The Application of Atomic Force Microscopy in Semiconductor Technology - Towards High-K Gate Dielectric Integration

Yordan Stefanov
2012

TECHNISCHE
UNIVERSITÄT
DARMSTADT

The Application of Atomic Force Microscopy in Semiconductor Technology – Towards High-K Gate Dielectric Integration

Vom Fachbereich Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von
Dipl.Ing. Yordan Stefanov

Geboren am 28 Dezember 1978 in Sofia, Bulgarien

Referent:	Prof. Dr. Udo Schwalke
Korreferent:	Prof. Dr. Henning Riechert
Tag der Einreichung:	01.02.2012
Tag der mündlichen Prüfung:	30.09.2011

D17
Darmstadt, 2012

Eidesstattliche Erklärung laut §9 PromO

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

Yordan Stefanov

Sofia, 01.02.2012

Acknowledgements

Prof. Udo Schwalke, first and foremost I would like to express my deepest gratitude to you. You gave me the opportunity to start my doctoral work at the Institute of Semiconductor Technology and Nanoelectronics and supported me one hundred percent from the very beginning all the way to the finish line. I really appreciate our many discussions for which you always seemed to have the time. Thank you for friendly and supportive guidance!

Karin Rose, thank you for your help and your friendship! You went way beyond your professional duties in helping me with all the beaurocratic troubles and all the small everyday problems. I enjoyed your kindness and positive attitude. And thanks for all the candy!

Dear colleagues at the ISTN, you have made my work in the institute not only easy but also a real pleasure. Thank you!

Thank you, dear students for doing all the hard work for me! My thesis would not have been possible without your help.

Last but not least, thank you, Federal Ministry of Education and Research for financing my work!

Abstract

Development of semiconductor technology over the last five decades has led to aggressive scaling down of integrated circuit (IC) device dimensions. ICs have become faster, denser and more power-efficient by continuous shrinking down of the metal-oxide-semiconductor field-effect transistor (MOSFET) size and implementation of complex integration schemes using novel materials. We are steadily approaching the physical limits of scaling and along the way more and more obstacles appear that need to be overcome in order to continue further. Traditional process control and device characterization techniques are becoming insufficient for addressing these problems. Novel techniques must be implemented for obtaining information about structural and electrical properties on materials and geometries with nanometer resolution. This is particularly relevant at the present transition from silicon dioxide gate dielectrics to ones with higher dielectric permittivity – high-*K* dielectrics.

The present work is a contribution to this search for novel suitable analytical techniques and their implementation in semiconductor technology. It exploits extensively the high resolution imaging possibilities of atomic force microscopy (AFM) as a key support technique from the selection of prospective high-K candidates to their integration into a suitable MOSFET fabrication process. Particular attention is paid to conductive atomic force microscopy (C-AFM) which offers the possibility of mapping simultaneously topography dimensions and electrical conductivity.

Initially, AFM and C-AFM are used for the development and optimization of a device isolation technology that is relevant in the context of high-*K* dielectrics in ultra large scale integration (ULSI) ICs – shallow trench isolation (STI). For the first time, reliable detection is obtained of the common problem related to STI – nitride erosion after the chemical planarization (CMP) step. Again with the help of C-AFM, two different techniques for planarity optimization are developed and evaluated – oxide etchback and reverse nitride masking.

Next, C-AFM supports the investigation of two principally different types of prospective high-*K* dielectric materials. First generation dual-stack dielectrics that consist of a high-*K* material on top of a thin interfacial silicon dioxide layer are the easier but less effective solution. C-AFM reveals imperfections in the investigated titanium oxide – silicon dioxide stacks related to the insufficient stability of such bilayer structures. Second generation high-*K* dielectrics in the face of epitaxial rare-earth metal oxides possess key advantages such as higher thermal stability and the possibility for engineered interface with silicon. C-AFM investigates their properties and proves the superiority of these materials. Imperfections are observed as well that show the need for growth and processing optimizations. For the first time, charge trapping is observed on the nanoscale directly on the high-*K* dielectric surface. Nonuniform leakages in rare-earth metal oxides grown under insufficiently optimized conditions presumably related to grain boundaries are discovered in some samples. Based on AFM measurements, predictions are made about the expected behavior of MOS devices incorporating these materials.

The compatibility of epitaxial rare-earth metal oxides with standard complementary metal-oxide-semiconductor (CMOS) processing is investigated next. Incompatibility with some steps such as for example cleaning with acid-containing solutions is determined and suitable replacement steps are chosen. Changes in film properties are determined during key steps that could indicate incompatibility of the dielectrics with the standard gate-first integration scheme.

In order to determine to what extent the observed microscopic changes affect macroscopic device behavior, epitaxial dielectric layers are integrated for the first time into complete devices. Rare earth metal oxide MOSFETs are fabricated into a modified gate-first process using different gate dielectrics. C-AFM is used for process control in critical steps. Electrical evaluation of the functional devices featuring praseodymium oxide (Pr_2O_3), including charge pumping, reveals that at this initial stage of

development the high-*K* gate dielectric devices suffer from degraded performance when compared to SiO_2 reference devices. Imperfections such as high density of interface states, susceptibility to charge trapping and gate leakages for large area devices are observed. Neodymium oxide (Nd_2O_3) integration after further optimization of the gate-first process fails to produce functional devices due to substantial degradation of the gate dielectric and excessive gate leakages. The MOSFET behavior for both materials as determined by macroscopic electrical characterization results is compared to AFM predictions and they coincide very well.

It is concluded that the imperfections of the gate dielectrics are at least partially a result of the integration process. Analysis is carried out and critical performance-reducing steps are identified. The gate structuring by reactive ion etch (RIE), the source/drain ion implantation and the high temperature source/drain activation anneal are responsible for the dielectric degradation to the largest extent. The inseparable link between these steps and conventional processing leads to the idea of implementing an entirely different approach for gentle integration of high-*K* dielectrics.

Once again with the help of AFM and C-AFM, a replacement gate technology (RGT) is developed and implemented for high-*K* gate dielectric MOS devices in order to prove this concept. By positioning the gate dielectric growth module after the source/drain implantation and anneal and avoiding the aggressive RIE through indirect gate patterning with CMP, the integration process is adapted to the sensitive high-*K* materials in order to preserve their as-grown state. Electrical evaluation of devices with Gd_2O_3 produced using RGT proves the advantage of RGT. The first integration attempt is compared to conventional fabrication technology and there are definite improvements in terms of threshold voltage stability and interface state distribution. The first RGT high-*K* devices still do not exhibit the mobility and low defect density of equivalent state-of-the-art SiO_2 devices but this is expected considering the 40-year-long optimization history behind silicon dioxide. Further optimization is needed for epitaxial rare-earth metal oxides as well, both in terms of growth conditions and process integration.

Kurzfassung

Die Entwicklung der Halbleitertechnologie hat in den zurückliegenden fünf Jahrzehnten zur aggressiven Verkleinerung der Bauelementdimensionen in integrierten Schaltkreisen (ICs) geführt. Durch die stetige Verkleinerung von Metall-Oxid-Halbleiter-Feldeffekt-Transistoren (MOSFETs) und durch Implementierung von hoch-komplexen Systemen unter Zuhilfenahme neuer Materialien sind ICs schneller und energieeffizienter geworden, zudem wurde die Packungsdichte erhöht. Jedoch ist man hierdurch immer näher an die physikalischen Grenzen der Skalierung gestoßen, die überwunden werden müssen, um diese Entwicklung weiter fortsetzen zu können.

Mit zunehmender Verkleinerung der Bauelemente in den Nanometerbereich sind konventionelle Techniken der Prozesskontrolle und Charakterisierung immer weniger geeignet. Daher müssen neue Verfahren für die Gewinnung von Informationen über strukturelle und elektrische Eigenschaften von Materialien und Geometrien mit Nanometerauflösung entwickelt werden. Dies ist gegenwärtig, beim Übergang von Siliziumdioxid Gate-Dielektrika zu solchen mit einer höheren Dielektrizitätskonstante (sogenannten High-*K*-Dielektrika), besonders relevant.

Die vorliegende Arbeit leistet einen Beitrag zur Auswahl geeigneter, neuartiger analytischer Methoden zur Prozesskontrolle und behandelt deren Implementierung in die Halbleitertechnologie. Hier findet die Rasterkraftmikroskopie (AFM) Anwendung, die aufgrund der hochauflösenden Abbildungsmöglichkeiten eine geeignete Technik zur Untersuchung neuartiger High-*K*-Dielektrika darstellt und Fragen zur Prozessintegration dieser Materialien in einen geeigneten MOSFET-Herstellungsprozess klären hilft. Besonderes Augenmerk wird auf die Rasterkraftmikroskopie mit elektrisch leitfähiger Messspitze (C-AFM) gesetzt, welches die simultane Messung von Topographie und elektrischer Leitfähigkeit an der Probenoberfläche ermöglicht.

Zuerst wird die Rasterkraftmikroskopie (AFM und C-AFM) zur Entwicklung und Optimierung einer modernen Isolationstechnologie für Bauelemente, der Grabenisolation (Shallow Trench Isolation, STI) eingesetzt, die für die Integration von High-*K*-Dielektrika in hoch-skalierten integrierten Schaltungen (ULSI) relevant ist. Erstmalig wird der Nachweis der Nitrid-Erosion, einem bekannten Problem im Zusammenhang mit STI nach dem chemisch-mechanischen Planarisieren (CMP), in zuverlässiger Weise erreicht. Die C-AFM-Technik wird außerdem zur Entwicklung und Evaluierung von zwei unterschiedlichen Techniken zur Optimierung der Planarität erfolgreich eingesetzt, dem Oxid-Rückätz-Verfahren (oxide etchback) und der inversen Nitrid-Maskierung (reverse nitride masking).

Weiterhin konnte die C-AFM-Technik für die Untersuchung von zwei grundsätzlich verschiedenen High-*K*-Dielektrika eingesetzt werden. Es konnte gezeigt werden, dass die erste Generation der High-Dielektrika, bestehend aus Titanoxid als High-*K*-Material auf einer dünnen Siliziumdioxid-Pufferschicht vom Si-Substrat getrennt, eine technologisch einfache aber elektrisch unbefriedigende Lösung darstellt. C-AFM Messungen zeigten Mängel in den untersuchten Titanoxid - Siliziumdioxid-Doppelschichtstrukturen hinsichtlich der Defektdichte und des Leckstromverhaltens. Die zweite Generation der High-*K*-Dielektrika, die epitaktisch gewachsenen kristallinen Seltenerd-Oxide, konnten deutlich bessere Eigenschaften mittels C-AFM nachgewiesen werden, insbesondere hinsichtlich der thermischen Stabilität. Dennoch wurden auch in diesen High-*K*-Materialien Unvollkommenheiten beobachtet. So konnte erstmalig der Ladungseinfang und die Generation von Einfangstellen in einem High-*K*-Dielektrikum im Nanometerbereich mittel C-AFM direkt an der Oberfläche beobachtet werden.

Des Weiteren wird die Kompatibilität der epitaktisch gewachsenen kristallinen Seltenerd-Oxide mit Standard-CMOS-Prozessen untersucht. So wurden z.B. der Reinigung mit säurehaltigen Lösungen, Unverträglichkeiten ermittelt, die eine Neuentwicklung geeigneter Reinigungsverfahren notwendig machte. Änderungen der Filmeigenschaften wurden auch bei anderen wichtigen Prozessschritten

festgestellt, die unter Umständen auf eine Unvereinbarkeit der High-*K*-Dielektrika mit der Standard-CMOS Prozessführung (Gate-First) hindeuten.

Um zu bestimmen, inwieweit die mittels C-AFM beobachteten nanoskaligen Veränderungen im Dielektrikum das Verhalten des makroskopischen Bauelements beeinflussen, wurden epitaktische High-*K*-Schichten aus Praseodymoxid (Pr_2O_3) erstmalig vollständig in MOS-Transistoren integriert. C-AFM wurde zusätzlich zur Prozesskontrolle bei kritischen Schritten verwendet. Die elektrische Charakterisierung der voll funktionsfähigen Bauelemente mit Praseodymoxid (Pr_2O_3), zeigte, dass diese ersten High-*K*-MOSFETs deutliche Leistungseinbußen im Vergleich zur SiO_2 Referenzbauelementen aufweisen. An großflächigen Bauelementen werden deutlich erhöhte Werte der Grenzflächenzustandsdichte, Gate-Leckströme und eine höhere Anfälligkeit für Ladungsträgereinfang beobachtet. Beim Versuch der Integration von anderen kristallinen Dielektrika, wie Neodymiumoxid (Nd_2O_3), konnten trotz Prozessoptimierung im Rahmen der Gate-First Technologie keine funktionierenden MOSFETs realisiert werden. Es wird gefolgert, dass die Degradation der High-*K* Gate-Dielektrika zumindest teilweise eine Folge des Integrationsprozesses sind.

Nach Durchführung von weiteren Analysen konnten die kritischen Prozessschritte identifiziert werden. Im Wesentlichen sind die Gatestrukturierung mittels reaktivem Ionenätzen (RIE), die Source/Drain-Ionen-Implantation und das Source/Drain-Aktivierung bei hoher Temperatur für die Degradierung des Dielektrikums verantwortlich. Der untrennbare Zusammenhang zwischen diesen Schritten und der konventionellen (Gate-First) Prozessierung führte zur Entwicklung eines neuartigen Gesamtprozesskonzeptes zur schonenden Integration der High-K-Dielektrika.

Unter Verwendung der Rasterkraftmikroskopie (AFM und C-AFM) wird eine sogenannte ‚Replacement-Gate-Technologie' (RGT) für MOS-Transistoren mit High-*K*-Dielektrika entwickelt und implementiert, um die Realisierbarkeit des Konzepts nachzuweisen und das Verbesserungspotential aufzuzeigen. Es konnte gezeigt werden, dass dank der geänderten Prozessfolge, in der das empfindliche High-*K*-Material erst nach allen aggressiven Prozessschritten eingefügt wird, die ursprünglich gute Qualität der High-*K* Dielektrika weitgehend erhalten bleibt. Es konnten deutliche Verbesserungen hinsichtlich der Stabilität der Einsatzspannung und der Grenflächenzustandsdichten erreicht werden.

Contents

List of figures

List of tables

List of equations

List of abbreviations

AFM	Atomic force microscopy
Al	Aluminum
APCVD	Atmospheric pressure chemical vapor deposition
Ar	Argon
BEOL	Back end of line
C-AFM	Conductive atomic force microscopy
Cl_2	Chlorine gas
CMOS	Complementary metal-oxide-semiconductor
CMP	Chemical mechanical planarization / polishing
CP	Charge pumping
CPU	Central processing unit
Cu	Copper
CVD	Chemical vapor deposition
DLC	Diamond-like carbon
EOT	Equivalent oxide thickness
F-N	Fowler-Nordheim
Gd	Gadolymium
Gd_2O_3	Gadolynium oxide
HCl	Hydrogen chloride
HF	Hydrogen fluoride
Hf	Hafnium
H_2O	Water
IC	Integrated circuit
IEMD	Institute of Electronic Materials and Devices
IHP	Innovations for High Performance Microelectronics GmbH
ILD	Interlayer dielectric
ISTN	Institute of Semiconductor Technology and Nanoelectronics
ITRS	International technology roadmap for semiconductors
KOH	Pottasium hydroxide
K_2SiO_2	Pottasium silicate
LDD	Lightly doped drain
LOCOS	Local oxidation of silicon
MBE	Molecular beam epitaxy
MIS	Metal-insulator-semiconductor
MOS	Metal-oxide-semiconductor
MOSFET	Metal-oxide-semiconductor field-effect transistor
Nd:YAG	Neodymium-doped yttrium aluminum garnet, $Nd:Y_3Al_5O_{12}$
NH_3	Ammonia
NH_4OH	Ammonium hydroxide
Nd	Neodymium
Nd_2O_3	Neodymium oxide
nMOS	Metal-oxide-semiconductor structure with a n-type doped semiconductor
nMOSFET	Metal-oxide-semiconductor field-effect transistor with a n-channel
nSi	n-doped silicon
O_2	Diatomic oxygen gas
O_3	Ozone
OE	Oxide etchback
PECVD	Plasma enhanced chemical vapor deposition
pH	Measure of alkalinity or acidity of a solution
PLD	Pulsed laser deposition

pMOS	Metal-oxide-semiconductor structure with a p-type doped semiconductor
pMOSFET	Metal-oxide-semiconductor field-effect transistor with a p-channel
Polysilicon	Polycrystalline silicon
Pr	Praseodymium
Pr_2O_3	Praseodymium oxide
pSi	p-doped silicon
Pt	Platinum
PtIr	Platinum-iridium alloy
RCS	Remote coulomb scattering
RGT	Replacement gate technology
RIE	Reactive ion etch
RMS	Root mean square
RNM	Reverse nitride masking
RTA	Rapid thermal annealing
Si_3N_4	Silicon nitride
SELOX	Selective oxide deposition
SEM	Scanning electron microscopy
SF_6	Sulfur hexafluoride
SiO_2	Silicon dioxide
$SiHF_3$	Trifluorsilane
SPM	Scanning probe microscopy
STI	Shallow trench isolation
STM	Scanning tunneling microscopy
TEM	Transmission electron microscopy
TEOS	Tetraethyl orthosilicate
Ti	Titanium
TiN	Titanium nitride
TiO_2	Titanium oxide
ULSI	Ultra large scale integration
VLSI	Very large scale integration
W	Tungsten
WTi	Tungsten-titanium alloy
XPS	X-ray photoelectron spectroscopy
XTEM	Cross-sectional transmission electron microscopy
Zr	Zirconium

List of symbols

Symbol	Unit	Description
C	F	Dynamic capacitance of a MOS capacitor
C_D	F	Depletion capacitance
C_i	F	Capacitance of the gate dielectric
C_s	F	Capacitance of the semiconductor substrate
d	nm	Gate dielectric thickness
d_{highK}	nm	Thickness of a high-K gate dielectric
d_{SiO2}	nm	Thickness of SiO_2 gate dielectric
$D_{it}(E)$	$cm^{-2}eV^{-1}$	Energy distribution of interface states
D_{it}	$cm^{-2}eV^{-1}$	Average interface state density
E_c	eV	Energy level of the conductance band edge of silicon
E_F	eV	Fermi level in silicon
E_{FM}	eV	Fermi level in a metal
E_G	eV	Energy bandgap of silicon
E_0	eV	Energy of an electron in vacuum
E_v	eV	Energy level of the valence band edge of silicon
E_g	eV	Energy bandgap of silicon
e	m	Depletion layer width
F	$V.m^{-1}$	Electric field
F_i	$V.m^{-1}$	Normal electric field across the gate dielectric
F_{lat}	$V.m^{-1}$	Lateral electric field along the MOSFET channel
f	Hz	Frequency
h	eV.s	Planck constant
I	A	Electric current
I_{CP}	A	Charge pumping current
I_{off}	A	MOSFET off-state current
I_{on}	A	MOSFET on-state current
I_{DS}	A	Drain-to-source current
I_{DSsat}	A	Saturation drain-to-source current
J	$A.cm^{-2}$	Electric current density
J_{FNT}	$A.cm^{-2}$	Fowler-Nordheim tunneling current density
J_{DT}	$A.cm^{-2}$	Direct tunneling current density
J_{PF}	$A.cm^{-2}$	Poole-Frenkel emission current density
J_{SE}	$A.cm^{-2}$	Schottky emission current density
J_{OC}	$A.cm^{-2}$	Ohmic conduction current density
K		Dielectric constant
K_{eff}		Effective dielectric constant of a dual-layer stack
K_{highK}		Dielectric constant of a high-K dielectric material
K_i		Dielectric constant of a gate dielectric
K_{IF}		Dielectric constant of an interface layer
K_s		Dielectric constant of a silicon substrate
K_{SiO2}		Dielectric constant of SiO_2
k	$m^2.kg.s^{-2}.K^{-1}$	Bolzmann constant
l	μm	Gate length
m^*	kg	Effective electron mass
N_{it}	cm^{-2}	Density of interface states per unit area
n_i		Intrinsic carrier concentration
P	$N.m^{-2}$	Pressure

Q	C	Electric charge
Q_f	C	Fixed charge
Q_G	C	Electric charge accumulated in the gate electrode of a MOS structure
Q_{it}	C	Interface state charge
q	C	Elementary charge
R	Ω	Resistance
r	$nm.min^{-1}$	Polish rate
S	$mV.dec^{-1}$	Subthreshold swing
s	$m.s^{-1}$	Linear relative velocity between wafer and polish pad
t_r	s	Pulse rise time
t_f	s	Pulse fall time
V	V	Voltage
V_{DS}	V	Drain-to-source voltage
V_{DSsat}	V	Saturation drain-to-source current
V_{FB}	V	Flat-band voltage
V_G	V	Gate-to-substrate voltage
V_{GS}	V	Gate-to-source voltage
V_T	V	Threshold voltage
v_d	$cm.s^{-1}$	Drift velocity
v_{th}	$cm.s^{-1}$	Carrier thermal velocity
w	μm	Gate width
z	$nm^2.N^{-1}$	Preston's coefficient
ΔC	F	Capacitane shift
ΔE_{ah}	eV	Hopping activation energy
ΔV	V	Voltage shift
g_m	$A.V^{-1}$	Transconductance
ε_i	$F.m^{-1}$	Electric constant
μ_n	$m^2.V^{-1}.s^{-1}$	Carrier mobility
μ_{eff}	$m^2.V^{-1}.s^{-1}$	Effective carrier mobility
μ_{FE}	$m^2.V^{-1}.s^{-1}$	Field effect mobility
σ_n	cm^{-2}	Capture cross-section for electrons
σ_p	cm^{-2}	Capture cross-section for holes
φ_B	V	Barrier height
φ_{BC}	V	Conduction band offset
φ_{BV}	V	Valence band offset
φ_F	V	Fermi potential of silicon
φ_M	V	Metal work-function
φ_T	V	Thermal potential
φ_{MS}	V	Work-function difference betweeen metal and silicon
χ	V	Electron affinity of silicon
ψ_s	V	Band bending
ω	s^{-1}	Angular velocity

1. Introduction

Microelectronics has witnessed an unparalleled growth since the early 1960s. Following the prodigious Moore's law [1], the number of transistors on an integrated circuit has doubled every 18 months for almost five decades. Integrated circuits (ICs) have become faster, denser and more power-efficient by continuous scaling down of transistor size and implementation of complex integration schemes using novel materials.

Traditional process control and device characterization techniques are already not sufficient for advanced ICs. Novel measurement technology must be developed and implemented in order to support the further development of microelectronics towards nanoelectronics. This is especially true at present when the transition in complementary metal-oxide-semiconductor (CMOS) fabrication from silicon dioxide to high-*K* gate dielectrics is taking place. Sufficient device-relevant knowledge of the electrical properties of these materials must be gained in order to integrate them successfully for electronic applications. The compatibility of these materials with standard CMOS processing steps must be investigated. Modified process steps and modules must be developed when needed. The electrical properties of devices incorporating the dielectrics must be evaluated. Eventually, development of novel integration schemes might be necessary in order to achieve optimal IC performance and reliability.

This work investigates the application of atomic force microscopy (AFM) in microelectronic technology. With its ultra-high resolution and the possibility for simultaneous topographic and electrical conductivity measurements, AFM makes it possible to obtain critical information about the integration of novel gate dielectrics in semiconductor technology. AFM is implemented for study of basic electrical and structural properties of selected high-*K* dielectrics (Chapter 5). Based on AFM measurements, predictions are made about the expected behavior of MOS devices incorporating these materials. The best prospective candidates are selected and integrated in a modified gate-first fabrication process. AFM is utilized for development of key process steps and modules in high-*K* integration (Chapter 4 and Chapter 6). MOSFETs are fabricated and evaluated electrically (Chapter 7). Their behavior as determined by macroscopic electrical characterization results is compared to AFM prediction. Analysis of the performance of the devices is carried out and critical steps are identified that reduce performance. The inseparable link between these steps and conventional processing leads to the idea of implementing an entirely different integration approach. Again with the help of AFM, a replacement gate technology is developed and implemented for high-*K* gate dielectric MOS devices in order to prove the concept (Chapter 8). First integration attempt is compared to conventional fabrication technology and the potential of the new approach is analyzed.

The work is carried out at the Institute for Semiconductor Technology and Nanotechnology (ISTN), Darmstadt University of Technology under the KrisMOS project. Own results are presented together with several results from project partners.

2. Atomic force microscopy

Atomic force microscopy (AFM), also known as scanning force microscopy, is one of several types of scanning probe microscopy (SPM) techniques that use a physical scanning probe to produce high-resolution images of the specimen of interest.

