interfaces. In the case of Pt electrode, the post metallization annealing (*PMA*) condition may affect the dipole formation at the *top* interface [5]. E_{vac} is the vacuum level, ϕ_m is the metal work function, $E_{F,m}$ is the metal Fermi level, and ϕ_b is the barrier height

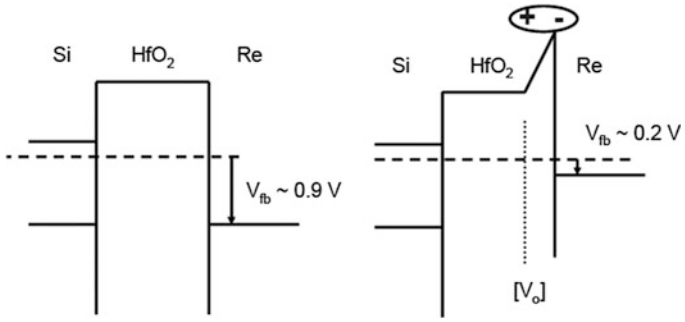


Fig. 6.12 Redox-induced V_{FB} shift model in metal/high-k gate stacks. Charged oxygen vacancy induced dipole formation at metal/high-k interface is proposed for anomalous V_{FB} shift observed in the experiment [8]

6.3.2 Interface Dipole at High-k/SiO_2 Interface

The typical structure of the high-k gate stack is shown in Fig. 6.13. To understand the V_{FB} anomaly, the Fermi-level pinning at the gate electrode/high-k interface has been so far discussed. The anomalous V_{FB}, however, can originate from several locations in this structure. In this part, we discuss a dipole effect at the high-k/SiO_2(IL: Interfacial Layer) interface (bottom interface).

It was recently claimed that the dipole layer at the $HfLaO_x/SiO_2$(IL) interface played a dominant role in the V_{FB} shift. Several kinds of bi-layer high-k films on SiO_2 with different La-concentration in $HfLaO_x$ were prepared. As shown in Fig. 6.14, V_{FB} in C–V curve is not determined by the top La concentration but by the bottom one in contact with SiO_2. And it was found that the V_{FB} shift was insensitive to the high-k layer thickness but was in proportion to the La concentration at the $HfLaO_x/SiO_2$ interface (not shown) [9]. In addition, the V_{FB} difference between Au/$HfLaO_x$ and Al/$HfLaO_x$ gate stacks was a constant close to the work function difference between Au and Al. This fact means that the top interface La concentration was not related to the V_{FB} shift in the $HfLaO_x$ system.

From those results, it was considered that this was not due to positive *or* negative charges but was due to an identical amount of positive *and* negative charges (dipole) located along the high-k/SiO_2(IL) interface. A schematic view of

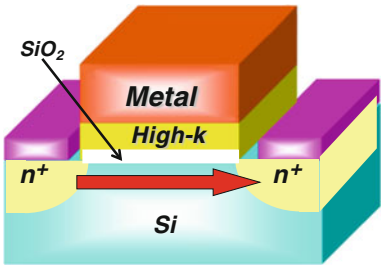


Fig. 6.13 Schematic of high-k MOSFET structure. Interfacial SiO_2 layer always exists between high-k film and silicon substrate

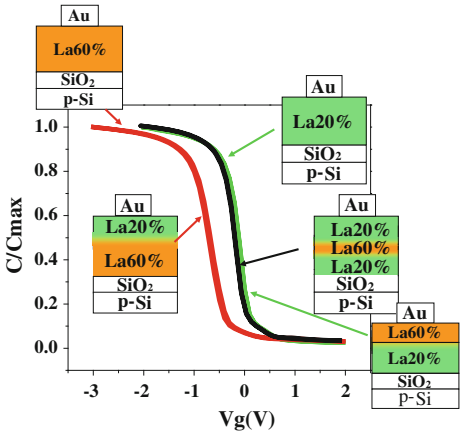


Fig. 6.14 C–V characteristics of MOS capacitors with $HfLaO_x$ dielectric films. La concentration was profiled in thickness direction. V_{FB} is characterized by *bottom* La concentration in contact with SiO_2-IL even for the tri-layer film case [15]

this gate stack system is shown in Fig. 6.15. This seems to be quite unusual because the dipole layer is formed at the insulator/insulator interface.

We further investigated several high-k materials on SiO_2(IL) from the viewpoint of the dipole layer formation at the bottom high-k/SiO_2 interface using the ALD technique, because in the Au/$HfLaO_x$/SiO_2(IL) gate stack in Fig. 6.14, high-k films were deposited by sputtering. It is important to clarify whether anomalous V_{FB} behaviors come from the particular processing step or from more general features of high-k/SiO_2 interfaces, and whether or not the behaviors are specific to La-oxide. A distinct difference should be observed between the dipole and trapped charges in the high-k film thickness dependence of V_{FB}. For dipoles, no thickness dependence of V_{FB} should be observed, while the effect of trapped charges can be detected as the film thickness dependence of V_{FB} depending on the charge distribution in the film. High-k films inherently contain traps, making it difficult in general to experimentally distinguish between dipole and trapped charge

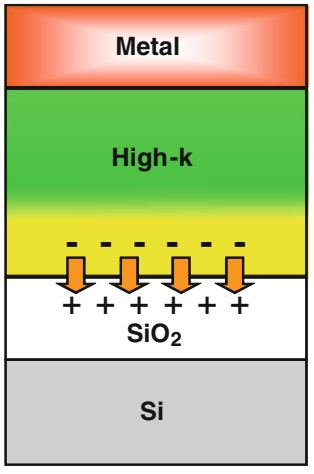


Fig. 6.15 Schematic representation of the *bottom* dipole layer formed at high-k/SiO_2 interface

behaviors. Therefore, high-k film thickness dependence must be investigated in detail to assign the V_{FB} anomaly as the dipole formed at high-k/SiO_2 interface.

Next, it is important to differentiate between the top and bottom interfaces. So, bi-layer high-k gate stack structures was used to determine which interface was critical for the V_{FB} shift. Figure 6.16 shows the V_{FB} shift in bi-layer high-k stacks grown by ALD as a function of both top and bottom high-k layer thicknesses independently [10]. In these experiments, both Al_2O_3/HfO_2/SiO_2/Si and HfO_2/Al_2O_3/SiO_2/Si gate stacks were employed. In Fig. 6.16a, only the top high-k layer thickness in the bi-layer structure is changed, while in Fig. 6.16b only the bottom one is changed for each bi-layer high-k gate stack. Figure 6.16a shows that V_{FB} values in HfO_2/SiO_2 and Al_2O_3/SiO_2 cases are not sensitive to the change of the top high-k material. On the other hand, V_{FB} in HfO_2 top case increases by inserting Al_2O_3 on SiO_2 and approaches the value in Al_2O_3/SiO_2 gate stack, while that in Al_2O_3 top case decreases by inserting HfO_2 on SiO_2, and approaches that in HfO_2/SiO_2 one. The results demonstrate unquestionably that the V_{FB} shift is caused not by the top-layer high-k insertion but by the bottom one in contact with SiO_2. Namely, the V_{FB} is significantly shifted by the bottom high-k layer material on SiO_2, and it gives us a clear evidence of the dipole formation at the high-k/SiO_2 (bottom) interface. Those results also point out that no dipoles are formed at HfO_2/Al_2O_3 interface. Therefore, most V_{FB} anomalies reported for the metal gate/high-k/SiO_2/Si system should be explained by considering that the direction and magnitude of the dipole are dependent on the high-k material deposited on SiO_2.

Furthermore, by defining the normalized V_{FB} (V_{FB}^{Norm}) in Fig. 6.16b by

$$V_{FB}^{Norm} = \frac{V_{FB}^{exp} - V_{FB}^{\infty}}{V_{FB}^{0} - V_{FB}^{\infty}} = exp\left(-\frac{T_{high-k}}{\lambda}\right), \tag{6.2}$$

it was also elucidated by using ALD that the dipole was formed in a monolayer high-k deposition from the plot of V_{FB}^{Norm} as a function of the bottom high-k film

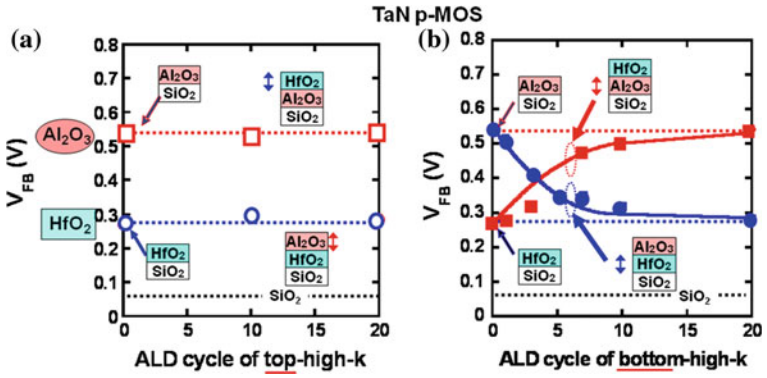


Fig. 6.16 V_{FB} shift of bi-layer high-k stack (Al_2O_3 and HfO_2) MOS capacitors with TaN gate electrode as function of ALD cycle of **a** *top* and **b** *bottom* high-k layers. Possible high-k layer dependent effect at top interface is not likely to be related to V_{FB} shift in this experiment

thickness as shown in Fig. 6.17, where $V_{FB}^{\exp}$, V_{FB}^{0} and V_{FB}^{∞} are V_{FB} values obtained experimentally at a given bottom high-k layer thickness, and a single layer high-k gate stack with no bottom high-k layer, and bi-layer high-k gate stack inserted in 20 ALD cycles on SiO_2 in Fig. 6.16b, respectively. The characteristic length is about 4 Å both for Al_2O_3 and HfO_2 [23]. This fact means that the dipole is not formed by the long-range atom diffusion but within the monolayer high-k at the interface. It is worth mentioning that the atomically precise control of high-k film thickness is possible using the ALD technique, in which one cycle of the ALD process enables the deposition of about 0.1 nm-thick film. This has enabled measurement of the gradual change of V_{FB} as a function of the bottom-layer high-k thickness. Thus, we can conclude that the V_{FB} shift is governed by the bottom high-k layer material in contact with SiO_2, within a distance of one or two monolayers in this system.

Here, it is worth mentioning that the dipole layer formation at the bottom interface should be considered even in the case of HfO_2/SiO_2 gate stack, though it has often been ignored in discussions. An interesting point observed experimentally is that the interface dipole direction of La_2O_3/SiO_2 or Y_2O_3/SiO_2 is opposite to that of Al_2O_3/SiO_2 or HfO_2/SiO_2.

This bottom layer dipole quantitatively explains most of the V_{FB} anomalies reported for the metal gate/high-k/SiO_2/Si system. By taking both the strength and direction of the dipole moment of interfacing high-k dielectrics into account, it should, in principle, be possible to adjust the V_{TH} values in both n- and p-MOSFETs in the gate last process including the high-k first process. In that sense, the dipole concept sheds new light on how V_{FB} is determined in metal/high-k MOS capacitors.

The internal dipole has not been taken into account for CMOS device design so far, but in practice the dipole-based correction to V_{TH} becomes much more important in setting an appropriate V_{TH} for advanced high-k CMOS particularly under a lower V_{DD} operation. The capping high-k technique might be understandable from the bottom dipole layer formation [12, 14].

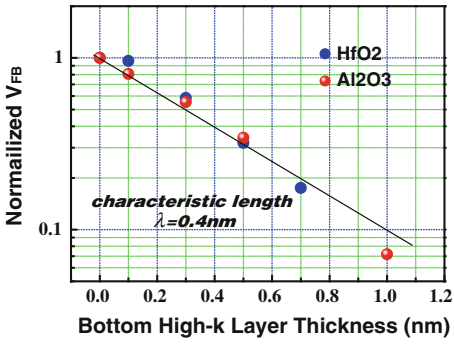


Fig. 6.17 V_{FB} shift (normalized V_{FB}) is simply described by exponential decay as a function of *bottom* high-k layer thickness with characteristic length of ~0.4 nm for both HfO_2 and Al_2O_3 [23]

6.3.3 Dipole Formation Model at High-k/SiO_2 Interface

Experimental evidence of the dipole formed at a high-k/SiO_2 interface has been obtained, but the physical origin of the dipole formation has not been clarified yet. In this part, we discuss a model for the dipole formation at high-k/SiO_2 interface. The dipole necessarily involves the same amount of positive and negative charges at the interface. Thus, it would not be appropriate to ascribe it to coincidental numbers of positive and negative charges at the interface.

The oxygen density has been paid attention to rather than metal ionicity at the high-k/SiO_2 interface because the common atom for both sides of the materials at the interface is not metal but oxygen [13]. Oxygen should bond to both Si and metal at the interface. In other words, oxygen is an interfacing atom and the oxygen density at the interface should be adjusted to be energetically consistent with both SiO_2 and high-k. The oxygen density of various oxides including SiO_2 is shown in Table 6.1. Note that there is a significant volume difference for each oxide "molecule" when we consider the volume of a unit structure defined by its chemical formula. This leads to the areal density difference of oxygen atoms between high-k and SiO_2 at the interface. The number of oxygen atoms per unit area (σ) is approximated by $V_u^{-2/3}$, where V_u is defined as the volume of the structure containing a single oxygen atom. The calculated ratio of $\sigma_{\text{High-k}}$ to σ_{SiO_2} is listed in Table 6.1.

A schematic diagram of dipoles for various high-k/SiO_2 interfaces in terms of V_{FB} shift is shown in Fig. 6.18, and the relative oxygen density for the various oxides listed in Table 6.1 is shown in Fig. 6.19 [13]. Of interest is the apparent coincidence between the results shown in Figs. 6.18 and 6.19. This suggests that the oxygen density difference might be a strong driving force for the dipole formation at the high-k/SiO_2 interface.

Next an intuitive explanation to understand the dipole formation process will be presented in terms of the oxygen density difference at the interface [13, 15]. The simple model is as follows. A charged oxygen atom can be transferred from the higher to the lower oxygen density side, and the resultant charge imbalance at the interface will induce the dipole. In other words, this is a vacancy-interstitial pairing formation at the interface. This model can reasonably explain the formation of an equal amount of charges with opposite sign at the interface. A slight motion of the charged oxygen will substantially shift a net charge distribution at the interface from the two respective neutral oxides. This net charge redistribution

Table 6.1 Ratios of $\sigma_{\text{High-k}}$ to σ_{SiO_2} estimated for various oxides [15]

Oxide	SiO_2	HfO_2	AI_2O_3	Y_2O_3	La_2O_3
Density (ρ) [g/cm^3]	2.2	9.6	4.0	4.8	6.5
Unit structure	$Si_{1/2}O$	$Hf_{i/2}O$	$AI_{2/3}O$	$Y_{2/3}O$	$La_{2/3}O$
V_u [Å^3]	22.7	18.3	14.2	26.0	27.7
σ/σ SiO_2	1	1.2	1.37	0.91	0.88

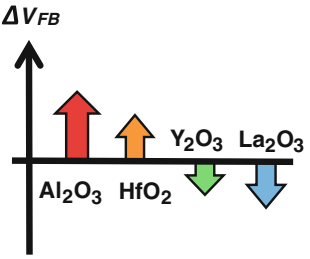


Fig. 6.18 Schematic representation of anomalous V_{FB} shift observed experimentally. *Upward* and *downward arrows* indicate apparent V_{FB} shift direction from the normal SiO_2 gate stack [15]

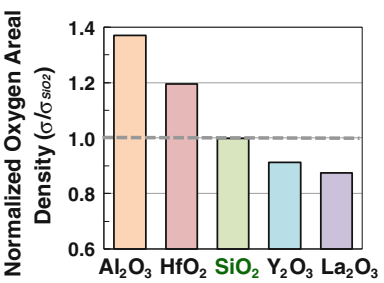


Fig. 6.19 Relative areal oxygen density of various oxides at interfaces normalized by that of SiO_2, which are calculated using their molar volume as shown in Table 6.1 [15]

seems to be identical to the dipole layer at the interface. The actual bonding structure should certainly be complicated from the atomistic viewpoint, but the net dipole layer is naturally formed at the high-k/SiO_2 interface due to the oxygen density difference.

It is considered that the increase of structural energy at the hetero-interface would be suppressed by the "dipole" formation, although the dipole formation increases the electronic energy. This is understandable by considering that the total free energy composed of electronic and structural components should be minimized to stabilize the interface, as shown in Fig. 6.20 [15].

On the other hand, it might be possible that the dipole formation at high-k/SiO_2 interface is associated with the interfacial oxygen coordination [24]. This model is actually also applicable for the oxygen density difference one mentioned previously. However, the interfacial oxygen coordination is not the driving force but the result of the dipole formation, so only from this view it is difficult to explain in general the dependency of the dipole direction on the high-k material. Furthermore, the metal electronegativity (*EN*) difference model between high-k and SiO_2 has been considered [13–15]. Pauling's *EN* value is obviously inappropriate for discussing the metal *EN* in ionic oxides because most of the electrons at metal atoms are already transferred to the oxygen due to the high iconicity of high-k oxides. Thus, Sanderson's *EN* [25] for compounds was considered. In Sanderson's definition, the effects of surrounding oxygen atoms on the *EN* of a metal (the principle of *EN*-equalization) are considered. The equalized *EN* value is calculated by the geometric average of all of the individual atoms in the oxide. It is, however, concluded that Sanderson's *EN* difference cannot explain the experimental results of negative V_{FB} shift because all high-k materials have smaller *EN* values than that

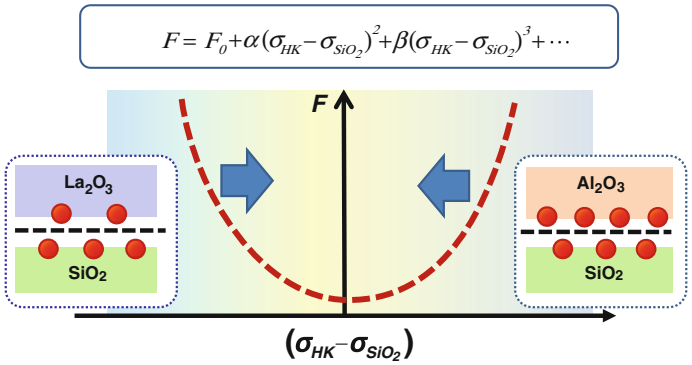


Fig. 6.20 Schematic description of the free energy, F, for high-k/SiO_2 interface as a function of $(\sigma_{HK} - \sigma_{SiO_2})$. The linear term of $(\sigma_{HK} - \sigma_{SiO_2})$ is ignorable because both *upward* and *downward* dipoles with opposite direction are observed. F_0, α and β are constants [15]

of SiO_2, irrespective of the definition of electronegativity. Thus, we rule out the dipole model on the basis of the *EN* difference between high-k and SiO_2, irrespective of the *EN* definition.

Nevertheless, it is interesting to discuss the relationship between our model and the *EN* model [15]. The oxygen density difference actually corresponds to the ionic radius difference, which is directly related to the *EN*. The relationship between Pauling's *EN* and the normalized oxygen areal density is shown in Fig. 6.21. The order of metal is basically identical to the oxygen density trend except for Si. The most important point here is that ionic radius differs in the meaning between high-k materials and SiO_2. The ionic radius in high-k oxides is determined by random close packing (RCP) of oxygen, while that in SiO_2 is determined by a continuous random network structure (CRN) through covalent bonding in a tetrahedron unit formation. The normalized oxygen areal density as a function of the ionic radius is shown in Fig. 6.22, and a qualitative difference between high-k oxides and SiO_2 is clearly indicated. The dipole direction and magnitude can be explained with regard to the SiO_2 line in the present model. This is a critical difference between our model and the *EN* one for explaining the dipole formation.

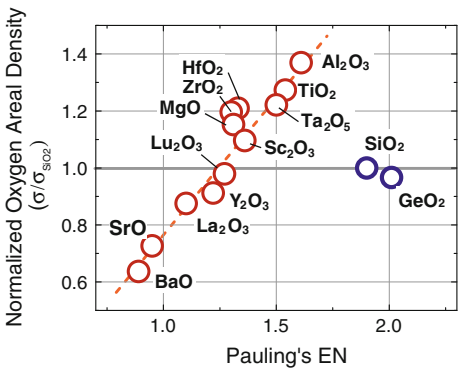


Fig. 6.21 Normalized oxygen areal density as function of Pauling's electronegativity in various oxides. Note that all high-k oxides are on the same line, while SiO_2 does not follow the trend [15]

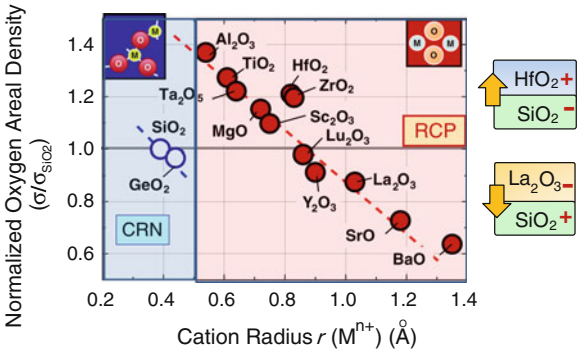


Fig. 6.22 Normalized oxygen areal density as function of cation ionic radius in various oxides. SiO_2 is far from the trend. A big difference between high-k oxides and SiO_2 is the structure forming oxides. High-k films are generally formed in the RCP (randomly close packing) while SiO_2 is in the (continuously random network) CRN. This difference is considered to play a critical role for forming the dipole at high-k/SiO_2 interface [15]

6.4 Top or Bottom Interface Dipole?

Because the bottom interface dipole is real, subtraction of the bottom contribution from the total V_{FB} shift is required to estimate the top effect on V_{FB} [11]. The relationship of the effective work function on high-k (HfO_2, Al_2O_3, and Y_2O_3) and on SiO_2, by changing gate metals, is shown in Figs 6.23a and b. There are unity slopes for the high-k films in Fig. 6.23a, which means a very little Fermi-level pinning, if any, at the top for these films (the pinning parameter $S \sim 1$). A finite offset from the ideal line corresponds to the magnitude and direction of the dipole for each high-k film on SiO_2, when the effective work function on high-k films is compared with that on SiO_2.

However, a significant degradation of the slope for Si-gates including FUSI is observed (Fig. 6.23b), where corrected $\phi_{Si,eff}$ means the effective work-function after subtracting the bottom dipole effect for each high-k dielectric on SiO_2. This indicates that Si has unique top interface effects. By decreasing Si content in $PtSi_x$, the effective work function on high-k shifts back to the ideal line as discussed in Sect. 6.2. Thus, V_{TH} will be generally described as follows.

$$V_{TH} = \phi_{MS} + \phi_{FLP}^{Top}(Si) + \phi_{Dipole}^{Bottom} + 2\phi_f + \frac{\sqrt{2k\varepsilon_0 q N_{Sub}(2\phi_f)}}{C_{ox}} \qquad (6.3)$$

Because we have only investigated the top dipole effect of limited kinds of metals on V_{FB} in the gate last process, we cannot simply conclude that only the bottom dipole works for all of the metals. In fact, the top dipole observation with X-ray Photoelectron Spectroscopy (XPS) has been reported [26]. Nevertheless, we think that the predominant contribution to V_{FB} comes from the bottom dipole in

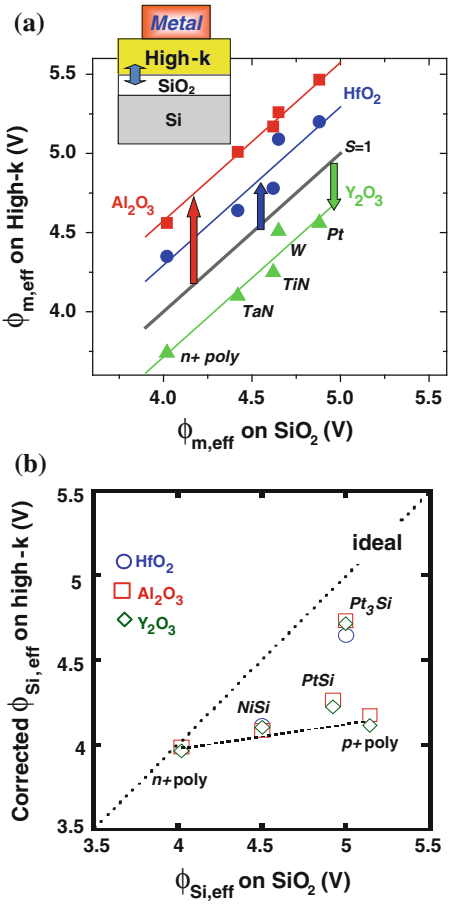


Fig. 6.23 **a** Relationship between $\phi_{m,eff,high-k}$ and ϕ_{m,eff,SiO_2} for various metal gates including n^+ poly-Si. The $\phi_{m,eff,high-k}$ shifts on Al_2O_3, HfO_2, and Y_2O_3 are well fitted by the parallel shift of "ideal-line" ($S = 1$) [15]. **b** Relationship between $\phi^C_{Si,eff,high-k}$ and ϕ_{m,eff,SiO_2}, where $\phi^C_{Si,eff,high-k}$ means effective work function on high-k film after correction by *bottom* dipole contribution to V_{FB}. It is interesting to see that $\phi^C_{Si,eff,high-k}$ with a given gate electrode material has almost the same value regardless of high-k material [11]

metal/high-k gate stacks with relatively thick SiO_2 interfacial layer, because almost the same V_{FB} shifts as those in Fig. 6.16 have been observed in the gate-last/high-k first process [11]. From this viewpoint, the results shown in Sect. 6.3.1 might not be general, because the redox annealing could affect the oxygen content of high-k films significantly in the case of particular metals. Moreover, even in the case of those metals, it was discussed that the bottom interface was directly related to the V_{FB} anomaly, though further systematic investigation is needed for understanding particular metal gate stack systems.

6.5 Summary

We conclude that a predominant contribution to anomalous V_{TH} is a dipole layer formed at the high-k/SiO_2 interface in metal/high-k gate stacks, while a substantial effect from the top interface on V_{TH} should be taken into consideration in

Si-related gates. Here, it should be noted again that the bi-layer high-k stack structure enabled us to clearly differentiate the contributions from the top and bottom interfaces to the V_{FB} shift.

At any hetero-interfaces, mismatches between two materials are expected. In one case, the electron transfer from one material to the other will relax the mismatch, while in another, a structural relaxation by atom motion at the interface will occur. In general, both may be possible. The issue is which is predominant. In high-k gate stacks, both views have been proposed for both the top and bottom interfaces.

It is strongly suggested that the origin of the top interface dipole is the charged oxygen vacancy formation in HfO_2, and that the origin of the bottom one is the oxygen movement model at the high-k/SiO_2 interface due to the oxygen density difference. Because the model is not sensitive to details of the high-k materials, it is substantially quite useful for predicting V_{TH} shift for new high-k oxides. The bottom interface dipole is the dipole formed at the ionic oxide/covalent oxide and has not so far been taken into account for SiO_2-based CMOS device design. Quantitative understanding of this interface is critically important both for setting the appropriate V_{TH} in advanced CMOS devices under a low V_{DD} (supply voltage) operation and for delving further into related material science.

Finally we would like to make a couple of comments. Experiments have been performed under various conditions by many researchers. We have mainly focused on relatively thick interfacial SiO_2 and high-k layers to understand what happens at the interfaces in metal/high-k gate stacks. With further scaling of the gate stack, all layers should be thinned. Thus it becomes very difficult to discriminate among various interfaces, because the interface in gate stacks come close to each other and the interface-interface interaction should be taken into account. The V_{TH} roll-off [27] is one of those examples. This will be the next challenge.

Acknowledgments We would like to thank collaboration members in The University of Tokyo (UT) and in the MIRAI project, particularly Koji Kita (UT) for his contribution to the bottom dipole formation model, and Masataka Hirose (MIRAI) for his useful discussion and continuous encouragement. This work was partly supported by a Grant-in-Aid for Scientific Research from MEXT in Japan and partly conducted in the MIRAI Project supported by NEDO.

References

1. J. Robertson, J. Vac. Sci. Technol. **B18**, 1785 (2000)
2. Y.C. Yeo, P. Ranade, T.J. King, C. Hu, IEEE Electron Dev. Lett. **23**, 342 (2002)
3. C.C. Hobbs, L. Fonseca, V. Dhandapani, S. Samavedam, B. Taylor, J. Grant, L. Dip, D. Triyoso, R. Hegde, D. Gilmer, R. Garcia, D. Roan, L. Lovejoy, R. Rai, L. Hebert, H. Tseng, B. White, P. Trobin, Symp. VLSI Tech. Dig. **9** (2003)
4. C.C. Hobbs, L.R.C. Fonseca, A. Knizhnik, V. Dhandapani, S.B. Samavedam, W.J. Taylor, J.M. Grant, L.G. Dip, D.H. Triyoso, R.I. Hegde, D.C. Gilmer, R. Garcia, D. Roan, M.L. Lovejoy, R.S. Rai, E.A. Hebert, H.-H. Tseng, S.G. H. Anderson, B.E. White, P.J. Trobin, IEEE Trans. Electron Dev. **51**, 971–978 (2004)

5. J.K. Schaeffer, L.R.C. Fonseca, S.B. Samavedam, Y. Liang, P.J. Tobin, B.E. White, Appl. Phys. Lett. **85**, 1826 (2004)
6. J.K. Schaffer, D.C. Gilmer, C. Capasso, S. Kalpat, B. Taylor, M.V. Raymond, D. Triyoso, R. Hegde, S.B. Samavedam, B.E. White Jr, Microelectron. Eng. **84**, 2196 (2007)
7. K. Shiraishi, K. Yamada, K. Torii, Y. Akasaka, K. Nakajima, M. Konno, T. Chikyo, H. Kitajima, T. Arikado, Symp. VLSI Tech. Dig. 108 (2004)
8. E. Cartier, F.R. McFeely, V. Narayanan, P. Jamison, B.P. Linder, M. Copel, V.K. Paruchuri, V.S. Basker, R. Haight, D. Lim, R. Carruthers, T. Shaw, M. Steen, J. Sleight, J. Rubino, H. Deligianni, S. Guha, R. Jammy, G. Shahidi, Symp. VLSI Tech. Dig. **230** (2005)
9. Y. Yamamoto, K. Kita, K. Kyuno, A. Toriumi, Jpn. J. Appl. Phys. **46**, 7251 (2007)
10. K. Iwamoto, Y. Kamimuta, A. Ogawa, Y. Watanabe, S. Migita, W. Mizubayashi, Y. Morita, M. Takahashi, H. Ota, T. Nabatame, A. Toriumi, Appl. Phys. Lett. **92**, 132907 (2008)
11. Y. Kamimuta, K. Iwamoto, A. Ogawa, Y. Watanabe, S. Migita, W. Mizubayashi, Y. Morita, M. Takahashi, H. Ota, T. Nabatame, A. Toriumi, Tech. Dig. IEDM **341** (2007)
12. P.D. Kirsch, P. Sivasubramani, J. Huang, C.D. Young, M.A. Quevedo-Lopez, H.C. Wen, H. Alshareef, K. Choi, C.S. Park, K. Freeman, M.M. Hussain, G. Bersuker, H.R. Harris, P. Majhi, R. Choi, P. Lysaght, B.H. Lee, H.H. Tseng, R. Jammy, T.S. Boscke, D.J. Lichtenwalner, J.S. Jur, A.I. Kingon, Appl. Phys. Lett. **92**, 092901 (2008)
13. K. Kita, A. Toriumi, Appl. Phys. Lett. **94**, 132902 (2009)
14. H. Jagannathan, V. Narayanan, S. Brown, ECS-Trans. **16**(5), 19 (2008)
15. A. Toriumi, T. Nabatame, ECS-Trans. **25**(6), 3 (2009)
16. M. Kadoshima, A. Ogawa, M. Takahashi, H. Ota, N. Mise, K. Iwamoto, S. Migita, H. Fujiwara, H. Satake, T. Nabatame, A. Toriumi, Symp. VLSI Tech. Dig. **70** (2005)
17. M. Koyama, Y. Kamimuta, T. Ino, A. Kaneko, S. Inumiya, K. Eguchi, M. Takayanagi, A. Nishiyama, IEDM Tech. Dig. **499** (2004)
18. J. Kedzierski, E. Nowak, T. Kanarsky, Y. Zhang, D. Boyd, R. Carruthers, C. Cabral, R. Amos, C. Laovie, R. Roy, J. Newbury, E. Sullivan, J. Benedict, P. Saunders, K. Wong, D. Canaperi, M. Krishnan, K.-L. Lee, B.A. Rainey, D. Fried, P. Cottrell, H.-S.P. Wong, M. Ieong, W. Haensch, IEDM Tech. Dig. **247** (2002)
19. T. Nabatame, M. Kadoshima, K. Iwamoto, N. Mise, S. Migita, M. Ohno, H. Ota, N. Yasuda, A. Ogawa, K. Tominaga, H. Satake, A. Toriumi, IEDM Tech. Dig. **83** (2004)
20. K. Takahashi, K. Manabe, T. Ikarashi, N. Ikarashi, T. Hase, T. Yoshihara, H. Watanabe, T. Tatsumi, Y. Mochizuki, IEDM Tech. Dig. **91** (2004)
21. W.L. Scopel, Antonio J.R. da Silva, W. Orellana, A. Fazzio, Appl. Phys. Lett. **84**, 1492 (2004)
22. V. Narayanan, V.K. Paruchuri, N.A. Bojarczuk, B.P. Limder, B. Doris, Y.H. Kim, S. Zafar, J. Stathis, S. Brown, J. Arnold, M. Copel, M. Steen, E. Cartier, A. Callegari, P. Jamison, J.-P. Locquet, D.L. Lacey, Y. Wang, P.E. Batson, P. Ronsheim, R. Jammy, M.P. Chudzik, M. Ieong, S. Guha, G. Shahidi, T.C. Chen, Dig. Symp. VLSI Tech. **178** (2006)
23. A. Toriumi, K. Kita, K. Tomida, Y. Zhao, J. Widiez, T. Nabatame, H. Ota, M. Hirose, Tech. Dig. IEDM **53** (2007)
24. O. Sharia, A.A. Demkov, G. Bersuker, B.H. Lee, Phys. Rev. **B75**, 035306 (2007)
25. R.T. Sanderson, *Chemical Bonds and Bond Energy*, 2nd edn. (Academic Press, New York, 1976)
26. S. Miyazaki, H. Yoshinaga, A. Ohta, Y. Akasaka, K. Shiraishi, K. Yamada, S. Inumiya, M. Kadoshima, Y. Nara, ECS Trans. **13**(2), 67 (2008)
27. B.H. Lee, J. Oh, H.H. Tseng, R. Jammy, H. Huff, Mater. Today **9**, 32 (2006)

Chapter 7
Channel Mobility

Chadwin Young

Abstract Effective channel mobility, μ_{eff}, of high-k gate dielectrics in various device technologies is discussed in detail. Initially, the background on mobility is provided with brief explanations on different scattering mechanisms. Next, the fast transient charge trapping effect, seen in hafnium-based dielectric films, is explained in conjunction with the pulse I–V technique, which is used to quantify the fast trapping. This fast trapping is shown to degrade device performance and thereby degrade the μ_{eff}. Then, how the components of high-k gate dielectric stack structure (i.e., interfacial layer quality and composition, high-k layer thickness and Hf content) affects the fast trapping is discussed. It is shown that increases in high-k layer thickness or Hf content lead to mobility degradation due to increased levels of fast trapping. However, an increase in the thickness of the interfacial layer, that screens the high-k dielectric from the device channel, can dramatically improve the mobility. Unfortunately, increasing the equivalent oxide thickness may not be ideal for some logic technologies. In order to understand the factors, there is a need to extract the mobility in the presence of fast trapping. Several methodologies of μ_{eff} extraction that remove the fast transient contribution to the μ_{eff} are discussed. Finally, the chapter discusses future technology possibilities: substrate orientation dependent mobility enhancement, process induced strain, germanium-based channels, and compound semiconductor substrates. All approaches demonstrate mobility enhancement over standard bulk silicon substrates, but they can also introduce other aspects of mobility degradation.

C. Young (✉)
Materials Science and Engineering Department, University of Texas at Dallas, 800 W. Campbell Road, RL 10, Richardson, TX 75080, USA
e-mail: chadwin.young@utdallas.edu

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5_7,

7.1 Introductory Remarks and Brief Outline of the Chapter

This chapter investigates device mobility, and the way high-k gate dielectrics can impact this mobility. The first several sections will explain mobility in metal-oxide-semiconductor field effect transistors (MOSFETs). The next sections discuss the impact of hafnium-based films on mobility. Here, focus is placed on the physical thickness, the hafnium content, and the underlying interfacial layer and their impact on the fast transient charge trapping aspect of these gate stacks. Next, mobility extraction techniques that remove the fast trapping effect on mobility are covered. The remaining sections discuss high-k mobility in relation to the current and potential mobility-boosting technologies such as substrate orientation, non-planar silicon, silicon germanium, and III–V compound devices.

7.2 Background on Mobility and the Different Carrier Scattering Mechanisms

The mobility of carriers in bulk materials has been well categorized. However, *effective* mobility involves the movement of carriers near the surface of the semiconductor. Effective mobility is a key parameter in evaluating transistors because the drive current and the device speed are directly proportional to it in MOSFETs, the movement of carriers in the inversion layer constitutes this mobility. Here inversion carriers traverse the surface of the semiconductor from source to drain at the interface between the dielectric and substrate (Fig. 7.1). The vertical electric field component brings the inversion charge to the surface of the semiconductor while the electric field between the source and drain creates a lateral inversion carrier flow. As carriers traverse the channel, they are subjected to

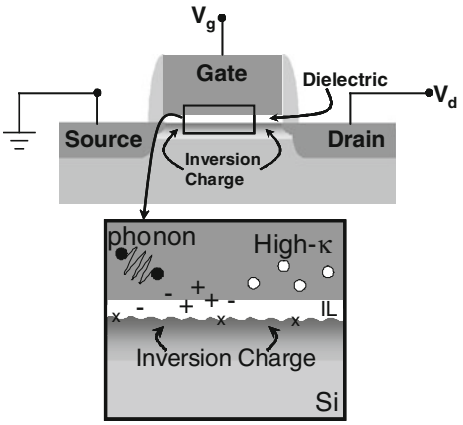


Fig. 7.1 Cross section of a MOSFET with a blow-up of the boxed channel region that contains several mobility scattering mechanisms (e.g., phonon scattering; coulomb scattering—fixed charges (−, +), interface traps (×), and high-k fast transient traps; and surface roughness)

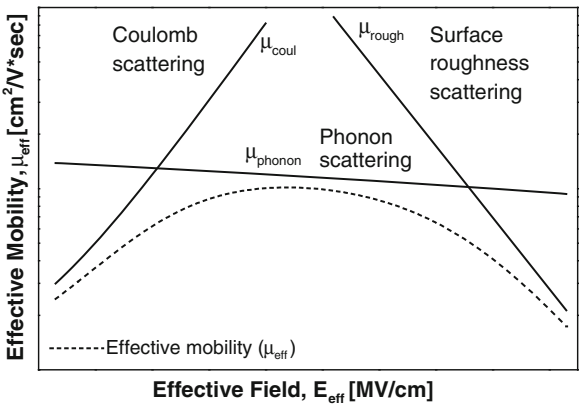


Fig. 7.2 Universal mobility characteristic versus the effective field at the silicon surface showing the effects of coulomb, phonon, and surface roughness scattering at different effective field strengths

various scattering mechanisms that make the effective mobility lower than the bulk material mobility. The effective carrier mobility, μ_{eff}, is defined as

$$\mu_{eff} = \frac{g_d}{\left(\frac{W}{L}\right) Q_{inv}} \tag{7.1}$$

where the channel conductance $g_d = \Delta I_D / \Delta V_D$, Q_{inv} is the inversion charge, W is the channel width, and L is the channel length. Effective mobility is typically plotted against effective field E_{eff}, which is defined as

$$E_{eff} = \frac{Q_{dep} + \eta Q_{inv}}{\varepsilon_s} \tag{7.2}$$

where Q_{dep} is the depletion charge and η is 1/2 for electrons and 1/3 for holes [1]; ε_s is the permittivity of silicon.

How different scattering mechanisms impact μ_{eff} as a function of the effective field is illustrated in Figs. 7.1 and 7.2 [2]. Figure 7.1 illustrates the inversion layer and various scattering mechanisms that impact μ_{eff} in a cross section of a MOSFET transistor. As can be seen in Fig. 7.2, μ_{eff} is affected by coulomb scattering, μ_{coul}, at the low field regime, while at high fields, surface roughness scattering, μ_{rough}, dominates. In the moderate effective field regime, μ_{eff} is subject to the effect of phonon scattering, μ_{phonon}. The Matthiessen's rule allows these scattering mechanisms to be included in the calculation of effective mobility:

$$\frac{1}{\mu_{eff}} = \frac{1}{\mu_{coul}} + \frac{1}{\mu_{phonon}} + \frac{1}{\mu_{rough}} \tag{7.3}$$

7.2.1 Phonon Scattering

Phonons can be characterized as quantum harmonic oscillators representing atomic vibrations in crystal lattices. These phonons can interact with electrons or holes

through scattering mechanisms that exchange kinetic energy and momentum just like particle interactions (i.e., collisions). These scattering mechanisms are instrumental in understanding electrical conductivity in crystalline solids [3].

7.2.2 Surface Roughness Scattering

As stated above, inversion carriers traverse the channel near the interface between the substrate and the dielectric. The roughness of a non-ideal interface (i.e., inherent roughness) scatters electrons particularly when a relatively large gate-to-channel vertical electric field is present in the MOSFET. This "pulls" the inversion charge to the interface as it attempts to travel across the device channel when there is a lateral electric field, caused by the potential difference between the source and drain. The inherent roughness resulting from a given fabrication process acts as scattering centers to inversion carriers, degrading the channel mobility more than the bulk mobility. Therefore, surface roughness scattering dominates at relatively high vertical electric field values.

7.2.3 Coulomb Scattering

Coulomb scattering is the result of several possible mechanisms: charged ionized impurity atoms, interface trapped charges, and oxide trap charges that are relatively near the device channel. The channel carriers can be scattered by attractive and repulsive forces of the charged traps that impede the progress of inversion channel carriers traversing the channel from source to drain. However, at a high enough electric field, the high density of inversion layer carriers can essentially "screen" away the effects of coulomb scattering [4]. Therefore, coulomb scattering has a greater impact at low vertical electric fields.

7.3 Different Factors of Mobility Degradation in High-k Gate Dielectrics

The fast transient charging effect (FTCE), due to substrate-injected electrons being trapped at sites just below the hafnia conduction band, is a major contributor to the instability and mobility degradation seen with pulsed current–voltage (I–V) characterization [5–13]. Factors such as reduction of the inversion charge from electron trapping in the high-k film [14, 15], quality of the interfacial SiO_2 layer [16], phonon scattering [17, 18], crystallization of the high-k [19, 20], and the impact of metal gates [21] have been studied. In addition, the effect of charge

trapping on the extraction of intrinsic channel mobility must be understood. As stated in [9, 12, 14, 15, 22], measured capacitance–voltage (C–V) data and DC drain current—gate voltage, $I_D - V_{G,}$ data include the effects of charges trapped in the high-k dielectric, which causes the inversion charge to be overestimated and, consequently, the extracted DC-based mobility to be lower than the correct ("intrinsic") mobility. In several studies in the literature [9, 15, 18, 22], corrections have been made for the effects of charge trapping when determining the inversion charge before the intrinsic mobility is extracted.

7.3.1 Fast Transient Charge Trapping in the High-k Bulk

High-k gate dielectrics can demonstrate the FTCE [12], which happens when substrate electrons are injected to and captured by shallow traps located just below the Hf-based dielectric conduction band [23]. Trapping may occur within hundreds of nanoseconds to microseconds [24, 25]. During the FTCE process, the trapped charges are well-localized within the dielectric, with their spatial distribution through the dielectric thickness being determined by the Fermi level of the channel electrons and the trap energy (presumably, a neutral or single-charged negative oxygen vacancy) [26–31]. Fast trapping, which is reversible with no residual damage [5, 8, 32], can severely impact device performance, causing threshold voltage (V_T) instability and mobility degradation in Hf-based gate stack nMOS transistors.

The pulsed I – V methodology was introduced to quantify the trapped charge during fast transient charging (FTC) as well as to measure "intrinsic" (i.e., trapping-free) device characteristics [5, 33–35]. Figure 7.3a and b illustrate drain current–gate voltage ($I_D - V_G$) (pulse voltage domain) and I_D-time (pulse time domain) characteristics, respectively, wherein the change of I_D during a "single" pulse measurement can be determined. The V_g pulse, along with its corresponding drain voltage (V_D) response, is captured and converted into I_D–V_G or I_D-time. From this, the hysteresis in the $I_D - V_G$ curve (Fig. 7.3a) or the decrease of I_D in the I_D-time plot (Fig. 7.3b) illustrates the change in threshold voltage due to charge trapping.

Figure 7.4 illustrates pulsed mobility (to be discussed in Sect. 7.4.3), extracted from the rise (i.e., trapping-free) and fall (i.e., post-trapping) time portions, showing similar effective mobility values. Even after 100 μs of charging time, the falling $I_D - V_G$ curve appears to be parallel to the rising curve, suggesting no significant impact due to remote coulomb scattering from the trapped electrons (Fig. 7.4, inset). This negligible effect on intrinsic mobility can be understood when one considers that the trapped charges are located in the bulk of the high-k [8, 36], relatively far from the channel.

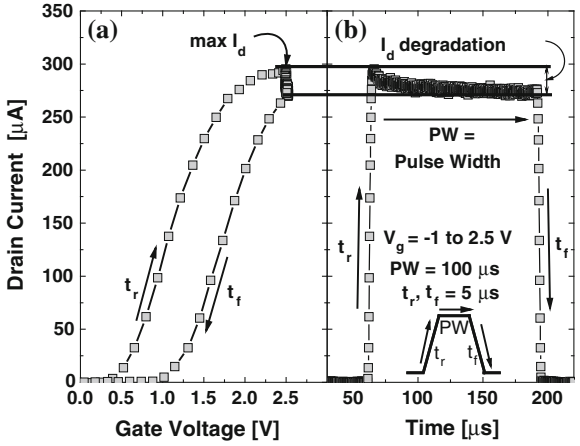


Fig. 7.3 Various forms of pulsed I_D data in the **a** pulse voltage, or **b** time domains based on the "single" pulse approach where the rise time (t_r) portion of the I_D trace represents a "trapping-free" characteristic. Results clearly show that I_D degrades at the top of the pulse (i.e., pulse width portion) due to FTC

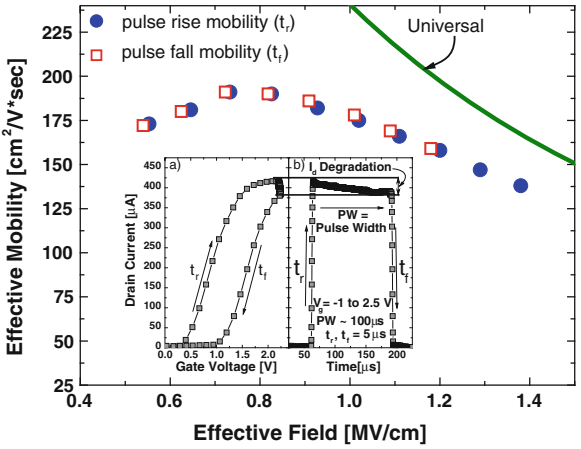


Fig. 7.4 Similar mobility values extracted from the rise (trapping-free) and fall (post-trapping) portions of the single pulse measurement (inset, a) demonstrating that remote coulomb scattering is negligible in high-k gate stacks due to bulk trapping beyond the 1 nm SiO_2-like interfacial layer

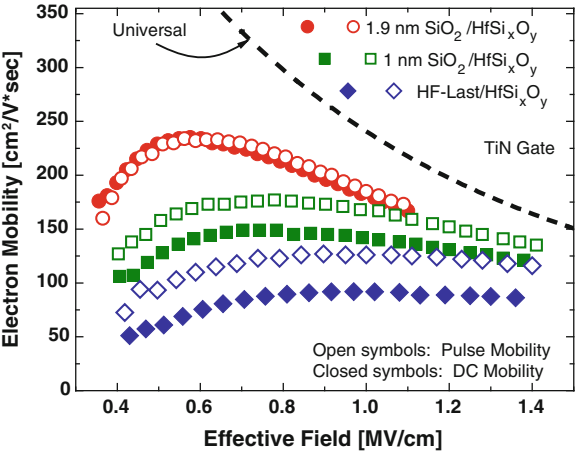


Fig. 7.5 Application of the pulse mobility extraction method [36] to gate stacks with interfacial layers (ILs) of varying thicknesses and fixed $HfSiO_x$ dielectric layer

7.3.2 Interface Quality

The impact of interface quality on effective mobility has been studied extensively [6, 13, 16, 19, 37]. For high-k dielectrics, the fabrication process typically creates an interfacial layer (IL). The quality, composition, and thickness of the IL can have a significant impact on the extracted mobility.

7.3.2.1 Quality/Composition

Most reports on high-k films directly deposited on silicon substrates have shown severe mobility degradation due to the formation of a poor quality interface between the dielectric and the substrate. Some of the factors attributed to this degradation include fluctuation of the k-value in the channel, soft optical phonon scattering induced from the overlying high-k layer [17], high interface state density, and substantial surface roughness.

To restore mobility, the interface between the dielectric and the substrate must be pristine and of high quality. Traditionally, this high quality interface has been created by thermally growing a SiO_2 gate dielectric. Hence, some of the best performing interfacial layers are SiO_2-like layers. However, compared to conventional SiO_2, this type of a layer has limits. Remember that in subsequent processing a high-k layer, typically hafnium-based, is deposited on top of this SiO_2-like layer with additional thermal cycles. In this situation, oxygen can be leached from the SiO_2 layer, making it oxygen-deficient and non-stochiometric. This can lead to fixed charges and electron traps near the device channel, which degrade mobility [16].

7.3.2.2 Interfacial Layer Thickness

While the presence of an interfacial layer can enhance mobility compared to no IL at all, the IL may not be an ideal SiO_2 dielectric. To study this, gate stacks with a fixed $HfSiO_x$ layer and various interfacial layers [38] were investigated to evaluate the non-electron trapping contribution to mobility degradation. Figure 7.5 illustrates the effect of interfacial layer thickness on the extracted effective mobility. DC and pulsed $I_D - V_G$ data were collected on these samples, and mobility was subsequently extracted from both sets of data. (Pulsed $I_D - V_G$ mobility extraction techniques will be discussed in Sect. 7.4.) The sample with a 1.9 nm SiO_2 interfacial layer shows essentially the same DC and pulsed mobility since with such a thick interface, fast transient charging is negligible, as indicated by similar DC and pulsed $I_D - V_G$ curves. As the interface becomes thinner, the difference between the pulsed and DC mobility for a given gate stack becomes greater. The overall degradation of the pulsed (intrinsic) mobility as the interface layer becomes thinner can be attributed to the poorer quality of the interfacial layer [16] and greater effects of soft optical phonons [18] and crystallinity [19, 20].

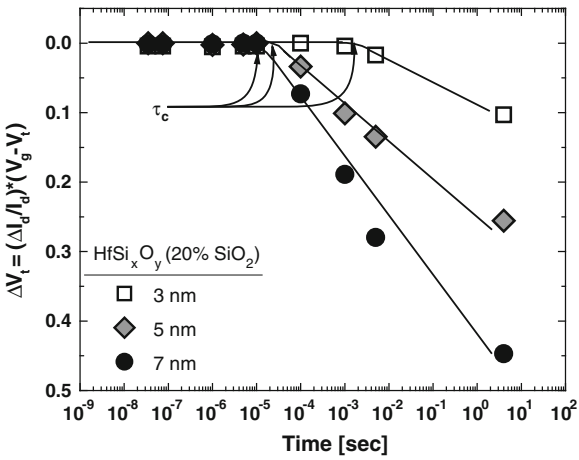


Fig. 7.6 Normalized ΔV_T for varying high-k physical thicknesses. Greater physical thickness of the high-k is subject to faster and more severe degradation

7.3.3 High-k Dielectric Properties

High-k gate dielectrics can have a significant impact on extracted mobility. Mobility degradation from fast charge trapping, phonon scattering [17, 18], and crystallization of the high-k [19, 20] were evaluated to explore this impact.

7.3.3.1 High-k Layer Thickness

Experimental results have shown that thicker high-k films lower the extracted mobility when using conventional DC mobility extraction techniques such as the split C–V (Capacitance–Voltage). As high-k physical thickness decreases, electron trapping at a given pulse duration also decreases (see Fig. 7.6), pointing to a scaling of the high-k film physical thickness as a potential solution to the charge trapping problem. The effective threshold voltage shift (ΔV_T) increases with increasing gate bias (Fig. 7.7) since electrons are trapped closer to the Si-substrate [39]. Lower ΔV_T and longer τ_c for thinner films are correlated with higher DC mobility (Fig. 7.8). Therefore, reducing the thickness of the high-k layer improves the mobility.

7.3.3.2 Hafnium Content

Incorporating more Hf, which was varied from pure HfO_2 to 80 % HfO_2, 60 % HfO_2, to pure SiO_2 (i.e., 0 % Hf), correlates to greater electron trap density (Fig. 7.9). This correlation between the electron trap density, roughly proportional to ΔV_T, and the Hf content, reflects an increase in the total volume of HfO_2 resulting from phase separation in the high-k film after a 1,000 °C dopant activation anneal [15]. The impact of Hf content on bias dependence and device

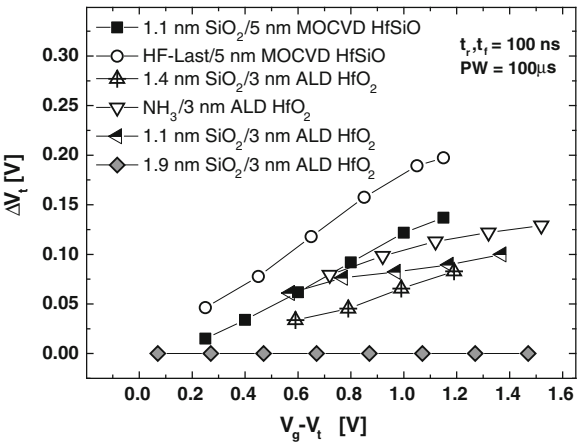


Fig. 7.7 ΔV_T versus pulse voltage adjusted to the transistor threshold voltage for the various gate stacks illustrating a ΔV_T increase with increasing V_G

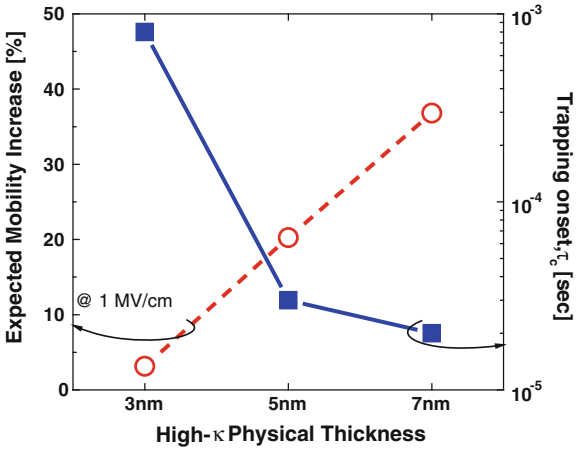


Fig. 7.8 Expected mobility increase (thinner high-k → larger pulse/DC $I_{d,lin}$ ratio) as a function of high-k layer thickness compared with the onset of the critical trapping time, τ_c, (i.e., the time for the onset of fast transient charging trapping)

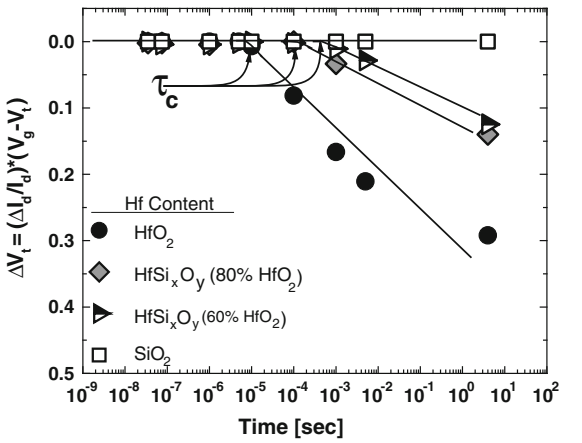


Fig. 7.9 Normalized ΔV_T for varying amounts of hafnium; more Hf results in faster and greater degradation; however, introducing a small amount of SiO_2 significantly reduces ΔV_T

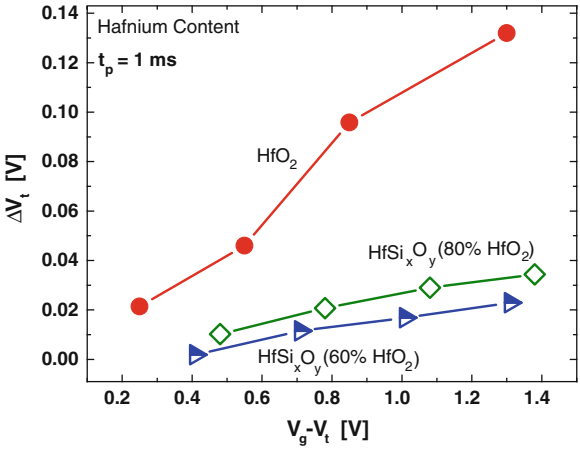


Fig. 7.10 Calculated ΔV_T for varying amounts of Hf as a function of $V_G - V_T$. Incorporating SiO_2 minimizes the increase in ΔV_T with increased gate bias compared to HfO_2 only

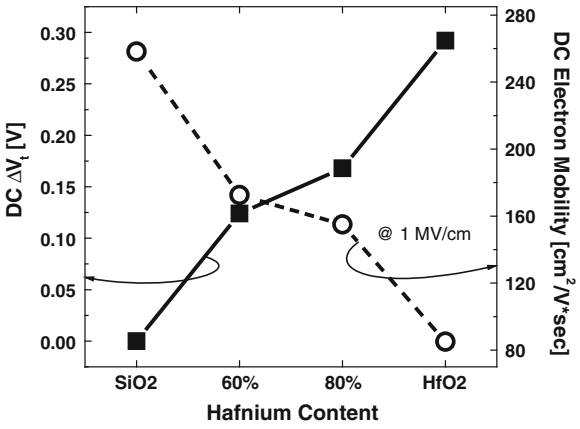


Fig. 7.11 DC ΔV_T increases and measured DC mobility decreases as the Hf content increases

degradation is summarized in Figs. 7.10 and 7.11, respectively, demonstrating that a silicate with a small amount of SiO_2 can significantly reduce ΔV_T and improve performance. The onset of fast transient charging trapping is mostly determined by the projected density of the electron traps in the high-k film, while the effect of charging (in particular, ΔV_T) is controlled by multiple factors such as the relative volume of HfO_2 in the film, high-k thickness, etc.

7.4 Different Techniques for the Extraction of Channel Mobility

In the linear regime, FTCE can have a significant impact on the extracted mobility [9, 15, 18, 22, 36, 37, 40–44]. In a conventional DC (Direct Current) sweep of the $I_D - V_G$ measurement, the threshold voltage can shift as the measurement

progresses, thereby decreasing the drive current at each bias sweep point and decreasing mobility. Several FTC correction techniques discussed in the literature can correct for this effect [9, 15, 36, 37].

7.4.1 Split C–V and Conventional I_D-V_G Correction Technique

The FTC correction techniques discussed below are derived from investigating high-k phenomena that can impact mobility extraction when using the split C–V technique [45] and extracting E_{eff} and μ_{eff} [15]. Measured CV data include trapped and inversion charges, which overestimate inversion carrier density and lower apparent mobility. The fundamental approach is to correct the measured split C–V and the gate-to-channel capacitance C_{gc} (integrated to get Q_{inv}), using a modeled (ideal) C_{gc}. By comparing the experimental C_{gc} with the ideal C_{gc}, the amount of trapped charge (Q_t) can be determined and subtracted from the measured Q_{inv} to provide the intrinsic mobility, which is benchmarked to the Hall mobility measurement.

7.4.2 Direct Measurement of Inversion Charge Using Charge Pumping (CP)

The fast transient mobility extraction algorithm in [46] demonstrates the impact of the trapped charge on the mobility of high-k gate stacks. The procedure uses charge pumping (CP) and pulsed $I_D - V_G$ measurements. CP was used to obtain the inversion charge, Q_{inv}; the rise time of the fast transient pulsed $I_D - V_G$ values was used to calculate the channel conductance, g_d. Use of the pulsed $I_D - V_G$ and inversion charge pumping as outlined in [46] allow the free-carrier mobility to be estimated. The CP measurement involves measuring a long channel device using fast rise and fall times (e.g., 5 μs) to take advantage of the geometric effect. This measures inversion charge as well as trapped charge because a long channel device with fast rise and fall times does not allow the inversion charge to release back to the source and drain (geometric effect), therefore becoming part of the measured charge pumping current, I_{cp}. The variable amplitude CP technique measures the trapped charge, qN_t, which subtracted from the long channel measurement charge yields the charge of the free inversion carriers, N_{inv}. Figure 7.12 illustrates the results of this approach for a MOCVD (Metal Organic Chemical Vapor Deposition) Hf-based gate stack. The figure clearly demonstrates an increase in the peak and high field mobility, with the high field mobility being close to the universal mobility when using this mobility extraction approach. For this particular gate stack, then, most of the observed mobility degradation is associated with charge trapping effects, which shift the flatband and threshold voltages of the device.

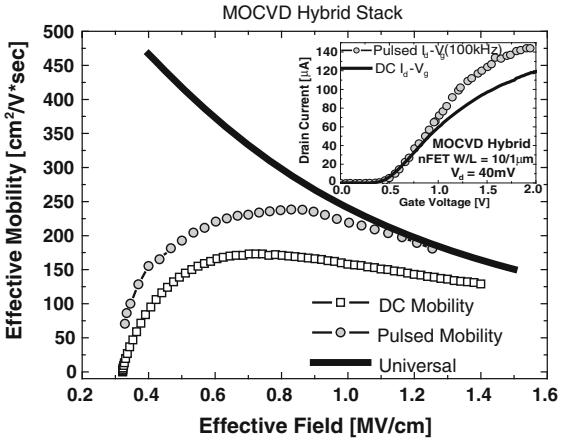


Fig. 7.12 Comparison of electron mobility from pulsed/CP methodology and conventional split C–V for a MOCVD hybrid stack. Inset: comparison of pulsed $I_D - V_G$ to DC $I_D - V_G$ for the mobility shown

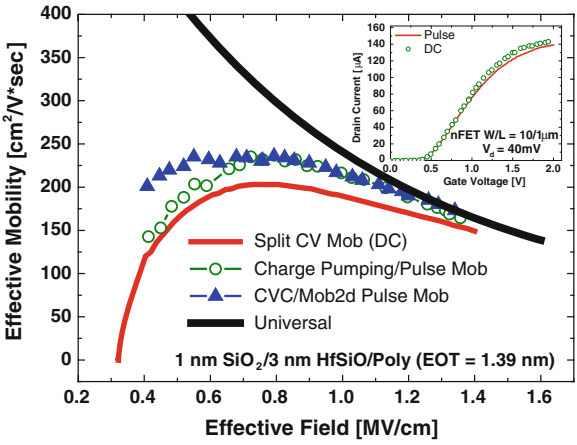


Fig. 7.13 Comparison of DC mobility to the novel pulse mobility technique for the $I_D - V_G$ data in the inset. The mobility values extracted using the proposed mob2d pulse methodology and the direct measurement of N_{inv} technique [46] are in excellent agreement

7.4.3 Pulsed I–V with Model Fitting and Parameter Extraction

One technique that requires only two measurements and minimal data analysis was presented in [36]. In this approach, pulsed $I_D - V_G$ is combined with North Carolina State University's CVC [47] and mob2d [1] modeling parameter extraction software. To extract the mobility, mob2d performs a non-linear, least squares fit to the experimental pulsed I_D data in conjunction with the equivalent oxide thickness (EOT) and the substrate doping data supplied from the CVC. EOT and substrate doping are negligibly impacted by FTC [48]. Therefore, the result is a trap-free mobility extraction. Figure 7.13 compares the impact of the trapped charge in the DC mobility data to the trap-free, pulsed-based extraction for the $I_D - V_G$ data in the inset, showing the pulse-based mobility is higher than its DC-based counterpart.

7.5 Process Optimization

Previous sections have discussed the importance of scaling the high-k layer and/or incorporating an interfacial layer to improve mobility. However, to sustain the scaling required to meet the International Technology Roadmap for Semiconductors (ITRS), continued process optimization is required. Because Chaps. 3, 4, 5, 9 and 10 discuss high dielectric constant materials, only a brief discussion is provided here with an emphasis on the mobility.

7.5.1 High-k Layer

In Sect. 7.3.3.1, the results clearly show that thinner high-k gate dielectrics improve mobility. However, the high-k gate dielectric was introduced to allow a thicker dielectric layer to ultimately reduce the gate leakage current. To slow the high-k scaling trend, a "higher-k" dielectric may be needed that would allow for a slightly thicker film, thereby reducing the gate leakage for the future technology nodes. Ti doping of hafnium-based dielectrics is a possible solution. Compared to the undoped dielectrics [49], Ti doping has been found to boost the k-value while maintaining mobility.

More recently, LaO_x and AlO_x capping has been reported to achieve the proper nMOS and pMOS threshold voltage values, respectively [50–57]. Processing was optimized to maintain mobility, unlike corresponding non-capped devices.

7.5.2 Interfacial Layer

The interfacial layer has also been subject to further investigation and optimization. In LaO_x-capped samples, La was identified in the IL, which in some cases has impacted mobility. However, having a SiON interfacial layer instead of a SiO_x layer prevented La from diffusing into the IL while also reducing the overall EOT due the higher k-value from incorporating nitrogen [50]. Moreover, compared to SiO_x, SiON improved mobility. Recent work has even looked at removing the interfacial layer [58].

7.6 High Mobility Channels and Substrates

As the scaling trend continues for future technology nodes, the need to incorporate high mobility channels has become significant. While effective channel mobilities have been higher in these substrates than in standard (100) silicon, these types of devices can have more scattering mechanisms than the standard ones (see Sect. 7.2).

7.6.1 Orientation Dependent Mobility Enhancement

Theory predicts and experiment verifies that taking advantage of surface orientation improves inversion carrier mobility. For example, pMOS devices with SiON gate dielectrics have been shown to have better mobility on (110) Si than on the standard (100) Si. However, nMOS devices have higher mobility values on (100). This orientation dependent mobility is also demonstrated with HfO_2 as the dielectric in Fig. 7.14.

Therefore, integrating n-channel MOSFETs (nFETS) on Si(100) and pFETS on (110) Si on the same wafer to maximize both electron and hole mobility, known as hybrid orientation technology (HOT) [59], has been demonstrated to be a production-worthy option. In one study, both HfO_2 and HfSiON dielectrics have been atomic layer deposited on (110) Si to study the mobility enhancement when using high-k dielectrics [60]. This work showed that HfO_2 hole mobility, μ_h, was significantly higher than HfSiON μ_h on (110) Si. On (100) Si, however, the μ_h of HfO_2 and HfSiON is similar. When comparing the carrier mobility on (110) Si to that on (100) Si, the (110) Si electron mobility enhancement was $\sim$0.5 times for both HfO_2 and HfSiON; however, μ_h enhancement was about 3.3 times and 2.9 times greater for HfO_2 and HfSiON, respectively. The authors conducted cryogenic temperature measurements to study the improved HfO_2 and HfSiON mobility on (110) Si. Here, the temperature dependence of μ_h at 1 MV/cm is much greater on Si(110) than on Si(100) for both dielectrics, which demonstrates that phonon scattering is more significant on Si(110) [61].

Furthermore, since the temperature dependence was greater for HfO_2 than for HfSiON, HfO_2 introduces an additional phonon scattering mechanism. This difference in dependence was attributed to the interaction of inversion channel carriers to "remote" phonons in the dielectric where Hf content plays a role. More

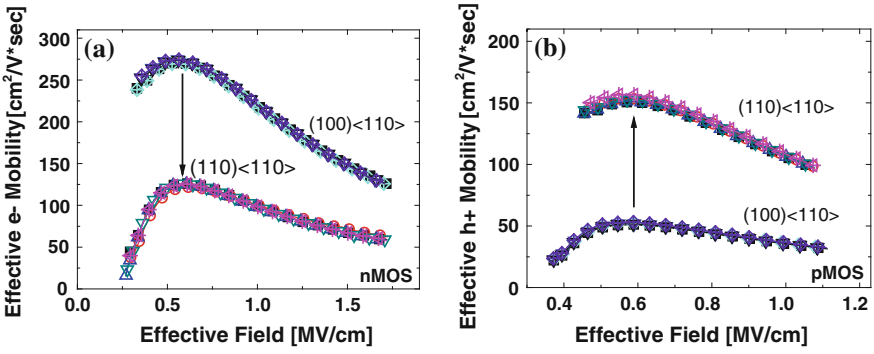


Fig. 7.14 Comparison of planar silicon: several **a** nMOS and **b** pMOS mobility data with (100) and (110) channel orientations, respectively, demonstrating repeatable results. As reported many times in the literature, nMOS mobility degrades when switching from (100) to (110) while pMOS mobility is enhanced

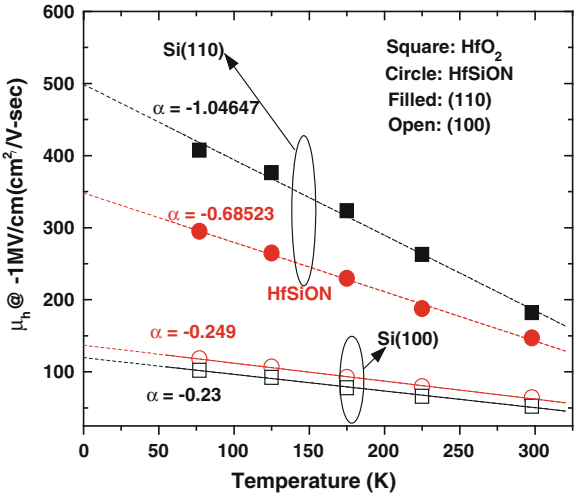


Fig. 7.15 Comparison of high field hole mobility versus T (K) for HfSiON and HfO_2 with different orientations. Si(110) devices have a higher dependence on temperature where α-HfO_2/Si(110) is greater than α-HfSiO/Si(110) [60]

remote phonon scattering is seen due to the higher Hf concentration in HfO_2 than in HfSiON [17, 18, 21]. At 77 K, a temperature at which phonon impact is minimal, HfO_2 mobility is appreciably higher than HfSiON, indicating greater coulomb and surface roughness scattering in HfSiON lowered its mobility, despite enhanced phonon scattering in HfO_2. Positive fixed charges can increase with the amount of Hf in the dielectric [62]. In (110) Si, these positive fixed charges push dominant light holes away from the interface, creating a "buried" channel and thereby increasing hole mobility. Also, the low effective mass electrons in (100) Si degrade mobility more in HfO_2 than in HfSiON, which could be the result of the Coulomb interactions of electrons near the interface, ultimately increasing the scattering. This could lead to the mobility "degradation" typically reported for channel electrons in high-k/(100) Si gate stacks [9, 15, 19, 36, 37, 40] (Fig. 7.15).

Although HOT is production-worthy, it would be much more costly than fabrication on only one orientation. Another investigation demonstrated comparable performance for nMOS on (110) and (100) orientations [63]. This work argues that nMOS low field mobility on long channel devices does not necessarily correlate to short channel nMOS saturation current performance, and shows that nMOS saturation drive currents for (100) and (110) orientations are comparable (Fig. 7.16) while the long channel mobility differed as typically seen (see Fig. 7.14) [63]. Thus, Fig. 7.17 demonstrates that fabricating on (110) orientation only, is possible because the expected pMOS enhancement is realized and nMOS (110) orientation is comparable to nMOS (100) orientation.