2.1. History

The first ancestor of the atomic force microscope is the stylus profiler invented in 1929 by Gustav Schmaltz [2]. This profiler monitors the fine movement of a sharp tip attached at the end of a lever arm over the investigated surface during a line scan by means of a light deflection system. A magnified profile of the surface is recorded on photographic paper. The magnification exceeds 1000.

In 1950 Becker introduces the idea of dynamic operation by oscillating the stylus in order to minimize lateral forces [3]. In 1971 Young used the exponential dependence of distance between a sharp metal probe and the conductive sample and the field emission current to keep the distance constant [4]. His TopoGraphiner works in non-contact mode and uses a piezoelectric element to move the sample in the x- and y-directions. This instrument is able of constructing three-dimensional images of sample surfaces, but its resolution is limited by mechanical vibrations.

A serious advance in high-resolution SPM imaging comes with the invention of the STM in 1981 by Gerd Binnig and Heinrich Rohrer at IBM [5; 6]. The STM uses a sharp conductive tip which is brought to the sample surface within electron quantum tunneling distance and scanned across the surface line by line (Fig. 2.1). Controlling vibrations, the group is able to image the atomic lattice of a sample. The STM is a revolution at the time of its invention and it earns the Nobel Prize in Physics to its inventors in 1986. It is a comparatively simple instrument that allows achieving atomic resolution on flat surfaces even in ambient atmosphere in a matter of minutes and without any particular sample preparation.

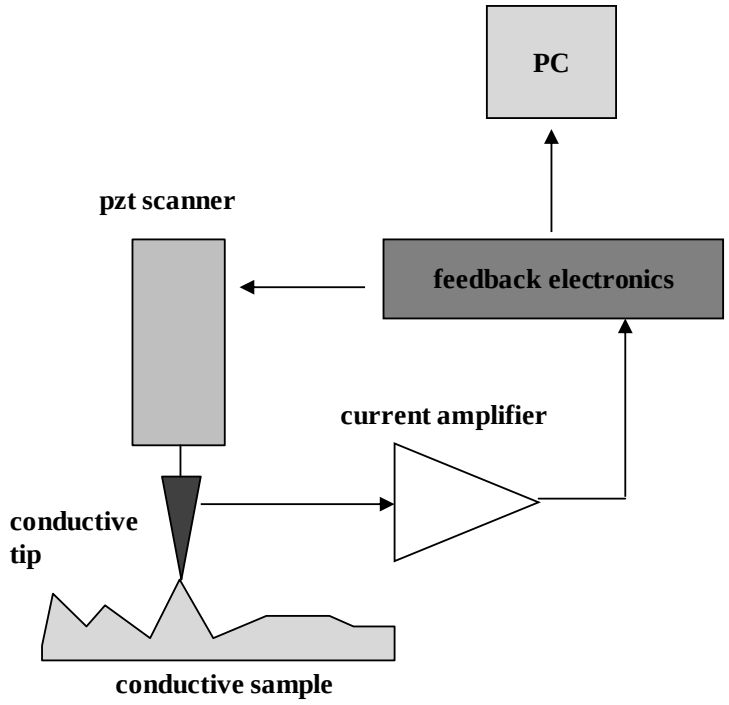


Fig. 2.1: Schematic of STM operation principle.

Unfortunately, STM has one serious flaw: as it relies on the tunneling current between the tip and the surface for tip positioning, its application is limited to investigating conducting samples. In an attempt to overcome this disadvantage, Gerd Binnig, Calvin Quate and Christoph Gerber develop in 1986 an instrument for nanoscale imaging whose measurement capabilities are independent of the sample electrical properties – the atomic force microscope [7].

2.2. Operation principle

The basic operation principle of a standard AFM system with optical feedback (Fig. 2.2) is the following: a probe with a sharp tip usually made of silicon or silicon nitride attached to a flexible cantilever is scanned over the sample surface. As the tip moves over areas of different height the deflection of the cantilever changes. A laser beam is reflected from the back side of the cantilever and directed into a position sensitive photodetector. The photodetector signal is used to control a piezoelectric scanner that keeps the probe-sample distance and hence interaction force constant. The piezoelectric scanner is attached to either the sample holder or the probe holder and moves it in all three directions. The voltages applied to the piezoelectric scanner to define the spacial position of the probe tip relative to the sample surface are used to create a topography map of the scanned surface.

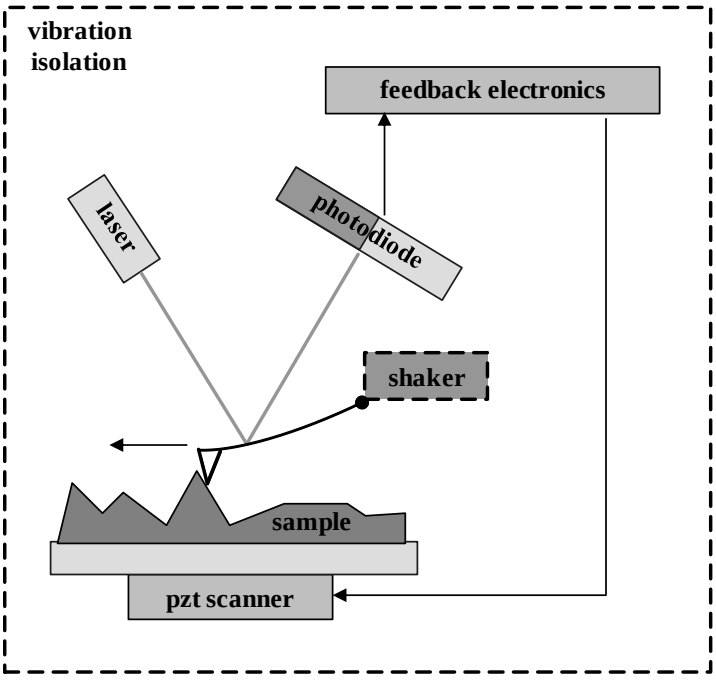


Fig. 2.2: Schematic of AFM operation principle.

The tip can also be scanned at a constant height with the feedback loop disabled and the deflection of the cantilever recorded. This regime is used for high-resolution imaging of very flat surfaces. It is not suitable for rougher surfaces as it causes higher tip wear.

Noise isolation is critical to achieving high resolution. AFM systems must be isolated from acoustic, electrical and optical interference from the surroundings.

Because AFM uses cantilever deflection due to physical contact and electrostatic or van der Waals forces for tip positioning, AFM can be used for investigation of virtually any solid surface (and in some cases - liquid surfaces [8]), making it a much more versatile tool than STM. With AFM it is possible to

achieve atomic resolution in vacuum [9] and near atomic resolution in ambient atmosphere which makes it a very important analytical technique in various areas of science and technology.

2.3. Primary modes of operation

There are two basic modes of imaging surface topography with an atomic force microscope: static or contact mode and dynamic mode. Dynamic mode is further divided into tapping or intermittent contact and non-contact mode depending on the interaction forces between the tip and the surface.

In contact mode, the probe tip is pressed against the scanned surface and in hard contact with it (Fig. 2.3a). The feedback system maintains a constant cantilever deflection and hence a constant interaction force. The tip-sample distance of several Angstrom places the interaction in the positive region of the Lennard-Jones potential [10] and accordingly, the interaction forces between the tip and the surface are repulsive (Fig. 2.4). Normally, the forces are in the range of 10^{-11}-10^{-7}N. Soft cantilevers with force constants <1N/m are usually used in order to minimize tip wear and surface damage and increase sensitivity. Because of the soft cantilevers, occasionally the tip sticks to the surface contamination layer that is present on most surfaces in ambient air. This causes image distortions and limits resolution. In ultra-high vacuum (UHV) and in liquids, the absence of capillary forces allows for higher resolution. Biological samples are hard to scan in contact mode because they are usually soft and loosely bound to the base surface.

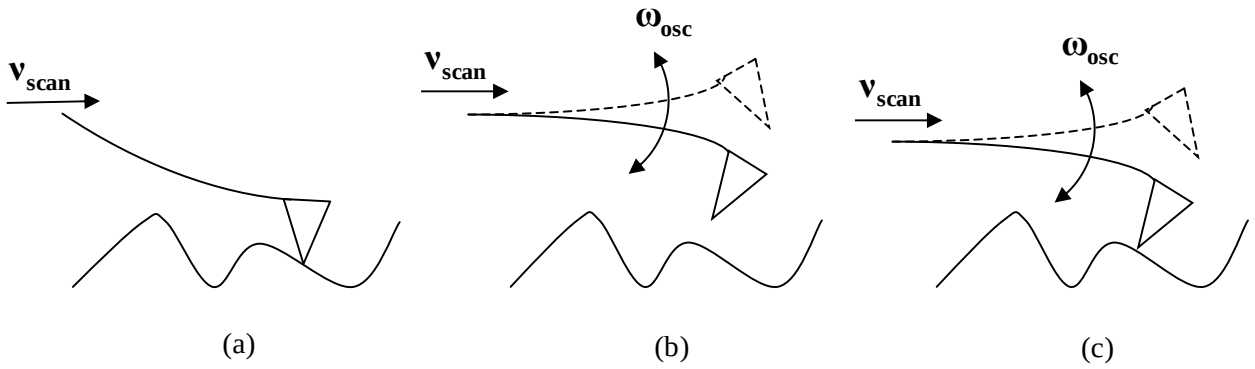


Fig. 2.3: Schematics of the three different modes of AFM operation: (a) contact mode, (b) non-contact mode and (c) tapping mode.

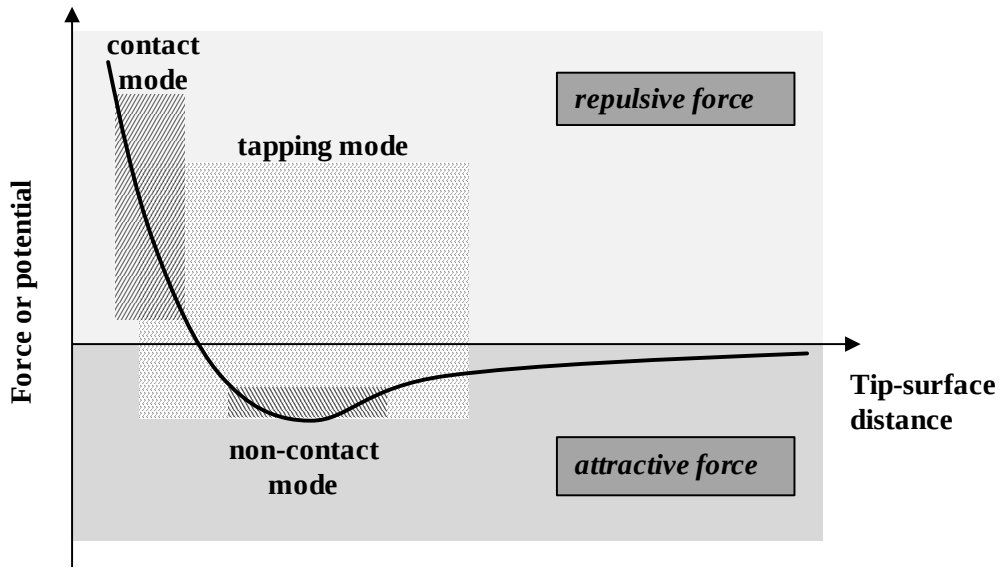


Fig. 2.4: Operation range of the different modes of AFM operation.

Non-contact mode is introduced in 1987 [11] in an effort for more successful biological sample imaging. In this mode a stiff cantilever with a force constant of several 1-100N/m is oscillated by a piezoelectric actuator at a frequency close to its resonance at a distance 1-10nm away from the sample surface (Fig. 2.3b) in the region of long-range attractive interaction forces (Fig. 2.4). The interaction forces are in the range of several pN and the oscillation amplitude is below 10nm. The feedback loop maintains a constant amplitude (amplitude modulation regime) or a constant frequency (frequency modulation regime). In frequency modulation very stiff cantilevers can be used (several tens of N/m) since frequency can be measured with very high sensitivity. In amplitude modulation, the phase information can be extracted in order to identify different materials. Non-contact mode offers the lowest possible interaction between the tip and the sample surface which is beneficial when imaging soft samples as well as considering probe longevity. On the downside, it is not very easy to work in this mode of operation. Sticking of the tip by meniscus forces is possible after accidental tip-sample contact when working at the small tip-sample distances needed for achieving high resolution. For this reason, high performance Z-scanner with a fast response is required.

Intermittent-contact mode, also known as dynamic-contact mode or tapping mode is developed in 1993 [12]. Tapping mode is similar to non-contact mode and differs from it in that the tip taps the sample surface gently at the lower end of each oscillation (Fig. 2.3c). The oscillation frequency is several tens to several hundred kilohertz and the oscillation amplitude is much larger than in non-contact mode – in the range of 100-200nm. The forces that dominate the interaction between the tip and the surface are repulsive (Fig. 2.4). Stiff cantilevers with force constants above 1N/m and usually in the range of 15-40N/m are used. Tapping mode is possible in ambient atmosphere, UHV and liquids [13].

The main benefits of tapping mode operation in comparison with contact mode are the reduced requirements towards the AFM system and the eliminated sticking of the tip. They come at the expense of marginally lower resolution and higher interaction forces. Compared to contact mode, tapping mode offers a much lower tip-sample interaction and eliminates lateral forces.

2.4. Important scan parameters

1. The interaction force between the tip and the sample determines how well the tip follows the surface topography. The parameter responsible for controlling this force is commonly referred to as the 'setpoint'. In tapping mode, the setpoint is related to the oscillation damping. The latter is the reduction of the oscillation amplitude during scanning relative to the free oscillation amplitude. In contact mode the setpoint defines the cantilever deflection during scanning. Stronger interaction allows for better and more stable scanning but also causes faster tip wear and more sample damage.

2. The combination of scan size and scan rate defines the velocity at which the tip moves along the surface. High velocities can lead to improper surface tracking with large variations of tip-sample interaction and faster tip wear. Low scan velocities prolong tip life and improve surface tracking at the expense of increased scan time.

3. Feedback settings also affect the way the tip tracks the surface. Too low feedback gains result in inaccurate tracking, while too high gains cause instability of the feedback loop.

4. The oscillation amplitude of the tip is an important parameter in tapping and non-contact mode operation. In tapping mode, common free oscillation amplitudes are in the order of a couple hundred nanometers. Increasing the amplitude helps improve tracking at the expense of higher tip-sample interaction. In non-contact mode low amplitudes in UHV conditions allow achieving true atomic resolution.

2.5. Secondary modes of operation

In addition to topography, numerous other sample and sample-tip interaction properties can be imaged and measured with nanometer resolution including elasticity, hardness, friction, thermal properties, component distribution, work-function, electrostatic potential, magnetic potential and others [14; 15]. AFM is additionally used for manipulating matter on the nanoscale by nanoindentation, scratching, wear testing, etc. [16]. Below is a list of the most common secondary modes of AFM.

1. Force-distance measurements: the tip can be moved vertically towards the surface past the initial contact and back with x-y scanning disabled. The piezo-displacement and the corresponding cantilever deflection yield information about the stiffness and elasticity of the material at the point of contact, van der Waal's forces, stretching forces for single molecules, etc. Force sensitivity is in the range of several pN.

2. Lateral force microscopy: by using a four-segment photodetector and scanning in contact mode, the microscope is capable of detecting not only the vertical deflection of the cantilever but also its lateral twisting that can be related to the frictional forces acting on the probe tip.

3. Phase imaging: recording the phase difference between the drive signal and the cantilever oscillation in dynamic mode gives additional information about material properties such as elasticity, adhesion, etc.

4. Force modulation microscopy: while scanning in contact mode, an additional high frequency low amplitude oscillation is applied to the cantilever, allowing the tip to test the elasticity and stiffness of the sample surface.

5. Electrostatic force microscopy: the variations in electric field gradient across the sample surface can be analyzed in non-contact mode by applying an additional voltage to conductive probes.

6. Magnetic force microscopy: probes with a hard magnetic coating are used to scan the magnetic field above the sample surface in non-contact mode.

7. Conductive atomic force microscopy: the local electrical resistivity of the surface can be probed with electrically conductive probes in contact mode by applying a voltage difference between the probe and the sample and measuring the resulting electrical currents. For more information, please check Chapter 2.6.

8. Scanning capacitance microscopy: localized capacitance measurements are performed with electrically conductive probes with the help of high frequency resonant capacitance sensors.

9. Scanning thermal microscopy: using a temperature sensitive probe, mapping of the temperature distribution of a surface and the local thermal conductivity can be measured.

10. Near-field scanning optical microscopy: operation with visible light and resolution beyond the limits of diffraction with the help of light-guiding probes is possible.

11. AFM in liquids: measurements of biological and other samples in aqueous solutions are possible in both contact and tapping mode. For measurements in liquids, silicon nitride probes are commonly used. Measurements in liquids help avoid tip sticking to the surface when using soft cantilevers.

12. Ultra high vacuum AFM: UHV AFM systems are also available for atomic resolution imaging or for protecting samples from contamination.

13. Nanolithography: not only imaging but also deliberate modification of the sample surface can be obtained by applying high enough forces with the probe tip. Nanolithography is a chemical-mechanical technique for patterning surfaces.

14. Nanoindentation: material properties can be evaluated by indenting a hard tip (usually diamond or diamond-like-carbon coated) into the surface of interest. Force displacement curves yield information about the hardness and the elastic modulus of the surface.

2.6. Conductive AFM

Conductive atomic force microscopy (C-AFM) is a powerful technique for nanoscale characterization of the electrical properties of conducting and semiconducting materials. Current suggested applications of C-AFM measurements include, among others, characterization of thin dielectric films, imaging of defects embedded in SiO_2 films, imaging of thin ferroelectric and piezoelectric films [17].

In Conductive AFM, an electrical voltage is applied between a conductive tip and the sample surface which it scans in contact mode. The resulting currents tip and the sample are sensed, passed through a low noise linear current amplifier and imaged (Fig. 2.5). Since topography and current images are measured simultaneously and independently from each other, direct correlation of the sample location and its local electrical properties is possible with nanometer resolution. In addition to imaging mode, C-AFM can be also used in spectroscopy mode. Local current-voltage spectra can be measured in point contact mode at points of interest.

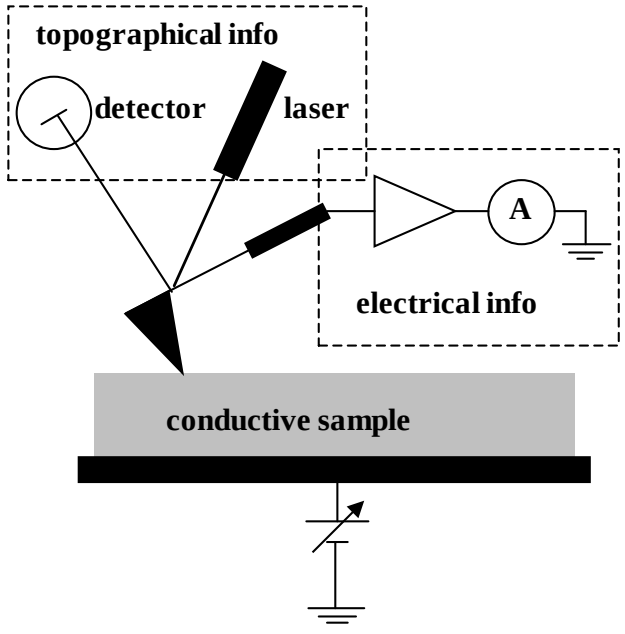


Fig. 2.5: C-AFM operation principle schematic.

C-AFM is the measurement technique in the focus of this thesis. It is implemented for structural and electrical characterization of several different high-*K* dielectric materials, helping the selection of suitable candidates for CMOS integration. Additionally, C-AFM is used for determining CMOS compatibility of the selected materials. Furthermore, the technique is utilized as a major tool for process control at various stages of microelectronic device fabrication.

The tool for all AFM measurements performed at the ISTN is a Veeco Dimension 3100 system with a Nanoscope V controller and the optional electrical measurement module 'TUNA'.

Fig. 2.6: Photograph of the Veeco Dimension 3100 AFM system at the ISTN.

2.7. Probe tip: resolution and tip-shape effects

AFM images are a result of the convolution of the tip and the sample surface. The smaller the surface features compared to tip dimensions, the less accurate the sample image on the computer screen. In order to understand how AFM images are formed and how they differ from the real surface topography, one needs to know about tip shape effects.

Achieving maximum lateral imaging resolution is not an easy task. Resolution depends on several factors including the measurement setup (microscope, acoustic isolation and environment), the initial quality of the AFM measurement probe and the experience of the user. In general, the capabilities of the microscope are higher than the resolution limitations imposed by probe tip geometry. A medium-level AFM system offers sub-nanometer lateral resolution and sub-angstrom vertical resolution. High quality uncoated probes have 1-10nm tip radii, allowing for nanometer resolution measurements of atomically flat surfaces and gradually decreasing resolution for surfaces with higher roughness. This resolution, however, can be achieved only if the original tip geometry is preserved during the measurement by limiting tip wear.

The two most important tip-related factors that affect the quality of AFM images are the apex radius of curvature and the sidewall angles (Fig. 2.7). The tip radius is the real limitation of the ultimate resolution that an AFM measurement can yield. Sidewall angles affect the accurate imaging of steep slopes.

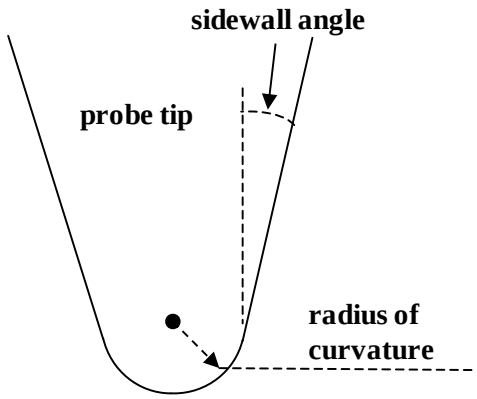


Fig. 2.7: Cross section schematic of a probe tip.

The main effects of tip geometry are shown in Fig. 2.8 and explained below. These effects determine how closely AFM images represent the real surface topography.

1. High surface features are broadened laterally. The broadening at the top of the feature is equal to the tip-curvature diameter and increases additionally down the sidewalls (Fig. 2.8 A) due to the tip sidewall angles (see point 3 below). The smaller the feature, the more noticeable this effect is.

2. Sharp feature edges are rounded. The rounding occurs for both acute and obtuse angles (Fig. 2.8 B, Fig. 2.8 D). Again, the smaller the edge's radius of curvature, the more noticeable this effect is.

3. Measurable sidewall angles are limited to the tip sidewall angles. All sidewalls with steeper slopes appear to have the tip sidewall angles (Fig. 2.8 C). Less steep sidewalls can be measured without distortions (Fig. 2.8 F).

4. Trenches become narrower. Lateral dimensions of the trenches shrink (Fig. 2.8 G) as a direct consequence of the broadening of high features (see point 1 above).

5. Trench depth between nearby structures can be reduced due to the inability of the tip to enter the narrow space in-between and touch the bottom (Fig. 2.8 H, Fig. 4). Step heights around broader trenches are measured correctly (Fig. 2.8 E).

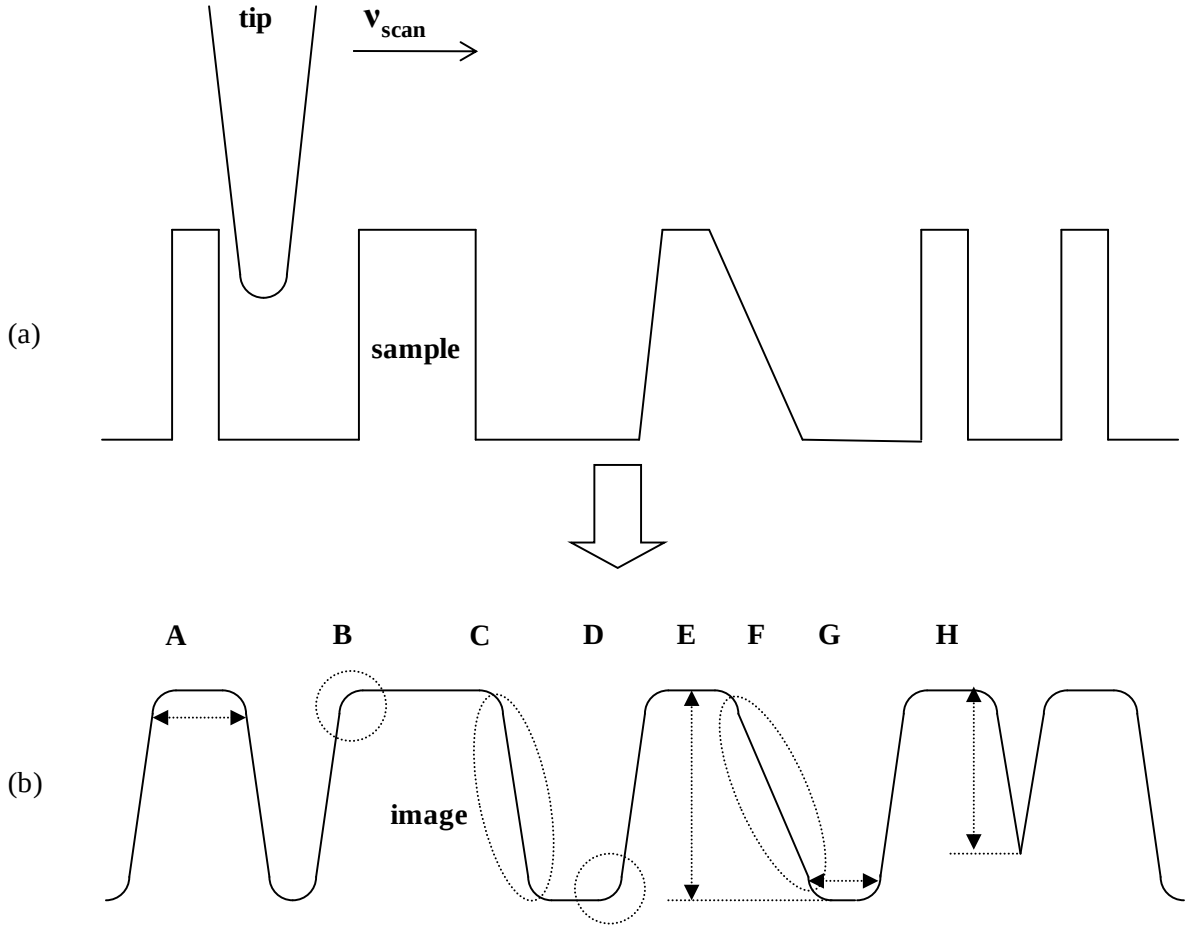


Fig. 2.8: Schematic cross sections of (a) a scanned surface and (b) its image illustrating different image distortion effects: lateral broadening (A), edge rounding (B,D), slope reduction (C), trench narrowing (G) and trench depth reduction (H).

2.8. AFM in relation to other measurement techniques

There are already numerous well established techniques available for process control during MOS fabrication and for evaluation of the fabricated devices. However, it is easy to see how AFM and C-AFM can contribute significantly to semiconductor technology. On one hand, there are the process control techniques: optical microscopy, scanning electron microscopy, tunneling electron microscopy, scanning tunneling microscopy, scanning transmission microscopy, elipsometry, interferometry, profilometry, etc. They have different operation principles and offer different levels of resolution. These measurement techniques are indispensible during MOS fabrication. For example, elipsometry allows precise measurements of thin film thicknesses. Interferometry is useful for step height evaluation over large distances. Scanning electron microscopy offers a good overview of three dimensional samples. However, none of these techniques can offer the combination of nanometer resolution imaging in all three dimensions, non-destructibility of the sample, ease of sample preparation, low maintenance, fast measurement time and the possibility for simultaneous measurements of topography and other surface properties.

Device evaluation techniques give important information about the electrical performance of MOS devices and from there information about the gate dielectric can be extracted. However, these techniques can be applied only after device fabrication is complete. MOS device fabrication is a time and effort consuming process. Considering this, all information available upfront that can give clues to optimization steps is welcome, because it could save entire optimization cycles. Additionally,

measurements yield information about the dielectric strength, breakdown field, leakage currents, charge trapping, etc. of the gate dielectric as a whole. The entire dielectric surface is simultaneously contacted by the gate electrode and the effects of different point peculiarities on the dielectric are averaged over the entire dielectric areas. For example, even a single weak spot in the entire gate dielectric area below the gate electrode will result in high leakage currents that will dominate the device leakage characteristics. The tunneling currents through the rest of the dielectric will be 'hidden' and therefore impossible to investigate. On the other hand, conductive AFM performed directly on the non-covered dielectric surface allows establishing a nanometer wide MOS structure using the probe tip as a gate. This gives the possibility of isolating and studying separately individual defects on the dielectric surface. This is a key advantage for studying novel gate dielectric materials because there is virtually no information about the charge transport mechanisms in them. These mechanisms could differ significantly from those in SiO_2.

3. MOS integrated circuit basics

3.1. Operation principles of MOS devices

Metal-oxide-semiconductor (MOS) devices, whose operation is based on the field effect, are the basic building elements of present-day digital integrated circuits.

3.1.1. The MOS capacitor

The MOS capacitor is the simplest MOS device. It has two terminals, one connected to a metal electrode called 'gate' and the other – to the silicon substrate. The two electrodes are separated by an isolating film, called the 'gate insulator' or 'gate dielectric', the latter term being more commonly used. The gate dielectric is usually a thin film of thermally grown silicon dioxide. Because of the simplicity of its fabrication and analysis, the MOS capacitor is a very useful structure for studying the properties of MOS stacks.

A p-type MOS (pMOS) capacitor is one with a p-doped substrate where holes are the majority carriers and electrons are the minority carriers. A n-type MOS (nMOS) capacitor has an n-doped substrate where electrons are majority carriers and holes are minority carriers. For simplicity, only pMOS capacitor operation is described here. All working principles can be applied analogically to nMOS capacitor operation.

The electron energy states of MOS structures are visualized schematically by energy band diagrams. The band diagram of a pMOS structure with an aluminum gate is shown in Fig. 3.1, where E_o is the energy of an electron in vacuum, E_{FM} - the Fermi level in the metal electrode, E_F - the Fermi level of Si, E_i – the intrinsic level of Si, χ - the electron affinity of Si, φ_M - the metal work-function, φ_{MS} – the work-function difference between the metal and Si, φ_F – the Fermi potential of Si, E_G – the Si bandgap, E_c – the energy level of the conductance band edge of Si and E_v - energy level of the valence band edge of Si [18].

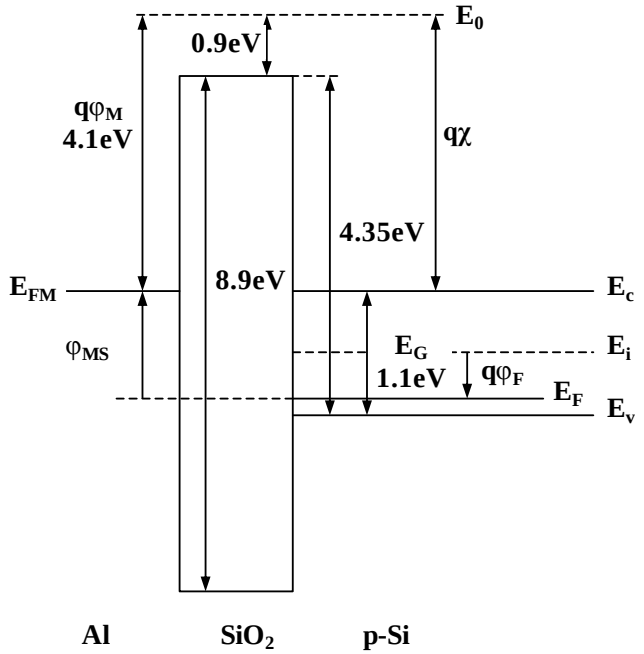
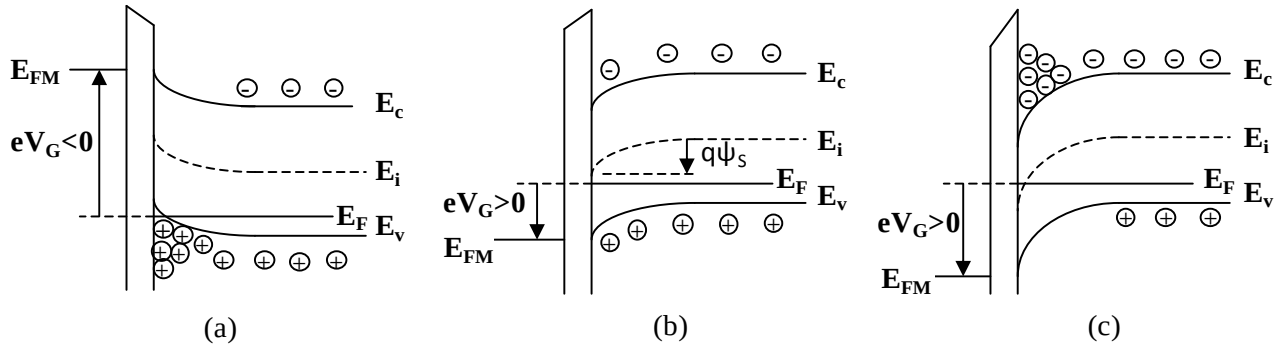


Fig. 3.1: Band diagram of an Al-SiO_2-pSi pMOS capacitor in flat-band condition.

When the gate voltage V_G is equal to φ_{MS}, the MOS capacitor is in so called 'flat-band' condition. In this case the energy levels in the semiconductor near the interface with the dielectric and away from it are equal and so is the concentration of majority and minority carriers. The large bandgap of SiO_2 and the band edge offsets between Si and SiO_2 ensure that there is no carrier transport between the gate and the substrate through the dielectric.

Depending on the applied voltage between the gate and the substrate of the pMOS capacitor, the semiconductor surface is in one of three different states. In case the gate voltage is more negative than the flatband voltage ($V_G<V_{FB}$), additional majority carriers are attracted towards the gate and their concentration at the semiconductor surface increases. This is the accumulation region (Fig. 3.2a). When the gate voltage becomes slightly more positive than the flatband voltage ($V_G>V_{FB}$), the holes are repelled from the surface, leaving uncompensated acceptor ions. At the same time, minority carriers start to build up. The surface is in depletion (Fig. 3.2b). For further more positive gate voltages the minority carrier concentration becomes larger than that of the majority carriers. This is the onset of inversion. Strong inversion is the region in which the minority carrier concentration at the surface substantially exceeds the majority concentration in the bulk of the semiconductor (Fig. 3.2c). The voltage at which the minority concentration at the surface equals the majority concentration in the bulk is called the threshold voltage (V_T).



(a) (b) (c)

Fig. 3.2: Band diagram of a pMOS capacitor in (a) accumulation, (b) depletion and (c) inversion.

The different regions according to the band bending ψ_S (Fig. 3.2) are defined in Table. 1.

$\psi_S<0$	accumulation
$\psi_S=0$	flatband condition
$\psi_S>0$ and $\psi_S<\varphi_F$	depletion
$\psi_S=\varphi_F$	intrinsic condition
$\psi_S>\varphi_F$ and $\psi_S<2*\varphi_F$	weak accumulation
$\psi_S=2*\varphi_F$	threshold voltage
$\psi_S>2*\varphi_F$	strong inversion

Table 1: Condition of the semiconductor surface in a MOS capacitor depending on band bending.

An important electrical characteristic of a MOS capacitor is the dependence of the voltage on the differential capacitance (from here on referred to simply as 'capacitance'). Unlike the case with metal-insulator-metal capacitors, the capacitance of a MOS capacitor varies with the applied voltage. It also depends on the substrate doping. The total differential capacitance of the MOS structure is the series

equivalent of the voltage-independent insulator capacitance C_i and the varying semiconductor capacitance C_s (Eq. 3.1).

$$C \equiv \frac{\partial Q_G}{\partial V_G} = \frac{1}{1/C_i + 1/C_s} \tag{3.1}$$

To measure the capacitance a sinusoidal AC signal with small amplitude (e.g. 20mV) is superimposed with the gate voltage. Typical pMOS capacitor capacitance-voltage characteristics are shown in Fig. 3.3. If V_G is swept sufficiently slowly so that the system is in equilibrium for every value of V_G (e.g. 0.2V/s) and the AC signal frequency is sufficiently low so that the generation-recombination rates of minority carriers at bulk traps can keep up with the small signal variations (e.g. 10kHz), the capacitance-voltage characteristic follows curve ***a*** in Fig. 3.3. In accumulation the AC signal changes the charge directly at the oxide interface and as a result the total capacitance C is close to the oxide capacitance C_i. As V_G increases, a depletion layer is formed that acts as a dielectric of thickness e in series with the oxide. Thus, C decreases according to Eq. 3.2, where K_i is the dielectric constant of the gate dielectric, K_s - the dielectric constant of Si and d – the dielectric thickness.

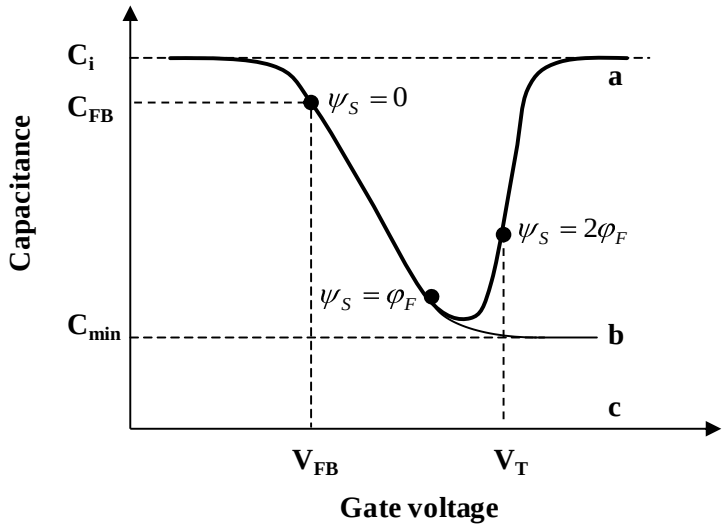


Fig. 3.3: Graph of the capacitance-voltage characteristic of a pMOS capacitor at low frequency (line a), high frequency (line b) and deep-depletion (line c).

$$C = \frac{\varepsilon_0}{d/K_i + e/K_s} \tag{3.2}$$

In inversion the capacitance increases again as a result of the attracted minority carriers whose concentration increases exponentially with the applied voltage. Finally, the total capacitance saturates to the oxide capacitance C_i. Once strong inversion is reached, e reaches a saturation value because the semiconductor is shielded from further penetration of the electric field by the inversion layer.

At higher frequencies of the AC signal (e.g. 1MHz) the generation-recombination rates of the minority carriers are not high enough to keep up the equilibrium. As a result, the high-frequency C-V curve shows no capacitance increase in inversion (curve ***b***). In case the gate voltage is swept at such a high rate, that the MOS system can not keep in equilibrium (e.g. 2V/s), i.e. faster than the minority carriers' response time, an inversion layer can not be formed and instead the depletion region keeps becoming

wider, resulting in continuous lowering of the capacitance without saturation - the so called 'deep depletion' (curve ***c***).

Charges

A non-ideality that is present in even the highest quality gate dielectrics is the presence of trapped charges in the gate dielectric itself as well at the dielectric – silicon interface where the chemical composition of the dielectric is not stoichiometric. The basic types of charges are described below [18].

Fixed charges (Q_f) are located close to the interface and do not exchange charges with the substrate. They cause parallel shifts ΔV of the capacitance-voltage characteristics according to Eq. 3.3.

$$\Delta V = -Q_f / C_i = -Q_f d / \varepsilon_i \tag{3.3}$$

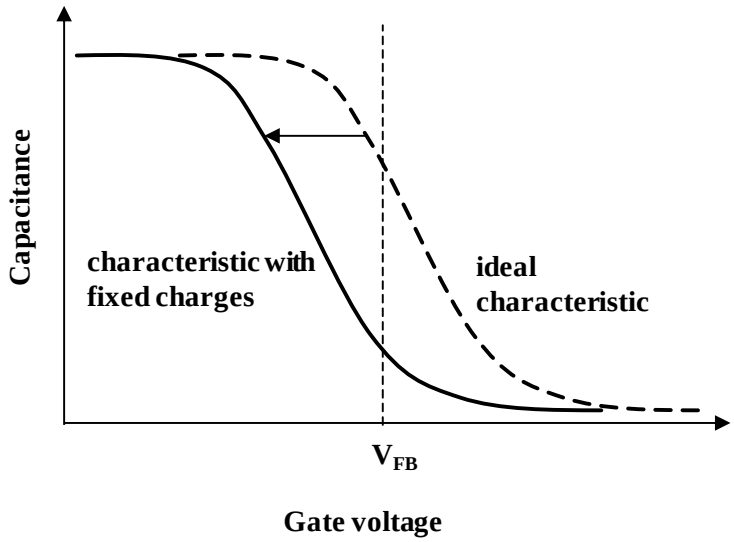


Fig. 3.4: Graph illustrating the effects of positive fixed charges on the capacitance-voltage curve of a pMOS capacitor.

Interface state charges (Q_{it}), also called interface traps, are located very close to the interface and have such energy levels that they are able to exchange charges with the semiconductor within moderately short times. They are distributed across the bandgap and their density D_{it} is defined as the number of charges per unit surface area per unit energy (Eq. 3.4).

$$D_{it} \equiv \partial N_{it} / \partial E \tag{3.4}$$

The interface state density distribution $D_{it}(E)$ for SiO_2 is U-shaped with a maximum near the conduction and valence band edges [18]. This type of charges can interact with the conduction band of silicon by emitting or capturing electrons and with the valence band by capturing or emitting holes. They cause stretch-out of the capacitance-voltage characteristic and an additional distortion of the low frequency *C-V* characteristic (capacitance increase ΔC) as shown in Fig. 3.5.

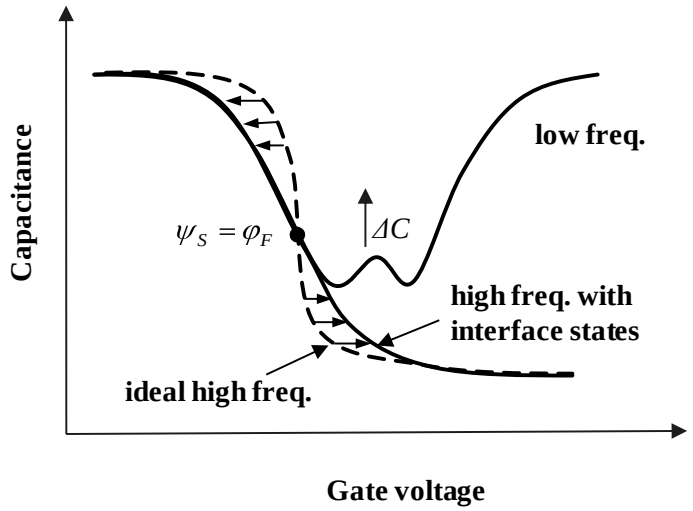


Fig. 3.5: Graph illustrating the effects of interface states on the capacitance-voltage characteristic of a pMOS capacitor.

Oxide traps are distributed within the dielectric layer and able to trap and emit charges but at much slower rates that interface traps. When charged, their effect on the capacitance-voltage characteristic is similar to that of fixed charges, except that the distance d in Eq. 3.3 is replaced by the distance of the oxide traps from the silicon – oxide interface.

Mobile ionic charges, as the name suggests, can move within the dielectric. The most common mobile charges in silicon dioxide are sodium ions.

Gate-dielectric current-voltage characteristic

Due to the non-ideality of the gate dielectric and quantum-mechanical effects, there are measurable gate currents even at relatively low electric fields across the oxide F_i despite the large bandgap of the dielectric and the large band edge offsets. Carrier transport occurs through several different mechanisms (Fig. 3.6) [18].

1. Fowler-Nordheim emission is due to field assisted tunneling to the conductance band of the dielectric at large normal electric fields (F_i>6-7MV/cm). F-N tunneling depends on F_i stronger than the other transport mechanisms.

2. Direct tunneling refers to transfer of electrons from the conduction band of the substrate to the conductor band of the gate or vice-versa directly through the oxide without change in their energy. Both F-N and direct tunneling occur in thin dielectrics ($d<10nm$). A simple criterion to determine that direct tunneling occurs rather than F-N tunneling is if the product of the electric field F_i multiplied by the dielectric thickness is lower than the barrier height for electrons at the metal-dielectric boundary φ_{BC} (Eq. 3.5).

$$F_i d < \varphi_{BC} \tag{3.5}$$

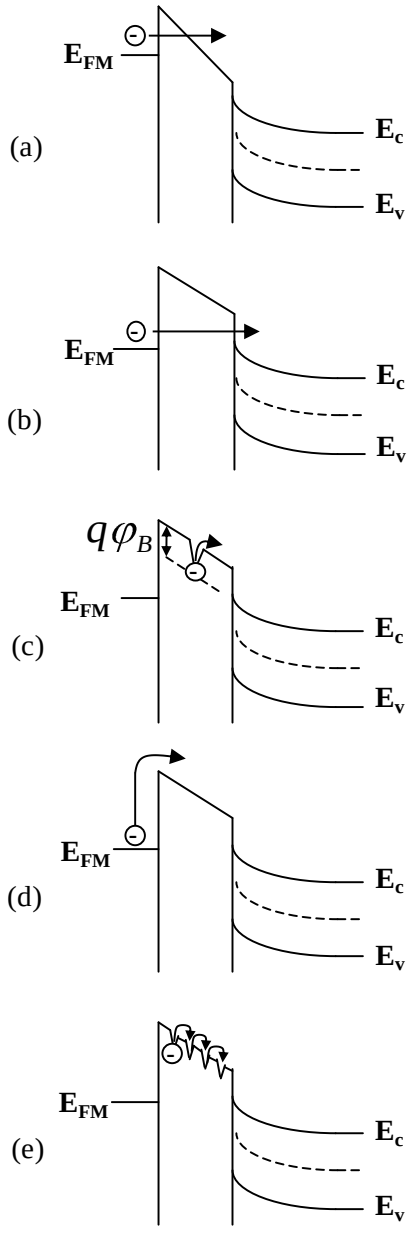
For SiO_2 Fowler-Nordheim tunneling is dominant for thicknesses d>4.5nm while direct tunneling is dominant for smaller thicknesses.

3. Poole-Frenkel emission occurs due to field-assisted thermal excitation of trapped electrons into the conduction band of the gate dielectric.

4. Schottky emission is the thermionic emission across the semiconductor-dielectric or metal-dielectric interface.

5. Ohmic or hopping conduction is the result of thermally excited electrons hopping between isolated traps.

For a given gate dielectric, different transport mechanisms can contribute to or dominate the dielectric leakage depending on F_i and temperature. The temperature dependence of the leakage can help in determining its origin. As seen from the formula in Fig. 3.6, direct tunneling and F-N tunneling are largely independent of temperature unlike the other mechanisms.



$$J_{FNT} = \frac{A}{4\varphi_B} F_i^2 \exp\left(-\frac{2B\varphi_B^{3/2}}{3F_i}\right)$$

$$J_{DT} = \frac{A}{d^2}\left[\left(\varphi_B - \frac{V}{2}\right)\exp\left(-Bd\sqrt{\varphi_B - \frac{V}{2}}\right) - \left(\varphi_B + \frac{V}{2}\right)\exp\left(-Bd\sqrt{\varphi_B + \frac{V}{2}}\right)\right]$$

$$J_{PF} \approx F_i \exp\left(-\frac{\varphi_B - 2\sqrt{F_i / C}}{\varphi_t}\right)$$

$$J_{SE} \approx T^2 \exp\left(-\frac{\varphi_B - 2\sqrt{F_i / C}}{\varphi_t}\right)$$

$$J_{OC} \approx F_i \exp\left(-\frac{\Delta E_{ah}}{kT}\right)$$

where $A = \frac{q^2}{2\pi h}$, $B = \frac{4\pi\sqrt{2m^* q}}{h}$, $C = \frac{4\pi\varepsilon_i}{q}$ and ΔE_{ah} – hopping activation energy

Fig. 3.6: Band diagrams and current density formula for different mechanisms of carrier transport through dielectrics: (a) Fowler-Nordheim tunneling, (b) direct tunneling, (c) Poole-Frenkel emission, (d) Schottky emission and (e) Ohmic conduction.

3.1.2. The MOSFET

The metal-oxide-semiconductor field-effect transistor or MOSFET (Fig. 3.7) has two additional terminals compared to the MOS capacitor. In its role as a voltage-controlled switch, the MOSFET is the single most important building block of digital electronic circuits.