Future technology nodes may call for the use of three-dimensional, multi-gate FETs. These devices can typically have silicon "fin" structures etched down to buried oxides (BOX) on silicon-on-insulator wafers. The gate electrode can wrap around the three sides of the fin creating two sidewall channels if the topside of the fin has a hardmask (inset, Fig. 7.18a), which is sometimes referred to as a FinFET. The sidewall crystallographic orientation is typically (110); however, fabricating

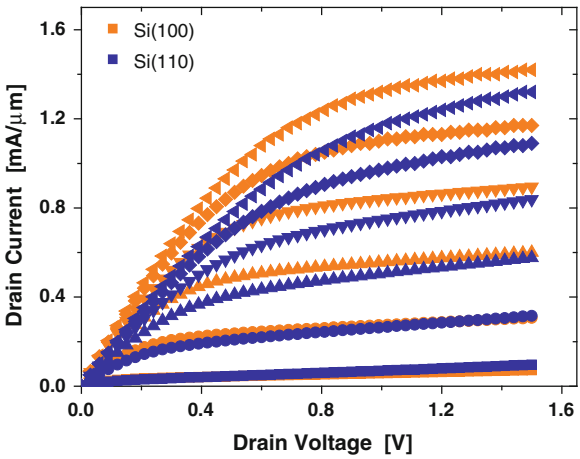


Fig. 7.16 Short channel nMOS $I_D - V_D$ characteristics for different surface orientations where comparable performance is seen for either orientation [63]

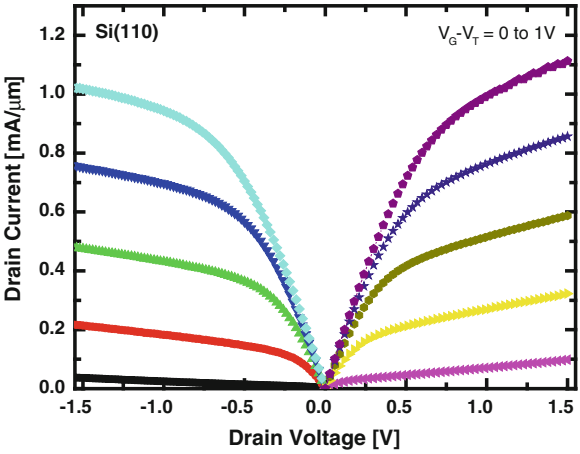


Fig. 7.17 Short channel CMOS $I_D - V_D$ characteristics for (110) surface orientation demonstrating quite symmetric performance on a single channel orientation [63]

devices rotated 45° from vertical can produce a sidewall crystallographic orientation of (100). Using the split-CV methodology, but taking into account the FinFET device structure in the inset of Fig. 7.18a, the effective mobility expression becomes:

$$\mu_{eff} = \frac{\Delta g_d}{\left(\frac{M \cdot 2 \cdot H_{fin}}{L_g}\right) \cdot Q_{inv}} \tag{7.4}$$

where M is the number of fins and H_{fin} is the height of the fin (actual device width). Figure 7.18 shows the sidewall mobility characteristics on p- and n-channel FinFETs with different sidewall surface orientations. Contrary to the results in Fig. 7.14a, FinFETs on either (100) or (110) have electron mobility values that are quite close to each other. This suggests that the FinFET conventional (110)

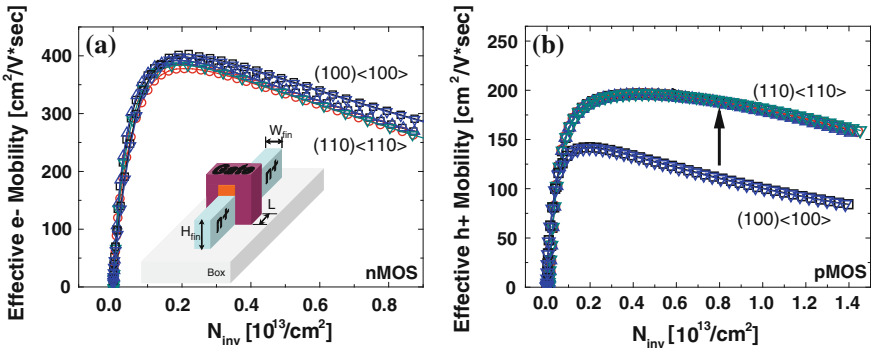


Fig. 7.18 Comparison of silicon FinFETs: **a** nMOS and **b** pMOS mobility with (100) and (110) channel orientations demonstrating repeatable results. In the nMOS case, mobility does not significantly degrade when moving to the (110) orientation. In addition, pMOS maintains its enhanced mobility suggesting that fabricating FinFETs on (110) only may be possible without the need for HOT. Inset: FinFET device structure with topside hardmask that creates a double-gate FinFET device

sidewall plane, is not as detrimental to the electron mobility. Additionally, one should notice the high value of the electron mobility ($\sim$400 cm^2/Vs) despite the sidewall conduction. For pMOS, the (110) holes enjoy the high enhanced mobility relative to (100) as shown in Fig. 7.18b similar to the improvement seen in Fig. 7.14b. Therefore, the plausibility on having FinFETs on (110) may be sufficient for future technologies without the need for HOT.

If the hardmask is not used, then a third topside orientation is introduced on (100) for non-rotated devices. The differences in the effective masses within these crystal planes and directions cause orientation-dependent mobility values. Further effective mobility extraction details, where top and sidewall mobilities can be separated, can be found in [64].

7.6.2 Strain and Germanium-Based Channels

Increased effective mobility has been demonstrated through channel strain engineering and germanium-based substrate materials in the channel region of MOSFETs, which has provided better device performance than conventional Si-based technologies.

7.6.2.1 Channel Strain Enhancement

Channel strain enhancement has been introduced through the actual fabrication process. For pMOS transistors, compressive strain enhances hole mobility over

unstressed silicon while tensile strain does the same for electron mobility [65]. Various fabrication techniques can enable both types of strain. For example, source and drain regions formed by selective epitaxial $Si_{1-x}Ge_x$ provide longitudinal uniaxial compressive stress that can increase hole mobility by 50 % in pMOSEFTs. A capping layer comprised of silicon nitride provides tensile strain in nMOSFETs, which enhances electron mobility by 20 % [65].

Continued work in this area has led to the development of different capping layers that serve as an etch stop layer but also impart compressive or tensile strain to short channel MOSFETs, thereby enhancing mobility [66]. One study using a hafnium-based dielectric that incorporates lanthanum achieved a 21 % improvement in drive current using "additive process-induced strain technology," which included a stressor contact etch stop layer and a stressor spacer in short channel ($L_g = 30$ nm) nMOSFETs [67].

7.6.2.2 Germanium-Based Channels

Fabricating a layer of silicon over a layer or substrate of silicon germanium (SiGe) introduces strain in the silicon (Si) layer because silicon atoms are stretched beyond their normal interatomic spacing. Since the SiGe lattice structure is spaced farther apart than conventional bulk silicon, the atoms in the overlying silicon layer "stretch" to match the atoms of the underlying silicon germanium layer, leading to strained silicon. When the silicon atoms are farther apart, the atomic forces that affect electron mobility are weakened resulting in better mobility and hence better performance and lower energy consumption. It has been shown that electrons in the SiGe channel can move significantly faster than in the silicon channel, allowing strained silicon transistors to switch much faster.

Another approach is to grow SiGe layer on silicon followed by deposition of a silicon capping layer [68]. This creates a Si/SiGe/Si quantum well device which has been shown to demonstrate the benefits of SiGe while having excellent short channel control [68]. Furthermore, the Si/SiGe/Si transistors can experience further mobility enhancements by taking advantage of the device orientation [69]. Figure 7.19 illustrates a 10 % improvement in peak SiGe (110)<110> electron mobility over Si (110)<110>. Meanwhile, SiGe (110)<110> pMOS experiences the highest enhanced mobility compared to any other SiGe orientation and higher than Si (100)<100> and (110)<110> (Fig. 7.20) [69]. Surface passivation and the related topics such as channel mobility of the Ge channels have been discussed in the chapter on high mobility channels.

7.6.3 Compound Semiconductors

Compound semiconductor (III–V) devices with high-k gate dielectrics are now being researched for their increased mobility [70–73]. Compared to the planar

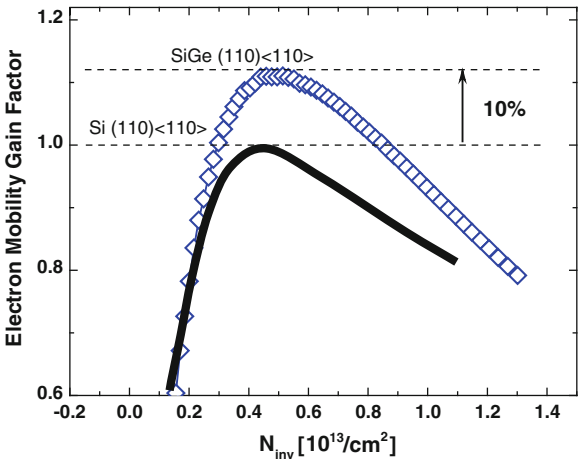


Fig. 7.19 Comparison of SiGe (110)<110> nMOS to Si (110)<110> nMOS where an improvement of 10 % peak electron mobility is seen for SiGe (110)<110> [69]

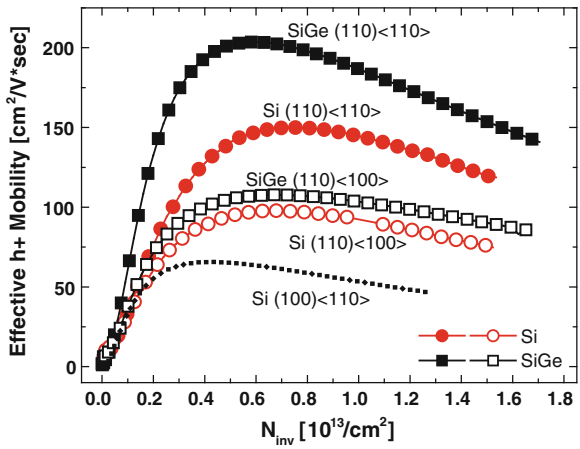


Fig. 7.20 Comparison of SiGe (110)<110> pMOS to other orientations on SiGe and Si. Peak hole mobility improved dramatically −210 and 130 % higher than Si (100)<110> and Si (110)<110>, respectively [69]

silicon technologies, III–V compound transistors can have significantly higher mobility. This is primarily due to the smaller effective mass of free carriers in the III–V material. Significant advances in interface quality and performance have been achieved on gallium arsenide (GaAs). However, device and performance issues with other III–V materials must first be overcome if a III–V compound device, such as indium gallium arsenide—which has higher mobility than GaAs, is to be used. Furthermore, III–V MOSFETs are still susceptible to scattering mechanisms, some of which are different from the mechanisms in the conventional Si-based devices. They include alloy, dislocation, polar optical phonon, and ac piezo-acoustic phonon scattering.

7.6.3.1 Current III–V Material Issues

Many III–V materials currently under investigation exhibit a highly disordered interface with the dielectric, a multi-layer epi-heterostructure substrate, highly resistive junctions, and a low density of states in the gamma valley of the conduction band. Additional factors affecting performance are device channel coupling with phonons in an overlying high-k dielectric layer; surface roughness; a high interface state defect density, border traps, and fixed charges; and multiple interfaces.

These material issues significantly impact the ability to model, characterize, and extract the "intrinsic" properties of the III–V devices. The key issue is the low density of states ($\sim 10^{11}$ cm^{-3} to 10^{13} cm^{-3}) in these materials because the interface state densities are currently high ($\sim 10^{11}$ cm^{-3} to 10^{13} cm^{-3}). Complex surface restructuring and poor bonding to the dielectric are two important factors contributing to the significant number of traps. These traps cause severe frequency dispersion in the capacitance—voltage (C–V) measurements [74, 75] of some III–V compound substrates, making it difficult to determine the correct split C–V capacitance to derive Q_{inv} values for mobility extraction [76]. These material concerns can degrade the $I_D - V_G$ characteristics, which further complicates mobility extraction.

7.6.3.2 Scattering Mechanisms

Scattering mechanisms are reviewed in detail in [77, 78]. In bulk III–V devices, such as Bipolar Junction Transistors and High Electron Mobility Transistors (HEMTs), mobility may be limited by ionized impurity scattering at low temperatures and by phonons at room and elevated temperatures, since the phonon scattering mechanism is thermally activated when phonons increase during temperature increases. Since III–V are polar materials, the carriers are scattered by the electromagnetic forces adding piezo-accoustic and polar optical phonon scattering to the deformation potential, which is typical for electron–phonon interactions in Si and Ge. In MOS III–V devices, the major scattering mechanisms are associated with the III–V/high-k interfaces. Remote charge scattering arises from the coulomb interaction between the carriers in the channel and the charged traps and fixed charges at the interface with the high-k and in the bulk of the high-k material. This scattering is "screened" as carrier density increases, but it might remain significantly strong even at a 10^{12} cm^{-2} carrier density. Remote phonon scattering is associated with the electron phonon interaction between the electrons in the channel and the low frequency optical phonons in the high-k material. This interaction is also electromagnetic [17]. The other important scattering mechanism, which plays a major role at high normal electric fields at the interface, is surface roughness (or remote surface roughness for MOSHEMTs) [79]. Obviously, the weighted contribution of different scattering mechanisms to the mobility degradation is dependent on the electrical bias values applied to the device, in

addition to the dependence on the material quality and the structure of the III–V/high-k interface. Surface passivation and the related topics such as channel mobility of the high mobility III–V compound semiconductor channels have been discussed in the chapter on high mobility channels.

7.7 Summary

This chapter has provided key information on effective channel mobility in devices with high-k gate dielectrics. A brief introduction to the effective mobility is discussed where various scattering mechanisms are highlighted that dictate the typical mobility behavior seen on conventional silicon transistors. The discussion continues with the impact that high-k dielectric gate stacks have on mobility. Mobility is shown to improve with reduction of high-k layer thickness and/or hafnium content. Finally, the chapter concludes with a discussion on the mobility enhancement experienced on high mobility channels and substrates with an emphasis placed on the differences seen when compared to conventional bulk Si-based technologies.

References

1. J.R. Hauser, Extraction of experimental mobility data for MOS devices. IEEE Trans. Electron Devices **43**, 1981–1988 (1996)
2. S. Takagi, A. Toriumi, M. Iwase, H. Tango, On the Universality of inversion layer mobility in Si MOSFETs: Part I: Effects of substrate impurity concentration. IEEE Trans. Electron Devices **41**, 2357–2362 (1994)
3. J.P. McKelvey, *Solid State Physics for Engineering and Materials Science* (Krieger Publishing Company, NY, 1993)
4. A. Tsividis, *Operation and Modeling of the MOS Transistors*, 2nd edn. (McGraw-Hill, NY, 1999)
5. A. Kerber, E. Cartier, L. Pantisano, M. Rosmeulen, R. Degraeve, T. Kauerauf, G. Groeseneken, H.E. Maes, U. Schwalke, Characterization of the V_T instability in SiO_2/HfO_2 gate dielectrics. in *41st Annual IEEE International Reliability Physics Symposium Proceedings*, pp. 41–45, 2003
6. C.D. Young, R. Choi, J.H. Sim, B.H. Lee, P. Zeitzoff, Y. Zhao, K. Matthews, G. A. Brown, G. Bersuker, Interfacial Layer Dependence of $HfSi_xO_y$ Gate Stacks on V_t Instability and Charge Trapping Using Ultra-short Pulse I–V Characterization. in *43rd Annual IEEE International Reliability Physics Symposium Proceedings*, pp. 75–79, 2005
7. C.D. Young, A. Kerber, T.H. Hou, E. Cartier, G.A. Brown, G. Bersuker, Y. Kim, J. Gutt, P. Lysaght, J. Bennett, C.H. Lee, S. Gopalan, M. Gardner, P.M. Zeitzoff, G. Groeseneken, R.W. Murto, H.R. Huff, Charge Trapping and Mobility Degradation in MOCVD Hafnium Silicate Gate Dielectric Stack Structures. in *203rd Fall Meeting of the Electrochemical Society, Physics and Technology of High-K Gate Dielectrics—II*, Orlando, FL, pp. 347–362, 2003
8. C.D. Young, G. Bersuker, G.A. Brown, P. Lysaght, P. Zeitzoff, R.W. Murto, H.R. Huff, Charge trapping and device performance degradation in MOCVD hafnium-based gate

dielectric stack structures. in *42nd Annual IEEE International Reliability Physics Symposium. Proceedings* pp. 597–598, 2004
9. A. Kerber, E. Cartier, L.A. Ragnarsson, M. Rosmeulen, L. Pantisano, R. Degraeve, Y. Kim, G. Groeseneken, Direct measurement of the inversion charge in MOSFETs: application to mobility extraction in alternative gate dielectrics. in *VLSI Symposium Technical Digest*, pp. 159–160, 2003
10. C.D. Young, Y. Zhao, M. Pendley, B.H. Lee, K. Matthews, J.H. Sim, R. Choi, G.A. Brown, R.W. Murto, G. Bersuker, Ultra-short pulse current–voltage characterization of the intrinsic charactistics of high-k devices. Jpn. J. Appl. Phys. **44**, 2437–2440 (2005)
11. C. Leroux, J. Mitard, G. Ghibaudo, X. Garros, G. Reimbold, B. Guillaumot, F. Martin, Characterization and modeling of hysteresis phenomena in high K dielectrics. in *IEEE International Electron Devices Meeting Technical Digest*, pp. 737–740, 2004
12. B.H. Lee, C.D. Young, R. Choi, J.H. Sim, G. Bersuker, C.Y. Kang, R. Harris, G.A. Brown, K. Matthews, S.C. Song, N. Moumen, J. Barnett, P. Lysaght, K.S. Choi, H.C. Wen, C. Huffman, H. Alshareef, P. Majhi, S. Gopalan, J.J. Peterson, P. Kirsh, H.-J. Li, J. Gutt, M. Gardner, H.R. Huff, P. Zeitzoff, R.W. Murto, L. Larson, C. Ramiller, Intrinsic characteristics of high-k devices and implications of fast transient charging effects (FTCE). in *IEEE International Electron Devices Meeting Technical Digest*, pp. 859–862, 2004
13. L. Pantisano, E. Cartier, A. Kerber, R. Degraeve, M. Lorenzini, M. Rosmeulen, G. Groeseneken, H.E. Maes, Dynamics of threshold voltage instability in stacked high-k dielectrics: role of the interfacial oxide. in *VLSI Symposium Technical Digest*, pp. 163–164, 2003
14. G. Bersuker, P. Zeitzoff, J. H. Sim, B.H. Lee, R. Choi, G.A. Brown, C.D. Young, Mobility Evaluation in High-k Devices. in *IEEE International Integrated Reliability Workshop Final Report*, pp. 141–144, 2004
15. W. Zhu, J.-P. Han, T.P. Ma, Mobility measurements and degradation mechanisms of MOSFETs made with ultrathin high-k dielectrics. IEEE Trans. Electron Devices **51**, 98–105 (2004)
16. G. Bersuker, P. Zeitzoff, J. Barnett, N. Moumen, B. Foran, C.D. Young, J.J. Peterson, P. Lysaght, Interface-induced mobility degradation in high-k transistors. Jpn. J. Appl. Phys. **43**, 7899–7902 (2004)
17. M.V. Fischetti, D.A. Neumayer, E.A. Cartier, Effective electron mobility in Si inversion layers in metal-oxide semiconductor systems with a high-k insulator: The role of remote phonon scattering. J. Appl. Phys. **90**, 4587–4608 (2001)
18. W.J. Zhu, T.P. Ma, Temperature dependence of channel mobility in HfO_2-gated NMOSFETs. IEEE Electron Device Lett. **25**, 89–91 (2004)
19. G. Bersuker, J.H. Sim, C.D. Young, R. Choi, B.H. Lee, P. Lysaght, G. A. Brown, P. Zeitzoff, M. Gardner, R. W. Murto, H.R. Huff, Effects of structural properties of Hf-based gate stack on transistor performance. in *2004 Spring Meeting of the Material Research Society*, pp. 31–35, 2004
20. T. Yamaguchi, R. Iijima, T. Ino, A. Nishiyama, H. Satake, N. Fukushima, Additional scattering effects for mobility degradation in Hf-silicate gate MISFETs. in *IEEE International Electron Devices Meeting Technical Digest*, pp. 621–624, 2002
21. R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, M. Metz, High-κ/metal-gate stack and its MOSFET characteristics. IEEE Electron Device Lett. **25**, 408–410 (2004)
22. D.V. Singh, P. Solomon, E.P. Gusev, G. Singco, Z. Ren, Ultra-fast measurements of the inversion charge in MOSFETs and impact on measured mobility in high-k MOSFETs. in *IEEE International Electron Devices Meeting Technical Digest*, San Francisco, CA, pp. 863–866, 2004
23. G. Bersuker, J. Sim, C.S. Park, C.D. Young, S. Nadkarni, C. Rino, L. Byoung Hun, Mechanism of electron trapping and characteristics of traps in HfO_2 gate stacks. IEEE Trans. Device Mater. Reliab. **7**, 138–145 (2007)

24. D. Heh, C.D. Young, G. Bersuker, Experimental evidence of the fast and slow charge trapping/detrapping processes in high-k dielectrics subjected to PBTI stress. IEEE Electron Device Lett. **29**, 180–182 (2008)
25. C.D. Young, Y. Zhao, D. Heh, R. Choi, B.H. Lee, G. Bersuker, Pulsed I_d–V_g methodology and its application to electron-trapping characterization and defect density profiling. IEEE Trans. Electron Devices **56**, 1322–1329 (2009)
26. P. Broqvist, A. Pasquarello, Oxygen vacancy in monoclinic HfO_2: A consistent interpretation of trap assisted conduction, direct electron injection, and optical absorption experiments. Appl. Phys. Lett. **89**, 262904 (2006)
27. J.L. Gavartin, D.M. Ramo, A.L. Shluger, G. Bersuker, B.H. Lee, Negative oxygen vacancies in HfO_2 as charge traps in high-k stacks. Appl. Phys. Lett. **89**, 082908 (2006)
28. D.M. Ramo, J.L. Gavartin, A.L. Shluger, G. Bersuker, Intrinsic and defect-assisted trapping of electrons and holes in HfO_2: An ab initio study. Microelectron. Eng. **84**, 2362–2365 (2007)
29. D.M. Ramo, J.L. Gavartin, A.L. Shluger, G. Bersuker, Spectroscopic properties of oxygen vacancies in monoclinic HfO_2 calculated with periodic and embedded cluster density functional theory. *Physical Review B (Condensed Matter and Materials Physics),* vol. 75, p. 205336, 2007
30. D.M. Ramo, A.L. Shluger, J.L. Gavartin, G. Bersuker, Theoretical prediction of intrinsic self-trapping of electrons and holes in monoclinic HfO_2. Phys. Rev. Lett. **99**, 155504 (2007)
31. J. Robertson, K. Xiong, S.J. Clark, Band gaps and defect levels in functional oxides. Thin Solid Films **496**, 1–7 (2006)
32. R. Degraeve, A. Kerber, P. Roussell, E. Cartier, T. Kauerauf, L. Pantisano, G. Groeseneken, Effect of bulk trap density on HfO_2 reliability and yield. in *IEEE International Electron Devices Meeting Techical Digest*, pp. 935–938, 2003
33. C. Leroux, J. Mitard, G. Ghibaudo, X. Garros, G. Reimbold, B. Guillaumor, F. Martin, Characterization and modeling of hysteresis phenomena in high K dielectrics. in *IEEE International Electron Devices Meeting Techical Digest*, pp. 737–740, 2004
34. C. Shen, M.F. Li, X.P. Wang, Y. Yee-Chia, D.L. Kwong, A fast measurement technique of MOSFET I_d–V_g characteristics. IEEE Electron Device Lett. **27**, 55–57 (2006)
35. C.D. Young, Y.G. Zhao, M. Pendley, B.H. Lee, K. Matthews, J.H. Sim, R. Choi, G.A. Brown, R.W. Murto, G. Bersuker, Ultra-short pulse current–voltage characterization of the intrinsic characteristics of high-kappa devices. Jap. J Appl. Phys. Part 1-Regular Pap. Short Notes Rev. Pap. **44**, 2437–2440 (2005)
36. C.D. Young, P. Zeitzoff, G.A. Brown, G. Bersuker, L. Byoung Hun, J.R. Hauser, Intrinsic mobility evaluation of high-κ gate dielectric transistors using pulsed I_d–V_g. IEEE Electron Device Lett. **26**, 586–589 (2005)
37. G. Bersuker, P. Zeitzoff, J.H. Sim, B.H. Lee, R. Choi, G. Brown, C.D. Young, Mobility evaluation in transistors with charge-trapping gate dielectrics. Appl. Phys. Lett. **87**, 042905 (2005)
38. J. Barnett, N. Moumen, J. Gutt, M. Gardner, C. Huffman, P. Majhi, J. J. Peterson, S. Gopalan, B. Foran, H.-J. Li, B.H. Lee, G. Bersuker, P. Zeitzoff, G.A. Brown, P. Lysaght, C.D. Young, R.W. Murto, H.R. Huff, Experimental study of etched back thermal oxide for optimization of the Si/high-k interface. in *2004 Spring Meeting of the Material Research Society*, pp. 341–346, 2004
39. C.D. Young, Y. Zhao, D. Heh, R. Choi, B.H. Lee, G. Bersuker, Pulsed I_d–V_g methodology and its application to electron-trapping characterization and defect density profiling. IEEE Trans. Electron Devices **56**, 1322–1329 (2009)
40. G. Bersuker, J. Barnett, N. Moumen, B. Foran, B. Young, P. Lysaght, J. Peterson, B.H. Lee, P.M. Zeitzoff, H.R. Huff, Interfacial layer-induced mobility degradation in high-k transistors. Jap. J. Appl. Phys. Part 1-Regular Pap. Short Notes Rev. Pap. **43**, 7899–7902 (2004)
41. G. Bersuker, P. Zeitzoff, J.H. Sim, B.H. Lee, R. Choi, G.A. Brown, C.D. Young, Mobility Evaluation in High-K Devices. in *IEEE International Integrated Reliability Workshop Final Report*, 2004, pp. 141–144

42. P.D. Kirsch, J.H. Sim, S.C. Song, S. Krishnan, J.J. Peterson, H.-J. Li, M. Quevedo-Lopez, C.D. Young, R. Choi, N. Moumen, P. Majhi, Q. Wang, J.G. Ekerdt, G. Bersuker, B.H. Lee, Mobility enhancement of high-κ gate stacks through reduced transient charging. in *35th Annual European Solid-State Device Research Conference*, Grenoble, France, pp. 367–370, 2005
43. M.A. Negara, K. Cherkaoui, P.K. Hurley, C.D. Young, P. Majhi, W. Tsai, D. Bauza, G. Ghibaudo, Analysis of electron mobility in HfO_2/TiN gate metal-oxide-semiconductor field effect transistors: The influence of HfO_2 thickness, temperature, and oxide charge. J. Appl. Phys. **105**, 024510 (2009)
44. M.A. Negara, K. Cherkaoui, P. Majhi, C.D. Young, W. Tsai, D. Bauza, G. Ghibaudo, P.K. Hurley, The influence of HfO_2 film thickness on the interface state density and low field mobility of n channel HfO_2/TiN gate MOSFETs. Microelectron. Eng. **84**, 1874–1877 (2007)
45. C.G. Sodini, T.W. Ekstedt, J.L. Moll, Charge accumulation and mobility in thin dielectric MOS transistors. Solid State Electron. **25**, 833–841 (1982)
46. A. Kerber, E. Cartier, L. A. Ragnarsson, M. Rosmeulen, L. Pantisano, R. Degraeve, T. Kauerauf, Y. Kim, and G. Groeseneken, Direct measurement of the inversion charge in MOSFETs: application to mobility extraction in alternative gate dielectrics. in *VLSI Technology Symposium*, Kyoto, Japan, 2003, pp. Session 12A-5
47. J.R. Hauser, K. Ahmed, Characterization of Ultrathin oxides using electrical C–V and I–V measurements. in *International Conferences on Characterization and Metrology for ULSI Technology*, pp. 235–239, 1998
48. C.D. Young, P. Zeitzoff, G. Bersuker, R. Choi, Comparison of trap-free mobility extraction techniques for high-κ gate dielectrics. in *International Workshop on Electrical Characterization and Reliability for High-κ Devices*, pp. 73–81, 2004
49. P. Sivasubramani, P.D. Kirsch, J. Huang, C. Park, Y.N. Tan, D.C. Gilmer, C. Young, K. Freeman, M.M. Hussain, R. Harris, S.C. Song, D. Hen, R. Choi, P. Majhi, G. Bersuker, P. Lysaght, B.H. Lee, H.H. Tseng, J.S. Jur, D.J. Lichtenwalner, A.I. Kingon, R. Jammy, Aggressively scaled high-k gate dielectric with excellent performance and high temperature stability for 32 nm and beyond. in *IEEE International Electron Devices Meeting*, pp. 543–546, 2007
50. J. Huang, P.D. Kirsch, D. Heh, C.Y. Kang, G. Bersuker, M. Hussain, P. Majhi, P. Sivasubramani, D.C. Gilmer, N. Goel, M.A. Quevedo-Lopez, C. Young, C. S. Park, C. Park, P.Y. Hung, J. Price, H.R. Harris, B.H. Lee, H.H. Tseng, R. Jammy, Device and reliability improvement of $HfSiON^+$ LaOx/metal gate stacks for 22 nm node application. in *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, pp. 1–4, 8164–2284, 2008
51. M. Inoue, Y. Satoh, M. Kadoshima, S. Sakashita, T. Kawahara, M. Anma, R. Nakagawa, H. Umeda, S. Matsuyama, H. Fujimoto, H. Miyatake, Impact of area scaling on threshold voltage lowering in La-containing high-k/metal gate NMOSFETs fabricated on (100) and (110)Si. in *VLSI Technology, 2009 Symposium on*, pp. 40–41, 2009
52. P.D. Kirsch, P. Sivasubramani, J. Huang, C.D. Young, M.A. Quevedo-Lopez, H.C. Wen, H. Alshareef, K. Choi, C.S. Park, K. Freeman, M.M. Hussain, G. Bersuker, H.R. Harris, P. Majhi, R. Choi, P. Lysaght, B.H. Lee, H.H. Tseng, R. Jammy, T.S. Boscke, D.J. Lichtenwalner, J.S. Jur, A.I. Kingon, Dipole model explaining high-k/metal gate field effect transistor threshold voltage tuning. in *Applied Physics Letters*, 2008
53. K. Okamoto, M. Adachi, K. Kakushima, P. Ahmet, N. Sugii, K. Tsutsui, T. Hattori, H. Iwai, Effective control of flat-band voltage in HfO_2 gate dielectric with La_2O_3 incorporation. in *Solid State Device Research Conference, 2007. ESSDERC 2007. 37th European*, pp. 199–202, 1930–8876, 2007
54. C.S. Park, J.W. Yang, M.M. Hussain, C.Y. Kang, J. Huang, P. Sivasubramani, C. Park, K. Tateiwa, Y. Harada, J. Barnett, C. Melvin, G. Bersuker, P. D. Kirsch, B.H. Lee, H.H. Tseng, R. Jammy, La-doped metal/high-K nMOSFET for sub-32 nm HP and LSTP application. in *VLSI Technology, Systems, and Applications, 2009. VLSI-TSA '09. International Symposium on*, pp. 59–60, 1930–885X, 2009

55. P. Sivasubramani, P.D. Kirsch, J. Huang, C. Park, Y.N. Tan, D.C. Gilmer, C. Young, K. Freeman, M.M. Hussain, R. Harris, S.C. Song, D. Hen, R. Choi, P. Majhi, G. Bersuker, P. Lysaght, B.H. Lee, H.H. Tseng, J.S. Jur, D.J. Lichtenwalner, A.I. Kingon, R. Jammy, Aggressively scaled high-k gate dielectric with excellent performance and high temperature stability for 32 nm and beyond. in *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pp. 543–546, 2007
56. K. Tatsumura, T. Ishihara, S. Inumiya, K. Nakajima, A. Kaneko, M. Goto, S. Kawanaka, A. Kinoshita, Intrinsic correlation between mobility reduction and V_t shift due to interface dipole modulation in $HfSiON/SiO_2$ stack by La or Al addition. in *IEEE International Electron Devices Meeting*, pp. 1–4, 2008
57. H.C. Wen, S.C. Song, C.S. Park, C. Burham, G. Bersuker, K. Choi, M.A. Quevedo-Lopez, B.S. Ju, H.N. Alshareef, H. Niimi, H.B. Park, P.S. Lysaght, P. Majhi, B.H. Lee, R. Jammy, Gate first metal-Aluminum-Nitride PMOS electrodes for 32 nm low standby power applications. in *VLSI Technology, 2007 IEEE Symposium on*, pp. 160–161, 2007
58. J. Huang, D. Heh, P. Sivasubramani, P.D. Kirsch, G. Bersuker, D.C. Gilmer, M.A. Quevedo-Lopez, M.M. Hussain, P. Majhi, P. Lysaght, H. Park, N. Goel, C. Young, C.S. Park, C. Park, M. Cruz, V. Diaz, P.Y. Hung, J. Price, H.H. Tseng, R. Jammy, Gate first high-k/metal gate stacks with zero SiOx interface achieving EOT = 0.59 nm for 16 nm application. in *VLSI Technology, 2009 Symposium on*, pp. 34–35, 2009
59. Y. Min, V.W.C. Chan, K.K. Chan, L. Shi, D.M. Fried, J.H. Stathis, A.I. Chou, E. Gusev, J.A. Ott, L.E. Burns, M.V. Fischetti, I. Meikei, Hybrid-orientation technology (HOT): opportunities and challenges. Electron Devices, IEEE Trans. **53**, 965–978 (2006)
60. S.A. Krishnan, H. Rusty Harris, P.D. Kirsch, C. Krug, C. Quevedo-Lopez, C. Young, L. Byoung Hun, R. Choi, N. Chowdhury, S. Suthram, S. Thompson, G. Bersuker, R. Jammy, High performing pMOSFETs on Si(110) for application to hybrid orientation technologies—comparison of HfO_2 and HfSiON. in *Electron Devices Meeting, 2006. IEDM '06. International*, pp. 265–268, 2006
61. F. Stern, Quantum properties of surface space-charge layers. CRC: Crit. Rev. Solid State Sci. **4**, 499–514 (1973)
62. G. Bersuker, C.S. Park, J. Barnett, P.S. Lysaght, P.D. Kirsch, C.D. Young, R. Choi, B.H. Lee, B. Foran, K.V. Benthem, S.J. Pennycook, P.M. Lenahan, J.T. Ryan, The effect of interfacial layer properties on the performance of Hf-based gate stack devices. J. Appl. Phys. **100**, 094108 (2006)
63. H.R. Harris, S.E. Thompson, S. Krishnan, P. Kirsch, P. Majhi, C.E. Smith, M.M. Hussain, G. Sun, H. Adhikari, S. Suthram, B.H. Lee, H.H. Tseng, R. Jammy, Flexible, simplified CMOS on Si(110) with metal gate/high k for HP and LSTP. in *IEEE International Electron Devices Meeting*, pp. 57–60, 2007
64. V.V. Iyengar, A. Kottantharayil, F.M. Tranjan, M. Jurczak, K. De Meyer, Extraction of the top and sidewall mobility in FinFETs and the impact of fin-patterning processes and gate dielectrics on mobility. IEEE Trans. Electron Devices **54**, 1177–1184 (2007)
65. S.E. Thompson, M. Armstrong, C. Auth, M. Alavi, M. Buehler, R. Chau, S. Cea, T. Ghani, G. Glass, T. Hoffman, C.H. Jan, C. Kenyon, J. Klaus, K. Kuhn, M. Zhiyong, B. McIntyre, K. Mistry, A. Murthy, B. Obradovic, R. Nagisetty, N. Phi, S. Sivakumar, R. Shaheed, L. Shifren, B. Tufts, S. Tyagi, M. Bohr, Y. El-Mansy, A 90-nm logic technology featuring strained-silicon. Electron Devices IEEE Trans. **51**, 1790–1797 (2004)
66. S. Ito, H. Namba, K. Yamaguchi, T. Hirata, K. Ando, S. Koyama, S. Kuroki, N. Ikezawa, T. Suzuki, T. Saitoh, T. Horiuchi, Mechanical stress effect of etch-stop nitride and its impact on deep submicron transistor design. in *Electron Devices Meeting, 2000. IEDM Technical Digest. International*, pp. 247–250, 2000
67. M.M. Hussain, K. Rader, C. Smith, C. Young, S. Suthram, C. Park, M. Cruz, P.D. Kirsch, R. Jammy, Additive process induced strain (APIS) technology for L_g = 30 nm band-edge high-k/metal gate nMOSFET. in *International Conference on Solid State Devices and Materials*, 2008

68. S.-H. Lee, P. Majhi, J. Oh, B. Sassman, C. Young, A. Bowonder, W.-Y. Loh, K.-J. Choi, B.-J. Cho, H.-D. Lee, P. Kirsch, H.R. Harris, W. Tsai, S. Datta, T. Hsing-Huang, S.K. Banerjee, R. Jammy, Demonstration of $L_g \sim 55$ nm pMOSFETs With $Si/Si_{0.25}Ge_{0.75}/Si$ channels, high I_{on}/I_{off} ($>5 \times 10^4$), and controlled short channel effects (SCEs). Electron Device Lett. IEEE **29**, 1017–1020 (2008)
69. J. Oh, S.-H. Lee, K.S. Min, J. Huang, B.G. Min, B. Sassman, K. Jeon, W.-Y. Loh, J. Barnett, I. Ok, C.Y. Kang, C. Smith, D.-H. Ko, P. Kirsch, R. Jammy, SiGe on (110) channel orientation with mobility boosters: Surface orientation, channel directions, and uniaxial strain. in *Symposium on VLSI Technology*, pp. 39–40, 2010
70. S. Datta, III–V field-effect transistors for low power digital logic applications. Microelectron. Eng. **84**, 2133–2137 (2007)
71. R.J. W. Hill, X. Li, H. Zhou, D.S. Macintyre, S. Thoms, M.C. Holland, P. Longo, D.A.J. Moran, A.J. Craven, C.R. Stanley, A. Asenov, R. Droopad, M. Passlack, I.G. Thayne, Deep sub-micron and self-aligned flatband III–V MOSFETs. in *2009 Device Research Conference*, pp. 251–252, 2009
72. J. Huang, N. Goel, H. Zhao, C.Y. Kang, K.S. Min, G. Bersuker, S. Oktyabrsky, C.K. Gaspe, M.B. Santos, P. Majhi, P.D. Kirsch, H.H. Tseng, J.C. Lee, R. Jammy, InGaAs MOSFET performance and reliability improvement by simultaneous reduction of oxide and interface charge in ALD (La)AlOx/ZrO2 gate stack. in *2009 IEEE International Electron Devices Meeting*, pp. 1–4, 2009
73. M. Passlack, P. Zurcher, K. Rajagopalan, R. Droopad, J. Abrokwah, M. Tutt, Y.B. Park, E. Johnson, O. Hartin, A. Zlotnicka, P. Fejes, R.J.W. Hill, D.A.J. Moran, X. Li, H. Zhou, D. Macintyre, S. Thorns, A. Asenov, K. Kalna, I.G. Thayne, High mobility III–V MOSFETs for RF and digital applications. in *2007 IEEE International Electron Devices Meeting*, pp. 621–624, 2007
74. H. Hasegawa, S. Kodama, K. Koyanagi, and M. Akazawa, Control of structure and properties of compound semiconductor interfaces by Si interface control layer. in *Indium Phosphide and Related Materials, 1993. Conference Proceedings., Fifth International Conference on*, pp. 289–292, 1993
75. C.L. Hinkle, M. Milojevic, B. Brennan, A.M. Sonnet, F.S. Aguirre-Tostado, G.J. Hughes, E.M. Vogel, R.M. Wallace, Detection of Ga suboxides and their impact on III–V passivation and Fermi-level pinning. Appl. Phys. Lett. **94**, 162101–162103 (2009)
76. C.L. Hinkle, A.M. Sonnet, R.A. Chapman, E.M. Vogel, Extraction of the Effective Mobility of $In_{0.53}$ $Ga_{0.47}$ as MOSFETs. Electron Device Lett. IEEE **30**, 316–318 (2009)
77. W. Walukiewicz, H.E. Ruda, J. Lagowski, H.C. Gatos, Electron mobility in modulation-doped heterostructures. Phys. Rev. B **30**, 4571 (1984)
78. M.V. Fischetti, S. Jin, The effects of defects on electron transport in nanometer-scale electronic devices: Impurities and interface roughness, in *Defects in Microelectronic Materials and Devices*, ed. by S. Pantelides, D. Fleetwood, R. Schrimpf (CRC Press, Boca Raton, 2008)
79. F. Gamiz, M.V. Fischetti, Remote Coulomb scattering in metal-oxide-semiconductor field effect transistors: Screening by electrons in the gate. Appl. Phys. Lett. **83**, 4848–4850 (2003)

Chapter 8
Reliability Implications of Fast and Slow Degradation Processes in High-k Gate Stacks

Gennadi Bersuker

Abstract Aggressive transistor scaling to achieve better chip functionality is driving the introduction of high-k dielectric materials into traditional device gate stacks. These advanced gate stacks are multilayer structures, the materials of which may strongly interact during high temperature processing, generating structural defects in the layers. Such complex structures pose new challenges in interpreting electrical measurements, which are sensitive to even extremely small concentrations of electrically active defects. A high density of pre-existing defects and defect precursors is shown to give rise to both fast and slow instabilities in the device characteristics, some of which may be reversible while others cause unrecoverable dielectric degradation. In this chapter, we focus on the reliability implications of instabilities associated with pre-existing structural defects in both the high-k film and the interfacial layer of gate oxide stacks subjected to a variety of reliability tests.

8.1 Introduction

The introduction of metal oxides (high-k dielectrics, high-k) in gate dielectric applications raises new issues for reliability evaluations due to their special material properties, which are different from the conventional silicon dioxide gate dielectric. A common electronic feature of high-*k* materials is the presence of *d*-shell states, which make their structural properties drastically different from those of SiO_2 gate dielectric [1]. One of the consequences of the *d*-electron bonding in high-*k* dielectrics is a relatively high density of as-grown defects, specifically oxygen vacancies, which, as shown by ab initio calculations [2, 3] for

G. Bersuker (✉)
SEMATECH, 257 Fuller Road, Albany, NY 12203, USA
e-mail: gennadi.bersuker@sematech.org

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5_8,

monoclinic hafnia, may exist in five charge states, from −2 to +2, and may function as both electron/hole traps and fixed charges. Thus, in high-*k* devices, the contribution to electrical instability from charge trapping/detrapping at pre-existing structural defects may be significant and may even dominate stress-induced defect generation traditionally considered the major cause of instability in SiO_2 gate dielectrics.

Understanding the fundamental reliability mechanisms is complicated by the fact that high-*k* gate stacks formed using Hf and Zr oxide (or their silicates) thin films are usually multilayer structures that include a SiO_2-like layer either spontaneously or intentionally formed at the interface with the substrate [4]. This interfacial layer (IL) exhibits a rather non-uniform elemental composition due to its interaction with the overlaying high-k film, the gate electrode, and the substrate. While SiO_2 is known to form an oxygen-deficient transitional layer near the interface with the Si substrate to accommodate the lattice constant mismatch between these materials, data from high resolution chemical and spectroscopic analyses suggest that the SiO_2 film adjacent to the high-k dielectric is also oxygen-deficient [5–8]. This deficiency is driven by the energetically advantageous diffusion of oxygen vacancies out of the metal oxide into the silicon dioxide [9–11]. The IL electrical characteristics support this understanding: ILs have been reported to exhibit *k* values significantly greater than that of a bulk stoichiometric SiO_2 [6, 12, 13], which can result from oxygen deficiency of the SiO_2 film, specifically due to contributions from the under coordinated Si^{2+} ions [14]. Although hafnia and silicon dioxide generally do not mix (under high temperatures, hafnium silicates in particular tend to phase separately into HfO_2 and SiO_2 [15]), and isolated, opportunistic Hf ions may diffuse into the IL [5, 16, 17], potentially forming charge traps. Overall, the IL in the high-k gate stack can be expected to contain a high density of electrically active defects and defect precursors.

A high density of as-grown process-related electrically active defects in the high-k gate stacks may introduce new reliability phenomena associated with charge trapping/detrapping at these defects. The effect of trapping on device parameters is determined by its characteristic time with respect to that of sense measurements and stress-induced defect generation. When the characteristic time of the charge trapping at pre-existing defects is much shorter than that of the sense measurement, the effect of this fast trapping manifests itself in intrinsic (rather than time-dependent) device characteristics. In other words, the contribution from the fast process is included in the time-zero parameter value. However, under actual use conditions, for instance, when employed as a high frequency switch, the device may experience much shorter bias pulses than the test measurements (in particular, DC measurements), which would lead to not only different time-zero values of device parameters but also a drift of these values with the total actual time of device operation.

On the other hand, slower charge trapping/detrapping at the pre-existing defects may add to the total measured time-dependent device instability. When the characteristic times of these slow processes are less than that of new defect generation, the contribution from the charge trapping at pre-existing defects may

dominate device instability, resulting in a time dependency of the electrical parameters. This is traditionally interpreted as, and thus could erroneously be assigned to, a dielectric degradation caused by the generation of new defects. As shown below, the presence of a large fast instability component may affect the evaluation of slower, long-term instability processes that are responsible for time-dependent degradation and, used for lifetime projection. Separating contributions from the fast and slow processes to total device instability is thus critically important to obtain correct reliability estimations.

The above considerations indicate that characterization of such a complex multilayer, multi-component, ultra-thin dielectric stack raises a set of questions that were not significant with conventional silicon dioxide-based gate stacks:

- Where are the defects contributing to instability located: in the high-k film or IL?
- Are these pre-existing (process-related) or stress-generated defects?
- How can a lifetime projection methodology for the contribution of fast charging/discharging processes to the total instability be corrected?

These questions, in turn, define requirements for electrical measurements, which should provide high time and spatial resolutions to address fast transient processes and differentiate signals from different regions through the depth of the gate dielectric.

In this work, we focus on the effects on reliability estimations of fast transient processes associated with pre-existing, as-grown defects in high-k stacks and outline an evaluation methodology to correct for the fast charging. We also explore mechanisms for slow charge trapping processes due to pre-existing defects, which are responsible for long-term device reliability characteristics. Gate stack degradation caused by stress-induced defect generation is considered next. We determine that the dielectric degradation can be reliably monitored by stress-induced leakage current (SILC) measurements and that under low voltage/moderate temperature stress conditions, defect generation in the IL represents a major degradation factor.

8.2 Instability Due to Pre-existing Defects

Reversible and repeatable instability of device parameters may indicate that the device is controlled by charge trapping in pre-existing defects. Indeed, electron trapping at low stress voltages relevant to circuit operation conditions is found to be reversible. By applying a voltage of the opposite polarity, the threshold voltage (V_t) can be returned to its initial pre-stress value as illustrated in Fig. 8.1a. De-trapping can be accelerated by applying a voltage of the opposite polarity [18–21], while the actual V_t value after detrapping is defined by the detrapping voltage (which, in the case in Fig. 8.1a is not high enough for complete detrapping). The V_t dependence on stress time, as well as the maximum V_t value, does not change

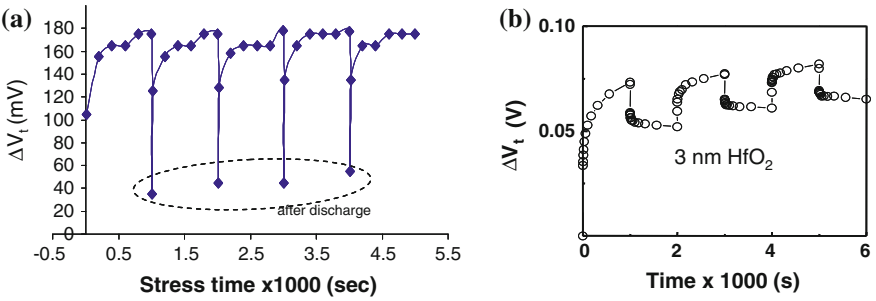


Fig. 8.1 **a** ΔV_t change during low voltage CVS on 1.1 nm SiO_2/3 nm HfO_2 gate stack in (**a**) nFETs at $V_g = 2.4$ V, with periodic discharge at $V_g = -1$ V/10 s and **b** pFETs at $V_g = -2$ V with discharge at +1 V

with each subsequent trapping/detrapping cycle. Therefore, electron trapping appears to occur in pre-existing (as-grown) defects and a low voltage electrical stress of practical importance does not generate new traps responsible for the V_t shift. This reversible behavior can be expected when the trapping occurs in the high-k dielectric in which the electrons, when trapped at the neutral oxygen vacancy sites, are delocalized over three neighboring metal ions [2]. In this case, the electron trapping causes no significant structural relaxation of the surrounding lattice [22], hence permanent structural damage associated with the electron trapping/detrapping should be less expected. For pFETs, however, ΔV_t increases with each stress/detrapping cycle, see Fig. 8.1b, which indicates the generation of new defects. Because of the SiO_2 IL, the increase in ΔV_t can be expected to be similar to conventional SiO_2 pFETs subjected to negative bias temperature instability (NBTI) stress.

8.2.1 Fast Transient Instability

The abrupt increase in threshold voltage instability (ΔV_t) when the stress is initiated (first data point in Fig. 8.1) points to a possible contribution from fast electron trapping, which could be significantly faster than the characteristic times of conventional DC measurements, Fig. 8.1. To monitor the kinetics of the fast trapping, pulsed drain current-gate voltage (I_d–V_g) measurements [23, 24] can be used, in which a trapezoid or triangular pulse in the nano- to micro second range is applied to the transistor gate while recording drain current variation during the rise and fall times of the pulse width. The change in I_d during a single pulse measurement can be presented either in a pulse voltage or pulse time domain; Fig. 8.2a and b illustrate I_d–V_g and I_d-time characteristics, respectively. The hysteresis of the I_d–V_g curve, as well as a current decrease in I_d–t dependency, is caused by the fully reversible electron trapping in the high-k layer during the pulse, cf. Fig. 8.2c, which effectively produces a threshold voltage shift, ΔV_t [24].

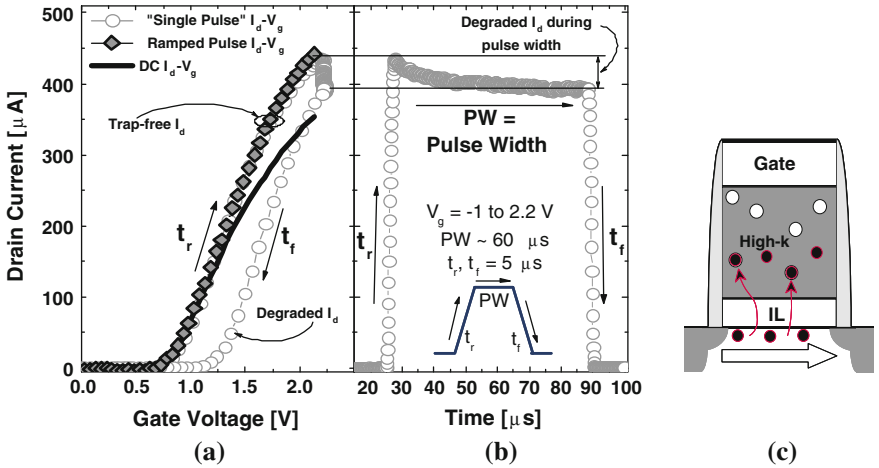


Fig. 8.2 Various forms of pulsed I_d data in the **a** pulse voltage and **b** pulse time domains. I_d degrades during the pulse width period due to fast transient charging. The I_d curves corresponding to the ramped pulse and the rising portion of the single pulse (the t_r portion) coincide, demonstrating trapping-free characteristics. DC included for comparison [24] **c** illustration of the electron trapping process during the I_d–V_g measurements

An example of the evolution of single pulse I_d–V_g curves during stress is presented in Fig. 8.3, [25]. In this particular high-k gate stack, a significant shift of the pulse I_d–V_g curve occurs during 10 μs stress, which is attributed to fast transient charging filling the resonant traps [26]. On the other hand, longer stresses led to only a negligible additional shift of the pulsed I_d–V_g curve. This result suggests that the overall stress-induced V_{th} change in high-k devices at a low bias condition is dominated by fast transient charging with characteristic times in the

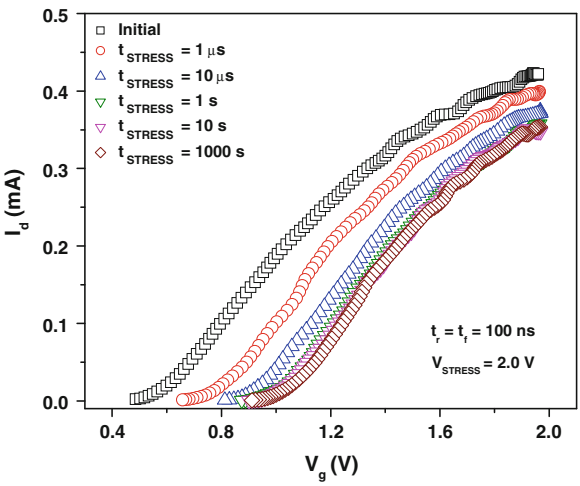


Fig. 8.3 Comparison of single pulse I_d–V_g curves before and after different stress times

micro-second range. However, multiple stress cycles, each including a long stress followed by an electron detrapping step using a small bias of opposite polarity, do not result in any detectable changes in threshold voltage, as measured by the pulse method, cf. Fig. 8.4, indicating that no new traps were generated in the high-k layer.

The devices used to study the kinetics of fast charge trapping were fully processed nMOSFETs fabricated with 3 nm atomic layer deposited (ALD) HfO_2 dielectric on 1 nm thermal SiO_2 and TiN electrodes using a standard CMOS process with a 1,000 °C, 10 s dopant activation anneal. Gate stacks were fabricated to have considerably high density of traps for better resolution of the threshold voltage changes during short increments of stress time. Devices were stressed from sub-microseconds to thousands of seconds at room temperature and 125 °C. The study was conducted using a pulse-based BTI technique, with which short stress times (<1 s) were achieved by applying a trapezoidal pulse with the desired pulse width. The difference between the V_{th} values extracted from the rising and falling V_g portions of the pulse then indicates the V_{th} change due to stress during the well controlled pulse width, as shown in the inset of Fig. 8.5. As follows from the data in Fig. 8.5, during short stresses, ΔV_{th} is not affected by temperature. This temperature-independent characteristic suggests that the V_{th} time dependency during a short stress is dominated by the tunneling process. During longer stress times (>1 s), an automated pulse-based BTI method was employed [27] in which the pulse I_d–V_g measurement followed a stress without delay. With longer stress, an additional increase in ΔV_{th} is observed at elevated temperatures, suggesting that ΔV_{th} in this stress time range is dominated by the slow temperature-assisted processes.

To verify that the ΔV_{th} time dependency might indeed represent trapping at pre-existing defects, the measured $\Delta V_{th}(t)$ values were fit to a first-order trapping kinetic equation [28] (see the solid line in Fig. 8.5), which describes the process of filling the pre-existing defects under constant flux conditions. The fitting was restricted to the first few μs of the stress pulse time, when the electrostatic repulsion between the trapped electrons is weak enough that the simple first order

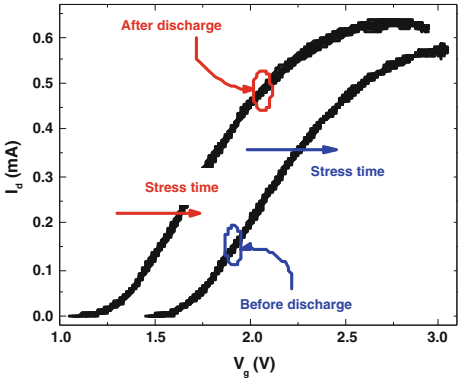


Fig. 8.4 Pulse I_d–V_g curves of 1.1 nm SiO_2/3 nm HfO_2/TiN devices measured after every 1,000 s intervals during 3 V/10^4 s CVS before and after discharge at $V_g = -1$ V for 30 s

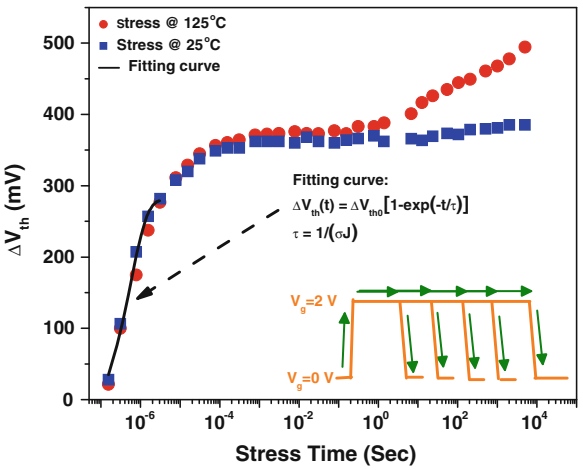


Fig. 8.5 Dependence of the threshold voltage shift (ΔV_{th}) on stress time at 25 and 125 °C stress temperatures. The solid line is the result of the data fitting by the first order electron trapping kinetic equation $\Delta V_{th}(t) = \Delta V_{th0}[1 - \exp(-t/\tau)]$, $\tau = 1/(\sigma J)$, where τ is the characteristic trapping time, σ is the trap capture cross section, and J is the flux density of the injected electrons

kinetic description remains valid. The fitting result suggests that charge trapping within this time range may indeed occur through a trap-filling process, which extracted density and capture cross-section on the order of 10^{14} and 10^{-13} cm^2, respectively [29], with a reasonably well defined characteristic time of 0.5 μs. An absence of temperature dependency points to a low activation energy associated with the fast trapping.

Fast trapping may be expected to be accompanied by fast detrapping, which would affect the magnitude of instability when monitored by DC measurements and result in an inaccurate prediction of device lifetime. It has been reported that in high-k devices the measured V_{th} under a bias temperature stress is strongly affected by the stress interruption time during sense measurements [27]. This V_{th} recovery has been attributed to the fast (μs range) charge relaxation (detrapping) [30, 31]. Such fast relaxation raises questions about the validity of conventional (relatively slow >10 ms) DC V_{th} monitoring during BTI stresses.

A charge relaxation was studied by superimposing negative voltage pulses of various pulse widths [25] with the positive constant voltage stress applied to the transistor gate (see insert in Fig. 8.6). A variety of trapped charge distributions in the dielectric resulting in different stress-induced V_{th} values was produced by varying the duration of the constant voltage stress (as defined by the moment at which the negative pulse was applied) and stress temperatures.

In Fig. 8.6, ΔV_{thr} represents the magnitude of the post-stress relaxation of the stress-induced V_{th} value (back to its initial pre-stress value) during the relaxation time defined by the pulse period, $\Delta V_{thr}(t) \equiv \Delta V_{th}(t = 0) - \Delta V_{th}(t)$ where $\Delta V_{th}(t = 0)$ is the threshold voltage shift (vs. the initial pre-stress V_{th} value) as extracted from the falling voltage portion of the pulse immediately after the stress, and $\Delta V_{th}(t)$ is the post-relaxation threshold voltage shift as measured from the rising voltage portion of the pulse of the width (relaxation time) t.

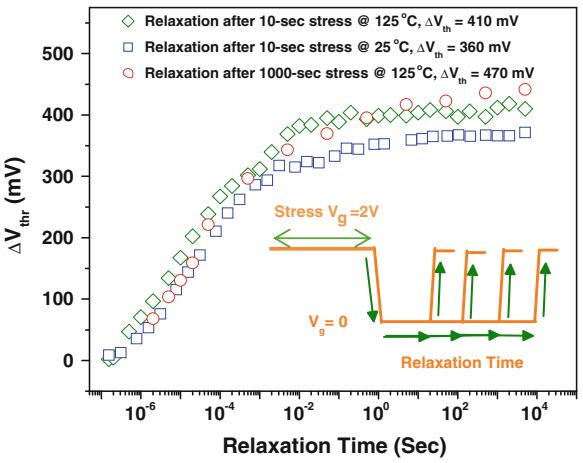


Fig. 8.6 Dependence of the magnitude of V_{th} relaxation, ΔV_{thr}, on the relaxation time for extremely different stress conditions: 10 s stress at room temperature and 1,000 s stress at 125 °C. ΔV_{th} values after the given stresses are also provided in the legend

Within a millisecond time range, ΔV_{thr} is independent of temperature and stress time. This suggests that the relaxation process in this time scale is dominated by the back tunneling of charges trapped through the fast transient process (see Fig. 8.5). The ΔV_{thr} saturation level reached at long relaxation times is determined by the characteristics of the stress applied before relaxation; higher relaxation ΔV_{thr} values are achieved with higher stress temperatures and longer stress times due to a higher stress-induced V_{th} shift (ΔV_{th}) caused by the additional temperature-assisted charge trapping [29] (see Fig. 8.5). Detrapping of these long-term stress-induced charges leads to a continuous ΔV_{thr} increase during longer relaxation times that could be associated with the temperature-assisted emission of the trapped electron to the conduction band of the high-k dielectric. As can be seen in Fig. 8.6, most of the ΔV_{th} value relaxes within milliseconds of stress interruption. Therefore, conventional DC-based measurements, which take at least a few milliseconds, may significantly underestimate PBTI.

The effect of the threshold voltage shift relaxation during stress interruption is illustrated in Fig. 8.7 where ΔV_{th} stress time dependency is monitored by applying single negative pulses of various widths as the sensing I_d–V_g measurements. V_{th} values extracted from the rising portion of the pulsed V_g include the effect of charge relaxation, which occurs when $V_g < V_{stress}$. The ΔV_{th} values obtained by the pulse measurements with shorter pulse widths are found to be higher due to less charge detrapping during the sense measurement. The downward vertical shift of $\Delta V_{th}(t)$ measured with wider pulses (longer relaxation times) is due to greater detrapping of the transient charges (that is, the charges that were trapped by the fast transient process), as was shown in Fig. 8.6. To eliminate the effect of discharging, V_{th} values were extracted from the falling portion of the pulse I_d–V_g curves. As can be expected in this case, the $\Delta V_{th}(t)$ values are both independent of the pulse duration and much larger than those obtained from the rising pulse voltage $\Delta V_{th}(t)$, Fig. 8.7.

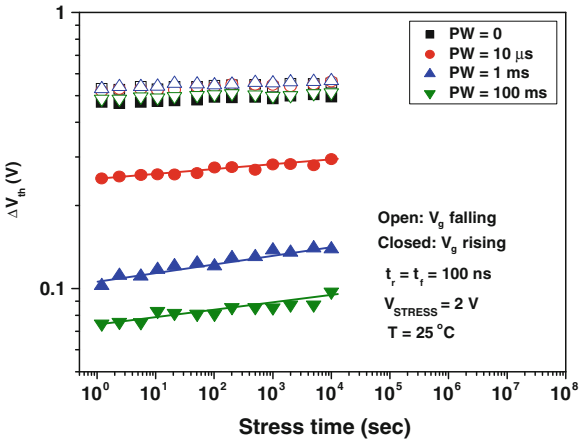


Fig. 8.7 The power law dependence of ΔV_{th} on stress time measured by a pulse-based BTI method with different pulse times. V_{th} values were extracted from the falling (without relaxation) and rising (includes relaxation) V_g portion of the pulse. See pulse schematic in Fig. 8.6

8.2.2 *Correction for Fast Instability Process*

As these data demonstrate, both the magnitude of the charging and its time dependency [the $\Delta V_{th}(t)$ slope] are strongly affected by sensing measurements. Fast charge detrapping during stress interruption, which is generally needed for sense measurements, results in greater power law exponent values [defined by the $\Delta V_{th}(t)$ slope] commonly used for long-term reliability projections. Even when the effect of relaxation on $\Delta V_{th}(t)$ is removed, as in the case of the ΔV_{th} values measured using the falling front of the V_g pulse in the proposed pulse BTI method, Fig. 8.7, the $\Delta V_{th}(t)$ slope is largely determined by the magnitude of the fast charging (i.e., the value of the first data point of the $\Delta V_{th}(t)$ dependency): the higher initial ΔV_{th}, which is controlled by the fast charging process, leads to a lower $\Delta V_{th}(t)$ slope, which determines long-term device degradation. This results in the reliability estimate being dependent on the specifics of the applied DC or pulse technique, which is undesirable from a practical standpoint.

To obtain a power law exponent value that is free from the parasitic effects of fast trapping/detrapping, their contribution should be subtracted from the total $\Delta V_{th}(t)$. Since the fast process is mostly completed within about 100 µs or less while the slow process kicks in after about 1 s, cf. Fig. 8.5, the ΔV_{th} value after about 1–2 s of stress includes the entire fast component with a minimal contribution from the slow process. After adjusting for the initial charging, which is determined by fast trapping/detrapping, $\Delta V_{th}(t) - \Delta V_{th}(t = 2\ s)$, the degradation rate becomes independent from the sense measurement times as demonstrated by identical slopes of the $\Delta V_{th}(t)$ data measured using different pulse rise and width times, as well as DC measurements, Fig. 8.8.

In the alternative method to address charge relaxation during sense measurements, V_{th} is calculated based on a drain current (I_d) degradation during stress that can be monitored without stress interruption, the so-called on-the-fly method [32]. However, this method does not allow measuring the initial I_d, which may be

affected by the fast transient charging that occurs within microseconds at the initial moment of stress in high-k nMOSFETs and causes significant errors in predicting device lifetime. Figure 8.9a schematically compares stress-sense schemes using the stress-interrupted DC I_d–V_g and single pulse methods to the on-the-fly method [33]. On-the-fly and DC measurements underestimate the initial ΔV_T, cf. Fig. 8.9b. However, by applying an adjustment methodology, all three techniques show identical degradation rates, see Fig. 8.9c. Here, all the adjusted V_{th} sets, $\Delta V_{th}'(t) \equiv \Delta V_{th}(t) - \Delta V_{th}(t = 1)$, exhibit identical time dependence even though the data were obtained by techniques with different characteristic relaxation times. This intrinsic time dependency, which is independent of the measurement technique, is attributed exclusively to slow charge trapping. The parallel shift of the curves is caused by the difference between the charge trapping/detrapping ratios during different sense measurements.

The $\Delta V_{th}(t)$ slope at long stresses can be used for reliability projections. To compare the estimate against the maximum allowed ΔV_{th} value, the adjusted $\Delta V_{th}(t)$ curves should be corrected up by the value of the initial fast trapping component.

A fast instability components can also be observed with hole injection stress (NBTI). Frequency-dependent charge pumping data (CP), see Fig. 8.10, exhibit a sharp increase in trap density at high CP frequencies (closer to the IL/Si interface) after the initial moment of stress, suggesting fast activation of the near-interface as-grown defects by the injected holes. This is consistent with expectations that the pre-existing defects active in the hole trapping process should be located in the dielectric close to the anode. Indeed, reported results [34, 35] suggest that the

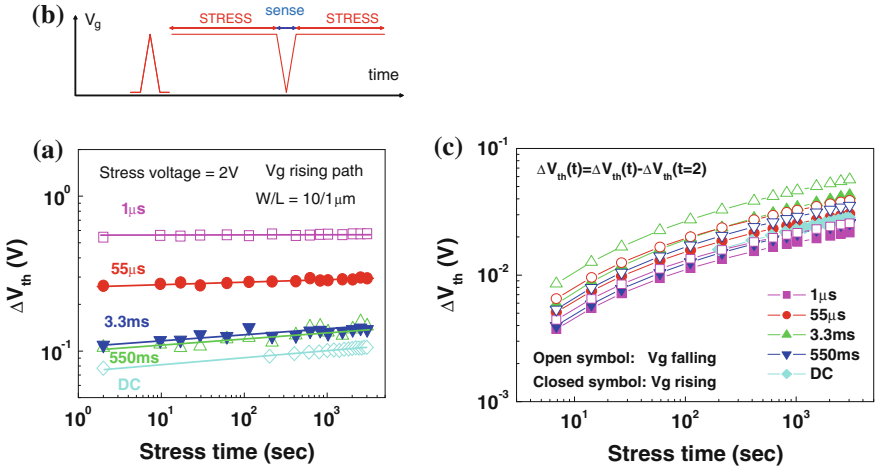


Fig. 8.8 **a** Stress time dependence of ΔV_{th} measured by a pulse-based BTI method with different sense times. V_{th} values were extracted from the rising (includes relaxation) V_g portion of the pulse as shown by the test diagram (**b**). **c** After adjusting for the initial fast charging, the degradation rate is independent of sensing time

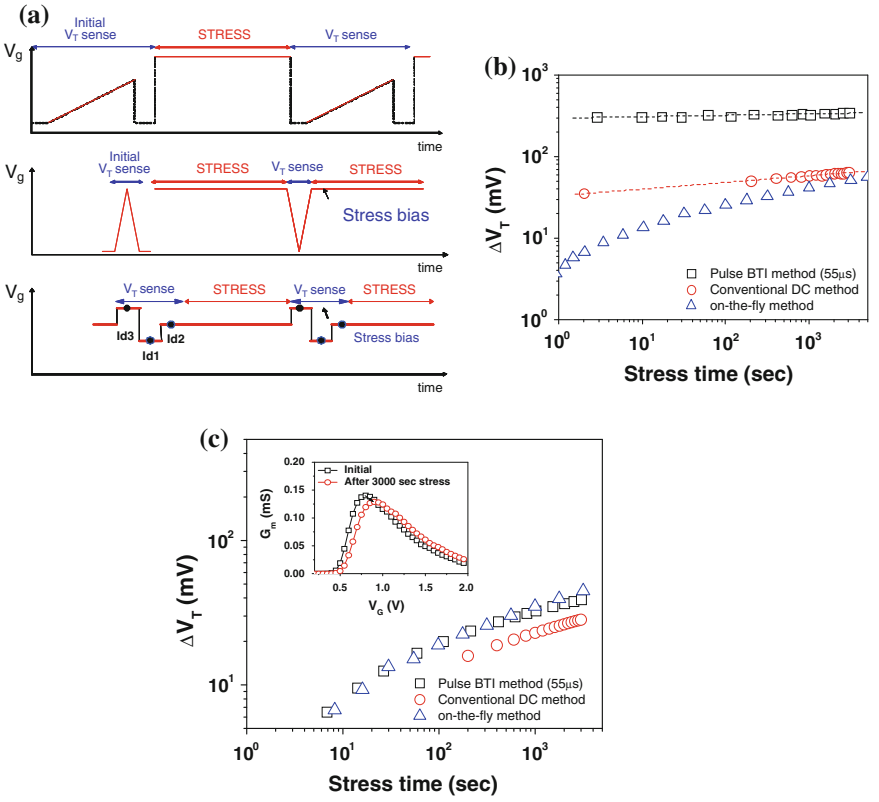


Fig. 8.9 **a** Schematic comparison of stress-sense schemes using the stress-interrupted DC I_d–V_g method, single pulse method, and on-the-fly method. **b** The comparison of the stress time dependence of ΔV_{th} measured using different characteristic methods. The stress time shown in the x-axis does not include sense measurement time. **c** Three techniques show identical degradation rate using adjustment methodology (Insert: G_m degradation might be a source of slightly lower values of DC data). Sample: 2 nm HfSiO (20 % SiO_2)/1 nm thermal SiO_2

NBTI in high-k gate stacks is largely caused by defects in the IL. Certain NBTI features in the high-k stacks were found to be different from those in conventional SiO_2 gate dielectrics due to modifications in the IL composition induced by its interaction with the overlaying high-k film. Both high-k and pure SiO_2 gate stacks show a similar long-term growth rate of stress-induced charging, $\Delta V_{th} \propto t^n$, $n \approx 0.2$, as monitored by the DCIV technique [35], see Fig. 8.11. However, for short stresses (<1 s), additional fast near-interface charging has been observed in high-k gate stacks, which could originate from the as-processed precursor defects, see Fig. 8.12. Spin-dependent recombination (SDR) measurements [36] indicate that these fast-activated hole traps near the Si interface could be related to Hf atoms, the density of which in the IL was found to be on the order of 10^{13} cm^{-2} [5].

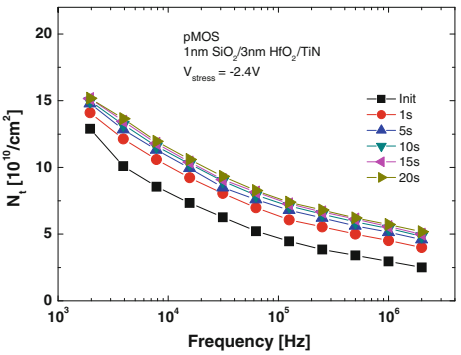


Fig. 8.10 Trap density as measured by the charge pumping method at different frequencies during the stress

To separate the fast trap generation process from the slow process, which is responsible for long-term ΔV_{th} degradation, we applied the approach in which the contribution of the fast transient trap generation process is approximated by the ΔV_{th} value after the first 1 s of stress ΔV_{th} (1 s). The slow, conventional trap generation is then obtained by subtracting ΔV_{th} (1 s) from the measured $\Delta V_{th}(t)$ dependence, Fig. 8.11. This procedure results in the high-k gate stack charging kinetics, which are similar to that of SiO_2. A similar adjustment procedure can be applied to the trapped charge relaxation kinetics after NBTI stresses. An example in Fig. 8.13 shows the kinetics of the interface trap recovery ΔD_{IT} at different relaxation voltages as measured by the DCIV current peak method [35]. The ΔD_{IT} relaxation values under the positive bias condition adjusted to the initial $\Delta D_{IT}(1\ s)$ transient (during the first 1 s of relaxation) are nearly identical to the 0 V bias curve, suggesting that the fast transient recover is field-driven.

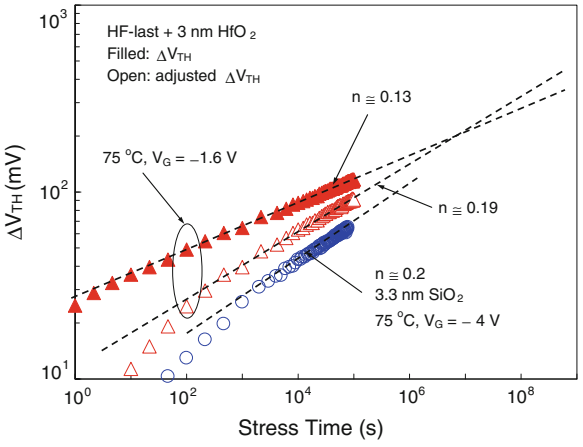


Fig. 8.11 Measured (*filled circles*) and adjusted for fast transient charging (*open circles*) V_{th} shift in HfO_2 and control SiO_2 devices

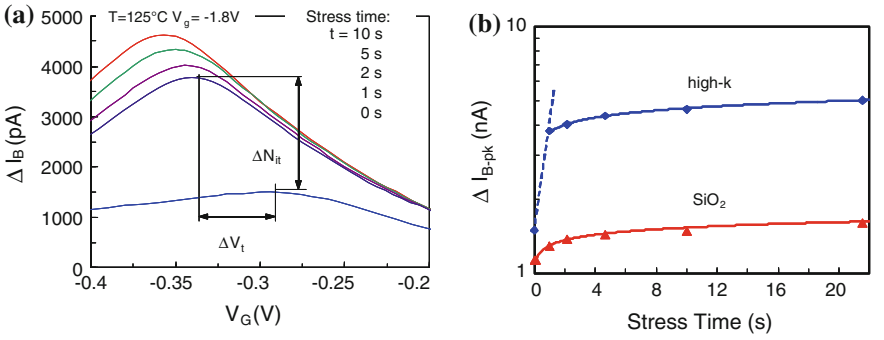


Fig. 8.12 **a** DCIV curves measured on pMOSFETs with 1.1 nm SiO_2/3 nm HfO_2 gate stack (EOT = 1.2 nm) for the stress conditions of $V_g = -1.8$ V/T = 125 C. The *bottom curve* is the pre-stress dependence; the other curves are measured at 1, 2, 5, and 10 s, respectively. The initial peak voltage shift $\Delta V_{GB\text{-}pk}(1\text{ s}) \cong 45$ mV is comparable to ΔV_{TH} (1 s) $\cong$ 50 mV. **b** Threshold shift and DCIV peak current increase ($\sim \Delta D_{IT}$) versus the stress time demonstrating contributions from fast and slow processes. For the stress times >1 s, each data set ($\Delta V_{TH}(t)$ and $\Delta I_{B\text{-}pk}(t)$) is fitted with a single power law function (*solid line*) for the stress interval from 1 to 10^4 s. Fitting the changes of the parameter values during the first 1 s of the stress time (*broken line*) requires much stronger time dependence (faster process)

The results presented here demonstrate that failure to correct for fast trapping/detrapping may result in incorrect lifetime estimations from both PBTI and NBTI tests.

8.2.3 Slow Instability

Data in Fig. 8.5 point to the existence of a slow, temperature-dependent process, which exhibits characteristic electron trapping time on the order of about

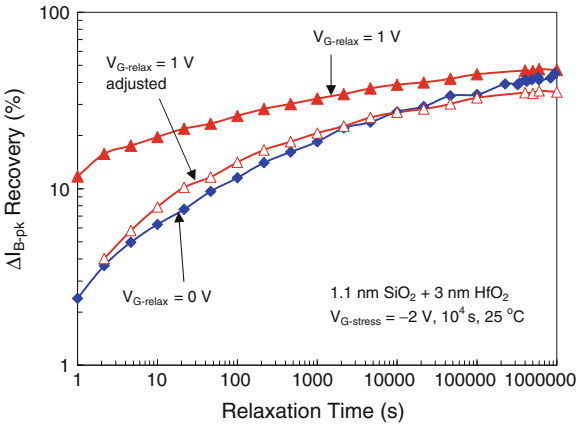


Fig. 8.13 Interface trap recovery, as monitored by the time dependency of the DCIV current peak value, $\Delta I_{B\text{-}pk}(t)$, at 0 V and +1 V relaxation voltages for high-k pMOSFETs stressed at −2 V/10^4 s at 25 °C, which resulted in $\Delta D_{IT} \sim 10^{11}$ cm^{-2}. The *open triangles* correspond to the adjusted $\Delta I_{B\text{-}pk}(t) - \Delta I_{B\text{-}pk}(1\text{ s})$ values

10^1–10^2 s (see Fig. 8.14), which is about six orders of magnitude higher than a fast process. An attempt to analyze slow charging similar to the way fast charging was examined (assuming that the pre-existing traps are filled directly by the current of injected electrons of the same density J as was used for describing the fast trapping in Fig. 8.5) yields an effective capture cross-section of $\sigma \approx 10^{-18}$–10^{-19} cm^2, indicating that the elastic capture of the electrons injected from the dielectric conduction band cannot be used to describe the slow trapping process. Such a significant difference in the σ values of the slow and fast ($\sigma \approx 10^{-14}$ cm^2) charging processes might suggest that the former is assisted by inelastic phonon-assisted processes.

Due to the significant difference in the characteristic charging times for fast (μs) and slow (sec) trapping, these processes can be separated within the adiabatic approximation when the slow process is considered assuming a steady-state population of traps filled by fast charging. Therefore, subtracting a magnitude of the parameter shift caused by the fast charging (during the initial stress period of about a second) can uncover the stress time dependency of the device parameter controlled by the slow instability processes.

Based on the assumption that the same traps contribute to both fast and slow electron trapping, a two-step charge trapping model was proposed to describe the PBTI temperature/voltage dependence observed in high-k gate stack transistors [28, 29]. The model suggests that a portion of the injected electrons may be captured through a tunneling process with a negligible activation energy (as follows from its weak temperature dependency) and localized in the shallow traps in the high-k dielectric. This is a fast transient charging process (in the ~μs range) responsible for the shifts of the pulse I_d–V_g curve during a short stress. The applied field then drives the temperature-activated migration of the trapped charges into the adjacent traps (Fig. 8.15, process B) during the BTI stress, which represents a much slower process (in the seconds range). The slow charging may then reflect the capture of secondary electrons, i.e., electrons supplied by the traps charged by the fast direct tunneling process [29]. A certain number of electrons temperature-activated from the traps into the conduction band can be recaptured by the nearby available traps before they gain kinetic energy and thus cannot be retrapped.

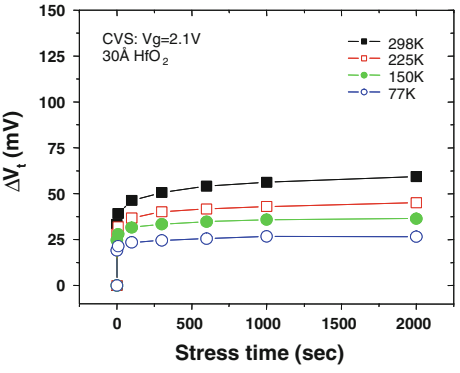


Fig. 8.14 Change of V_t values during $V_g = 2.1$ stress in the temperature range of 298–77 K

The large trap dimension, on the order of 1 nm as was estimated above, results in high delocalization of the trapped electrons allowing them to migrate along the conduction band by hopping via the trap-conduction band-trap process, without gaining kinetic energy (since trapped electrons are not subject to electric field acceleration), see Fig. 8.15a. The traps accessible by fast tunneling are instantly refilled after their detrapping, which provides a constant supply of secondary electrons to sustain the slow charging process.

By solving the first order kinetic equation, the time dependence of slow trapping can be obtained in the following form,

$$N = N_s \sum_i \left(1 - e^{-p_s(i)t}\right), \tag{8.1}$$

where N_s is now the density of the unoccupied secondary traps near the filled traps and $p_s(i) \equiv \sigma_s J_s(i)$ is the probability of the electron capture by these traps. Here J_s is the flux of the electrons available for this trapping process and σ_s is the trap capture cross-section for these electrons. In (8.1), it is taken into consideration that more than one type of electron traps (as indicated by the summation over i) may contribute to capturing the emitted electrons. In this model, J_s is determined by a rate of the electron emission from the fast-filled traps by thermal excitation:

$$J_s(i) = n\frac{1}{\tau}\exp(-E_i/kT) \tag{8.2}$$

where n is the density of the filled fast traps, τ is the detrapping time constant, and E_i is the effective detrapping activation energy (note that the activation energy

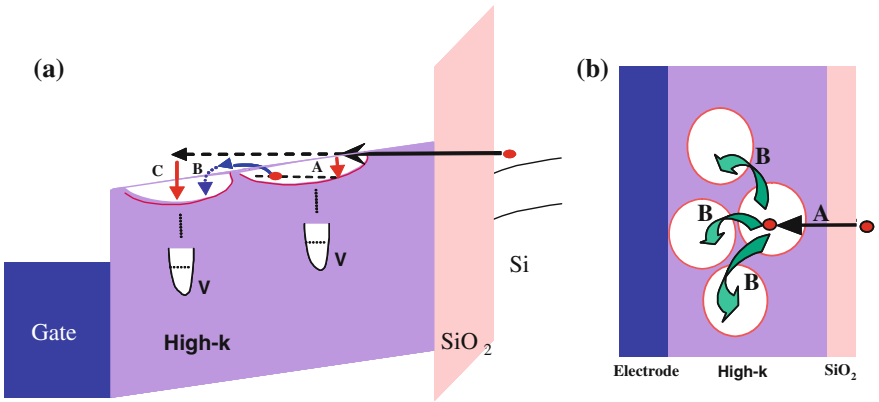


Fig. 8.15 **a** Schematic of the energy band diagram illustrating the proposed electron trapping processes: *A*—electron injection from the conduction band and trapping at a shallow defect (presumably an O-vacancy (V)), accompanied by lattice relaxation (*down arrow*); *B*—temperature-assisted electron detrapping/retrapping; *C*—capture of an electron accelerated in the conduction band is ineffective. **b** Schematic top view of the gate stack illustrating the proposed primarily (*A*) and secondary (*B*) electron trapping processes

depends on the lattice relaxation energy associated with atomic rearrangements caused by the charge localization/delocalization at a trap site) .

Fitting the $\Delta V_{t,}$ values after subtracting the fast trapping component ($\sim \Delta V_t$ after 1 s stress) for different stress voltages using (8.1) is shown in Fig. 8.16. The iterative fitting begins with a single term in the sum in (8.1) and increases the number of terms until a satisfactory fit is obtained. The fitting for all the voltages in the entire temperature range requires three terms in (8.1).

The extracted activation energies E_i are on the order of 0.35 and 0.45 eV, close to the reported value of 0.35 eV obtained from detrapping measurements [37]. The third extracted energy, $E_i < 0.01$ eV, may correspond to the polaron-type motion of the electrons in the conduction band. The N_s value, which corresponds to the density of the available traps appropriately located to recapture secondary electrons, is on the order of 10^9–10^{10} cm^{-2}, suggesting that most of the electrons activated from the primary traps are not recaptured. This offers insight into the origin of the high characteristic time constant of electron trapping during long-term constant voltage stress.

8.3 Defect Generation

Stress-induced degradation of the dielectric stack may originate from defect generation in both high-k and SiO_2 layers, where degradation is expected to be controlled by different mechanisms. Hence, the resulting time dependency of a given device characteristic represents a convolution of those associated with each active mechanism. Since relative contributions of these mechanisms to the overall degradation may vary with the stress conditions (voltage/temperature), a failure to separate these contributions and identify the origin of the leading degradation component—high-k or IL—might lead to erroneous reliability projections.

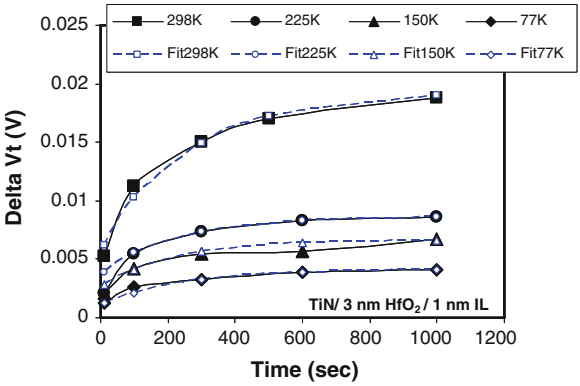


Fig. 8.16 Measured and modeled (open symbols) using (1.2) V_{th} shift associated with slow trapping during the $V_g = 1.95$ V stress at different temperatures

8.3.1 SILC as Gate Stack Degradation Monitor

Stress-induced leakage current (SILC) characteristics exhibited clear and unambiguous correlations with the major features usually present in the stress time dependency of the gate leakage current, namely, soft breakdown (SBD) and progressive breakdown (PBD), which can be demonstrated by studying stress-induced degradation in gate stacks of various high-k and IL thicknesses. nMOS transistors were fabricated using ALD HfO_2 gate dielectrics (with O_3 oxidation cycles) of 2, 3, 5, and 7 nm physical thicknesses [38]. These films were deposited on either an O_3-treated (001) epi Si substrate or 1 nm and 2.1 nm SiO_2 layers thermally grown by in situ steam generation (ISSG). The gate electrode was formed by a 10 nm TiN film capped with an N-doped poly-Si layer. The fabrication included a 1,000 °C/10 s dopant activation anneal and 485 °C forming gas anneal. The test samples also included MIM capacitors with 4.0 nm HfSiO (10 % SiO_2) film and TiN electrodes.

Constant voltage stress (CVS) in inversion interrupted for I_d–V_g, frequency-dependent (1 kHz–1 MHz) CP and SILC measurements was performed on the 10^{-7}–10^{-8} cm^2 ring capacitors and transistors, which exhibited correct area scaling of gate leakage currents. It was verified that the stress interruption for sense measurements—and associated stress interruption/resumption V_t changes caused by fast electron trapping/detrapping—did not change the kinetics of the dielectric degradation [38].

To prevent hard breakdown (HBD), gate current compliance limits were applied during CVS. As can be seen for the gate stack with a thick interfacial layer, 3 nm HfO_2/2.1 nm SiO_2, see Fig. 8.17, SILC increases after SBD although the gate current was limited by a $I_g = 65$ nA compliance, until HBD was reached, around 9.875×10^3 s. After HBD, SILC exhibited further growth only when the I_g compliance was increased to 500 mA.

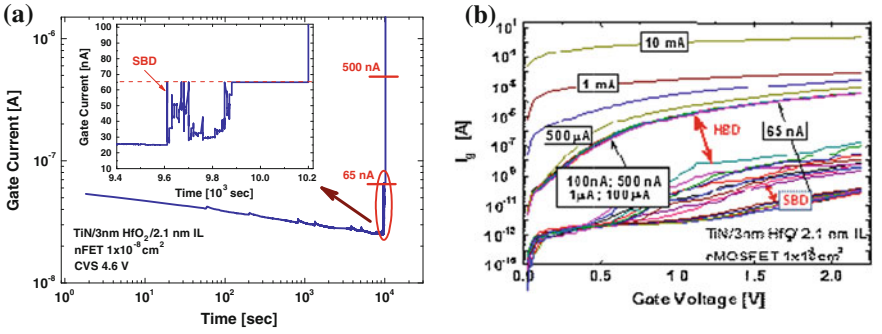


Fig. 8.17 **a** Gate leakage current during CVS in 3 nm HfO_2/2.1 nm SiO_2 gate stack nFET with current compliance limits. **b** SILC data collected under the current compliance limits (as labeled) during the stress presented in (**a**). *Double arrows* indicate SBD and HBD (HBD has occurred under the 65 nA I_g compliance condition: the flat region in the insert in a)

To understand the dynamics of the BD process, we developed a SILC analysis method based on the differential resistance $R_{diff}(V_g)$ and the slope of the differential resistance with respect to V_g, S_{Rdiff}, see Fig. 8.18 a and b, respectively, calculated from the SILC data. For pure ohmic conductance, resistance is independent of the applied voltage, $R_{diff}(V_g) = \text{Const}$ (i.e., $S_{Rdiff} = 0$); thus, the slope value $S_{Rdiff} \neq 0$ can be used as a figure of merit of ohmic versus non-ohmic (tunneling, hopping) conductance. Changes in R_{diff} and S_{Rdif} values (taken, for consistency, at a fixed $V_g = 1.5$ V) during the total stress time are plotted in Fig. 8.19. Detailed variations of S_{Rdiff} during stress before and after the SBD moment are shown in Fig. 8.20 a and b, respectively. Before SBD, the differential resistance decreased only slightly (Fig. 8.19) while the conduction mechanism (as reflected by S_{Rdiff}) remained mostly unchanged (non-ohmic). This points to the generation of isolated traps (which increased the total current) as a dominant degradation mechanism. At SBD, S_{Rdiff} abruptly decreased by several orders of magnitude, indicating that the conductance had qualitatively changed, most likely due to the formation of a conductive percolation path through the dielectric. Post-SBD evolution of S_{Rdiff} (Fig. 8.20b) and R_{diff} demonstrated a relatively rapid change (more than an order of magnitude within a few seconds) in conductance towards the ohmic regime that was indicative of continued degradation of the conductive percolation path, i.e., PBD. This degradation was primarily driven by stress time rather than stress current (since it occured under the current compliance limit). When the current compliance limit was raised to 1 mA and above, the runaway current formed a near-ohmic conductive path, which is a characteristic of the thermal HBD, accompanied by a several orders of magnitude drop in the S_{Rdiff} and R_{diff} values.

For the 5 nm HfO_2/1 nm SiO_2 stack, small step-wise increases in the leakage current, which correlate to the SILC increases, see Fig. 8.21a, could be qualified as SBD events since they were associated with an abrupt decrease of both R_{diff} and S_{Rdiff}, Fig. 8.21b. After the second SBD, the rate at which S_{Rdiff} decreased

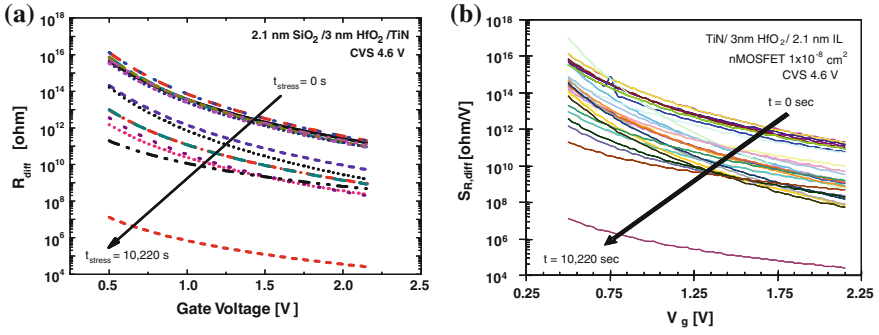


Fig. 8.18 Differential resistance (**a**) and slope of differential resistance (**b**) for SILC data (*smoothed*) in Fig. 8.17

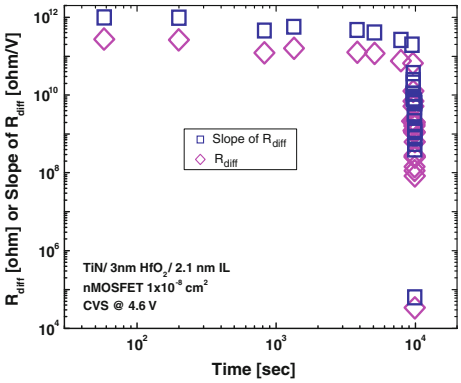


Fig. 8.19 Stress-induced evolution of the differential resistance, R_{diff}, and slope of the differential resistance, S_{Rdiff}

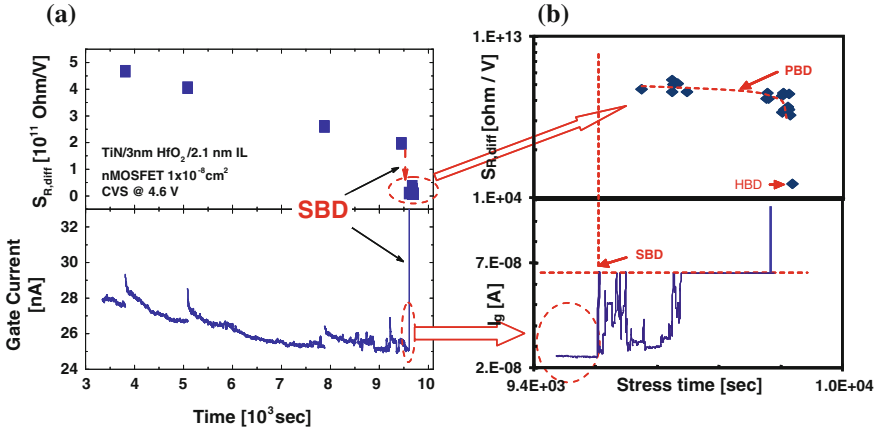


Fig. 8.20 **a** Evolution of the slope of the differential resistance and leakage current during CVS before SBD. **b** Post-SBD evolution of the slope of the differential resistance, S_{Rdiff}, and leakage current

remained unchanged until it abruptly decreased, suggesting that the subsequent increase in leakage current could be considered a manifestation of PBD.

The 7 nm HfO_2/1 nm SiO_2 sample demonstrated similar CVS characteristics with continuously increasing leakage current and interface trap density (N_{it}) correlated to decreasing R_{diff} and S_{Rdiff}, see Fig. 8.22. On the other hand, in the gate stacks with a smaller ratio of the equivalent oxide thicknesses (EOT) of the high-k to IL, as, for example, in the 2 nm HfO_2/1 nm SiO_2 and 3 nm HfO_2/1 nm SiO_2 stacks, see Fig. 8.23, BD was more abrupt. Similar to the above cases, the breakdown occurred, as manifested by the abrupt reduction of the differential resistance and its slope, under the I_g compliance. In our data, PBD was readily observed in thicker gate stacks with less of a drop in voltage across the IL while

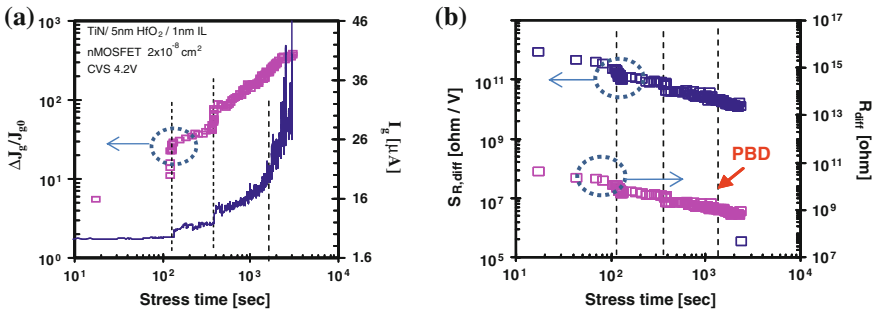


Fig. 8.21 Evolution of **a** SILC and stress leakage current and **b** differential resistance, R_{diff}, and its slope, S_{Rdiff}, during CVS in 5 nm HfO_2 stack (*PBD line* indicates onset of progressive BD regime)

the duration of the PBD quickly diminishes in thinner samples (when stress voltage was kept constant). Analysis of the CVS data showed that in the gate stacks with larger ratios, which resulted in proportionally more applied voltage to be dropped across the ILs, BD became more abrupt, with almost no signs of the PBD features.

8.3.2 SILC Origin

To understand which portion of the gate stack—IL or/and high-k—provides the dominant contribution to SILC, and hence to BD, the evolution of SILC was compared to that of the trap density as measured by the frequency-dependent CP, which probes traps in the dielectric at different distances from the substrate.

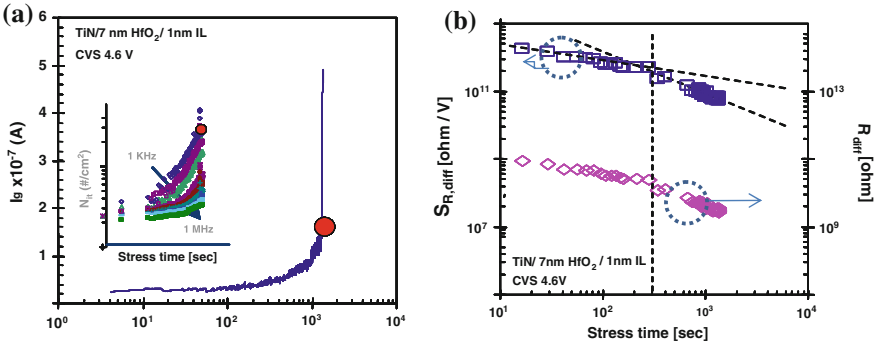


Fig. 8.22 **a** Variation of the gate leakage current in 7 nm HfO_2/1 nm SiO_2 gate stack during CVS at $V_g = 4.6$ V Insert: Variation of N_{it} measured at various CP frequencies. **b** Stress-induced evolution of the differential resistance, R_{diff}, and slope of the differential resistance, S_{Rdiff}. *Broken vertical line* indicates onset of progressive BD regime

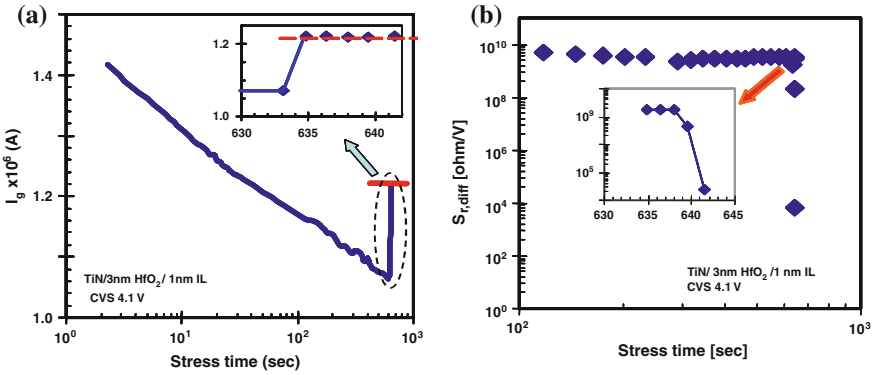


Fig. 8.23 **a** Variation of the gate leakage current in 3 nm HfO_2/1 nm SiO_2 gate stack during CVS at $V_g = 4.1$ V under the I_g compliance limit of 1.2 uA. **b** Stress-induced evolution of the slope of differential resistance, S_{Rdiff}

Normalized ΔN_{it}, measured within the CP frequency range of 1 MHZ–1 kHz in the samples with the 2.1 nm thick IL, is plotted with respect to the stress time in Fig. 8.24. During the measurements, the amplitude of the pulse was kept at 1.4 V, whereas the base was varied from −1.3 to 0 V with a step voltage of 0.1 V. The pulse rise/fall time was set at 100 ns. Under these test conditions, the CP probing depth in the dielectric was no more than 1.2 nm from the Si substrate [39, 40], which is entirely within the thickness of the 2.1 nm interfacial SiO_2 layer. While more defects were generated in the IL at the initial moment of stress closer to the high-k dielectric layer, the defect generation rate (the slope of the $N_{it}(t)$ curves) was independent of the CP frequency and, hence, constant throughout the probed IL thickness. This meant that the number of stress-generated defects at a given depth was proportional to the initial defect density, consistent with the suggestion that defects are generated at defect precursors [41]. The correlation of the changes in leakage current, SILC, and CP data during stress points to a common underlying physical cause. For a thin (∼1.1 nm) IL, the growth rates of SILC and interface

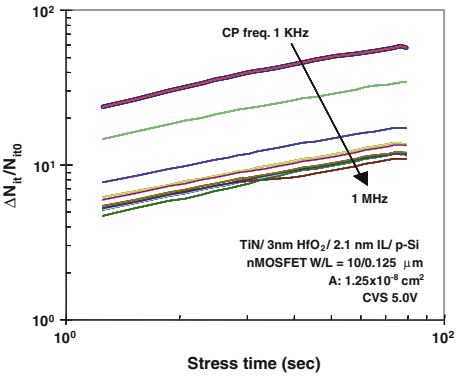


Fig. 8.24 Change of normalized N_{it} measured at different frequencies versus stress time interrupted for V_{th} measurements

trap density in each gate stack of a given high-k thickness are similar (as shown by the slopes of their time dependencies), see Fig. 8.25, indicating that changes in these characteristics were likely driven by the same defects. The fact that the N_{it} stress time dependency extracted from the CP measurements was similar for high and low CP frequencies, which probes traps in the IL near the interface with the Si substrate and deeper in the dielectric, respectively, suggested that the deeply embedded generated traps were also primarily located within the IL. The N_{it} (by low frequency CP, see Fig. 8.26a) and SILC (see Fig. 8.26b) growth measured during the same stress sequence correlated extremely well at all stress voltages, (see Fig. 8.27) [42], indicating that both characteristics are controlled by the same defects. To obtain such a correlation, only one layer, either the high-k or IL, can contribute to the SILC since these dielectrics should exhibit different voltage acceleration factors for defect generation. According to the CP simulation results [39], CP within the 1 MHz–1 kHz frequency range probes traps primarily within the IL and IL/high-k interface, see Fig. 8.28. (Here the distances represented the upper limit of the probing depth since the lattice structural relaxation associated with the charge trapping was not taken into account in [40]). Indeed, the trap density exhibited a monotonous growth within the total frequency range, as should be expected when the probed traps are located within a single material. Since the high frequency CP measurements are known to probe the IL, we could conclude that the low frequency data reflected on the traps within the IL are consistent with simulation results. For comparison, a rather uniform trap density profile was obtained in the transistors with a so-called "zero interface" stack of the ~0.3 nm IL [40]. In this case, the CP measurements probe exclusively (besides the Si interface) the HfO_2 film, and the trap density was found to be rather uniform through the entire range of the CP frequencies. Therefore, data in Fig. 8.27 suggest that SILC is controlled by the trap generation in the IL near its interface with the high-k. Alternatively, MIM high-k dielectric capacitors (no IL) did not exhibit an appreciable SILC up to the BD moment at any stress voltage, see Fig. 8.29, S_{Rdiff} and R_{diff} changes, or signs of PBD, see Fig. 8.30.

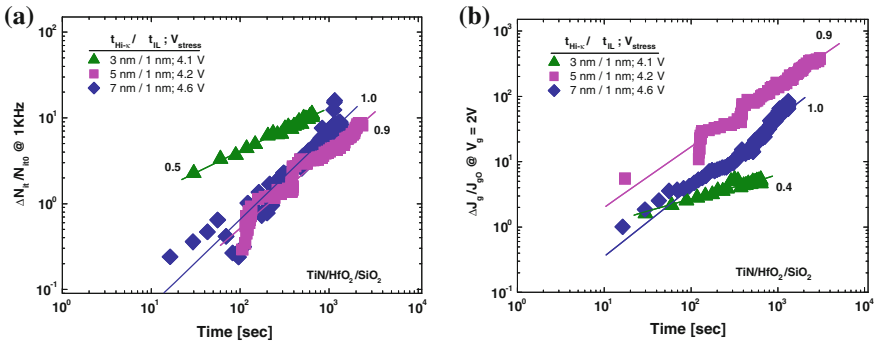


Fig. 8.25 SILC (**a**) and N_{it} (**b**) growth rates in the 3, 5, and 7 nm HfO_2 gate stacks. Similar SILC and N_{it} slopes are observed for each high-k thickness

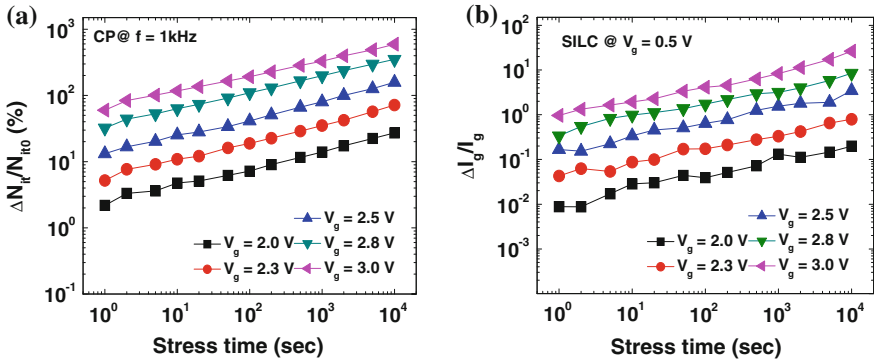


Fig. 8.26 **a** Stress time dependence of N_{it} at different stress voltages and **b** Stress time dependence of low voltage SILC at different stress voltages in 1.1 nm SiO_2/3 nm HfO_2/TiN nFET

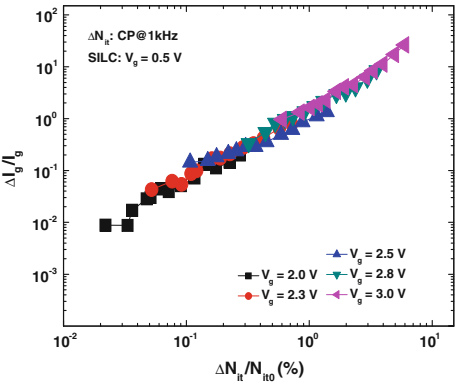


Fig. 8.27 Correlation of low voltage SILC and low frequency N_{it} stress time dependencies at various stress voltages (data from Fig. 8.26)

In fresh devices, a higher trap density in the IL closer to the high-k [41] was reported to be caused by the high-k/IL interaction during high-k deposition and subsequent device processing [5]. In nFETs stressed in inversion, the rate of trap generation, which is greater in the IL closer to the high-k layer, decreased with stress time, see Fig. 8.31. This observation is consistent with the proposed understanding that this trap generation was driven by the capture of electrons by precursor defects in the IL, most likely oxygen vacancies. Electron energy-loss spectroscopy (EELS), electron spin resonance (ESR), and X-ray photoelectron spectroscopy (XPS) data [43] showed these to be induced by the interaction of the IL with high-k/metal films. Ab initio calculations [44, 45] demonstrated that the electron captured by the Si–Si vacancy formed a specific E′ center (a negative O-vacancy), whose energy lay within the 2.2–3.5 eV range below the SiO_2 conduction band (depending on the specifics of the local amorphous SiO_2 matrix around the defect). The calculated g-matrix components of this negative O-vacancy, $g = 2.0023$ and 2.0027 [45], matched well with the zero crossing

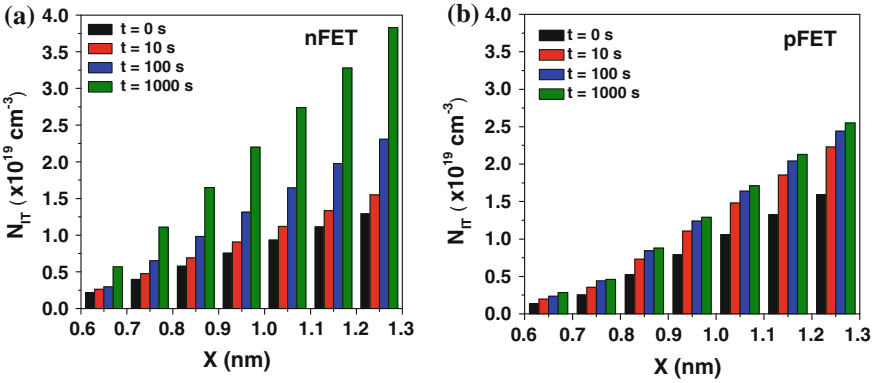


Fig. 8.28 Variation of the trap density through the thickness of the interfacial SiO_2 layer during stress at **a** $V_g = 3$ V in nFET (data from Fig. 8.2) and **b** $V_g = -3$ V in pFET

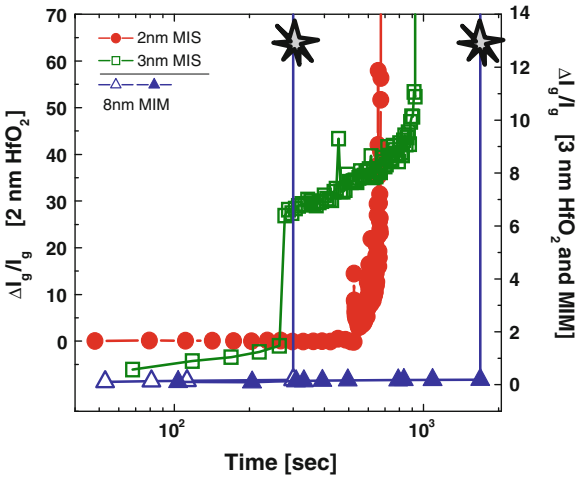


Fig. 8.29 Examples of typical SILC in 2 nm (CVS at $V_g = 2.7$ V) and 3 nm ($V_g = 3.7$ V) MIS gate stacks and 8 nm MIM ($V_g = 3.4$ and 3.3 V)

$g = 2.0026 \pm 0.0002$ observed by ESR in the IL signal, see Fig. 8.32a. These E′ defects were clearly present in the as-processed stack, and their density increased as a result of substrate injection stress, which was performed by applying a corona charge on a blanket film stack used for the ESR measurements, see Fig. 8.32a. The results indicated that these E′ centers were stress-activated from the defect precursors. The ESR measurements could underestimate the total number of these defects but not overestimate them since the measurements were sensitive to the centers in only a paramagnetic state. No such E′ centers were generated in the SiO_2 samples (see Fig. 8.32b), confirming the high-k-induced origin of these IL traps. In pFETs, CP data indicated, Fig. 8.10, that stress in inversion instantaneously activated as-grown defect precursors near the Si interface by injected holes (their possible atomic structure was discussed above).

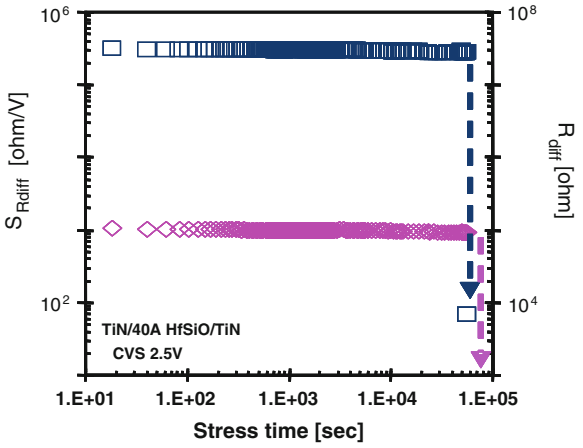


Fig. 8.30 Evolution of the differential resistance, R_{diff}, and its slope, S_{Rdiff}, during CVS of MIM caps

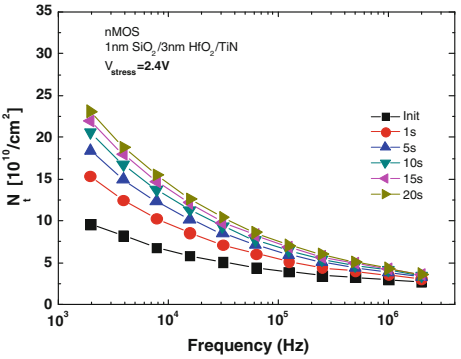


Fig. 8.31 Trap generation rate decreases with time; Trap generation greater with closer proximity to high-k layer

8.3.3 Verification of SILC Factors

To verify that SILC is associated exclusively with the bulk electron traps in the IL and that trap generation near the Si interface could be responsible for the gate leakage current increase, a substrate hot carrier (SHC) stress was performed under different carrier energy conditions. The SHC resulted in trap generation exclusively at the IL/Si interface, see Fig. 8.33, with no SILC, see Fig. 8.34, which is consistent with the above conclusion that SILC is controlled by the traps within the bulk of the IL. These data indicated that the bulk IL trap generation due to the possible release of H from the anode interface did not occur in the metal gate stacks as opposite to the poly-Si gates [46]. CVS and SHC results suggest that the IL bulk traps apparently were activated by electron trapping, which was much less effective during SHC stress, probably due to a significant mismatch between the precursor defect energies and those of the high energy electrons injected into the SiO_2 conduction band.

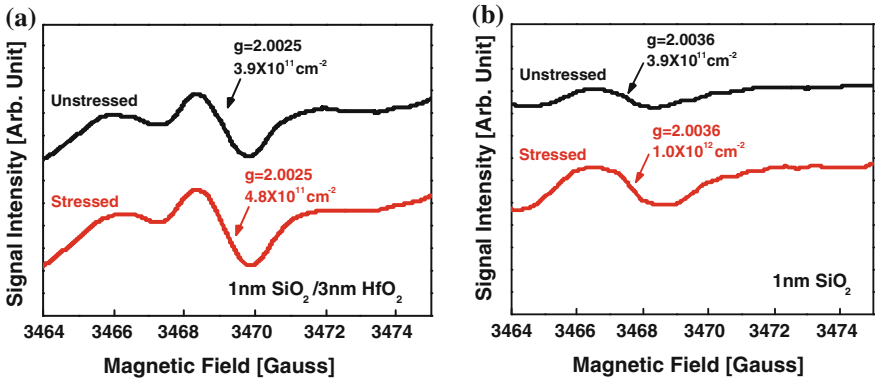


Fig. 8.32 ESR signal from the **a** IL of the 1.1 nm SiO_2/3 nm HfO_2 stack before and after corona stress and **b** 1.1 nm thermal SiO_2 before and after corona stress (only the P_b centers generation is observed)

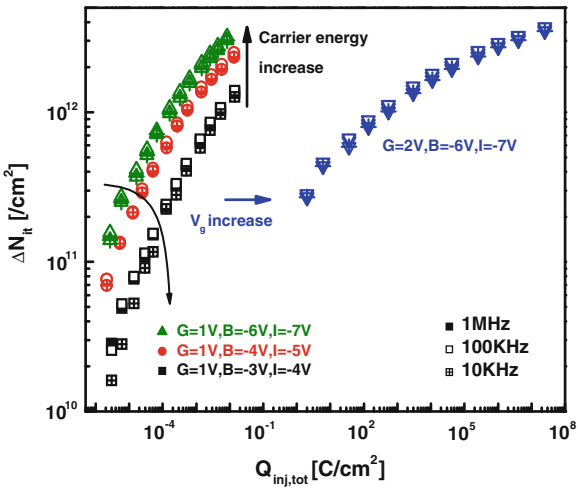


Fig. 8.33 Trap generation in IL (by CP at different frequencies) versus injected charge during HSC stress under various electron energies and gate bias conditions (*B*—substrate, *I*—injector, *G*—gate)

The stress time-dependent trap distributions in Fig. 8.28 were used to simulate SILC (measured during the same stress/sense sequence as CP) using the model, which considers the multi-phonon trap-assisted tunneling conduction mechanism, including effects of lattice distortion caused by electron trapping, random defect generation, and barrier deformation induced by the charged traps [47, 48]. The excellent agreement between the experimental and simulated I_g–V_g curves in the whole range of stress times, see Fig. 8.35, confirmed the above conclusion that SILC is controlled by defect generation in the IL. The trap energies were determined to be uniformly distributed within the 2.2–2.6 eV range from the bottom of the SiO_2 conduction band in both n- and pFETs (see Fig. 8.36), which correctly reproduced the maximum SILC increase, indicating that defect energy was aligned

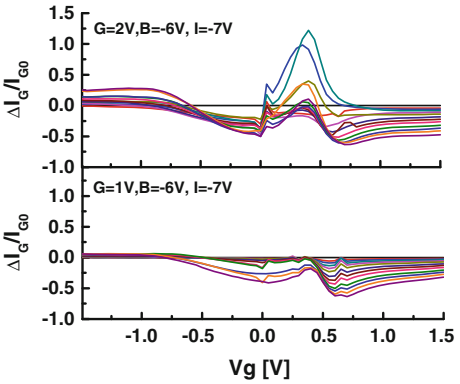


Fig. 8.34 SILC during HSC stress at high and low gate biases (*B*—substrate, *I*—injector, *G*—gate)

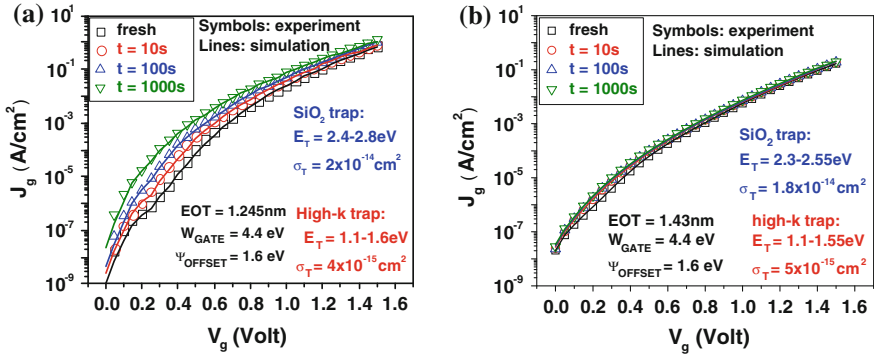


Fig. 8.35 **a** Measured (*symbols*) and simulated (*lines*) I_g–V_g curves during 1.1 nm SiO_2/3 nm HfO_2/TiN nFET stress $V_g = 3$ V. **b** Measured (*symbols*) and simulated (*lines*) I_g–V_g curves during 1.1 nm SiO_2/4 nm HfO_2/TiN pFET stress $V_g = -3$ V

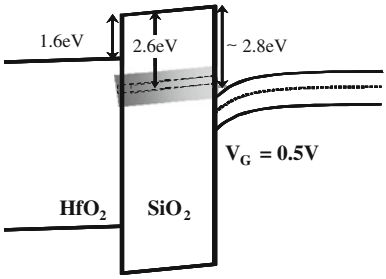


Fig. 8.36 Band diagram for 1.1 nm SiO_2/3 nm HfO_2/TiN nFET gate stack. Energy distribution of the stress-generated defects exhibiting a non-uniformed depth profile is shown by a *shaded rectangular area*

with the electron-quantized levels in the Si channel. No trap generation in high-k film was required in order to fit the data, and the high-k density of traps supporting the trap assisted tunneling (TAT) current was found by simulations to remain

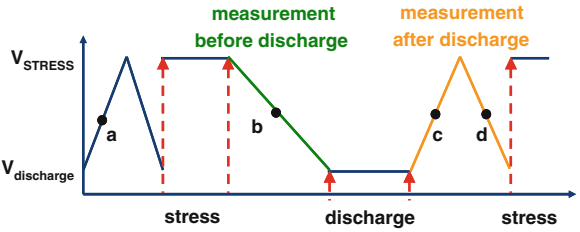


Fig. 8.37 Schematic of the stress-discharge-sense sequence with the pulse (*triangle* pulse waveform with $t_r = t_f = 100$ ns) sense measurement used in this work. V_{tha}: trap-free V_{th} value (intrinsic). V_{thb}: may include contribution from filled pre-existing and generated electron traps. V_{thc}: may include contribution from filled hole traps and positive fixed charges. V_{thd}: may include contribution from transient electron traps only

constant at $\sim 2 \times 10^{19}$ cm^3. In agreement with the simulation results, the pulse I–V measurements, which were sensitive to the density and characteristics of the electron traps in the high-k dielectrics (see detailed discussion above), did not show any trap generation within the sensitivity of this description, see Fig. 8.4.

These results demonstrated that the IL is the major factor controlling the overall degradation and breakdown of the metal/high-k gate stacks in inversion. Defects contributing to gate stack degradation were associated with the metal/high-k-induced oxygen vacancies in the IL.

8.3.4 Fast Defect Generation Caused by Hole Injection

A relatively high density of as-grown defects in high-k stacks, which activation under stress could manifest as time-dependent degradation of the device parameters, might mask the generation of new stress-induced defects. In addition to filling pre-existing traps in the high-k film and generating new electron traps in the IL, electron injection during PBTI stress can generate defects that act as hole acceptors. These traps, although not directly observable in PBTI measurements, can change transistor characteristics during off-state cycles. To identify potential hole traps, we applied a specially developed stress/discharge/sense scheme implemented using both DC and pulse measurements, see Fig. 8.37 [49].

The data in Fig. 8.38 demonstrates the effect of the discharge step on the shift of the threshold voltage, ΔV_{th}, and transconductance, Δg_m, with respect to their initial values. As widely reported in the literature, the magnitude of ΔV_{th} measured with no discharge increased during PBTI stress. However, when the sense measurements were performed after the discharge (a low voltage stress of the opposite polarity applied for a few seconds), ΔV_{th} decreased with stress time, which points to a build-up of positive charges during stress. The identical Δg_m trends measured either with or without discharge, see Fig. 8.38, suggested that the discharge step did not contribute to the generation of these positive defects. The absence of any

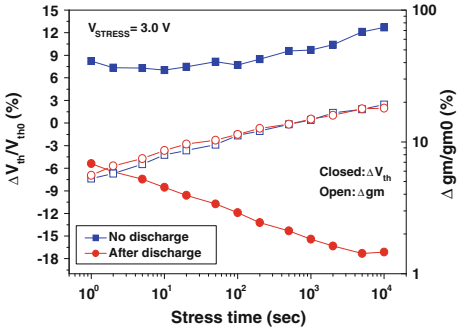


Fig. 8.38 ΔV_{th} and Δg_m measured without and after discharge at $V_g = -1$ V for 30 s using DC I_d–V_g

observable difference in the ΔV_{th} and ΔN_{it} trends with discharge time indicates that the low discharge bias did not degrade the dielectric. This negative shift could be attributed to the generation of hole-trapping defects within the interfacial SiO_2 layer in the high-k gate stack.

To delineate reversible and permanent PBTI contributions, V_{th} was measured at the end of each stress period before and after the discharge step. This was intended to empty both pre-existing and stress-generated electron traps. When the sense measurements followed the discharge, ΔV_{th} was seen to decrease with PBTI stress time, pointing to a build-up of positive charges during stress, see Fig. 8.38. The stress time dependence of ΔV_{th}, ΔN_{it}, and Δg_m measured with different discharge times confirmed that the discharge operation did not noticeably increase the defect density and that the positively charged defects were indeed caused by the PBTI stress.

To understand whether the defects responsible for the positive charge build-up during discharge represented fixed positive charges or hole traps, which might be reversibly filled and detrapped during the positive and negative bias cycles, the stress/discharge cycles were performed every 1 s during the first 10 s, then every 10 s during the subsequent 100 s, etc., as shown in Fig. 8.39. The stress was applied under a low voltage condition to minimize electron trap generation. The ΔV_{th} measured before the discharge may include contributions from the trapped

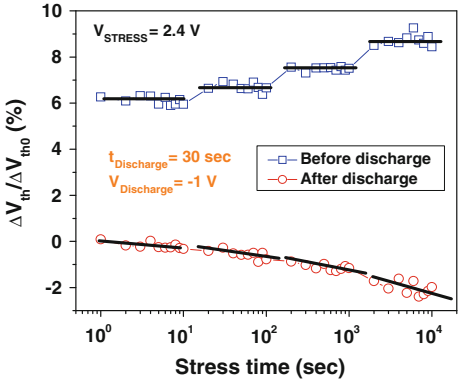


Fig. 8.39 ΔV_{th} measured before and after discharge from groups with different stress intervals

electrons and the fixed positive charges (the hole traps would be detrapped under the positive bias condition); the post-discharge ΔV_{th} may include the contribution from positive fixed charges and/or hole traps.

In each group of ten measurements within a given stress cycle interval (1–10 s, 10–10^2 s, 10^2–10^3 s, etc.), the values of ΔV_{th} measured before the discharge operation are approximately constant (a zero slope in the trend of the ΔV_{th} versus number of measurements in Fig. 8.39) while the ΔV_{th} values measured after the discharge were progressively negative (a negative slope of the ΔV_{th} versus number of measurements trend), indicating that the hole traps, rather than the fixed positive charges, were created by the positive bias stress.

Note that the hole trapping at the available traps occurred through the fast transient process, as follows from the pulse data. Therefore, the slow time dependency of the post-discharge ΔV_{th} seen in Figs. 8.39 was determined by the hole trap generation (N_t) during the PBTI stress rather than by the fast filling of the hole traps during the discharge, i.e., $\Delta V_{th}(t)$ is linearly proportional to $N_t(t)$. Thus, the kinetics of the hole trap generation by PBTI stress could be addressed by analyzing the stress time dependency of the post-discharge ΔV_{th}. The latter was fitted with the solution of a simple kinetic equation describing the charge trapping at pre-existing defects: $dN_t(t)/dt = \alpha(N_0 - N_t(t))$. Here N_t and N_0 are the densities of the PBTI-generated hole traps and available precursor defects, which could be converted to hole traps by trapping electrons during the PBTI stress, and α is the electron trapping rate at the precursor sites. A reasonably good fit suggests that the hole traps might be generated from the as-grown defect precursors (Fig 8.40).

A common feature for all samples exhibiting negative ΔV_{th} is the presence of nitrogen in the high-k gate stacks [43]. Nitrogen can be incorporated as a result of the pre-gate NH_3 treatment of the Si surface, the high-k PDA in the NH_3 ambient, or the ALD nitrogen-containing precursors. The ΔV_{th} was measured after the discharge on high-k samples with and without nitrogen and on the SiON gate dielectric devices. The absence of a negative ΔV_{th} in the SiON samples as well as in the nitrogen-free high-k sample strongly suggested that the negative post-discharge ΔV_{th} shift in nMOSFETs, which is caused by PBTI stress, is associated with only the high-k dielectric stacks containing nitrogen.

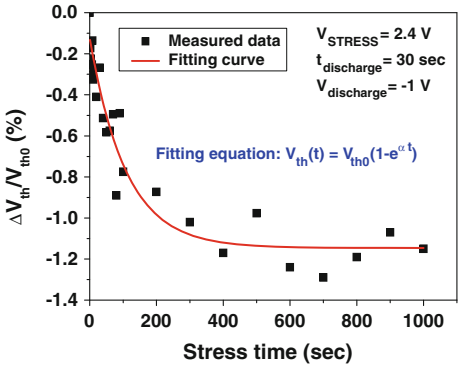


Fig. 8.40 The stress time dependence of the negative ΔV_{th} measured after discharge and the data fitting using the solution of the a simple kinetic equation for hole trap generation

8.4 Concluding Remarks

Multi-component high-k gate stack structures, which usually include silicon dioxide and transition metal oxide dielectric layers in contact with the metal electrode, exhibit electrical properties that might significantly differ from those of conventional SiO_2-based gate dielectrics. Inter-material interactions in the gate stack were shown to result in a high density of as-grown defects and defect precursors, giving rise to fast transient charging/discharging instability and spatially dependent trap generation, respectively. Consequently, interpretations of electrical data that depend on resolving time and spatial components of the electrically active processes require refining the existing and developing new analysis methodologies. Since a leading degradation mechanism, which depends on the properties of the dielectric layer contributing the most to the overall gate stack degradation, may change with the stress conditions, in particular, stress voltage and temperature, a widely used accelerated stress approach may cause erroneous lifetime estimates and, therefore, should be exercised with caution. Successfully identifying the weak link in the multi-component gate stack allows the governing degradation mechanism to be determined and, thus, enables reliability to be evaluated correctly.

References

1. G. Bersuker, B.H. Lee, H.R. Huff, Int. J. High Speed Electron. Syst. **16**, 221 (2006)
2. D. Muñoz Ramo, J.L. Gavartin, A.L. Shluger, G. Bersuker, Phys. Rev. B **75**, 205336 (2007)
3. J. Robertson, Rep. Prog. Phys. **69**, 327 (2006)
4. P. Kirsch et al., J. Appl. Phys. **99**, 023508 (2006)
5. G. Bersuker, C.S. Park, J. Barnett, P.S. Lysaght, P.D. Kirsch, C.D. Young, R. Choi, B.H. Lee, J. Appl. Phys. **100**, 094108 (2006)
6. G. Bersuker, J. Barnett, N. Moumen, B. Foran, C.D. Young, P. Lysaght, J. Peterson, B.H. Lee, P.M. Zeitzoff, H.R. Huff, Jap. J. Appl. Phys. **43**, 7899 (2004)
7. S.J. Wang, P.C. Lim, A.C.H. Huan, C.L. Liu, J.W. Chai, S.Y. Chow, J.S. Pan, Q. Li, C.K. Ong, Reaction of SiO_2 with hafnium oxide in low oxygen pressure. Appl. Phys. Lett. **82**, 2047 (2003)
8. D.Y. Cho, S.J. Oh, Y.J. Chang, T.W. Noh, R. Jung, J.C. Lee, Role of oxygen vacancy in HfO2/SiO2/Si(100) interfaces. Appl. Phys. Lett. **88**, 193502 (2006)
9. W.L. Scopel, A.J.R. da Silva, W. Orellana, A. Fazzio, Comparative study of defect energetics in HfO2 and SiO2. Appl. Phys. Lett. **84**, 1492–1794 (2004)
10. N. Capron, P. Broqvist, A. Paquarello, Migration of oxygen vacancy in HfO2 and across the HfO2/SiO2 interface: A first-principles investigation. Appl. Phys. Lett. **91**, 192905 (2007)
11. M. Ikeda, G. Kresse1, T. Nabatame and A. Toriumi, Theoretical analysis of oxygen diffusion in monoclinic HfO2. Mat. Res. Soc. Symp. Proc. **786**, E5.4.1 (2004)
12. Y.-S. Lin, R. Puthenkovilakam, J.P. Chang, C. Bouldin, I. Levin, N.V. Nguyen, J. Ehrstein, Y. Sun, P. Pianetta, T. Conrad, W. Vandervorst, V. Venturo, S. Selbrede, J. Appl. Phys. **93**, 5945 (2003)
13. D. Chi, P.C. McIntyre, Appl. Phys. Lett. **85**, 4699 (2004)
14. A. Pasquarello et al. Appl. Sur. Sci. **104/105**, 317 (1996)

15. S. Stemmer, Z. Chen, C.G. Levi, P.S. Lysaght, B. Foran, J.A. Gisby, J.R. Taylor, Application of metastable phase diagrams to silicate thin films for alternative gate dielectrics. Jpn. J. Appl. Phys. **42**, 3593–3597 (2003)
16. K. van Benthem et al., Appl. Phys. Lett. **87**, 034104 (2005)
17. A.G. Marinopoulos, K. van Benthem, S.N. Rashkeev, S.J. Pennycook, S.T. Pantelides, Phys. Rev. B **72**, 195317 (2008)
18. J.H. Sim, R. Choi, Y.H. Lee, C. Young, P. Zeitzoff, D.L. Kwong, G. Bersuker, Jap. J. Appl. Phys. **44**, 2420 (2005)
19. S. Zafar, A. Kumar, E. Gusev, and E. Cartier, IEEE Trans. Device Mater. Reliab 545, (2005)
20. G. Ribes, J. Mitard, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy, E. Vincent, G. Ghibaudo, IEEE Trans. Device Mater. Reliab. **5**, 5 (2005)
21. R. Choi, S.J. Rhee, B.H. Lee, J.C. Lee, G. Bersuker, IEEE Electron. Device Lett. **26**, 197 (2005)
22. D. Munoz Ramo, A.L. Shluger, J.L. Gavartin, G. Bersuker, Phys. Rev. Lett. **99**, 155504 (2007)
23. A. Kerber, E. Cartier, L. Pantisano, M. Rosmeulen, R. Degraeve, T. Kauerauf, G. Groeseneken, H. E. Maes, and U. Schwalke, in *Proceedings of 41st IEEE International Reliability Physics Symposium* (2003), pp. 41–44
24. C.D. Young, Y. Zhao, D. Heh, R. Choi, B.H. Lee, G. Bersuker, IEEE Trans. Electron Device **56**, 1322 (2009)
25. D. Heh, C.D. Young, G. Bersuker, IEEE Electron Device Lett. **29**, 180 (2008)
26. B.H. Lee, C.D. Young, R. Choi, J.H. Sim, G. Bersuker, C.Y. Kang, R. Harris, G.A. Brown, K. Matthews, S.C. Song, N. Moumen, J. Barnett, P. Lysaght, K.S. Choi, H.C. Wen, C. Huffman, H. Alshareef, P. Majhi, S. Gopalan, J. Peterson, P. Kirsh, H-J. Li, J. Gutt, M. Gardner, H.R. Huff, P. Zeitzoff, R.W. Murto, L. Larson, and C. Ramiller, IEDM Tech. Dig. 859 (2004)
27. D. Heh, R. Choi, C.D. Young, B.H. Lee, G. Bersuker, IEEE Electron Device Lett. **27**, 849 (2006)
28. G. Bersuker, J. H. Sim, C. S. Park, C. D. Young, S. Nadkarni, R. Choi, and B. H. Lee, Proc. Int. Reliab. Phys. Symp. 179 (2006)
29. G. Bersuker, J.H. Sim, C.S. Park, C.D. Young, S. Nadkarni, R. Choi, B.H. Lee, IEEE Tran. Device Mater. Reliab. **7**, 138 (2007)
30. G. Bersuker, J.H. Sim, C.D. Young, R. Choi, P.M. Zeitzoff, G.A. Brown, B.H. Lee, R.W. Murto, Microel. Reliab. **44**, 1509 (2004)
31. R. Choi, S.J. Rhee, J.C. Lee, B.H. Lee, G. Bersuer, IEEE Electron Device Lett. **26**, 197 (2005)
32. M. Denais, C. Parthasarathy, G. Ribes, Y. Rey-Tauriac, N. Revil, A. Bravaix, V. Huard, and F. Perrier, IEDM Tech. Digest, 109 (2004)
33. D. Heh, R. Choi, G. Bersuker, IEEE Elec. Dev. Lett. **28**, 245 (2007)
34. A. Neugroschel, G. Bersuker, R. Choi et al. IEDM Tech Digest, 317 (2006)
35. A. Neugroschel, G. Bersuker, R. Choi, B.H. Lee, IEEE Trans. Device Mater. Reliab. **8**, 47 (2008)
36. C.J. Cochrane, P.M. Lenahan, J.P. Campbell, G. Bersuker, A. Neugroschel, Appl. Phys. Lett. **90**, 123502 (2007)
37. G. Ribes, M. Muller, S. Bruyere, D. Roy, M. Denais, V. Huard, T. Skotnicki, and G. Ghibaudo, Presented at European Solid-State Device Research Conference, (2004), p. 89
38. G. Bersuker, N. Chowdhury, C. Young, D. Heh, D. Misra, and R. Choi, Proc. IEEE IRPS 49 (2007)
39. D. Heh, C.D. Young, G.A. Brown, P.Y. Hung, A. Diebold, E.M. Vogel, J.B. Bernstein, G. Bersuker, IEEE Trans. Electron Devices **54**, 1338–1345 (2007)
40. D. Veksler, G. Bersuker, A. Koudymov, C.D. Young, M. Liehr, and B. Taylor, in *Proceedings of IEEE International Reliability Physics Symposium* GD.4.1 (2011)
41. C.D. Young, D. Heh, S. Nadkarni, R. Choi, J.J. Peterson, J. Barnett, B.H. Lee, P. Zeitzoff, G.A. Brown, G. Bersuker, IEEE Trans. Device Mater. Reliab. **6**, 127 (2006)
42. G. Bersuker, D. Heh, C. Young et al. IEDM Tech. Digest (2008)

43. P. Lysaght, J. Barnett, G. Bersuker, J.C. Woicik, D.A. Fischer, B. Foran, H.-H. Tseng, R. Jammy, J. Appl. Phys. **101**, 024105 (2007)
44. P.V. Sushko, S. Mukhopadhyay, A. Mysovsky, V.B. Sulimov, A. Taga, A.L. Shluger, J. Phys. Condens. Mater **17**, S2115 (2005)
45. A.V. Kimmel, P.V. Sushko, A.L. Shluger, G. Bersuker, ECS Trans. **19**(2), 3 (2009)
46. D.J. DiMaria, J. Appl. Phys. **86**, 2100 (1999)
47. L. Larcher, IEEE Trans. Electron Devices **50**, 1246 (2003)
48. L. Vandelli, A. Padovani, L. Larcher, R.G. Southwick III, W.B. Knowlton, G. Bersuker, IEEE Trans. Electron Devices **58**, 2878 (2011)
49. D. Heh, P.D. Kirsch, C.D. Young, C.Y. Kang and G. Bersuker, Proc. IEEE IRPS 347 (2008)

Chapter 9
Lanthanide-Based High-k Gate Dielectric Materials

Daniel J. Lichtenwalner

Abstract This chapter covers selected issues related to lanthanides (or lanthanoids) used in oxide gate dielectrics. In general, lanthanides offer key material property advantages for gate dielectric applications. These include high dielectric constants, stable amorphous silicate (or ternary) phases, thermodynamic stability with Si, interface properties allowing threshold voltage tuning, and crystalline properties making possible epitaxial growth on Si and other semiconductors. Although not as heavily researched as Hf-based materials, lanthanide materials continue to hold promise for device scaling on Si, and as dielectrics on other high-mobility semiconductors.

9.1 Introduction to Lanthanide Dielectrics

Lanthanides (or lanthanoids, or rare-earths) have atomic properties making them key components for materials designed for a wide range of material properties. For example, lanthanides are key components for enhancing magnetic, optical, catalytic, and electrical properties of many materials. The dielectric properties of their respective oxides make them of interest as candidates for high dielectric constant (high-k) gate dielectrics. For example, initial interest was generated by binary compounds such as lanthanum oxide or lanthana (La_2O_3) and ternary compounds such as lanthanum aluminate ($LaAlO_3$) and lanthanum scandate ($LaScO_3$) found to have dielectric constants (k) in the range of 25–30, and sufficiently large band gaps for use as dielectrics on silicon, in the 5.0–6.0 eV range [1–3]. Additionally, from

D. J. Lichtenwalner (✉)
CREE, Inc., 4600 Silicon Drive, Durham, NC 27703, USA
e-mail: daniel_lichtenwalner@cree.com

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5_9,

a thermodynamic viewpoint, it appears that lanthanide oxides are chemically stable in contact with Si [2]. Additional reasons fueling interest in these materials as metal–oxide–semiconductor field-effect transistor (MOSFET) gate dielectrics include the following: (1) many of these lanthanide oxide materials have not been comprehensively studied in bulk form, certainly not the many potential ternary alloy options; (2) the properties of metastable amorphous thin films of these binary or ternary alloys were in many cases unknown; and (3) the fact that some properties are 'tunable' as one moves across the series due to f-orbital filling levels opens possibilities for desirable properties to be obtained.

This interest in lanthanides in the arena of high-κ gate dielectrics has continued to strengthen as it has been shown that lanthanide materials: can provide MOSFET threshold voltage (V_T) control to achieve band-edge values for n-channel devices [4]; and are promising as dielectrics on alternate channel materials such as GaAs or other III–V compounds [5], as lanthanides are chemically similar to group III elements in some respects.

As is the case for all materials and applications, a host of factors must be considered when selecting particular materials for a given application. Although the lanthanides are similar in many respects (all have electronegativity values of $\sim$1.1, most have +3 valence), and have some properties which vary smoothly across the series (such as ionic radius), they differ in various ways. For example, for stable dielectric properties which are insensitive to electric field or temperature or oxidation environment, it is typically essential to have cation elements having a single stable oxidation state. This would favor the use of La, Nd, Gd, Dy, Ho, Er, and Lu which have stable +3 oxidation states (disregarding the radioactive Pm) over the other lanthanide elements which have some tendency towards mixed +3 and +2 oxidation states. Furthermore, La, Gd, and Lu might be grouped due to their similarity in electron configuration, with $5d^16s^2$ available bonding electrons and their f-orbital being empty, full, or half-filled; however, other properties differ along this group such as ionic radius (decreasing), and melting point (increasing). Thus, we will discuss only some key lanthanide elements primarily from those listed above with a single oxidation state, with the hope of covering key materials having properties of interest as gate dielectrics.

9.2 Lanthanide Materials Properties

Most of the known properties of lanthanide dielectrics derive from studies on bulk crystalline materials, to be briefly summarized here, with some information related to thin films included as appropriate. Discussed here are some prerequisite dielectric properties for use as a gate dielectric; detailed issues specific to particular thin film gate stacks will be discussed later.

9.2.1 Physical/Structural Properties

As covered in reviews by Adachi and Imanaka [6] and others [7, 8], all the lanthaniods oxidize readily, typically forming +3 valence sesqiuoxide (Ln_2O_3) lanthanide compounds, with melting points ranging from 2,200 to 2,500 °C. The f-orbital electrons do not participate directly in chemical bonding, but do facilitate valence changes to +2 or +4 that occur in some elements such as Ce, Pr, or Tb. F-orbital electrons also affect other properties, such as magnetic or optical properties, not discussed here.

In the presence of water vapor, many of these oxides will form lanthanide hydroxides, with the hydroxide tendency decreasing with increased lanthanide atomic number, in general [6]. For example, La_2O_3 readily forms hydroxides [9], while for Dy_2O_3 or Ho_2O_3, hydroxide formation is minimal [7]. Carbonates will also form, but these are more limited to the outer surfaces. Hydroxide and carbonate formation will generally result in a lower density material, with lower dielectric constant (k), and higher loss and leakage. Thus in situ capping of these binary dielectrics is critical, and partly explains the desire to move towards ternary compounds such as silicates and aluminates, which are very stable with respect to atmospheric exposure [9, 10].

The binary lanthanide sesquioxides can exist in 5 crystal structure variations, as shown in the Fig. 9.1 phase-relationship diagram adapted from Adachi and Imanaka [6]. For temperatures of interest here (up to ~1,000 °C) typically either a hexagonal (termed 'A' form) or a cubic (termed 'C' form) phase are observed. Table 9.1 summarizes key materials and dielectric properties of select lanthanides. Below ~400 °C the cubic phase has been generally believed to be the stable phase for all lanthanides [6], but the hexagonal 'A' phase persists in practice for La through Nd (for example, Zhao et al. [11] and Schroeder et al. [12]). Marsela and Fiorentini [13]. have shown by structure calculations that the hexagonal phase of La_2O_3 has 0.2 eV lower energy per formula unit than the cubic bixbyite phase, and thus expected to be the equilibrium phase. It appears that the cubic bixbyite phase is only observed for La_2O_3 if impurities or appropriate nucleation sites are present

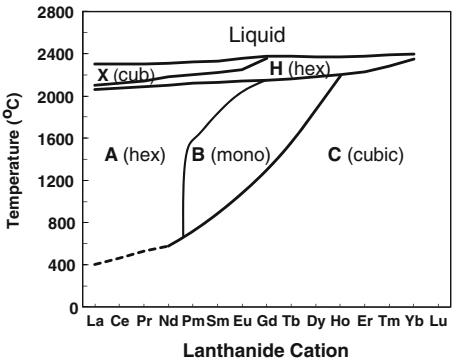


Fig. 9.1 Phase relationship diagram of binary lanthanide sesquioxides (adapted from Adachi [6])

to stabilize this structure, as occurs for some thin films formed by decomposition of molecular precursors [14]. In the hexagonal 'A' phase (La_2O_3 through Nd_2O_3), the cations have a coordination number of 7. The cubic 'C' phase is the bixbyite structure, in which cations are 6-fold coordinated in a matrix which can be described as having ordered vacant sites on the oxygen sub-lattice.

Ternary compounds are typically found in perovskite-like orthorhombic (pseudo-cubic) crystal structures (such as $LaAlO_3$) or pyrochlore phases ($La_2Hf_2O_7$). Lanthanide silicates may exist in a variety of crystalline phases, but typically in either Ln_2SiO_5 or $Ln_2Si_2O_7$ compositions [15, 16].

For use as a gate dielectric, the tendency has been to focus on amorphous dielectrics (following on the success of SiO_2 and SiO_xN_y), and to attempt to frustrate dielectric crystallization, as defects within grains or at the grain-boundaries have been linked with reliability degradation [17]. For standard (gate first) silicon-based transistor application, specific materials-related properties of interest are: (1) the amorphous phase stability temperature; and (2) the interface properties of lanthanides in contact with silicon or silicon dioxide.

From a thermodynamic viewpoint, the lanthanide oxides appear to be very stable oxides, with very large (negative) enthalpy of formation [6]. From available data, these oxides appear to be stable in contact with Si [2]. However, in the presence of excess oxygen, the interface will oxidize to SiO_2, and the mixing of silica and lanthanide oxides to form silicates is energetically favorable, as Liang et al. [16], Marsela and Fiorentini [13] discuss. The mixing of silica and lanthanide oxides to form a silicate lowers the overall free energy compared to having separate oxide layers, although reaction temperatures will depend on diffusion coefficients and other kinetic considerations. A comparison of amorphous La and Hf silicate crystallization temperatures and dielectric constants [1] are shown in Fig. 9.2. The higher temperature amorphous phase stability favors amorphous lanthanum silicate over Hf silicates. Additionally, the crystallization of HfSiOx dielectric corresponds with a decomposition reaction (phase separation) [18] resulting in low-k SiO_2 formation, while the lanthanide-silicates maintain a silicate phase without phase-separation [18]. Lanthanide silicates are also more stable with respect to atmospheric exposure [9] than the binary oxides, and the silicate reaction can be used to minimize the low-k silica interface layer thickness in the stack [19]. Thus the silicate is in many respects advantageous compared to the binary lanthanide. For a gate-last approach, with limited process temperatures and less restriction on the amorphous phase stability or interface reactivity, dielectrics with higher k than the silicates may be utilized.

9.2.2 Electrical/Dielectric Properties

In their crystalline forms, binary lanthanides have band gaps and dielectric constants making them attractive choices as gate dielectrics; select values are listed in Table 9.1 (some data is that of amorphous films, when crystalline reference data is

Table 9.1 Materials and dielectric properties of selected lanthanides. Structure and lattice parameters from JCPDS (1998), unless indicated

Material	Equil. structure	Lattice parameter (Å)	Dielectric constant (k)	Band gap (eV)	CB offset (eV)	ΔH_f° kJ/mol O_2
Ln_2O_3						
La_2O_3	Hex. 'A' type	a 3.94 c 6.13	23 [49] 27 [36]	5.5 [6]	2.3 [3]	−1,196 [6]
Nd_2O_3	Hex. 'A' type	a 3.83 c 5.99	12–14 [20]	4.7 [6]		−1,205 [6]
Gd_2O_3	Cubic 'C' type, bixbyite	10.81	12–14 [20], 24 [21]	5.4 [6]		−1,211 [6]
Dy_2O_3	Cubic 'C' type, bixbyite	10.67	12–16 [20]	4.9 [6]		−1,242 [6]
Ho_2O_3	Cubic 'C' type, bixbyite	10.61	12–13 [20]	5.3 [6]		−1,254 [6]
Er_2O_3	Cubic 'C' type, bixbyite	10.55	12–14 [20]	5.3 [6]		−1,265 [6]
Lu_2O_3	Cubic 'C' type, bixbyite	10.39	11–13 [20]	5.5 [6]		−1,247 [6]
Sc_2O_3	Cubic 'C' type, bixbyite	9.85	14 [25]	6.3 [25]		−1,271 [6]
Y_2O_3	Cubic 'C' type, bixbyite	10.60	11.4	5.7 [26]	1.6 [3]	−1,270 [6]
$LnMO_3$						
$LaAlO_3$	Rhomb., Perovskite-like	3.79 pseudo-cubic	24 [29]	5.5 [30]	1.8 [3]	
$LaScO_3$	Orthorhombic Perov-like	a,b ~ 5.84, c ~ 8.06 [52]	24–27 [52]	5.5–5.8 [52]	2.0 [54]	
$GdScO_3$	Orthorhombic Perov-like	5.49, 5.75, c 7.93	20 [27, 52]	5.5–5.8 [52]	2.0 [54]	
$LnLnO_3$						
$LaLuO_3$	Orthorhombic Perov-like	5.8, 6.0, c 8.35 [33]	32 (amor) [32]	5.2 [32]	2.1 [32]	
$Ln_2M_2O_7$						
$La_2Hf_2O_7$	Cubic, ~Pyrochlore	10.77	20 (amor) [28]	5.9 [54]	2.0 [54]	
$Ln_xSi_yO_z$						
~La_2SiO_5	monoclininc	9.33, 7.51, c 7.03	20 (amor) [49]	6.5 [76]		−1,106
~$La_2Si_2O_7$	monoclininc	13.19, 8.79, c 5.41	12 (amor) [49]	7.0 [76]		
SiO_2	~amor		3.9	9.0	3.1 [3]	−910.7

not readily available). La_2O_3 has possibly the largest dielectric constant (~23–28), with the expected dielectric constants in the ~12–15 range for most of the binary oxides [20]. However, some reports for Gd_2O_3 [21], CeO_2 [22] and Pr_2O_3 [23] describe dielectric constants in the 20–30 range or higher, presumably due to effects of crystal structure and orientation, requiring further validation. Scarel et al. [24] reports that hexagonal La_2O_3 has a k of 27, whereas the cubic phase gives a value of 17. This may be the key explaining why many different values of k are reported, especially for the lanthanides composed of La, Ce, Pr, and Nd, which can exist in either hexagonal or cubic structures, in various orientations. Indeed, the higher atomic number lanthanides of the cubic bixbyite phase (and bixbyite Sc_2O_3 and Y_2O_3) almost universally have k from 10 to 15, indicating that crystal structure may be more influential than lanthanide atomic number for k determination.