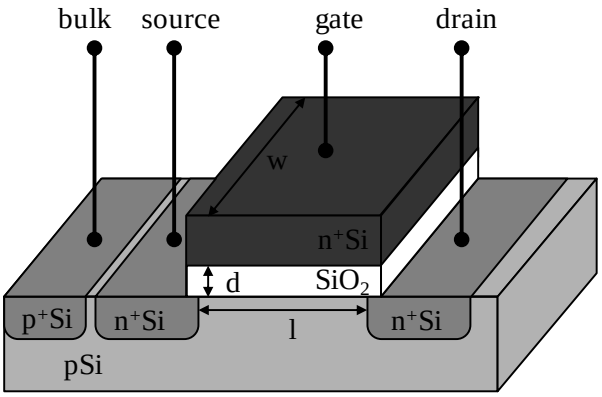


Fig. 3.7: nMOSFET schematic.

An n-type MOSFET or nMOSFET has a p-type substrate and a p-type MOSFET or a pMOSFET has an n-type substrate. Only nMOSFET operation is discussed for simplicity. If a large enough gate voltage is applied so as to induce an inversion layer, an n-channel is formed between the source and the drain. If at the same time a voltage is applied also between the source and drain regions, an electric current flows between these two electrodes. For small drain-to-source voltages V_{DS}, the channel is almost uniform in thickness along its length. In this region the channel has ohmic conduction and the drain-to-source current I_{DS} is linearly proportional to V_{DS} (Fig. 3.8). When $V_{DS} << \varphi_F$ (gradual channel approximation) the drain current is described by Eq. 3.6 [18], where l is the channel length, w - the channel width, μ_n - the carrier mobility [19], V_{GS} - the gate-to-source voltage and V_{DS} – the drain-to-source voltage.

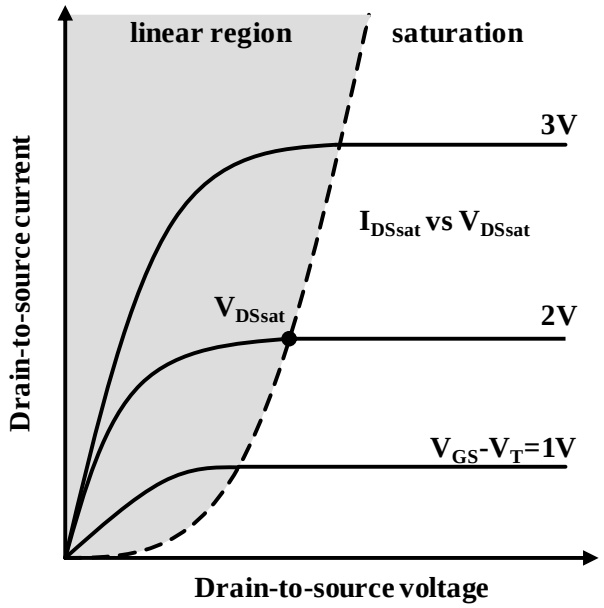


Fig. 3.8: Graph of nMOSFET output characteristic.

$$I_{DS} = \frac{w}{l}\mu_n C_i (V_{GS} - V_T - V_{DS}/2)V_{DS} \tag{3.6}$$

For higher drain voltages, the inversion layer near the drain decreases and becomes zero at a certain point, the so called 'pinch-off'. For even higher voltages, the pinch-off moves towards the source, leaving a depletion region near the drain. The current in this case is due to drift in the conducting channel and injection into the depletion region. After pinch-off occurs, the drain current does not increase further since the potential drop on which it depends is fixed at the value of the saturation voltage V_{DSsat}. This is the saturation region.

For a given V_{GS}, I_{DS} gradually goes into saturation at V_{DSsat} defined in Eq. 3.7.

$$V_{DSsat} \equiv V_{GS} - V_T \tag{3.7}$$

At this point the saturation current I_{DSsat} is calculated by Eq. 3.8.

$$I_{DSsat} = \frac{w}{l}\mu_n C_i \frac{(V_{GS} - V_T)^2}{2} \tag{3.8}$$

The dependence of V_{GS} on I_{DS} is called the transfer characteristic of the MOSFET and it is shown in Fig. 3.9. In the subthreshold region I_{DS} increases exponentially with V_G starting from the onset of weak inversion. Unlike the on-state drift current, the subthreshold current is diffusion based.

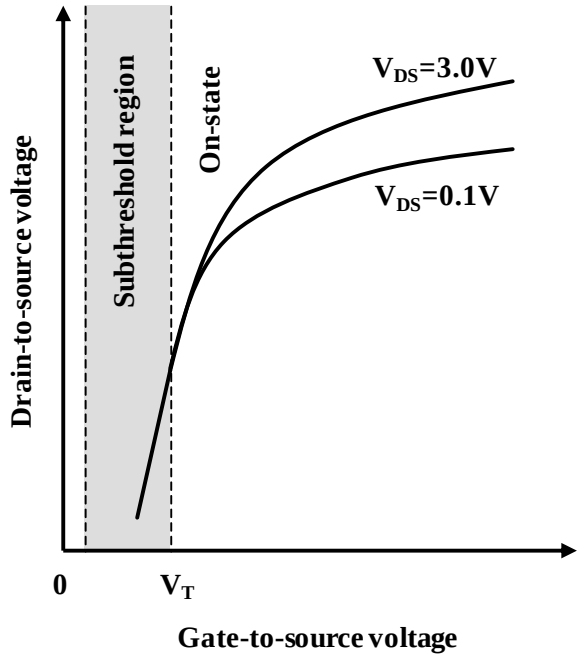


Fig. 3.9: Graph of nMOSFET transfer characteristic.

Important MOSFET parameters

Important parameters for MOSFETs are the subthreshold swing, drift velocity and carrier mobility [18]. The subthreshold swing characterizes the subthreshold behavior of MOSFETs. It is defined as the

voltage needed to increase the current tenfold (Eq. 3.9), where C_D is the depletion capacitance. Typical values of the subthreshold swing for ULSI are 80±10mV/dec.

$$S \equiv \ln(10)\frac{\partial V_{GS}}{\partial \ln I_{DS}} \cong \varphi_t \ln(10)(1+\frac{C_D}{C_i}) \tag{3.9}$$

Interface states increase the swing according to Eq. 3.10.

$$\Delta S = \varphi_t \ln(10) q D_{it} / C_i \tag{3.10}$$

Drift velocity v_d is the velocity at which charge carriers travel under the influence of a lateral electric field F_{lat} (Eq. 3.11).

$$v_d = \mu_n F_{lat} \tag{3.11}$$

In Eq. 3.11 μ_n is the so called 'mobility'. Mobility can be defined in different ways, the two most common of which are the effective mobility μ_{eff} (Eq. 3.12) and the field-effect mobility μ_{FE} (Eq. 3.13), where g_m is the transconductance.

$$\mu_{eff} \equiv \frac{1}{(w/l)V_{DS}C_i}\frac{I_{DS}}{V_{GS}-V_T} \tag{3.12}$$

$$\mu_{FE} \equiv \frac{1}{(w/l)V_{DS}C_i}\frac{\partial I_{DS}}{\partial V_{GS}} = \frac{1}{(w/l)V_{DS}C_i} g_m \tag{3.13}$$

Measurement of interface states with charge pumping

Charge pumping (CP) is an electrical characterization technique for the determination of the state density at the interface between the substrate and the gate dielectric in small geometry MOSFETs [20]. Charge pumping offers low detection limits below 10^{-9}cm^{-2}eV^{-1}. The measurement setup and a schematic band diagram explanation for an nMOSFET are shown in Fig. 3.10.

When an nMOSFET is pulsed from accumulation to inversion, holes are emitted from the interface states towards the valence band (region 1 in Fig. 3.10b) and flow towards the substrate. Initially, the rate of emission of trapped charge imposed by the emission process is able to keep up with the voltage sweep and the channel is in steady-state condition. At a certain moment rate becomes smaller than the one necessary for keeping the dynamic equilibrium. The channel is now in non-steady state and the emptying of traps is limited by the emission process (region 2). The transition from steady to non-steady state occurs at $V_G \approx V_{FB}$ [20]. When $V_G \approx V_T$ the interface states that are not yet emptied of holes are filled with electrons coming from the source and drain (region 3). When the gate is pulsed back into accumulation, electrons are emitted from the interface states in a steady state and flow back to the source and drain (region 4). Below the threshold voltage, a non-steady state emission of electrons from the interface states to the conduction band occurs. When the gate voltage reaches approximately the V_{FB}, holes fill the remaining traps occupied with electrons. The result of the entire cycle is a net current flow of negative charge from the source and drain to the substrate (Fig. 3.11).

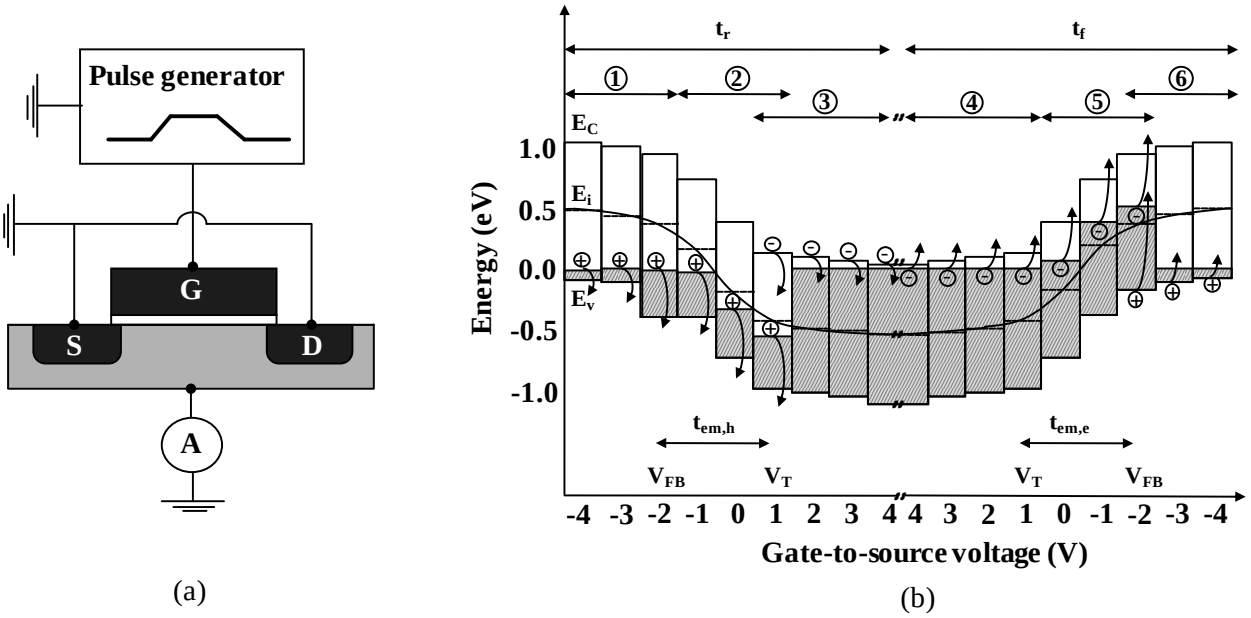


Fig. 3.10: Charge pumping (a) measurement setup schematic and (b) band-diagram explanation of one charge pumping cycle [20].

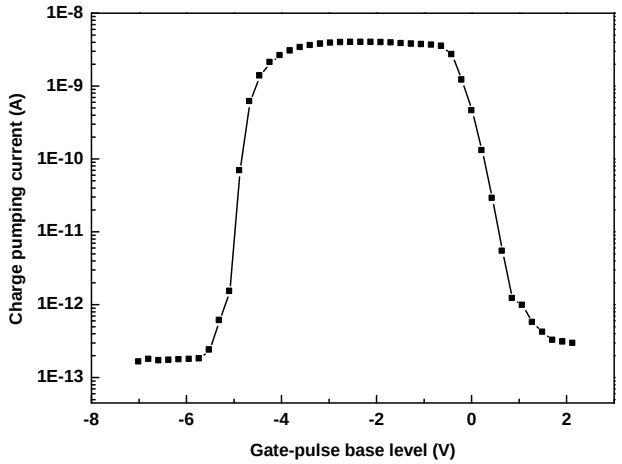


Fig. 3.11: Graph of a charge pumping measurement on a SiO_2 gate dielectric nMOSFET with a constant pulse amplitude and varying base level.

The interface state density D_{it} in case of trapezoid gate pulses is determined by Eq. 3.14 [20], where q is the elementary charge, I_{CP} – the charge pumping current measured at the substrate, f – the pulse frequency, t_r and t_f – the pulse rise and fall times, v_{th} – the thermal velocity of the carriers, n_i – the intrinsic concentration of carriers, and σ_n and σ_p – the capture cross sections for electrons and holes.

$$I_{CP} = 2qD_{it} f . A_G . kT \left[\ln(v_{th} n_i \sqrt{\sigma_n \sigma_p}) + \ln(\frac{|V_{FB} - V_T|}{|\Delta V_G|} \sqrt{t_f t_r}) \right] \tag{3.14}$$

By varying the rise and fall times of the gate pulses, it is possible to extract information about the energy distribution of the interface traps [20]. The trap densities for energy levels in the upper and lower part of the bandgap are determined by Eq. 3.15 and Eq. 3.16 respectively.

$$D_{it}(E_1) = -\frac{t_f}{qA_G kTf}\frac{\partial I_{CP}}{\partial t_f} \tag{3.15}$$

$$D_{it}(E_2) = -\frac{t_r}{qA_G kTf}\frac{\partial I_{CP}}{\partial t_r} \tag{3.16}$$

The energy levels E_1 and E_2 in Eq. 3.15 and Eq. 3.16 are determined by Eq. 3.17 and Eq. 3.18 respectively.

$$E_1 = E_i - kT\ln\left[\upsilon_{th}\sigma_n n_i \frac{|V_{FB} - V_T|}{|\Delta V_G|} t_f\right] \tag{3.17}$$

$$E_2 = E_i + kT\ln\left[\upsilon_{th}\sigma_p n_i \frac{|V_{FB} - V_T|}{|\Delta V_G|} t_r\right] \tag{3.18}$$

Knowing more about the energy distribution of the traps is important because it indicates at what voltages the defects are active (change their occupancy) and contribute more strongly to the deviations from ideal transistor behavior. Furthermore, if the defect energy levels of a certain type of defect (for example oxygen deficiencies on the dielectric-Si interface) have been previously determined by other analytical techniques, energy resolved charge pumping can help going backward and identifying the origin of the traps by their energy distribution signature.

Charge pumping also helps obtaining information about the depth distribution of the interface states by varying the gate pulse frequency and amplitude [21]. High amplitudes and low pulse frequencies allow traps further away from the Si-dielectric interface to change their occupancy states. In this way, it is possible to study 'slow traps' that are located in the oxide at a tunneling distance from the substrate. On the other hand, measuring the frequency dependency of the charge pumping current for high frequencies and extrapolating the measured charge pumping currents to even higher frequencies can give an idea of the effective trap densities of the corresponding devices for real-world high-frequency operation.

3.2. Conventional gate-first integration process flow

MOS capacitors and MOSFETs are manufactured through a series of carefully selected and scrutinized process steps. Although no two manufacturers' processes are completely the same, the basic conventional gate-first CMOS manufacturing process is relatively standardized [22]. Below, a simplified CMOS process flow with n^+/p^+ dual work-function polysilicon gates is outlined and illustrated schematically (Fig. 3.12). This process is suitable for polysilicon gates and silicon dioxide gate dielectrics. For the reasons explained in Chapter 4, significant modifications are required for the process integration of most metal gates and high-*K* gate dielectrics. An alternative processing concept is described in Chapter 8.2.

1. P-well and N-well formation
Processing begins with a bare silicon p- or n-doped (100) wafer with doping the order of $10^{15}cm^{-3}$. p-wells and n-wells are formed on the wafer surface by two lithographic steps and ion implantation using SiO_2 and Si_3N_4 as hard masks. The wells define the nMOSFET bulk doping profile. The wells are driven in by a high-temperature anneal. The final doping is in the range of $10^{16}cm^{-3}$.

2. Device isolation formation
A thin pad oxide and a thicker nitride layer are deposited. Using a resist mask, boron is implanted through the oxide and nitride layers in order to form a channel stopper region under the isolation areas. The next lithography step defines the isolation regions. In these regions the hard mask is removed and after removing the resist, a reactive ion etch (RIE) creates trenches with several hundred nanometer depth. SiO_2 is deposited in order to fill the trenches. Next, the SiO_2 on top of the active areas is polished away by chemical mechanical planarization (CMP). Finally, the nitride layer is removed by a wet etch. This type of isolation is called shallow trench isolation (STI). STI is a comparatively novel technique for device isolation which replaced the older local oxidation of silicon (LOCOS). More details about both types of device isolation techniques can be found in Chapter 3.3.

3. Threshold voltage adjustment
A thin oxide layer is grown. With the help of two more lithography steps, threshold voltage adjustment impurities (boron and phosphorus) are implanted through the thin oxide by a low energy implantation. The implantation defines the nMOSFET surface doping.

4. Gate stack formation
The gate oxide is grown by thermal oxidation and the polysilicon layer is deposited by chemical vapor deposition (CVD) right after pad oxide removal. The polysilicon layer is degenerately doped with phosphorus for single work-function gate processes and undoped for double work-function gate processes. Lithography and RIE define the gates. After resist removal, the gates are reoxidized in order to reinforce the oxide near the gate edges.

5. Source/drain formation
Two lithography steps with an implantation after each one define the source and drain regions. Arsenic and boron are implanted as the n+ and p+ source/drain dopants. For dual gate processes, n+ and p+ gate doping is carried out simultaneously with the source/drain implantation. A high-temperature anneal follows. Its purpose is to active the dopants by repositioning them in the silicon crystal lattice and to cure the damage caused to the silicon surface by the ion bombardment.

6. Interlayer dielectric deposition and structuring. Source/drain activation
The rest of the processing is the so called 'back end of line' (BEOL) processing that seals off the MOS devices and creates contacts among the MOS devices and between them and the IC terminals. A CVD oxide is deposited as an interlayer dielectric (ILD). Contact holes are defined by a lithography step and an etch step.

7. Interconnects/contacts formation
Metals (Ti and Al or Cu) are deposited on the ILD. Interconnects and contacts are defined by a lithography and an etch step.

8. Forming gas anneal and nitride passivation
After back-end formation, a forming gas anneal is performed (95%Ar, 5%H) for passivation of the defects at the Si-SiO_2 interface. A final nitride passivation may be carried out for contamination protection, mainly against humidity and positive alkali ions.

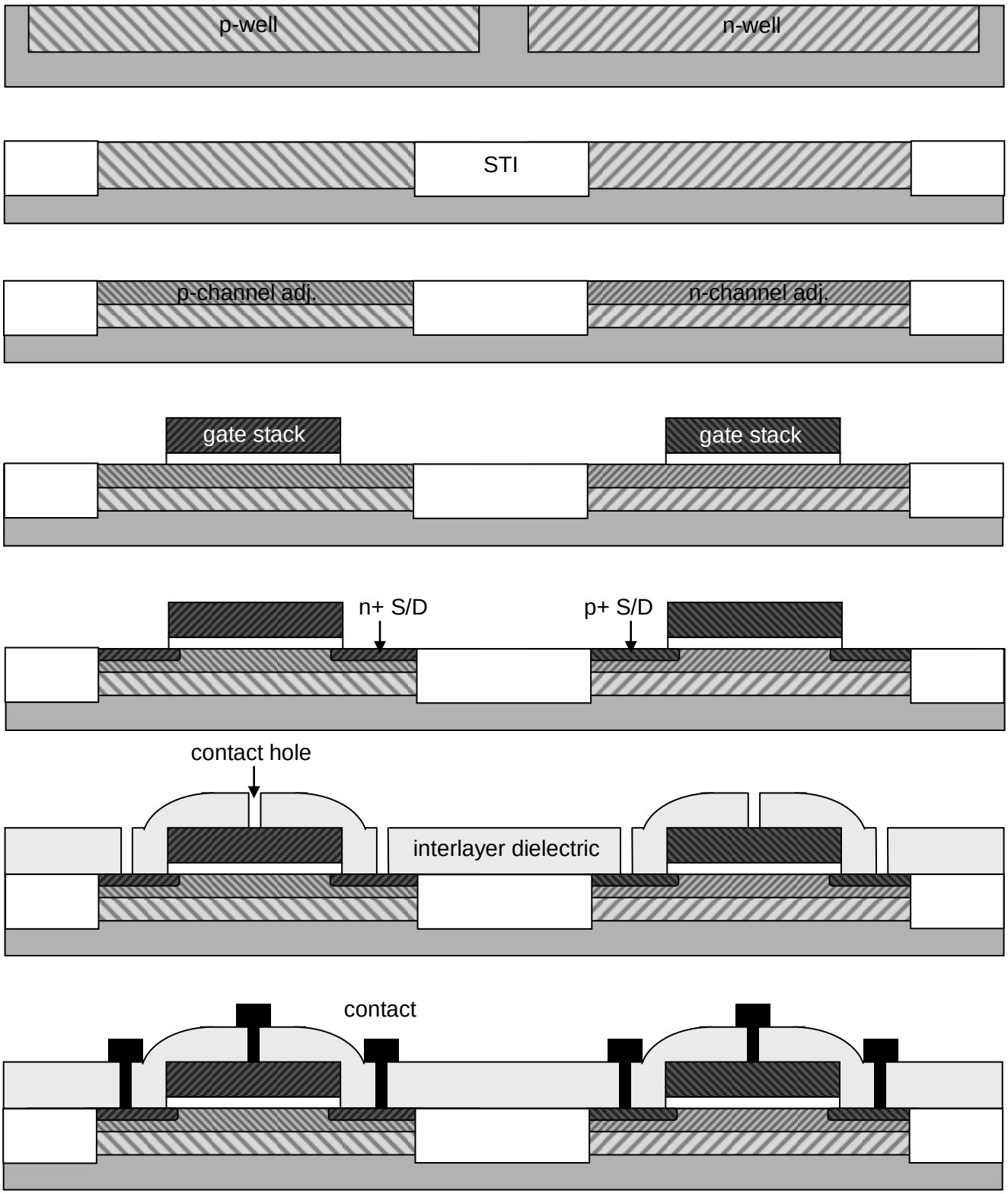


Fig. 3.12: Schematics of integration modules in conventional CMOS integration with STI device isolation.

3.3. Device isoaltion: from LOCOS to STI

As device dimensions shrink, devices are brought closer together. This leads to electrically related problems such as crosstalk and surface leakages [23]. In 1959 it is discovered that a thermally grown SiO_2 on the Si surface dramatically reduces surface leakages by surface passivation [23]. This gives the

birth of the first device isolation process – the Planox [24] that utilizes oxidation of the entire silicon surface. In 1970 it is discovered that Si_3N_4 is stable against oxidation. Si_3N_4 is then applied for oxidizing Si selectively and the so-called local oxidation of silicon (LOCOS) device isolation technology is developed [25]. The non-scalability of the LOCOS brings the need for a novel device isolation scheme for ultra large scale integration (ULSI) ICs, namely – shallow trench isolation [26]. With its abrupt active area – to – isolation transition STI allows scaling to continue further to deep submicrometer technology nodes.

LOCOS relies on local thermal oxidation of parts of the silicon surface for device isolation [27; 28]. It remains the main isolation technique for many years because of its simplicity and low cost. A standard process flow for manufacturing LOCOS proceeds as follows (Fig. 3.13a). Starting with a blanket preimplanted wafer, windows are opened with lithography and etching in a deposited nitride layer on the silicon surface in the places where the isolation areas are to be created. A long high-temperature wet oxidation follows which oxidizes the open silicon areas according to the reaction in Eq. 3.19, where '(S)' indicates a solid phase and '(V)' – a vapor phase.

$$Si(S) + 2H_2O(V) \rightarrow SiO_2(S) + 2H_2(V) \tag{3.19}$$

The selective oxidation is based on the much lower diffusion rates of oxygen atoms through silicon nitride compared to silicon dioxide. Because of this difference, the nitride layer is an efficient mask for oxidation. Since the growth of a certain amount of SiO_2 consumes approximately 44% as much silicon, as the oxidation process takes place, the oxide front moves deeper below the original silicon surface, separating the active silicon areas by oxide isolation areas. After nitride removal, LOCOS isolation is complete (Fig. 3.13b) and processing continues with threshold voltage implantation and gate stack formation.