As one moves across the lanthanide binary oxide series, the band-gap has a maximum value of about 5.5–6.0 eV for La_2O_3, Gd_2O_3, and Lu_2O_3, corresponding to empty, half-filled, or a full f-orbital. Thus, due to dielectric constant and band-gap, La-based materials have been most heavily researched, although due to its hygroscopic nature some researchers have focused on other lanthanides such as Gd or Lu which are more moisture resistant, or related Sc and Y compounds having similar band gaps [25, 26] although lower k values.

Although somewhat more complex when considering film growth and composition control, ternary compounds typically offer higher temperature amorphous phase stability [1], as shown in Fig. 9.2 for silica alloying of La_2O_3 and HfO_2. This has been believed to be the key for maintaining smooth interfaces and low leakage after source/drain activation anneals. Silicates also offer dielectrics with wider band gap, but lower dielectric constant, shown in Fig. 9.2 comparing La and Hf silicate dielectric constants as the SiO_2 content is increased. As shown in Fig. 9.2, these amorphous dielectrics (processed below the crystallization temperature), show dielectric constant scaling nearly following a simple component mixing rule considering density effects, between the silica and lanthanide end points [1]. Mixing binary lanthanides with other lanthanides or group III or IV binary oxides, besides Si, typically also increases the amorphous phase stability, while producing

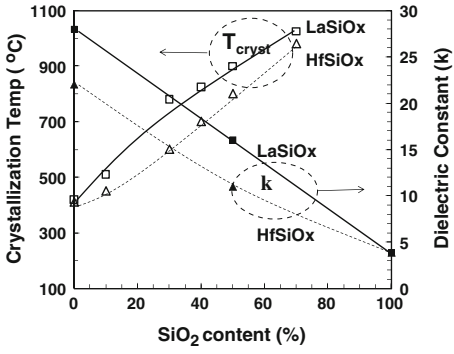


Fig. 9.2 Comparison of La and Hf silicate crystallization T and dielectric constant with silica content (adapted from Kingon et al. [1])

slightly higher dielectric constants, with little change in the band gap. Examples include lanthanide scandates [27], hafnates [28], aluminates [29–31], and mixed lanthanides such as $LaLuO_3$ [32, 33]. Additionally, for epitaxial dielectric growth, alloying can often reduce lattice mismatch with the substrate. For all these reasons, ternary compounds offer more ultimate promise than binary lanthanide oxides. The dielectric properties of La_2O_3 based ternary compounds—LaYO and LaAlO—have been briefly analyzed in Sect. 10.4.

9.3 Thin-Film Deposition and Processing

Commonly used growth techniques for lanthanide films include sputtering, thermal or e-beam evaporation (physical vapor deposition (PVD) techniques), and more recently chemical vapor deposition (CVD) techniques such as atomic layer deposition (ALD). A comprehensive listing of the growth techniques used by various researchers for lanthanide materials can be found elsewhere [34]. The gate dielectric, on silicon substrates, is deposited on either hydrogen terminated (hydrofluoric acid treated, or HF-last) oxide-free Si (001), or an 'engineered' SiO_2 or SiO_xN_y surface layer for interface defect minimization and channel mobility optimization. Thin film growth techniques typically must not damage either of these surfaces, so sputtering techniques are less preferable to other options. A key issue in general for lanthanides is the hygroscopic nature of the materials, especially La through Nd. This means that special precaution against atmospheric (water vapor and carbon dioxide) exposure is a critical issue for all deposition techniques. Additionally, magnetic properties of many of the lanthanide metals can affect the deposition process if using metallic sources. Some features specific to each deposition approach are discussed below. The basics of the deposition techniques for gate dielectrics have been described also in Chap. 4: atomic layer deposition (Sect. 4.2.2); precursors (Sect. 4.2.3); and physical vapor deposition (Sect. 4.2.4).

9.3.1 Physical Vapor Deposition

Utilizing physical vapor deposition (PVD) techniques, lanthanide-based gate dielectrics films have been reported using sputtering [11], pulsed-laser ablation deposition (PLD) [35], e-beam evaporation [36, 37], or thermal evaporation from effusion cells in molecular beam epitaxy (MBE) systems [10, 31]. Details are rarely given in the research reports, but materials must be carefully handled. Source materials should be purchased sealed in an inert gas ambient. As the lanthanide metals are very reactive with oxygen and water vapor, metal powders will quickly form oxides and hydroxides, and thus larger pieces of elemental materials should be used for e-beam or effusion cell evaporation source charges.

For e-beam evaporation of La, Al_2O_3 crucibles appear to be satisfactory [38], while we have found W crucibles effective for effusion cells. Although La melts at ~920 °C, a vapor pressure of 10^{-4} torr is not reached until 1,368 °C; practically we find that an effusion cell temperature of ~1,700 °C is necessary for La evaporation. The other lanthanides tend to evaporate more easily (typically 200–400 °C lower [38]), while oxides of any of the lanthanides typically require evaporation temperatures of 1,400–1,600 °C, with O loss occurring during heating. For effusion cell evaporation or sputtering of elemental lanthanides, reactive evaporation in oxygen ambient is sufficient to deposit oxide films. Growing the elemental lanthanide and subsequently oxidizing is an option, but rapid interface reactions can produce silicides and roughened interfaces [39]. During e-beam evaporation, it may be problematic to operate in high enough oxygen pressures, since the electron emission decreases with oxygen increases. For PLD or sputtering, elemental or oxide target materials may be used (although oxide targets require RF sputtering due to the insulating target). Again, even oxide targets convert to hydroxide very easily, and care must be taken to minimize atmospheric exposure.

9.3.2 Chemical Vapor Deposition and Atomic Layer Deposition

The advent of atomic layer deposition (ALD) has brought the general technique of chemical vapor deposition to the realm of monolayer growth control, and thus is of key importance for the growth of nm-scale gate dielectrics. This is an emerging field in relation to lanthanide oxide growth, and new or recently formulated precursor chemicals have made possible successful implementation of ALD for these compounds. Ultimately ALD is expected to be the technique of choice for virtually all high-κ dielectrics, although lack of suitable precursors has resulted in more exotic dielectric materials being studied initially by PVD techniques, with appropriate ALD precursor materials closely following.

A variety of precursors for lanthanides have been utilized by various groups [34]. Aspinall et al. [40] report that complexes of the early (lower atomic number) lanthanides with the donor-functionalized alkoxide ligand mmp (Ln (mmp)3 complexes; mmp referring to methoxy-methylpropanol) are excellent precursors for metal–organic CVD or ALD; while Paivasaari et al. [14]. utilize volatile β-diketonate type chelates Ln $(thd)_3$. Others report the use of various Ln-cyclopentadienyl complexes (Ln $(Cp)_3$) [41, 42], which result in uniform deposition at low rates, and may be advantageous due to their low melting temperatures [41].

Ozone is often used as an oxidant (instead of H_2O or O_2) to reduce C content and hydroxide formation [14, 41–43]. Complex chemistries are at play, and understanding the chemistries of the ALD process remains an area of extensive research, beyond the extent of this chapter. However, using an O_3 oxidant, these

Ln-based ALD precursors can be utilized to give high-quality low EOT (0.68 nm) device properties for La-capped HfO_2 dielectrics [41].

9.4 Lanthanide-Based Dielectric Gate Stacks

9.4.1 Lanthanum Oxides and Silicates

Extensive research has been performed on La-based oxides and silicates, as it was recognized [1] that these had the most promising values of dielectric constant, band gap, and amorphous phase stability. Key to the successful application of high-κ materials is the ability to fabricate the transistor gate with capacitance values equivalent to sub-nm SiO_2 oxide thickness. When utilizing high-κ dielectrics, gate capacitance properties are compared by referencing to the dielectric properties of ideal SiO_2 dielectric, thus determining a capacitor's 'equivalent silica oxide thickness' (EOT). Making use of the silicate reaction provides a route towards aggressive sub-nm EOT scaling [19] by thinning the SiO_2 interface layer (IL). La-based dielectrics have resulted in some of the lowest EOT gate stacks reported (0.5–0.75 nm) when processed under moderate temperatures [19, 36, 39, 41]. Discussed below are properties of La_2O_3 and lanthanum silicate ($LaSiO_x$) gate stacks deposited by reactive evaporation of La in an O_2 ambient, typically on a thin SiO_2 oxide on Si.

9.4.1.1 Gate Stack Electrical Properties

One approach to obtain low EOT devices is to deposit a thin lanthanide oxide on a silica chemical oxide, and anneal to allow interface SiO_2 consumption by the silicate phase formation. Capacitance versus gate voltage, and gate leakage data are shown in Fig. 9.3a, b for metal insulator semiconductor (MIS) capacitors with

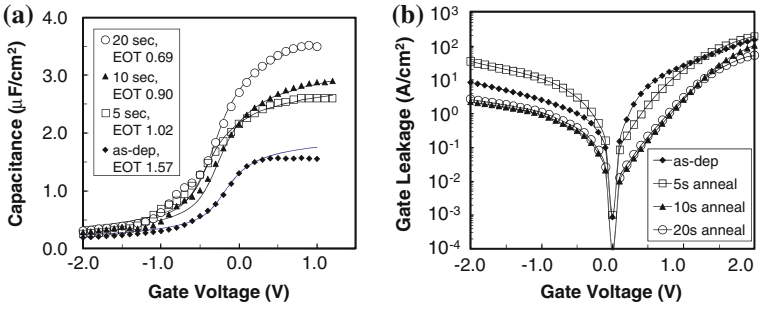


Fig. 9.3 **a** Gate capacitance and **b** leakage curves of a La_2O_3/SiO_2 bilayer stack as a function of 400 °C RTA time, in N_2

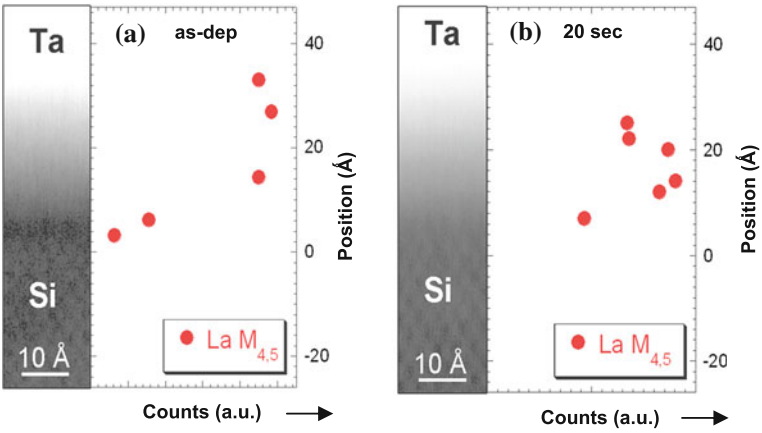


Fig. 9.4 HAADF images and corresponding EELS La $M_{4,5}$-edge intensity (increasing to the right), with position indicated as distance from the Si interface (taken as zero). The figures show La_2O_3/SiO_2 bilayer gate stacks (**a**) as-deposited, and (**b**) after 20 s RTA in N_2. SiO_{-2} interlayer is consumed by gate stack anneal

e-beam evaporated 60 nm Ta electrodes, capped with 60 nm W and reactive-ion etched (RIE) [19]. The data are shown for a bilayer dielectric of 1.6 nm La_2O_3 deposited on a ~0.8 nm SiO_2 chemical oxide on Si, measured at 1 MHz, with a 50 mV signal. As the gate stack undergoes a post-metallization anneal (PMA), the stack EOT decreases from an initial value of 1.57–0.69 nm after 20 s at 400 °C in N_2 [19]. Also, the gate leakage correspondingly decreases, as in Fig. 9.3b. As corroborated by the cross-sectional transmission electron microscope (TEM) image, the EOT lowering corresponds to dielectric thinning, caused by a combination of silicate formation, densification, and some oxygen gettering of the SiO_2 layer. Note the flatband voltage decrease as the silicate reaction proceeds, which may explain the flatband voltage shifts observed by Kuriyama et al. [44] for post-deposition anneal (PDA) of La_2O_3 on HF-last Si.

Figure 9.4a, b shows high-angle annular dark-field (HAADF) images and La electron energy loss spectra (EELS) peak intensity of the gate stack before (a) and after (b) the 400 °C 20 s PMA. The initial SiO_2 interface layer (dark in Fig. 9.4a HAADF) is seen to disappear after the anneal (Fig. 9.4b), corroborated by the La EELs signal uniformity after annealing. The lanthanum silicate bonding is determined by X-ray photoelectron spectroscopy (XPS) analysis [10] of La 3d and O1 s peaks in films with no metal capping; O gettering by the gate metal is a possible secondary EOT reduction mechanism, in conjunction with the silicate formation. The key aspect is that this silicate reaction can be a route to achieve low (sub nm) EOT, by minimizing SiO_2 IL thickness, ideal for the gate-last fabrication processes.

Under only low temperature processing [without forming gas (H_2 in N_2) anneal (FGA)], although low EOT can be obtained, high fixed charge level and high interface trap density (D_{it}) are observed as shown in Fig. 9.5. In Fig. 9.5a

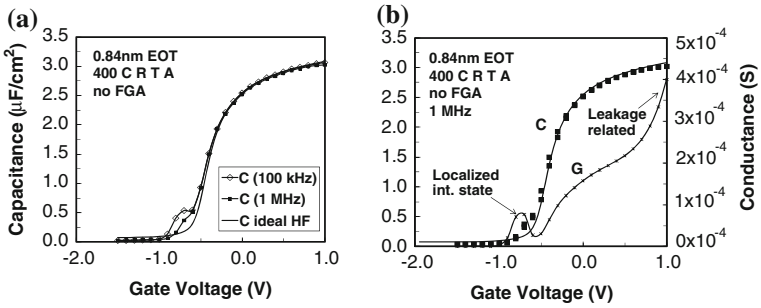


Fig. 9.5 **a** Frequency dispersion in low temperature processed $LaSiO_x$ on n-type Si. **b** C–V and G–V curves at 1 MHz showing effects of a localized interface state

frequency dispersion comparing 100 kHz, 1 MHz, and an 'ideal' HF curve are shown, revealing a sharp capacitance (C) 'hump' indicative of a localized trap state. Figure 9.5b shows the 1 MHz, 50 mV C–V and conductance versus voltage (G–V) signals, showing that the C 'hump' correlates with the conductance peak as expected. Comparing the frequency dispersion of low temperature processed La-SiO_x (1.4 nm EOT) on p- and n-type Si, we note a much greater frequency dispersion on n-type Si than on p-type Si, and determine that electron trapping is much more prevalent than hole trapping [45]. The lanthanum silicate reaction process is believed to result in this high interface defect density, under low temperature processing. This is expected to be due in part to the low temperature (forming an imperfectly bonded silicate), and partly due to the silicate reaction virtually eliminating the SiO_2 at the Si interface, and new La–O–Si interface bonding taking its place at (or very near to) the Si interface. In general, interface bonding constraints due to differing valence and O coordination number will tend to increase trap densities when introducing lanthanide oxides (or silicates) on Si or SiO_2/Si [46].

Typical gate-first processing requires that the complete gate stack undergo a high temperature anneal. Results for PMA of gate stacks which have received a prior 400 °C in situ silicate reaction anneal (to form silicate before gate metal deposition) are shown in Fig. 9.6. For capacitors with W-capped TaN electrodes, PMA anneals up to 1,000 °C for 5 s (simulating the source/drain anneal) result in a gradual increase in EOT, from 0.6 to 1.0 nm, as shown in the Fig. 9.6a C–V curves (1 MHz, 50 mV signal). In this case, starting with a silicate, further anneals grow an additional SiO_2-rich interface region [19] if excess O is available, often coming from the gate metals such as W [47]. This lowers the interface state density and the total stack charge, as revealed by the C 'hump' and V_{FB} shifting. Modeling these C–V curves and their frequency dependence [45] allows D_{it} distributions and total fixed charge densities to be extracted. After the 1,000 °C RTA, the net effective charge is reduced by 5×10^{12} cm^{-2} (V_{FB} shift), and the peak interface trap density (as shown in Fig. 9.6b) decreases from 3×10^{13} to 6×10^{11} cm^{-2} eV^{-1} without FGA. C–V data fits indicate device effective work function ($\phi_{m,eff}$) values

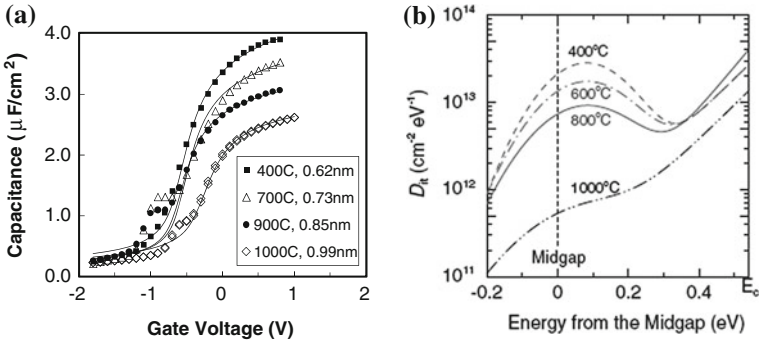


Fig. 9.6 TaN/LaSiO$_x$/Si gate stack after RTA in N_2. **a** Annealing results in some SiO_2 interface re-growth (excess O present), increasing EOT from 0.6 to 1.0 nm, but reducing interface states. **b** D_{it} extracted from C–V for each temperature

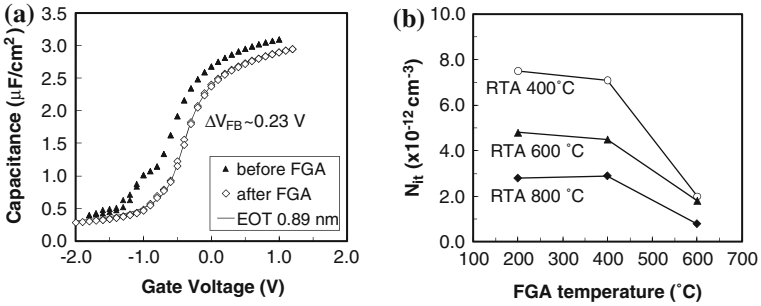


Fig. 9.7 **a** C–V before and after FGA for W/TaN/LaSiO$_x$/Si stack. **b** total N_{it} as a function of PMA RTA temperature, and subsequent FGA temperature

in the range of 3.6–4.0 eV. This makes these stacks of interest for n-channel (nMOSFET) devices; as will be discussed in Sect. 9.5.

After a high-temperature PMA of 1,000 °C for 10 s in N_2, a subsequent 450 °C 30 min FGA reduces both the total charge states reflected in a V_{FB} shift, and the interface states reflected in the lower C 'hump', as shown in Fig. 9.7a. However, typically FGA temperatures above 450 °C are required to significantly lower total interface state density (N_{it}) levels, as shown in Fig. 9.7b [45]. Because aggressive gate stack annealing typically results in some interface SiO_2 growth, D_{it} characteristics with increasing anneal temperature should approach that of SiO_2/Si [48]. So although high temperature processing can clearly result in low interface trap levels, obtaining low trap densities while maintaining sub-nm EOT remains a key process challenge, as excess O is hard to eliminate.

High temperature processing also significantly improves reliability under gate voltage stress (negative (or positive) bias temperature instability, NBTI (PBTI))

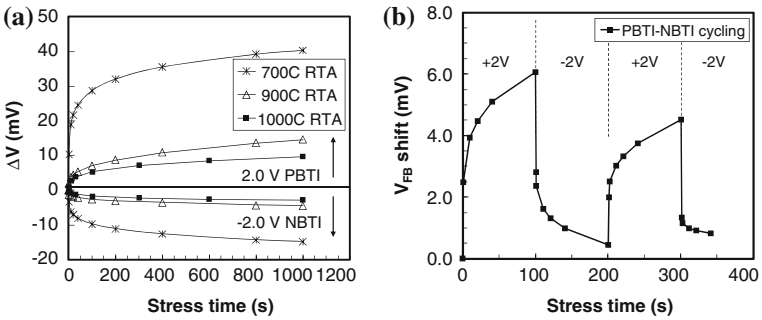


Fig. 9.8 V_{FB} shift under 2 V (PBTI) or −2 V (NBTI) gate bias, for TaN/$LaSiO_x$ MOS capacitors on n-Si with EOT ≤ 1.0 nm. **a** V shift as a function of PMA temperature, and **b** V_{FB} shift reversibility for sample with 1,000 °C PMA

stresses. Figure 9.8a, b shows the room temperature flat-band voltage (V_{FB}) shift of sub-nm EOT TaN/$LaSiO_x$/n-Si MOS devices as a function of stack RTA temperature, comparing −2 V and +2 V gate bias [49]. It is clear from Fig. 9.8a that lower T processing results in increasing amounts of V_{FB} shift, consistent with the increased defect densities as shown in Figs. 9.6, and 9.7. Also, the V_{FB} shifts are clearly recoverable as shown in Fig. 9.8b, although some net $+V_{FB}$ shift occurs revealing faster e- trapping (PBTI) than detrapping (NBTI). This reveals that PBTI stress is a more urgent problem than NBTI especially for electron-channel devices, due to large amounts of pre-existing electron traps [50].

Table 9.2 summarizes La_2O_3 and $LaSiO_x$ dielectric properties, comparing MIM to MIS devices, and various processing temperatures. While TaN MIM capacitors are useful to demonstrate potential dielectric constants, MIS devices typically will contain more Si than expected (due to additional O and subsequent silicate reaction), and thus lower k values for $LaSiO_x$. This reveals the difficult processing balance in which low temperature processing promotes aggressive EOT scaling (0.5–0.7 nm) while moderately high temperature processing is required to minimize D_{it} levels. Aggressively pursuing means to eliminate excess O in the gate electrode is critical for the gate first processes.

9.4.1.2 Gate Stack Materials Characterization

To better understand the obtained electrical properties, it is of importance to study the materials properties of these gate stacks. Critical for aggressive EOT scaling of typical gate-first processing of gate stacks are: (1) amorphous phase stability of the dielectric to source/drain activation anneal temperatures; (2) no cation diffusion into the Si channel; and (3) minimizing the low-k SiO_2 growth at the dielectric-Si interface. These have been investigated for $LaSiO_x$ gate stacks using cross-sectional TEM, back side secondary ion mass spectrometry (SIMS), and medium-energy ion-scattering (MEIS).

Table 9.2 Lanthanum-based dielectric film property summary

Film	Process temp (°C)	Thickness (nm)	EOT (nm)	k	Peak D_{it} (cm^{-2} eV^{-1})	Leakage V_{FB} +1 V (A/cm^2)
La_2O_3	400	5.6		23		**MIM** capacitor
La_2SiO_5	400	5.6		20		**MIM** capacitor
La_2SiO_5	400	4.5	1.1	16		1×10^{-4}
$LaSiO_x$	400	1.8	0.63	11	1×10^{13}	1×10^{-1}
$LaSiO_x$	1,000	~2.5	1.0	~10	4×10^{12}	3×10^{-2}
$LaSiO_x$	1,000, 600 FGA	~2.8	1.2	~9	1×10^{11}	1×10^{-2}

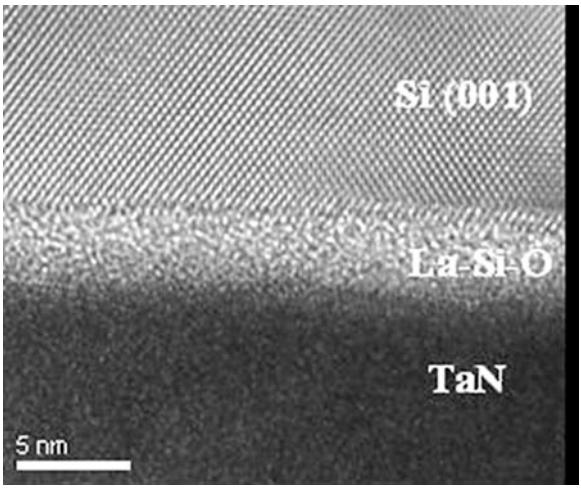


Fig. 9.9 Cross-sectional TEM of TaN/$LaSiO_x$/Si stack after RTA at 1,000 °C for 10 s in N_2

As shown in Fig. 9.9, cross-sectional TEM clearly shows that the $LaSiO_x$ dielectric remains amorphous after a 1,000 °C 10 s RTA in N_2 [51]. This is a key to keeping smoother interfaces and low gate leakage. There is no indication of phase separation with lanthanum silicate as there is with Hf or Zr silicates [18], another key advantage of lanthanide silicates in general. Using back-side SIMS analysis, it is clear that La does not diffuse into the Si [51], and thus the stack shows good thermal stability. Although the stack is chemically stable, an EOT increase is typical with annealing, as Fig. 9.6 shows. MEIS analysis of $LaSiO_x$/Si without metal capping shows clearly that there is dielectric thickening after a 1,000 °C 10 s RTA, as Fig. 9.10 reveals [49]. As these were uncapped, presumably O from the anneal chamber diffuses to the Si interface, silica forms, which subsequently reacts to form silicate. Thus annealing ultimately lowers the La content, thus lowering the dielectric constant as well. Annealing at high temperature also creates a dielectric which is stable against hydroxide formation. Figure 9.11 shows the O1s XPS spectra of uncapped $LaSiO_x$ before and after 1,000 °C 10 s RTA, from stacks corresponding to the MEIS data of Fig. 9.10. Fits to the spectra clearly show that with ~1 day of air exposure, the 400 °C formed

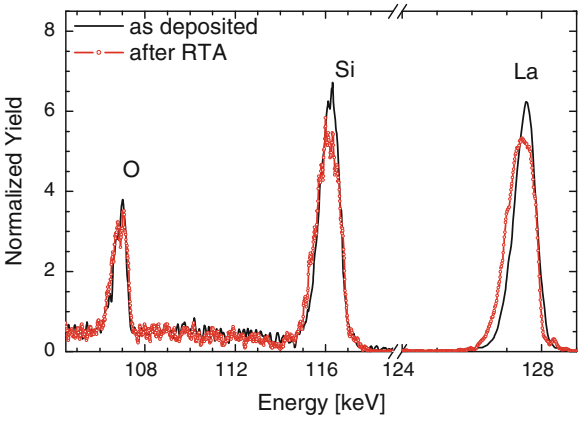


Fig. 9.10 MEIS spectra of $LaSiO_x$/Si comparing silicate without and with a 1,000 °C RTA (wider peaks, thicker $LaSiO_x$ with high-T anneal)

silicate reacts to form some hydroxide (20 % of O as hydroxide), while after the 1,000 °C RTA only about 3 % of O is bonded as hydroxide. Thus a dense silicate is very stable against hydroxide formation [10].

If excess O is properly eliminated from the gate electrode, there appear to be no fundamental reasons why these materials cannot be implemented into scaled MOSFETs, although the obtained V_{FB} (effective work function <4 eV) indicates that nMOSFET devices will be more practical, while pMOSFET devices will require stack modifications.

9.4.2 Aluminates and Scandates

The addition of a second +3 valence cation to a simple binary lanthanide oxide has important implications. Typically this will: (1) result in higher-temperature stability for an amorphous film compared to the binary oxide; (2) render the lanthanide more stable against hydroxide formation; and (3) result in crystalline structures more suitable to epitaxial growth on Si (001). Lanthanide aluminates (e.g. $LaAlO_3$) and scandates (e.g. $LaScO_3$, $GdScO_3$) have perovskite-like orthorhombic structures when crystallized. These materials have received attention especially as epitaxial oxide candidates (to be discussed in Sect. 9.6), since they have cubic lattices allowing some matching with Si, and the crystalline structures lead to large k values (of about 24–27 as reported by Heeg et al. [52], while $LaScO_3$ may have a k as high as 33 even for amorphous films [53]) and large conduction band offsets (of ~2.0 eV [54]) with respect to Si. In comparison to silicates described previously, the dielectric constants of these ternaries without Si will be greater, while the amorphous phase stability limit is typically lower.

Considering $LaAlO_3$, amorphous films have been observed to have a k of only 16 [31], lower than the crystalline value of 24. However, upon PMA at above

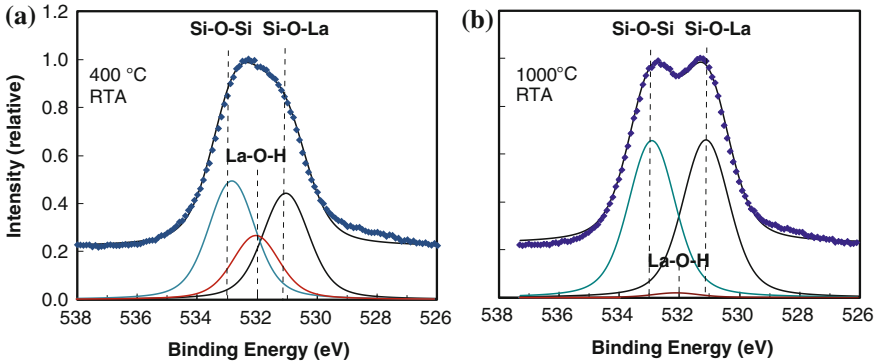


Fig. 9.11 XPS O1s peak shape of uncapped $LaSiO_x$ (**a**) with 400 °C 20 s RTA, or (**b**) after 1,000 °C 10 s RTA. High temperature silicate is impervious to hydroxide formation

900 °C simulating a source/drain activation anneal, films crystallize and partially decompose, allowing both Al and La to diffuse into the Si substrate [55]. Amorphous mixtures of La_2O_3–Al_2O_3 have an increased tendency to separate, rather than crystallize into the perovskite phase, especially in the presence of SiO_2, in which case Al_2O_3 separation occurs, and lanthanide silicate forms [56]. The amorphous phase stability can be increased to at least 1,000 °C by depositing a LaAlON film [57], in which case the phase stability rivals that of $LaSiO_x$, and the k should be higher. Cation diffusion into Si appears to correspond with the crystallization event, such that no La or Al diffusion into Si is observed for amorphous LaAlON. This is reasonable, in that this amorphous-crystalline structural change to the equilibrium phase acts as an indicator of increased diffusivity. Thus amorphous LaAlON maintains promise as an alternative dielectric, while amorphous $LaAlO_3$ would appear restricted to gate-last processing routes.

9.4.3 Hafnates and Zirconates

Key improvements in the properties of dielectrics such as HfO_2 can be achieved by alloying with lanthanides, such as increasing the amorphous phase stability [58], and lowering of defect levels [59]. Due to the maturity of the use of HfO_2 dielectric, Ln-modified Hf-based dielectrics have been implemented in scaled MOSFET devices, thus better characterized than other lanthanide binary or ternary compounds.

When alloying lanthanides with Hf or Zr oxides, the resulting compound retains a high dielectric constant, in contrast to the silicates. The +3 valence Ln cation addition to the +4 Hf or Zr results in a mixed cation coordination alloy, which helps to frustrate crystallization. If formed, the crystalline phase is a very stable pyrochlore-type phase, such as the $La_2Hf_2O_7$ composition. In addition, the Ln

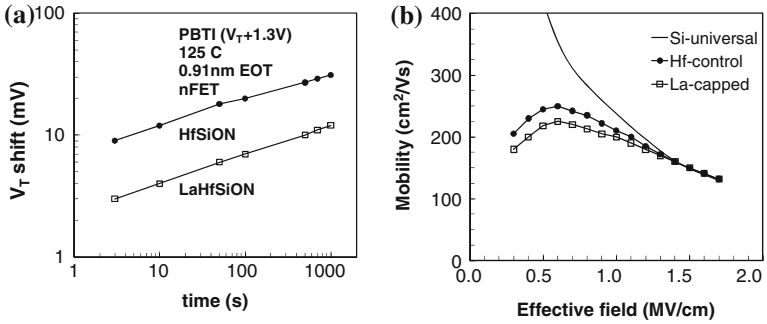


Fig. 9.12 MOSFET properties comparing effects of La addition, on **a** PBTI reliability, and **b** electron mobility (from Kirsch et al. [64], and Kang et al. [65])

present plays the important role of converting (at least some of) the SiO_2 interface layer to a higher-k Ln-silicate, to further lower EOT beyond that possible with HfO_2 alone [60].

Another key improvement obtained by alloying HfO_2 with lanthanides (expected to apply to zirconia as well) is the attainment of nMOSFET band-edge effective work function [4, 61–63]. Effects of Ln addition on V_T, a critical issue with regard to MOSFET device properties, will be discussed in Sect. 9.5.

It has been observed that HfSiON-based MOSFETs (EOT $\sim$0.9 nm) with a lanthanide addition improves the PBTI device reliability [63–65]. Using a La_2O_3 capping layer results in a $\sim 3\times$ reliability improvement under PBTI stress compared to a control $HfSiO_x$ dielectric without lanthanide, shown in Fig. 9.12a, and listed in Table 9.3. Additionally, there is little mobility degradation upon adding La, indicated in Fig. 9.12b and Table 9.3. The PBTI reliability enhancement may be linked with the ability of the Ln addition to enhance the amorphous phase stability, thus limiting defects associated with crystalline dielectrics [17]. Maximizing device mobility is another complex issue relating to the entire gate stack including the metal gate electrode, but it is clear that the Ln addition can result in very little mobility degradation.

Table 9.3 nMOSFET properties with and without La addition to $HfSiO_x$ (Kirsch et al. [64].)

	LaHfSiON	HfSiON
EOT	0.9 nm	0.9 nm
V_T	0.33 V	0.66 V
Mobility (1MV/cm)	200 $cm^2/(V{\cdot}s)$	210 $cm^2/(V{\cdot}s)$
PBTI (V_T + 1.3 V) (125 °C, 1,000 s)	12 mV	31 mV

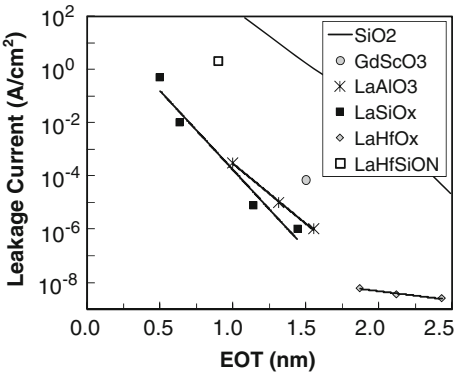


Fig. 9.13 Comparison of leakage and EOT of representative multi-component lanthanides. Data from Wagner ($GdScO_3$ [27]), Edge ($LaAlO_3$ [31]), Lichtenwalner ($LaSiO_x$ [49]), Rittersma ($LaHfO_x$ [59]), and Kirsch (LaHfSiON [64])

9.4.4 Multi-Component Dielectrics Summary

In comparison to binary lanthanides or other binary oxides such as HfO_2, dielectrics with at least one additional cation (ternaries) offer several potential advantages for use as a gate dielectric. These include: (1) a higher-temperature amorphous phase-stability; (2) a potential for limiting IL SiO_2 formation; (3) an ability to control device V_T; and (4) a means of reducing V_T shift during PBTI stress. A comparison of capacitor data of various ternary dielectrics from various groups is shown in Fig. 9.13 (while LaHfSiON represents MOSFET data). These ternary dielectrics all result in much lower-leakage gate stacks in comparison with SiO_2, while maintaining the additional benefits described above. As the search for higher k is expected to move towards epitaxial oxide growth, multi-component systems allow the ability to tune the lattice parameter for enhancing epitaxial lattice match.

The particular choice of which dielectric is the best as a gate dielectric has to consider the many aspects of device properties such as EOT, V_T, mobility, ease of processing, etc.; and the choice may differ for n versus p MOSFET, or for gate-first versus gate-last processing.

9.5 Threshold Voltage Control

Although V_T control will be covered in a separate chapter, key issues related to V_T control using Ln elements in the dielectric layer is mentioned briefly here. (Sects. 2.3 and 2.8.1—covers the theoretical basis for flat-band and threshold voltages and the complications in the case of the high-k gate stacks. The threshold voltage anomaly and the role of the interface dipoles in it have been analyzed in Chap. 6. Work function considerations and threshold voltage tuning by interface dipoles and oxygen vacancies has been outlined in Chap. 5.) Initial attempts to control V_T

were based primarily on finding appropriate electrode alloy work functions. More recently it has been understood that any region of the gate stack may be influential in controlling device V_T, especially the regions near the Si interface. It is worth noting, although not covered here, that Ln elements alloyed in the metal gate can influence the device work function [66], but here we focus on Ln (or similar) elements within the dielectric layer.

9.5.1 nMOSFET V_T Control Strategies

Early work with capacitors showed that lanthanide dielectrics resulted in very low effective work function devices (originally perceived to be problematic, as it was expected to degrade channel mobility), while for Hf-based dielectrics mid-gap work functions were prevalent. This was taken advantage of in a simple way by Alshareef et al. [4]. by applying a thin lanthana cap layer to a Hf-based dielectric to obtain nMOSFET band-edge V_T devices. It was shown that with a 0.5 or 1.0 nm La_2O_3 cap, a V_T reduction of more than 0.4 eV was obtained, with very little mobility loss (achieving 92 % of the SiO_2 universal mobility). Split C–V measurements of nMOSFETs are shown in Fig. 9.14a, revealing the V_T lowering by a La_2O_3 capping of the HfSiO dielectric. Backside SIMS analysis (Fig. 9.14b) reveals that after a spike anneal processing, the final device has the La reaching the bottom of the HfO_2 layer, although initially deposited as a cap layer [4].

The V_T lowering has been correlated to the presence of Ln cations at the SiO_2 top interface (as is corroborated by depositing lanthana directly on SiO_2 [49]), and is adequately described by an interface dipole model [67] which considers electronegativity and ionic radii of the various cation components in the dielectrics, and their influence at the SiO_2/high-κ interface. It has been shown that V_T lowering occurs with various Ln elements as well as other group II or III cations mixed with HfO_2 [62, 67].

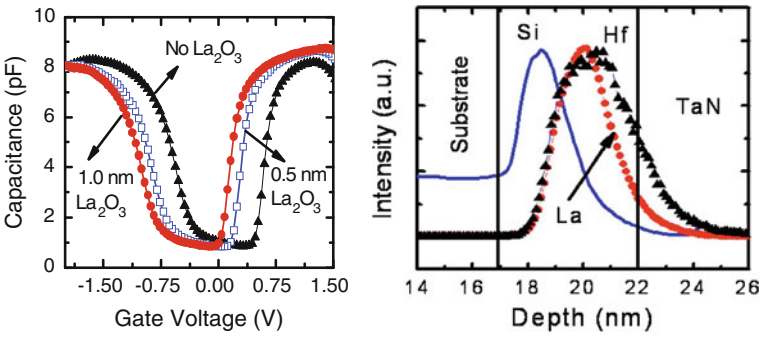


Fig. 9.14 MOSFET split CV (*left*) showing the V_T lowering with La_2O_3 capping. After spike anneal processing, the La reaches the bottom of the HfO_2 layer, shown by backside SIMS (*right*). From Alshareef et al. [4]

Other critical aspects of the V_T control are expected to relate to the silicate-like bonding at the interface, which determines bond density and directionality at the interface, and thus contributes to determining the net dipole moment beyond the issues of ionic radius and electronegativity. Without considering effects of local interface chemistry, a simple addition of a series of interfaces would not be expected to result in the significant net dipole effect as is seen in practice. Kita and Toriumi [68] present a model describing how the O density difference in the dielectrics at the SiO_2 interface may be the root origin of the interface dipole. Although there remains debate concerning specifics, there is wide consensus [62, 67, 69] that the V_T control relates to effects predominantly at the bottom high-k interface with SiO_2, and is not simply a top-electrode work function change.

9.5.2 pMOSFET V_T Control Strategies

As any high-temperature treatment typically results in the Ln element diffusing to the SiO_2 IL, typical gate first processing will produce nMOSFET V_T values. To obtain pMOSFET V_T with a Ln element present, the Ln would need to remain at the top interface, necessitating gate-last processing and additional dielectric layering schemes [70]. However, the use of elements such as Al [68] would be more suitable, and Ln elements are typically not pursued for pMOSFET applications.

9.6 Epitaxial Lanthanide High-*κ* Gate Dielectrics

As device scaling continues to push EOT to the values required for the 22 nm device node and beyond, dielectric constants in the range of 25 and higher will be required. This drives research efforts into epitaxial dielectrics compatible with the semiconductor channel material, as epitaxial dielectrics should have virtually no low-k interface layer, and crystalline oxides have higher k than their amorphous counterparts. This is another area where lanthanide-based dielectrics appear to again be advantageous, based on lattice structure and chemical compatibility. Issues related to epitaxial growth of lanthanide oxides on semiconductors is covered in detail in Chap. 11, in particular that of Gd_2O_3 on Si.

9.7 Lanthanide Dielectrics on High-Mobility Semiconductors

Scaling beyond the limits of Si devices has many researchers looking into the use of high-mobility semiconductors for the device channel, in device structures which allow higher drive current to be realized. Candidates of present interest include Ge,

GaAs and related III–V materials such as InGaAs. Whether the high mobility channel is formed from an epitaxial layer on Si, or an alternative semiconductor bulk substrate, the dielectric must be chemically compatible with the channel. Lanthanides are expected to be suitable dielectrics on GaAs based on the work of Passlack [5], and should result in acceptably high band-offsets [71] to minimize the gate leakage.

In order to realize higher device currents, the key challenge for MOSFETs on alternative semiconductors has been to achieve a low defect-level interface with the gate dielectric. On GaAs, extensive research has been performed on the use of Gd_2O_3 dielectric by Passlack [5]. They have shown the ability to obtain a low leakage dielectric and reduce interface states to acceptable levels, through an optimized process to obtain $GaGdO_x$ ($k \sim 20$) on a $\sim$1 nm Ga_2O_3 interface layer. Eliminating As oxides while precisely controlling Ga interface oxides appear to be key factors for obtaining good MOS behavior. Additionally, many groups are investigating the use of sulphur-passivation treatments [72, 73] or deposited interface layers [74] in attempts to lower D_{it} on III–V semiconductors.

As HfO_2 directly on GaAs appears to result in unacceptably high interface state densities [74], there is an impetus to look at various interface treatments, and other dielectrics, such as $LaSiO_x$ [75, 76]. Shown in Fig. 9.15a–b is an XPS comparison of 3 nm $LaSiO_x$ films on p-type GaAs (001) subjected to three different surface treatments (native oxide, HCl dip, or HCl dip and S-passivation) [75]. It can be seen that, under the restriction of a 400 °C process temperature limit, interface As-oxides are reduced after dielectric deposition, and the presence of Ga-oxide depends on the GaAs surface treatment. The S-treated interface, composed of minimal Ga-O (and/or Ga-S) bonding with no As-O, is conducive to low interface state densities [5, 72, 77]. MOS capacitor C–V characteristics as show in Fig. 9.16 indicate that the S-treatment results in the thinnest dielectric, with small hysteresis and lowest V_{FB} [75]. It has also been shown that $LaSiO_x$ on GaAs results in good values of band gap (6.5 eV) and conduction band offset of 2.7 eV [76], and thus lanthanide dielectrics appear to show promise for practical application to III–V

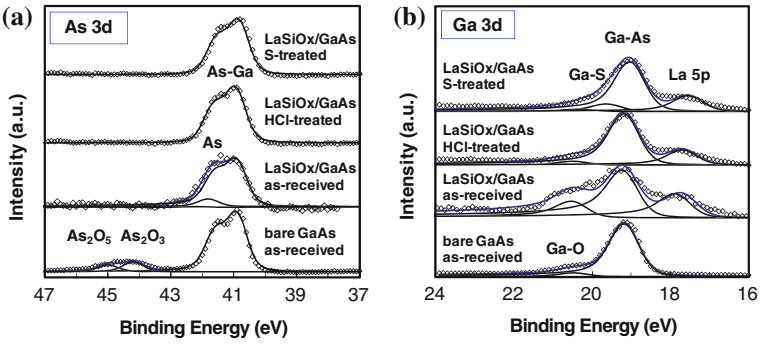


Fig. 9.15 XPS spectra of bare GaAs, and $LaSiO_x$ on GaAs comparing three surface treatments. **a** As 3d, and **b** Ga 3d spectra shown

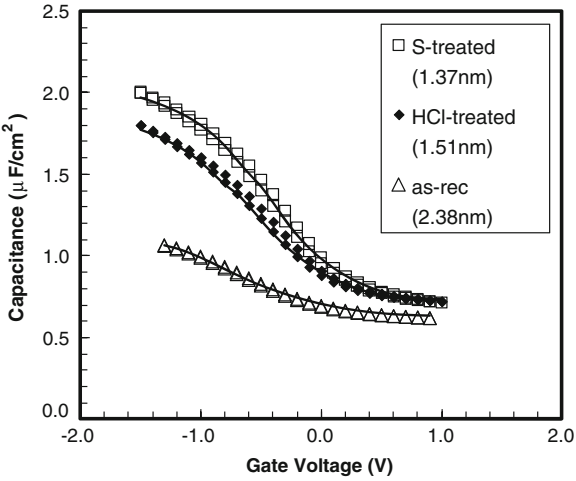


Fig. 9.16 MOS C–V characteristics of $LaSiO_x$/GaAs corresponding to the XPS spectra in Fig. 15

semiconductors. These effects of surface treatments and processing on the interface oxides is similar to that observed during ALD deposition of HfO_2 on GaAs, termed 'self-cleaning' [78].

However, as is the case for all dielectrics on III–V semiconductors, much more work needs to be done to eliminate or control defect levels near band edges, as well as deep levels, which result in channels with low carrier concentration, and low mobility. This can be done to a large extent by utilizing semiconductors such as $In_xGa_{1-x}As$, but the interface with the dielectric must be optimized simultaneously, while maintaining a high dielectric constant stack. Passivation of Ge, GaAs, and InGaAs surfaces has been covered in detail in Chap. 12.

9.8 Processing, Scaling, and Integration Issues

While lanthanide-based oxides are thermodynamically very stable oxides, as mentioned the binary oxides have tendencies towards hydroxide and carbonate formation. Consideration of the oxide integrity must be taken into account when considering wet process methods. However, ternary lanthanides or silicates do not suffer from these limitations; and gate stacks typically have some interface mixing which renders the hydroxide formation issue moot if the dielectric is extremely thin.

Relatively successful small gate length MOSFET processing of lanthanide-containing dielectrics has been demonstrated by both SEMATECH [4] and IMEC [79]. Wet etching of Ln-based oxides in weak HCl solutions gives good selectivity w/r to HfO_2 etching [79]. However, Ln-based oxide caps on HfO_2 for nMOSFET V_T tuning must be etched off in regions requiring pMOSFET V_T. When etched, it is observed that some Ln element remains in HfO_2 due to Ln intermixing [80].

Thus the ability to attain the pMOS WF will be affected. To get around this, some HfO_2 etch and re-deposition may be required, thus complicating the processing. Although processing must be optimized, there appears to be no inherent limitation to the use of Ln-based dielectrics or electrode materials. Indications are that, to obtain the lowest possible EOT gate stacks, and low effective work function for nMOSFETs, Ln materials will indeed be required in the gate stack for future technology nodes.

9.9 Summary of Lanthanide Materials and Properties

Lanthanide materials will continue to gain interest as device features shrink, requiring higher-k dielectrics with thinner interface layers, and possibly alternative channel materials. To date lanthanides have proven key components in dielectrics providing for: (1) sub-nm EOT gate stacks due to high dielectric constant and SiO_2 interface reduction, while maintaining good MOSFET mobility; (2) nMOSFET V_T control and PBTI reduction of Hf-based dielectrics; and (3) obtaining low leakage, low D_{IT} dielectrics on GaAs. Application areas which remain open for further study for lanthanide dielectrics include epitaxial oxide integration on Si or SiGe, and oxides on alternative high-mobility (III–V based) substrates for high speed applications, as well as use as a gate dielectric or passivation layer for high power FETs (for example, on SiC or GaN).

References

1. A.I. Kingon, J.-P. Maria, S.K. Streiffer, Alternative dielectrics to silicon dioxide for memory and logic devices. Nature **406**, 1032 (2000)
2. D.G. Schlom, J.H. Haeni, A thermodynamic approach to selecting alternative gate dielectrics. MRS Bull. **27**(3), 198 (2002)
3. J. Robertson, Electronic structure and band offsets of high-dielectric-constant gate oxides. MRS Bull. **27**(3), 217 (2002)
4. H.N. Alshareef, H.R. Harris, H.C. Wen, C.S. Park, C. Huffman, K. Choi, H.F. Luan, P. Majhi, B.H. Lee, R. Jammy, D.J. Lichtenwalner, J.S. Jur, A.I. Kingon, Thermally stable N-metal gate MOSFETs using La-incorporated HfSiO dielectric, IEEE Symp. on VLSI Technology (2006), pp. 7–8
5. M. Passlack, *Methodology for development of High-κ stacked gate dielectrics on III-V semiconductors, materials fundamentals of gate dielectrics*, ed. by A.A. Demkov, A. Navrotsky (Springer, 2005), pp. 403–467
6. G.-Y. Adachi, N. Imanaka, The binary rare earth oxides. Chem. Rev. **98**, 1479–1514 (1998)
7. S. Bernal, G. Blanco, J.J. Calvino, J.A. Perez Omil, J.M. Pintado, Some major aspects of the chemical behavior of rare earth oxides: An overview. J. Alloy. Compd. **408–412**, 496–502 (2006)
8. G. Scarel, A. Svane, M. Fanciulli, *Scientific and technological issues related to rare earth oxides: An Introduction*, eds. by M. Fanciulli, G. Scarel. *Rare earth oxide thin films, topics applied Physics* (Heidelberg, 2007), 106:1–14

9. T. Gougousi, G.N. Parsons, Postdeposition reactivity of sputter-deposited high-dielectric-constant films with ambient H2O and carbon-containing species. J. Appl. Phys. **95**(3), 1391 (2004)
10. D.J. Lichtenwalner, J.S. Jur, S. Novak, V. Misra, A.I. Kingon, Reliability and Stability Issues for Lanthanum Silicate as a High-K Dielectric. ECS Trans. **3**(3), 245 (2006)
11. Y. Zhao, M. Toyama, K. Kita, K. Kyuno, A. Toriumi, Moisture-absorption-induced permittivity deterioration and surface roughness enhancement of lanthanum oxide films on silicon. Appl. Phys. Lett. **88**, 072904 (2006)
12. T. Schroeder, T.-L. Lee, L. Libralesso, I. Joumard, J. Zegenhagen, P. Zaumseil, C. Wenger, G. Lupina, G. Lippert, J. Dabrowski, H.-J. Müssig, Structure and strain relaxation mechanisms of ultrathin epitaxial Pr2O3 films on Si (111). J. Appl. Phys. **97**, 074906 (2005)
13. L. Marsella, V. Fiorentini, Structure and stability of rare-earth and transition-metal oxides. Phys. Rev. B **69**, 172103 (2004)
14. J. Paivasaari, M. Putkonen, L. Niinisto, A comparative study on lanthanide oxide thin films grown by atomic layer deposition. Thin Solid Films **472**, 275–281 (2005)
15. I.A. Bondar, Rare-earth silicates. Ceram. Int. **8**(3), 83 (1982)
16. J–.J. Liang, A. Navrotsky, T. Ludwig, H.J. Seifert, F. Aldinger, Enthalpy of formation of rare-earth silicates Y2SiO5 and Yb2SiO5 and N-containing silicate Y10(SiO4) 6N2. J. Mater. Res. **14**(4), 1181–1185 (1999)
17. P.D. Kirsch, M.A. Quevedo-Lopez, S.A. Krishnan, B.H. Lee, G. Pant, M.J. Kim, R.M. Wallace, B.E. Gnade, Mobility and charge trapping comparison for crystalline and amorphous HfON and HfSiON gate dielectrics. Appl. Phys. Lett. **89**, 242909 (2006)
18. J.-P. Maria, D. Wickaksana, J. Parrette, A.I. Kingon, Crystallization in SiO2–metal oxide alloys. J. Mater. Res. **17**, 1571 (2002)
19. D.J. Lichtenwalner, J.S. Jur, A.I. Kingon, M.P. Agustin, Y. Yang, S. Stemmer, L.V. Goncharova, T. Gustafsson, E. Garfunkel, Lanthanum silicate gate dielectric stacks with sub-nanometer EOT utilizing an interfacial silica consumption reaction. J. Appl. Phys. **98**(2), 4314 (2005)
20. D Xue, K Betzler, H Hesse, Dielectric constants of binary rare-earth compounds, J. Phys.: Condens. Matter **12**, 3113–3118 (2000)
21. H.J. Osten, M. Czernohorsky, E. Bugiel, D. Kuehne, A. Fissel, Interface engineering during epitaxial growth of high-K lanthanide oxides on silicon. Mater. Res. Soc. Symp. Proc. **917**, E10–04 (2006)
22. L. Tye, N.A. El-Masry, T. Chikyow, P. McLarty, S.M. Bedair, Electrical characteristics of epitaxial CeO2 on Si (111). Appl. Phys. Lett. **65**, 3081 (1994)
23. A. Fissel, H.J. Osten, E. Bugiel, Towards understanding epitaxial growth of alternative high-K dielectrics on Si (001): Application to praseodymium oxide. J. Vac. Sci. Technol., B **21**, 1765 (2003)
24. G. Scarel, A. Debernardi, D. Tsoutsou, S. Spiga, S.C. Capelli, L. Lamagna, S.N. Volkos, M. Alia, M. Fanciulli, Vibrational and electrical properties of hexagonal La2O3 films. Appl. Phys. Lett. **91**, 102901 (2007)
25. A.M. Herrero, B.P. Gila, C.R. Abernathy, S.J. Pearton, V. Craciun, K. Siebein, F. Ren, Epitaxial growth of Sc2O3 films on GaN. Appl. Phys. Lett. **89**, 092117 (2006)
26. S.Y. Chiam, W.K. Chim, C. Pi, A.C.H. Huan, S.J. Wang, J.S. Pan, S. Turner, J. Zhang, Band alignment of yttrium oxide on various relaxed and strained semiconductor substrates. Appl. Phys. Lett. **103**, 083702 (2008)
27. M. Wagner, T. Heeg, J. Schubert, C. Zhao, O. Richard, M. Caymax, V.V. Afanas'ev, S. Mantl, Preparation and characterization of rare earth scandates as alternative gate oxide materials. Solid-State Electron. **50**, 58–62 (2006)
28. Y.F. Loo, S. Taylor, R.T. Murray, A.C. Jones, P.R. Chalker, Structural and electrical characterization of amorphous lanthanum hafnium oxide thin films. J. Appl. Phys. **99**, 103704 (2006)
29. R. Schwab, R. Sporl, P. Severloh, R. Heidinger, J. Halbritter, Temperature and orientation dependent dielectric measurements of LaAlO3-substrates, Appl. Supercond. 1/2. **61** (1997)

30. S.G. Lim, S. Kriventsov, T.N. Jackson, J.H. Haeni, D.G. Schlom, A.M. Balbashov, R. Uecker, P. Reiche, J.L. Freeouf, G. Lucovsky, Dielectric functions and optical bandgaps of high-K dielectrics for metal-oxide-semiconductor field-effect transistors by far ultraviolet spectroscopic ellipsometry. J. Appl. Phys. **91**, 4500 (2002)
31. L.F. Edge, D.G. Schlom, P. Sivasubramani, R.M. Wallace, B. Holländer, J. Schubert, Electrical characterization of amorphous lanthanum aluminate thin films grown by molecular-beam deposition on silicon. Appl. Phys. Lett. **88**, 112907 (2006)
32. J.M.J. Lopes, M. Roeckerath, T. Heeg, E. Rije, J. Schubert, S. Mantl, V.V. Afanas'ev, S. Shamuilia, A. Stesmans, Y. Jia, D.G. Schlom, Amorphous lanthanum lutetium oxide thin films as an alternative high-K gate dielectric. Appl. Phys. Lett. **89**, 222902 (2006)
33. K.L. Ovanesyan, A.G. Petrosyan, G.O. Shirinyan, C. Pedrini, L. Zhang, Single crystal growth and characterization of LaLuO3. Opt. Mater. **10**, 291–295 (1998)
34. M. Leskelä, K. Kukli, M. Ritala, Rare-earth oxide thin films for gate dielectrics in microelectronics. J. Alloys Compd. **418**, 27–34 (2006)
35. M.F. Vignolo, S. Duhalde, M. Bormioli, G. Quintana, M. Cervera, J. Tocho, Structural and electrical properties of lanthanum oxide thin films deposited by laser ablation. Appl. Surface Sci. **197–198**, 522–526 (2002)
36. Y.H. Wu, M.Y. Yang, A. Chin, Senior Member, IEEE, W.J. Chen, C.M. Kwei, Electrical characteristics of high quality La2O3 gate dielectric with equivalent oxide thickness of 5 Å, IEEE Electron Dev. Lett., **21**(7), 341 (2000)
37. M. Copel, E. Cartier, F.M. Ross, Formation of a stratified lanthanum silicate dielectric by reaction with Si (001). Appl. Phys. Lett. **78**(11), 1607 (2001)
38. E. Graper, (2009), Thin Film Evaporation Guide. http://www.tfi-telemark.com/telemrk/cat_evapguide.htm. Accessed Jan 4 2009
39. H. Watanabe, N. Ikarashi, F. Ito, La–silicate gate dielectrics fabricated by solid phase reaction between La metal and SiO2 underlayers. Appl. Phys. Lett. **83**(17), 3546–3549 (2003)
40. H.C. Aspinall, J.F. Bickley, J.M. Gaskell, A.C. Jones, G. Labat, P.R. Chalker, P.A. Williams, Precursors for MOCVD and ALD of rare earth oxides-complexes of the early lanthanides with a donor-functionalized alkoxide ligand. Inorg. Chem. **46**(15), 5852–5860 (2007)
41. S. Kamiyama, T. Miura, E. Kurosawa, M. Kitajima, M. Ootuka, T. Aoyama, Y. Nara, Band edge gate first HfSiON/metal gate n-MOSFETs using ALD-La2O3 cap layers scalable to EOT = 0.68 nm for hp 32 nm bulk devices with high performance and reliability, *IEEE Inter. Electron Devices Meeting* (IEDM). 539–542 (2007)
42. S. Schamm, P.E. Coulon, S. Miao, S.N. Volkos, L.H. Lu, L. Lamagna, C. Wiemer, D. Tsoutsou, G. Scarel, M. Fanciulli, Chemical/Structural nanocharacterization and electrical properties of ALD-grown La2O3/Si interfaces for advanced gate stacks. J. Electrochem. Soc. **156**, H1 (2009)
43. J. Niinisto, N. Petrova, M. Putkonen, L. Niinisto, K. Arstila, T. Sajavaar, Gadolinium oxide thin films by atomic layer deposition. J. Cryst. Growth **285**, 191–200 (2005)
44. A. Kuriyama, S.-I. Ohmi, K. Tsutsui, H. Iwai, Effect of post-metallization annealing on electrical characteristics of La2O3 gate thin films. Jpn. J. Appl. Phys. **44**(2), 1045–1051 (2005)
45. N. Inoue, D.J. Lichtenwalner, J.S. Jur, A.I. Kingon, Analysis of interface states in LaSixOy metal–insulator–semiconductor structures, Jpn. J. Appl. Phys. **46**(10A), 6480–6488 (2007)
46. G. Lucovsky, J.-P. Maria, J.C. Phillips, Interfacial strain-induced self-organization in semiconductor dielectric gate stacks. II. Strain-relief at internal dielectric interfaces between SiO2 and alternative gate dielectrics. J. Vac. Sci. Technol. **B22**, 2097 (2004)
47. J.S. Jur, D.J. Lichtenwalner, A.I. Kingon, Processing impact on electrical properties of lanthanum silicate thin films. Mater. Res. Soc. Sym. Proc. **917**, E10–03 (2006)
48. P.K. Hurley, K. Cherkaoui, S. McDonnell, G. Hughes, A.W. Groenland, Characterisation and passivation of interface defects in (100)-Si/SiO2/HfO2/TiN gate stacks. Microelectronics Reliab. **47**, 1195–1201 (2007)
49. D.J. Lichtenwalner, J.S. Jur, N. Inoue, A.I. Kingon, Overview of materials processing and properties of lanthanum-based high-κ dielectrics. ECS Trans. **11**(4), 319 (2007)

50. T.P. Ma, X.W. Wang, L. Song, H. Bu, Charge trapping—a major reliability challenge for High-κ gate dielectrics. ECS Trans. **1**(5), 733 (2006)
51. J.S. Jur, D.J. Lichtenwalner, A.I. Kingon, High temperature stability of lanthanum silicate dielectric on Si (001). Appl. Phys. Lett. **90**, 102908 (2007)
52. T. Heeg, J. Schubert, C. Buchal, E. Cicerrella, J.L. Freeouf, W. Tian, Y. Jia, D.G. Schlom, Growth and properties of epitaxial rare-earth scandate thin films. Appl. Phys. A **83**, 103–106 (2006)
53. J.M.J. Lopes, U. Littmark, M. Roeckerath, S.T. Lenk, J. Schubert, S. Mantl, A. Besmehn, Effects of annealing on the electrical and interfacial properties of amorphous lanthanum scandate high-κ films prepared by molecular beam deposition. J. Appl. Phys. **101**, 104109 (2007)
54. V.V. Afanas'ev, A. Stesmans, C. Zhao, M.Caymax, T. Heeg, J. Schubert, et al., Band alignment between (100) Si and complex rare earth/transition metal oxides, Appl. Phys. Lett. **85**(24) (2004)
55. P. Sivasubramani, J. Kim, B.E. Gnade, R.M. Wallace, L.F. Edge, D.G. Schlom, H.S. Craft, J.-P. Maria, Outdiffusion of La and Al from amorphous LaAlO3 in direct contact with Si (100). Appl. Phys. Lett. **86**, 201901 (2005)
56. B. Lee, D.J. Lichtenwalner, M. Agustin, R. Arghavani, X. Tang, S. Gandikota, V. Ku, V. Misra, Investigation of VT shift mechanisms of High-K dielectrics caused by lanthanum capping for NMOS and tantalum capping for PMOS devices. ECS Trans. **13**(1), 123 (2008)
57. P. Sivasubramani, J. Kim, M.J. Kim, B.E. Gnade, R.M. Wallace, Effect of nitrogen incorporation on the thermal stability of sputter deposited lanthanum aluminate dielectrics on Si (100). Appl. Phys. Lett. **89**, 152903 (2006)
58. T. Wang, J.G. Ekerdt, Atomic layer deposition of lanthanum stabilized amorphous hafnium oxide thin films. Chem. Mater. **21**, 3096–3101 (2009)
59. Z.M. Rittersma, J.C. Hooker, G. Vellianitis, J.-P. Locquet, C. Marchiori, M. Sousa, J. Fompeyrine, L. Pantisano, W. Deweerd, T. Schram, M. Rosmeulen, S. De Gendt, A. Dimoulas, Characterization of field-effect transistors with La2Hf2O7 and HfO2 gate dielectric layers deposited by molecular-beam epitaxy. J. Appl. Phys. **99**, 024508 (2006)
60. K. Choi, H. Jagannathan, C. Choi, L. Edge, T. Ando, M. Frank, P. Jamison, M. Wang, E. Cartier, S. Zafar, J. Bruley, A. Kerber, B. Linder, A. Callegari, Q. Yang, S. Brown, J. Stathis, J. Iacoponi, V. Paruchuri, V. Narayanan, Extremely scaled gate-first High-k/metal gate Stack with EOT of 0.55 nm using novel interfacial layer scavenging techniques for 22 nm technology node and beyond, IEEE Sym. on VLSI Technology (2009), 138
61. X.P. Wang, M.-F. Li, C. Ren, X.F. Yu, C. Shen, H.H. Ma, A. Chin, C.X. Zhu, J. Ning, M.B. Yu, D.-L. Kwong, Tuning effective metal gate work function by a novel gate dielectric HfLaO for nMOSFETs. IEEE Electron. Device Lett. **27**(1), 31 (2006)
62. V. Narayanan, V.K. Paruchuri, N.A. Bojarczuk, B.P. Linder, B. Doris, Y.H. Kim, S. Zafar, J. Stathis, S. Brown, J.Arnold, M. Copel, M. Steen, E. Cartier, A. Callegari, P. Jamison, J.-P. Locquet, D.L. Lacey, Y. Wang, P.E. Batson, P. Ronsheim, R. Jammy, M.P. Chudzik, M. Ieong, S. Guha, G. Shahidi, T.C. Chen, Band-Edge High-Performance High-κ/Metal Gate n-MOSFETs using cap layers containing group IIA and IIIB elements with gate-first processing for 45 nm and beyond, IEEE Symp on VLSI Technology (2006), pp. 178–179
63. H.R. Harris, H. Alshareef, H.C. Wen, S. Krishnan, K. Choi, H. Luan, D. Heh, C.S. Park, H.B. Park, M. Hussain, B.S. Ju, P.D. Kirsch, S.C. Song, P. Majhi, B.H. Lee, R. Jammy, Simplified manufacturable band edge metal gate solution for NMOS without a capping Layer, IEEE Inter. Electron Devices Meeting (IEDM) (2006), pp. 1–4
64. P.D. Kirsch, M.A. Quevedo-Lopez, S.A. Krishnan, C. Krug, H. AlShareef, C.S. Park, H.R. Harris, N. Moumen, A. Neugroschel, G. Bersuker, B.H. Lee, J.G. Wang, G. Pant, B.E. Gnade, M.J. Kim, R.M. Wallace, J.S. Jur, D.J. Lichtenwalner, A.I. Kingon, R. Jammy, Band Edge n-MOSFETs with High-k/metal gate stacks scaled to EOT = 0.9 nm with excellent carrier mobility and high temperature stability, IEEE Inter. Elec. Dev. Mtg. (IEDM) (2006), pp. 1–4
65. C.Y. Kang, C.D. Young, J. Huang, P.D. Kirsch, D. Heh, P. Sivasubramani, H.K. Park, G. Bersuker, B.H. Lee, H.S. Choi, K.T. Lee, Y.-H. Jeong, D.J. Lichtenwalner, A.I. Kingon, H.-

H. Tseng, R. Jammy, The Impact of La-doping on the reliability of low VT High-k/Metal gate nMOSFETs under various gate stress conditions, IEEE Inter. Elec. Dev. Mtg. (IEDM) (2008), pp. 1–4
66. C. Ren, D.S.H. Chan, M.-F. Li, W.-Y. Loh, S. Balakumar, C.H. Tung, N. Balasubramanian, D.-L. Kwong, Work function tuning and material characteristics of lanthanide-incorporated metal nitride gate electrodes for NMOS device applications. IEEE Trans. Electron. Dev. **53**, 1877 (2006)
67. P.D. Kirsch, P. Sivasubramani, J. Huang, C.D. Young, M.A. Quevedo-Lopez, H.C. Wen, H. Alshareef, K. Choi, C.S. Park, K. Freeman, M.M. Hussain, G. Bersuker, H.R. Harris, P. Majhi, R. Choi, P. Lysaght, B.H. Lee, H–.H. Tseng, R. Jammy, T.S. Böscke, D.J. Lichtenwalner, J.S. Jur, A.I. Kingon, Dipole model explaining high-k/metal gate field effect transistor threshold voltage tuning. Appl. Phys. Lett. **92**, 092901 (2008)
68. K. Kita, A. Toriumi, Origin of electric dipoles formed at high-k/SiO2 interface. Appl. Phys. Lett. **94**, 132902 (2009)
69. Y. Yamamoto, K. Kita, K. Kyuno, A. Toriumi, Study of La-induced flat band voltage shift in metal/HfLaOx/SiO2/Si capacitors. Jpn. J. Appl. Phys. **46**, 7251 (2007)
70. B.M. Lee, D.J. Lichtenwalner, V. Misra, Unpublished research results
71. J. Robertson, B. Falabretti, Band offsets of high κ gate oxides on high mobility semiconductors. Mater. Sci. Eng., B **135**, 267–271 (2006)
72. N.G. Wright, C.M. Johnson, A.G. O'Neill, Sulphur based surface passivation for high voltage GaAs Schottky diodes. Solid State Electron. **42**, 437 (1998)
73. M.V. Lebedev, T. Mayer, W. Jaegermann, Sulfur adsorption at GaAs (1 0 0) from solution: role of the solvent in surface chemistry. Surf. Sci. **547**, 171–183 (2003)
74. S. Oktyabrsky, V. Tokranov, M. Yakimov, R. Moore, S. Koveshnikov, W. Tsai, F. Zhu, J.C. Lee, High-k gate stack on GaAs and InGaAs using in situ passivation with amorphous silicon. Mater. Sci. Eng., B **135**, 272–276 (2006)
75. D.J. Lichtenwalner, R. Suri, V. Misra, Effect of GaAs surface treatments on lanthanum silicate high-K dielectric gate stack properties. Mat. Res. Soc. Symp. Proc. **1073**, H06–04 (2008)
76. J.-K. Yang, W.S. Kim, H–.H. Park, Chemical bonding states and energy band gap of SiO2-incorporated La2O3 films on n-GaAs (001). Thin Solid Films **494**, 311–314 (2006)
77. R. Suri, D.J. Lichtenwalner, V. Misra, Impact of elemental arsenic on electrical characteristics of metal-oxide-semiconductor capacitors on GaAs using atomic-layer deposited HfO2 gate dielectric. Appl. Phys. Lett. **92**, 243506 (2008)
78. R.M. Wallace, P.C. McIntyre, J. Kim, Y. Nishi, Atomic layer deposition of dielectrics on Ge and III–V materials for ultrahigh performance transistors. MRS Bull. **34**, 493 (2009)
79. R. Vos, S. Arnauts, I. Bovie, B. Onsia, S. Garaud, K. Xu, Y. HongYu, S. Kubicek, E. Rohr, T. Schram, A. Veloso, T. Conard, L.H.A. Leunissen, P.W. Mertens, Challenges with respect to high-κ/metal gate stack etching and cleaning. ECS Trans. **11**(4), 275 (2007)
80. T. Schram, S. Kubicek, E. Rohr, S. Brus, C. Vrancken, S.-Z. Chang1, V.S. Chang, R. Mitsuhashi, Y. Okuno, A. Akheyar, H.-J. Cho, J.C. Hooker, V. Paraschiv, R. Vos, F. Sebai, M. Ercken, P. Kelkar, A. Delabie, C. Adelmann, T. Witters, L-A. Ragnarsson, C. Kerner, T. Chiarella, M. Aoulaiche, Moon-Ju Cho, T. Kauerauf, K.De Meyer, A. Lauwers, T. Hoffmann, P.P. Absil, S. Biesemans, Novel Process to pattern selectively dual dielectric capping layers using soft-mask only, IEEE Symp on VLSI Technology (2008), pp. 44–45

Chapter 10
Ternary HfO_2 and La_2O_3 Based High-*k* Gate Dielectric Films for Advanced CMOS Applications

Akira Toriumi and Koji Kita

Abstract We first discuss HfO_2-based ternary high-*k* dielectric films. We emphasize that ternary materials do not only exhibit average properties expected by the mean-media model, but that they also reveal unexpected properties due to structural phase transformation. We also discuss La_2O_3-based ternary high-*k* films, where dopant atoms play a key role in stabilizing the hexagonal phase that is inherently the high-*k* phase of La_2O_3. Finally, ternary dielectrics for preparing amorphous gate insulators are discussed from the viewpoint of forming random network structure and suppressing long-range ordering. Understanding of ternary systems will hopefully guide us to higher-*k* dielectric materials and/or those that are highly reliable.

10.1 Introduction

HfO_2 is considered to be a realistic high-k dielectric film for advanced complementary metal-oxide-semiconductor (CMOS) applications [1]. However, several drawbacks with HfO_2 have been reported in scaling CMOS further and new candidates for advanced dielectric films are being searched for [2]. Ternary high-*k* dielectrics have a twofold purpose in the engineering of CMOS gate stacks. The first involves how they improve dielectric properties, and the second involves how they tune the flat-band voltage, V_{FB}, in MOS gate stacks. The latter is discussed in Chap. 6, and we focus on the former in this chapter. There are several key issues in

A. Toriumi (✉) · K. Kita
Department of Materials Engineering, The University of Tokyo,
7-3-1 Hongo, Tokyo 113-8656, Japan
e-mail: toriumi@material.t.u-tokyo.ac.jp

K. Kita
e-mail: kita@adam.t.u-tokyo.ac.jp

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5_10,

terms of dielectric film properties such as (1) dielectric constant, (2) band gap and band offset, (3) poly-crystalline/amorphous properties, and (4) defect generation and passivation. From the high-k engineering viewpoint (k: dielectric constant), it is critically important to find how they are directly and indirectly correlated with one another not only from the viewpoint of gate-stack engineering but also from that of materials science.