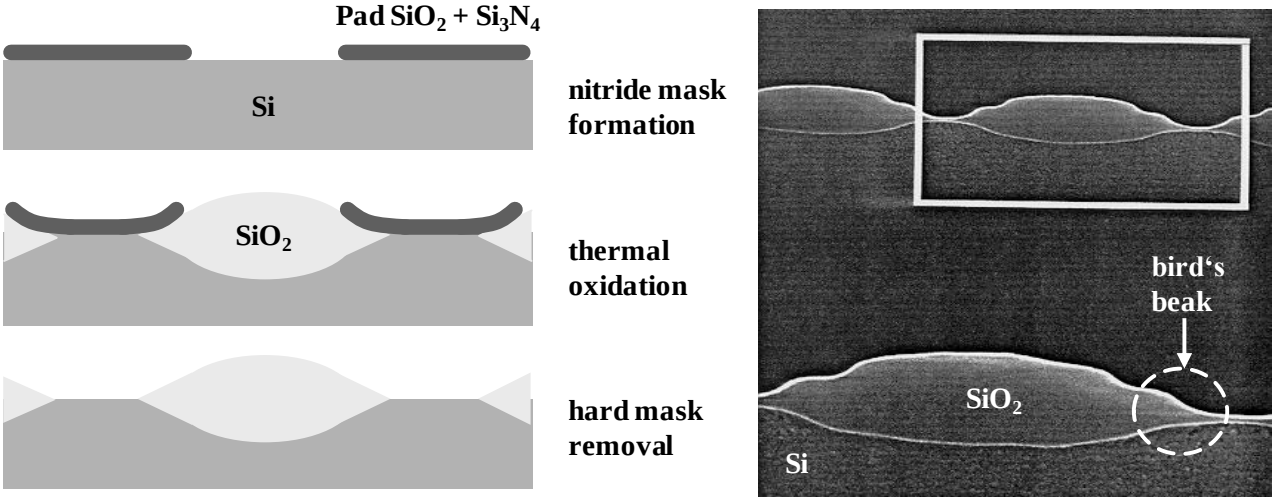


Fig. 3.13: (a) Schematic cross sections illustrating the major LOCOS fabrication steps and (b) SEM image of partially fabricated wafer with LOCOS isolation and zoom-in.

Thermal oxidation is both the reason why this technology has been so successful and the cause for its shortcomings and eventual replacement by STI. On one hand, oxidation makes LOCOS easy to implement because of the fact that the isolation areas are created with a single oxidation step. On the other hand, the geometry is not easily controllable. First of all, 56% of the oxide is situated above the silicon surface where it does not serve as isolation, but nevertheless creates additional topography of several hundred nanometers on the wafer surface. As devices scale down, more advanced lithography tools are required for imprinting smaller and more precise images on the surface. However, these tools

are only applicable for very planar surfaces because of their limited depth of focus. Therefore, they are not compatible with LOCOS isolated wafers, projecting unsharp images on the surface in consecutive lithography steps (for example the critical gate lithography with the gates running from isolation to isolation area through active areas). A second important disadvantage of LOCOS is the lateral encroachment [27; 28] of the oxide into the active area below the nitride layer, which leads to effective reduction of the active area dimensions by forming a 'bird's beak' at the active-isolation transition. This transition, which is typically 100-200nm in length even for advanced LOCOS integration schemes (and sometimes several times that), limits the packing density of LOCOS isolated ICs. Thirdly, there is uncontrollable thinning of the oxide in small isolation areas [29] that additionally reduces the isolating properties of LOCOS for high density designs. There are also additional limitations to using LOCOS, including doping encroachment, susceptibility to latch-up and oxidation induced stress in the substrate [27; 28; 29].

STI relies on RIE etched and CVD oxide filled trenches as isolation rather than areas formed by oxidation. In the 1990s STI gradually replaces LOCOS as the state-of-the-art device isolation technology for ULSI ICs, largely thanks to the development and maturing of the CMP process. Of course, LOCOS is not simply left behind and forgotten. As every technology that has been developed and applied over such a long period of time, LOCOS continues its development and various modifications allow its application in further technology generations than initially expected. Although there are doubts about the scalability of the technology beyond 1μm, ingenuous approaches such as poly-buffered LOCOS [30] and poly-buffered recessed LOCOS [31] extend it even at 0.25μm [29]. Still, complexity for advanced LOCOS schemes increased dramatically and did not offer any significant advantages over STI. The latter isolation technology presented a more direct approach for meeting the different requirements for sub-half-micrometer nodes.

A schematic overview of the STI fabrication module [26; 28; 29] is shown in Fig. 3.14. Manufacturing starts with a thin pad oxide growth and nitride deposition. The nitride serves as a hard mask for etching and the oxide is a stress reducer. Typical values for the thicknesses of the pad oxide and the nitride are 15-25nm and 100-150nm respectively. The thickness of the nitride layer is determined by the desired post-CMP step height between the active and isolation areas [26]. Lithography defines the isolation areas in which the nitride and oxide are then etched with RIE. Next, isolation trenches with depth of 250-700nm are etched into the substrate with chloride based (e.g. $SiCl_4$) plasma. The pad oxide is underetched slightly. The trench sidewalls are chosen to be at 80-86° to horizontal [32]. The angle must be sufficiently steep in order to avoid reduction of trench depth for small isolation spaces and still not too steep so that optimal trench fill and trench corner properties can be achieved [26]. After trench etch a multipurpose liner oxidation is performed [32]. It reduces RIE damage to the trench walls [33], rounds the bottom trench corners thus reducing mechanical stress and rounds the top trench corner for improved device performance [34]. Liner oxide thickness is in the range of 20-50nm. On top of the thermal oxide a thicker oxide is deposited [32]. The total trench oxide thickness is higher than the combined trench depth plus nitride layer thickness, sometimes as much as twice as high. Oxide deposition technique is chosen in such a way as to allow conformal void-free filling of high-aspect ratio trenches. Tetraethyl orthosilicate – ozone (TEOS:O_3) and high-density plasma oxides are commonly used [32; 35; 36]. A high temperature densification is performed in order to lower the etch rate of the deposited oxide to values close to those of thermal oxide [26; 37]. This is important for avoiding trench oxide loss at the end of the STI module during pad oxide etch and cleaning procedures before gate oxidation.

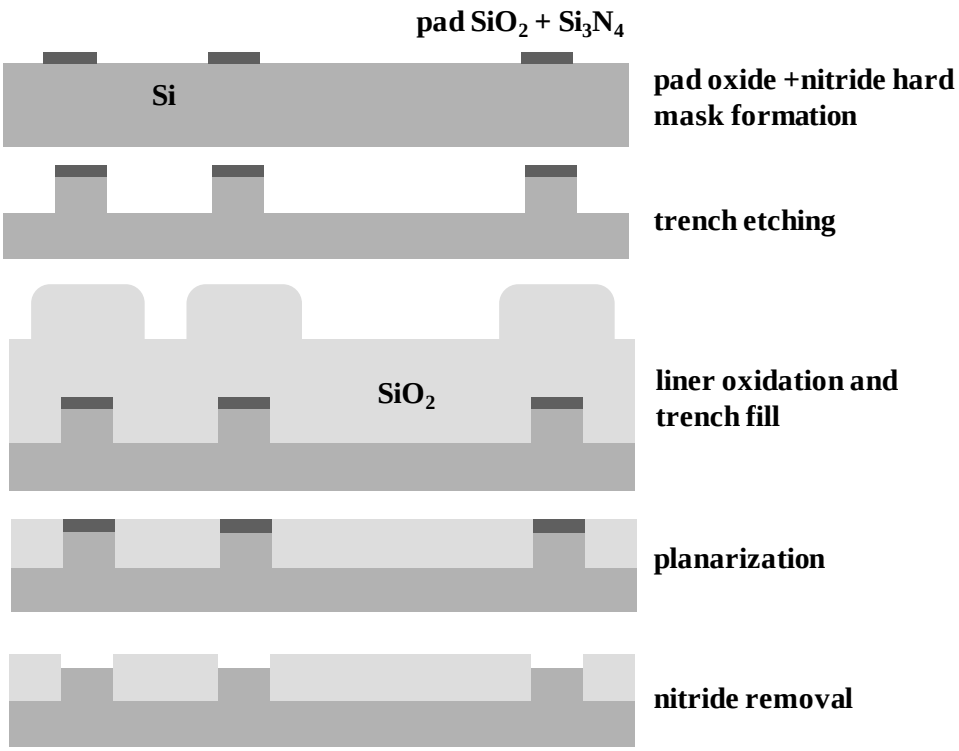


Fig. 3.14: Schematic cross sections showing the major STI fabrication steps.

Chemical mechanical planarization comes next. It removes the oxide on top of the active areas and leaves it only in the isolation trenches [32] by combined chemical and mechanical forces. It is important that oxide removal stops at the level of the nitride stop-layer without reaching the silicon surface in the active areas in order to avoid damage of the silicon substrate and performance degradation of the MOS devices. More information about CMP can be found in Chapter 4.

It should be mentioned here that CMP is not the only option for planarization. More cost effective solutions have been studied and used such as multi layer resist processes and spin-on glass in combination with RIE (Fig. 3.15) [38]. However, the obtained global planarity for such non-CMP processes is inferior to CMP planarity. The former are therefore more suitable for ILD planarization than for STI. There have also been investigations of using multilayer resist/spin-on material techniques for planarization with a combination of RIE and CMP [39].

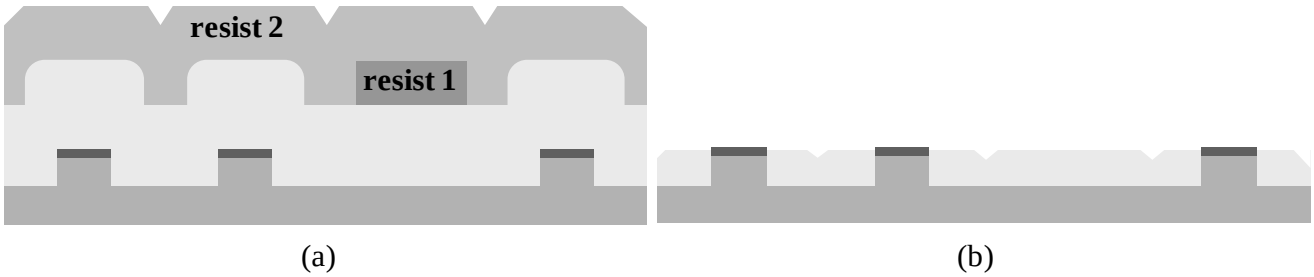


Fig. 3.15: Schematic cross sections of a wafer with block resist-RIE for non-CMP planarization (a) before reactive ion etch and (b) after reactive ion etch.

Cleaning must always follow directly after CMP. A good post-CMP cleaning is very important for removing all slurry and polished material residues from the wafer surface. The procedure is usually a combination of brush and ultrasonic cleaning with water, diluted ammonium hydroxide and diluted HF [40]. Finally, the nitride is stripped with hot phosphoric acid while trying to prevent as much loss

of the trench oxide and damage to the silicon surface as possible. This concludes the STI fabrication module (Fig. 3.16).

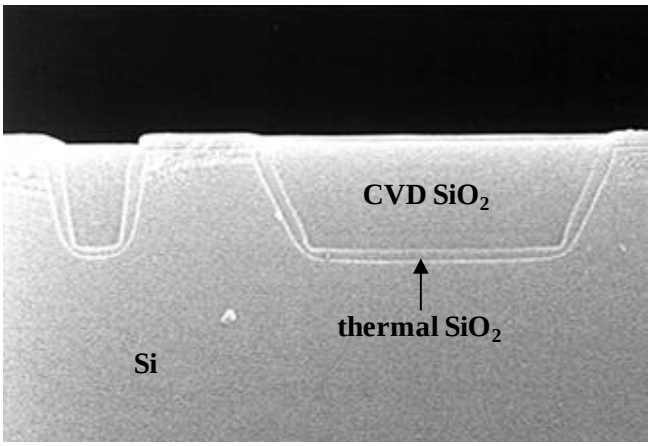


Fig. 3.16: SEM image of a cross section of a wafer with STI.

Looking at the process steps that are responsible for the geometry of the active and isolation areas, it is easy to see how much more scalable STI is compared to LOCOS. Lithography defines the top lateral dimensions of the trenches and therefore they can be scaled with each technology node. RIE defines the depth and sidewall angle controllable. Finally, CMP patterns the oxide laterally and eliminates vertical topography. As additional advantages of STI, the additional degree of latch-up immunity, the lower junction capacitance and the reduced substrate stress can be mentioned.

Achieving a good STI process is a challenging task. In particular, it is difficult to achieve the needed degree of planarity after the polishing step. In order to demonstrate this, a comparison is made between an ideal planarization process and a real-life one.

The chemical mechanical planarization step

An ideal process is characterized by ideal the following features: the two elements in relative motion to each other - the polish pad and the wafer are infinitely stiff; the slurry selectivity is infinitively high (it removes SiO_2 and does not affect Si_3N_4) and wafer pre-CMP geometry is ideal (i.e. oxide and nitride thicknesses and trench depth are constant throughout the wafer and there is no nanotopography on the wafer surface). During the planarization process the pressure distribution across the wafer is constant because of the stiffness of the pad and wafer. This guarantees uniform polish rates of all elevated areas throughout the process. The removal of material from the surface is governed by Preston's equation (Eq. 3.20) where r is the material removal rate, z – Preston's coefficient, P – the applied pressure on the wafer and s – the relative linear velocity between the wafer and the polish pad.

$$r = zPs \tag{3.20}$$

Because no pad compression occurs, the polish rate in the isolation areas (low features) is zero until step height is reduced to zero (Fig. 3.16a, A-B). The planarization rate defined as the difference in the polish rates of elevated and low features (Eq. 3.21) equals the polish rate of elevated features. When the initial step height between elevated and low areas is reduced to zero (B), the planarization phase ends and afterwards the polish rates everywhere on the wafer are constant – there is uniform blanket film removal (B-C). Planarization rate is zero and the polish rate in active areas is reduced due to the increased pattern density (100%) and consequently decreased pressure. When polishing reaches the nitride layer (C), polish rates are reduced to zero and polishing stops (C-D). Overpolishing in case of longer polish times is not possible as the process is self arresting once end-point is reached and again, due to pad stiffness, no further reduction of the trench oxide occurs. The infinitively high slurry

selectivity guaranties no thinning of the stop layer. After planarization a perfectly planar surface is realized. The only topography left at the end of the STI module is the vertical step in the active-isolation transition after nitride removal equal to its thickness.

$$planarization\ rate = elevated\ feature\ polish\ rate - low\ feature\ polish\ rate \tag{3.21}$$

In reality, the planarization process deviates from the ideal case significantly (Fig. 3.17b). The polish pad properties have without a doubt the most pronounced effect on the quality of planarization. The pad has certain compressibility because of which polish rates are not averaged over the wafer, but over smaller areas. Removal rates vary within the die and across the wafer. Areas with higher pattern densities are polished more slowly that areas with lower pattern densities [41]. This leads to reaching end-point (nitride exposure) in some areas earlier than in others. In the STI CMP process there is zero tolerance for remaining oxide on top of the nitride, because it would prevent subsequent nitride strip. There is therefore an the overpolishing step required for exposing the nitride in slow polishing areas, during which nitride and oxide thinning [42] occurs in faster polishing areas due to the limited selectivity of real STI slurries.

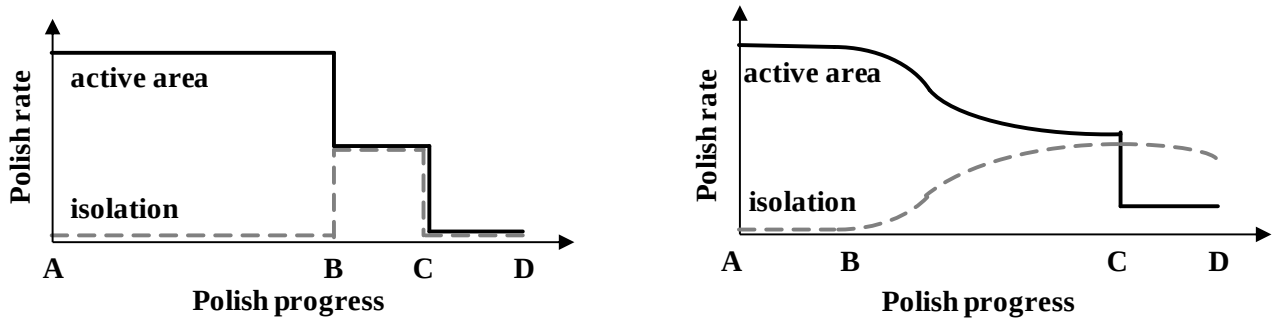


Fig. 3.17: Graphs of (a) ideal and (b) real STI CMP polish rates in active and isolation areas.

Additionally, on a more local level between two structures, again due to pad compressibility, polishing of the isolation area starts earlier than in the ideal case (Fig. 3.17b, B), before the initial step is reduced to zero. This slows down planarization rate after this point (B-C) and generally, planarity is not reached at the moment the nitride layer is exposed (C). During the overpolishing step the polishing rate of the active areas is reduced because of the nitride stop layer, but is nevertheless different from zero, especially near the corners of the active area. Pad compressibility allows polishing to continue in the isolation area with polish rates that are highest farthest away from the active areas. As a result, bowl shaping of the oxide occurs – dishing, as well as erosion of the stop layer [42] that is enhanced along the periphery of the active areas (Fig. 3.18). The larger the distance between the active areas, the more pronounced these effects are and the closer the polish rate in the low area is to the one of the elevated area. Furthermore, pre-CMP wafer nonidealities such as surface nanotopography [43] and nonuniformity in trench depth and layer thicknesses contribute to additional deviations of the polish rates across the wafer.

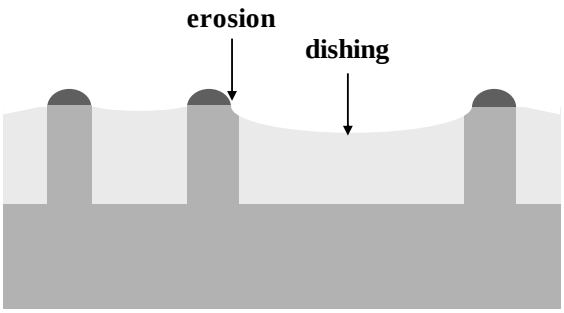


Fig. 3.18: Schematic cross section of a wafer after CMP showing oxide dishing in large isolation areas and nitride erosion in active area corners next to them.

3.4. Scaling and the need for alternative gate dielectrics

Over the last almost five decades, MOSFETs have been continuously scaled from their initial gate lengths of around 10 micrometers down to 45nm at present. The goal of scaling is to produce more complex, faster and more power-efficient ICs. With constantly decreasing transistor dimensions, however, problems arise, which need to be addressed in order to retain the performance advantage of successive technology nodes over the previous ones.

Some of the major scaling problems arise from the increasing electric fields within the scaled MOSFETs. Since the power supply voltage is scaled much more slowly than the gate length and the thickness of the nanometer–thin thermally grown SiO_2 gate dielectric, the lateral field across the channel and the normal electric field across the dielectric inevitably increase. The increased lateral field results in carriers with enough energy to generate electron-hole pairs in the substrate, which have enough energy to surmount the interfacial potential energy and get injected into the dielectric, causing a reliability issue referred to as hot carrier induced device degradation. To solve this issue, lightly doped drain architectures are developed. These are outside of the scope of this work.

The increased normal electric field on the other hand causes carriers to tunnel from the substrate or the gate into the gate dielectric or directly through it resulting in time dependent dielectric breakdown (TDDB) and increased power consumption. In the past several years the aggressive scaling has led to the exponential increase in leakage currents. Scaling laws require the specific oxide capacitance C_{SiO2} (Eq. 3.22) where ε_o is the electric constant, K_{SiO2}=3.9 - the dielectric constant of SiO_2 and d_{SiO2} – its thickness, to increase for each successive technology node.

$$C_{SiO2} = \varepsilon_0 K_{SiO2} / d_{SiO2} \tag{3.22}$$

For decades, this has been achieved by reducing the SiO_2 thickness. From the initial 100nm the dielectric thickness is reduced to below 2nm. At such ultra scaled thicknesses quantum mechanical effects start playing an important role and direct tunneling through the dielectrics became significant. A ‘revolution’ in the microelectronics world is taking place at present, because for the first time in the history of microelectronics, SiO_2 as a gate dielectric is being replaced by a different material, one with higher dielectric permittivity than that of SiO_2 (high-*K* material). This high-*K* material will allow reaching equivalent oxide thickness (EOT) in the sub-nanometer range. EOT is a parameter that relates the thickness of the high-*K* material d_{highK} to the equivalent thickness of SiO_2 yielding the same capacitance as a high-*K* material layer of a certain thickness (Eq.3.23). Materials with relative permittivity higher than that of silicon dioxide offer the same capacitive performance as a thinner oxide and the additional benefit of reduced tunneling leakage currents because of the higher physical thickness.

$$EOT_{highK} \equiv d_{highK} K_{SiO2} / K_{highK} \tag{3.23}$$

A lot of research and large amounts of resources are being invested in order to make the transition. Nevertheless, this transition happens slowly and not without problems. This is due to the fact that silicon dioxide is a unique material in terms of physical and electrical properties. The novel gate dielectric must at least come close to, if not surpass, the properties of silicon dioxide, so that the benefits of its higher permittivity are not outweighed by integration and performance related drawbacks.

Various materials are being investigated. Until now, no material has been found that comes close to silicon dioxide in terms of thermal stability, resistance to crystallization and interface defect density.

3.4.1. Requirements for high-*K* materials

The low permittivity of thermally grown amorphous SiO_2 is practically its only shortcoming. Thanks in part to its intrinsic properties and in part to the long optimization history behind its back, silicon dioxide has been perfectly integrated and optimized in present-day CMOS electronics. SiO_2 is a very stable material with a high melting point, high resistance to crystallization and it forms a stable interface with silicon with very low defect density. Additionally, it is fabricated easily and cost-effectively by simple thermal oxidation of the silicon substrate. In order for a high-K candidate to successfully replace SiO_2, the benefits of its higher permittivity must not be outweighed by its inferiority to SiO_2 in the above mentioned aspects. Below is a short summary of the most important criteria, which a high-*K* material must meet in order to be integrated into a CMOS process without significant changes in the fabrication process compared to SiO_2.

1. Appropriate *K*-Value
EOT must continue scaling down in the sub-nanometer region for future technology generations. This means that dielectric constants above those of SiO_2 (3.9) and nitride oxides (3.9-6) are desired. On the other hand, the *K* value must also not be too high [44]. There is an inverse relationship between dielectric permittivity and bandgap, which means that a material with a very high *K* value is not a good insulator. Furthermore, high *K* values enhance the fringing fields from the gate to the source and drain thus increasing the severity of short channel effects [44]. Therefore, *K* values in the range 20-50 are desirable.

2. Stable interface with silicon
The high-*K* material must not react with the silicon substrate and form silicides or oxides, because this will change the chemical composition of the interface and inevitably also lead to changes in its electrical properties. Unstable interfaces can be counterfeited by using a thin SiO_2 interface layer between the silicon substrate and the high-*K* dielectric. However, in this case the total permittivity of the dielectric stack (K_{eff}) is reduced according to Eq. 3.24 which is a simple calculation of the equivalent capacitance for two capacitors in series.

$$\frac{d_{IF} + d_{highK}}{K_{eff}} = \frac{d_{IF}}{K_{IF}} + \frac{d_{highK}}{K_{highK}} \tag{3.24}$$

For a total EOT in the Angstrom range, the interface SiO_2 must be in the range of 1-3 atomic layers. Such thin layers suffer from non-uniformities and can not be used as reliable barriers.

3. Compatibility to CMOS technology
CMOS integration comprises hundreds of processing steps. The high-*K* dielectric must be compatible with every single one of them. Therefore, each new candidate must be tested extensively in order to

obtain information the effect of different cleaning, etch, implantation, anneal, etc. steps. For example, the ability to withstand high temperatures without deterioration in quality is vital to whether the material can be successfully integrated in a standard gate-first fabrication process. The anneal temperatures for source/drain dopant activation are often in excess of 1000°C. The challenge to most known dielectric materials is that they tend to crystallize which leads to enhancement of leakage paths along grain boundaries and interfacial layer formation.

4. Low defect density of the material and the interface

Defect densities are critical to MOSFET performance. High defect densities increase undesired leakages and decrease carrier mobility. The degradation in performance of high-*K* gate dielectric devices is attributed to a significant degree to defects at the silicon-gate dielectric interface [45].

5. Sufficiently large bandgap

Sufficiently high potential barriers for electrons and holes are needed in order for the material to be a good insulator. The bandgap of a material is usually inversely proportional to its *K* value (Fig. 3.19). Therefore, a compromise must be made between the *K* value and the bandgap when choosing the dielectric.

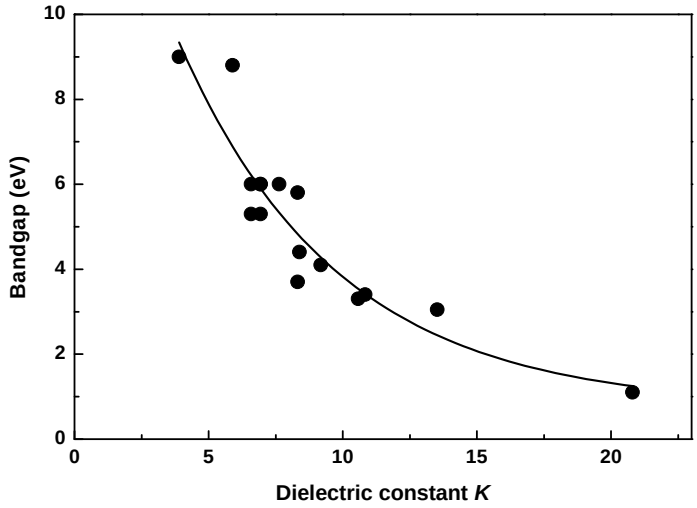
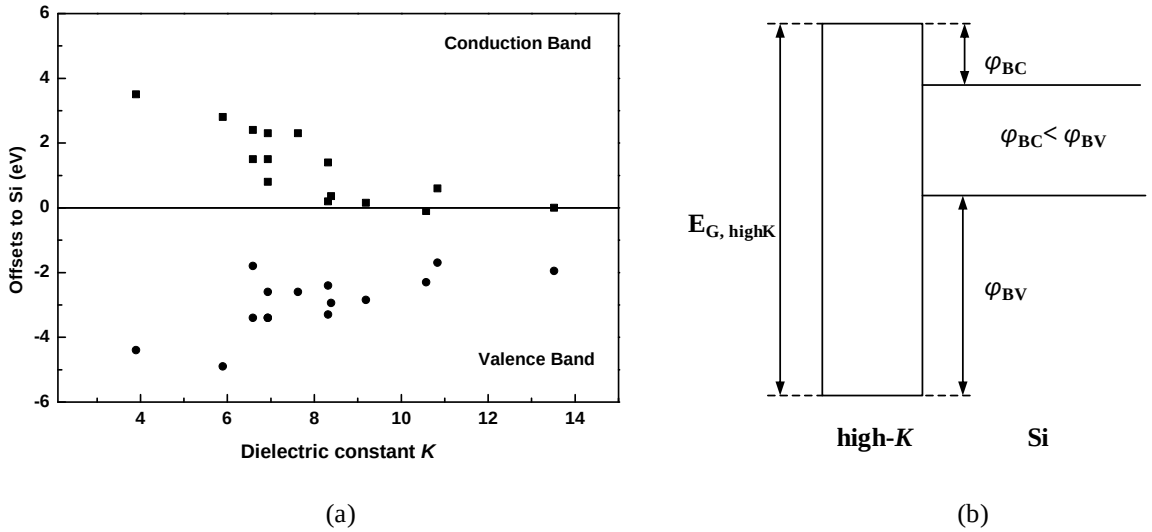


Fig. 3.19: Diagram of dependence between dielectric constant and bandgap for several investigated high-*K* materials [46].

According to previous studies [46] the conduction band offset φ_{BC} at the interface between silicon and a number of materials may be small although they have a large enough bandgap and a sufficient valence band offset φ_{BV} (Fig. 3.20). A small conduction band offset means that the particular material is not a very efficient barrier for the electrons in the substrate. They can tunnel through the dielectric directly into the conduction band of the gate or jump over the barrier to the conduction band of the gate dielectric by thermal excitation. The latter effect is enhanced at the presence of normal fields across the MOS structure due to F-N tunneling. All of these phenomena lead to enhanced substrate-to-gate leakage which in turn results in increased power consumption and decreased reliability.



(a) (b)

Fig. 3.20: (a) Graph of the conduction and valence band offsets to silicon for different high-*K* dielectrics [46] and (b) band diagram.

3.4.2. Amorphous deposition versus epitaxial growth of high-K materials

Following the example of the amorphous SiO_2 most researchers have focused their efforts on searching for other amorphous materials to replace it. The advantages of amorphous materials are clear – simpler deposition techniques and no grain boundaries that enhance leakages. Unfortunately, SiO_2 seems to be an exception to the rule, as all other 'amorphous' materials seem to crystallize to a higher or lesser extent before front-end processing is complete [47]. This is the situation with both the $Hf_xO_yN_z$ family and the Zr_xO_y family which have been the most popular candidates in the near past. As a matter of fact, Hf-based gate dielectrics are now state of the art as Intel introduced them in its 45nm technology node CPUs. Additionally, these first-generation high-*K* materials are not chemically stable against the silicon surface and therefore need a thin interfacial SiO_2 layer. Unfortunately, the thus formed dual-layer structure results in a reduced K_{eff} value according to Eq. 3.18 and the EOT is increased as illustrated in Fig. 3.21.

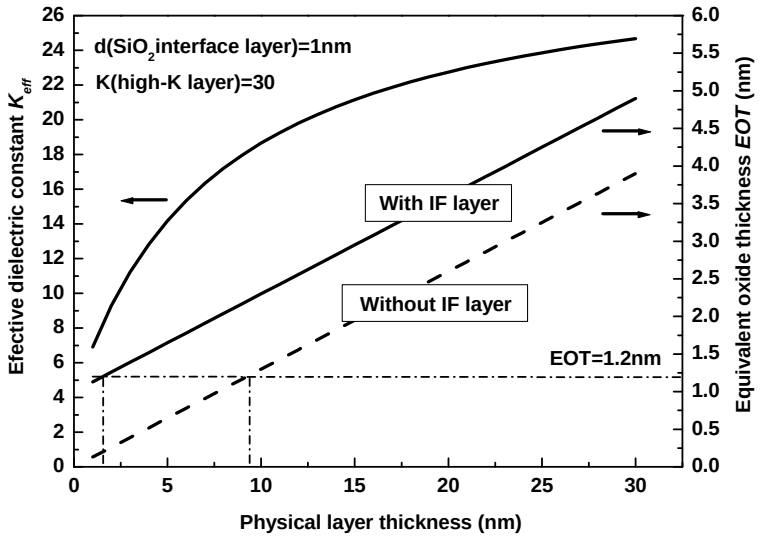


Fig. 3.21: Graph showing the influence of interfacial SiO_2 on the effective dielectric constant K_{eff} and the equivalent oxide thickness EOT in dual-layer high-K dielectric stacks.

Second generation dielectrics are now in the focus of high-*K* research. Epitaxial oxides offer a potential for solving problems associated with first generation materials. Epitaxial growth of crystalline oxides on silicon substrates has been reported by a number of researchers [48; 49]. The idea is to accept crystallization and actually use it as an advantage. Purposeful growth of epitaxial dielectric films on silicon substrates offers several advantages:

1. Epitaxy allows very precise control on growth conditions. The interface properties can be engineered [50]. In addition, the chemical composition of the dielectric with depth can be controlled.

2. Epitaxial films are much more resistant to high temperatures because the crystalline structure is already formed during growth.

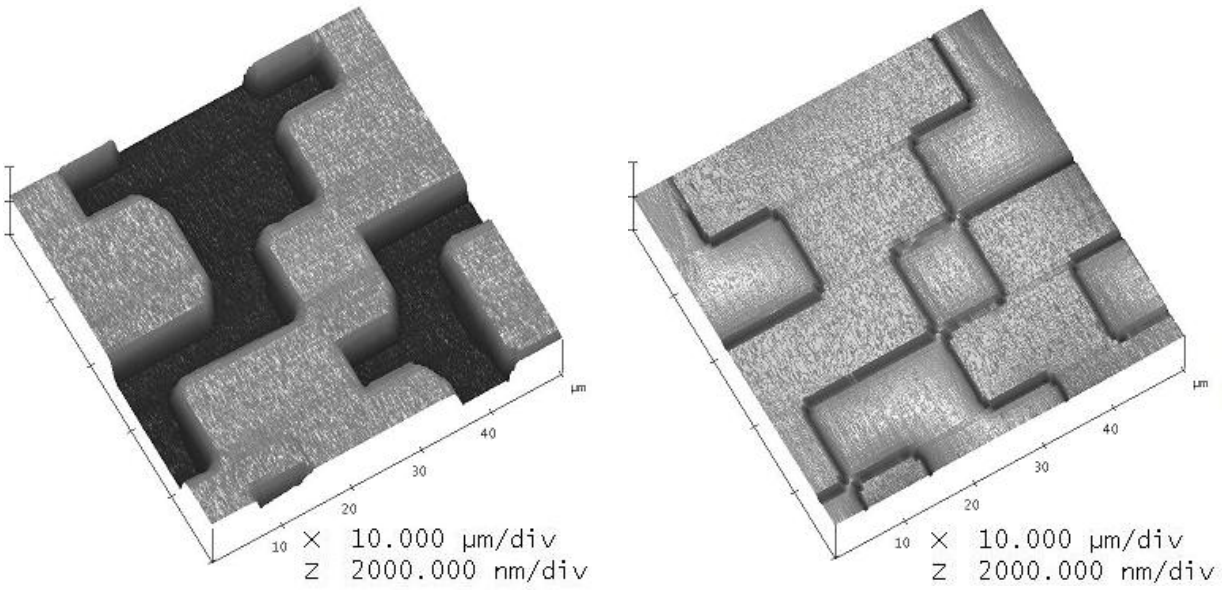
3. Eventually, perfect monocrystalline films can be grown with epitaxy.

4. Studies indicate that some epitaxial rare-earth metal oxides are chemically stable on silicon [51].

4. Development and optimization of shallow trench isolation

Integration of advanced MOS ICs requires state of the art device isolation techniques. Therefore, it is necessary to make the transition from local oxidation of silicon (LOCOS) to the novel and more scalable shallow trench isolation. Before the beginning of this work there has been no previous experience with either STI technology or its enabling technique – chemical mechanical planarization at the ISTN. For this reason, the initial effort is concentrated on the development of stable CMP process and the development and optimization of a reliable STI processing module.

AFM is a very suitable and useful technique for CMP control and evaluation. With its help the process can be examined in detail. Detailed images (Fig. 4.1) or just several scan lines profiles can be used to control the polish progress.



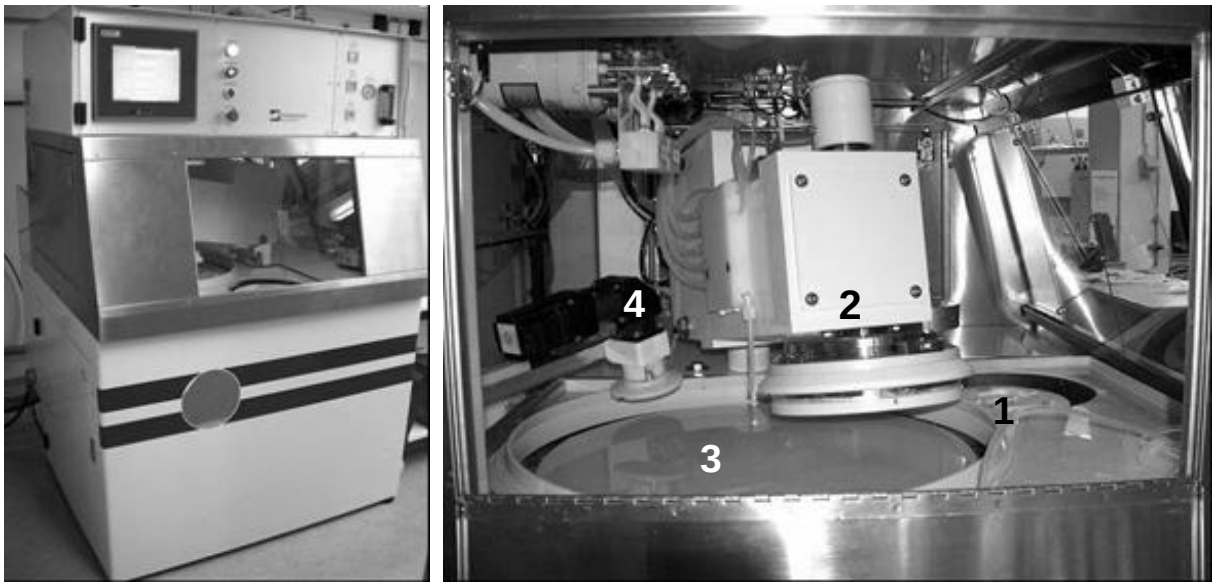
(a) (b)

Fig. 4.1: AFM images of a structured wafer (a) before and (b) after CMP.

4.1. Chemical mechanical planarization preparatory work

Carrying out the STI CMP step is crucial to the success of device isolation formation and the functionality of the MOS devices. Therefore, good control over the polish process must be established.

All CMP work is carried out on a small research-type rotary polisher – Strasbough 6EC (Fig. 4.2). The polisher is an easy to use, automated system adapted to 4" wafers. The wafer to be polished is loaded on the loading station, from where the overarm engages it and brings it over the polish table. During polishing, the overarm presses the rotating carrier against the rotating polish pad, soaked with polishing slurry with a certain predefined force. Combined mechanical and chemical forces lead to selective material removal from the wafer surface predominantly in elevated areas, thus reducing overall wafer topography.



(a) (b)

Fig. 4.2: Photographs of (a) the Strasbough 6EC polisher and (b) its work chamber with loading station (1), overarm and carrier (2), polish table and pad (3) and pad conditioner (4).

CMP introduction includes several steps. Work begins from the very basics. Initially, blanket wafers are polished in order to investigate and optimize polish parameters. Structured wafers are polished next to examine dependence between pattern size and polish rates. Considering the significant variations in pattern density in tests chip designs used for test device fabrication, optimization techniques are utilized in order to improve planarity. Finally, optimization techniques are developed for STI CMP. The results are fully applicable to RGT oxide CMP.

Initial parameter evaluation

The goal of this part of the work is to establish the effects of different process parameters such as back pressure, down-force, retainer ring pressure, slurry flow, etc on the polish results and more specifically on polish rate uniformity [52]. For successful integration of CMP into the manufacturing process, the removal rate of material during polishing has to be constant across the 4" wafers within several percent.

The initial tests are carried out with ESM-U pads by Universal Photonics and Levasil 50CK /30% silica slurry by H. C. Starck. Blanked wafer with uniform thermal SiO_2 films are polished because they allow easy and fast elipsometric (and sometimes even visual) evaluation of the results.

Practically all parameters have an effect on center-to-edge polish uniformity (Fig. 4.3). Sets of process parameter combinations or 'recipes' that give a good balance between polish rates, global uniformity and surface damage are established. Although the initial experiments are carried out exclusively on non-patterned oxide films, the results and dependencies hold true for different materials and topographies and with small corrections are used throughout the work.

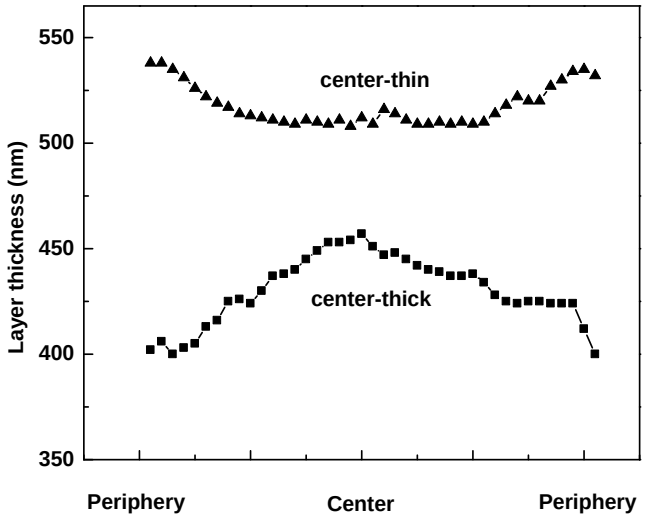


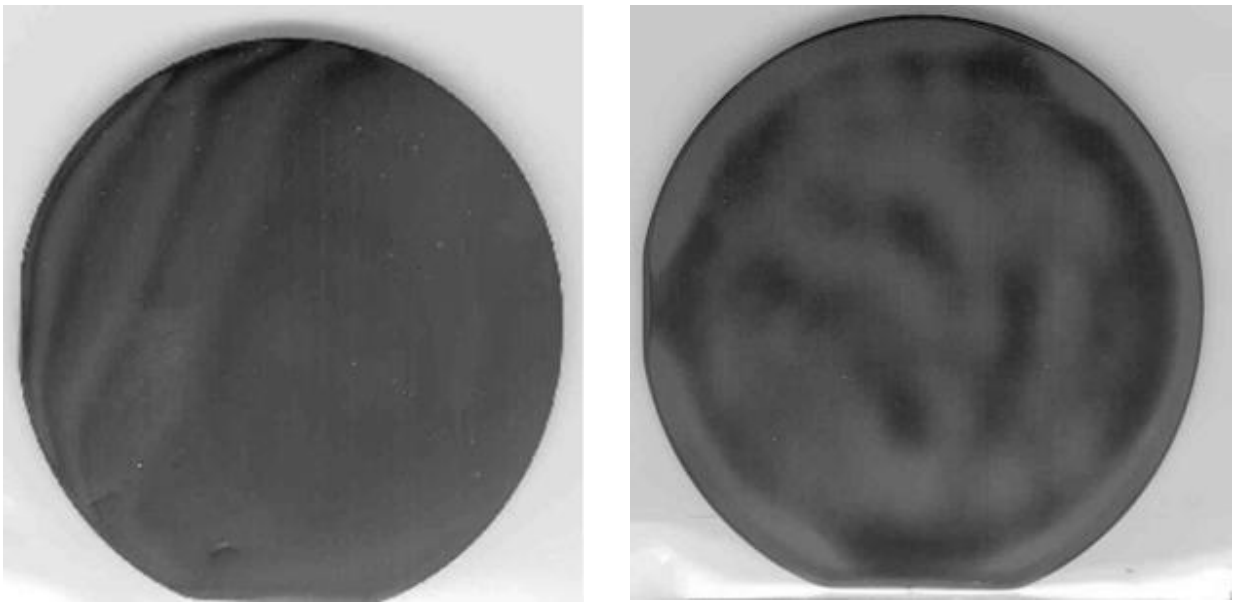
Fig. 4.3: Graph of elipsometric measurement results of post-CMP SiO_2 film thickness on two wafers, polished with different process parameters, showing center-to-edge polish rate non-uniformity.

A commonly used oxide polishing set of parameters is shown in Table 2. The polish rates in this case are 100nm/min for thermal SiO_2 and 120nm/min for as-deposited atmospheric pressure chemical vapor deposition (APCVD) SiO_2. They vary linearly with the applied down-force.

Parameter	Value	Unit
Retainer ring pressure	1	psi
Back pressure	4	psi
Downforce	6	psi
Chuck rotation speed	40	rpm
Table rotation speed	50	rpm
Slurry flow rate	125	mL/min

Table 2: Optimal parameter set for uniform CMP on the Strasbough 6EC machine.

Lack of central symmetry of polishing (Fig. 4.4a) is initially observed. It is determined that it is caused by unsymmetrical distribution of the applied pressure across the wafer by the wafer carrier. Eliminating crystallization of silica particles inside the carrier by utilizing the correct carrier idle and storage procedures solves this problem. Last but not least, wafer nanotopography also affects post-CMP planarity. Deviations in the thickness of the polished layer of 10-100nm (Fig. 4.4b) can occur due to non-planarity of the underlying silicon surface [43].



(a) (b)

Fig. 4.4: Photographs of wafers after CMP showing different problems: (a) non-symmetric polishing (upper-left region of the wafer) and (b) nanotopography (spots of different colors are areas with different thickness of the SiO_2 layer due to nanotopography of the underlying silicon substrate).

Structured wafers CMP

The effects of feature size and height on the planarization process [53] are examined next. Wafer designs with different structure patterns defined in SiO_2 are polished (Fig. 4.5). AFM is used prior to polishing as well as after each polish step to evaluate the polish progress at well defined position on the wafer.

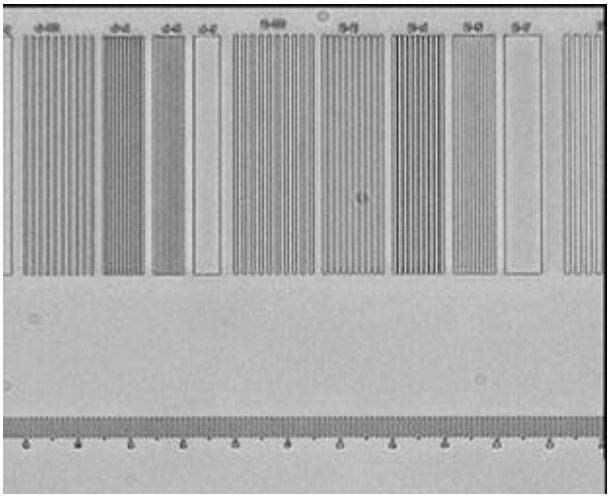


Fig. 4.5: A micrograph of a wafer surface with different etched patterns in SiO_2 used for CMP analysis.

In order to accelerate work when measuring line patterns, only a small number of scan lines are recorded rather than complete images. They are sufficient for obtaining the necessary profiles (Fig. 4.6).

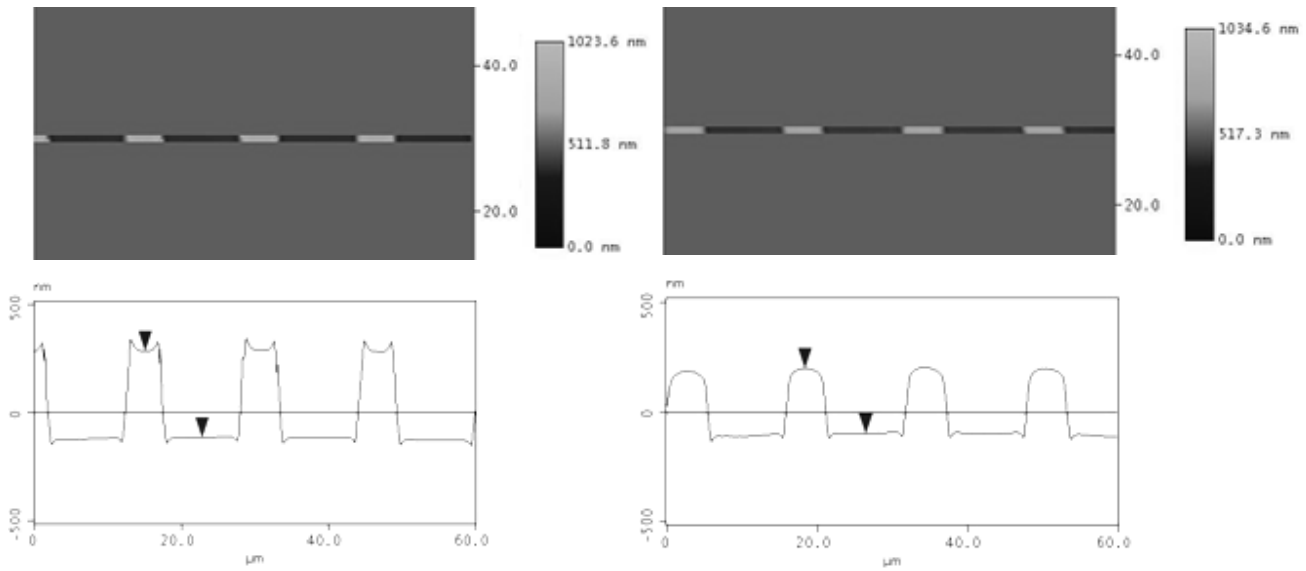


Fig. 4.6: Reduced AFM images and corresponding cross sections of a wafer surface (a) before and (b) after a 10sec. polish step.

Polish results indicate strong dependence structure dimensions and spacing on local polish rates (Fig. 4.7). It is concluded that patterns of varying density within a wafer pose a serious challenge to direct one-step planarization. Critical areas are defined and the need for design level optimization is established.