Here, we discuss ternary oxides focusing on both HfO_2-based and La_2O_3-based oxides, because both are currently the most promising high-k materials in the Si community.

10.2 Dielectric Films for CMOS Applications

10.2.1 Figure-of-Merit for High-k Dielectric Films

This section describes the basic properties of dielectric film. Generally speaking, a figure-of-merit (FOM) in using high-k dielectric film in CMOS should include both dielectric constant k, and band-offset Φ_b, against the metal and Si, because the most relevant merit of the high-k dielectric film for gate insulators is to reduce the gate leakage current. This section discusses the phenomenological relationship between k and Φ_b.

The leakage current in an ultra-thin gate insulator is intrinsically determined by quantum mechanical tunneling (cf. Chap. 2). Simple analysis with first-order approximation for the direct tunneling in Metal Insulator Metal (MIM) structures reveals that the direct tunneling current density, J_{DT}, can be expressed as

$$J_{DT} \propto \exp(-\alpha\sqrt{m^*\Phi_b} \cdot T_{OX}) \tag{10.1}$$

in the very low field region, where m^* and T_{OX} are the effective carrier mass and the physical oxide thickness, respectively. In considering the tunneling current, m^* plays a significant role in (10.1). However, since there have only been a few reports on tunneling effective mass [8], m^* has usually been assumed to be a constant.

Since the incentive for using high-k is that thicker T_{OX} can be used to keep a given capacitance, Φ_b and m^* are particularly important as well as the k value in reducing gate-leakage current. The electronic energy structures of various high-k materials have been experimentally reported and compared with those that have been theoretically calculated [3]. The band gap and offset values obtained experimentally have been slightly different in various X-ray photoemission spectroscopy (XPS) experiments [4] and internal photoemission (IPE) measurements [5]. A clear message from calculations on transition metal oxides is the relatively low conduction-band edge of d-band states, schematically depicted in Fig. 10.1 [6]. This is quite different from SiO_2 and Al_2O_3 with no d-electrons. The valence-band maximum, on the other hand, is generally determined by the non-bonding O2p band, and it is also reduced from SiO_2 because O atoms are more

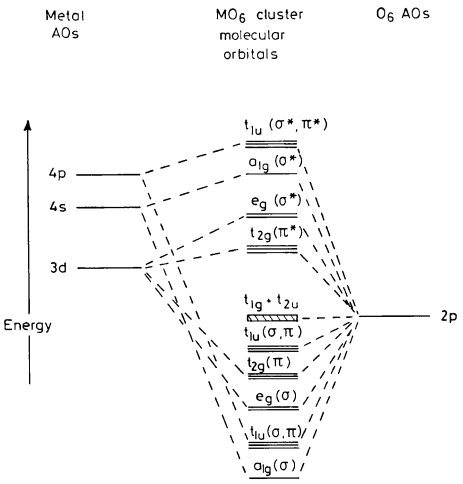


Fig. 10.1 Schematic of molecular orbital-energy level for group IVB metal in octahedral bonding arrangement with six oxygen neighbors. The top of the valence band is assigned to non-bonding 2p orbital of oxygen atom p-states, and the first two conduction bands are associated with transition metal d-states [6]

negatively charged in high-k dielectrics. Thus, the band gap for high-k films is significantly reduced to 3–6 eV, compared to $\sim$9 eV for SiO_2, and both the Φ_b (electron) and Φ_b (hole) become lower. The valence band edge of high-k films is considered to further shift upward with the introduction of nitrogen [7].

It is well known that a larger band gap causes a smaller dielectric constant at optical region k_{el} due to a smaller transition probability over the forbidden gap. However, k_{el} is not the main contributor to the dielectric constant of high-k dielectrics. Thus, a smaller k_{el}, i.e., a smaller n (the optical refractive index), is better for obtaining higher insulating barriers to suppress the tunneling current.

Further details on the electronic structure in high-k dielectrics are not discussed in this chapter but their implications are briefly discussed from the viewpoint of the FOM of high-k materials.

In (10.1) for J_{DT}, T_{ox} is the physical oxide thickness, and

$$T_{OX} = \frac{k}{k_{\mathrm{SiO_2}}} \cdot EOT \tag{10.2}$$

obtained by using the k value and equivalent oxide thickness (EOT) of high-k films, where EOT denotes the thickness that is a capacitive equivalent to the SiO_2 MIM capacitor case. Note the difference from CET which is a capacitive equivalent in MIS structure. Therefore CET includes the capacitance in semiconductor.

Thus, (10.1) can be rewritten as

$$J_{DT} \propto \exp\left(-\alpha \cdot k\sqrt{m^*\Phi_b} \cdot \frac{EOT}{k_{\mathrm{SiO_2}}}\right). \tag{10.3}$$

Therefore, $k\sqrt{m^*\Phi_b}$ is an FOM for suppressing the leakage current at a certain EOT. Figure 10.2 shows the relationship between the FOM defined above and the k-value, where k and Φ_b are the values reported in the literature. The results

indicate that a large high-k will not be appropriate for Si-CMOS applications, but a medium high-k will provide an optimum point in terms of FOM [2].

Here, three points should be taken into consideration in the analysis of actual leakage current. First, we have thus far focused on the conduction band offset (electron tunneling), but it is not always true that electron tunneling is more dominant than hole conduction. Second, attention should be focused on the interface layer on Si substrate. Thus, a sequential tunneling process should be considered to estimate the leakage current in a realistic energy-band diagram in a high-k gate stack, as shown in Fig. 10.3. This fact is also related to the experimental determination of electronic structures in Fig. 10.2. In most experiments, the relevant energy positions in high-k gate stacks have actually been estimated for high-k/SiO_2/Si stacks, so careful consideration by taking into account the formation of dipoles at the high-k/SiO_2 interface, as discussed in Chap. 6, is needed to enable a quantitative discussion. Third, Pool-Frenkel-type conduction via defect sites in the film has often been reported from the temperature and electric-field dependencies in the leakage current of high-k dielectric films. Therefore, it is not that simple to quantitatively fix the optimum point; nevertheless, the results in Fig. 10.2 suggest an intrinsic trend in estimating the lowest leakage currents in high-k gate stacks.

10.2.2 Dielectric Constant and Molecular Polarization

It is crucial to have a fundamental understanding of dielectric constant to overcome the general trend in Fig. 10.2. Here, let us briefly review the dielectric response to an electric field [9, 10].

The macroscopic polarization, $\vec{P}$ is described by a volume addition of the microscopically induced polarization, $\vec{m}$. By defining the local field, $\vec{E}_i$, which works on microscopic polarization

$$\vec{P} = N_m \vec{m} = N_m \alpha_m \vec{E}_i, \tag{10.4}$$

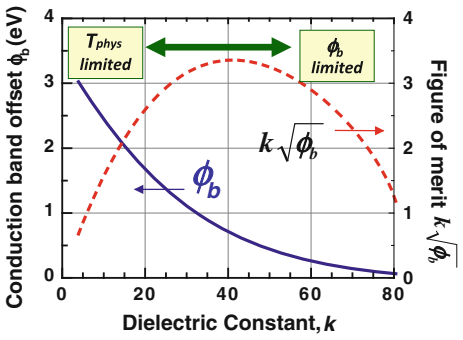


Fig. 10.2 Relationship between the FOM, $k\sqrt{\Phi_b}$, and k, estimated from the reported conduction-band-offset (Φ_b) values (eV). Very high-k will not always be advantageous for Si-CMOS applications

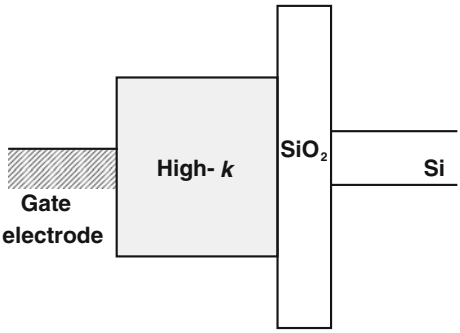


Fig. 10.3 Realistic energy-band structure of high-k MIS structure. The interfacial SiO_2 layer is inevitably formed on the Si substrate

where α_m is the microscopic polarizability and N_m is the number density of microscopic polarization.

The dielectric constant, k, is defined as follows, where $\vec{D}$ is an electric displacement vector.

$$\vec{D} = \vec{E} + 4\pi \cdot \vec{P} = k\vec{E}. \tag{10.5}$$

Thus,

$$\vec{P} = \frac{(k-1)}{4\pi} \cdot \vec{E}. \tag{10.6}$$

Here, the difference between $\vec{E}_i$ in (10.4) and the applied electric field $\vec{E}$ in (10.6) should be noted. The relationship between $\vec{E}$ and $\vec{E}_i$ is usually described by the Lorentz field as

$$\vec{E}_i = \vec{E} + \frac{4\pi\vec{P}}{3} = \frac{\vec{E}}{3}(k+2). \tag{10.7}$$

This relationship is quite important to enable what happens inside the dielectric materials to be understood, because the internal field of high-k film is quite different from the field applied between the two electrodes. Thus, from (10.4), (10.6), and (10.7)

$$\frac{(k-1)}{4\pi}\vec{E} = N_m\alpha_m\vec{E}_i = N_m\alpha_m\left(\frac{k+2}{3}\right)\vec{E}. \tag{10.8}$$

The Clausius–Mossotti ((called C–M after this)) relation is obtained from these equations, i.e.,

$$\frac{k-1}{k+2} = \frac{4\pi}{3}\frac{\alpha_m}{V_m} \tag{10.9}$$

The C–M relation is in practice quite useful for connecting macroscopic dielectric constants with microscopic polarization. Namely, the C–M relation

associates microscopically defined polarizability α_m with the dielectric constant, k, defined macroscopically. Here, it should be remembered that dielectric constant k is not linearly dependent on molar polarizability, but that it is determined by both α_m and V_m (molar volume, $1/V_m = N_m$), as shown in Fig. 10.4. In the larger α_m/V_m region, k is very sensitive to a small modulation of α_m/V_m.

The C-M relation can be applied to most dielectric materials. In fact, molar polarizability α_m is not independent of molar volume V_m. Since α_m also generally increases with the increase in V_m, α_m/V_m does not change very much for various materials. Therefore, decreasing V_m while retaining α_m is another way of overcoming the previously mentioned trend.

10.2.3 Origin of Molecular Polarization

Dielectric properties can be understood from classical physics on the basis of quantum mechanical description of polarization. Figure 10.5 is a schematic of three kinds of origins for the dielectric properties of materials. These are typically permanent, ionic, and electronic polarizations. In the high-k materials for CMOS, no permanent polarization is assumed, and both (1) ionic polarization and (2) electronic polarization are considered.

First, let us discuss the ionic contribution to microscopic polarization. For simplicity, let us assume a two-atom system. From the equation of motion for each atom with effective charge Z^*e and mass M^+ or M^- under a certain electric field, ionic polarizability is obtained as

$$\alpha_{ion} = \frac{(Z^*e)^2}{c} = \frac{(Z^*e)^2}{M^*\omega_0^2}, \tag{10.10}$$

where c, M^*, and ω_0 correspond to the force constant, reduced mass, and characteristic phonon frequency. When M^+ is much larger than M^-, M^* approaches M^-

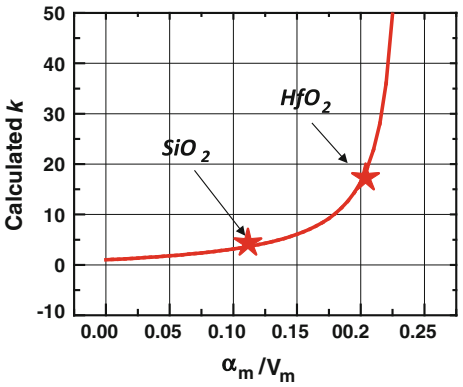


Fig. 10.4 Relationship between k and α_m/V_m in C-M relation. Both SiO_2 and HfO_2 are plotted on the line

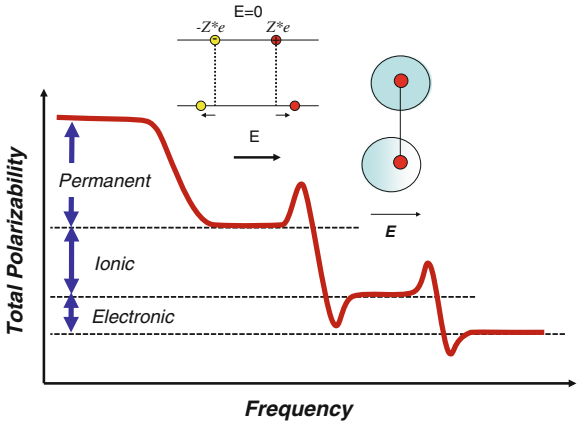


Fig. 10.5 Dependence of three kinds of polarizability on frequency

(an atom of smaller mass is generally oxygen). And, low phonon frequency modes in a material are beneficial for increasing α_{ion}.

Thus, α_{ion} is determined by Z^*e and ω_0 in the simplest case. This view is actually too simple but very useful for understanding and controlling the dielectric constant of high-k materials.

Infrared absorption studies are quite useful for studying ω_o, particularly for investigating the softness or hardness of the bonding characteristics of relevant atoms. Figure 10.6 shows typical infrared absorption spectra in monoclinic HfO_2 [11]. We can understand that the low-energy phonons of HfO_2 in the far-infrared region really contribute to high-k of HfO_2, as expected from (10.10).

Next, let us consider the electronic contribution to microscopic polarizability. Electron polarization is quite fast and corresponds to the optical response. Electron displacement around the atom is involved with the quantum mechanical transition probability of the wave function with no electric field to the state with the electric field. This can be calculated by using the second-order perturbation theory of quantum mechanics. First-order perturbation does not provide electronic polarizability, α_e, because polarization is linearly dependent on the distance. For insulators or semiconductors, α_e is inversely proportional to the band gap to excite electrons from the valence to the conduction bands. This means that a larger band-

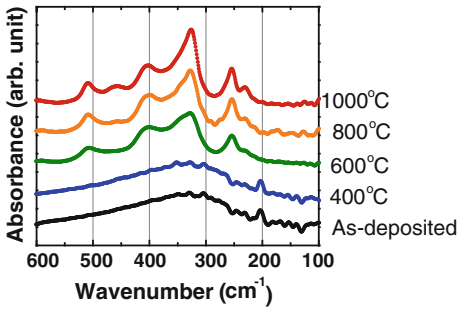


Fig. 10.6 Far-infrared absorption spectra of HfO_2 at different annealing temperatures.

gap implies a smaller α_e. In terms of the applications of high-k films to electron devices, a higher band-gap material is actually required to suppress the leakage current, which is inconsistent with a larger α_e. The ionic contribution in high-k materials is more dominant than the electronic contribution. Thus, a larger α_e would be better, but from the viewpoint of high-k dielectric insulators for CMOS, a larger α_{ion} and a smaller α_e are better in many cases. As $k_{el}=n^2$ is well known in the Lorentz-Lorenz relation in optics, the electronic contribution to dielectric constant in a high frequency region can be obtained by the optical refractive index, n, of a material. Therefore, a larger refractive index is not always appropriate for CMOS applications.

A straightforward relationship between ionic polarization and the energy-band structure may not be expected in high-k materials. Both covalent and ionic bonding are involved in high-k materials as well. Simply speaking, the bonds in ionic materials are longer than those in covalent ones. With increased bond length, the energy separation of bonding–anti-bonding decreases, and then the band-gap energy decreases [12]. This view is obviously too simple since it does not include the ionic contribution to the energy-band gap. Specific material properties should definitely be included to enable a quantitative discussion. However, the overall trend indicates that the above view can be applied to the actual representation of high-k materials from the engineering viewpoint.

10.3 HfO_2-Based Ternary Oxides

10.3.1 Dielectric Constant and Structural Phase of Crystalline HfO_2

A number of high-k materials besides pure HfO_2 have recently been reported. Other transition metal oxides or rare earth metal oxides are typical examples. A ternary system is another approach to reaching a new high-k material. The ternary system is expected to provide a new function and/or alleviate the disadvantages of the binary ones. This subsection focuses on HfO_2-based ternary oxides.

ZrO_2 has been intensively studied in the ceramics community from various points of view [13], even though the structural properties of ZrO_2 have been focused on in preference to the dielectric ones. Although HfO_2 is very similar to ZrO_2, in Si microelectronics, HfO_2 has thus far been the most promising candidate for CMOS applications in terms of thermal stability. It is well known that HfO_2 has several crystalline structures (polymorphs), as seen in Fig. 10.7. Since the low temperature phase of HfO_2 is monoclinic, pure HfO_2 films generally have monoclinic patterns in X-ray diffraction (XRD) in the Ultra Large Scale Integration (ULSI) process. This characteristic can also be observed in the IR absorption spectrum, as seen in Fig. 10.6.

Both phonon mode structures and dielectric constants in both ZrO_2 and HfO_2, including anisotropy have been reported by using the first-principles calculations of dielectric properties in three polymorphs [14, 15]. These results have demonstrated that the lowest k was obtained for the monoclinic phase, while the highest one was for the tetragonal phase ($k \sim 90$!). This calculation was further improved quantitatively [16], but the trend is the same as that previously reported [14, 15]. It is interesting to note that the dielectric properties greatly depend on the structural phase, because the dielectric properties can be tuned by modulating the structural phase. The issue is how to change the structural phase of HfO_2 in low-temperature processes for Si-LSI fabrication.

10.3.2 Control of Structural Phase in HfO_2

A typical example of ZrO_2-based ternary oxides is yttria-stabilized zirconia (YSZ), which has a cubic or tetragonal structure. Since we are interested in cubic or tetragonal rather than monoclinic HfO_2 from the viewpoint of dielectric constant, Y-doped HfO_2 may be an interesting material. The dielectric constant of a ternary oxide, to the first-order approximation, might be the average of those of the two binary oxides. Consequently, it is interesting to see the competition between structural phase transition and average constant.

Y-doped HfO_2 (YDH) was grown by co-sputtering HfO_2 and Y_2O_3 targets in Ar, followed by thermal annealing in 0.1 % O_2 for 30 s at various temperatures. Au electrodes were evaporated for the top gate and Al electrodes were evaporated for the bottom contact. The dielectric constant was determined using both MIS and MIM structures. Figure 10.8 plots the dielectric constant of YDH as a function of the Y atomic concentration in HfO_2 [17], where there are several characteristic features to note. The first most-interesting feature is the increase in the k value with the increase in Y concentration. The second is that the k value has a peak as a function of Y concentration. Finally, the pure HfO_2 k value decreases below the amorphous value with crystallization in the monoclinic phase. Figure 10.9 has XRD patterns for pure HfO_2 and YDH, where a large structural difference is observed between them. By annealing at 600 °C, the pure HfO_2 is crystallized to the monoclinic phase, while YDH tends to be "cubic" as expected from the phase diagram [18]. This clarifies that a structural transformation from a monoclinic to a

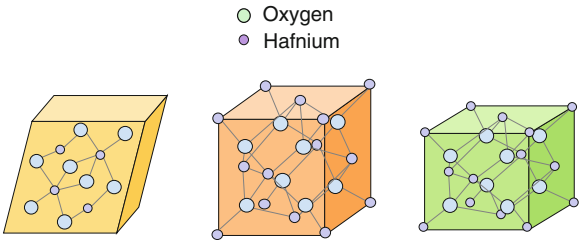


Fig. 10.7 Schematic of atomic arrangement of three polymorphs of HfO_2

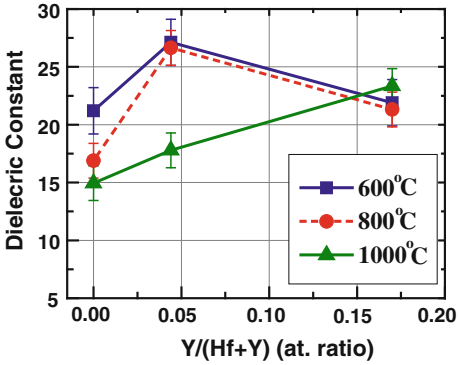


Fig. 10.8 Change in dielectric constant by Y-doping into HfO_2 films annealed at 600–1,000 °C evaluated with MIS capacitors. For the films annealed at 600 and 800 °C, the dielectric constant first increases with Y-doping ($k \sim 27$) and then gradually decreases with further Y-doping. The k value for the films annealed at 1,000 °C does not have a peak but increases monotonically

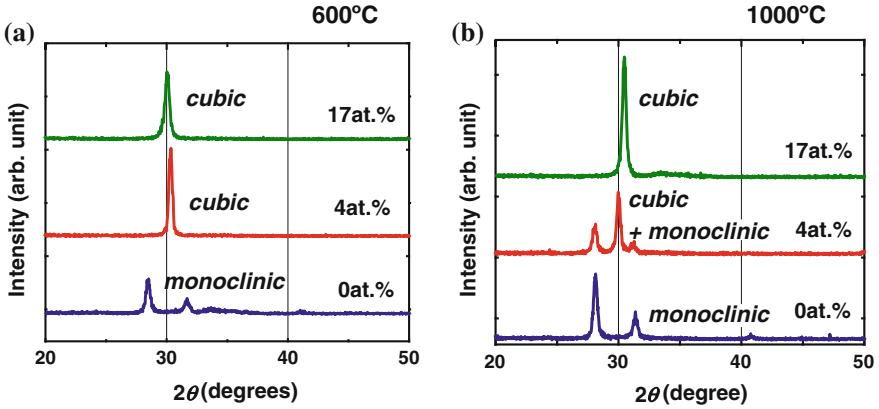


Fig. 10.9 XRD patterns for Y-doped and undoped 30 nm-thick HfO_2 films annealed at (**a**) 600 °C and (**b**) 1,000 °C in 0.1 % $O_2 + N_2$ ambient. At temperatures as low as 600 °C, the films exhibit a clear phase transformation from the monoclinic to the cubic phase by Y-doping. The 17 at.%-doped film has a peak at a slightly lower angle than that of 4 at.%-doped film, since the increase in Y concentration induces the lattice to gradually expand. At 1,000 °C, however, they demonstrate phase separation attained by 4 at.% Y-doping [19]

cubic phase substantially induces a change in dielectric constant. The increase in the Y ratio of HfO_2 eventually brings about a decrease in dielectric constant in the sense of effective media [17,19]. That is, at a lower concentration of Y_2O_3 in HfO_2, the structural phase transition of HfO_2 determines the dielectric constant, while at higher concentrations the approximation of effective media dominates the dielectric properties because the k value of Y_2O_3 is around 12.

Another interesting example is SiO_2-doped HfO_2. The Hf–Si–O system is usually regarded as Hf-silicate, but this is not the case. By introducing a small

amount of SiO_2 into HfO_2, structural phase transformation to tetragonal HfO_2 has also been observed. Dielectric constant in this case should also be affected by the phase transformation of HfO_2. Figure 10.10 surprisingly indicates that dielectric constant increases up to 27 with a slight introduction of SiO_2 into HfO_2 [20]. Although there is a great difference between Y_2O_3 and SiO_2 from both their unit structure and bonding characteristics, the doping effect on dielectric constant seems to be the same with small amounts of Y_2O_3 or SiO_2 doping. That is, the dielectric constant of their ternary systems is modified by "doping" according to the structural properties inherent to HfO_2.

First-principles calculations favoring experimental results have recently been reported for YDH and Si-doped hafnia (SDH) from the viewpoint of the doping-induced stabilization of higher temperature phases [21]. Here, the C-M relation has been used to explain the experimental results. The dielectric constant in the C-M relation is a function of α_m/V_m. Therefore, the question is which is the main contributor to the k enhancement of doped ternary HfO_2. Let us think about the YDH system. Figure 10.11 shows both V_m and α_m as a function of Y content, where V_m was measured by XRD and α_m was calculated from both k and V_m using the C-M relation. We can clearly see that k enhancement comes from reduced V_m rather than enhanced α_m.

The energy-band gap and band offset against Si were investigated by using spectroscopic ellipsometry (SE) [22] and XPS [23] to investigate the electronic structure of YDH. The XPS results indicated no substantial changes in the electronic structure irrespective of the significant increase in the dielectric constant of YDH, as seen in Fig. 10.12. This may be due to the fact that the conduction band and valence band edges are basically determined by the isolated d-band for the

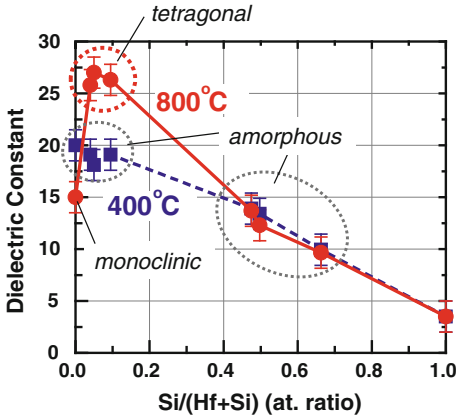


Fig. 10.10 Dielectric constant of SiO_2-doped HfO_2 films annealed at 400 and 800 °C in 0.1 % $O_2 + N_2$ ambient. With 5–10 at. %, SiO_2-doped HfO_2 films annealed at 800 °C have a phase transformation from the monoclinic to the tetragonal phase, and a drastic increase in dielectric constant, while the films annealed at 400 °C remain amorphous and do not reveal an increase in dielectric constant [20]

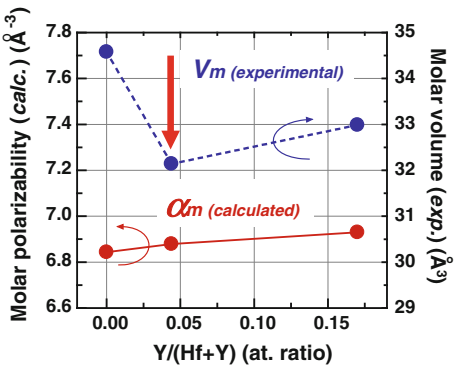


Fig. 10.11 Change in molar volume (V_m) and molar polarizability (α_m) by Y-doping for films annealed at 800 °C. The V_m values were determined experimentally from XRD measurements, and the α_m values were estimated by substituting the experimental k and V_m values into the C-M relation [19]

former and by the non-bonding O2p band for the latter. The unchanged electronic structure was also confirmed by the following experiments. Very thin SiO_2-doped HfO_2 was prepared to study the mechanism responsible for leakage current at an EOT of ~0.5 nm. Figure 10.13a plots the J_g–V_g characteristics in MIM structures with various HfO_2 films that have the same physical thickness [24]. The leakage current is definitely determined by the physical thickness, while the SDH benefit is obviously seen as a function of EOT in Fig. 10.13b. This result clearly demonstrates no change in the band offset of HfO_2 by doping with SiO_2.

Another way to change the k value is to modulate ionic polarizability by decreasing the force constant of the ionic system. Since the force constant is $M^*\omega_0^2$, lower vibration-frequency materials would be better. Figure 10.14 shows the IR absorption spectra of TiO_2, CeO_2, and HfO_2 [19]. The typical absorption

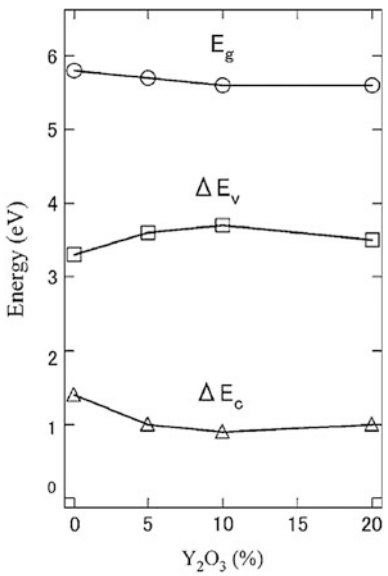


Fig. 10.12 Changes in band-gap (E_g), valence-band offset (ΔE_v), and conduction-band offset (ΔE_c) as a function of Y_2O_3 composition, determined by XPS measurements [23]

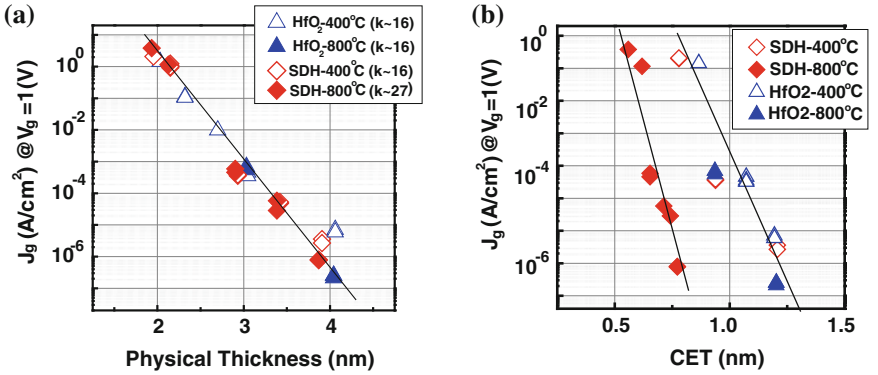


Fig. 10.13 Leakage current of HfO_2 and SiO_2-doped HfO_2 films as functions of (**a**) physical thickness and (**b**) CET. SDH annealed at 800 °C only exhibits higher-k benefit in J_g-CET relationship in (**b**) [24]

peaks of these oxides are located in the far-IR region as expected. Figure 10.15a plots dielectric constant as a function of Ti concentration in (Hf–Ti)O_2 annealed at 600 and 800 °C [19]. We can clearly see that the k value greatly depends on the Ti concentration in both cases. The dielectric constant simply increases with an increase in Ti content up to around 40, for the 800 °C-annealed sample with 60 at. % Ti. However, it should be noted that there is a crucial drawback, i.e., the reduced energy-band gap in this system, as shown in Fig. 10.15b. This should severely degrade the leakage current. On the other hand, the dependence of the k value on frequency was observed in the Hf–Ce–O system due to significant dielectric loss, even though high dielectric constant was obtained in the low-frequency region. This suggests that a smaller force constant results in structural instability, even though the static k value is relatively high. Of course, it is not certain whether optimizing the process would improve the present poor qualities of these systems or not. Further research obviously needs to be done.

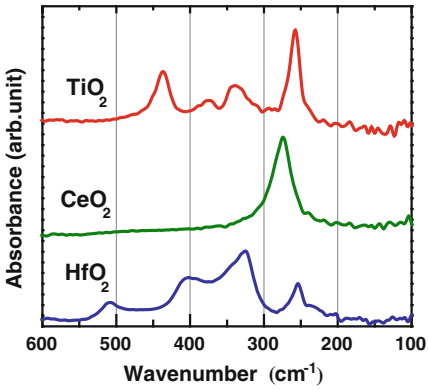


Fig. 10.14 Far-infrared absorption spectra for 20-nm-thick HfO_2, CeO_2, and TiO_2 films, annealed at 800 °C. They provide direct evidence of low-frequency lattice vibration in these films [19]

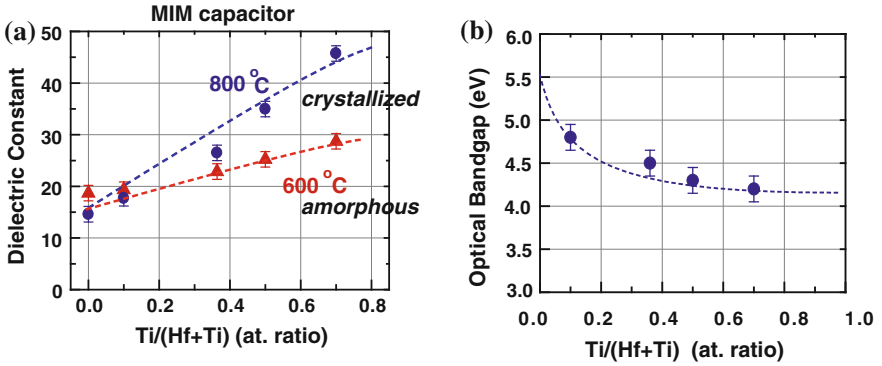


Fig. 10.15 **a** Dielectric constant of $HfTiO_x$ films annealed at 600 and 800 °C. It monotonically increases up to $k \sim 30$ at 600 °C, and $k \sim 45$ at 800 °C. The films with Ti/(Hf + Ti) = 0.36, 0.5, and 0.68, are crystallized in the orthorhombic $HfTiO_4$ phase at 800 °C, but become amorphous at 600 °C. The introduction of Ti into HfO_2 films enhances molar polarizability and increases the dielectric constant. **b** Optical band gap of $HfTiO_x$ films annealed at 600 °C monotonically decreases with Ti introduction

Doping of other elements such as La, Dy, Sc, Er, and Gd into HO_2 can also alter dielectric constant [25, 26]. The mechanism seems to be the same as in the Y doping case.

ZrO_2 is similar to HfO_2, and the $HfZrO_x$ system is also interesting to observe to find what happens by mixing the same kinds of materials. Figure 10.16 shows the monotonic decrease in capacitive-equivalent thickness (CET) due to the addition of ZrO_2 to HfO_2 associated with the formation of tetragonal-phase HfO_2. This indicates that the dielectric constant increases with the addition of ZrO_2 [27].

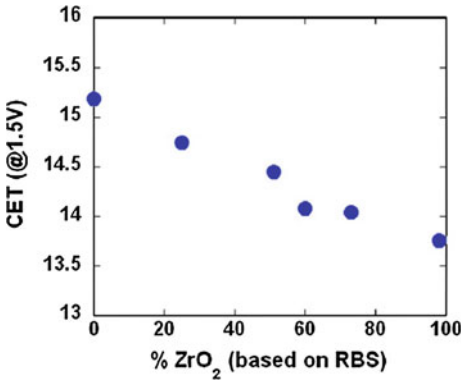


Fig. 10.16 CET monotonically decreases with addition of ZrO_2 to HfO_2, which increases the dielectric constant [27]

10.4 La_2O_3-Based Ternary Oxides

La_2O_3 has been anticipated as a promising candidate beyond HfO_2 [28]. However, only a few papers on high-k La_2O_3 have actually been published. The k values obtained experimentally have thus far not been particularly high, partly because La_2O_3 is hygroscopic and its dielectric properties are easily degraded in air. The SiO_2 capping on La_2O_3 effectively protected La_2O_3 from moisture, resulting in dielectric constant of roughly 24 [29]. Another approach has been to stabilize La_2O_3 by introducing other materials.

(a) $(La–Y)_2O_3$

Yttrium (Y) has similar chemical properties as La in the general sense. Therefore, we can expect that introducing Y_2O_3 into La_2O_3 would cause little change in its dielectric properties. The dielectric constant and resistance against moisture of $(La\text{-}Y)_2O_3$ films annealed at 600 °C with different Y concentrations were investigated. Figure 10.17 shows that the dielectric constant of 40-at%Y–$(La\text{-}Y)_2O_3$ and 70-at%Y–$(La\text{-}Y)_2O_3$ films are higher than 25. The high dielectric constant originates from the well-crystallized hexagonal phase of $(La\text{-}Y)_2O_3$ films, since hexagonal-phase rare-earth oxides have higher dielectric constant than cubic-phase ones due to the smaller molar volume of the hexagonal phase. Furthermore, the high dielectric constant 40-at %Y–$(La\text{-}Y)_2O_3$ and 70-at %Y–$(La\text{-}Y)_2O_3$ are extremely resistant to moisture. This indicates that due to the introduction of Y_2O_3, $(La\text{-}Y)_2O_3$ films have much larger lattice energy than La_2O_3 film [30]. In addition to that, the energy-band gap slightly increases with the introduction of Y_2O_3, because of the higher energy-band gap of Y_2O_3 [31].

$(La\text{-}Y)_2O_3$ has also been reported to be a lattice-matched insulator on Si (111). This may enable us to achieve Si/insulator/Si hetero-structures for advanced applications [32].

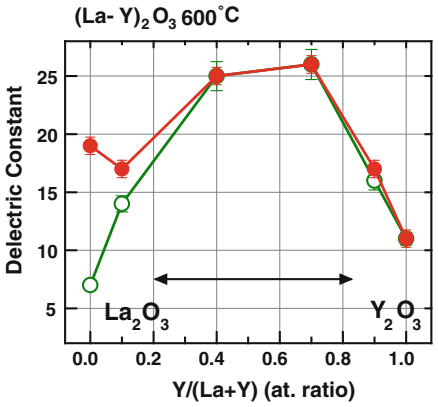


Fig. 10.17 Effect of absorption of moisture on dielectric constant of $LaYO_x$ films with different Y concentrations. The films were annealed at 600 °C and the dielectric constant was measured with MIM capacitors [30]. The closed and open circles represent two kinds of samples exposed to air for less than 1 min and 24 h, respectively

(b) $LaAlO_3$

$LaAlO_3$ can easily be obtained in single crystalline form for epitaxial crystallization of high-Tc superconductors, ferroelectric films, and/or oxide electronics materials. Only a few applications for gate dielectric film, however, have been reported [33]. According to the literature, the method of pulsed laser deposition is promising for preparing well crystallized $LaAlO_3$ films. Other film-deposition techniques such as physical vapor deposition (PVD), atomic layer deposition (ALD), or chemical vapor deposition (CVD) have thus far not been likely to yield well crystallized and stoichiometric $LaAlO_3$ films. Figure 10.18 demonstrates the minimum EOT thus far achieved by PLD on Si [34].

Other ternary rare-earth oxides can also be found [35]. Lanthanide based high-*k* oxides have been discussed also in Chap. 9.

10.5 Amorphous Ternary High-*k* Dielectrics

Amorphous dielectric films are more appreciated in the industry because SiO_2 is very robust and amorphous with no grain boundaries that might become leakage paths. HfO_2 is easily crystallized by 600 °C PDA (Post Deposition Annealing) as has been discussed. Therefore, Hf-silicate and Hf-aluminate have been considered to be promising candidates to increase the crystallization temperature of HfO_2 from the initial phase of high-*k* research.

The stability of glass has been investigated for a long time [36, 37]. In the silicon microelectronics community, SiO_2 has generally been considered to be a typical amorphous material. However, it should be noted that the amorphous nature of high-*k* films is different from SiO_2 in terms of its network structure. Thermally grown SiO_2 has a continuous-random-network (CRN), while HfO_2 has

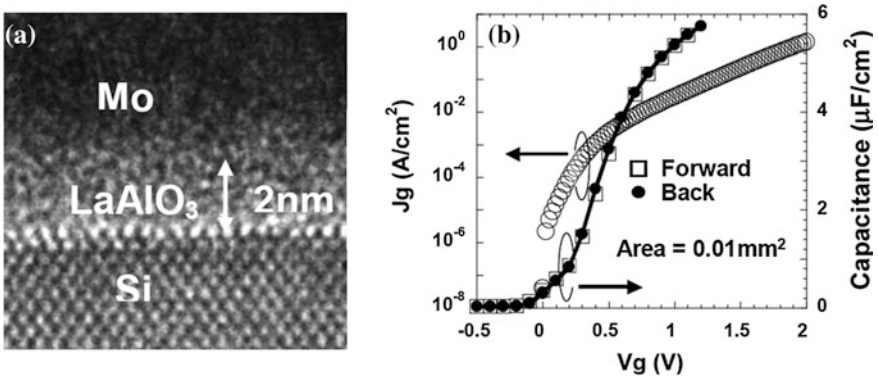


Fig. 10.18 **a** TEM image of 2-nm-thick Mo/$LaAlO_3$/Si stack. **b** C–V_g (capacitance–gate voltage) and J_g–V_g (gate leakage current density–gate voltage) characteristics. EOT was estimated to be 0.31 nm [34]

a random-close-packed (RCP) structure. The latter is basically determined by oxygen-atom packing. By using molecular-dynamics calculations, the difference between the two types of amorphous structures has been pointed out from both the viewpoints of the molar volume and the coordination number versus. metal–oxygen (M–O) distance [38], as shown in Fig. 10.19. We can clearly see that there are two types of amorphous materials. One is classified into the amorphous SiO_2-like group, while the other is classified into high-*k* materials. This change causes a striking difference in structural stability, and seems similar to the stability criterion that was discussed from the viewpoint of the balance in the average coordination number between the freedom of space and bonding constraints for atom motion [39].

(a) $(Hf–Si)O_x$ and $(Hf–Al)O_x$

The most well researched ternary systems are $(Hf–Si)O_x$ and $(Hf–Al)O_x$. Each of these is an HfO_2 film compounded with a very robust amorphous dielectric, SiO_2 or Al_2O_3. We should first refer to the phase diagram of the relevant systems. Figure 10.20 shows (a) HfO_2–SiO_2 [40] and (b) HfO_2–Al_2O_3 [41] phase diagrams, from which one can predict trends in the stability of materials.

A very strong advantage of Hf-silicate is that its ultimate limit (the lowest Hf concentration limit) is SiO_2, and its affinity with the Si substrate seems to be naturally good. The crystallization temperature increases with the introduction of Si into HfO_2, but a large amount of Si in $(Hf–Si)O_x$ is needed to maintain the amorphous state at around 1,000 °C, while SiO_2 decreases the dielectric constant significantly and reduces the advantage of high-*k* film substantially.

Since many sections of the book have dealt with Hf-silicate CMOS, we will not discuss these in detail, but simply comment on this system from the viewpoint of phase separation. It is necessary to consider what effects crystallization and phase separation have on gate dielectric properties. There seems to be no quantitative evidence for crystallization-enhanced leakage current in ultra-thin films. As high-*k* dielectrics are not thick films, we have to compare intrinsic leakage current with crystallization-induced leakage in the ultra-thin region. We think that the crucial effect is not crystallization, but phase separation in this material. Both HfO_2 and SiO_2 have been identified in samples annealed at high temperatures with Fourier

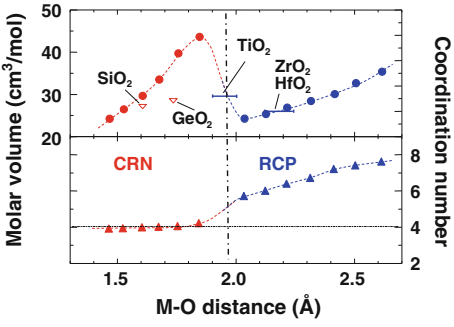


Fig. 10.19 Molar volume and coordination number of amorphous metal oxides, calculated with classical molecular dynamics as a function of metal–oxygen (M–O) distance. A shorter M–O distance creates a SiO_2-like CRN, while a longer M–O distance results in a RCP structure [38]

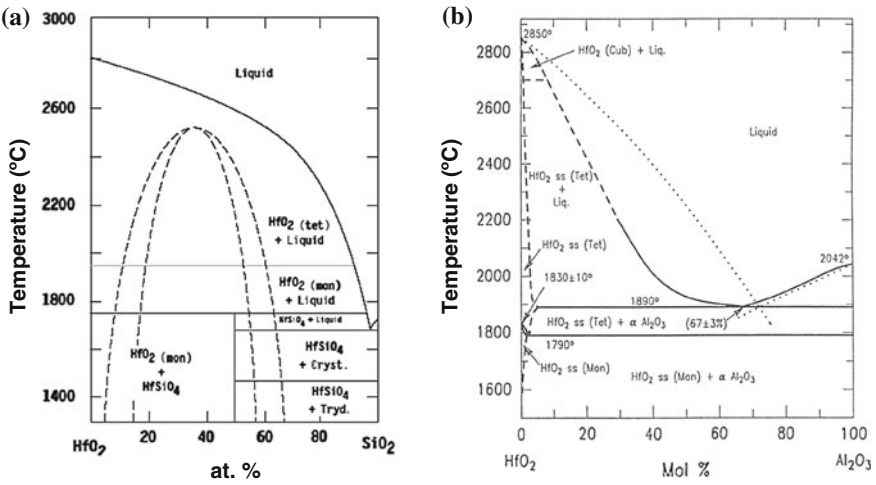


Fig. 10.20 **a** Phase diagram for HfO_2–SiO_2 compositional system, with liquid–liquid immiscibility zone (*dotted line*) [40]. **b** Phase diagram for HfO_2–Al_2O_3 compositional system [41]

transform infrared (FT-IR) spectroscopy, as shown in Fig. 10.21 [11]. Phase separation may result in spatial variations in the dielectric properties of gate dielectric film. Thus, there is a concern that Hf-silicate with high temperature PDA may cause fluctuations in the dielectric properties of short-channel devices, which will cause huge problems in ULSI systems in terms of statistical *V*th fluctuations.

(Hf–Al)O_x is another example of a high-temperature system of crystallization. To determine what the crystallization process was in this system, periodically

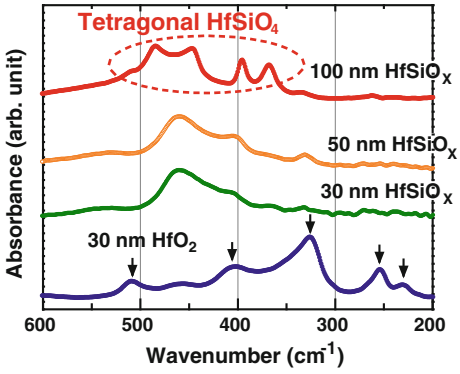


Fig. 10.21 Far-IR absorption spectra of $HfSiO_x$ (Hf:Si ~ 6:4) films with various thicknesses and HfO_2 film annealed at 1,000 °C in O_2. The 30- and 50-nm-thick $HfSiO_x$ films have a broad peak at around 450 cm^{-1} in addition to the relatively weak peaks from monoclinic-HfO_2 (*the arrows* denotes characteristic peaks in monoclinic HfO_2), while the 100 nm-thick film has a different pattern because of the formation of the tetragonal $HfSiO_4$ phase

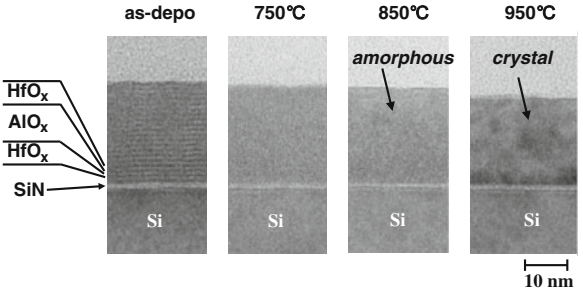


Fig. 10.22 Cross-sectional TEM images of as-deposited multi-stack HfO_x/AlO_x (3 Å/9 Å) film, annealed at 750, 850 and 950 °C

grown multi-stack films, as shown in Fig. 10.22, were prepared. Relatively thick films ($\sim$24 nm) consisting of periodically stacked $HfOx/AlO_x$ layers with various ratios of Hf to Al were grown by ALD. A thin SiN layer was first grown on Si (100) to restrict the Si diffusion into the high-k film. The PDA was carried out at various temperatures in N_2. An as-deposited film has the long-range periodicity like that of a crystalline artificial superlattice. Figure 10.23 shows the XRD patterns of $HfOx/AlO_x$ (3Å/9Å) stacked films as a function of annealing temperature, in which a single peak at around $2\theta = 7.6°$ in the as-deposited film is observed [42]. This value corresponds to d = 11.59Å, which is in good agreement with one set of HfO_x/AlO_x thicknesses (about 12Å). The superlattice structure can barely be observed after annealing at 750 °C, and the stacked film structure changes into one that is homogeneously amorphous. The crystalline structure was systematically analyzed as a function of the Hf/Al ratio and the annealing temperature. A phase diagram of the Hf–Al–O system was then constructed as a function of Al content, as shown in Fig. 10.24 [43]. Both monoclinic and cubic phases were observed for smaller Al content, while only the cubic phase was observed with increased Al content.

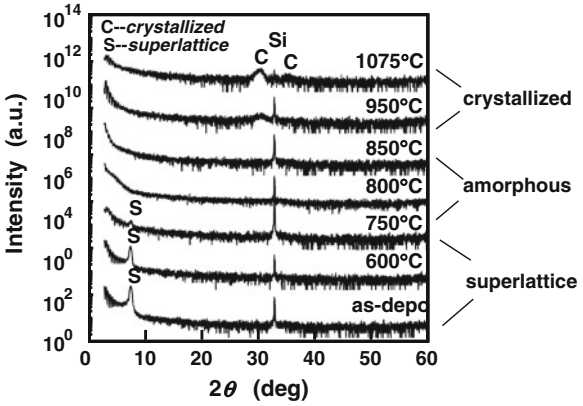


Fig. 10.23 XRD patterns of HfO_x/AlO_x (3 Å/9 Å) stacked films after annealing at different temperatures in N_2 for 30 s, in which the single peak at around $2\theta = 7.6°$ in as-deposited film corresponds to d = 11.59 Å [42]

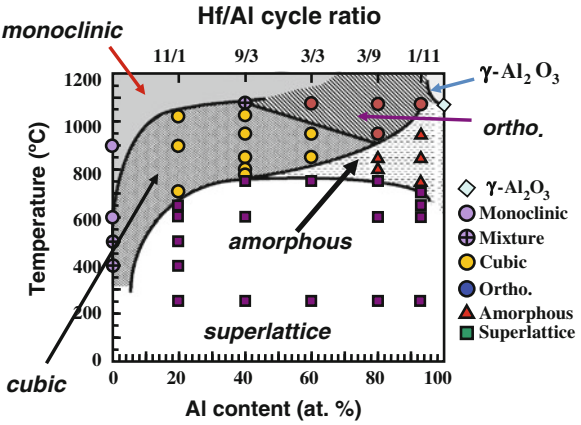


Fig. 10.24 Phase diagram of $(Hf–Al)O_x$ system with temperature and Al content as parameters [43]

(b) $(Hf–La)O_x$

SiO_2 and Al_2O_3 were incorporated into HfO_2 to suppress crystallization, but their dielectric constant values decreased. Thus, it is obvious that further EOT scaling requires new dielectrics with both high dielectric constant and high crystallization temperature.

It has recently been reported that the introduction of La_2O_3 to nanophase HfO_2 powders can raise the crystallization temperature ($\sim$900 °C) [44]. Furthermore, the incorporation of La_2O_3 is not expected to degrade the dielectric constant, because La_2O_3 is a more ionic oxide and has higher dielectric constant than SiO_2 or Al_2O_3 [45]. Both HfO_2 and La_2O_3 films start to crystallize under 600 °C, while with increased La concentration in $(Hf–La)O_x$, the film has a higher crystallization temperature, as seen in Fig. 10.25 [45]. The XRD peak of crystallized $(Hf-La)O_x$ films around $2\theta = 29°$ corresponds to the (222) planes of the pyrochlore $La_2Hf_2O_7$ structure [47]. Considering that the ionic radius of Hf^{4+} and La^{3+} are 0.71Å and 1.03 Å (for six-fold coordination), respectively, a significantly large difference in the ionic radius may substantially suppress long-range ordering [44].

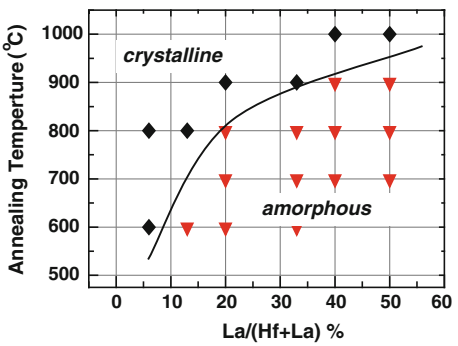


Fig. 10.25 Crystallization temperature of $(Hf–La)O_x$ films with various La compositions [46]

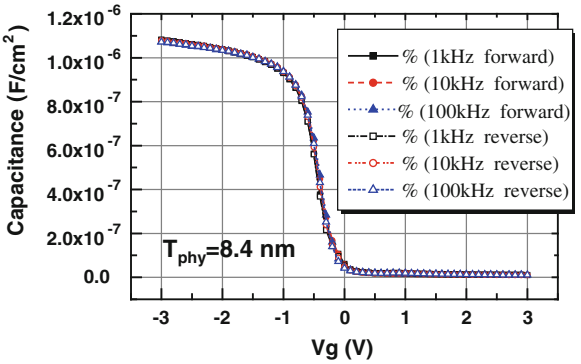


Fig. 10.26 Bi-directional C–V characteristics of Au/40-at%La-(Hf–La)O_x/p-Si MIS capacitor annealed at 600 °C. The film is 8.4-nm thick. Neither hysteresis nor frequency dispersion is observed [46]

Figure 10.26 plots the bi-directional C–V characteristics at various frequencies for Au/40-at%La-(Hf-La)O_x/p-Si MIS capacitors. The C–V curves indicate neither hysteresis nor frequency-dispersion. From good linear relationships with very small dispersion in the CET versus physical thickness, the dielectric constant of (Hf-La)O_x, which is roughly above 20, does not change greatly as a function of La concentration in (Hf–La)O_x, as shown in Fig. 10.27, which is also advantageous in terms of its very small dielectric fluctuations.

It has also been reported that $La_2Hf_2O_7$ is a promising material for epitaxial dielectric film due to its match of lattice parameters to the Si (001) substrate [47]. These results suggest that the (Hf–La)O_x system will become a potential candidate for advanced CMOS applications. First-principles calculations have recently suggested fewer oxygen vacancies inside (Hf–La)O_x film [48].

(c) (La–Ta)O_x, and (La–Lu)O_x

(La–Ta)O_x not only has high-k($k \sim 30$) but also amorphous characteristics in addition to a much larger band gap than Ta_2O_5 films, obtained by adjusting the Ta concentration in the films. Figure 10.28 shows XRD patterns of 30-nm-thick 35-at%Ta-(La–Ta)O_x films annealed at 800, 900, and 1,000 °C. The XRD patterns of La_2O_3 and Ta_2O_5 films annealed at 800 °C indicate clear crystallization peaks, but

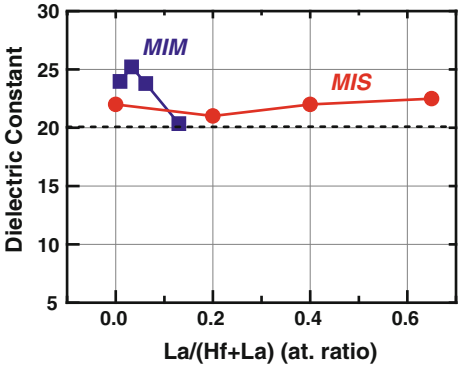


Fig. 10.27 Dielectric constant as function of La concentration in (Hf–La)O_x films, evaluated in MIM capacitors (Au/(Hf-La)O_x/Pt) and MIS capacitors (Au/(Hf-La)O_x/Si). The films were annealed at 600 °C. With > 20-at.% La, the films are amorphous but $k \sim 22$ is maintained.

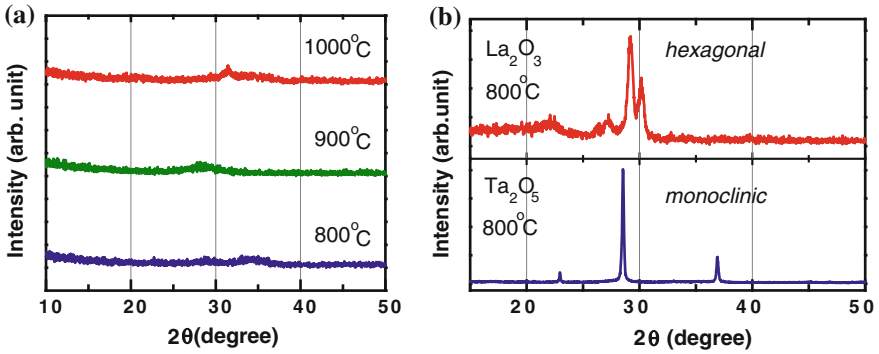


Fig. 10.28 **a** XRD patterns of 30-nm-thick 35-at%Ta-(La-Ta)O_y films annealed at 800, 900, and 1,000 °C. No crystallization is observed. **b** XRD patterns of La_2O_3 and Ta_2O_5 films annealed at 800 °C [49]

35-at%Ta-(La–Ta)O_x does not exhibit any diffraction patterns [49]. The high crystallization temperature of this film is also attributed to the large difference of the ionic radius between La^{3+} and Ta^{5+}, the same as in the (Hf-La)O_x case. The crystallization temperature is sensitive to the Ta concentration in the film, while up to $x = 0.35$, the energy band gap is not very sensitive to the Ta concentration.

(La–Lu)O_x is also a candidate for amorphous high-*k* dielectric film. La and Lu are at both ends of the lanthanide metals. La^{3+} has the largest ionic radius, while Lu^{3+} has the smallest due to the lanthanide contraction. While both La_2O_3 and Lu_2O_3 are easily crystallized and their dielectric constants are about 20, $LaLuO_3$ is reported to be amorphous up to 1,000 °C and its dielectric constant is about 30 [50]. For the present, $LaLuO_3$ is not a hot material but its potential is interesting.

Finally, it may be worth mentioning again that there is a difference between (Hf–Si)O_x and (Hf-La)O_x from the viewpoint of forming amorphous film. Amorphous (Hf–Si)O_x is formed by changing the close packing to the network forming structure, where in the case of (Hf-La)O_x long-range ordering is suppressed by introducing different sized atoms. This fact should be taken into account in the discussion of amorphous high-*k* materials.

10.6 Summary

We started this chapter with the general theory of dielectric material properties, and then discussed HfO_2-based ternary high-*k* dielectric films. We need to emphasize that ternary materials do not only exhibit average properties expected by the mean-media model, but that they also reveal unexpected properties due to structural phase transformation. This will hopefully guide us to higher-*k* dielectric materials and/or those that are highly reliable. We also discussed La_2O_3-based ternary high-*k* films, where dopant atoms play a key role in stabilizing the

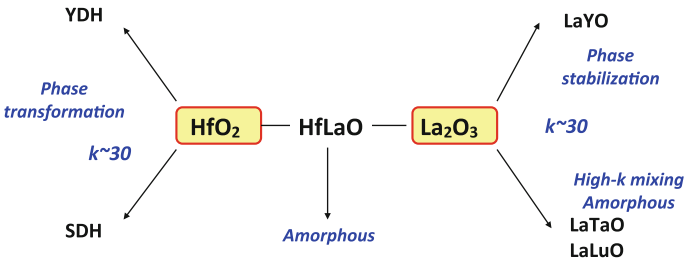


Fig. 10.29 Schematic for changing dielectric properties of high-k HfO_2- and La_2O_3-based ternary dielectric films by introducing other oxides to the host materials

hexagonal phase that is inherently the high-k phase of La_2O_3. Finally, ternary dielectrics for preparing amorphous gate insulators were discussed from the viewpoint of forming random network structure and suppressing long-range ordering. The defect levels are also modulated due to the electron-negativity difference of cations, though details have not been discussed in this chapter.

Along with these results and discussion, we illustrate a schematic in Fig. 10.29 for engineering high-k film properties by doping other oxides into HfO_2 or La_2O_3. This diagram provides guidelines for designing new ternary or more complex materials for advanced higher-k CMOS.

Acknowledgments This work was partly supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

1. M. T. Bohr, R. S. Chau, T. Ghani, K. Mistry, Spectrum **10**, 29–35 (2007)
2. A. Toriumi, K. Kita, K. Tomida, Y. Zhao, J. Widiez, T. Nabatame, H. Ota, M. Hirose, IEDM Tech. Dig. 53–56 (2007)
3. P. W. Peacock, J. Robertson, J. Appl. Phys. **92**, 4712 (2002)
4. S. Miyazaki, J. Vac. Sci. Technol., B **19**, 2212 (2001)
5. V. V. Afanas'ev, A. Stesmans, W. Tsai, Appl. Phys. Lett. **82**, 245 (2003)
6. P. A. Cox, *Transition Metal Oxides* (Oxford Press, Oxford, 1992)
7. G. Shang, P. W. Peacook, J. Robertson, Appl. Phys. Lett. **84**, 106 (2004)
8. W. J. Zhu, T.-P. Ma, T. Tamagawa, J. Kim, Y. Di, IEEE Elec. Dev. Lett. **23**, 97 (2002)
9. C. Kittel, *Introduction to Solid State Physics*, 6th edn. (Wiley, London, 1986)
10. H. Fröhlich, *Theory of Dielectrics*, 2nd edn. (Oxford University Press, Oxford, 1958)
11. K. Tomida, H. Shimizu, K. Kita, K. Kyuno, A. Toriumi, Mat. Res. Symp. Proc. **8111**, D10.9.1 (2004)
12. J. Robertson, P. W. Peacock, Electronic structure and band offsets of high electric constant gate, in *High-k Gate Dielectrics*, ed. by M. Houssa (IOP, London, 2004), pp. 372
13. M. W. Pitcher, S. V. Ushakov, A. Navrotsky, B. F. Woodfield, G. Li, J. Boero-Goates, B. M. Tissue, J. Am. Cer. Soc. **88**, 160 (2005)
14. X. Zhao, D. Vanderbilt, Phys. Rev. B. **65**, 233106 (2002)
15. X. Zhao, D. Vanderbilt, Phys. Rev. B. **65**, 075105 (2002)
16. G.-M. Rignanese, X. Gonze, G. Jun, K. Cho, A. Pasquarello, Phys. Rev. **B69**, 184301 (2004)

17. K. Kita, K. Kyuno, A. Toriumi, Appl. Phys. Lett. **86**, 102906 (2005)
18. D. W. Stacy, D. R. Wilder, J. Am. Ceram. Soc. **58**, 285 (1975)
19. A. Toriumi, Y. Yamamoto, Y. Zhao, K. Tomida, K. Kita, ECS Trans. **1**, 185 (2006)
20. K. Tomida, K. Kita, K. Kyuno, A. Toriumi, Appl. Phys. Lett. **89**, 142902 (2006)
21. D. Fisher, A. Kersch, Appl. Phys. Lett. **92**, 012908 (2008)
22. J. Widiez, K. Kita, K. Tomida, T. Nishimura, A. Toriumi, Jpn. J. Appl. Phys. **47**, 2410 (2008)
23. M. Komatsu, R. Yasuhara, H. Takahashi, S. Toyoda, H. Kumigashira, M. Oshima, D. Kukuruznyak, T. Chikyow, Appl. Phys. Lett. **89**, 172107 (2006)
24. K. Tomida, K. Kita, A. Toriumi, Jpn. J. Appl. Phys. **50**, 11502 (2011)
25. C. Adelmann, V. Sriramkumar, S. Van Elshocht, P. Lehnen, T. Conard, S. De Gendt, Appl. Phys. Lett. **91**, 162902 (2007)
26. S. Govindarajan, T. S. Boscke, P. Sivasubramani, P. D. Kirsch, B. H. Lee, U. Schroder, S. Ramanathan, B. E. Gnade, Appl. Phys. Lett. **91**, 062906 (2007)
27. D. H. Triyoso, R. I. Hegde, J. K. Schaeffer, R. Gregory, X.-D. Wang, M. Canonico, D. Roan, E. A. Hebert, K. Kim, J. Jiang, R. Rai, V. Kaushik, S. B. Samavedam, J. Vac. Sci. Technol. B **25**, 845 (2007)
28. A. Chin, Y. H. Wu, S. B. Chen, C. C. Liao, W. J. Chen, Dig. VLSI Symp. **16**, (2000)
29. Y. Zhao, K. Kita, K. Kyuno, A. Toriumi, Appl. Phys. Lett. **88**, 072904 (2006)
30. Y. Zhao, K. Kita, K. Kyuno, A. Toriumi, Appl. Phys. Lett. **89**, 252905 (2006)
31. Y. Zhao, K. Kita, K. Kyuno, A. Toriumi, Appl. Phys. Lett. **94**, 042901 (2009)
32. S. Guha, N. A. Bojarczuk, V. Narayanan, Appl. Phys. Lett. **80**, 766 (2002)
33. X.-B. Lu, Z-G Liu, Y-P Wang, Y. Yang, X-P Wang, H.-W Zhou, B-Y Nguyen. J. Appl. Phys. **94**, 1229 (2003)
34. M. Suzuki, T. Yamaguchi, M. Koyama, Tech. Dig. IEDM **2005**, 445–448 (2005)
35. M. Fanciulli, G. Scarel (eds.), *Rare Earth Oxide Thin Films* (Springer, Berlin, 2007)
36. W. H. Zachariasen, J. Am. Chem. Soc. **54**, 3841 (1932)
37. J. C. Philips, J. Non-crystalline Solids **34**, 153 (1979)
38. H. Inoue, F. Utsuno, I. Yasui, J. Non-Cryst, Solids **349**, 16–21 (2004)
39. G. Lucovsky, Y. Wu, H. Niimi, V. Misra, J. C. Phillips, J. Vac. Sci. Tech. **B17**, 1806–1812 (1999)
40. A. I. Kingon, J-P. Maria, D. Wicaksana, C. Hoffnan, in *Extended Abstract of International Workshop on Gate Insulator (IWGI)*, vol 36 (2001)
41. R. S. Roth, J. R. Dennis, H. F. McMurdie, *Phase Diagrams for Ceramists*, vol. 6 p. 145, Fig. 6441 (The American Ceramics Society, 1987)
42. A. Toriumi, T. Nabatame, T. Horikawa, Mat. Res. Soc. Symp. Proc. **786**, 135 (2004)
43. A. Toriumi, K. Iwamoto, H. Ota, M. Kadoshima, W. Mizubayashi, T. Nabatame, A. Ogawa, K. Tominaga, T. Horikawa, H. Satake, Microelectronic Engineer. **80**, 190 (2005)
44. V. Ushakov, C. E. Brown, A. Navrotsky, J. Mater. Res. **19**, 693 (2004)
45. A. Dimoulas, G. Vellianitis, G. Mavrou, G. Apostolopoulos, A. Travlos, C. Wiemer, M. Fanciulli, Z. M. Rittersma, Appl. Phys. Lett. **85**, 3205 (2004)
46. Y. Yamamoto, K. Kita, K. Kyuno, A. Toriumi, Appl. Phys. Lett. **89**, 032903 (2006)
47. L. Bixner, Mater. Res. Bull. **19**, 143 (1984)
48. N. Umezawa, K. Shiraishi, S. Sugino, A. Tachibana, K. Phmori, K. Kakushima, H. Iwai, T. Chikyow, T. Ohno, Y. Nara, K. Yamada, Appl. Phys. Lett. **91**, 132904 (2007)
49. Y. Zhao, K. Kita, K. Kyuno, A. Toriumi, J. Appl. Phys. **105**, 034103 (2009)
50. J. M. J. Lopes, M. Roeckerath, T. Heeg, E. Rije, J. Schubert, S. Mantl, V. V. Afanas'ev, S. Shamulia, A. Stesmans, Y. Jia, D. G. Schlom, Appl. Phys. Lett. **89**, 222902 (2006)

Chapter 11
Crystalline Oxides on Silicon

H. Jörg Osten

Abstract The ability to integrate crystalline metal oxide dielectric layers into silicon structures can open the way for a variety of novel applications which enhances the functionality and flexibility, ranging from high-k gate dielectric replacements in future Metal Oxide Semiconductor (MOS) devices to oxide/silicon/oxide heterostructures for nanoelectronic application in quantum-effect devices. We present results for crystalline gadolinium oxides on silicon in the cubic *bixbyite* structure grown by solid source molecular beam epitaxy. On Si (100) oriented surfaces, crystalline Gd_2O_3 grows as (110)-oriented domains, with two orthogonal in-plane orientations. Layers grown under best vacuum conditions often exhibit poor dielectric properties due to the formation of crystalline interfacial silicide inclusions. Additional oxygen supply during growth improves the dielectric properties significantly. Experimental results for Gd_2O_3-based MOS capacitors grown under optimized conditions show that these layers are excellent candidates for application as very thin high-k materials replacing SiO_2 in future MOS devices. Epitaxial growth of lanthanide oxides on silicon without any interfacial layer has the advantage of enabling defined interfaces engineering. We will show that the electrical properties of epitaxial Gd_2O_3 thin films on Si substrates can further be improved significantly by an atomic control of interfacial structures. The incorporation of few monolayers of Ge chemisorbed on the Si surface has been found to have significant impact on the electrical properties of crystalline Gd_2O_3 grown epitaxially on Si substrates. Efficient manipulation of Si(100) 4° miscut substrate surfaces can lead to single domain epitaxial Gd_2O_3 layer. Such epi-Gd_2O_3 layers exhibited significant lower leakage currents compared to the commonly obtained epitaxial layers with two orthogonal domains. For capacitance equivalent thicknesses below 1 nm, this difference disappears, indicating that for ultrathin layers, direct tunneling becomes dominant. Further, we investigate the effect of post-growth annealing on layer properties. We show that a

H. J. Osten (✉)
Institute of Electronic Materials and Devices, Leibniz University Hannover,
Schneiderberg 32 D-30167 Hannover, Germany
e-mail: osten@mbe.uni-hannover.de

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5_11,

standard forming gas anneal can eliminate flat-band voltage instabilities and hysteresis as well as reduce leakage currents by saturating the dangling bonds caused by the bonding mismatch. In addition, we investigated the impact of rapid thermal anneals on structural and electrical properties of crystalline Gd_2O_3 layers grown on Si. Finally, we will present a new approach for nanostructure formation which is based on solid-phase epitaxy of the Si quantum-well combined with simultaneous vapor-phase epitaxy of the insulator on top of the quantum-well. Ultra-thin single-crystalline Si buried in a single-crystalline insulator matrix with sharp interfaces was obtained by this approach on Si(111). In addition, structures consisting of a single-crystalline oxide layer with embedded Si nano-clusters for memory applications will also be demonstrated.

11.1 Introductory Remarks

Many materials systems are currently under consideration as potential replacements for Si-based gate dielectrics for sub-100-nm Complimentary MOS (CMOS) technology. The gate leakage current, at least for direct quantum–mechanical tunneling, is exponentially dependent upon the thickness of the dielectric, while the capacitance is only linearly dependent on the thickness. Short-channel performance degradation is caused by the fringing fields from the gate to the S/D (Source/Drain) regions, depending on the gate dielectric thickness-to-length aspect ratio. Therefore, materials with too high dielectric constant will create additional problems. A desirable alternative gate insulator material should have a dielectric constant of five to ten times higher than SiO_2 that is between 20 and 40. The use of such high-k dielectrics would provide scaled-down values of Equivalent Oxide Thickness (EOT) required to maintain the Field Effect Transistor (FET) current drive, while reducing tunneling current through an increase in the film thickness [1]. At first glance, this would seem to be a winning proposal, since a substantial reduction in the current should be possible with only small increases in thickness. There is, however, another exponentially dependent term in the tunneling current—the barrier height between the semiconductor and the conduction band of the insulator. The tunnel current is exponentially dependent upon the barrier height. Therefore, not only is a material with higher dielectric constant required, but this material must also have a suitably large barrier height, to keep the gate leakage currents within reasonable limits. However, physical thickness and interfacial band alignment are only two factors that affect the device performance. Other basic limitations for high-k gate dielectric substitutions include those associated with the chemical and the physical bonding constrains. Bonding at the $Si–SiO_2$ (and also $Si–Si_3N_4$) interface is *isovalent* with bond charge exactly matching the nuclear charge. Each dangling bond on a Si surface has one electron, and this contributes to an interfacial two electron pair bond with an oxygen (or nitrogen) atom. Most high-k materials such as the binary metal oxides are generally more ionic, and

additionally more highly coordinated on the average than SiO_2. In a simple picture, one would expect, that these dielectrics will form *heterovalent* interfaces with Si with a quantitative mismatch between bond and nuclear charge. This mismatch can give rise to interfacial traps and/or fixed charges.

Another limitation on the material choice concerns stability problems [2]. The alternative gate dielectrics have to be compatible with a CMOS process flow. That poses severe restrictions on the materials due to the post-growth processing steps (like thermal stability, or manufacturability). The high-k dielectric will most likely be deposited on a Si channel. Therefore, it should not show any tendency to form silicides, silicates, silicon dioxides, or mixtures of those during post-deposition annealing. Thermodynamic stability consideration reduces the number of possible candidates significantly.

Many dielectrics appear favorable in some of these fields, but very few materials are promising with respect to all of these guidelines. The most promising of these are the simple binary metal oxides. Unfortunately, a number of these materials are not thermally stable on silicon [2]. The formation of SiO_2 and/or metal silicate interfacial layers often occurs when these materials are deposited upon silicon. Further growth of silicon dioxide or a silicate at the interface takes place during subsequent annealing steps. It is important to note, however, that the occurrence of any interfacial layer of SiO_2 or another low permittivity material, will limit the highest possible gate stack capacitance, or equivalently, the lowest achievable EOT value. This effect of reduced capacitance can be seen by noting that when the structure contains several dielectrics in series, the lowest capacitance layer will dominate the overall capacitance and also set a limit on the minimum achievable t_{eq} value. For example, the total capacitance of two dielectrics in series is given by

$$1/C_{tot} = 1/C_1 + 1/C_2, \tag{11.1}$$

where C_1 and C_2 are the capacitances of the two layers, respectively. If one considers a dielectric stack structure such that the bottom layer (layer 1) of the stack is SiO_2, and the top layer (layer 2) is the high-k alternative gate dielectric (11.1) is simplified (assuming equal areas) to

$$t_{eq} = t_{SiO2} + \left(k_{ox}/\, k_{high-k}\right) t_{high-k}. \tag{11.2}$$

For example, to obtain an EOT of 1.2 nm one could use a hypothetical 7.7 nm thick layer with $k = 30$. The presence of a 0.8 nm interfacial SiO_2 layer would reduce this thickness down to only 1.5 nm. For the same required EOT the physical thickness of the high-k material decreases significantly due to the interfacial layers; thus, the leakage current through the film also increases substantially. Therefore, much of the expected increase in the gate capacitance associated with the high-k dielectric is compromised.

The common approach for alternative high-k gate dielectrics has involved the amorphous metal oxides. In general, attempts have been made to keep these materials amorphous, in particular, after post-deposition high temperature

processing in order to avoid increased surface roughness and additional leakage due to the formation of grain boundaries. Generally, all known high-k materials are highly ionic. There are some general trends related to the ionicity of an oxide: The crystallization temperature decreases and the dielectric constant increases with the ionic character. Due to the relatively low re-crystallization temperature of highly ionic materials, these materials are often not compatible to a CMOS process. It is possible to reduce the ionicity (and thereby increase the crystallization temperature) by alloying the metal oxides with Al or Si. These aluminates or silicates are thermally more stable on silicon but have lower dielectric constants. Consequently, for a given EOT, the physical layer thickness has to be reduced, leading to an increase in the leakage current.