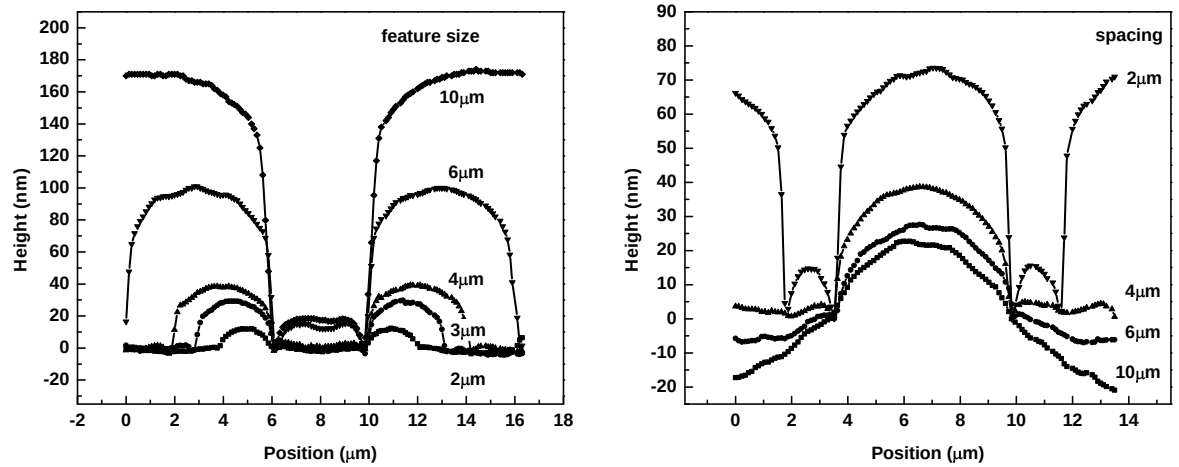
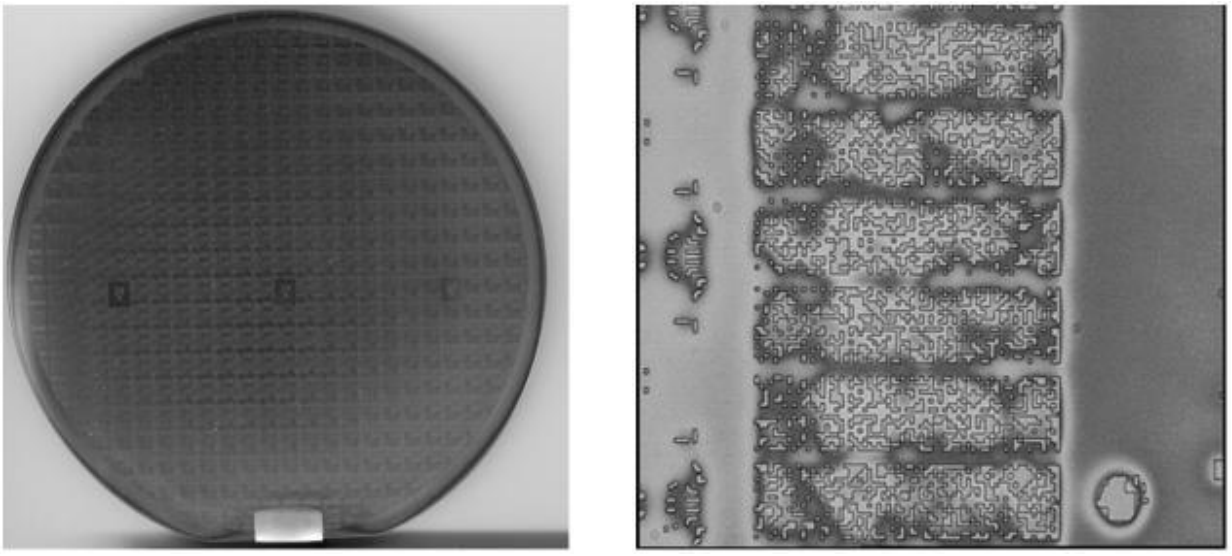


Fig. 4.7: AFM-measured profiles of post-CMP wafer surface features illustrating pattern density dependence of polish rates for different patterns: (a) different feature size, identical spacing and (b) different spacing, identical feature size.

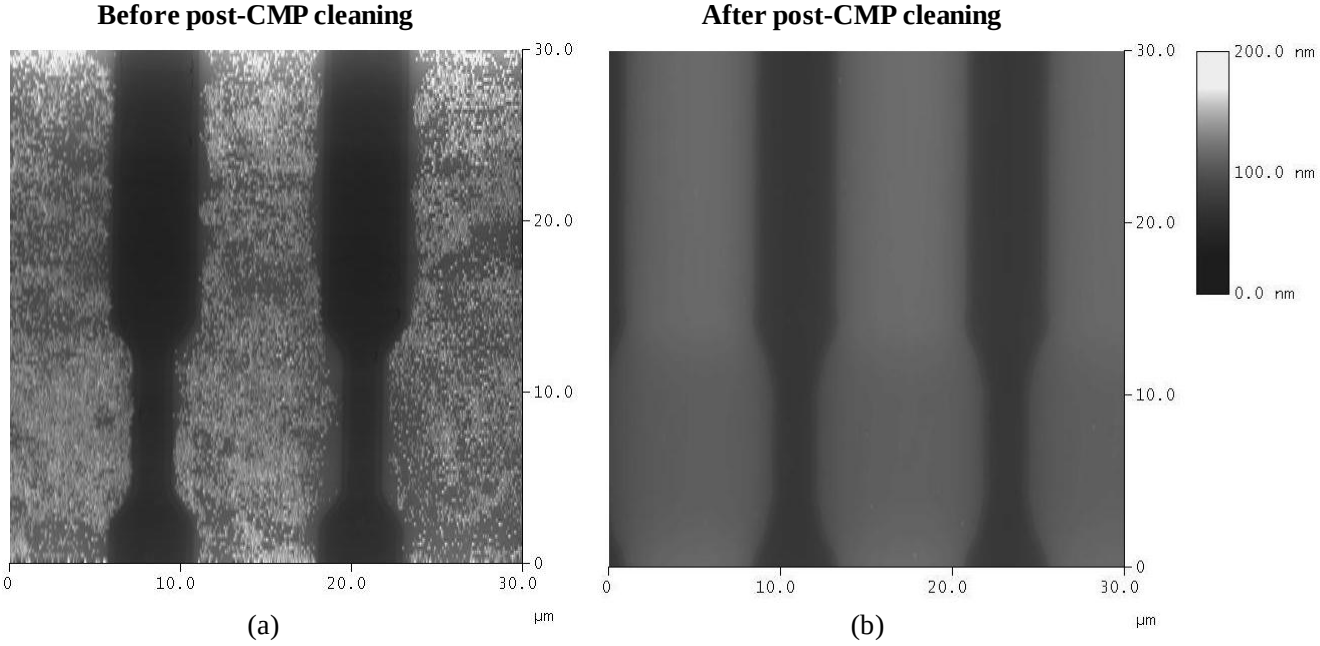
Other serious issues encountered, which are directly connected with the pattern density dependence of polish rates, are the oxide dishing in large low areas and the erosion of the edges of high areas next to large low areas. Wafers with SiO_2-covered trenches in silicon are polished [53]. The edges of the Si_3N_4 layer protecting the elevated silicon areas in some cases are fully eroded. The problem is not solved even when slurry with very high selectivity of several hundred to one is used (HS-STI-D11 from H. C. Starck).



(a) (b)

Fig. 4.8: (a) Photograph of a structured wafer after CMP. Center-to-edge polish uniformity is good which is evident by the small deviations in color among chips. (b) Micrograph of a part of the wafer with noticeable oxide dishing. The differences in the color of SiO_2 close to and away from the silicon high-areas can be easily distinguished.

After a CMP there must always be post-CMP cleaning. CMP is an inherently dirty process. With the help of AFM, large amount of Silica particles from the CMP slurry are observed on the wafer surface after STI CMP. They are concentrated on the Si_3N_4 surface. A combined brush and ultrasonic cleaning on a Corwet 100 cleaner with water, ammonium hydroxide (NH_4OH) and hydrochloric acid (HCl) is implemented in order to guarantee clean surfaces in the critical active areas. AFM indicates that the contaminations are completely removed from the surface without mechanical damage of the surface itself (Fig. 4.9).



(a) (b)

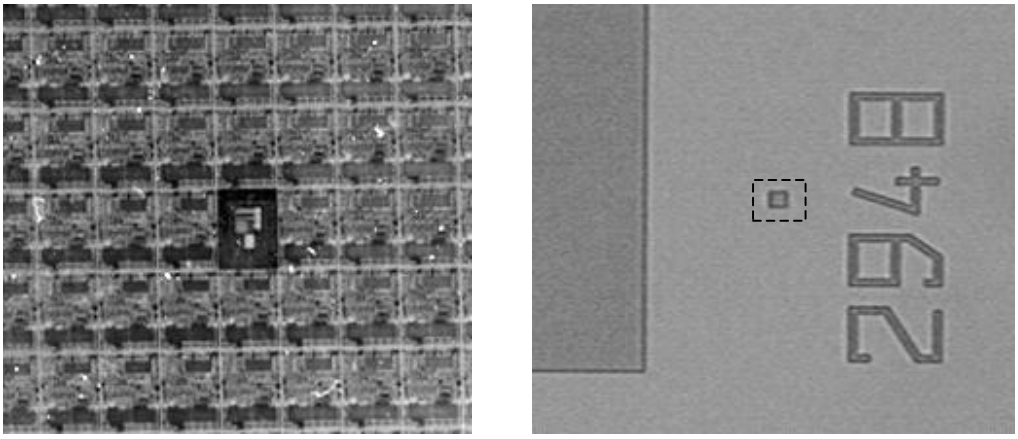
Fig. 4.9: Topography AFM images of a wafer surface (a) before and (b) after post-CMP cleaning. The brighter areas are Si_3N_4 and the darker ones – SiO_2.

4.2. Application of C-AFM for Si_3N_4 erosion detection

It is a well known fact that Si_3N_4 erosion occur often for chip designs with significantly varying pattern densities. Damaging the Si substrate has very negative effects on device performance of the finished MOS device. Therefore, reliable metrology must be available for erosion detection.

Post-CMP topography is commonly evaluated by profilometry and topographic atomic force microscopy [54]. Both techniques are suitable for dishing measurements with the latter having superior resolution. However, neither of these techniques provides a simple, non-destructive and efficient means of detecting fully eroded areas on the device level. Here, the application of C-AFM for such a purpose is described and its advantages over other measurement methods are discussed.

The samples used for performing the erosion measurements are fabricated on 4" Si wafers. The processing starts with chemical vapor deposition of a 120nm silicon nitride layer on the silicon substrate. For photolithography, a non-optimized for CMP mask design with large deviations in structure size and pattern density is used in order to accentuate post-CMP problems (Fig. 4.10a). Reactive ion etch of the nitride with CHF_3-O_2 follows. $SiCl_4$ RIE of the silicon defines the active area and isolation trench areas with depth of 350nm. Next, a plasma enhanced chemical vapor deposition (PECVD) oxide of thickness 600nm is deposited. CMP is performed for removing the excessive oxide over the nitride stop-layer in the active areas followed by post-CMP cleaning.

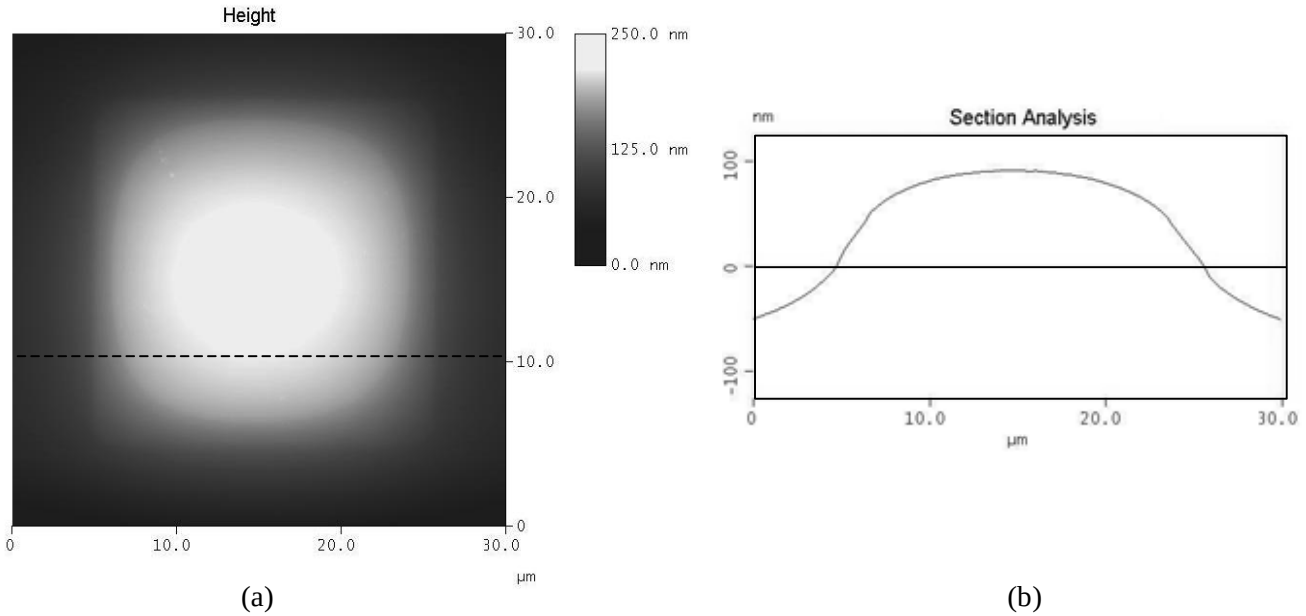


(a) (b)

Fig. 4.10: Micrographs of (a) a test wafer surface after shallow trench isolation chemical mechanical planarization with areas of different pattern density and (b) an area with very low pattern density where AFM measurements are performed (dashed square area, 20*20μm).

Topography AFM images are obtained in different areas of the polished wafers. As expected, erosion is most pronounced in a region containing a single isolated active area with dimensions of 20*20μm, surrounded by a large isolation area of at least 60μm on all sides (Fig. 4.10b) and this region is used for illustrating the measurement technique.

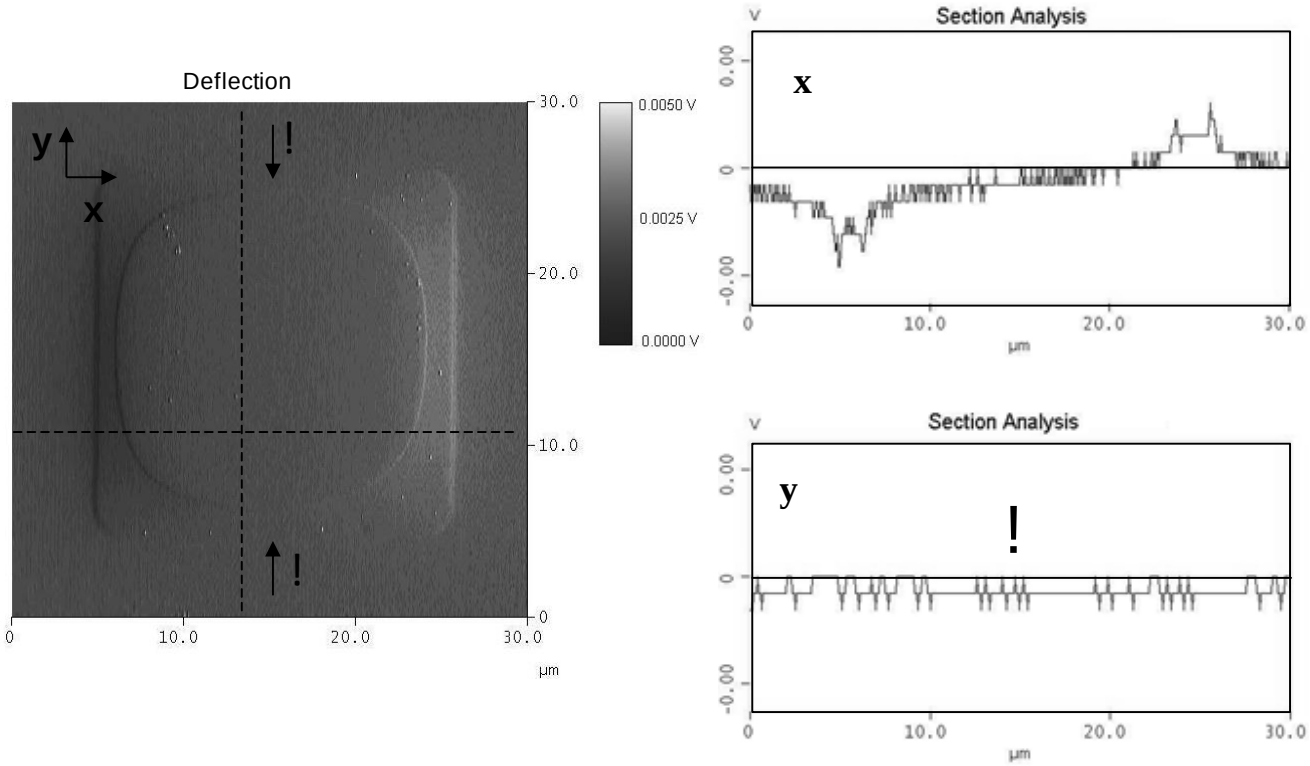
Fig. 4.11a shows a topography image of the selected region. Dishing in the oxide trench can be seen clearly in the cross section. The curve in the central nitride area indicates higher erosion at the edges of the nitride stop-layer, caused by the different pressure distribution in the central and the edge areas [55]. However, there is no clear indication whether there is full erosion, i.e. whether the silicon surface has been exposed to and damaged by CMP. There is a gradual transition from the elevated nitride area to the lower oxide isolation area (Fig. 4.11b). Therefore, height measurements alone are not an efficient solution for distinguishing partial from full erosion.



(a) (b)

Fig. 4.11: (a) Topography AFM image and (b) cross section.

On Fig. 4.12a a different set of data collected during the same scan is shown. The deflection of the scanning probe during the measurement is indicative of the height derivative in the direction of the scan (along the x-axis). One can see two pairs of lines on the left and on the right side of the active area, showing places where the slope of the surface relief is slightly higher than elsewhere. Since these high slope areas are narrow, there are no large height steps that can be clearly observed in the height image.



(a) (b)

Fig. 4.12: (a) Probe deflection (error signal) AFM image obtained simultaneously with the topography image in Fig. 4.12 and (b) cross sections in the x- and y- directions.

The distance between the lines is a few micrometers. The outer one coincides with the edge of the active area on the left and on the right and shows the transition from the trench oxide to the nitride or eventually silicon. The inner one most likely indicates a transition from nitride to silicon which means that there is an open silicon area of a few micrometers on the left and on the right side of the square. However, the deflection image can detect only transitions that are perpendicular to the scan direction (Fig. 4.12b). This is why the deflection image is also not applicable for erosion investigation, since misleading conclusions would be drawn from it.

Simultaneously with the topographic information, electric currents are measured during the scan at sample voltage of 2V. Exposed silicon areas are expected to be detectable by their electrical conductivity. Fig. 4.13a shows the electric current image. The brighter area in the picture is the region where the tunneling currents are different from zero. Together with the topography image the electric current image shows clearly that there are fully eroded areas on all four sides of the active area including top and bottom areas. As suggested by the deflection image, the open silicon areas are within a few micrometers from the edges of the active area. However, the current measurement shows precisely the entire eroded area because it is not dependent on scan direction (Fig. 4.13b) unlike the deflection data but only on the electrical properties of the scanned surface (Fig. 4.14). The current image indicates complete nitride erosion around the edges of the active area and none or partial erosion in the central area. With the help of the current image, the 'base' silicon surface level can be determined, revealing also the nitride thickness throughout the active area in the height image.

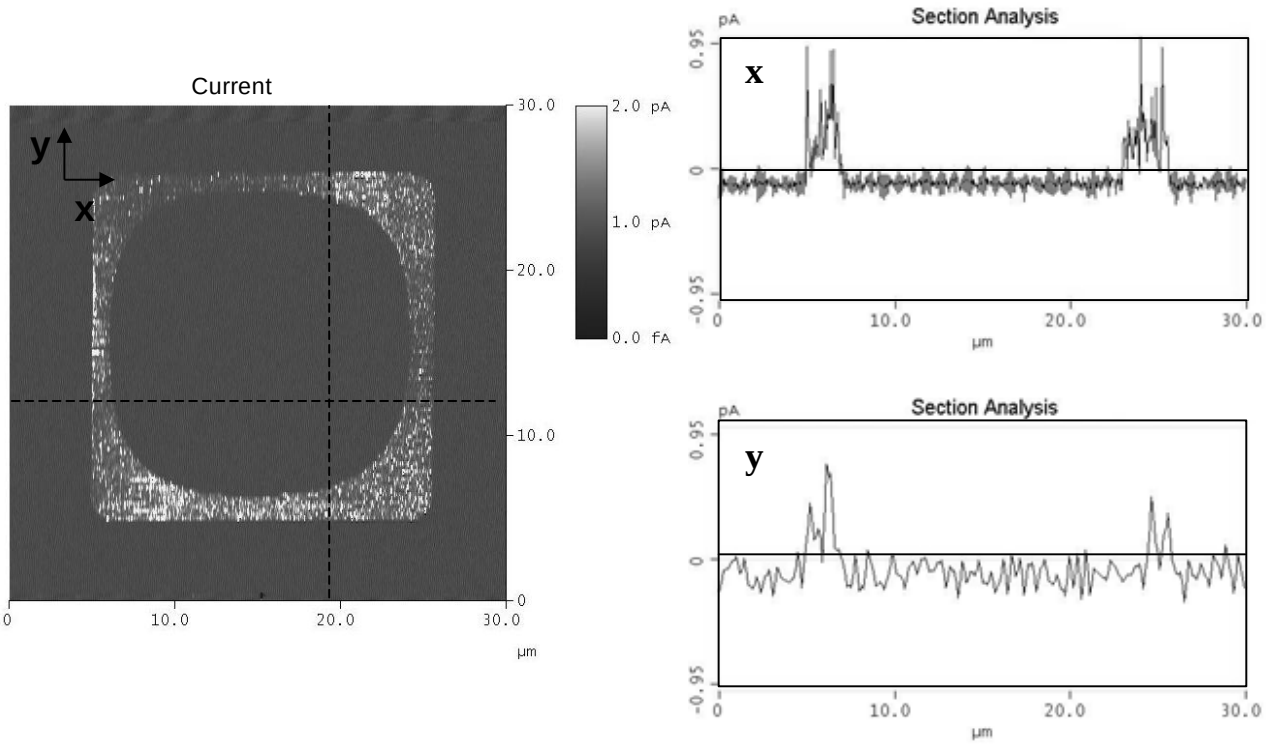


Fig. 4.13: (a) Electric current AFM image obtained simultaneously with the topography image in Fig. 4.11 showing the eroded periphery of the active area and (b) cross sections in the x- and y-direction.

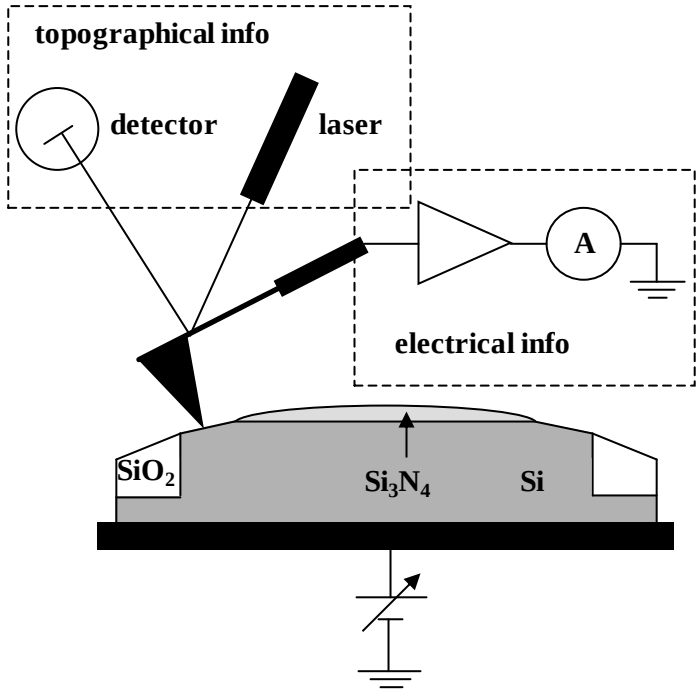


Fig. 4.14: Schematic of C-AFM for STI erosion evaluation.

The relative symmetry of the currents in the current image is in accordance to expectations since the trench around the structure is sufficiently large in all directions to ensure symmetrical pressure distribution. In Fig. 4.13a it can be seen that the currents in the eroded area are not uniform. This can be attributed to CMP related surface roughness and the structure of native oxide that grows on the silicon surface in ambient atmosphere.