The upcoming generation of high-k dielectrics will most likely be formed by amorphous hafnium-based alloys. As discussed above, however, the existence of an interfacial layer of SiO_2 or another low permittivity material, will limit the highest possible gate stack capacitance, or equivalently, the lowest achievable EOT value (see 11.2). Thus, any interfacial lower-permittivity layer should be minimized in the future generations. However, it is essential to keep the channel carrier mobility high. In addition, the increased process complexity for the deposition and control of additional ultrathin dielectric layers, as well as scalability to later technology nodes, remains a concern.

The 2nd generation of high-k materials have to be formed without any interfacial layer to realize EOT <0.8 nm. In such a case, the high-k dielectric/Si interface properties influence the device performance significantly. A good interface requires either that the oxide is amorphous, or that it is epitaxial and lattice-matched to the underlying silicon. Amorphous dielectrics are expected to be able to adjust the local bonding to minimize the number of Si dangling bonds at the interface. The alternative is to use an epitaxial oxide. This involves more effort, but it has the advantage of enabling defined interface engineering. Generally, there are two groups of possible candidates for epitaxial growth on Si, namely (a) perovskite-type structures and (b) binary metal oxides, in particular lanthanide oxides [3, 4]. In the following, we will concentrate on the binary lanthanide oxides.

11.2 Lanthanide Oxides on Silicon

Lanthanide oxides (LnO's) form a very interesting group of insulators for epitaxial growth on silicon [5]. It is extremely important and desirable to integrate these highly functional metal oxides into mature semiconductor technologies. The LnO's can have different oxygen compositions LnO_x, with x ranging from 1 to 2 due to the multiple oxidation states (+2, +3, and +4) of the rare-earth metals [6, 7]. This leads to oxides with different stoichiometries (LnO, Ln_2O_3, LnO_2). All known Ln (II) oxides, like EuO, are not insulating. Therefore, they will not be considered here any further. For application in a Si-based device fabrication process, all lanthanide oxides exhibiting more than one valence state (+3 and +4) are not the

best choice as epitaxial high-k materials because of the coexistence of phases with different oxygen content. For example, cerium(IV) oxide (CeO_2) can release oxygen under reducing conditions forming a series of reduced oxides with stoichiometric cerium(III) oxide (Ce_2O_3) as an end product, which in its turn easily takes up oxygen under oxidizing conditions, turning the cerium(III) oxide back into CeO_2. In addition, stable mixed valence-state structures can occur for some LnO's. For example, the mixed valence-state Pr_6O_{11} is the most stable phase for praseodymium oxide.

For the highly ionic oxides, the position of the charge neutrality level depends strongly on the stoichiometry [8]. Thus, also the band alignment to silicon and, finally, the leakage behavior becomes strongly dependent on the oxygen content. All lanthanide oxides displaying only one valence state are easier to handle due to the absence of transitions between phases with different oxygen content. Based on that argument, we will focus our discussion mainly on lanthanide (III) oxides (occurring as Ln_2O_3).

The Ln_2O_3 oxides can occur in different structural phases, like the manganese oxide (Mn_2O_3) or *bixbyite* structure. Some of these oxides also crystallize in the hexagonal lanthanum oxide structure, which is suitable for epitaxy only on Si(111). Also, monoclinic phases are known for various lanthanide(III) oxides. Different crystallographic structures are accompanied by different dielectric properties. Several lanthanide oxides can undergo structural phase transformation within a temperature range, typical for CMOS processing [9], see Fig. 11.1; these oxides are not very well suited for technological applications.

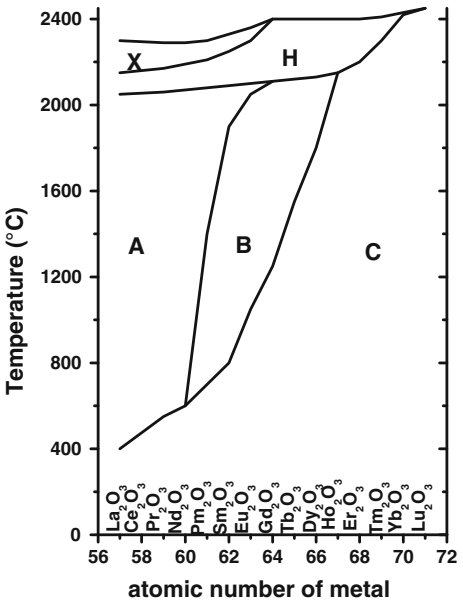


Fig. 11.1 Structural phase transitions in lanthanide oxides (based on Ref. [9]): *A*: hexagonal (P63/m) *B*: monoclinic (C2/m) *C*: cubic (Ia-3) *H*, *X* high-T modifications

Epitaxial growth on a clean surface requires matching in symmetry as well as in atomic spacing. Here, we will present results for crystalline gadolinium oxide on silicon with the Gd_2O_3 composition in the cubic *bixbyite* structure which has a large band gap of about 6 eV and nearly symmetrical band offsets to Si [10] as well as a low lattice mismatch of about 0.5 %. The *bixbyite* structure is based on the calcium fluorite structure, where 1/4 of the oxygen atoms have been removed from specific lattice sites. That structure has a lattice symmetry suitable for epitaxial growth on Si(100) and Si(111).

Commonly, epitaxial heterostructures are evaluated on the basis of lattice matching, with the misfit defined as the relative difference in the lattice constants (a_{film} and a_{Si}). Here, the lattice mismatch would be identical for all three epitaxial relationships, i.e. (100)//(100), (110)//(110), and (111)//(111). All other combinations violate symmetry matching. This concept is misleading, because epitaxial growth is also governed by the surface and interface energetics. In the case of the rare-earth oxides, the surface energy of the (100) surface is much higher than that of the Si(100) surface [11]. That could result in a 3-dimensional growth mode or in a change of the growth direction towards a low-energy surface orientation. The latter is observed for the growth on Si(100) for most of the lanthanide oxides, where the oxides were found to grow in the (110) orientation. For the understanding of the Ln_2O_3(110)/Si(100) interface formation we have to consider that the lattice is made up of metal atoms occupying the positions of a face-centered cubic lattice with a lattice constant a_{film}, where the tetrahedral holes are occupied by the oxygen atoms. However, the growth of LnO's is based on the deposition of metal oxide molecules. Due to the existence of highly ionic Ln–O bonds in combination with the high bonding strength of the covalent Si–O bonds, we can assume that the interface is predominantly formed by Si–O–Ln bonds. Therefore, the matching of the oxygen atoms is the important parameter (Fig. 11.2).

The complete crystallographic structure can also be described by two non-identical metal and oxygen lattices, respectively. The arrangement of the oxygen atoms forms a simple cubic lattice with a lattice constant of $a_{film}/2$. Thus, the interesting matching condition for epitaxial growth is the Ln_2O_3(110)[100]//

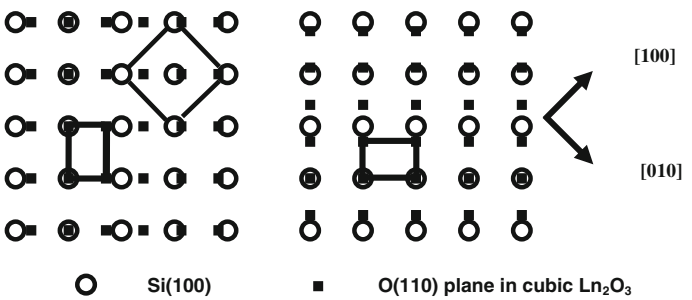


Fig. 11.2 Ln_2O_3(110)//Si(100) alignment (schematically): 1:1 matching occurs along one direction. In the other direction, there is roughly a 3:2 matching relation. Two types of (110)-oriented domains are possible, with two orthogonal in-plane orientations

Si(100)[110] relation. In that case, nearly 1:1 matching should occur along one direction. In the other direction, there would be roughly a 3:2 matching relation. As shown schematically in Fig. 11.2, layers grown in this orientation exhibit mostly two types of (110)-oriented domains, with two orthogonal in-plane orientations (found experimentally for a large variety of binary metal oxides on Si(100), like Er_2O_3 [12], Sm_2O_3 [13], Lu_2O_3 [14, 15], Sc_2O_3 [16], Gd_2O_3 and Y_2O_3 [17]). That is even enforced by the Si(100) dimer (2 × 1) surface reconstruction. Dimer rows on adjacent terraces are oriented perpendicular to each other and nucleation of the oxide on Si(100) follows the dimer orientation, which results in the domains on stepped surfaces [18]. Due to the 45° rotation of the (110) plane relative to the substrate, the mismatch for the 1:1 matching is identical to that one obtained from applying the lattice constant concept.

11.3 Epitaxial Growth of Lanthanide Oxides on Silicon

All experiments were performed in a multi-chamber ultrahigh-vacuum system (*DCA Instruments*) capable of handling 8" wafers. This system includes a growth, an annealing, and an analysis chamber connected by an ultra-high vacuum (UHV) transfer system. The layers were grown on 4" Si substrates with different orientations. Substrates were cleaned *ex situ* using as the last step diluted HF etch ($HF:H_2O = 1:10$) followed by a dilution rinse, and then were immediately inserted into the vacuum system. Substrates were annealed in situ to transform the initial hydrogen-terminated (1 × 1) surface structure into the (2 × 1) superstructure indicating a clean and well-ordered surface. Commercially available, granular Gd_2O_3 material was evaporated using an electron-beam evaporator. Growth temperatures were in the range 800–1,000 K. Typical growth rates were 0.005–0.01 nm/sec. The surface and layer structure was evaluated during the growth by reflection of high energy electron diffraction (RHEED).

After growth, the wafers can be transferred into the analysis chamber without leaving the UHV to perform x-ray photoelectron spectroscopy (XPS) investigations. Non-monochromatized Al Kα radiation ($h\upsilon = 1{,}486.6$ eV) was used for the excitation of photoelectrons. All measured wafers were electrically grounded to eliminate charging effects during long-time measurements. To minimize experimental uncertainties associated with energy variations caused by spectrometer instabilities and to improve the signal to noise ratio, the XPS data were collected by repeatedly scanning the silicon, the lanthanide, and the oxygen level. A multipeak Gaussian deconvolution procedure was used to extract the exact line position and intensities.

The layer thickness was measured *ex vacuo* by X-ray reflectivity (XRR) using a standard single crystal diffractometer with graphite monochromator in front of the detector. Layers were also characterized by X-ray diffraction (XRD) ($\Theta/2\Theta$, ω and Φ-scans) and transmission electron microscopy (TEM), i.e., high resolution cross-section and plan-view images combined with selected area diffraction (SAD).

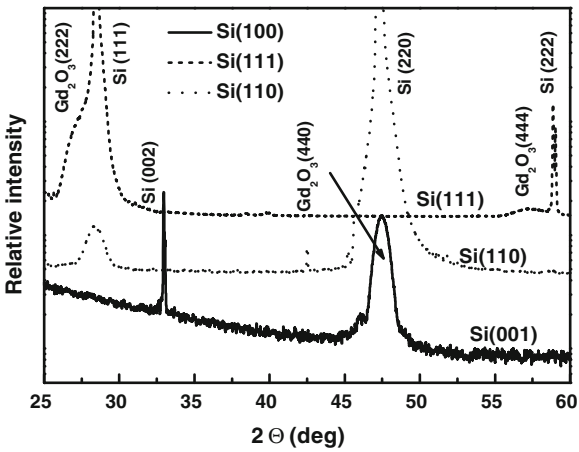


Fig. 11.3 X-ray diffraction patterns of Gd_2O_3 thin films grown on silicon substrates with different orientations at a substrate temperature of 700 °C

On the Si(100) substrate, the grown layers exhibit the known two types of (110)-oriented domains, with two orthogonal in-plane orientations [19]. On (111) oriented Si surface, *bixbyite* Gd_2O_3 grows epitaxially when the substrate temperature is above 600 °C. Here, the oxide exhibits an A/B twining relationship where B is related to the substrate twinning orientation A by a 180° rotation around the Si(111) surface normal. On a Si(110) surface, grown layers exhibit pronounced faceting by the development of low-energy {111} facets [20].

Figure 11.3 shows the X-ray ($\Theta - 2\Theta$ scan) diffraction patterns of Gd_2O_3 thin films grown on different Si substrates. Gd_2O_3 layer on the Si(100) surface exhibits a distinct peak at $2\theta = 46.4^{o}$ corresponding to the d(440) inter planar spacing of cubic Gd_2O_3 along the [110] orientation. The layers on the Si(111) substrate also exhibit single orientation without any indication of disoriented crystallites. The peaks at 28.5° and 59.4° corresponding to d(222) and d(444) interplanar spacing are hidden under the appropriate Si peaks. The $\Theta - 2\Theta$ scan for Gd_2O_3 thin films grown on Si(110) displays a peak at 28.5° confirming the preferential growth along the (111) direction as observed in RHEED images (not shown).

11.4 Electrical Characterization

Figure 11.4 compares the capacitance versus voltage (C–V) characteristics of Gd_2O_3 thin films deposited on Si substrates with different orientations. For the comparison, films with similar thickness in the range of 4–5 nm have been chosen [20]. As demonstrated in the C–V characteristics, there is no distinct difference in the accumulation capacitance for the films grown on differently oriented Si substrates. Despite having a different crystalline layer structure, the films on Si(110) substrate exhibit electrical properties similar to other orientations. The accumulation capacitance measured for Gd_2O_3 on Si(110) was close to the value of those

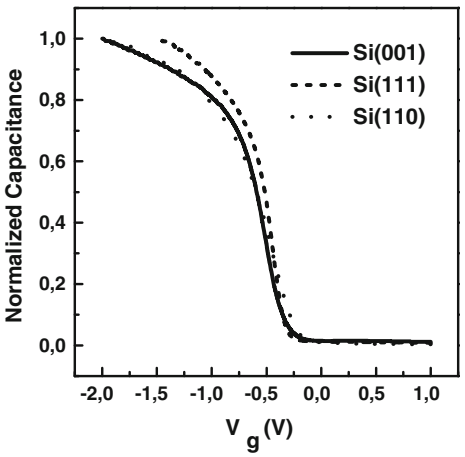


Fig. 11.4 Normalized capacitance–voltage characteristics of epitaxial Gd_2O_3 thin films on different silicon substrates

grown on other two substrates. The capacitance equivalent thicknesses (CET) estimated for all three films were around 0.9 nm. No hysteresis was observed in the C–V measurements. The normalized capacitance for all three cases shows also no differences in the flat band voltage (~ -0.25 V). The flat-band voltage shift is generally due to the flat-band interface charge, to which both the fixed charge as well as the flat-band interface trap charge contributes. It infers that the same amount of positive charges is involved in each case, i.e. that the substrate orientation has also no significant influence on fixed charge concentration.

Interface trap density (D_{it}) and leakage current density (J) are also important parameters which characterize the gate dielectrics (Figs. 11.5 and 11.6). Among the various techniques available, we have used the conductance method for evaluating the D_{it}. This method is generally considered to be most sensitive for this purpose. The equivalent parallel conductance (G_p) of MOS capacitor is measured as a function of the gate bias and the frequency ($2\pi f$), the details of which are described in Ref. [21]. G_p calculated from the measured conductance, shown in Fig. 11.6 as a function of the frequency, displays a maximum at a particular

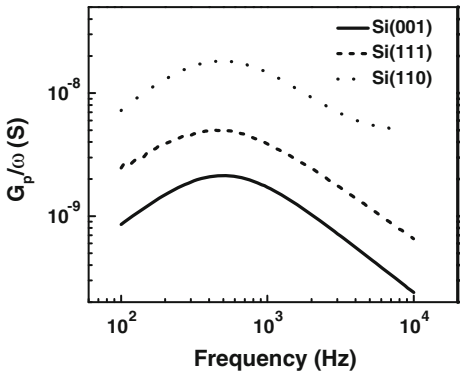


Fig. 11.5 Equivalent parallel conductance (G_p/ω) estimated from measured conductance (G_m) as a function of frequency at $V_g = -0.5$ V for three different substrates

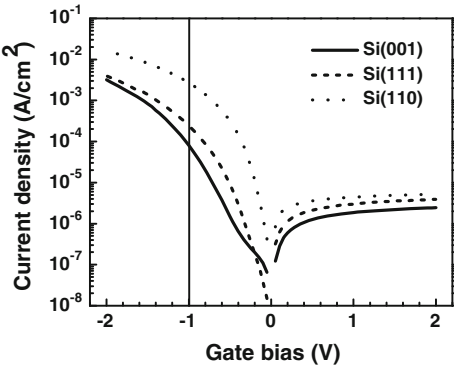


Fig. 11.6 Leakage current characteristics of epitaxial Gd_2O_3 thin films grown on different silicon substrates

frequency. The values of D_{it} estimated from the peak in the G_p/ω versus ω plot are $1.0 \cdot 10^{12}$, $2.4 \cdot 10^{12}$ and $9.0 \cdot 10^{12}$ $cm^{-2}eV^{-1}$ for Si(100), Si(111) and Si(110), respectively. However, broadening in the peak could be attributed to the interface traps having distribution of relaxation times. The current density in Pt/Gd_2O_3/Si(110) MOS structure is found to be slightly higher in comparison to other substrates; nevertheless, the value is much lower than the expected value for SiO_2 capacitor with the same equivalent thickness. We suppose that the larger leakage current in Gd_2O_3/Si(110) capacitor could be due to structural defects in the bulk and at the interface. Reference [10] contains further investigations of the electronic structure at interfaces of crystalline and amorphous Gd_2O_3 thin layers with silicon substrates of different orientations.

11.5 Impact of Oxygen Concentration on Layer Properties

Recently, we investigated the interface and the layer formation processes of Ln_2O_3 films on Si(100) grown under ultra-clean ultra-high vacuum (UHV) conditions of Molecular Beam Epitaxy (MBE) [22]. Layers grown under best vacuum conditions often exhibit very high leakage currents. We found that the partial oxygen pressure during the interface formation and during the growth is a very crucial parameter. Too low oxygen content can lead to the formation of silicide-like inclusions. For Ln_2O_3 growth, the formation and stability of the silicide-like phase depends on the oxygen chemical potential [23]. Considering the low oxygen partial pressure under the UHV conditions (as used in MBE growth), the chemical potential of oxygen can become negative. Thus, silicide formation will be favorable compared to oxide formation. That is one of the most crucial points for the growth of dielectric layers, because the silicide growth can continue as long as the oxygen content remains low enough or the oxygen chemical potential remains strongly negative. This occurs even faster at a surface or interface region where the thermodynamic equilibrium is distorted, for example, due to stress. The appearance of silicide inclusions seems to be a serious drawback for a lot of possible material candidates

for future high-k applications. For the epitaxial growth of Ln_2O_3 oxides, a well-controlled interface and appropriate engineered growth procedures are necessary to realize the required electrical properties. The available oxygen concentration can be controlled by modifying the MBE growth processes using an additional oxygen supply. We found, that MBE growth under defined oxygen partial pressures during the interface formation and/or during the subsequent growth can prevent any kind of silicide inclusions and void formation [24]. On the other hand, too high oxygen content might oxidize the Si surface, leading to a lower-k interfacial SiO_x layer, which finally limits the achievable minimal equivalent oxide thickness.

In a set of experiments, we varied the growth temperature while keeping all the other parameters constant. Figure 11.7 shows the extracted CET values versus the physical oxide thickness. The oxygen partial pressure during growth was kept at $1 \cdot 10^{-7}$ mbar. In a simple picture, a system including an interfacial layer and a high-k layer can be described by a linear relation between CET and the physical thickness (11.2).

The experimental results shown in Fig. 11.7 follow that relation. The slope for all three temperatures is nearly equal resulting in an intrinsic dielectric constant of around 20 for the high-k layer in all the three cases. Only the intercept for $t = 0$ varies with the growth temperature, yielding 0.4, 0.9, and 1.4 nm, respectively. Based on (11.2), this intercept is often attributed to the physical thickness of an interfacial layer only. However, an increase of the thickness with the increasing growth temperature was not supported by XRR and XTEM (Cross-sectional TEM) investigations. Instead, we suppose that the dielectric constant (k_{IF}) at the interface decreases due to the transformation of the interfacial layer from a silicate-like type to a more silicon-oxide like layer. That can be explained by the fact that gadolinium, in contrast to cerium or praseodymium, can only exist in the +3 oxidation state. Thus under equilibrium conditions, the Gd_2O_3 bulk cannot act as an effective source for oxygen supply to the Si/dielectric interface. For an interfacial layer other than silicon dioxide (11.2) transforms into

$$CET = t_{IF}(3.9/k_{IF}) + \left(3.9/k_{high-k}\right)t_{high-k}. \tag{11.3}$$

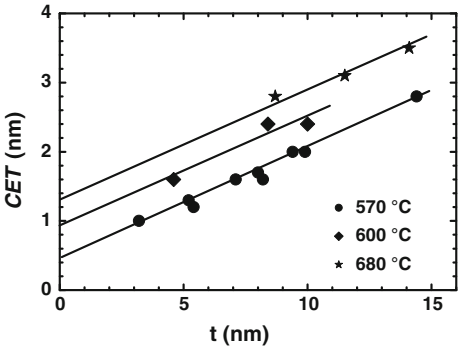


Fig. 11.7 CET as a function of the physical layer thickness for Gd_2O_3 layers grown on p- and n-type Si(100) at 570, 600, and 680 °C, respectively. The oxygen partial pressure during growth was kept at $1 \cdot 10^{-7}$ mbar

Next, we investigated the influence of different oxygen partial pressures. The results are very similar for all investigated growth temperatures. For the layer grown with the lowest oxygen pressure ($1 \cdot 10^{-7}$ mbar), we always found strong hysteresis which we attribute to the presence of oxygen vacancies. Increasing the oxygen pressure to $5 \cdot 10^{-7}$ mbar results in a significant reduction of the hysteresis. Further increase in oxygen partial pressure p_{O2} does not lead to a further improvement of the electrical properties of the layer, but increases the physical thickness of the interfacial layer, as detected by XTEM. At this pressure, the oxygen concentration becomes supersaturated at the growth front, and oxygen atoms diffuse to the interface.

We obtained the best electrical results for growth at 600 °C and $p_{O2} = 5 \cdot 10^{-7}$ mbar. Figure 11.8 shows a cross-sectional TEM micrograph of a sample with a 15 nm thick Gd_2O_3 layer grown under these conditions on Si(100). No pronounced interfacial layer between the Si and the oxide layer can be detected.

Layers grown by an optimized MBE process display a sufficiently high-k value to achieve equivalent oxide thickness values <1 nm (Fig. 11.9), combined with ultra-low leakage current densities, good reliability, and high electrical breakdown voltage. This makes epitaxial Gd_2O_3 layers excellent alternative for replacing SiO_2 as a gate dielectric. First electrical characteristics of fully-depleted n- and p-type Silicon On Insulator (SOI)-MOSFETs with epitaxial Gd_2O_3 and TiN metal gate electrodes demonstrate the feasibility of this novel gate insulator [25]. Electrical properties of the gate stack have been extracted from the devices and CMOS process compatibility has been addressed. The oxide interface state density has

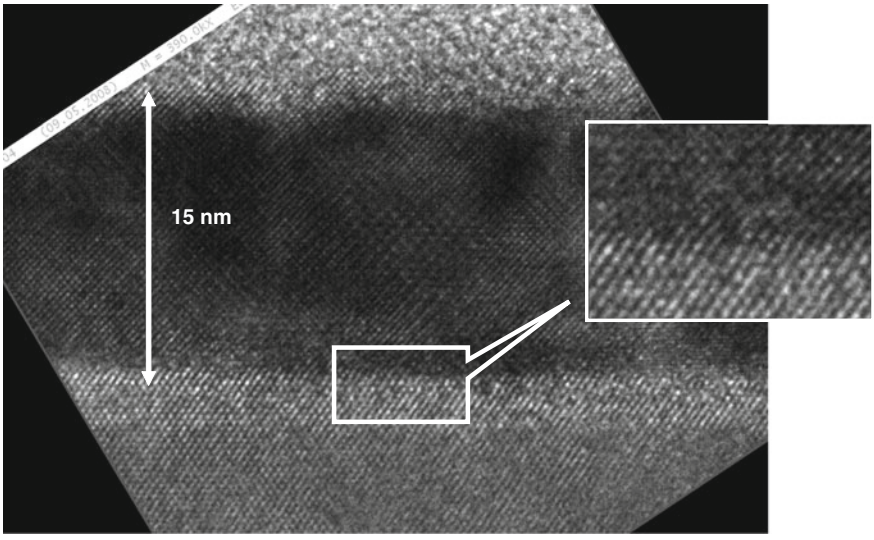


Fig. 11.8 High-resolution cross-sectional TEM image of a Gd_2O_3/Si(100) stack. The oxide layer was grown at 600 °C and $p_{O2} = 5 \cdot 10^{-7}$ mbar. No pronounced interfacial layer is visible

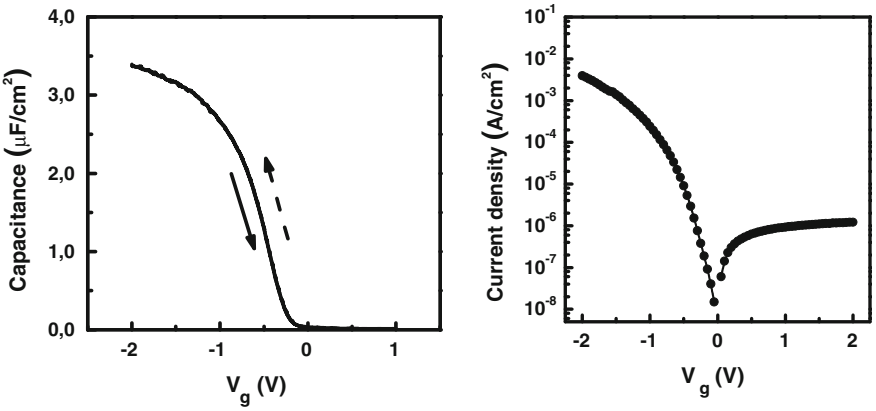


Fig. 11.9 MOS capacitor with Gd_2O_3 on p-Si(100): EOT <0.7 nm, hysteresis <10 mV, J @ ($V_{FB} - 1$ V) = 0.5 mA/cm^2

been found to decrease after rapid thermal annealing. On the other hand, mobile oxide charges and oxide traps have been significantly reduced by Rapid Thermal Annealing (RTA). The p-MOSFETs exhibit higher saturation currents compared to n-MOSFETs as they were accumulation-mode transistors with a p^+–p^-–p^+ doping structure.

In accumulation-mode p-MOSFETs, higher drain currents can be achieved as conduction appears at the front interface as well as in the body region below compared to an inversion channel which is confined at the front interface [25]. In order to minimize process induced oxide damage, MOSFET's with W/Gd_2O_3 gate stack have been fabricated in a virtually damage-free damascene metal process [26]. The major process steps of the replacement gate process are the following: Initially, dummy gate stacks are formed followed by self-aligned S/D ion implantation. An alignment-oxide is deposited by chemical vapor deposition (CVD), annealed and planarized by chemical–mechanical polishing (CMP) down to the gate level. The dummy gates are removed completely by wet chemical etching, leaving a self-aligned imprint of the gate stack on the Si-wafer. Subsequently, the gate dielectric (Gd_2O_3) is grown. A metal layer is deposited on top of the gate dielectric and CMP is used to pattern the damascene metal gates. Standard back-end processing completes the fabrication. This was the first successful attempt to integrate crystalline high-k dielectrics into a "gentle" damascene metal gate process in order to reduce process induced oxide damages.

The lattice parameters of Gd_2O_3 and Nd_2O_3 in their *bixbyite* phase are 1.081 and 1.108 nm, respectively. Thus, $2a_{si}$ is 0.5 % larger than Gd_2O_3 and 2.1 % smaller than Nd_2O_3, respectively. Therefore, a combination of these two oxides would create a system exhibiting exact lattice matching with Si especially at the deposition temperature. Therefore, such a ternary system would provide much more flexibility in choosing the epitaxial high-k oxide for next generation CMOS devices. Recently, we reported the growth and electrical properties of crystalline

mixed $(Nd_{1-x}Gd_x)_2O_3$ thin films and compared the results with those of binary Gd_2O_3 and Nd_2O_3 thin films, respectively [27]. Also the ternary films show excellent electrical properties with the CET values far below 1 nm, and associated leakage currents below 1 mA/cm^2, respectively. The electrical properties of the multi component oxides are comparable to those of the binary oxides. This work reveals that such mixed oxides would enlarge the window for material selection suitable for the upcoming generations of CMOS devices.

11.6 Effect of Domain Boundaries

Despite having a suitable crystalline structure, epitaxial rare earth oxides, when grown on Si(100) substrates, usually exhibit two (110) oriented orthogonal domains and hence, one could anticipate a great impact of domain boundaries on the final electrical performance. However, to elucidate the impact of such domain boundaries is more complicated because the growth of epitaxial lanthanide oxides in the common *bixbyite* structure on standard Si(100) substrates always leads to the formation of these two types of [110]-oriented domains, with two orthogonal in-plane orientation, each of them exhibiting two fold symmetry. It was suggested by Kwo et al. [17, 28] that the use of vicinal (4° miscut along <110>azimuth) Si(100) substrate could be a viable way to grow single-domain (SD) epitaxial layer, thus to eliminate domain boundary effect on electrical properties similar to what was also known from the growth of III–V compounds on Si(100) [29].

Recently, we showed that not only the use of vicinal surface is necessary but also the preparation of silicon surface prior to the layer growth is a very important step to achieve single domain epitaxial Ln_2O_3 layer on vicinal (4° miscut) Si(100) substrates [30]. The reason could be understood in the following way. It was reported earlier that the presence of dimers on a (2 × 1) reconstructed Si(100) surface forces the overgrown lanthanide oxide layer to form orthogonal oriented ad-dimers in line with Si dimer rows [31]. Normally, dimer rows on adjacent terraces are oriented perpendicular to each other and nucleation of the oxide on (2 × 1) Si(100) follows the dimer orientation, which results in the domains. If, this makes an arguable explanation; then, it seems to be possible to grow a single domain (SD) epi-layer on Si substrate if and only if all dimers on the surface are parallel to each other. Obviously, the next question is how to achieve such Si(100) surfaces with only one dimer orientation? Double-atomic steps result in the Si-dimers arranged only in one orientation across the whole surface. Such steps are easier to achieve on 4°-miscut surface; however, it demands a careful preparation of the Si(100) surface (details can be found in Ref. [30]).

Various layers were grown under identical deposition conditions (substrate temperature of 675 °C with oxygen partial pressure of $5 \cdot 10^{-7}$ mbar) on Si(100) 4° off substrates with and without surface preparation. The detailed structural quality of the layers was investigated by x-ray diffraction technique. The asymmetric 360° x-ray phi (Φ) scans (azimuthal rotation) were carried out to determine

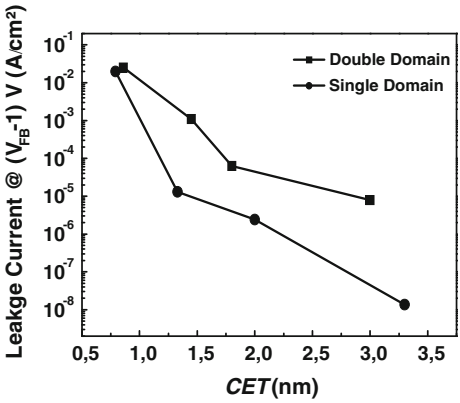


Fig. 11.10 Leakage current density of as-grown Gd_2O_3 SD and DD layers as a function of capacitance equivalent oxide thickness (CET)

the in-plane symmetries of the Gd_2O_3 layers and therefore, to confirm the presence of only a single or/and double domains in the layers. From these investigations (and additional TEM investigations) we can confirm that domain boundaries could be eliminated completely in the epitaxial single crystalline lanthanide oxide grown on carefully prepared 4°-off oriented Si(100) substrates. Next we performed electrical evaluations on those layers (Fig. 11.10). They clearly demonstrate that the as-grown SD Gd_2O_3 layers exhibit much lower leakage current for similar CET values, inferring that the domain boundaries in double domain Gd_2O_3 layers act as the leakage paths for the charge carriers. However, this disparity in the leakage current could only be observed for the thicker layers (>6 nm). For capacitance equivalent thickness below 1 nm, such differences disappear, indicating that for ultra thin layers, direct tunneling becomes dominating.

11.7 Influence of Interface Engineering

Epitaxial growth of lanthanide oxides on silicon without any interfacial layer has the advantage of enabling defined interface engineering. Here, we will show that a controlled preparation of the clean (2 × 1) reconstructed silicon surface prior to the Gd_2O_3 growth can create two completely different interfaces; one of them we call as "oxygen-enriched or oxide-like" interface while the other is a "stoichiometric or silicate-like" interface. The influence of the two different interfaces on the electrical properties of the metal/oxide/silicon stacks will be demonstrated [32].

Oxide-like interfaces were prepared as follows. The substrate temperature was increased to 650 °C and once the surface was clean, i.e. (2 × 1) reconstructed, molecular oxygen was introduced into the deposition chamber to adjust the background partial oxygen pressure to $2 \cdot 10^{-7}$ mbar. During this procedure the substrate temperature was held at 650 °C. Just after the (2 × 1) reconstruction was converted to (1 × 1), the Gd_2O_3 was grown with a growth rate of 0.006 nm/sec

under an oxygen partial pressure of $5 \cdot 10^{-7}$ mbar. For the formation of silicate-like interfaces, we performed the following steps: After appearance of the (2×1) reconstruction the substrate temperature was reduced from 650 to 300 °C. Once the temperature was stabilized, Gd_2O_3 was deposited with very slow deposition rate. The deposition was interrupted when the (2×1) reconstruction was converted into (1×1) after a sub-monolayer growth of Gd_2O_3. The substrate temperature was then increased back to 650 °C with ramp rate of 50 deg/min under an oxygen partial background pressure of $1 \cdot 10^{-7}$ mbar to stabilize the oxide surface. At 650 °C, growth of Gd_2O_3 was continued with the growth rate of 0.006 nm/sec under an oxygen partial pressure of $5 \cdot 10^{-7}$ mbar. The diffraction patterns for the epitaxial Gd_2O_3 films grown do not show any differences for the two types of surface modifications.

Figure 11.11 shows the XPS Si *2p* core-level spectra for thin Gd_2O_3 films grown on the two Si surfaces. The spectra exhibit two peaks, the Si bulk peak with the maximum intensity at around 100 eV and a second peak at higher energy corresponding to interfacial silicon. The peak position of the second peak is at higher energy for the oxide-like surface in comparison to the layer grown on the silicate-like silicon substrate. This confirms an oxygen-rich (oxide–like) bonding configuration at the interface for the first case and a metal-O–Si (silicate-like) bonding in the second case [22]. We also investigated the O *1s* and Gd *4d* spectra of Gd_2O_3 thin films with two different interfaces (not shown). Although the interface has minimum impact on those spectra, however, these confirm that Gd metal (140.0 eV) is fully oxidized (Gd^{+3}, 143.5 eV) with corresponding chemical shift of 3.5 eV toward higher binding energy. Results of ab initio model calculations for these two interfacial structures can be found in Ref. [33]. From the corresponding valence band spectra, the valence band maximum (VBM) was estimated following a procedure described elsewhere [12]. The VBM was found to be greater by 0.6 eV in case of the silicate-like interface, which is in complete agreement with a previously reported result [33].

Figure 11.12 shows the capacitance–voltage (C–V) characteristics of two epitaxial 4.2 nm thick Gd_2O_3 films with oxide-like and silicate-like interfaces, respectively, measured at 10 kHz. As anticipated, an epi-Gd_2O_3 film with silicate-like interface layer exhibits much higher accumulation capacitance and hence shows lower capacitance equivalent thickness. The CET values estimated from accumulation capacitance are 1.1 and 0.76 nm for the films with oxide- and silicate-like interface, respectively, while calculated effective dielectric constants are 12.0 and 18.9, respectively. The disparity in the dielectric constants for two differently processed thin films can be understood from the parasitic effect of interfacial layers where oxide-like interface has much lower dielectric constant than silicate-like. The EOT value (including quantum–mechanical corrections) is always significantly lower than *CET*, thus, one would finally be able to achieve an ultra-low EOT for such epitaxial Gd_2O_3 layer with a silicate-like interface. We also compare the leakage current density of epitaxial Gd_2O_3 thin films grown on the two surfaces. The current density estimated at $(Vg - V_{FB}) = 1$ V for two 4.2 nm thick films are 15.0 and 30.5 mA/cm^2 with silicate-like and oxide-like

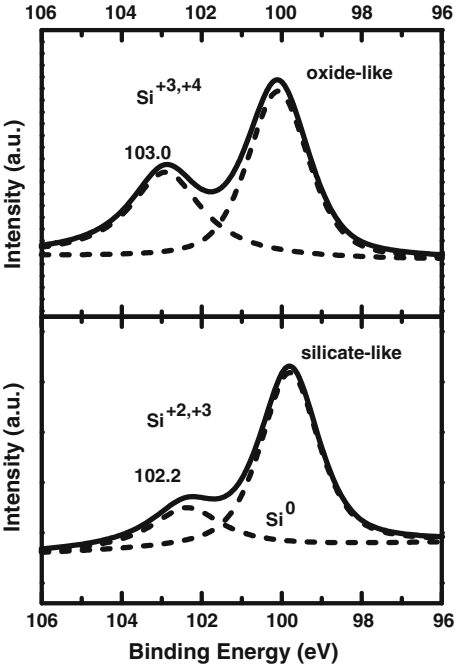


Fig. 11.11 Core-level XPS spectra of Si-2*p*, for Gd_2O_3 layers grown on the two differently prepared Si surfaces

interface, respectively. The films with silicate-like interface show reasonably lower leakage current compared to the films with oxide-like interface. That behavior can be understood from the formation of dipoles at the Gd_2O_3/Si interface originating from the accumulation and/or depletion of oxygen atoms for two the different growth procedures. Hence, the difference in stoichiometry at the interface eventually influences the band offset values [18, 34]. For example, the valence band offset for epi-Pr_2O_3 layer with silicate-like interface was found to be larger than that with oxide-like interface [18]. This likely resulted in a reduced hole current observed in our layers.

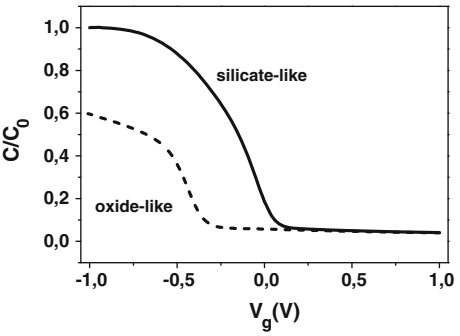


Fig. 11.12 Capacitance–Voltage (C–V) characteristics of epitaxial 4.2 nm thick Gd_2O_3 thin films with oxide- and silicate-like interfaces on p-type Si substrates, measured at 10 kHz

The incorporation of few monolayers of Ge chemisorbed on the Si surface has been found to have significant impact on the electrical properties of crystalline Gd_2O_3 grown epitaxially on Si substrates [35]. Although the Ge coverage on Si surface does not show any influence on the epitaxial quality of Gd_2O_3 layers, however, it exhibits a strong impact on their electrical properties (Fig. 11.13). We show that by incorporating few monolayers of Ge at the interface between Gd_2O_3 and Si, the capacitance–voltage characteristics, fixed charge and density of interface traps of Pt/Gd_2O_3/Si capacitor are much superior to those layers grown on clean Si surfaces.

Summarizing, we have shown that the electrical properties of epitaxial Gd_2O_3 thin films on Si substrates were improved significantly by an atomic control of interfacial structures due to a proper treatment of silicon surface prior to the Gd_2O_3 growth. The interface structure and its chemical composition play therefore the key role on the electrical performance of any MOS devices also in the future.

Recently, we reported on the effect of carbon doping on electrical properties of epitaxial gadolinium oxide grown on Si substrates [36]. The incorporation of small amounts of carbon (0.2–0.5 vol. %) into epitaxial Gd_2O_3 has been found to be very useful in improving the electrical properties especially by reducing the leakage current behavior. The doping has a negligible impact on the structural quality of the oxide layer. We also showed that the very often found adverse effect of oxygen vacancy induced defects in oxides grown at higher temperature can be eliminated by moderate amount of carbon doping during growth. The incorporated C amount should be below 1 % as higher C concentrations (>2 %) into Gd_2O_3 degrades the electrical properties by forming additional type of defects.

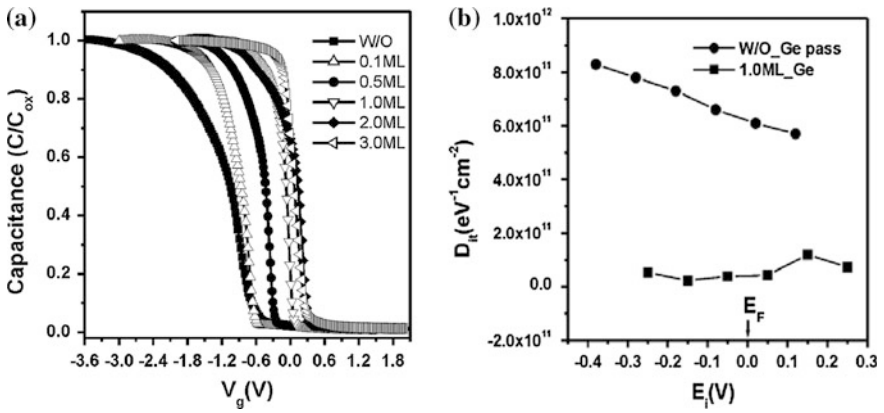


Fig. 11.13 **a** The C–V characteristics of Pt/Gd_2O_3/Si(111) MOS capacitors as a function of Ge coverage. The flat band voltage increases monotonically with increasing Ge coverage, inferring a strong influence of Ge adatoms on Si surface. **b** Density of interface traps of capacitors fabricated on clean and Ge passivated wafers, exhibiting significant improvement on D_{it} values

11.8 Impact of Post-Growth Processing

The *bixbyite* structure of Gd_2O_3 is based on the calcium fluorite structure, where 1/4 of the oxygen atoms have been removed from specific lattice sites. That structure has a lattice symmetry suitable for epitaxial growth on Si(100) with a lattice mismatch of $\sim$0.5 %. Although the mismatch is small, we have to face a bonding mismatch, i.e. a large number of dangling bonds. Recently, we introduced the traditional forming gas anneal to study its impact when applying to the novel single crystalline gate stacks [37].

To fabricate MOS structures, tungsten metal electrodes with an area of 10^{-3} cm^2 were evaporated by e-beam evaporation through a shadow mask in a separate chamber. The wafer was cut into small pieces for the annealing experiments. All structures were annealed in forming gas for 10 min at 450 °C. Electrical characterizations such as *C–V* and *I–V* measurements were performed at room temperature.

Figure 11.14 shows the normalized capacitance–voltage (*C–V*) characteristics of single crystalline Gd_2O_3 on Si(100) substrate with W as top electrode. Although the accumulation and depletion regions are clearly visible, the curves suffer from flat-band voltage instabilities. The flat-band voltage is determined by the slope of $d[1/(C_{hf}/C_{ox})^2]/dV_G$. From sweep 1 to the sweep 8, the flat-band voltage reduced from −0.44 to −1.36 V. A big amount of positive fixed charge is created during the *C–V* sweeping.

Unlike as-grown samples, the V_{FB} of the forming gas annealed samples is quite stable (−0.44 V) indicating the passivation of the slow interface states. The fixed oxide charges are calculated to be $2.3 \cdot 10^{12}$/cm^2 based on the equation: $Q_f = C_{ox} \cdot (\Phi_{ms} - V_{FB})$. Figure 11.15 depicts the hysteresis of the structure. The cycles are completed by from 1.5 to −2.5 V and −2.5 to 1.5 V with the step of 10 mV for 5 times. As we can see from the figure, there is almost no hysteresis (<10 mV) detectable during the five sweeps (only the first sweep is shown in

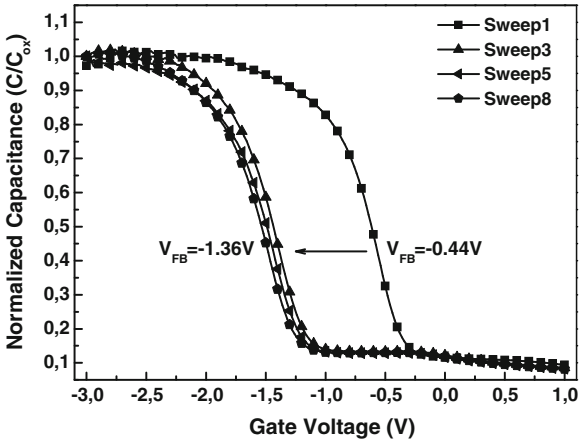


Fig. 11.14 Normalized Capacitance-Voltage (*C–V*) characteristics of single crystalline Gd_2O_3 on Si(100) substrate with W as top electrode. The flat-band voltage reduced from −0.44 to −1.36 V with the number of sweeps

Fig. 11.14 for clarity). Based on above results, the conventional forming gas annealing is compatible to novel single crystalline $W/Gd_2O_3/Si(100)$ structure.

The leakage current of $W/Gd_2O_3/Si(100)$ (with CET = 2.2 nm) is only $2 \cdot 10^{-6}$ A/cm$_2$ at $V_{FB} - 1$ V, four orders of magnitude lower than the as-grown samples, which is also much lower than that of SiO_2 at the same CET.

To evaluate D_{it} we measured the equivalent parallel conductance (G_p) of MOS capacitor as a function of gate bias and frequency (the details of the calculation method have been described elsewhere [21]). The conductance in the present case was measured at flat-band voltage condition. D_{it} was calculated to be only $8.8 \cdot 10^{11}/cm^2$ eV after the forming gas anneal.

In summary, the as-grown single crystalline Gd_2O_3 thin film on Si(100) substrate suffers from flat-band voltage instability and large hysteresis which are possibly due to the intrinsic dangling bonds induced by the existing bonding mismatch at the oxide/silicon interface. The instability of flat-band voltage and hysteresis can be fully eliminated by the introduction of traditional forming gas annealing.

In the following, we will discuss changes in structural and dielectric properties of uncapped and capped crystalline Gd_2O_3 layers on Si as the result of different rapid thermal annealing (RTA) process conditions commonly used in MOSFET fabrication [38]. Layers of different thicknesses were grown on p-type Si(100) and Si(111) substrates. On Si(100) substrate, the grown layers exhibit the known two types of (110)-oriented domains, with two orthogonal in-plane orientations. On (111) oriented Si surface, *bixbyite* Gd_2O_3 can grow single crystalline. After growth, the wafers were cut into smaller pieces. Each piece was first structurally characterized in the as-grown state using x-ray techniques. All variations across a wafer were found to be within the error ranges of the used techniques, thus, we have a sufficiently high homogeneity. Then, the uncapped pieces were annealed in different ambients for 30 s in a commercial RTA system. For each annealing batch, we used pieces from the same wafer. The layer thickness was measured by x-ray reflectivity (XRR). To obtain the best fit, we used a three layer model, where

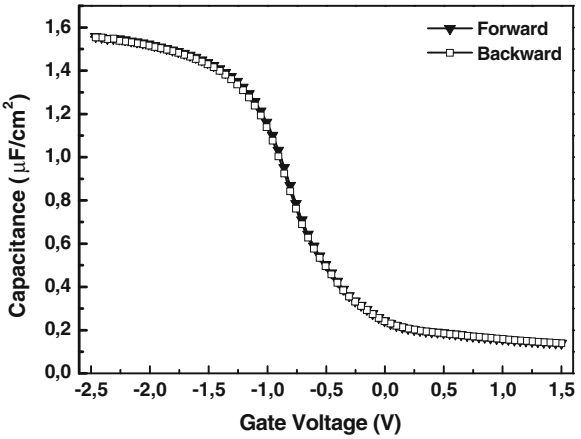


Fig. 11.15 Hysteresis characteristics of single crystalline Gd_2O_3 on Si(100) substrate after forming gas anneal at 450 °C for 10 min with W as top electrode. V_{FB} is stabilized at −0.44 V. Only the first sweep is shown for clarity

the first layer describes the interfacial transition between substrate and the gadolinium oxide layer, the second the high-k layer, and the third considers possible reactions of the oxide with atmospheric H_2O and CO_2 [39]. We always used the overall thickness of all three layers for evaluating annealing effects on layer thicknesses. Figure 11.16 summarizes the results for annealing the uncapped films in pure nitrogen. Up to 800 °C, we do not observe any changes in the overall layer thickness. At higher temperature, the layer thickness increases with annealing temperature. Below 1,000 °C RTA, substrate orientation has no effect on the observed thickness changes.

Layers were also characterized by X-ray diffraction (XRD) and transmission electron microscopy (TEM). Figure 11.17a shows XRD $\theta/2\theta$ scans for a 8 nm thick layer grown on Si(100) and annealed at different temperatures in nitrogen. Besides the substrate peaks (not shown), the only detectable peak around 47.4° can be assigned to the (440) reflection of Gd_2O_3 in the *bixbyite* structure. Up to 900 °C, we observe only minor changes in that peak caused most likely by relaxation effects.

The corresponding cross-sectional TEM images (not shown) reveal the following: In contrast to the as-grown crystalline Gd_2O_3 film, this layer shows an amorphous region between the silicon and the crystalline Gd_2O_3 after RTA at 900 °C. That structureless transition region consists of two layers with different contrast in TEM; the lower one can be attributed to a silicon oxide-like phase, whereas the upper one is most likely a silicate-like phase. For even higher annealing temperatures, the crystalline structure of the layer disappears completely. Thus, the high-k layer will be completely structurally degraded after annealing at 1,000 °C.

The formation of silicon oxide-like and/or silicate-like transition regions requires the presence of additional oxygen. Gadolinium, however, can only exist in the +3 oxidation state [40]. Thus under equilibrium conditions, the Gd_2O_3 bulk layer cannot act as an effective source for oxygen supply to the Si/dielectric

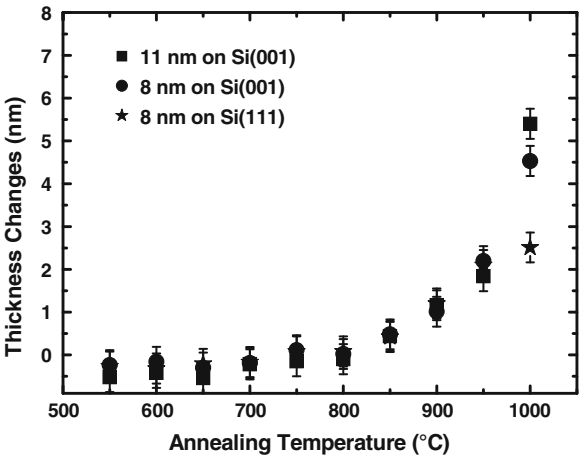


Fig. 11.16 Change in layer thickness (evaluated by XRR) for crystalline Gd_2O_3 layers grown on Si after an RTA in nitrogen for 30 s

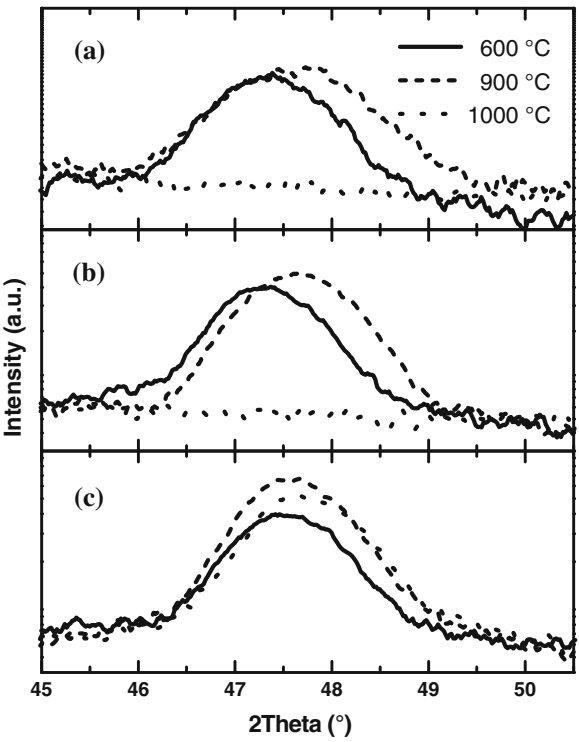


Fig. 11.17 XRD scans for Gd_2O_3 layers on Si(001) annealed at different temperatures for 30 s. in **a** nitrogen, **b** oxygen, and **c** nitrogen (with an a-Si capping layer)

interface. The observed changes can therefore only be attributed to oxygen supplied by the annealing ambient. To verify that hypothesis, we repeated the same annealing experiments in pure oxygen. In another set of experiments, we sealed the high-k layer by an amorphous, 100 nm thick Si layer in order to suppress oxygen supply from the ambient. For experiments with Si capping layer we used a four layer model for the XRR evaluation. Figure 11.18 summarizes the layer thickness variation for these two series.

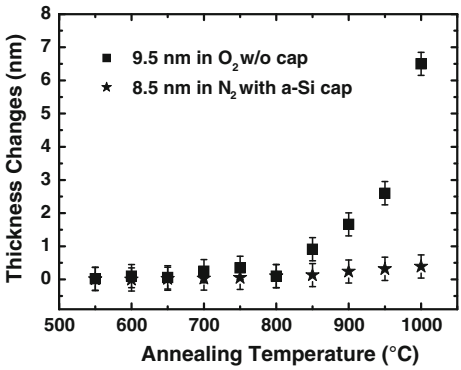


Fig. 11.18 Change in layer thickness (evaluated by XRR) for an open crystalline Gd_2O_3 layer grown on Si after a 30 s RTA in oxygen and for a layer capped with 100 nm a-Si and annealed in nitrogen

Similar to RTA in nitrogen, the layer thickness remains constant up to 800 °C also for the oxygen anneal. For higher temperature, it increases nearly exponentially with temperature. However, for the capped layers, we do not see any significant increase in layer thicknesses with temperature even up to 1,000 °C. Figure 11.16b and c show corresponding XRD measurements. Analogous to N_2 RTA, the (440) peak of the Gd_2O_3 layer disappears completely after the 1,000 °C anneal in oxygen. In contrast, the sealed high-k layer remains fully crystalline after such high temperature RTA. Here, an additional peak at 28.4° occurs (not shown in Fig 11.17c) which originates from the crystallization of the Si cap layer. All structural results support our hypothesis of oxygen supply from the annealing environment. Avoiding such oxygen in-diffusion leads to a significant stabilization of the grown layers.

For the electrical characterization of the high-k films, we prepared fully silicided (FUSI) NiSi metal electrodes on the annealed samples as described in detail elsewhere [41]. We extracted the capacitance equivalent thicknesses (CET) for the various layers from the accumulation capacitance measured at 100 kHz (Fig. 11.19). For the open layer, the CET increases as a function of annealing temperature. At the same time, the extracted effective dielectric constant remains constant within the error limits. Such behavior can be explained in two ways: Either the dielectric properties of the high-k layer degrade due to annealing or the impact of the interfacial transition increases. If a structure contains several dielectrics in series, the lowest capacitance layer will dominate the overall capacitance, and also set a limit on the minimum achievable CET value.

Electrical measurements of such a multilayer stack with the total thickness t_{total} will yield always an effective dielectric constant, k_{eff}. Detailed evaluations of the XRR measurements yield also the electronic density of the individual layers. Up to 800 °C anneals, that density of the high-k layer does not change. Let us assume that $k_{high\text{-}k}$ also does not change within that temperature range. Also from the XRR evaluation (and various cross-sectional TEM investigations, not shown here), the interfacial layer thickness remains constant around 1.5 nm up to 800 °C. The as-grown layers did not show any pronounced interfacial layer.

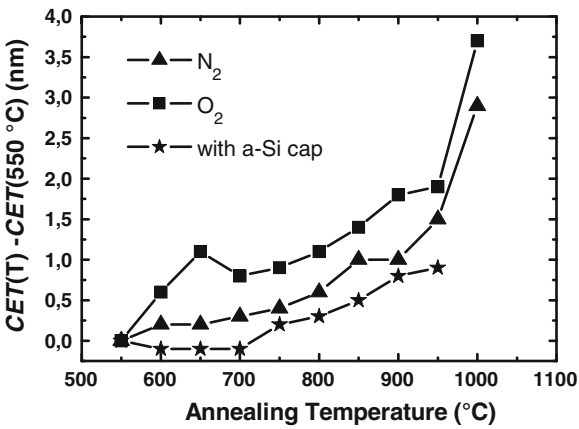


Fig. 11.19 Changes in the capacitance equivalent thickness (*CET*) extracted from C–V measurements after different annealing

Summarizing, the observed structural and electrical results up to 800 °C RTA can only be understood by a gradual transformation of parts of the crystalline oxide into a structure-less silicate-like phase with lower permittivity. The situation looks different for the sealed layers. Here the effective permittivity is nearly twice as high compared to the open systems annealed in nitrogen or oxygen, and CET increases only slightly due to annealing. The exclusion of additional oxygen retards the transformation of the Gd_2O_3 layer into a silicate layer.

11.9 Further Applications of Crystalline Lanthanide Oxides

There are several attempts to increase the material variety that is compatible with Si technologies. The ability to integrate crystalline dielectric barrier layers into silicon structures can open the way for a variety of novel applications ranging from high-k replacements in future MOS devices to oxide/silicon/oxide heterostructures for nano-electronic application in quantum-effect devices. Quantum-effect devices based on silicon are promising for future nanoelectronic application. Here, double-barrier structures using epitaxial insulators as barriers and Si as quantum-well are particularly interesting for resonant tunneling devices. The fabrication of such an epitaxial insulator/Si/insulator heterostructure requires the growth of ultra-thin, atomically flat and defect-free insulator barriers on Si, and the growth of epitaxial Si quantum-well layers on the insulator. However, this can not be achieved straightforward because of the differences in the surface free energy between the insulator and silicon. Often, the surface free energy of the insulator is much lower than that of silicon leading to Si cluster formation on the insulator. In our case, the growth of Si under common epitaxial conditions on Gd_2O_3 leads to island formation.

To obtain smooth layers, the growth process has to be modified. First, we tested low temperature deposition and surfactant-mediated epitaxy using boron, followed by a subsequent solid phase epitaxy (SPE). Our experiments clearly showed that Si does not wet the oxide layer and that Si island formation can not be prevented using conventional SPE. Also different modifications of surfactant-mediated epitaxy lead to Si islanding. We developed a new approach for nanostructure formation which is based on solid-phase epitaxy of the Si quantum-well combined with simultaneous vapor-phase epitaxy of the insulator on top of the quantum-well [42]. Ultra-thin single-crystalline Si buried in a single-crystalline insulator matrix with sharp interfaces was obtained by this approach on Si(111). Successful fabrication of single crystalline oxide/Si/oxide double-barrier nanostructure enable us to design Si-based resonant tunneling diodes (RTD), which might open up an exciting opportunity to integrate advanced quantum device functionalities into mainstream silicon based technology. Such devices utilizing epi-Gd_2O_3 as an insulator and Si as the quantum well offer many more functionalities in the future.

Apart from application of thin epitaxial layers, lanthanide oxides could also find another application in silicon nano-cluster (Si–NC) flash memories due to their thermodynamic stability on Si substrates at higher temperatures. Such nano-crystal memories represent one of the most promising candidates for future nonvolatile, high density, and low operating-voltage memory applications. Originally, the concept of Si-NC flash memories has been based upon embedding single crystalline Si clusters of few nanometers in size into the insulating gate oxide (SiO_2) of field effect transistor (FET), where Si–NCs could be used as deliberately generated trap centers for electrons and/or holes [43]. The entrapment of the charges (e.g. electrons) by these clusters embedded in the gate oxide eventually shifts the threshold voltage during the device operation by screening the gate charges and hence reduces conduction in the channel inversion layer. Electrical performance of these clusters strongly depends on their physical properties such as their size, density, and spatial distributions into the oxide as well. Thus, the most challenging task to improve the device performance has been the formation of nano-clusters with constant size, high density and uniform distribution. There are several approaches reported recently to form Si–NCs into SiO_2 and also into high-k oxides [44–46]. Replacing SiO_2 with high-k oxide in floating gate memory is advantageous since the larger capacitance enhances the drive current while high breakdown voltage and low leakage current could be maintained [27, 47] due to thicker physical thickness.

Recently, we demonstrated the controlled growth of Si-NCs incorporated into epitaxial rare earth oxide using MBE [48]. Exploiting the advantages of different thin film growth mechanisms, we were able to control the size and density of Si–NCs embedded in epitaxial Gd_2O_3 grown on Si substrates. The epitaxial rare earth oxide, if deposited by MBE, exhibits superior crystal quality and therefore allows better control over the scaling in future devices [49]. C–V characteristics of Pt/Gd_2O_3/Si MOS capacitors with embedded Si–NCs grown on different silicon substrates were measured at different frequencies. Significant shift of flat band voltage (ΔV_{FB}) due to a large number of electrons and holes captured by the Si–NCs could be observed for all cases. The positive shift in V_{FB} (for n-substrates) observed after charging of clusters indicates the trapping of electrons, which effectively screen the electric field at the gate and hence the measured flat band voltage is shifted. The reverse effect is observed for p-type substrates where the majority of trapped charges are holes. It is worth to remember here that the Gd_2O_3 layers with no Si-NCs display negligible hysteresis in C–V measurements. The density of the nano-cluster estimated from the flat band shift is about $4.2 \cdot 10^{12}$ cm^{-2} assuming one electron per cluster. This value turns out to be twice than that estimated from Atomic Force Microscopy (AFM) investigation, implying that the one-electron per cluster assumption does not fit to our devices. However, if we assume that the average number of electrons captured by one single cluster is two, it would result in a cluster density of around $2.1 \cdot 10^{12}$ cm^{-2}, in good agreement with the AFM results.

In Ref. [49], we showed in detail structural and electrical properties of high performance nonvolatile Si nano-cluster memories with epitaxial Gd_2O_3 as a

control and tunneling layer. Clusters with average size of 5 nm and density of $2 \cdot 10^{12}$ cm^{-2} exhibit excellent charge storage capacity with competent retention ($\sim 10^5$ s) and endurance (10^5 write/erase cycles) characteristics. The Pt/Gd_2O_3/Si MOS capacitors comprised with Si nano-clusters displays large hysteresis ($\sim$1.5–2 V) in capacitance–voltage measurements. These Si nano-clusters are bonded via Gd–O–Si bonds to the metal oxide. Thus, Si nano-clusters embedded in an epitaxial rare earth oxide could be a potential contender for future non-volatile memory devices.

So far we have demonstrated the major areas where rare earth oxide thin films could offer optimistic solution for future devices with ultra-scaled dimension. However, there are many more areas where these oxide thin films could find potential application too. Here, we will briefly overview some of the devices, which also could be realized with crystalline lanthanide oxide thin films: (a) for spintronics: The presence of localized strongly correlated 4f electrons is responsible for the magnetic behavior. Thus, rare earth materials can be considered as candidates for future spin based electronic devices [50–52]. The strong exchange coupling between the localized rare earth 4f spins and valence and conduction electrons leads to interesting magnetic properties [53, 54]. However, such coupling may often be destroyed by the presence of any impurity inside the crystal. Recent demonstration of single crystalline thin films grown by molecular beam epitaxy on semiconductor substrates may pave the way to build the future devices utilizing rare earth compounds. Combining the advantage of large dielectric constant and high magnetic moment of rare earth materials one could able to design a system, which might exhibits high magnetic moment in coexistence with large dielectric constant. (b) For optical application: Materials combining selective emission with thermo-structural properties such as a good resistance to thermal shocks or to large temperature cycles are very interesting for a variety of different applications ranging from aerospace to energy conversion [55]. Among the most effective materials with selective emission properties at high temperature (1,000–2,000 °C), rare earth oxides are certainly the most interesting ones, thanks to their high melting temperature and to peculiar electronic properties of rare earth elements that have particularly empty 4f levels inside the 5p and 5s shells and the 6s valence states and therefore retain some atomic like optical properties in a variety of different bonding environments. As rare oxide materials emit thermal radiation in narrow spectral region, they can be used for a variety of different high temperature application such as generation of electricity by thermo photovoltaic conversion of thermal radiation.

11.10 Summary and Outlook

We described the use of molecular beam epitaxy to grow epitaxially lanthanide oxide thin films on silicon. We presented results for crystalline gadolinium oxide in the cubic *bixbyite* structure. On Si(100) oriented surfaces, crystalline Gd_2O_3 grows

as (110)-oriented domains, with two orthogonal in-plane orientations. We obtain perfect epitaxial growth of cubic Gd_2O_3 on Si(111) substrates. Layers grown under best vacuum conditions often exhibit very high leakage currents. Deliberate oxygen supply during growth improves the dielectric properties significantly; however, too high oxygen partial pressures lead to an increase in the lower permittivity interfacial layer thickness, and finally limit the achievable minimal equivalent oxide thickness. Experimental results for Gd_2O_3-based MOS capacitors grown under optimized conditions show that these layers are excellent candidates for application as very thin high-k materials replacing SiO_2 in future MOS devices.

We also reported the impact of interface layer composition on the electrical properties of epitaxial Gd_2O_3 thin films on Si(100) substrates. The electrical properties of epitaxial Gd_2O_3 thin films were improved significantly by controlled modification of the interface layer composition. The minimum capacitance equivalent thickness estimated for $Pt/Gd_2O_3/Si$ MOS structures was as low as 0.76 nm with leakage current density of 15 mA/cm^2 at $(V_g - V_{FB}) = 1$ V.

We investigated the effect of post-growth annealings on the layer properties. We showed that a standard forming gas anneal can eliminate flat-band voltage instabilities and hysteresis as well as reduce leakage currents by saturating the dangling bonds caused by the bonding mismatch. In addition, we investigated the impact of rapid thermal anneals on structural and electrical properties of crystalline Gd_2O_3 layers grown on Si with different orientations. The degradation of layers can be significantly reduced by sealing the layer with a-Si prior to annealing. For the capped layers, the effective capacitance equivalent thickness increases only slightly even after a 1,000 °C anneal.

Finally, we demonstrated that the epitaxial rare earth lanthanide layers can be successfully used to realize several novel devices. Double-barrier insulator/Si/insulator nanostructures on Si(111) have been grown using a new approach which is based on solid-phase epitaxy of the Si quantum-well combined with simultaneous vapor-phase epitaxy of the insulator on top of the quantum-well. Other interesting structures are formed by single-crystalline high-k oxide layers with embedded Si nano-clusters. They present one of the most promising candidates for future nonvolatile, high density, and low operating-voltage memory applications.

Acknowledgments This chapter summarizes the work we have been doing over the last years. I am in particular grateful to A. Laha, A. Fissel, E. Bugiel, M. Czernohorsky, D. Schwendt, R. Dargis, R. Ranjith, Q.Q. Sun, D. Tetzlaff, D. Kühne, G. Glowatzki, and T. Wietler for their various contributions. I am also grateful to our partners all over the world for their support and collaboration. Part of this work was supported by the German Federal Ministry of Education and Research (BMBF) under the KrisMOS and the MegaEpos projects.

References

1. J. Robertson, Rep. Prog. Phys. **69**, 327 (2006)
2. K.J. Hubbard, D.G. Schlom, J. Mater. Res. **11**, 2757 (1996)
3. D.P. Norton, Mat. Sci & Engineer. R **43**, 139 (2004)

4. G.-Y. Adachi, N. Imanaka, Chem. Rev. **98**, 1479 (1998)
5. H.J. Osten, M. Czernohorsky, R. Dargis, A. Laha, D. Kühne, E. Bugiel, A. Fissel, Microelectron. Eng. **84**, 2222 (2007)
6. G.V. Samsonov (ed.), *The Oxide Handbook*, 2nd edn. (IFI/Plenum, New York 1982)
7. K.A. Gschneidner, Jr., L. Eyring, G.H. Lander and G.R. Choppin (eds). *Handbook on the Physics and Chemistry of Rare Earths*, Lanthanides/Actinides: Chemistry, vol. 18, (Elsevier Science B.V, 1994)
8. J. Robertson, K. Xiong, Top. Appl. Phys **106**, 313 (2007)
9. M. Foëx, J.P. Traverse, Rev. Int. Hautes Temp. Refract **3**, 429 (1966)
10. M. Badylevich, S. Shamuilia, V.V. Afanas'ev, A. Stesmans, A. Laha, H.J. Osten, A. Fissel, Appl. Phys. Lett. **90**, 252101 (2007)
11. M. Nolan, S. Grigoleit, D.C. Sayle, St.C. Parker, G.W. Watson, Surf. Sci. **576**, 217 (2005)
12. V. Mikhelashvili, G. Eisenstein, F. Edelmann, J. Appl. Phys. **90**, 5447 (2001)
13. V.A. Rozhkov, A.Y. Trusova, I.G. Berezhnoy, Thin Solid Films **325**, 151 (1998)
14. P. Delugas, V. Fiorentini, Microelectron. Reliab. **45**, 831 (2005)
15. G. Seguini, E. Bonera, S. Spiga, G. Scarel, M. Fanciulli, Appl. Phys. Lett. **85**, 5316 (2004)
16. W. Cai, S.E. Stone, J.P. Pelz, L.F. Edge, D.G. Schlom, Appl. Phys. Lett. **91**, 042901 (2007)
17. J. Kwo, M. Hong, A.R. Kortan, K.L. Queeny, Y.J. Chabal, R.L. Opila, D.A. Müller, S.N.G. Chu, J. Appl. Phys. **89**, 3920 (2001)
18. A. Fissel, H.J. Osten, E. Bugiel, J. Vac. Sci. Technol., B **21**, 1765 (2003)
19. H.J. Osten, E. Bugiel, M. Czernohorsky, Z. Elassar, O. Kirfel, A. Fissel, Top. Appl. Phys **106**, 101 (2007)
20. A. Laha, H.J. Osten, A. Fissel, Appl. Phys. Lett. **89**, 143514 (2006)
21. E.H. Nicollian, J.R. Brews, *MOS Physics and Technology* (Wiley, New York, 1982)
22. A. Fissel, Z. Elassar, E. Bugiel, M. Czernohorsky, O. Kirfel, H.J. Osten, J. Appl. Phys. **99**, 074105 (2006)
23. D. Schmeisser, J. Dabrowski, H.-J. Muessig, Mater. Sci. Engin. B **109**, 30 (2004)
24. M. Czernohorsky, A. Fissel, E. Bugiel, O. Kirfel, H.J. Osten, Appl. Phys. Lett. **88**, 152905 (2006)
25. T. Echtermeyer, H.D.B. Gottlob, T. Wahlbrink, T. Mollenhauer, M. Schmidt, J.K. Efavi, M.C. Lemme, H. Kurz, Solid-State Electron. **51**, 617 (2007)
26. R. Endres, Y. Stefanov, U. Schwalke, Microelectron. Reliab. **47**, 528 (2007)
27. A. Laha, A. Fissel, E. Bugiel, H.J. Osten, Appl. Phys. Lett. **88**, 172107 (2006)
28. J. Kwo, M. Hong, A.R. Kortan, K.T. Queeney, Y.J. Chabal, J.P. Mannaerts, T. Boone, J.J. Krajewski, A.M. Sergent, J.M. Rosamilia, Appl. Phys. Lett. **77**, 130 (2000)
29. H. Kroemer, in *Heteroepitaxy on Si, MRS Symposia Proceedings No. 67*, ed. by J.C.C. Fan, J.M. Poate (Materials Research Society, Pittsburgh, 1986), and references therein
30. A. Laha, E. Bugiel, J.X. Wang, Q.Q. Sun, A. Fissel, H.J. Osten, Appl. Phys. Lett. **93**, 182907 (2008)
31. H.J. Osten, J.P. Liu, E. Bugiel, H.J. Müssig, P. Zaumseil, Mat. Sci. & Engin. B **87**, 297 (2001)
32. A. Laha, A. Fissel, H.J. Osten, Appl. Phys. Lett. **90**, 113508 (2007)
33. A. Fissel, J. Dabrowski, H.J. Osten, J. Appl. Phys. **91**, 8986 (2002)
34. C.C. Fulton, G. Lucovsky, R.J. Nemanich, J. Appl. Phys. **99**, 063708 (2006)
35. A. Laha, A. Fissel, H.J. Osten, Appl. Phys. Lett. **96**, 072903 (2010)
36. A. Laha, B. Ai, P.R.P. Babu, A. Fissel, H.J. Osten, Appl. Phys. Lett. **99**, 152902 (2011)
37. Q.-Q. Sun, A. Laha, S.-J. Ding, D.W. Zhang, H.J. Osten, A. Fissel, Appl. Phys. Lett. **92**, 152908 (2008)
38. M. Czernohorsky, D. Tetzlaff, E. Bugiel, R. Dargis, H.J. Osten, H.D.B. Gottlob, M. Schmidt, M.C. Lemme, H. Kurz, Semicond. Sci. & Technol **23**, 035010 (2008)
39. G. Adachi, N. Imanaka, Z.C. Kang, *Binary rare Earth Oxides* (Kluwer Academic Publishers, Dordrecht, 2004)
40. R. Lo Nigro, V. Raineri, C. Bondiorno, R. Toro, G. Malandrino, I.L. Fragala, Appl. Phys. Lett. **83**, 129 (2003)

41. H.D.B. Gottlob, T. Echtermeyer, M. Schmidt, T. Mollenhauer, J.K. Efavi, T. Wahlbrink, M.C. Lemme, M. Czernohorsky, E. Bugiel, A. Fissel, H.J. Osten, H. Kurz, IEEE Electron Device Lett. **27**, 814 (2006)
42. A. Fissel, D. Kühne, E. Bugiel, H.J. Osten, Appl. Phys. Lett. **88**, 153105 (2006)
43. S. Tiwary, F. Rana, H. Hanafi, A. Hartsten, E. F. Crabbe, and K. Chan, Appl. Phys. Lett. **68**, 1377 (1996)
44. K. Ichikawa, P. Punchaipetch, H. Yano, T. Hatayama, Y. Uraoka, T. Fuyuki, E. Takahashi, T. Hayashi, K. Ogata, Jpn. J. Appl. Phys. **44**, L836 (2005)
45. C. Dumas, J. Grisolia, L. Ressier, A. Arbouet, V. Paillard, G. Ben Assayag, A. Claverie, M. A. van den Boogaart, and J. Brugger, phys. stat. sol. (a) **204**, 487 (2007)
46. O.M. Nayfeh, D.A. Antoniadis, K. Mantey, M.H. Nayfeh, Appl. Phys. Lett. **90**, 153105 (2007)
47. P. Punchaipetch, Y. Uraoka, T. Fuyuki, A. Tomyo, E. Takahashi, T. Hayashi, A. Sano, S. Horii, Appl. Phys. Lett. **89**, 093502 (2006)
48. A. Fissel, D. Kühne, E. Bugiel, H.J. Osten, J. Vac. Sci. Technol., B **24**, 2041 (2006)
49. A. Laha, D. Kühne, E. Bugiel, A. Fissel, and H. J. Osten, Semicond. Sci. Technol. **23,** 085015 (2008)
50. S.D. Sarma, Am. Sci. **89**, 516 (2001)
51. S.A. Wolf, D. Treger, IEEE Trans. Magnetics **36**, 2748 (2000)
52. F. Mireles, G. Kirczenow, Europhys. Lett. **59**, 107 (2002)
53. K.A. Gschineidner, L.Eyring (eds), *Handbook on Physics and Chemistry of Rare Earth* (Amsterdam, North-Holland, 1979)
54. P.A. Dowben, D.N. McIlroy, D. Li, *Handbook on Physics and Chemistry of Rare Earth*, vol. 24, ed. by K. A. Gschineidner, and L. Eyring (Amsterdam, North-Holland, 1997)
55. G. Torsello, M. Lomascolo, A. Licciulli, D. Diso, S. Tundo, M. Mazzer, Nat. Mater. **3**, 632 (2004)

Chapter 12
High Mobility Channels

Michel Houssa, Peide Ye and Marc Heyns

Abstract The need for high-κ gate dielectrics and metal gates for advanced integrated circuits has re-opened the door to germanium and III–V compounds as potential replacements for silicon channels, offering the possibility to further increase the performances of future CMOS generations, as well as adding new circuit functionalities. In this chapter, we discuss the most important issues related to high-mobility channels, and highlight recent advances in the field, focusing on promising approaches for the passivation of their interface with high-κ gate dielectrics, as required for the fabrication of high-performance inversion mode metal-oxide-semiconductor field effect transistors.

12.1 Introduction

As the scaling of advanced metal-oxide-semiconductor field effect transistors (MOSFETs) is approaching its technological and fundamental limits, new materials are needed to further improve the performances of integrated circuits (IC). As

M. Houssa (✉)
Semiconductor Physics Laboratory, Department of Physics and Astronomy,
Katholieke Universiteit Leuven, Celestijnenlaan 200D B-3001 Leuven, Belgium
e-mail: michel.houssa@fys.kuleuven.be

P. Ye
School of Electrical and Computer Engineering and Birck Nanotechnology Center,
Purdue University, West Lafayette, IN 47907, USA
e-mail: yep@purdue.edu

M. Heyns
IMEC, Kapeldreef 75 B-3001 Leuven, Belgium
e-mail: heyns@imec.be

M. Heyns
Department of Metallurgy and Materials Engineering, Katholieke Universiteit Leuven,
B-3001 Leuven, Belgium

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5_12,

an example, a lot of research efforts have focused recently on the replacement of ultra-thin SiO_2 or SiON gate insulators by high-κ gate dielectrics [1–4]. The use of physically thicker high-κ materials, with equivalent gate capacitance lower than 1 nm thick SiON layer, result in a large decrease of the gate leakage current flowing through the device, allowing for drastic reduction of the static power consumption in the circuit. In addition, high-κ dielectrics are currently integrated in advanced devices in combination with metal gates, instead of the "classical" poly-Si gates [4].

Other materials of much interest for high performance IC applications are high-mobility semiconductors, such as germanium and III–V compounds, which are currently investigated for the replacement of the silicon channel [5–7], in order to further enhance the drive current of the devices. Though germanium presents both increased electron and hole mobility over Si (see Table 12.1), functioning Ge nFETs still needs to be demonstrated. For this reason, III–V compounds are considered as promising materials for nFETs, due to their much higher electron mobility over Si, as indicated in Table 12.1. Consequently, future Complimentary MOS (CMOS) technology is envisioned as co-integrating III–V compounds (n-channel) and Ge (p-channel) on Si substrates [8], as schematically presented in Fig. 12.1. In addition, most III–V compounds are direct bang-gap semiconductors with useful optoelectronic properties, their integration with the Si technology potentially enabling to add new circuit functionalities (more than Moore concept).

Table 12.1 Energy band-gap and carrier mobility of Si, Ge, and few III–V compounds

Semiconductor	Energy band-gap at 300 K (eV)	Electron mobility (cm^2/V s)	Hole mobility (cm^2/V s)
Si	1.12 (indirect)	1,500	450
Ge	0.66 (indirect)	3,900	1,900
GaAs	1.42 (direct)	8,500	400
InAs	0.36 (direct)	33,000	460
InSb	0.17 (direct)	80,000	1,250
InP	1.35 (direct)	4,600	150

Appendix III contains a detailed list of semiconductors and their physical properties

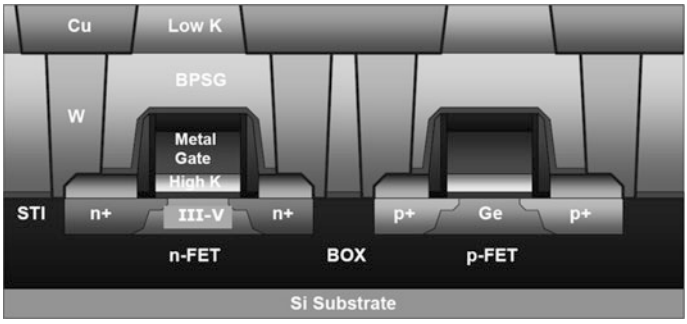


Fig. 12.1 Schematic illustration of a Ge/III–V CMOS technology

In this chapter, we highlight the most important challenges and progress recently achieved on high-mobility channels, in combination with high-κ/metal gate stacks. The chapter is organized as follows. In Sect. 12.2, the challenges facing high-mobility/high-κ gate stacks will be described. The most promising approaches for the passivation of the Ge and III–V interfaces will be reviewed in Sects. 12.3 and 12.4, respectively. A brief summary is given in Sect. 12.5, also highlighting the perspectives of high-mobility channels in future CMOS technologies.

12.2 Challenges Facing High-κ Dielectrics on High Mobility Substrates

From an historical point of view, the first (bipolar) transistor was built on bulk Ge in 1947 by Bardeen, Brattain and Shockley from Bell Laboratories. This scientific and technological breakthrough was awarded by a Nobel Prize in Physics in 1956. About 12 years later, the first integrated circuit was invented by Kilby, also using a Ge substrate, and for which he received the 2000 Nobel Prize in Physics. Concerning the III–V channels, the advantage of GaAs MOSFET over its Si counterpart was early recognized because the electron mobility in GaAs is more than five times higher than that in Si. The first GaAs MOSFET work using SiO_2 as gate dielectric was reported by Becke and White by the Radio Corporation of America in 1965 [9]. But it was quickly realized that SiO_2 was not the right gate dielectric for GaAs, which sparked an enormous research effort in the following decades with a search for a low-defect, thermodynamically stable gate dielectric for GaAs.

Si supplanted Ge and III–V materials for MOSFET applications, mainly thanks to the remarkable properties of silicon dioxide, used as a Si surface passivation layer, high quality gate insulator as well as field isolation between adjacent devices. Indeed, a major issue common to Ge and III–V compounds is the passivation of their interface, i.e., achieving a sufficiently low density of interface states, comparable to the reference Si/SiO_2 interface, which typically lies below 10^{10} cm^{-2}, after anneal in N_2/H_2 ambient at typically 400–450 °C; let us recollect that the natural density of dangling bonds at the (100)Si/SiO_2 interface, before N_2/H_2 post-deposition anneal, is typically 1×10^{12} cm^{-2} [10, 11]. Most Ge and III–V channels present a much larger density of interface states, lying in the 10^{13} cm^{-2} range [12, 13], which hampers the electrical properties of surface channel devices, like inversion-mode MOSFETs.

In the specific case of Ge, recent progress has been achieved regarding interface passivation, mainly using an ultrathin Si capping layer (few monolayers thick) [14, 15], a thermally grown GeO_2 or GeON interfacial layer [16–18], or using PH_3 [19] or H_2S [20] surface treatments. Few high-κ dielectrics, particularly rare-earth oxides, have also been shown to provide improved Ge surface passivation, as compared to HfO_2 or ZrO_2 [6, 21]. However, good device characteristics using these passivating approaches have so far been demonstrated on pMOSFET, most

of nMOSFET being either non or barely functional, with much lower electron mobility than expected.

As far as III–V channels are concerned, researchers at Bell Laboratories discovered in 1987 that a class of sulfides (e.g., Li_2S, $(NH_4)_2S$, $Na_2S{\cdot}9H_2O$) was able to passivate the GaAs surface and provided excellent electronic properties to GaAs surfaces [22, 23]; sulfur passivation is still used in today's research. Another interesting approach used a very thin amorphous or crystalline Si layer as an interfacial control layer between GaAs and SiO_2 or Si_3N_4 [24–27]. This approach was intensively studied in the 1990s [28–30] and was recently adopted to integrate high-κ HfO_2 gate dielectrics on GaAs [31–34]. More interesting work was reported recently using silane based PECVD Si-layer to passivate GaAs surface [35, 36].

In 1996, Passlack and Hong at Bell Laboratories reported that in situ deposition of $Ga_2O_3(Gd_2O_3)$ dielectric film on a GaAs surface by electron-beam evaporation from single-crystal $Ga_5Gd_3O_{12}$ produced MOS structures on GaAs with a low interface trap density (D_{it}) [37–39]. Because the experiment was realized in an ultra-high-vacuum (UHV) connected multi-chamber Molecular Beam Epitaxy (MBE) system, it was called MBE-grown $Ga_2O_3(Gd_2O_3)$. A series of $Ga_2O_3(Gd_2O_3)$ dielectric device work, mainly led by Hong, were done at Bell Laboratories [40–44], and continued at Agere Systems [45] and National Tsing Hua University in Taiwan [46]. In 2003, Passlack and co-workers at Motorola/Freescale modified the previous $Ga_2O_3(Gd_2O_3)$ dielectric process, using a Ga_2O_3 template in $Gd_xGa_{0.4-x}O_{0.6}/Ga_2O_3$ dielectric stacks on GaAs [47]. A series of accumulation-type enhancement-mode devices using high work-function metal gates were reported since 2006 [48, 49]. Currently, the research is focused on the conventional atomic layer deposition (ALD) high-κ dielectric integration on III–V compound semiconductors [50–54].

12.3 Passivation of Ge for p-MOSFETs

Various high-κ dielectrics have been studied as possible gate insulators for Ge channel MOSFETs [55–61]. HfO_2 and ZrO_2 have received particular attention, mainly because these materials are serious candidates for the replacement of SiO_2 in Si-based MOSFETs. To illustrate the Ge surface passivation issue, let us first consider the case of HfO_2 deposited on an HF-last treated Ge surface, leaving one to two monolayers of GeO_x at the interface [62]. A high-resolution transmission electron microscope (TEM) image of a 9 nm HfO_2 layer deposited by atomic layer deposition (ALD) on such a HF-treated Ge interface is shown in Fig. 12.2. One observes the presence of an ultra-thin GeO_x interfacial layer, typically less than 3 Å. Note that this interfacial layer is much thinner than the typical SiO_2 interfacial layer formed during high-temperature processing of Si devices, which is typically about 1 nm. From a scaling point of view, the Ge/HfO_2 system thus presents a potential advantage over its Si/HfO_2 counterpart.

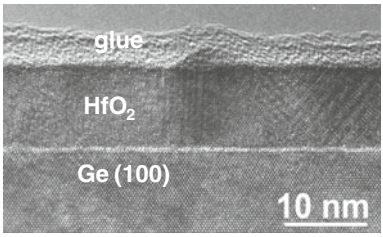


Fig. 12.2 High resolution cross-sectional transmission electron microscopic image of a $Ge/GeO_x/HfO_2$ gate stack

Unfortunately, the electrical properties of such a gate stack are rather poor, as illustrated in Fig. 12.3. The high-frequency capacitance–voltage (C–V) characteristics of the structure do not present the typical transition (decrease in capacitance) from accumulation to depletion observed in ideal MOS capacitors. These flat C–V characteristics rather suggest that the Fermi level at the Ge surface is pinned at a fixed value, due to the presence of a very high density of interface states (in the 10^{13} to 10^{14} cm^{-2} range). The leakage current flowing through the gate dielectric is also very high, and current–voltage characteristics are not reproducible from one device to the other. These poor electrical results are possibly caused by the formation of Ge–Hf bonds at the interface, as recently predicted from ab initio simulations [63, 64] as well as by the diffusion of Ge across the HfO_2 layer [62], either during the deposition of the HfO_2 layer or after post-deposition anneals.

Similar results were reported for HfO_2 layers deposited by Metal Organic Chemical Vapor Deposition (MOCVD) [62] on an HF-last cleaned Ge surface, or by atomic beam deposition on a clean (2×1) reconstructed Ge surface [65]. Proper passivation of the Ge surface is thus required before the deposition of the high-κ gate dielectric layer.

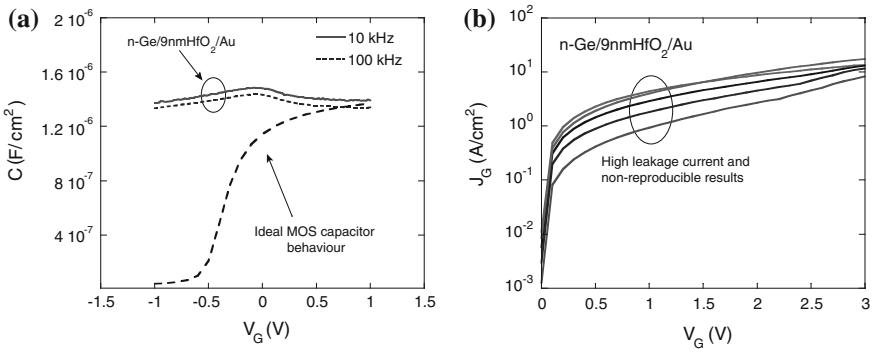


Fig. 12.3 High frequency C–V (**a**), and I–V (**b**) characteristics of Ge/HfO_2-based capacitors

12.3.1 Epi-Si Layer Passivation

One of the earliest promising Ge surface passivation approaches consists in depositing an ultra-thin Si epitaxial interlayer, followed by its (partial) chemical or thermal oxidation [14, 15]. The advantage of this approach is that it converts the issue of Ge surface passivation to the much better understood Si passivation.

Typically, the Ge surface is first etched in an HF solution, to remove most of the native GeO_x layer. The wafer is then transferred to an epitaxial reactor, where it is baked in H_2 at a typical temperature of 600–650 °C, to obtain a clean Ge surface. An ultra-thin Si epitaxial layer is next deposited on the Ge surface, using SiH_4 at a typical temperature of 500 °C. This interlayer is next partly oxidized in, e.g., an O_3-based solution, to form an ultra-thin SiO_2 interlayer, typically 0.5 to 1.0 nm thick.

A high resolution cross-sectional TEM picture of a $Ge/Si/SiO_x/HfO_2$ stack is shown in Fig. 12.4. The HfO_2 layer was deposited by ALD at 300 °C, using $HfCl_4$ and H_2O as precursors. A well defined gate stack, with sharp and smooth interfaces is observed from the TEM picture.

As shown in Fig. 12.5a, the C–V characteristics of such a $Ge/Si/SiO_2/HfO_2$ gate stack are much improved compared to the Ge/HfO_2 gate stack discussed above. Both the high-frequency gate-to-channel capacitance and gate-to-bulk capacitance of a typical Si-passivated Ge-pMOSFET are well behaved, suggesting a much improved Ge surface passivation. Consistently, the average density of interface states, estimated from charge-pumping current measurements, is about 1.5×10^{11} cm^{-2} near mid-gap, see Fig. 12.5b.

Interestingly, the Equivalent Oxide Thickness (EOT) extracted from C–V measurements depends on the gate bias polarity, as shown in Fig. 12.6a [66]. In addition, the EOT extracted under positive gate bias does not depend on the number of Si monolayers remaining after the partial oxidation of the epi-Si layer (estimated from total-X ray fluorescence measurements), while the EOT extracted under negative gate bias does. These results can be explained using the schematic band diagram shown in Fig. 12.6b. Due to the valence band offset at the Ge/Si interface [67], holes attracted at the surface under negative gate bias are confined in the Ge layer. The strained undoped epi-Si layer then acts as an additional

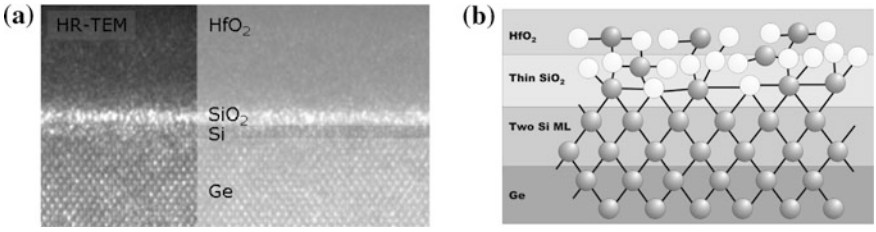


Fig. 12.4 **a** High resolution cross-sectional TEM picture of a gate stack formed by a Ge substrate, an ultra-thin epi-Si layer, partly oxidized into SiO_x, and a HfO_2 layer deposited by ALD. **b** Schematic *ball* and *stick* drawing of the $Ge/Si/SiO_x/HfO_2$ gate stack

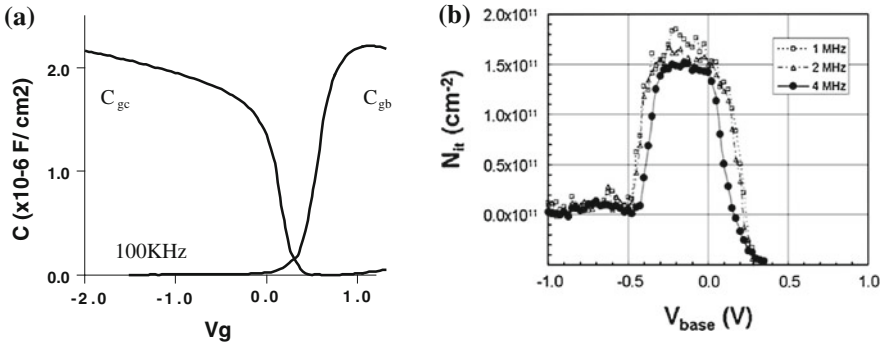


Fig. 12.5 **a** Gate-to-channel capacitance (C_{gc}) and gate-to-bulk capacitance (C_{gb}) versus gate voltage of a Si-passivated Ge-pMOSFET, measured at 100 kHz. **b** Density of interface states, extracted from charge pumping measurements, as a function of base-level sweep of a similar device

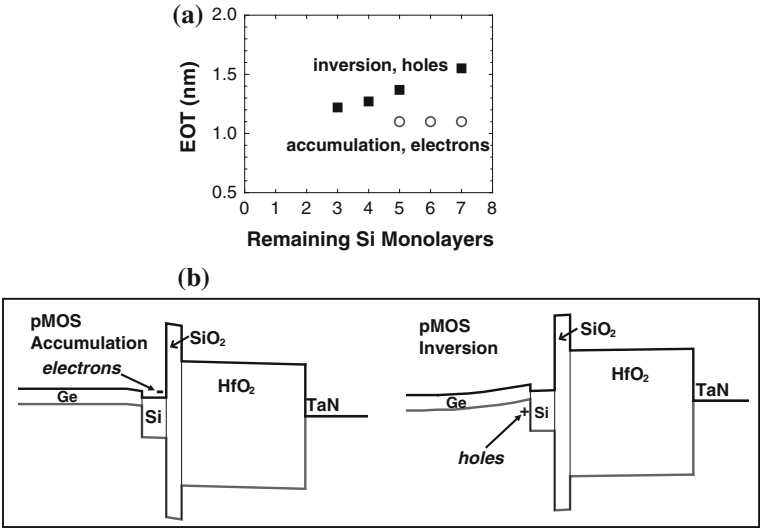


Fig. 12.6 **a** EOT as a function of the number of remaining Si monolayers of Ge/Si/SiO_2/HfO_2/TaN pMOSFET, under *positive gate* bias (*accumulation of electrons*) and under *negative gate* bias (*accumulation of holes*). **b** Schematic energy band structure of the device under both polarities

capacitor in series with the gate stack capacitance, increasing the total EOT of the system and explaining the almost linear dependence of EOT on the number of Si monolayers; the relative dielectric constant of the strained Si layer can be estimated from the slope of the linear fit to the data, yielding a reasonable value of about 9. On the other hand, the conduction band offset at the Ge/Si interface is such that electrons are attracted at the Si/SiO_2 interface, the EOT being then

determined by the SiO_2/HfO_2 gate stack, and is independent of the number of Si layers, as observed in Fig. 12.6a.

The much improved passivation of the Ge surface achieved by the epi-Si layer deposition allowed fabrication of functioning pMOSFET with promising performances [15, 68]. The typical drain current-gate voltage characteristics and transconductance (g_m) of a Si-passivated pMOSFET are presented in Fig. 12.7, before and after a post-metallization anneal in H_2 at 400 °C. The slight improvement in peak g_m and sub-threshold slope after the H_2 anneal are likely due to the passivation of dangling bonds at the Si/SiO_2 interface. The ratio between the on and off-current of the device is about 10^5, which is a very good result for Ge-based devices, taking into account the small band gap of Ge. In addition, the extracted peak hole mobility (using the split C–V technique) for this device is about 350 cm^2/V s, i.e., almost 3 times the universal hole mobility of Si devices, which is among the best values reported so far for Ge-based MOSFETs.

The performances of these devices also depend much on the number of Si monolayers [68], as clearly observed from Fig. 12.8. The drive current of a 1 μm channel length device is indeed found to present an optimal value at around 5 remaining Si monolayers (after partial oxidation), increasing almost by a factor two from one monolayer to five Si monolayers. These results can be attributed to the interplay between the observed increase in hole mobility with the number of Si monolayers (Fig. 12.8b)—which tend to increase the drive current—and the increase of EOT with the number of Si monolayers (Fig. 12.6a)—which tend to decrease the drive current, resulting in an optimal value for On current (I_{on}) at about 5 Si monolayers. Note that the increase in hole mobility with the number of Si monolayers is related to the decrease in sub-threshold slope, shown in Fig. 12.8b, hence to the decrease of the effective number of interface states with the number of Si monolayers. This latter result can be explained as follows. Assuming that most of the defects involved are located at or close to the Si/SiO_2 interface, these defects are located further away from the Ge channel as the

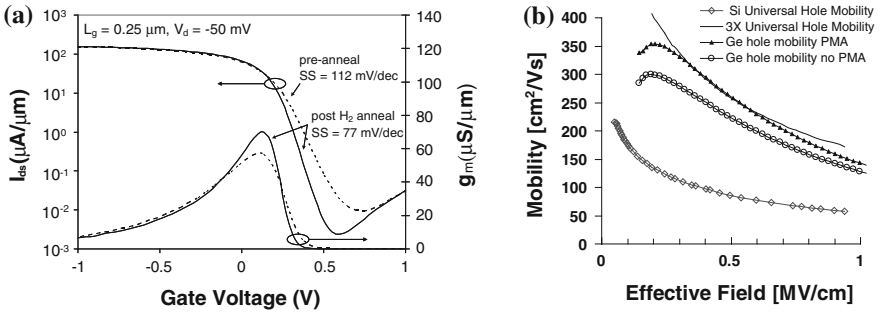


Fig. 12.7 **a** Drain current I_{ds} and transconductance g_m of a $Ge/Si/SiO_x/HfO_2$ pMOSFET as a function of the gate voltage, before and after a post-metallization anneal in H_2. **b** Hole mobility of the same devices, as a function of the effective electric field, extracted from split C–V measurements. The universal Si hole mobility is shown for comparison

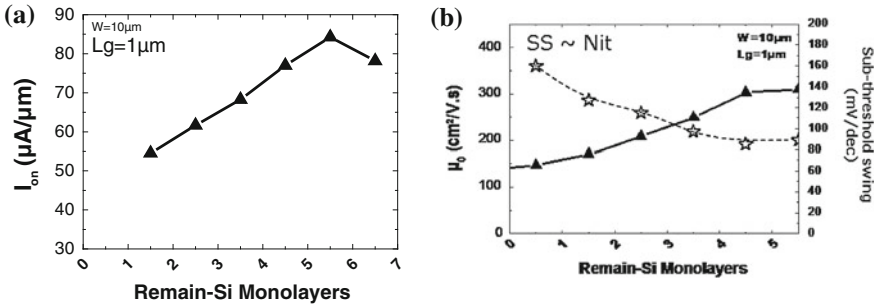


Fig. 12.8 **a** Drive current as a function of the number of remaining Si monolayers of epi-Si passivated Ge-pMOSFETs. **b** Peak hole mobility and sub-threshold slopes of the same devices

number of Si monolayer increases, leading to an enhanced mobility and reduced sub-threshold slope.

The threshold voltage (V_{th}) of these devices also depends much on the number of Si monolayers, as shown in Fig. 12.9. Between one and three monolayers, V_{th} of pMOSFETs is positive, i.e., the device is not turned off at zero gate bias. Part of this positive V_{th} shift could arise from the formation of an outward dipole at the Ge/Si interface, as predicted recently from ab initio simulations [69]. The strong dependence of V_{th} on the number of Si monolayers further suggests that defects are possibly responsible for this trend; it has been shown that Ge diffuses through the Si layer during deposition and subsequent processing [70], possibly producing defects that could lead to the shift of V_{th} from its ideal value (~ -0.3 V). Interestingly, this ideal value is observed at 5–6 Si monolayers, which also corresponds to the optimal value for the drive current of these devices. However, a strong variation of V_{th} with the number of Si monolayers requires very precise control of the thickness of this layer over large areas, in order to avoid important variations in the device characteristics over a chip and from chip to chip.

One should also notice that up to now, only high performance pMOSFETs have been realized using the epi-Si passivation approach, nMOSFETs being either non functional or showing very low electron mobility, typically 10–30 cm^2/V s. Another issue concerns aggressive EOT scaling; achieving an EOT below typically 1 nm appears to be rather difficult with this approach.

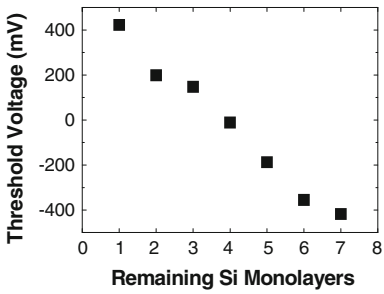


Fig. 12.9 Threshold voltage as a function of the number of remaining Si monolayers of epi-Si passivated Ge-pMOSFETs

12.3.2 Thermal Germanium Oxide as a Passivating Layer

The passivation of Ge by its thermal oxide (GeO_2) has been recently revisited. Though GeO_2 is hygroscopic and desorbs as volatile GeO at about 430 °C (in vacuum) [71], GeO_2 has been recently demonstrated to provide efficient passivation of the Ge surface [16–18]. Besides, capping the GeO_2 layer with a metal layer or with a high-κ gate dielectric, provides increased robustness to GeO desorption [72]. Thermal oxidation of Ge can be performed in atomic oxygen [73], dry O_2 [18] or in ozone. We will discuss here in more details results about thermal oxidation of Ge in dry O_2, at temperatures typically between 350 and 450 °C, followed by in situ Al_2O_3, ZrO_2 or HfO_2 deposition using ALD [18, 74].

The evolution of Ge oxide thickness, estimated from X-ray photoelectron spectroscopy (XPS) measurements, is shown in Fig. 12.10a as a function of the oxidation time at different temperatures. A linear growth regime is observed, suggesting that the reaction of Ge with O_2 is the rate-limiting step for the growth of very thin GeO_2 layers. One also observes the increase of the Ge oxidation growth rate with the temperature.

The chemical composition of the GeO_2 layer was studied by XPS (using Ge*3d* and Ge*3p* spectra). Figure 12.10b shows the fraction of Ge^{4+} oxidation state as a function of the Ge oxide thickness. It appears that the as-grown layer is mainly GeO_2 with the Ge^{4+} contribution increasing with the oxide thickness. Interestingly, deposition of a high-κ dielectric on the GeO_2 layer induces a reduction of Ge^{4+} density, e.g., the GeO_2 layer is partially reduced to lower (sub-4) oxidation states.

The elemental depth profiles of $Ge/GeO_2/ZrO_2$ and $Ge/GeO_2/Al_2O_3$ gate stacks are shown in Fig. 12.11. These profiles were reconstructed from angle-resolved XPS measurements, using the Ge*3p* and Ge*3d* intensities on the ZrO_2- and Al_2O_3-based gate stack, respectively. These profiles clearly reveal the intermixing between GeO_2 and ZrO_2, as opposed to a much sharper interface between GeO_2 and Al_2O_3.

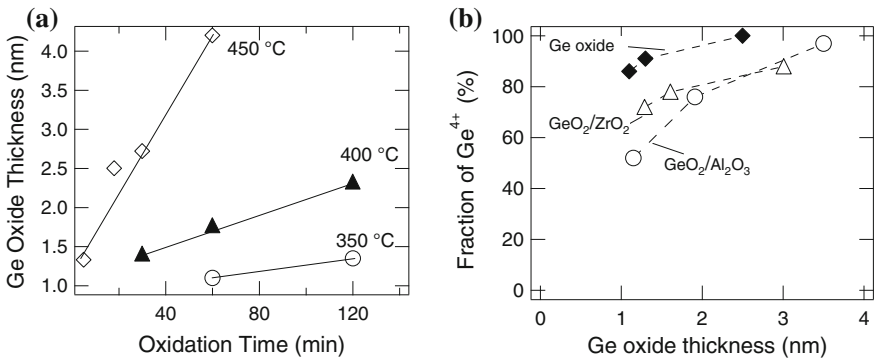


Fig. 12.10 **a** Ge oxide thickness versus oxidation time at different temperatures. **b** Fraction of Ge^{4+} oxidation state as a function of the Ge oxide thickness, estimated from XPS measurements

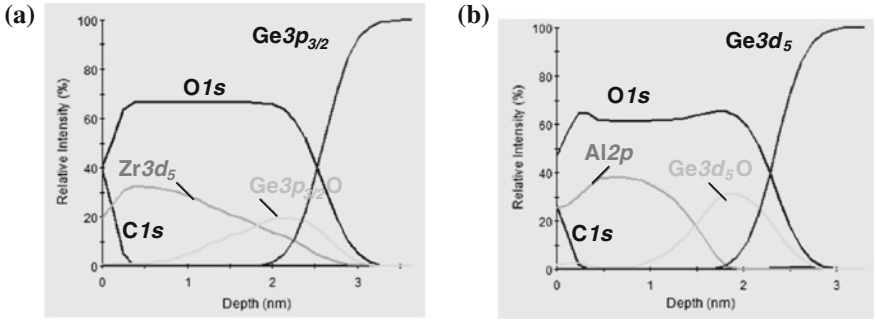


Fig. 12.11 Elemental depth profiles of $ZrO_2/GeO_2/Ge$ (**a**) and $Al_2O_3/GeO_2/Ge$ (**b**) gate stacks

The C–V characteristics of n and p-$Ge/GeO_2/Al_2O_3$ gate stacks measured at different frequencies are shown in Fig. 12.12a, b, respectively. Well-behaved C–V curves are observed, without flat-band voltage shift between the low frequency and high frequency measurements and bumps near the flat-band voltage, indicating that the interface state density, D_{it}, near the band edge is small and the surface Fermi level is likely unpinned. The minority carrier response is also observed at low frequency. All these observations suggest an efficient electrical passivation of the Ge/GeO_x interface, as will be discussed further below.

The EOT of the gate stack was calculated by C–V curve fitting, using a Ge C–V simulator, taking into account quantum–mechanical corrections. The EOT values for GeO_2/HfO_2 and GeO_2/Al_2O_3 stacks range between 1.5 and 3.7 nm, depending on the GeO_2 interface thickness, as shown in Fig. 12.13; from these measurements, the dielectric constant of the GeO_2 layer was estimated to be about 6, with κ values of 10 and 20 for Al_2O_3 and HfO_2, respectively.

The interface state density was extracted using the conductance method of Nicollian and Goetzberger, i.e., by measuring the ratio between the conductance G and the angular frequency $\omega = 2\pi f$ at different frequencies and gate voltages. Figure 12.14a shows the extracted interface state density, D_{it}, as a function of the position in the Ge bandgap for $Al_2O_3/GeO_2/p$-Ge and $HfO_2/GeO_2/n$-Ge gate stacks and for different thermal oxidation conditions. D_{it} is found to be around $1–2 \times 10^{11}$ eV^{-1} cm^{-2} near midgap, approaching state-of-the-art Si/high-κ/metal gate devices after N_2/H_2 post-metallization anneal (mid to high 10^{10} eV^{-1} cm^{-2}), confirming the efficient passivation of Ge by its thermal oxide. One also observes that D_{it} is slightly reduced in samples exposed to a higher oxidation temperature and time. Caution should be taken when extracting the interface state density of low-band gap semiconductors, like Ge, from conductance measurements performed at room temperature [75]. Indeed, minority carriers can respond to the Alternating Current (AC) signal applied during the conductance measurement, possibly leading to peaks erroneously assigned to interface states. The D_{it} values shown in Fig. 12.14a and extracted from room temperature measurements, were confirmed by conductance and capacitance measurements performed at 77 K [74].

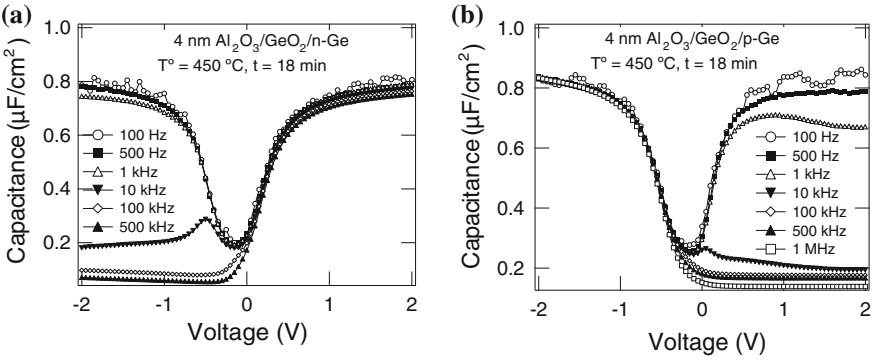


Fig. 12.12 C–V characteristics of Al_2O_3/GeO_2/Ge MOS capacitors at different frequencies on n (**a**) and p-type (**b**) Ge substrate

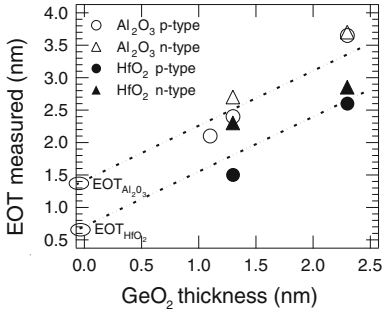


Fig. 12.13 EOT measured as a function of the GeO_2 layer thickness for HfO_2/GeO_2 and Al_2O_3/GeO_2 gate stacks

The surface potential of $Ge/GeO_2/Al_2O_3$ stacks for n- and p-type substrates, estimated from the Berglund integral method, is presented in Fig. 12.14b; the low-frequency C–V curve at 100 Hz was assumed to be an equilibrium C–V, i.e., assuming that the minority carriers are in equilibrium with the applied AC signal; such an assumption seems reasonable for low-band gap semiconductors like Ge. For p-type substrate, at negative gate bias (accumulation), one observes that the surface potential is negative. The Fermi level is close to the valence band and accumulation of holes takes place. At positive bias, the surface potential reaches positive values up to about 0.6 eV. The bands are bending down in order to allow accumulation of electrons at the interface, with the Fermi level close to the conduction band. The modulation of the surface potential with the gate bias indicates that the Fermi level at the surface is swept through the entire Ge band gap ($\sim$0.66 eV), from the valence band to the conduction band, i.e., the Ge/GeO_2 surface is very likely unpinned. Interestingly, the passivation of the Ge surface by its oxide layer has been recently predicted from first-principles calculations [63].

Hysteresis of the flat-band voltage (ΔV_{FB}) of HfO_2/GeO_2/p-Ge and Al_2O_3/GeO_2/p-Ge gate stacks is shown in Fig. 12.15a, b, respectively, when the gate voltage is swept from inversion to accumulation and back to inversion. The

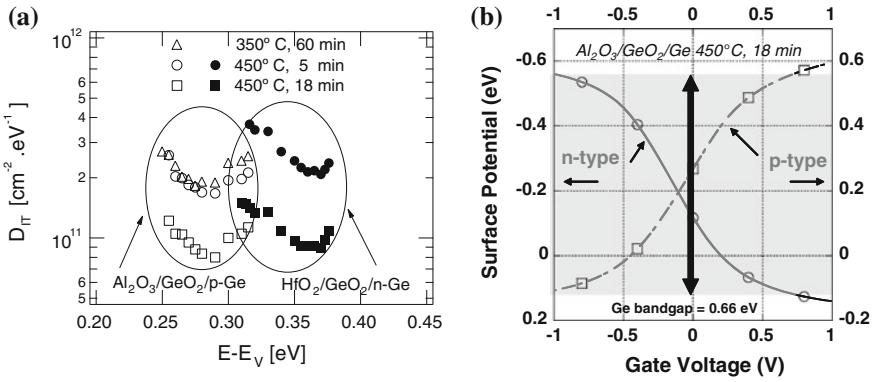


Fig. 12.14 **a** Energy distribution of traps at the interface of (100) Ge with stacks of GeO_2/HfO_2 and GeO_2/Al_2O_3. **b** Surface potential as a function of gate bias of $Ge/GeO_2/Al_2O_3$ gate stacks, on n- and p-type Ge substrate

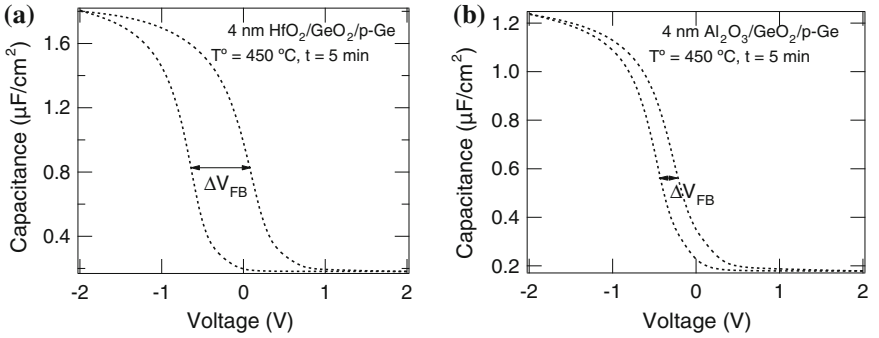


Fig. 12.15 C–V hysteresis *curves* for **a** $HfO_2/GeO_2/Ge$ and **b** $Al_2O_3/GeO_2/Ge$ gate stacks

maximum gate voltage reached in accumulation during the C–V measurements corresponds to the same electric field in the GeO_2 interfacial layer (fixed at $\sim$10 MV/cm), assuring a fair comparison of flat-band voltage hysteresis between gate stacks with different EOT. One observes a much larger hysteresis ($\Delta V_{FB} \sim 900$ mV) in the HfO_2 gate stack, as compared to the Al_2O_3 one ($\Delta V_{FB} \sim 200$ mV).

C–V hysteresis in high-κ based MOS devices is usually attributed to the trapping of charge carriers injected in the gate stack, during the gate bias sweep from inversion to accumulation [4]. In the case of GeO_2/high-κ stacks deposited by ALD (using H_2O as an oxidizing ambient), it has been recently suggested that the formation of an hydroxyl rich germanium oxide phase, with a small energy band-gap (typically 4.3 eV), leads to the injection of a high density of charge carriers in the high-κ layer when a gate bias is applied to the structure [76]. The larger hysteresis observed on the HfO_2-based stacks could also be related to the more

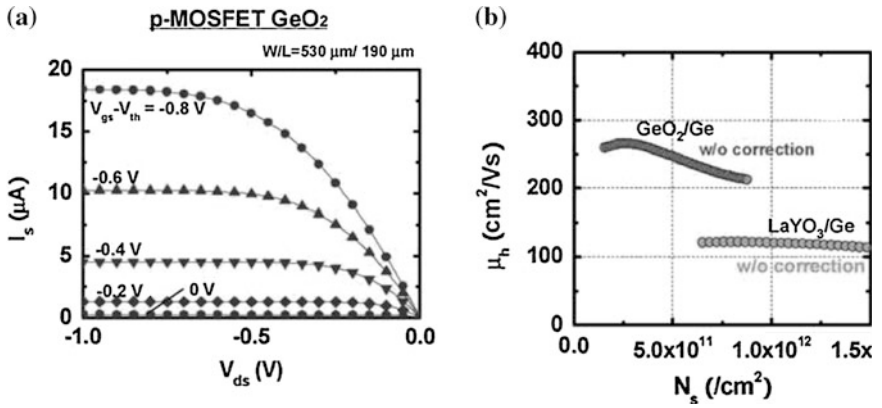


Fig. 12.16 **a** Output characteristics of a GeO_2 passivated Ge-pMOSFET [72]. **b** Extracted hole mobility versus inversion charge of Ge-pMOSFETs with GeO_2 or $LaYO_3$ gate stacks [72]

pronounced intermixing between HfO_2 and GeO_2, as indicated by the elemental depth profiles in Fig. 12.11 (assuming that HfO_2 and ZrO_2 interact similarly with GeO_2). Broader intermixing could potentially create more electrically active defects (slow states) near the interface between GeO_2 and HfO_2, leading to a larger hysteresis.

Very recently, well-behaved pMOSFETs with Ge/GeO_2 gate stacks were demonstrated, with very promising characteristics, as shown in Fig. 12.16 [72]. Functioning nMOSFETs with similar gate stacks were also reported by the same authors, with peak electron mobilities of about 270 cm^2/V s. These results confirm the efficient passivation of the Ge surface by its thermal oxide, provided that the desorption of GeO_2 is suppressed during device processing [72].

12.3.3 Possible Alternative High-κ Gate Dielectrics for Ge Substrates

Alternative high-κ dielectrics to HfO_2 and ZrO_2 have been investigated recently for their potential use in Ge-based MOSFETs. Among these materials, a few rare-earth based oxides, like La_2O_3 [77] and CeO_2 [78] have received particular attention, because these oxides can be deposited directly on Ge, apparently providing efficient passivation of the surface. The physico-chemical characterization of these interfaces reveals that the deposition of these rare-earth oxides promotes the formation of a stable germanate interlayer, likely responsible for the Ge surface passivation.

Recent ab initio calculations on the incorporation of La on GeO_x interlayer further indicated that La tends to be fourfold coordinated in the GeO_2 matrix, only forming La–O–Ge (germanate) bonds, and leading to a Ge surface free of surface states [64]. These simulations can explain the well-behaved C–V characteristics of

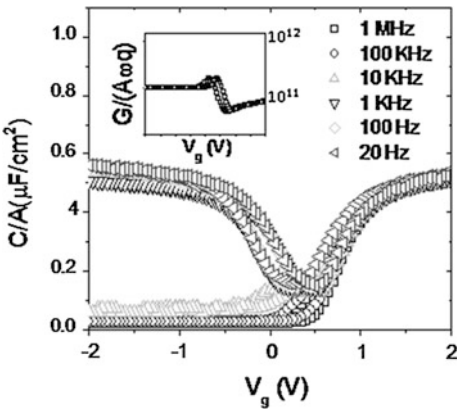


Fig. 12.17 Capacitance–voltage characteristics of a Ge/La_2O_3/Pt gate stack measured at different frequencies [77]. Inset shows the normalized conductance-voltage characteristics, from which a peak interface state density of about 4×10^{11} cm^{-2} is extracted

Ge/La_2O_3 gate stacks [77], shown in Fig. 12.17; the La_2O_3 layer was deposited by molecular beam epitaxy on a clean (2 × 1) reconstructed (100) Ge surface, and received a post-deposition anneal in O_2 at 200 °C, which promotes the formation of a La–germanate layer [77].

These rare-earth oxides have been successfully integrated into Ge-based MOSFETs, with promising electrical characteristics [79, 80]. As an illustration, the Drain current–drain voltage (I_D–V_D) characteristics and hole mobility of a La_2O_3-based pMOSFETs are shown in Fig. 12.18a, b, respectively [80]. Functioning transistors are demonstrated, with peak mobility higher than the (reference) Si/SiO_2 interface.

12.4 Passivation of GaAs and InGaAs for n-MOSFETs

III–V MOSFET has been a subject of study for several decades. The renewed research thrust is for advanced CMOS technology beyond the 22 nm node by using III–V compound semiconductors as n channels to replace traditional Si,

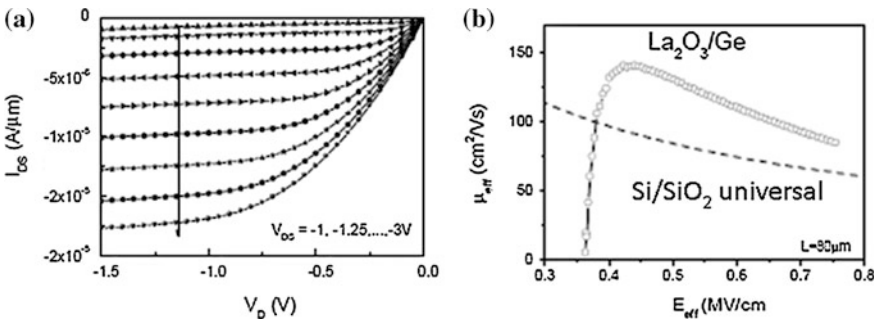


Fig. 12.18 Output characteristics (**a**) and hole mobility versus effective electric field (**b**) of a Ge/La_2O_3-based pMOSFET [80]

possibly in combination with Ge pMOSFETs. As mentioned already, the main obstacle to realizing a III–V MOSFET technology is the lack of high-quality, thermodynamically stable gate dielectrics on GaAs or III–V that can match device criteria similar to SiO_2 on Si. Our understanding of the interface physics and chemistry of III–V's is still quite limited, though enormous research efforts have been invested in this field. In this section, we mainly focus on the progress related with *ex-situ* ALD high-κ dielectrics/III–V integration, compared to the available work on in situ molecular-beam-expitaxy (MBE) grown $Ga_2O_3(Gd_2O_3)$ gate dielectric on III–V.

12.4.1 Surface Chemistry and Integration of Atomic Layer Deposition

The strong investment in research on deposited high-κ dielectrics for Si CMOS technologies began in the late 1990s [1–3]. High-κ research has since flourished—in particular the use of ALD method for fabricating Si-based high-κ MOS devices. This approach is now a manufacturable technology applied to the 45 nm CMOS node and beyond since 2007. At the end of 2001, one of the authors (Ye) and Wilk at Bell Labs/Agere Systems started to work on ALD Al_2O_3 and HfO_2 on GaAs and other III–V materials, using the commercially available ASM Pulsar 2000 ALD module [50–54]. The integration of this deposition technique with III–V compound semiconductors attracted wider interest in academia and industry, especially since a new cycle of interests in III–V MOSFETs was initiated by Intel in 2005 [81, 82].

It is widely believed that the surface pretreatment critically impacts interfacial layer thickness and composition between GaAs and ALD high-κ dielectrics. The successful integration of ALD high-κ on III–V materials also depends on the uniqueness of ALD base temperature and used precursors. In the past years, we developed simplified wet-chemical surface pretreatments for ALD high-κ process on III–V materials [83]. Before surface treatment, all wafers underwent standard degrease process using acetone, methanol and iso-propanol. The NH_4OH treatment was carried out by soaking the samples in NH_4OH (29 %) solution for 3 min to remove native oxide and rinsing in flowing de-ionized (DI) water followed by gently drying the surface using N_2 blow-gun. The NH_4OH etching step removed most of arsenic and gallium oxides from the surface. The subsequent ALD process resulted in the disappearance of arsenic oxides and self-cleaning of GaAs surface. The interface quality could be further improved sometimes by $(NH_4)_2S$ treatment. The role of sulfur passivation is to form the S–S, S–As, and S–Ga bonds at the surface. These bonds are chemically stable and are known to reduce the subsequent surface/interface reactions. The process consists in soaking the sample in $(NH_4)_2S$ for 10 min at room temperature and subsequently drying it using N_2 gun. The NH_4OH- or $(NH_4)_2S$-last surface is hydrophilic and good for two-dimensional ALD growth mode, compared to HF- or HCl- last surfaces. Detailed contact-angle

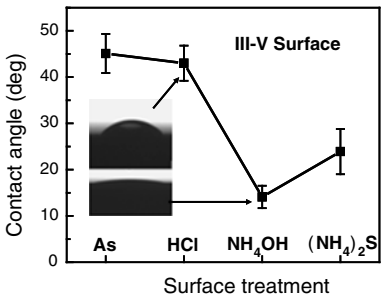


Fig. 12.19 Contact angle measurement of III–V *surfaces* after different surface treatments. NH_4OH and $(NH_4)_2S$ pretreated surfaces are hydrophilic with small contact angles

experiments on different chemically treated III–V surfaces unveil this point as shown in Fig. 12.19. It is particularly important for ultrathin oxide layers.

The thermal treatment is also an important factor. For GaAs at 300–460 °C, mixed gallium/arsenic oxides are converted into pure gallium oxide with As precipitates according to the reaction $As_2O_3 + 2\ GaAs \rightarrow Ga_2O_3 + 4\ As\uparrow$ (with partial As desorption in the form of As_2 and As_4). The typical ALD base temperature (300 °C) is also critical since this conversion process can be significantly enhanced by introducing chemical ligands such as Tri Methyl Aluminum (TMA) or Tetrakis Ethyl Methyl Amino Hafnium (TEMAH). Figures 12.20a, b show the high-resolution TEM images of the GaAs substrate with a 'native oxide' layer (containing e.g., Ga_2O_3, As_2O_3, As_2O_5, etc.) and the GaAs substrate after amorphous ALD Al_2O_3 deposition, respectively. It is clear that the native oxide on GaAs surface was "etched" away by the ALD process. Depositions were performed using alternating exposures of the common ALD precursors $Al(CH_3)_3 + H_2O$ in an N_2 carrier gas at 300 °C. This ALD "self-cleaning" effect

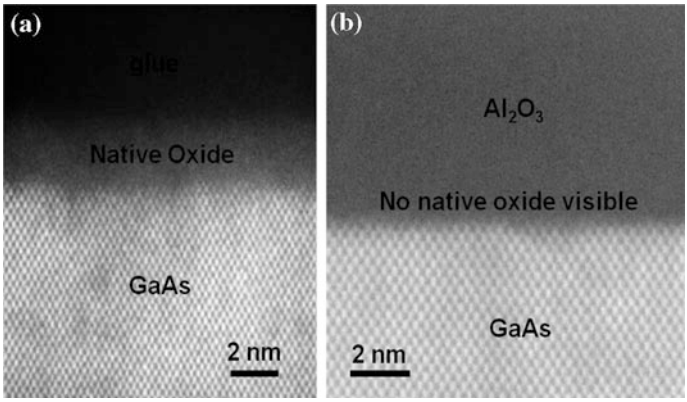


Fig. 12.20 **a** High-resolution TEM images of a GaAs substrate exposed to air with a 2–3 nm thick native oxide on the surface. **b** High-resolution TEM image of the same sample after ALD process. 2–3 nm native oxide was consumed or removed during ALD Al_2O_3 process. Native oxide is not desorbed at 300 °C. The chemical reaction during ALD causes the consumption. The images were taken by Dave Muller at Bell Labs of Lucent Technologies

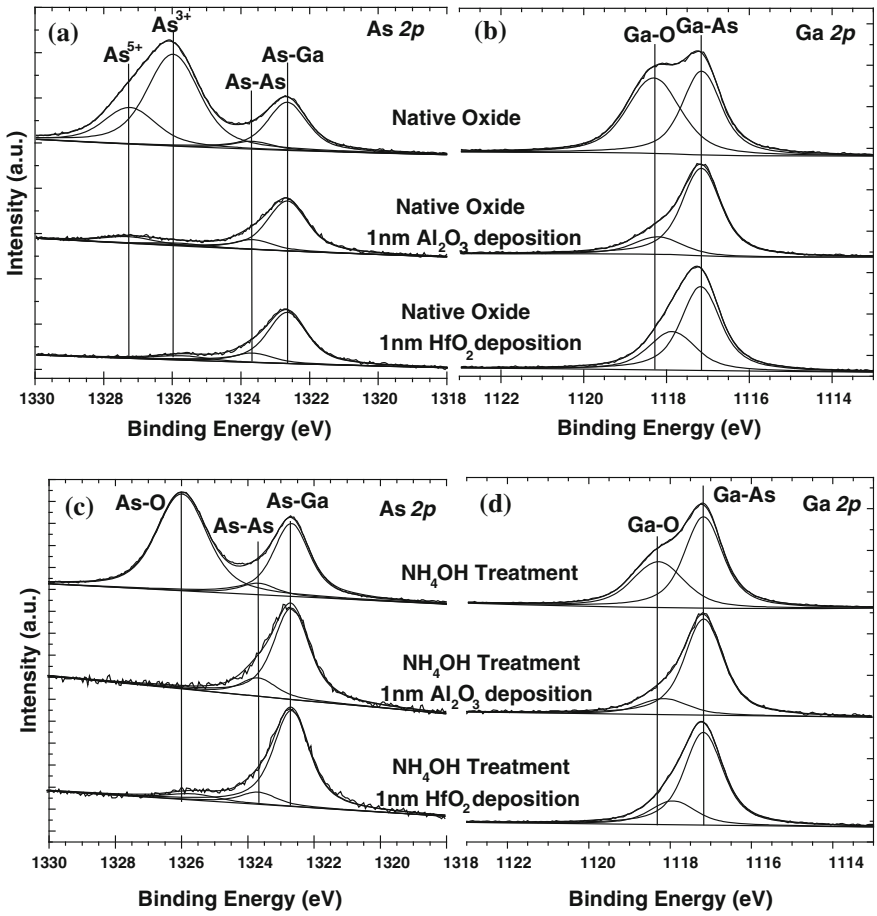


Fig. 12.21 **a** As *2p* 3/2 and **b** Ga *2p* 3/2 spectra showing oxidation state differences for a native oxide and subsequent 1 nm Al_2O_3 and HfO_2 depositions on a GaAs surface. **c** As *2p* 3/2 and **d** Ga *2p* 3/2 spectra showing oxidation state differences following NH_4OH surface treatment and subsequent 1 nm Al_2O_3 and HfO_2 depositions on a GaAs surface. The data was taken by C. L. Hinkle et al. at University of Texas at Dallas [87]

was discovered at the very beginning of the experiments and briefly addressed by Ye et al. [51], and further discussed in greater details by Frank et al. [84].

These ALD "self-cleaning" effects were confirmed by many different groups [85–89]. Figure 12.21 shows the clearest XPS work on this effect from Wallace's group at University of Texas at Dallas, associated with NH_4OH pretreatment [87]. Figure 12.21a shows that the As_2O_3 and As_2O_5, which are thought to be the oxides responsible for the pinning of the Fermi level at the GaAs interface, are mostly removed after 1 nm Al_2O_3 (using TMA) or HfO_2 (using TEMAHf) ALD deposition process. This effect is further improved on the NH_4OH etched GaAs surface with no visible arsenic oxide at Al_2O_3/GaAs interface, as shown in Fig. 12.21c.

These results are consistent with our electrical measurements on devices with similar ALD processes, showing mid-gap D_{it} of 8×10^{11}–$1 \times 10^{12}/cm^2$ eV for Al_2O_3/GaAs and mid-gap D_{it} of 1–2 $\times$ $10^{12}/cm^2$ eV for HfO_2/GaAs.

12.4.2 III–V Substrates Engineering for Majority and Minority Carrier MOSFETs

If we consider all semiconductor electronic devices, the most important one is the field-effect transistor, which can be divided into two categories: (1) majority carrier devices and (2) minority carrier devices. Si works nicely as a minority carrier device, with large inversion currents at the "perfect" SiO_2/Si interface (after forming gas annealing, used for the passivation of dangling bonds at this interface). This minority carrier Si MOSFET is the building block of our modern microelectronic industry. On the other hand, GaAs works only as a majority carrier device, such as high electron mobility transistors (HEMTs), used in our cell-phones. However, such devices consume lot of power and are not suited for digital applications. If we could fabricate a minority carrier GaAs (or III–V) MOSFET, the microelectronic industry could be reshaped.

In the past years, we studied the integration of ALD Al_2O_3 on GaAs [50, 51], InGaAs [52] and GaN [54], and demonstrated high performance depletion-mode (majority carrier device) III–V MOSFETs. For example, a 1 μm gate-length depletion-mode n-channel $In_{0.2}Ga_{0.8}As$/GaAs MOSFET with an Al_2O_3 gate oxide of 160 Å showed a gate leakage current density less than 10^{-4} A/cm^2, a maximum transconductance $\sim$105 mS/mm, and a strong accumulation current at $V_{gs} > 0$ in addition to buried-channel conduction. The maximum drain current reached 350 mA/mm at 1 μm gate length. Together with longer gate-length devices, we achieved electron accumulation surface mobility for $In_{0.2}Ga_{0.8}As$ as high as 660 cm^2/V s at the $Al_2O_3/In_{0.2}Ga_{0.8}As$ interface. Depletion-mode devices, which are majority carrier devices with doped transport carriers, are relatively easy to demonstrate because the required surface potential movement is limited. In most cases, the surface potential lies near mid-gap for n-type doped channels, and D_{it} is supposed to be the lowest at mid-gap for oxide/III–V interfaces.

For high-speed digital applications, enhancement-mode (E-mode) III–V MOSFETs within the Si CMOS platform are the real devices of interest. Due to the existing Fermi-level pinning issue or large interface trap density near conduction or valence band edges at oxide/III–V interface, it is very difficult to realize high-performance inversion-mode or surface channel III–V MOSFETs. In other words, the interface trap density (D_{it}) at the conduction/valence band edge of oxide/III–V interface is too high to have large inversion current. One alternative approach is to implement III–V hetero-junction or buried channel design, together with a high-κ dielectric layer (to reduce the gate leakage current) and a high work-function metal gate electrode, in order to realize the "normally-off" E-mode operation. The

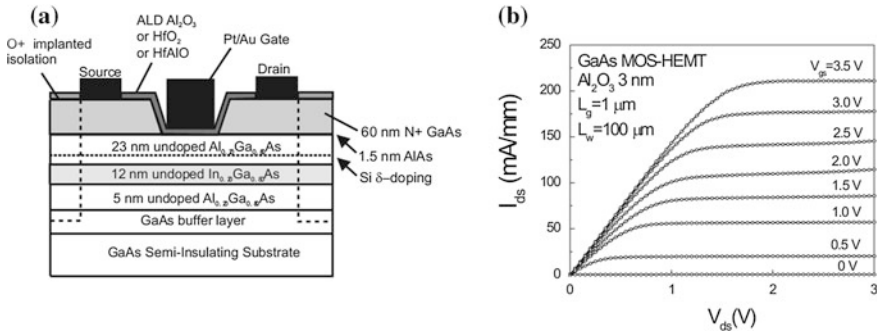


Fig. 12.22 **a** A cross-section of an E-mode high-κ/GaAs MOS-HEMT. **b** *I–V* characteristic of a 1 μm gate-length GaAs MOS-HEMT with a 3 nm ALD Al_2O_3 as a gate dielectric

devices as shown in Fig. 12.22 are operated in *accumulation region* [90], being different from the traditional inversion-type E-mode devices. The E-mode device is still a majority carrier device since the transport carriers (electrons) are δ-doped in AlGaAs by Si dopant. The high-mobility electrons are confined in semiconductor heterojunctions or quantum wells instead of oxide–semiconductor interfaces. Due to the implementation of the thick AlGaAs layer, the devices are hard to scale down to 100 nm for ultimate CMOS applications. But these devices may find a wide range of applications in the traditional III–V fields, such as wireless electronics or optoelectronics.

High-performance or high-drain-current/high transconductance inversion-type E-mode III–V MOSFET is a big challenge in the device research community for decades. Figure 12.23 shows a successful demonstration by implanting MBE p-$In_{0.53}Ga_{0.47}As$ as surface channel and ALD Al_2O_3 as the insulating layer [91]. The ALD process on III–V compound semiconductors enables the formation of high-quality gate oxides and unpinning of Fermi-level on In-rich InGaAs. A 0.5 μm gate-length MOSFET with an Al_2O_3 gate oxide thickness of 8 nm shows a gate leakage current less than 10^{-4} A/cm^2 at 3 V gate bias, a threshold voltage of 0.25 V, a maximum drain current of 367 mA/mm and a transconductance of 130 mS/mm at a drain voltage of 2 V. Mid-gap interface trap density of regrown Al_2O_3 on $In_{0.53}Ga_{0.47}As$ is $\sim 1.4 \times 10^{12}$/cm^2 eV determined by low-frequency and high-frequency C–V method. The peak effective mobility is $\sim$1,100 cm^2/V s from DC measurements, $\sim$2,200 cm^2/V s after interface trap correction, i.e., about a factor of 2–3 times higher than Si universal mobility in the range of 0.5–1.0 MV/cm effective electric field. Similar device performances have also been realized on ALD HfO_2 and HfAlO dielectrics [92]. The highest drain current is $\sim$430–450 mA/mm and transconductance is $\sim$180 mA/mm.

High-performance inversion-type enhancement- mode n-channel $In_{0.65}Ga_{0.35}As$ MOSFETs with ALD Al_2O_3 as gate dielectrics have also been demonstrated [93]. $In_{0.65}Ga_{0.35}As$ MOSFETs, with 12 % increase in In concentration, leads to a factor of 3 increased drain currents, compared to $In_{0.53}Ga_{0.47}As$ MOSFETs. A 0.4 μm

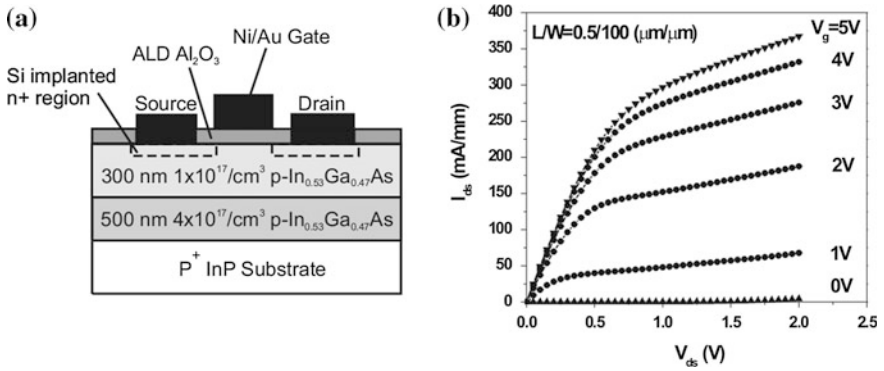


Fig. 12.23 **a** A cross section of an inversion-type E-mode Al_2O_3/InGaAs MOSFET. **b** *I–V* characteristic of a 0.5 μm mask gate length InGaAs MOSFET with an 8 nm ALD Al_2O_3 as gate dielectric

gate-length MOSFET with an Al_2O_3 gate oxide thickness of 10 nm shows a gate leakage current less than 5×10^{-6} A/cm^2 at 4.0 V gate bias, a threshold voltage of 0.4 V, a maximum drain current of 1.05 A/mm and a transconductance of 350 mS/mm at a drain voltage of 2.0 V. The maximum drain current and transconductance scales linearly from 40 μm down to 0.7 μm. The peak effective mobility is ~1,550 cm^2/V s at 0.3 MV/cm and decreases to ~650 cm^2/V s at 0.9 MV/cm. The maximum drain current and transconductance obtained are all record high values in 42 years of enhancement-mode III–V MOSFET research [93]. Figure 12.24 shows the device structure and representative I–V characteristics of such a device. Similar performances have also been realized on ALD HfO_2 dielectrics [94]. In-rich InGaAs is thus well positioned for high-speed low-power logic applications due to its high electron mobility and saturation velocity, and also its narrow bandgap required for low supply voltage. In addition, deep sub-

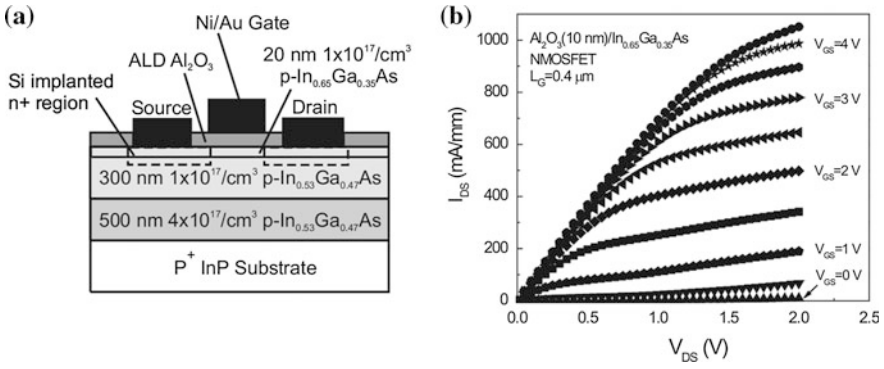


Fig. 12.24 **a** A cross section of an inversion-type E-mode $Al_2O_3/In_{0.65}Ga_{0.35}As$ MOSFET. **b** *I–V* characteristic of a 0.4 μm-gate-length $In_{0.65}Ga_{0.35}As$ MOSFET with a 10 nm ALD Al_2O_3 as gate dielectric

micron inversion-type enhancement-mode $In_{0.75}Ga_{0.25}As$ MOSFET with ALD Al_2O_3 as gate dielectric has been recently demonstrated experimentally. The n-channel MOSFET with gate length from 100 to 200 nm have been fabricated by a full electron-beam lithography process and characterized with drain saturation current (I_{ds}) from 260 to 630 μA/μm, transconductance from 540 μS/μm to 950 μS/μm at a maximum supply voltage of only 0.8 V [95]. One of the most important characteristics of this type of surface-channel MOSFET is its scalability. So far the scalability on drain current and transconductance have been demonstrated down to 100 nm on $In_{0.75}Ga_{0.25}As$ MOSFETs. Other important scaling metrics, such as sub-threshold swing (SS) and drain induced barrier lowering (DIBL), which are seriously affected by the short channel effects, also show promise. Benchmarking these transistors with state-of-the-art planar Si logic transistors is considered with several key device metrics, including intrinsic gate delay (CV/I) and energy-delay product ($CV/I{\cdot}CV^2$). Promising results demonstrate that In-rich InGaAs MOSFET is of great potential for the future high-speed low-power logic applications.

During the past decades, the research community focused mainly on dielectric materials and III–V surface chemistry study, and paid less attention to device physics of III–V MOSFET. We emphasize that device physics is also extremely important as the second determinant for realizing high inversion current, due to the smaller energy-level separation between the charge neutrality level (CNL) of In-rich InGaAs and the conduction-band minimum (CBM), compared to the same for GaAs. The simple CNL-based model below can explain all experimental work on III–V MOSFETs using *ex-situ* ALD high-κ dielectrics and also in situ MBE-grown $Ga_2O_3(Gd_2O_3)$ gate dielectric [96, 97]. There are many theoretical studies using sophisticated techniques (such as the metal-induced gap state model, the unified defect model, and the disorder-induced gap state model) to quantitatively calculate the electronic energies of the system. To simplify our explanation, we choose one internal reference level (CNL), as illustrated in Fig. 12.25, to phenomenologically

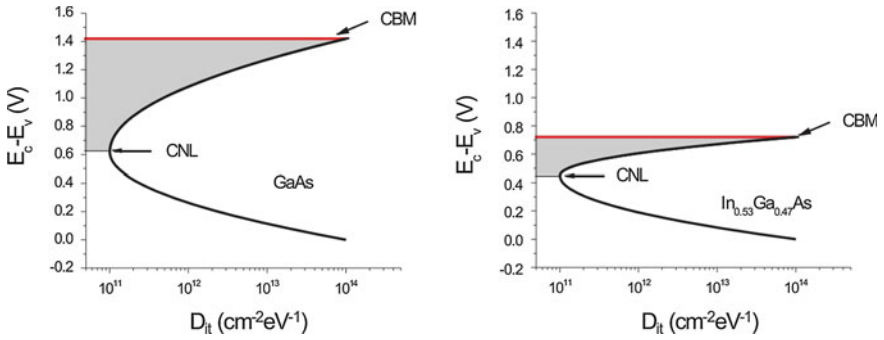


Fig. 12.25 Schematic for the parabolic D_{it} distribution within the energy band of GaAs and $In_{0.53}Ga_{0.47}As$. The CNL is aligned 0.8 eV below CBM for GaAs and 0.27 eV below CBM for $In_{0.53}Ga_{0.47}As$. The *shadow area* shows the built-up of negative charges at interface traps when the Fermi level moves from CNL to CBM

explain all our experimental observations. Presumably, every interface has donor-type and acceptor-type interface traps. For convenience, let us consider that the contribution of these two types of defect results to an equivalent D_{it} distribution, characterized by the charge-neutrality-level E_{CNL}. If the Fermi-level E_F is above E_{CNL}, the states are of acceptor type and negatively charged (when occupied). If Fermi-level E_F is below E_{CNL}, the states are of donor type and positively charged (when empty).

Figure 12.25 illustrates the basic idea to qualitatively understand the transistor output characteristic of Figs. 12.23 and 12.24. The model is extremely simplified to highlight the fundamental point with a few assumptions. First, the D_{it} distribution from valence-band maximum (VBM) to CBM is parabolic on a logarithmic scale. This is a reasonable assumption because the interface traps at III–V interfaces are more significant, compared to the SiO_2/Si interface with a U-shaped distribution. Second, the D_{it} value at CBM is fixed at 10^{14}/cm^2 eV for simplification. Third, Fermi-level stabilization energy is chosen as CNL. It is obvious now that the dominant factor is the energy difference between CNL and CBM. For GaAs, with the CNL and CBM difference as large as $\sim$ 0.8 eV, it builds up 1.02×10^{-6} coul/cm^2 negative charges to prevent a strong inversion charge to participate in transport. In contrast, the CNL and CBM potential difference for $In_{0.65}Ga_{0.35}As$ is only 0.15 eV assuming a linear extrapolation in $In_xGa_{1-x}As$ ternary alloys (CNL for InAs being at 0.2 eV above CBM). The built-up negative charge is only 1.91×10^{-7} coul/cm^2, a factor of 5 smaller than that in GaAs. That explains why an $Al_2O_3/In_{0.20}Ga_{0.80}As$ MOSFET has much less maximum drain current than an $Al_2O_3/In_{0.53}Ga_{0.47}As$ MOSFET and $Al_2O_3/In_{0.65}Ga_{0.35}As$ MOSFETs. It also explains why $Ga_2O_3(Gd_2O_3)$/GaAs MOSFETs have only less than 1 mA/mm drain current and $Ga_2O_3(Gd_2O_3)/In_{0.53}Ga_{0.47}As$ MOSFET has 360 mA/mm drain current. InP is another interesting material to check the applicability of the above model. InP has a bandgap of 1.35 eV, which is very similar to GaAs (1.42 eV). But the difference between CNL and CBM for InP is only $\sim$0.5 eV. It is thus much easier to realize an inversion-mode InP MOSFET than a GaAs one. Indeed, 100 mA/mm drain current InP MOSFETs have been demonstrated using ALD high-κ dielectrics [98]. Note that the interface quality of Al_2O_3/III–V is usually better than that of HfO_2/III–V in general. But the built-in charge increases less than a factor of 2 from the mid-gap D_{it} of 10^{11}/cm^2 eV 10^{12}/cm^2 eV at CBM. That explains why Al_2O_3, HfO_2, HfAlO, and even in situ-grown $Ga_2O_3(Gd_2O_3)$ have similar device performances on the same III–V substrate. The ALD process is obviously much more favorable, because it is *ex-situ*, robust, manufacturable, and widely used in the Si industry.

12.4.3 Methodology for III–V Interface Characterization

III–V interface characterization is a challenging topic. Some well established methods in Si are not suitable for III–V due to the different band-gap and minority

carrier concentrations [99]. Except for the commonly used capacitance methods and conductance method, III–V interface quality can also be studied by subthreshold characteristics [95, 100, 101], charge pumping [101] and optical photoluminescence [102]. Here, we mainly discuss the capacitance methods.

As already mentioned above, the major objective of GaAs MOS device research is to realize high-performance enhancement-mode surface channel devices with scalable gate dielectrics. The low frequency (LF) or quasi-static (QS) C–V curve is useful to analyze the properties of GaAs MOS structures, such as interface trap density, inversion charge, etc. However, it is difficult to observe LF or QS C–V on GaAs MOS device under normal conditions (room temperature and in the dark). According to the Shockley–Read–Hall statistics and the low intrinsic carrier concentration (n_i) of GaAs ($n_i = 10^6/cm^3$), the expected ac frequency to observe LF C–V in dark and at room temperature is very low ($\sim$0.002 Hz) [103]. For QS C–V, the required gate leakage current should be much less than the displacement current. This condition is hard to fulfill for high-κ dielectrics that are only a few nanometers thick. Three alternative approaches [104] to study LF C–V on GaAs are: (1) inversion-type MOS field-effect transistors with implanted source and drain, where minority carriers can be injected into the surface channel; (2) measurements at elevated temperature to increase the generation-recombination rates of minority carriers in GaAs; (3) photoillumination to increase the minority carrier concentration.

The high-frequency (1 kHz) C–V curves for a capacitor with a 30 nm ALD Al_2O_3 on $In_{0.20}Ga_{0.80}As$ is shown in Fig. 12.26a. This capacitor went through all device processes and was annealed with a Rapid Thermal Annealing (RTA) step of 800 °C for 10 s in nitrogen ambient. The solid curve is measured from −5 to +3 V with the sweep rate $\sim$4 V/min., while the dashed line is taken from +3 to −5 V. The C–V traces exhibit almost "zero" hysteresis in the voltage sweep loop, with the maximum shift less than 20 mV. Even with such 'good looking' C–V characteristics, no n-channel inversion is observed at frequency as low as 100 Hz at the

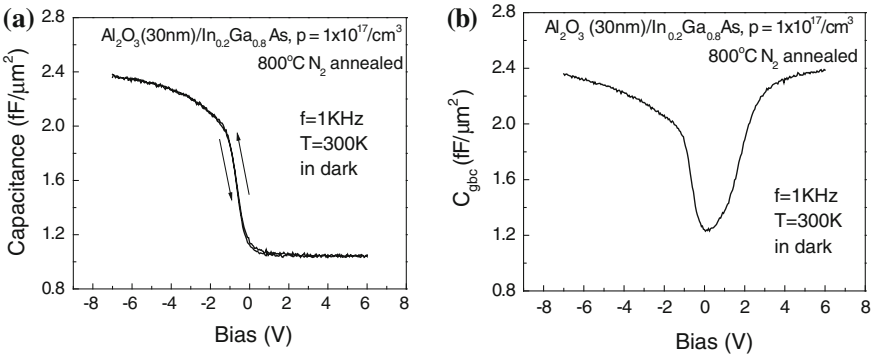


Fig. 12.26 **a** High-frequency (1 kHz) capacitance–voltage versus for the ALD Al_2O_3/InGaAs MOS capacitor subjected to a RTA step of 800 °C, 10 s in nitrogen. The hysteresis is negligible. **b** C_{gb} by split-CV measurement on a MOSFET with a 40 μm gate length and 100 μm gate width at a frequency of 1 kHz

conventional MOS configuration. Clear n-channel inversion (more accurate LF-CV) on p-type InGaAs is realized at Al_2O_3/InGaAs interface, which is demonstrated by measuring C_{gb} (split-CV method) on InGaAs MOSFETs as shown in Fig. 12.26b. The C–V curve is taken on a MOSFET with 40 μm gate length and 100 μm gate width at frequency as high as 1 kHz. It is understood that the minority carriers (electrons) are provided by injection from the source and not by thermal generation in the depletion region. In the split-CV or MOSFET configuration with source and drain grounded at the same time, majority (holes) and minority (electrons) carrier capacitances can be measured independently at the same frequency. Using a mid-gap low-frequency capacitance $C_{LF} = 5.5$ pF, a mid-gap high-frequency capacitance $C_{HF} = 5.0$ pF and an oxide capacitance $C_{ox} = 11.0$ pF, the mid-gap interface trap density D_{it} is determined to be $2.9 \times 10^{11}/cm^2$ eV. In this case, the minority carriers (electrons) are provided by the implanted source instead of the inversion process of the bulk semiconductor. The device is thus not under equilibrium condition. Strictly speaking, it does not correspond to a 'real' inversion of the MOS structure.

The second method consists in measuring the C–V characteristics of a MOS capacitor at an elevated temperature, to increase the generation-recombination rates of minority carriers in GaAs. Temperature dependent C–V measurements with a wide range of small ac signal frequencies are also widely used to study the minority-carrier recombination kinetics. The C–V measurements at 100 Hz–1 MHz and at temperatures from 300–500 K are systematically studied. Figure 12.27 shows the results measured at 1 kHz as a function of bias at different temperatures. The major effect of temperature occurs during depletion to inversion transition at positive biases. At room temperature (300 K), minority-carriers do not follow the 1 kHz signal for GaAs, thus a high frequency curve is measured as shown in Fig. 12.27. However, as temperature is increased from 300 to 500 K, minority carriers begin to follow the AC signal because generation and recombination rates both increase with temperature, and the transition from a high frequency curve to a low frequency curve is clearly observed. By systematically varying both the temperature and the frequency, the temperature dependence of the

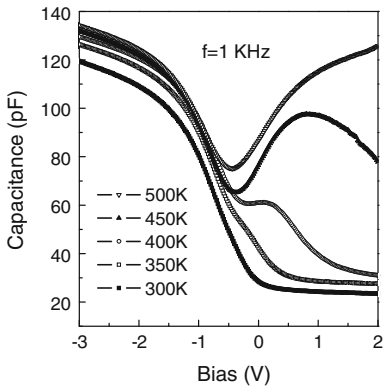


Fig. 12.27 Capacitance measurement at 1 kHz as a function of bias with different ambient temperatures. LF-CV is clearly observed at elevated temperatures (above 400 K)

transition frequency can be determined. The activation energy (E_A) of minority-carrier generation and recombination for Al_2O_3/GaAs MOS structures is $\sim 0.71 \pm 0.05$ eV, which is just about half the energy band-gap of GaAs. This activation energy is that of intrinsic electrons in GaAs. Therefore, the dominant mechanism for minority carrier response must be generation and recombination through semiconductor bulk traps or interface traps in the measured MOS structures over the temperature range 300–500 K. The significant change from high-frequency C–V at 300 K to low-frequency C–V at 500 K is due to the dramatic increase in minority carrier concentration in the bulk GaAs through thermal generation, as opposed to improved interface properties at high temperatures. This is why the "inversion" features at quasi-static C–V or low-frequency C–V should not be used as conclusive evidence to characterize the MOS interface, because it could be related to the minority carriers in the semiconductors, and not to D_{it}.

The third approach is to use photo-illumination to increase the minority carrier concentration. However, for a conventional, few 100 nm thick metal gate (i.e., Ni/Au or Ti/Au on GaAs), the gate absorbs most of the photons. Under photo-illumination, only the edge and a small area within a diffusion length can be illuminated by photons. The large central area, typically tens to hundreds of microns, remains in the dark. In most cases, no LF C–V is observed on GaAs even when measured under illumination at 1 kHz–1 MHz as shown in Fig. 12.28a. Clear LF C–V is demonstrated under photo-illumination at room temperature and at conventional measurement frequencies by using transparent conducting indium tin oxide (ITO) as metal gate for GaAs MOS structure, as shown in Fig. 12.28b [105]. ITO is a mixture of indium oxide (In_2O_3) and tin oxide (SnO_2), typically 90 % In_2O_3, 10 % SnO_2 by weight. It is transparent and colorless for thin layers.

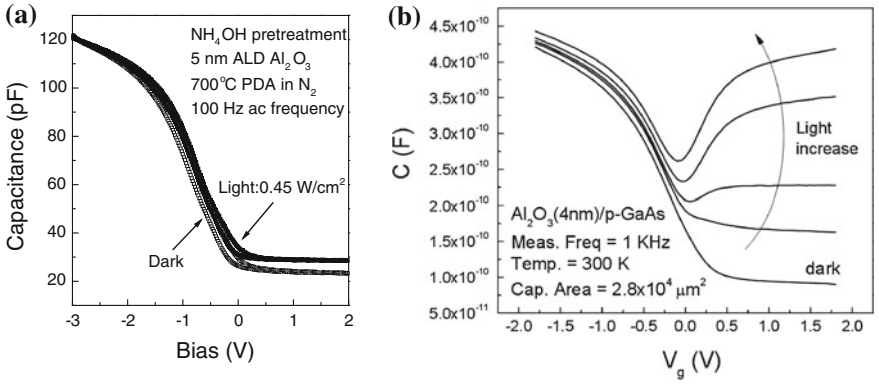


Fig. 12.28 **a** 100 Hz C–V loops for a MOS capacitor with 5 nm ALD Al_2O_3 film on p-type GaAs substrate measured in dark and with illumination. The inversion layer is difficult to form in GaAs even if the device with opaque gate electrode is illuminated with a light intensity of 0.45 W/cm^2. **b** Capacitance measurement at 1 kHz as a function of gate bias with increased light intensity from the microscope. Indium tin oxide is used as transparent conducting electrodes to improve light illumination on GaAs. The light intensity is simply controlled by the electrical power supplied to the microscope lamp

The main feature of this material is the combination of electrical conductivity and optical transparency, so that it can be used as a transparent gate for GaAs MOS capacitors. When light is on, the whole gate area is illuminated by photons, in contrast to only a small edge area that can be illuminated when conventional metals are used. As shown in Fig. 12.28b, the C–V still shows high frequency C–V characteristic in the dark. However, as light is shone on the device by controlling the input current of the power supply (the light intensity of the microscope lamp), the inversion capacitance (due to electrons) increases. LF C–V characteristics can be clearly observed when the lamp is fully on. Photo-illumination generates a large number of electron–hole pairs in the depletion layer, so that the minority carrier concentration (electrons) in GaAs is significantly increased. The carrier generation time becomes much shorter under photo-illumination; therefore it is much easier for the carriers to follow the relatively high frequency ac signals, and show low frequency C–V characteristics. The photo-generation of minority carriers and its response time for thermal generation-recombination could play a more important role on the "inversion" feature than the interface trap density in the C–V measurement. The observed LF or "inversion" feature in C–V measurement is still not a very clear conclusive evidence for Fermi-level unpinning on GaAs, because the pool of electrons are not directly from the band-bending to realize the inversion of bulk semiconductor. A 'good exercise' is to integrate low-frequency capacitance and obtain the surface potential ψ_s–V relationship, as discussed in Sect. 12.3.2. The surface potential change on GaAs of Fig. 12.28b can be calculated to be as large as 1.1 eV by Berglund equation, which is comparable to the GaAs bandgap itself. It might indicate that the Fermi level at ALD Al_2O_3/GaAs interface moves across nearly the whole bandgap of GaAs and it is not pinned at the mid-gap. However, Berglund equation is valid only for 'real inversion' in equilibrium condition.

All these three approaches are at non-equilibrium conditions or not at room temperature. The best way to study MOS is still the standard quasi-static (QS) CV. For QS C–V, the required gate leakage current should be much less than the displacement current. This condition is hard to fulfill for high-κ dielectrics that are only a few nanometers thick. Using 8 nm thick Al_2O_3, quasi-static/high frequency C–V measurement is presented at a very slow sweep rate, at room temperature, and in dark in Fig. 12.29. The origin of the clear inversion feature on QS C–V at positive biases is still under discussions [106]. However, some simple analysis of the data is helpful. Using Berglund equation to integrate quasi-static capacitance and obtain the Ψ_s–V relationship, the surface potential Ψ_s changes by 1.2 eV, nearly the entire band-gap of GaAs. Berglund equation is used to describe the real inversion and it may not be suitable if the LF-CV feature is from trap-related origin. Taking one step back conservatively, we calculate the surface potential change from accumulation to flat band condition; for that condition, the Berglund equation should be valid. The surface potential movement is 0.8 eV, more than half band gap of GaAs. The Fermi level at ALD Al_2O_3/GaAs interface seems not to be pinned, at least based on this simple analysis. Similar conclusion is also obtained by detailed experiments on metal work function study. In these

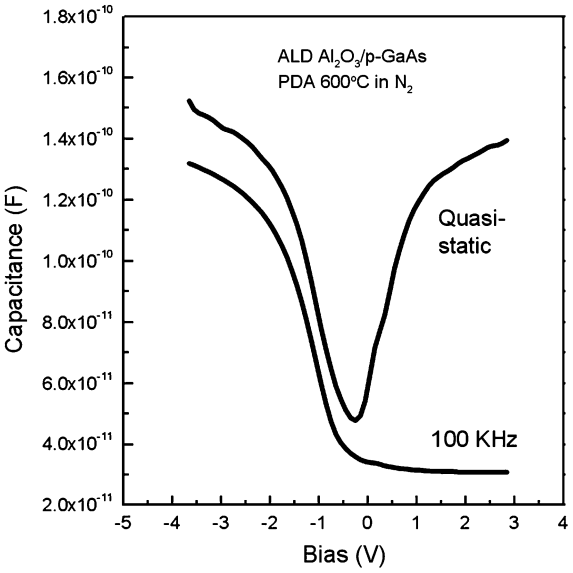


Fig. 12.29 Quasi-static and high frequency C–V measurement for a MOS capacitor with 8 nm ALD Al_2O_3 film directly on p-type GaAs substrate. Post deposition annealing was done at 600 °C for 30 s in nitrogen ambient. Gate metal is Ni. The data was taken at room temperature and in dark using Keithley 595 Quasi-static CV Meter and Keithley 590 CV Analyzer. The quasi-static C–V was measured with 0.5 s delay time on each data point. The capacitor area is $\sim$17,600 μm^2

experiments, we studied the flat-band voltage shift versus different metal gates on ALD Al_2O_3/GaAs, including Ti, Al, Ni, Pd, and Pt [107]. This so-called 'S-factor method', where $S_{\mathrm{Metal\_A,Metal\_B}} = \left(V_{\mathrm{FB,A}} - V_{\mathrm{FB,B}}\right)/\left(\Phi_{\mathrm{m,A}} - \Phi_{\mathrm{m,B}}\right)$, is the most direct way to determine if the Fermi-level is pinned or not, avoiding the complication of interpretation of the measured C–V data. The S-factor for $(NH_4)_2S$ treated samples is nicely aligned between 0.71 and 0.73, with the flat-band voltage shift $\sim$0.91 V, i.e., more than half bandgap of GaAs. Similar conclusion is also obtained on NH_4OH treated samples with flat-band shift $\sim$0.86 V.

However, one should admit that our understanding is still very limited after four decades of research on these 'mysterious' GaAs MOS interfaces. More studies are still needed to understand how to accurately characterize theses interfaces and, more importantly, how to improve it up to the SiO_2/Si interface quality. The physics and chemistry we have learned from the GaAs MOS system is also very valuable to III–V MOS in general.

12.5 Summary

The need for high-κ gate dielectrics and metal gates in advanced integrated circuits has re-opened the door to germanium and III–V compounds as potential replacements for silicon channels, offering the possibility to further increase the performance of future CMOS generations, as well as adding new circuit functionalities (more than Moore concept). One of the major challenges of Ge and III–V compound MOS technology is the passivation of their interfaces, required

for the proper functioning of field-effect transistors. Recent progress in Ge and III–V surface passivation has been reviewed in this chapter.

In the specific case of Ge, a number of passivation approaches have been successfully developed recently, including the deposition of a very thin epi-Si layer, the thermal oxidation of Ge, and the deposition of rare-earth oxides on clean Ge surfaces. All these approaches allow the fabrication of functioning pMOSFETs, with peak hole mobilities exceeding that of reference Si/SiO_2 interfaces by a factor of 2–3. Though some reports have also demonstrated functioning nMOSFETs with excellent electron mobilities, especially for Ge/GeO_2-passivated interfaces [72], most of the results reported in the literature indicate poor nFET performances. This could be related to a fundamental issue, i.e., the charge neutrality level of Ge lying close to the valence band edge, making it difficult to form an electron channel at the surface [108], or related to an asymmetric distribution of interface states in the Ge band-gap, with a very high density (up to 10^{13}/eV cm^2) of interface states near the Ge conduction band edge.

The lack of high-performance Ge-nFET requires the development of III–V compound-based devices, for the fabrication of complimentary MOS circuits. In that respect, significant progress in the realization of inversion-type enhancement mode III–V MOSFETs has been achieved recently, using ALD high-κ gate dielectrics, allowing for improved GaAs and InGaAs surface passivation. High-In content (50–60 %) InGaAs channels, combined with Al_2O_3 or HfO_2 ALD gate stacks, show very promising performances, with record drive current and transconductance. The improved performances of In-rich devices, compared to low In-content ones, can be explained by using the CNL concept, the CNL of In-rich devices being much closer to the conduction band minimum, allowing for a much easier substrate inversion.

Caution must be taken when assessing the quality of high-mobility semiconductors/high-κ dielectric interfaces, since methodologies developed for Si-based MOS structures cannot always be applied in a straightforward way, due to the different energy band-gap and associated carrier concentrations in high-mobility semiconductors. In the case of GaAs, which has a higher band-gap compared to Si, C–V and G–V measurements performed at higher temperatures (around 500 K) allow to extract D_{it} near mid-gap, where the Fermi level is usually pinned. For InGaAs with a high In-content or Ge, which have a lower band-gap compared to Si, measurements at low-T (e.g., liquid nitrogen) are better suited for the extraction of D_{it}, in order to suppress the "parasitic" response of minority carriers to the applied AC signals during C–V or G–V measurements.

Finally, let us point out that the recent progress achieved in the passivation of Ge and InGaAs surfaces, in combination with high-κ gate stacks, are very promising for the possible co-integration of these high-mobility channels into sub-22 nm CMOS generations. Though much work is still needed to fundamentally understand their surface passivation, further boost device performance, and integrate Ge and III–V compounds on large Si substrates, these high-mobility channels should help in fabricating very high-performing IC circuits, with added

functionalities and drastic reduction in power supply and power consumption, opening the door to the so-called "green" IC technology.

Acknowledgments We are indebted to our colleagues for their valuable contribution to this work: B. De Jaeger, A. Delabie, F. Bellenger, J. Mittard, F. Leys, B. Kaczer, T. Conard, M. Caymax, M. Meuris, G. Pourtois (IMEC), D. Brunco (Intel), G. Wilk (ASM), B. Yang (AMD), K.K. Ng (SRC), M. Hong (NTHU), Y. Xuan, H.C. Lin, Y.Q. Wu, T. Yang, T. Shen, M. Xu (Purdue), C.L. Hinkle, E.M. Vogel, and R.M. Wallace (UT Dallas). Stimulating discussions with Professor V.V. Afanas'ev, Professor A. Stesmans (KU Leuven), Professor J. Woodall, and Prof. M.S. Lundstrom (Purdue) are gratefully acknowledged. The work on Ge has been financially supported by the IMEC Industrial Affiliation Program on Ge and III–V devices, the European Project IST-ET4US-2048 "Epitaxial Technologies for Ultimate Scaling". The work on III–V has been supported by US National Science Foundation and SRC FCRP MSD Center.

References

1. G.D. Wilk, R.M. Wallace, J.M. Anthony, J. Appl. Phys. **89**, 5243 (2001)
2. M. Houssa (ed.), *High-κ Gate Dielectrics* (Institute of Physics Publishing, London, 2004)
3. H. Huff, D. Gilmer (eds.), *High-κGate Dielectric Materials for VLSI MOSFET Applications* (Springer, 2004)
4. M. Houssa, L. Pantisano, L.Å. Ragnarsson, R. Degraeve, T. Schram, G. Pourtois, S. De Gendt, G. Groeseneken, M.M. Heyns, Mater. Sci. Eng. R **51**, 37 (2006)
5. C. Claeys, E. Simoen (eds.), *Germanium Based Technologies: From Materials to Devices* (Elsevier, Oxford, 2007)
6. A. Dimoulas, E. Gusev, P.C. McIntyre, M. Heyns (eds.), *Advanced Gate Stacks for High-Mobility Semiconductors* (Springer, 2007)
7. R. Chau, B. Doyle, S. Datta, J. Kavalieros, K. Zhang, Nat. Mater. **6**, 810 (2007)
8. M. Passlack, M. Heyns, I. Thayne, *Compound Semiconductor*, May issue, vol 21 (2008)
9. H. Becke, R. Hall, J. White, Solid State Electron. **8**, 813 (1965)
10. E. Poindexter, P. Caplan, B. Deal, R. Razouk, J. Appl. Phys. **52**, 879 (1981)
11. A. Stesmans, V.V. Afanas'ev, J. Vac. Sci. Technol. B**16**, 3108 (1998)
12. C.W. Wilmsen (ed.) *Physics and Chemistry of III-V Compound Semiconductor Interfaces* (Plenum Press, New York, 1985)
13. V.V. Afanas'ev, Y.G. Fedorenko, A. Stesmans, Appl. Phys. Lett. **87**, 032107 (2005)
14. N. Wu, Q. Zhang, C. Zhu, D.S. Chan, M.F. Li, N. Balasubramanian, A. Chin, D.L. Kwong, Appl. Phys. Lett. **85**, 4127 (2004)
15. P. Zimmerman et al., IEDM Tech. Dig. (IEEE, Piscataway), 655 (2006)
16. T. Sugawara, Y. Oshima, R. Sreenivasan, P.C. McIntyre, Appl. Phys. Lett. **90**, 112912 (2007)
17. S. Takagi et al., Microelectron. Eng. **84**, 2314 (2007)
18. A. Delabie, F. Bellenger, M. Houssa, T. Conard, S. Van Elshocht, M. Caymax, M. Heyns, M. Meuris, Appl. Phys. Lett. **91**, 082904 (2007)
19. S.J. Whang, S.J. Lee, F. Gao, N. Wu, C.X. Zhu, J.S. Pan, L.J. Tang, D.L. Kwong, IEDM Tech. Dig. (IEEE, Piscataway), 307 (2004)
20. M. Houssa et al., Appl. Phys. Lett. **90**, 222105 (2007)
21. G. Mavrou, S. Galata, P. Tsipas, A. Sotiropoulos, Y. Panayiotatos, A. Dimoulas, E.K. Evangelou, J.W. Seo, Ch. Dieker, J. Appl. Phys. **103**, 014506 (2008)
22. E. Yablonovitch, C.J. Sandroff, R. Bhat, T. Gmitter, Appl. Phys. Lett. **51**, 439 (1987)
23. B.J. Skromme, C.J. Sandroff, E. Yablonovitch, T. Gmitter, Appl. Phys. Lett. **51**, 2022 (1987)

24. H. Hasegawa, M. Akazawa, K.I. Matsuzaki, H. Ishii, H. Ohno, Jpn. J. Appl. Phys, L**27**, L2265 (1988)
25. S. Tiwari, S.L. Wright, J. Batey, IEEE Electron Device Lett. **9**, 488 (1988)
26. G.G. Fountain, S.V. Hattangady, D.J. Vitkavage, R.A. Rudder, R.J. Markunas, Electron. Lett. **24**, 1134 (1988)
27. G.G. Fountain, R.A. Rudder, S.V. Hattangady, R.J. Markunas, J.A. Hutchby, IEDM Tech. Dig. (IEEE, Piscataway), 887 (1989)
28. M. Akazawa, H. Ishii, H. Hasegawa, Jpn. J. Appl. Phys. **30**, 3744 (1991)
29. D.S.L. Mui, H. Liaw, A.L. Demirel, S. Strite, H. Morkoc, Appl. Phys. Lett. **59**, 2847 (1991)
30. A. Callegari, D.K. Sadana, D.A. Buchanan, A. Paccagnella, E.D. Marshall, M.A. Tischler, M. Norcott, Appl. Phys. Lett. **58**, 2540 (1991)
31. S. Koveshnikov, W. Tsai, I. Ok, J.C. Lee, V. Torkanov, M. Yakimov, S. Oktyabrsky, Appl. Phys. Lett. **88**, 022106 (2006)
32. I. Ok, H. Kim, M. Zhang, C.Y. Kang, S.J. Rhee, C. Choi, S.A. Krishnan, T. Lee, F. Zhu, G. Thareja, J.C. Lee, IEEE Electron Device Lett. **27**, 145 (2006)
33. S.J. Koester, E.W. Kiewra, Y. Sun, D.A. Meumayer, J.A. Ott, M. Copel, D.K. Sadana, D.J. Webb, J. Fompeyrine, J.-P. Locquet, C. Marchiori, M. Sousa, R. Germann, Appl. Phys. Lett. **89**, 042104 (2006)
34. D. Shahrjerdi, M.M. Oye, A.L. Holmes, S.K. Banerjee, Appl. Phys. Lett. **89**, 043501 (2006)
35. H.C. Chin, M. Zhu, C.H. Tung, G.S. Samudra, Y.C. Yeo, IEEE Electron Device Lett. **29**, 553 (2008)
36. J.P. de Souza, E. Kiewra, Y. Sun, A. Callegari, D.K. Sadana, G. Shahidi, D.J. Webb, J. Fompeyrine, R. Germann, C. Rossel, C. Marchiori, Appl. Phys. Lett. **92**, 153508 (2008)
37. M. Passlack, M. Hong, J.P. Mannaerts, Appl. Phys. Lett. **68**, 1099 (1996)
38. M. Passlack, M. Hong, J.P. Mannaerts, R.L. Opila, S.N.G. Chu, N. Moriya, F. Ren, J.R. Kwo, IEEE Trans. Electron Devices **44**, 214 (1997)
39. M. Hong, J. Kwo, A.R. Kortan, J.P. Mannaerts, A.M. Sergent, Science **283**, 1897 (1999)
40. F. Ren, M. Hong, W.S. Hobson, J.M. Kuo, J.R. Lothian, J.P. Mannaerts, J. Kwo, S.N.G. Chu, Y.K. Chen, A.Y. Cho, Solid State Electron. **41**, 1751 (1997)
41. Y.C. Wang, M. Hong, J.M. Kuo, J.P. Mannaerts, J. Kwo, H.S. Tsai, J.J. Krajewski, Y.K. Chen, A.Y. Cho, IEEE Electron Device Lett. **20**, 457 (1999)
42. F. Ren, J.M. Kuo, M. Hong, W.S. Hobson, J.R. Lothian, J. Lin, W.S. Tseng, J.P. Mannaerts, J. Kwo, S.N.G. Chu, Y.K. Chen, A.Y. Cho, IEEE Electron Device Lett. **19**, 309 (1998)
43. M. Hong, J.N. Bailargeon, J. Kwo, J.P. Mannaerts, A.Y. Cho, in *Proceeding of IEEE International Symposium on Compound Semiconductors*, pp. 345–350 (2000)
44. Y.C. Wang, M. Hong, J.M. Kuo, J.P. Mannaerts, H.S. Tsai, J. Kwo, J.J. Krajewski, Y.K. Chen, A.Y. Cho, Electron. Lett. **35**, 667 (1999)
45. B. Yang, P.D. Ye, J. Kwo, M. Frei, H. Gossmann, J.P. Mannaerts, M. Sergent, M. Hong, K. Ng, J. Bude, J. Cryst. Growth **251**, 837 (2003)
46. T.D. Lin, H.C. Chiu, P. Chang, L.T. Tung, C.P. Chen, M. Hong, J. Kwo, W. Tsai, Y.C. Wang, Appl. Phys. Lett. **93**, 033516 (2008)
47. M. Passlack, N. Medendorp, R. Gregory, D. Braddock, Appl. Phys. Lett. **83**, 5262 (2003)
48. K. Rajagopalan, J. Abrokwah, R. Droopad, M. Passlack, IEEE Electron Device Lett. **27**, 959 (2006)
49. R.J.W. Hill, D.A.J. Moran, X. Li, H. Zhou, D. Macintyre, S. Thoms, A. Asenov, P. Zurcher, K. Rajagopalan, J. Abrokwah, R. Droopad, M. Passlack, I.G. Thayne, IEEE Electron Device Lett. **28**, 1080 (2007)
50. P.D. Ye, G.D. Wilk, J. Kwo, B. Yang, H.-J.L. Gossmann, M. Frei, S.N.G. Chu, J.P. Mannaerts, M. Sergent, M. Hong, K. Ng, J. Bude, IEEE Electron Devices Lett. **24**, 209 (2003)
51. P.D. Ye, G.D. Wilk, B. Yang, J. Kwo, H.-J.L. Gossmann, S.N.G. Chu, S. Nakahara, J.P. Mannaerts, M. Sergent, M. Hong, K. Ng, J. Bude, Appl. Phys. Lett. **83**, 180 (2003)
52. P.D. Ye, G.D. Wilk, B. Yang, J. Kwo, H.-J.L. Gossmann, M. Hong, K. Ng, J. Bude, Appl. Phys. Lett. **84**, 434 (2004)

53. P.D. Ye, G.D. Wilk, B. Yang, S.N.G. Chu, K.K. Ng, J. Bude, Solid State Electron. **49**, 790 (2005)
54. P.D. Ye, B. Yang, K. Ng, J. Bude, G.D. Wilk, S. Halder, J.C.M. Hwang, Appl. Phys. Lett. **86**, 063501 (2005)
55. D.P. Norton, J.D. Budai, M.F. Chisholm, Appl. Phys. Lett. **76**, 1677 (2000)
56. C.O. Chui, H. Kim, D. Chi, B.B. Triplett, P.C. McIntyre, K.C. Saraswat, IEDM Tech. Dig. (IEEE, Piscataway), 437 (2002)
57. H. Kim, C.O. Chui, K. Saraswat, P.C. McIntyre, Appl. Phys. Lett. **83**, 2647 (2003)
58. H. Shang, H. Okorn-Schmidt, J. Ott, P. Kozlowski, S. Steen, E.C. Jones, H.S. Wong, W. Hanesch, IEEE Electron Device Lett. **24**, 242 (2003)
59. A. Ritenour, S. Yu, M.L. Lee, N. Lu, W. Bai, A. Pitera, E.A. Fitzgerald, D.L. Kwong, D. Antoniadis, IEDM Tech. Dig. (IEEE, Piscataway), 433 (2003)
60. W.P. Bai, N. Lu, J. Liu, A. Ramirez, D.L. Kwong, D. Wristers, A. Ritenour, L. Lee, D. Antoniadis, VLSI Tech. Dig. (IEEE, Piscataway), 121 (2003)
61. E.J. Preisler, S. Guha, B.R. Perkins, D. Kazazis, A. Zaslavsky, Appl. Phys. Lett. **86**, 223504 (2005)
62. M. Caymax et al., Mater. Sci. Eng. B **135**, 256 (2006)
63. M. Houssa, G. Pourtois, M. Caymax, M. Meuris, M.M. Heyns, Surf. Sci. **602**, L25 (2008)
64. M. Houssa, G. Pourtois, M. Caymax, M. Meuris, M.M. Heyns, Appl. Phys. Lett. **92**, 242101 (2008)
65. A. Dimoulas, G. Mavrou, G. Vellianitis, E. Evangelou, N. Bukos, M. Houssa, M. Caymax, Appl. Phys. Lett. **86**, 032908 (2005)
66. B. Kaczer, B. De Jaeger, G. Nicholas, K. Martens, R. Degraeve, M. Houssa, F. Leys, M. Meuris, G. Groeseneken, Microelectron. Eng. **84**, 2067 (2007)
67. C.G. Van de Walle, R.M. Martin, Phys. Rev. B **34**, 5621 (1986)
68. D.P. Brunco et al., J. Electrochem. Soc. **155**, H552 (2008)
69. G. Pourtois, M. Houssa, B. De Jaeger, B. Kaczer, F. Leys, M. Meuris, M. Caymax, G. Groeseneken, M.M. Heyns, Appl. Phys. Lett. **91**, 023506 (2007)
70. F.E. Leys et al., Mater. Sci. Semicond. Process. **9**, 679 (2006)
71. K. Prabhakaran, F. Maeda, Y. Watanabe, T. Ogino, Appl. Phys. Lett. **76**, 2244 (2000)
72. T. Takahashi, T. Nishimura, L. Chen, S. Sakata, K. Kita, and A. Toriumi, IEDM Tech. Dig. (IEEE, Piscataway), 697 (2007)
73. A. Molle, M.D. Nurul Kabir Bhuiyan, G. Tallarida, M. Fanciulli, Appl. Phys. Lett. **89**, 083504 (2006)
74. F. Bellenger, M. Houssa, A. Delabie, V.V. Afanas'ev, T. Conard, M. Caymax, M. Meuris, K. De Meyer, M.M. Heyns, J. Electrochem. Soc. **155**, G33 (2008)
75. P. Batude et al., J. Appl. Phys. **102**, 034514 (2007)
76. V.V. Afanas'ev, A. Stesmans, A. Delabie, F. Bellenger, M. Houssa, M. Meuris, Appl. Phys. Lett. **92**, 022109 (2008)
77. G. Mavrou, S. Galata, P. Tsipas, A. Sotiropoulos, Y. Panayiotatos, A. Dimoulas, E.K. Evangelou, J.W. Seo, Ch. Dieker, J. Appl. Phys. **103**, 014506 (2008)
78. D.P. Brunco et al., J. Appl. Phys. **102**, 024104 (2007)
79. G. Nicholas et al., IEEE Trans. Electron Devices **54**, 1425 (2007)
80. A. Dimoulas, *Presented at the E-MRS Spring Meeting* (Strasbourg, France, 2008)
81. R. Chau, S. Datta, A. Majumdar, Proc. IEEE CSIC Dig (IEEE, Piscataway), 17 (2005)
82. S. Datta, T. Ashley, J. Brask, L. Buckle, M. Doczy, M. Emeny, D. Hayes, K. Hilton, R. Jefferies, T. Martin, T.J. Phillips, D. Wallis, P. Wilding, R. Chau, IEDM Tech. Dig. (IEEE, Piscataway), 763 (2005)
83. Y. Xuan, H.C. Lin, P.D. Ye, IEEE Trans. Electron Devices **54**, 1811 (2007)
84. M.M. Frank, G.D. Wilk, D. Starodub, T. Gustafsson, E. Garfunkel, Y.J. Chabal, J. Grazul, D.A. Muller, Appl. Phys. Lett. **86**, 152904 (2005)
85. M.L. Huang, Y.C. Chang, C.H. Chang, Y.J. Lee, P. Chang, J. Kwo, T.B. Wu, M. Hong, Appl. Phys. Lett. **87**, 252104 (2005)

86. C.H. Chang, Y.K. Chiou, Y.C. Chang, K.Y. Lee, T.D. Lin, T.B. Wu, M. Hong, J. Kwo, Appl. Phys. Lett. **89**, 242911 (2006)
87. C.L. Hinkle, A.M. Sonnet, E.M. Vogel, S. McDonnell, G.J. Hughes, M. Milojevic, B. Lee, F.S. Aguirre-Tostado, K.J. Choi, H.C. Kim, J. Kim, R.M. Wallace, Appl. Phys. Lett. **92**, 071901 (2008)
88. G.K. Dalapati, Y. Tong, W.Y. Loh, H.K. Mun, B.J. Cho, IEEE Trans. Electron Devices **54**, 1831 (2007)
89. D. Shahrjerdi, E. Tutuc, S.K. Banerjee, Appl. Phys. Lett. **91**, 063501 (2007)
90. H.C. Lin, T. Yang, H. Sharifi, S.K. Kim, Y. Xuan, T. Shen, S. Mohammadi, P.D. Ye, Appl. Phys. Lett. **91**, 212101 (2007)
91. Y. Xuan, Y.Q. Wu, H.C. Lin, T. Shen, P.D. Ye, IEEE Electron Device Lett. **28**, 935 (2007)
92. Y. Xuan, Y.Q. Wu, T. Shen, T. Yang, P.D. Ye, IEDM Tech. Dig. (IEEE, Piscataway), 637 (2007)
93. Y. Xuan, Y.Q. Wu, P.D. Ye, IEEE Electron Device Lett. **29**, 294 (2008)
94. Y. Xuan, T. Shen, Y.Q. Wu, M. Xu, and P.D. Ye, *in proceeding of IEEE 66th Device Research Conference*, p.37 (2008)
95. Y.Q. Wu, W.K. Wang, O. Koybasi, D.N. Zakharov, E. Stach, S. Nakahara, J.C.M. Hwang, P.D. Ye, IEEE Electron Device Lett. **30**, 700 (2009)
96. Y. Xuan, P.D. Ye, T. Shen, Appl. Phys. Lett. **91**, 232107 (2007)
97. P.D. Ye, J. Vac. Sci. Technol., A **26**, 697 (2008)
98. Y.Q. Wu, Y. Xuan, T. Shen, P.D. Ye, Z. Cheng, A. Lochtefeld, Appl. Phys. Lett. **91**, 022108 (2007)
99. G. Brammertz, K. Martens, S. Sioncke, A. Delabie, M. Caymax, M. Meuris, M. Heyns, Appl. Phys. Lett. **91**, 133510 (2007)
100. Y. Xuan, T. Shen, M. Xu, Y.Q. Wu, P.D. Ye, IEDM Tech. Dig. (IEEE, Piscataway) (2008)
101. D. Varghese, Y. Xuan, Y.Q. Wu, T. Shen, P.D. Ye, M.A. Alam, IEDM Tech. Dig. (IEEE, Piscataway) (2008)
102. G. Brammertz, M. Heyns, M. Meuris, M. Caymax, D. Jiang, Y. Mols, S. Degroote, M. Leys, G. Borghs, Appl. Phys. Lett. **90**, 134102 (2007)
103. S.R. Hofstein, G. Warfield, Solid State Electron. **8**, 321 (1965)
104. Y. Xuan, H.C. Lin, P.D. Ye, ECS Trans. **3**, 59 (2006)
105. T. Yang, Y. Liu, P.D. Ye, Y. Xuan, H. Pal, M.S. Lundstrom, Appl. Phys. Lett. **92**, 252105 (2008)
106. G. Brammertz, K. Martens, H.C. Lin, C. Merckling, J. Penaud, C. Adelmann, S. Sioncke, W.E. Wang, M. Caymax, M. Meuris, M. Heyns, Material Science in Semiconductor Processing, to be published
107. M. Xu, Y.Q. Wu and P.D. Ye, Presented at *the IEEE Semiconductor Interface Specialists Conference* (San Diego, 2008)
108. A. Dimoulas, P. Tsipas, A. Sotiropoulos, E.K. Evangelou, Appl. Phys. Lett. **89**, 252110 (2006)

Appendices

Appendix I: Fundamental Constants

We have included those fundamental constants which are frequently used when dealing with MOSFETs, IC technology, and dielectric materials.

The values of fundamental constants quoted are accurate.

Constant	Symbol	Value	Unit	Remarks
Avogadro constant	N_A	$6.022\ 141\ 79 \times 10^{23}$	mol^{-1}	
Boltzmann constant	k	$1.380\ 6504 \times 10^{-23}$	JK^{-1}	
Electron mass	m	$9.109\ 382\ 15 \times 10^{-31}$	kg	
Elementary charge	e, q	$1.602\ 176\ 487 \times 10^{-19}$	C	
Permeability of vacuum (magnetic constant)	μ_0	$12.566\ 370\ 614 \times 10^{-7}$	NA^{-2}	Exact
Permittivity of vacuum (electric constant)	ε_0	$8.854\ 187\ 817 \times 10^{-12}$	Fm^{-1}	Exact
Planck constant	h	$6.626\ 068\ 96 \times 10^{-34}$	Js	
		$4.135\ 667\ 33 \times 10^{-15}$	eVs	
	$h/2\pi$	$1.054\ 571\ 628 \times 10^{-34}$	Js	
		$6.582\ 118\ 99 \times 10^{-16}$	eVs	
Speed of light in vacuum	c	299 792 458	ms^{-1}	Exact

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5,

Appendix II: Periodic Table of the Elements

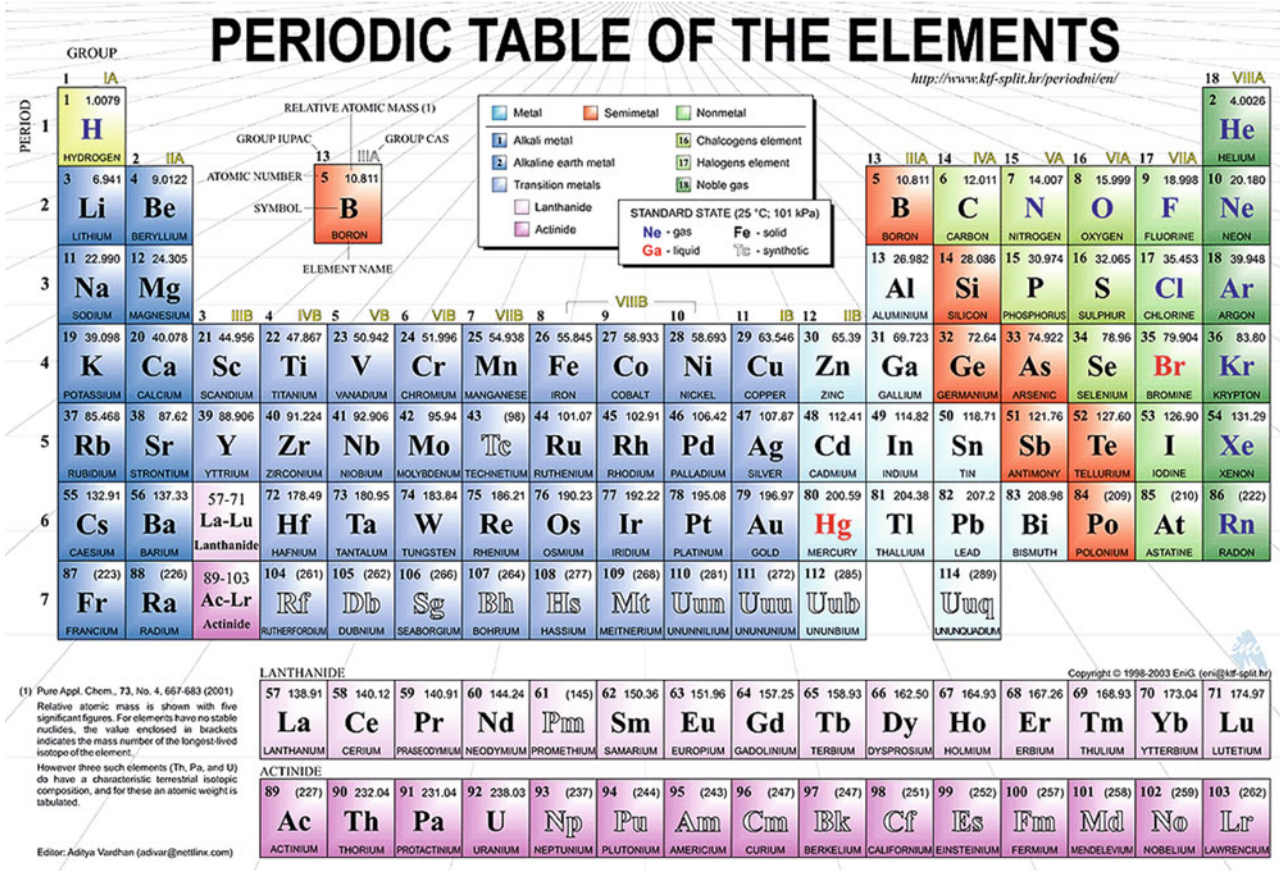
A visual display of the periodic table can greatly help in the study of and search for the high permittivity materials for the gate dielectrics, high mobility semiconductors for the MOSFET channel, and high and low work function metals for the gate electrode. A thorough understanding of and familiarity with the numerous salient features of the periodic table could inspire the appropriate ideas for research.

Notes and Comments

1. There are labels and notes on the periodic table explaining the features and the arrangement of the periodic table.
2. There are two nomenclatures in use for the designation of the groups: the classical Chemical Abstracts Service (CAS) group number and the more recent International Union of Pure and Applied Chemistry (IUPAC) group number.
3. A guide in the search of elements for the semiconductor is the four electrons per average atom criterion. This criterion leads to the elements of group IVA (elemental semiconductors), groups IVA and IVA (IV-IV compound semiconductors), groups IIIA and VA (III-V compound semiconductors), groups IIB and VIA (II-VI compound semiconductors), groups IB, IIIA, and VIA (I-III-VI_2 compound semiconductors), and groups IIB, IVA, and VA (II-IV-V_2 compound semiconductors), cf. Appendix III. There are exceptions to this criterion, e.g. the IV-VI compound semiconductors, cf. Appendix III.
4. Among group IVA natural elements, C, Si, and Ge are semiconductors, Sn is a semi-metal, whereas Pb is a metal.
5. The genesis of the criterion of an average valency of four is the fact that semiconductors are covalent or largely covalent material; a valency of four can be linked to covalent bonds, unless the electro-negativity difference is too high. The I-VII compounds are not semiconductors because these are ionic or largely ionic materials on account of the large difference between the anion and the cation electro-negativity.
6. In contrast to the search for the semiconductor behavior, we need to look for ionic oxides/insulators when searching for the high-k gate dielectrics. The genesis behind the search for high ionicity is the following. At the MOSFET operating frequency of a few GHz, only electronic and ionic polarizations count; as the former is not large enough even for the heavy elements, one has mainly to depend upon the ionic polarization to contribute to a high value of the dielectric constant k. Generally, large electro-negativity difference may lead to large ionic polarization, although the latter also depends upon the molecular volume and the phonon frequency. In a high-k material, the anion is generally oxygen, which has a Pauling electro-negativity of 3.44 (one of the highest values of electro-negativity), cf. Appendix VI. This means that to enhance the electro-negativity difference, we have to look for a cation (metals) with a small

value of the electro-negativity; in other words an element from the left side of the periodic table—from groups IIIB, IVB, and VB, including the lanthanides, cf. Appendix VI.

7. A guide in the search for the gate electrode metal is the value of the vacuum work function of the metal; however, the vacuum work function is less relevant on the high-k gate stack than on the SiO_2 gate dielectric, because of the very high trap density at the high-k/metal interface and the resultant Fermi level pinning. Still the vacuum work function of about 4.0 eV for the n-channel metal and of about 5.0 eV for the p-channel metal can be a starting point. There appears to be some correlation between the vacuum work function of the metal and the position of the metal on the periodic table. The vacuum work function tends to increase going from the left to the right columns of metals in the periodic table cf. Appendix VII.



Source for the periodic table of the elements: http://www.periodni.com

Appendix III: Physical Constants of Semiconductors

Notes

1. The values in the table are mostly those which obtain at 300 K and in the intrinsic semiconductor.
2. The values have been taken from multiple sources; a list of some important sources is given under the references.
3. Where two values of the electron effective mass occur, the first represents the longitudinal and the second the transverse electron effective mass. Where two values of the hole effective mass occur, the first represents the light and the second the heavy hole effective mass.
4. Where two values of the lattice constant occur, the first represents lattice constant a and the second the lattice constant c.
5. Many semiconductors may crystallize in more than one atomic arrangement. For example, many III-V compound semiconductors may solidify with either the zinc blende or the wurtzite crystal structure. We have presented values for the crystal structure with the more reliable and complete values.

Observations

1. The most common crystal structure among the semiconductors is the tetrahedral crystal structure with four nearest neighbors. This promotes covalent chemical bonding. The tetrahedral crystal structure is equivalent to two face-centered-cubic (fcc) structures displaced from each other by one–fourth of its body diagonal.
2. The covalent bond is strongly directional in nature; this does not promote close packing; hence most of the semiconductors have low packing density and expand on solidification.
3. The average valency of most of the semiconductors is four. This is true for the group IV elemental, group III-V compound, group II-VI compound, group I-III-V_2 compound, and group II-IV-VI_2 compound semiconductors.
4. Compound semiconductors are not completely covalent, but are partly ionic. The ionicity in general increases going from group IV elemental to group IV-IV compound to III-V compound to II-VI compound semiconductors.
5. The defect density increases going from group IV elemental to group IV-IV compound to group III-V compound to group II-VI compound to group II-IV-V_2 compound to group I-III-VI_2 compound semiconductors; the doping efficacy decreases in this order because of a variety of reasons including the self-compensation effect. Silicon is one of the easiest semiconductors to dope. In contrast, many of the II-VI compound semiconductors can be doped either p- or n-type but not both.

Semiconductor	Band gap (eV)	Gap transition	Mobility (cm^2 V^{-1} s^{-1})		Effective mass		Electron affinity (eV)	Dielectric constant	Crystal structure	Lattice constant (Å)	Melting point (°C)
			μ_e	μ_h	m_e^*/m	m_h^*/m					
Group IV elements											
C (Diamond)	5.46–5.60	Indirect	1,800–2,200	1,600–1,800	1.40/0.36	0.70/2.12		5.5–5.7	Diamond	3.56–3.57	3,800
Si	1.11–1.12	Indirect	1,350–1,500	450-600	0.97–0.98/ 0.19	0.16/0.49–0.50	4.01–4.24	11.7–12.0	Diamond	5.43	1,410–1,420
Ge	0.66–0.67	Indirect	3,600–3,900	1,800–1,900	1.60-1.64/ 0.08	0.04/0.30–0.33	4.00–4.13	16.0–16.3	Diamond	5.64-5.66	937–958
Group IV-IV compounds											
4H-SiC	2.99–3.23	Indirect	400-900	50–120	0.29/0.42			10.0–10.2	Wurtzite	3.07/10.05	2,830
3C-SiC	2.36	Indirect	800	320	0.68/0.25	1.00–1.20		9.66–9.72	Zinc Blende	4.36	2,830
Group III-V compounds											
BN	6.10–6.40	Indirect	200	500	0.35/0.24	0.15/0.38	4.50	7.1	Zinc Blende	3.62	2,973
AlN	6.02	Direct	300	14	0.40	0.24/3.53	0.60	8.5–9.1	Wurtzite	3.11/4.98	3,000
AlP	2.45–3.00	Indirect	80					9.8–11.6	Zinc Blende	5.46	1,500–2,000
AlAs	2.11–2.16	Indirect	180–280				3.50	8.5–12.0	Zinc Blende	5.66	1,600
AlSb	1.52–1.65	Indirect	200-900	400–420	0.30–0.39	0.40	3.65	10.1–14.4	Zinc Blende	6.13-6.14	1,050
GaN	3.25–3.50	Direct	1,000	200	0.19–0.20	0.30/1.40	4.10	8.9–12.2	Wurtzite	3.16–3.18/ 5.16–5.18	1,500
GaP	2.24–2.26	Indirect	110–300	75–150	0.12–0.35	0.14/0.79	3.80–4.30	8.5–11.1	Zinc Blende	5.45	1,465
GaAs	1.42–1.43	Direct	8,500–8,600	250–420	0.06–0.07	0.08–0.12/ 0.51–0.68	4.07	10.9–13.1	Zinc Blende	5.63–5.65	1,238
GaSb	0.67–0.78	Direct	3,000–5,000	850–1,400	0.04–0.05	0.05–0.06/ 0.30–0.50	4.06	14.0–15.7	Zinc Blende	6.09–6.10	712
InN	2.05–2.40	Direct	3200		0.11	0.27/1.63	4.38	15.3–19.3	Wurtzite	3.53/5.69	1,100
InP	1.27–1.35	Direct	4,500–4,600	100–150	0.07–0.08	0.08/0.40–0.60	4.38–4.40	9.6–12.4	Zinc Blende	5.67–5.87	1,070
InAs	0.33–0.36	Direct	30,000–40,000	450–500	0.03	0.03/0.41	4.90–4.96	12.3–15.1	Zinc Blende	6.05–6.06	943
InSb	0.16–0.18	Direct	77,000–80,000	450–1,250	0.01	0.02/0.39–0.43	4.59	15.9–17.9	Zinc Blende	6.47–6.48	536

(continued)

(continued)

Semiconductor	Band gap (eV)	Gap transition	Mobility ($cm^2\ V^{-1}\ s^{-1}$)		Effective mass		Electron affinity (eV)	Dielectric constant	Crystal structure	Lattice constant (Å)	Melting point (°C)
			μ_e	μ_h	m_e^*/m	m_h^*/m					
Group II-VI Compounds											
ZnS	3.58–3.68	Direct	120-165	5	0.25	0.50–1.00	3.90	8.0–8.3	Zinc Blende	5.41–5.42	1,830
CdS	2.42	Direct	340–400	9–50	0.10–0.17	0.40–0.80	4.50	8.3–9.0	Zinc Blende	5.82–5.83	1,750
HgS	2.00	Direct	250					30.7	Zinc Blende	5.85	1,450
ZnSe	2.60–2.67	Direct	530	16	0.15–0.17	0.60	4.09	8.1–9.1	Zinc Blende	5.67	1,515
CdSe	1.70–1.80	Direct	600–800	9–10	0.13	0.45	4.95	10.0–10.7	Zinc Blende	6.05	1,258
CdSe	1.70–1.74	Direct							Wurtzite	4.30/7.01	1,350
HgSe	0.60	Direct	20,000		0.04–0.05	0.02–0.08		25.6–25.8	Zinc Blende	6.08	800
ZnTe	2.26	Direct	530	130–900	0.20	0.10–0.30	3.50–4.80	10.1	Zinc Blende	6.10	1,295
CdTe	1.44–1.56	Direct	300–1,050	65–100	0.14	0.37	4.28–4.30	9.6–11.0	Zinc Blende	6.48	1,098
HgTe	0.15	Direct	22,900		0.02	0.003		48.0	Zinc Blende	6.46–6.52	670
Group IV-VI Compounds											
PbS	0.41	Indirect	600	700	0.66	0.50		205/17	NaCl	5.93–5.94	1,077
PbSe	0.27	Direct	900	700	0.33	0.34		280	NaCl	6.15	1,062
PbTe	0.29–0.32	Indirect	2,500–6,000	1,000–4,000	0.22	0.29		400/30	NaCl	6.46	904
Group I-III-VI_2 Compounds											
$CuInSe_2$	1.00–1.04	Direct	320	10			4.15		Chalcopyrite	5.77/11.54	990
$CuInTe_2$	0.95	Direct	200	12					Chalcopyrite	6.16/12.32	780
Group II-IV-V_2 Compounds											
$CdSnP_2$	1.17	Direct	2,000		0.04				Chalcopyrite	5.90/11.51	570
$CdSnAs_2$	0.26		11,000						Chalcopyrite	6.08/11.91	593–595

6. Doping ease correlates well with covalency.
7. For the semiconductors with the same class of crystal structure, the band-gap in an approximate manner is inversely proportional to the lattice constant. Hence, on an average, smaller lattice constant correlates with higher band-gap, stronger bond, and higher melting point.
8. Mixed semiconductors can be formed easily, particularly in the case of the III-V compound semiconductors as the rules of solid solutions are well satisfied for many combinations.
9. Nearly equal values of electron and hole mobility is rare in semiconductors. Only diamond and PbTe are among the exceptions with reasonable values of both electron and hole mobility. This rare property makes these two semiconductors potentially ideal channel materials for CMOSFETs.
10. Most semiconductors have disappointingly low values of the hole mobility. Diamond, Ge, GaSb, and PbTe are among the rare semiconductors with reasonable values of hole mobility.
11. Some III-V compound semiconductors and a few other compound semiconductors (InSb, InAs, HgTe, HgSe, $CdSnAs_2$, GaAs) have enormous values of electron mobility. Equally striking are their disproportionately low values of hole mobility. While InSb has the highest value of electron mobility, GaAs—a leading candidate for high mobility MOSFET channel—has much lower electron mobility than the other semiconductors with the highest electron mobility.
12. For most semiconductors, the value of hole mobility is lower than its value of electron mobility. There are some exceptions to this rule; among the exceptions are ZnTe and PbS in which the electron mobility is lower than the hole mobility.
13. PbTe is a rare semiconductor with high values of both electron and hole mobilities—6,000 and 4,000 $cm^2\ V^{-1}\ s^{-1}$ respectively.
14. No clear correlation can be observed between electron mobility and crystalline perfection including its chemical purity and absence of mechanical imperfections. IC grade silicon is the most perfect crystal available today, yet values of its electron and hole mobility are among the lower values.
15. Silicon has been and still is by far the leading CMOSFET channel material, but values of its electron and hole mobility remain perhaps its most potent weakness for this application.
16. Carrier mobility is proportional to the mean free time and inversely proportional to the effective mass. Many semiconductors with high electron mobility exhibit low effective mass for electrons; but the same semiconductors also exhibit low effective mass for holes as well; yet values of hole mobility are abysmally low.

References

1. A.G. Milnes, D.L. Feucht, *Heterojunctions and Metal-Semiconductor Junctions* (Academic Press, New York, 1972)
2. B.L. Sharma, R.K. Purohit, *Semiconductor Heterojunctions* (Pergamon Press, Elmsford, New York, 1974)
3. C. Hilsum, A.C. Rose-Innes, *Semiconducting III-V Compounds* (Pergamon Press, Elmsford, New York, 1961)
4. F.F.Y. Wang, *Introduction to Solid State Electronics* (North Holland, Amsterdam, 1980)
5. H.F. Wolf, *Semiconductors* (Wiley, New York, 1971)
6. J.L. Shay, J.H. Wernick, *Ternary Chalcopyrite Semiconductors: Growth, Electrical Properties, and Applications* (Pergamon Press, Elmsford, New York, 1975)
7. M. Aven, J.S. Prerier, *Physics and Chemistry of II-VI Compounds* (North-Holland, Amsterdam, 1967)
8. M. Lavinstein, S. Rumyantsev, M. Shur (eds.), *Handbook Series on Semiconductor Parameters* (World Scientific, London, 1999)
9. O. Madelung, *Physics of III-V Compounds* (Wiley, New York, 1964)
10. R.K. Willardson, A.C. Beer (eds.), *Semiconductors and Semimetals* (Academic Press, New York, 1966)
11. S.J. Fonash, *Solar Cell Device Physics* (Academic Press, New York, 1981)

Appendix IV: Physical Constants of Si, Ge, GaAs (Values at 300 K)

Physical constant (unit)	Ge	Si	GaAs	Remarks
Acceptors	B, Al, Ga, In	B, Al, Ga, In	Zn, Si, C, Ge, Sn	
Atomic density (cm^{-3})	4.42×10^{22}	5.00×10^{22}	4.42×10^{22}	
Atomic number	32	14	31, 33	
Atomic weight	72.60	28.09	144.63	
Avalanche (breakdown) field ($V\ cm^{-1}$)	2×10^5	3×10^5	3.5×10^5	
Band-gap—direct (eV) E_G	0.81	2.5	1.43	
Band-gap—indirect (eV) E_G	0.66	1.11	—	
Bulk modulus ($N\ cm^{-2}$)	7.5×10^6	9.8×10^6	7.53×10^6	
Cleavage plane	{001}	{111}	{110}	
Crystal structure	Diamond	Diamond	Zinc blende	
Density ($g\ cm^{-3}$)	5.327	2.329	5.317	
Dielectric constant k	16.3	11.7	11.5	
Diffusion coefficient ($cm^2\ s^{-1}$): electrons	93	35	200	
Diffusion coefficient ($cm^2\ s^{-1}$): holes	47	12.5	8	
Donors	P, As, Sb	P, As, Sb	Te, Si, Ge, S, Sn, Se	
Effective density of states in the conduction band (cm^{-3}) N_c	1.04×10^{19}	2.8×10^{19}	4.7×10^{17}	
Effective density of states in the valence band (cm^{-3}) N_v	6.0×10^{18}	1.04×10^{19}	7.0×10^{18}	
Effective mass (m_0): holes m_{hh}*, m_{lh}*	0.33, 0.043	0.49, 0.16	0.51, 0.082	
Effective mass (m_0): electrons m_l*, m_t*	1.6, 0.08	0.98, 0.19	0.07	
Electron affinity (eV) χ_s	4.13	4.05	4.07	
Hardness (Mohs scale)	6	7	4-5	
Intrinsic carrier density (cm^{-3}) n_i	2.4×10^{13}	1.45×10^{10}	9×10^6	
Intrinsic resistivity (Ω cm)	50	2.5×10^5	3.3×10^8	
Lattice constant (Å) a	5.658	5.431	5.654	

(continued)

(continued)

Physical constant (unit)	Ge	Si	GaAs	Remarks
Melting point (°C)	937	1,412	1,238	
Mobility ($cm^2\ V^{-1}\ s^{-1}$): electrons μ_e	3,900	1,350	8,600	
Mobility ($cm^2\ V^{-1}\ s^{-1}$): holes μ_h	1,900	480	250	
Reflectivity (%)	48	35	30	Normal incidence
Refractive index n	4.0	3.42	3.3	
Specific heat ($J\ g^{-1}\ K^{-1}$)	0.31	0.7	0.35	
Thermal coefficient (K^{-1})	5.75×10^{-6}	2.33×10^{-6}	5.73×10^{-6}	
Thermal conductivity ($W\ cm^{-1}\ K^{-1}$)	0.58	1.5	0.55	
Thermal velocity ($cm\ s^{-1}$): electrons	3.1×10^7	2.3×10^7	4.4×10^7	
Thermal velocity ($cm\ s^{-1}$): holes	1.9×10^7	1.65×10^7	1.8×10^7	
Vapor pressure (Pa)	1 (1,330 °C)	5×10^{-3} (1,300 °C)	1 (900 °C)	

Appendix V: Physical Constants of High Permittivity Dielectrics

Please see Sect. 1.4 for observations on the data presented below.

Table A Experimental values of important material constants of various high-k dielectrics

Material	Si	Al_2O_3	SiO_2	TiO_2	Y_2O_3	ZrO_2
Atomic number	14	13–8	14–8	22–8	39–8	40–8
Cation valency		3	4	4	3	4
Crystal structure (Alternate structure)	Diamond	Hexagonal	Cubic (β Cristobalite)	Tetragonal (Rutile: A; B)	Cubic	Monoclinic (T; C)
Ionicity, I (Pauling)	0	0.57	0.45	0.59	0.71	0.67
Coordination number, CN_c:CN_a	4	6:4	4:2	6:3	6:4	7:3 or 4
Cation-anion distance, $d_{c\text{-}a}$ (Å)	2.35(4X)	1.86(3X) 1.97(3X)	1.61(4X)	1.932(4X) 1.979(2X)	2.25(2X), 2.28(2X) 2.34(2X)	2.051, 2.057, 2.151, 2.163, 2.189, 2.220, 2.285
Specific density, ρ (g cm^{-3})	2.33	3.9–4.1	2.2–2.3	3.9–4.3	4.8–5.0	5.7–5.9
Melting point, m.p. (°C)	1,414	2,020–2,072	1,670–1,728	1,830–1,850	2,376–2,464	2,670–2,710
Thermal exp. coeff., α ($\times 10^{-6}$ K^{-1})	2.5	6.5-8.8	0.5–0.6	7.1–9.2	7.9–9.3	7.0–8.0
Bandgap, E_G (eV)	1.12	8.6–8.8	8.9–9.3	3.1–3.5	5.5–6.0	5.4–5.8
Electron affinity, χ_{di} (eV)	4.05	1.2–2.0	0.9–1.1	3.0–3.2	1.8–2.0	2.0–2.5
Effective mass m_h^*/m	0.59	—-	3–10	0.3–0.5	—-	4
m_e^*/m	1.06	0.35-0.50	0.3–0.5	1.3–13.0	—-	0.3
Dielectric constant, k (frequency range, Hz)	11.9 (10^7–10^9)	9.0–10.0 (10^2–10^{11})	3.8–3.9 (10^{-5}–10^{11})	89–173 (10^2–10^8)	14.0–18.0 (10^3–10^7)	14–18 (10^2–10^{11})

(continued)

Table A (continued)

Material	Si	Al_2O_3	SiO_2	TiO_2	Y_2O_3	ZrO_2
Refractive index, n	3.44	1.6-1.7	1.475	2.3–2.7	1.7–1.9	2.1–2.2
n^2	11.83	2.89	2.22	6.40	3.28	4.71
Relaxation frequencies (non-electronic) (Hz)	—-	Nil	Nil	10^2, 10^3, 10^{10}	—-	10, 10^4

In a majority of cases, for the values of each constant, five independent references were used. The frequency range given in the row for the dielectric constant k, indicates the range over which the measurements were made and the values of k were found to be frequency-independent. A ≡ Anatase(T); B ≡ Brookite(O); C ≡ Cubic; H ≡ Hexagonal; M ≡ Monoclinic; O ≡ Orthorhombic; T ≡ Tetragonal

Table B Experimental values of important material constants of various high-k dielectrics

Material	Si	$ZrSiO_4$	$BaZrO_3$	La_2O_3	HfO_2	Ta_2O_5
Atomic number	14	40-14-8	56-40-8	57-8	72-8	73-8
Cation valency		4;4	2;4	3	4	5
Crystal structure (Alternate structure)	Diamond	Tetragonal (b.c.)	Perovskite	Hexagonal (M;C)	Monoclinic (T;C)	Orthorhombic (H)
Ionicity, I (Pauling)	0	0.57	0.74	0.75	0.68	0.61
Coordination number, CN_c:CN_a	4	(8:2); (4:1)	(12:4); (6:2)	7:3 or 4	7:3 or 4	6:2 or 3
Cation-anion distance, $d_{c\text{-}a}$ (Å)	2.35(4X)	Zr-O: 2.10(4X), 2.24(4X), Si-O: 1.61(4X)	Ba-O: 2.957, Zr-O: 2.094	2.38(3X), 2.45(1X) 2.72(3X)	2.031, 2.052, 2.162, 2.170, 2.174, 2.202, 2.254	2.60, 1.98, 2.13, 1.92, 2.14, 1.97
Specific density, ρ (g cm^{-3})	2.33	4.0–4.9	5.3–6.3	6.5–6.6	9.7–10.0	7.5–8.3
Melting point, m.p. (^oC)	1,414	2,340–2,667	2,500–2,700	2,256–2315	2,780–2,810	1,800–1,877
Thermal exp. coeff., α ($\times 10^{-6}$ K^{-1})	2.5	4.0–5.0	6.3–8.7	5.8–6.6	5.8–6.5	2.5–4.0
Bandgap, E_G (eV)	1.12	6.0–7.6	5.3	5.5–5.8	5.6–5.9	4.0–4.5
Electron affinity, χ_{di} (eV)	4.05	2.2–2.6	2.5–3.5	1.8	2.2–2.8	3.2–3.8
Effective Mass m_h*/m	0.59					0.5
m_e*/m	1.06			0.26	0.18-0.20	0.5
Dielectric constant, k (frequency range, Hz)	11.9 (10^7–10^9)	10.7–12.7	28–43	21–27 (10^4–10^6)	21–25 (10^4–10^6)	22–26 (10–10^6)

(continued)

Table B (continued)

Material	Si	$ZrSiO_4$	$BaZrO_3$	La_2O_3	HfO_2	Ta_2O_5
Refractive index, n	3.44	1.9–2.0	2.0	2.0–2.1	2.0–2.2	2.0–2.2
n^2	11.83	3.80	4.00	4.00	4.41	4.41
Relaxation frequencies (non-electronic) (Hz)	—-	—-	—-	—-	—	10^5

In a majority of cases, for the values of each constant, five independent references were used. The frequency range given in the row for the dielectric constant k, indicates the range over which the measurements were made and the values of k were found to be frequency-independent. A ≡ Anatase(T); B ≡ Brookite(O); C ≡ Cubic; H ≡ Hexagonal; M ≡ Monoclinic; O ≡ Orthorhombic; T ≡ Tetragonal

Appendix VI: Electronegativity Table of the Elements

Notes and Comments:

1. Electronegativity represents roughly the strength of an atom to attract electrons to itself.
2. No precise definition or mathematical relation exists for electronegativity. This concept originated from Linus Pauling while he was working on the valence bond theory.
3. Factors influencing electronegativity include the atomic number and the covalent radius.
4. There are several ways of estimating the scale of electronegativity: the classical and the earliest is the Pauling scale, followed by the Allred-Rochow scale, the Mulliken scale, the Sanderson scale, and the Allen scale.
5. In the table below, we have indicated below the symbol of the element, the Pauling electronegativity scale to the left and the Allen electronegativity scale to the right of the solidus.
6. In general, electronegativity increases along the diagonal from the lower left to the upper right of the periodic table.
7. In some cases, there is wide divergence between the Pauling and the Allen electronegativity scale.
8. In the table below, both the IUPAC (to the left of the solidus) and the CAS (to the right of the solidus) group numbers have been indicated.
9. The electronegativity difference ΔEN is often used to classify the type of chemical bond between atoms. The following is an illustration of this kind of an empirical scheme; the range indicated can be considered only an approximate one. Non-polar covalent bond: $\Delta EN < 0.5$; Polar covalent bond: $\Delta EN = 0.5–1.6$; Ionic bond: $\Delta EN > 2.0$.

	Group																	
Period	1 / IA	2 / IIA	3 / IIIB	4 / IVB	5 / VB	6 / VIB	7 / VIIB	8 / VIIIB	9 / VIIIB	10 / VIIIB	11 / IB	12 / IIB	13 / IIIA	14 / IVA	15 / VA	16 / VIA	17 / VIIA	18 / VIIIA
1	H 2.20/2.30																	He 0/4.60
2	Li 0.98/0.91	Be 1.57/1.58											B 2.04/2.05	C 2.55/2.54	N 3.04/3.07	O 3.44/3.61	F 3.98/4.19	Ne 0/4.79
3	Na 0.93/0.87	Mg 1.31/1.29											Al 1.61/1.61	Si 1.90/1.92	P 2.19/2.25	S 2.58/2.59	Cl 3.16/2.87	Ar 0/3.24
4	K 0.82/0.73	Ca 1.00/1.03	Sc 1.36/1.19	Ti 1.54/1.38	V 1.63/1.53	Cr 1.66/1.65	Mn 1.55/1.75	Fe 1.83/1.80	Co 1.88/1.84	Ni 1.91/1.88	Cu 1.90/1.85	Zn 1.65/1.59	Ga 1.81/1.76	Ge 2.01/1.99	As 2.18/2.21	Se 2.55/2.43	Br 2.96/2.69	Kr 3.00/2.97
5	Rb 0.82/0.71	Sr 0.95/0.96	Y 1.22/1.12	Zr 1.33/1.32	Nb 1.60/1.41	Mo 2.16/1.47	Tc 1.90/1.51	Ru 2.20/1.54	Rh 2.28/1.56	Pd 2.20/1.59	Ag 1.93/1.87	Cd 1.69/1.52	In 1.78/1.66	Sn 1.96/1.82	Sb 2.05/1.98	Te 2.10/2.16	I 2.66/2.36	Xe 2.60/2.58
6	Cs 0.79/0.66	Ba 0.89/0.88	La 1.10/	Hf 1.30/1.16	Ta 1.50/1.34	W 2.36/1.47	Re 1.90/1.60	Os 2.20/1.65	Ir 2.20/1.68	Pt 2.28/1.72	Au 2.54/1.92	Hg 2.00/1.76	Tl 2.04/1.79	Pb 2.33/1.85	Bi 2.02/2.01	Po 2.00/2.19	At 2.20/2.39	Rn 2.20/2.60
7	Fr 0.70/0.67	Ra 0.89/0.89	Ac 1.1/	Rf	Db	Sg	Bh	Hs	Mt	Uun	Uuu	Uub						
	Lanthanides				Ce 1.12/	Pr 1.13/	Nd 1.14/	Pm 1.13/	Sm 1.17/	Eu 1.20/	Gd 1.20/	Tb 1.10/	Dy 1.22/	Ho 1.23/	Er 1.24/	Tm 1.25/	Yb 1.10/	Lu 1.27/
	Actinides				Th 1.30/	Pa 1.50/	U 1.38/	Np 1.36/	Pu 1.28/	Am 1.13/	Cm 1.28/	Bk 1.30/	Cf 1.30/	Es 1.30/	Fm 1.30/	Md 1.30/	No 1.30/	Lr 1.30/

Color Key:	0-0.66	0.66-1.00	1.00-1.33	1.33-1.66	1.66-2.00	2.00-2.33	2.33-2.66	2.66-

Appendix VII: Work Function Table of the Elements

Notes and Comments:

1. The work function of a metal is the difference in energy between its Fermi level and the vacuum level.
2. The values of work function in the table are in eV.
3. For many metals, there is a wide scatter in the values of the work function reported in the literature.
4. The work function is sensitive to the surface orientation.
5. Group VIIIB elements appear to have the highest work functions, whereas group IA elements seem to have the lowest work functions.

		Group																
Period	1/IA	2/IIA	3/IIIB	4/IVB	5/VB	6/VIB	7/VIIB	8/VIIIB	9/VIIIB	10/VIIIB	11/IB	12/IIB	13/IIIA	14/IVA	15/VA	16/VIA	17/VIIA	18/VIIIA
1	H																	He
2	Li 2.93	Be 4.98											B 4.45	C 5.00	N	O	F	Ne
3	Na 2.36	Mg 3.66											Al 4.06-4.26	Si 4.60-4.85	P	S	Cl	Ar
4	K 2.29	Ca 2.87	Sc 3.50	Ti 4.33	V 4.30	Cr 4.50	Mn 4.10	Fe 4.67-4.81	Co 5.00	Ni 5.04-5.35	Cu 4.53-5.10	Zn 3.63-4.90	Ga 4.32	Ge	As 3.75	Se 5.90	Br	Kr
5	Rb 2.26	Sr 2.59	Y 3.10	Zr 4.05	Nb 3.95-4.87	Mo 4.36-4.95	Tc	Ru 4.71	Rh 4.98	Pd 5.22-5.60	Ag 4.52-4.74	Cd 4.08	In 4.09	Sn 4.42	Sb 4.55-4.70	Te 4.95	I	Xe
6	Cs 2.14	Ba 2.52-2.70	La 3.50	Hf 3.90	Ta 4.00-4.80	W 4.32-5.22	Re 4.72	Os 5.93	Ir 5.00-5.67	Pt 5.12-5.93	Au 5.10-5.47	Hg 4.48	Tl 3.84	Pb 4.25	Bi 4.34	Po	At	Rn
7	Fr	Ra	Ac	Rf	Db	Sg	Bh	Hs	Mt	Uun	Uuu	Uub						

	Lanthanides	Ce 2.90	Pr	Nd 320	Pm	Sm 2.70	Eu 2.50	Gd 2.90	Tb 3.00	Dy	Ho	Er	Tm	Yb 2.60	Lu 3.30
	Actinides	Th 3.40	Pa	U 3.633.90	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr

Index

S. Kar (ed.), *High Permittivity Gate Dielectric Materials*,
Springer Series in Advanced Microelectronics 43, DOI: 10.1007/978-3-642-36535-5,

G

H