One can conclude that topographic data by itself is clearly insufficient for erosion evaluation and its interpretation can be very misleading. The gradual transition from the high nitride layer to the low dished trench oxide gives no clear indication of what is between them. The only way to positively identify problematic areas is by applying electrical measurements that utilize the different electrical properties of the used materials.

Finally, a short comment is made to explain what advantages C-AFM offers compared to several other measurement techniques. Scanning tunneling microscopy is also capable of measuring electrically conductive areas. However, this technique is not applicable to surfaces including both conductive and non-conductive areas since it utilizes current amplitude data for keeping the scanning probe at a certain distance above the sample surface. On the other hand, AFM is independent from the electrical characteristics of the sample for performing scans. Elipsometry allows precise measurement of thin film thicknesses but unfortunately it lacks the lateral resolution to do a nanometer scale measurement. SEM is a different technique that relies on electron imaging. Together with a short sample etch prior to measurement for enhancing contrast, the electron microscope gives good evaluation of the condition of the nitride layer. SEM however is a destructive technique and is therefore inapplicable for process control.

To summarize, combined topographic and electric current measurements with an AFM make it possible to detect excessive nitride erosion in STI CMP. The measurements are non-destructive and easy to implement.

4.3. Development of optimization techniques for shallow trench isolation

Taking advantage of the accurate nanoscale measurements of surface topography and the reliable erosion detection technique, with the help of C-AFM it becomes possible to optimize STI planarity. The goal is to reduce dishing to a minimum and to completely eliminate erosion by modifying pre-CMP topography with additional process steps [56].

Considering ways towards achieving optimum results, two functional optimization methods are developed – reverse nitride masking (RNM) and oxide etchback (OEB). These methods are chosen instead of the commonly used dummy gate insertion method which results in capacitative coupling of the dummies to the functional structures of the ICs and increased noise [57] especially pronounced in mixed-signal ICs [58]. The magnitude of these negative effects is going to increase in the future as devices shrink in size and come closer to each other. On the other hand, RNM and OEB are well scalable with technology node and applicable to arbitrary chip design. With these optimization techniques, additional isolating material is structured in order to optimize polishing. In the case of RNM, this material is a second thin nitride layer and in the case of OEB, it is the deposited isolation oxide itself. Both techniques avoid performance degradation due to capacitive coupling and noise at the expense of increased process complexity by introducing additional lithography and etching steps.

Reverse nitride masking

The goal of RNM is to effectively control overpolishing in low pattern density regions by providing an additional stop-layer in low areas as illustrated in Fig. 4.15. The structuring of the nitride layer is performed using an additional mask [59]. The mask is characterized by a specific spacing between the structures on it and on the original STI mask. This spacing determines how close to the active areas the surface is protected from overpolishing. Smaller spacing is expected to give better results but also requires more precise alignment.

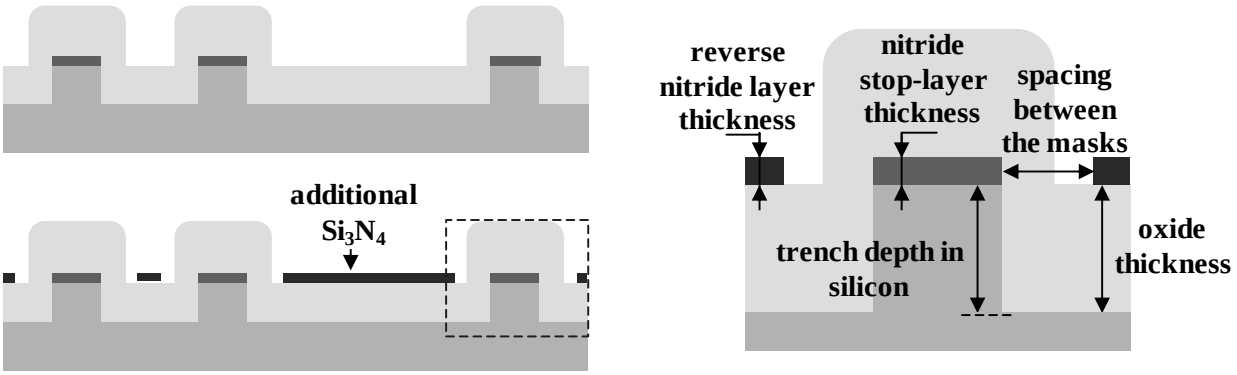


Fig. 4.15: (a) Schematic cross sections of a wafer before and after reverse nitride masking optimization and (b) zoom-in of the dashed rectangle with important dimensions.

Oxide etchback

Unlike RNM, the goal of oxide etchback is to equalize pattern density by preliminary oxide etching over high areas. In this case the result is maximizing pattern density throughout the wafer (Fig. 4.16). The same mask produced for RNM is also used for OEB [59]. The different spacing in this case determines to what extend pattern density variations are minimized after etchback. Again, smaller spacing gives better results, but requires better alignment.

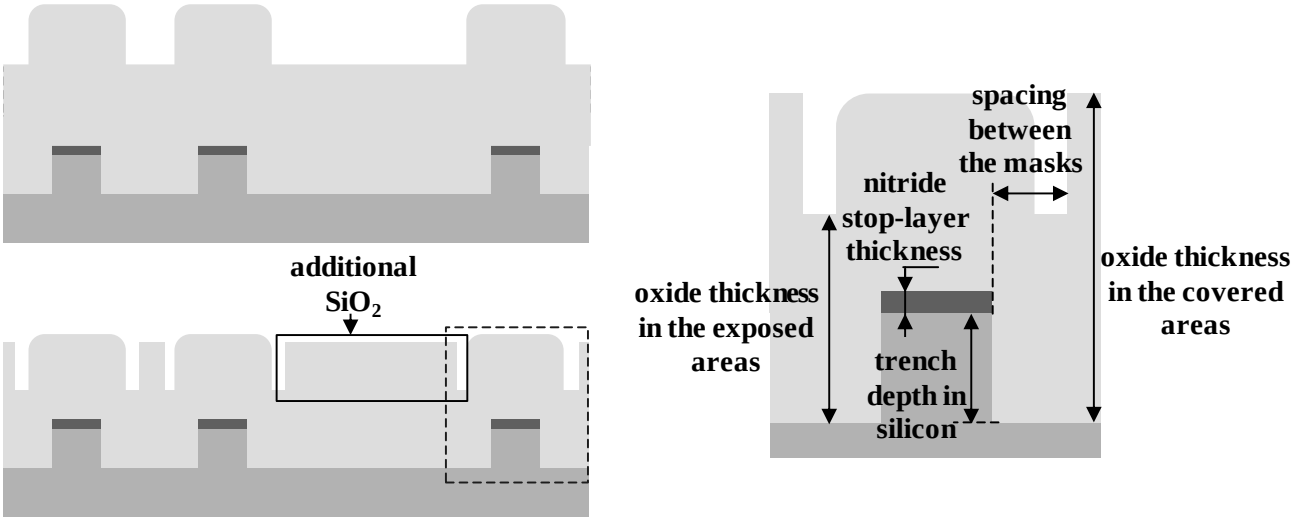


Fig. 4.16: (a) Schematic cross sections of a wafer before and after oxide etchback optimization and (b) zoom-in of the dashed rectangle with important dimensions.

Test wafers are fabricated for evaluation of the two optimization methods [56]. Initially, the new reverse masks needed for the second photolithography step are designed and produced. Masks with three different spacing distances are manufactured: 1, 2 and 3 μm. 4" pSi(100) wafers are used. A 100nm silicon nitride layer is deposited on all wafers by atmospheric pressure chemical vapor deposition (APCVD). STI photolithography (Fig. 4.17), CHF_3-O_2 RIE of the nitride and $SiCl_4$ RIE of the silicon bulk define trenches with a depth of approximately 300nm below the silicon surface i.e. 400nm below the top nitride surface. The batch is then split in two for testing the two optimization techniques.

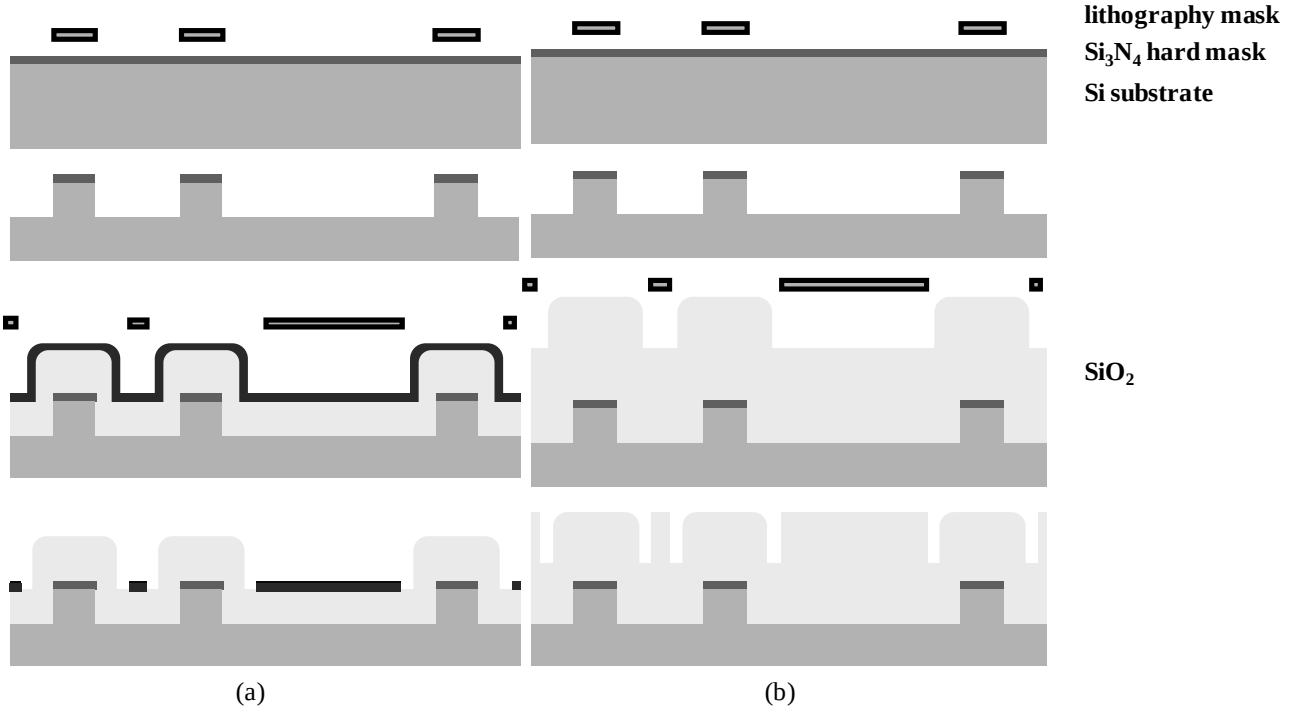


Fig. 4.17: Schematic cross sections of wafers showing fabrication sequence for test structures with (a) reverse nitride masking and (b) oxide etchback.

The first sub-batch is optimized with RNM. A 300nm thick plasma-enhanced chemical vapor deposition (PECVD) oxide is deposited (equal to the silicon trench depth), followed by a second APCVD nitride deposition (100nm). A second photolithography is performed next using the reverse masks. Finally, the second nitride layer is structured by another RIE step.

The second sub-batch receives OEB optimization. A PECVD oxide of around 1μm is deposited (in this case the oxide must be thicker than twice the total trench depth including the nitride thickness). Photolithography with the reverse masks follows and an oxide RIE of 400nm (equal to the total trench depth) completes the process, ensuring equal height in isolation and active areas.

After manufacturing is complete, CMP is carried out. Planarization is performed on a Strasbough 6EC polish machine equipped with Universal Photonics ESM-U polish pads using a high-selectivity STI slurry from H.C. Starck (HS-STI-D11). CMP is carried out in several steps per wafer with atomic force microscopy control in-between the steps in order to observe the progress of the planarization process. For both optimization techniques CMP continues until the first nitride stop-layer is reached everywhere on the wafers as shown by the AFM measurements.

Following CMP, wafers undergo post-CMP cleaning using brush and ultrasonic cleaning with water, ammonium hydroxide (NH_4OH) and hydrochloric acid (HCl).

To evaluate the planarization results conductive atomic force microscopy is used. As demonstrated in the previous chapter, C-AFM provides an accurate way of detecting nitride erosion with nanometer resolution, even in case of very smooth transitions between various materials where conventional topographic AFM fails to detect erosion. This is because of the additional electrical information which is supplied simultaneously and independently from the topographic information.

Reverse nitride masking results

Global planarity measured after nitride removal is below 100nm across the 4" wafers. Dishing is eliminated completely in large isolation areas (Fig. 4.18) because of the nitride layer protection and is only present within the spacing between the structures on the two masks. Concerning nitride erosion, it is absent for spacing of 1μm. With 2 and 3μm spacing, there is erosion in certain areas due to misalignment of the masks (Fig. 4.19). This is because misalignment causes different actual spacing in different positions on the wafer and within one structure itself (Fig. 4.20a). This misalignment is beneficial for the tests because it allows determining the precise critical distance at which erosion first appears. As shown in Fig. 4.20b, for the parameters chosen in the test process, erosion occurs at spacing distances above 2.7μm. It should be noted that erosion also depends on the vertical difference between the two nitride levels (Fig. 4.21) - the higher the reverse nitride level is, the smaller the erosion. This protection comes at the expense of additional topography.

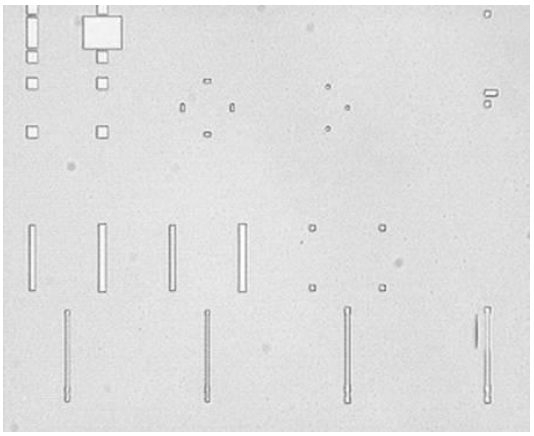
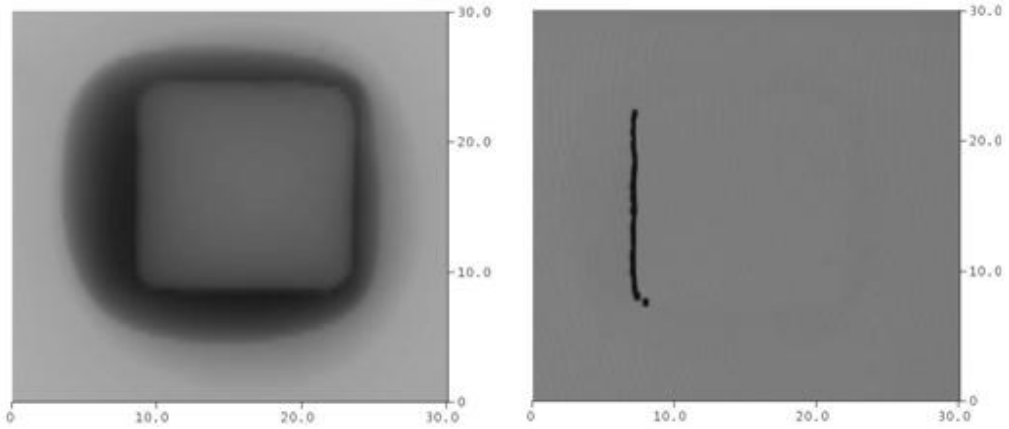
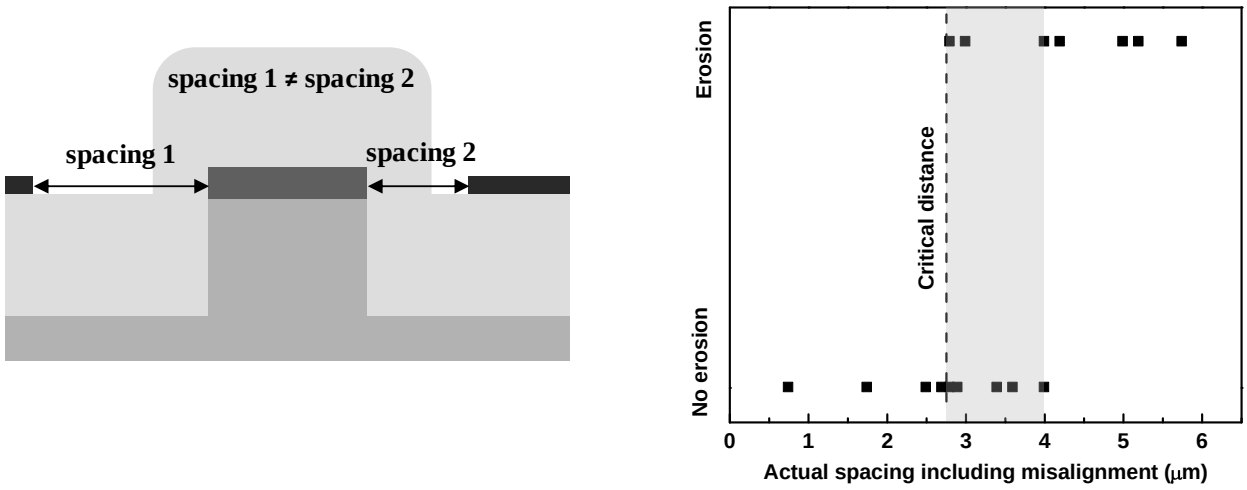


Fig. 4.18: Micrograph of test chip with reverse nitride masking optimization after CMP. The color uniformity in the isolation area is proof of dishing elimination with RNM (compare to 4.24).



(a) (b)

Fig. 4.19: (a) Topography and (b) corresponding electric current AFM images of an area on a wafer with reverse nitride masking optimization after CMP, indicating nitride erosion at the left edge of the active area due to mask misalignment.



(a) (b)

Fig. 4.20: (a) Schematic cross section of a wafer with reverse nitride masking with mask misalignment resulting in different length of the unprotected regions around active areas for reverse-nitride-masking optimized wafers and (b) graph showing dependence of misalignment on erosion as determined by C-AFM.

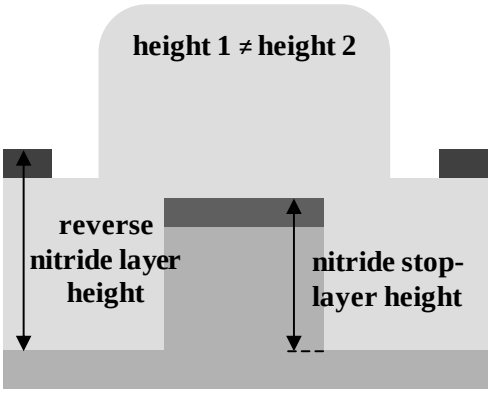
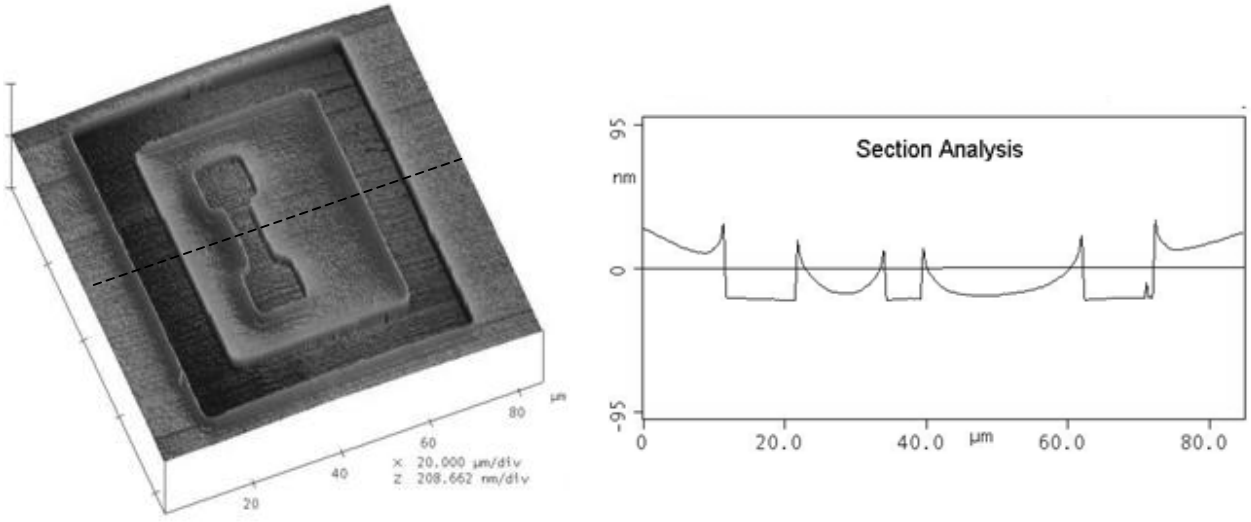


Fig. 4.21: Schematic cross section of a wafer with reverse nitride masking with vertical misalignment of the two nitride layers.

Oxide etchback results

Planarity after nitride removal is below 100nm across the wafers. Dishing is strongly reduced (Fig. 4.22, Fig. 4.23). Actually, for very large isolation areas where there is negative dishing, i.e. the area is slightly elevated above active area level (Fig. 4.24). As for erosion, OEB optimized wafers are found to be less sensitive to spacing. Erosion is fully eliminated for all tested spacing distances, even with the additional misalignment. This positive result can be explained with the fact that, as AFM measurements in-between polishing steps indicate, planarity is reached before reaching end-point of polishing. This leads to a much shorter overpolishing time needed and therefore there is no excessive thinning of the stop-layer.



(a) (b)

Fig. 4.22: (a) Topography AFM image of a transistor area on a wafer with oxide etchback optimization after CMP and nitride removal and (b) cross section along the dashed line.

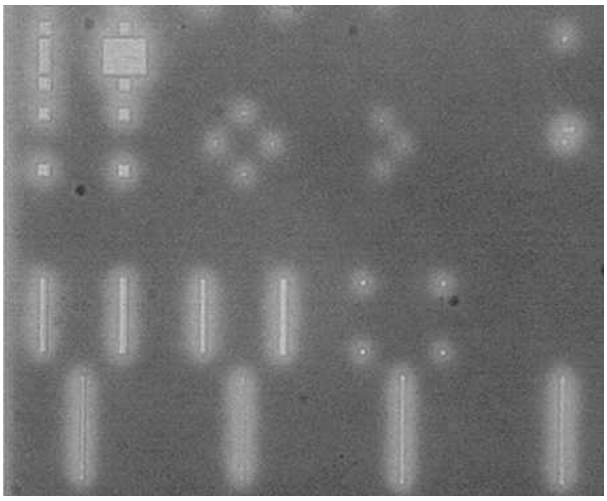


Fig. 4.23: Micrograph of a test chip area with oxide etchback optimization after CMP.

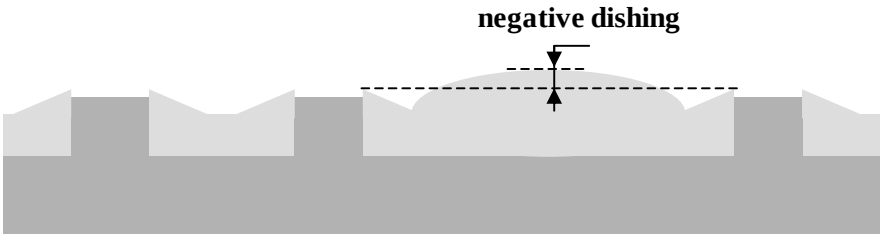


Fig. 4.24: Schematic cross section of a wafer with oxide etchback optimization after CMP and nitride removal illustrating negative dishing in large isolation areas.

Global planarity obtained with both optimization techniques is below 100nm. There is no doubt that by using more advanced equipment even better results can be obtained. The limiting factors for planarity in the case of RNM are stop layer thickness and spacing. Using spacing below 1μm in combination with a thinner nitride layer will presumably bring planarity below 50nm. In case of OEB, planarity can be improved further by reducing nitride thickness, lowering spacing and possibly using tiled support oxide structuring which will yield further reduction in pattern density distribution. For both optimization techniques, another useful addition would be the application of flexible spacing in order to change spacing within a wafer depending on pattern density. Using lower spacing in less dense regions can compensate for the higher polish rates there. Planarity in the presented tests is also limited due to non-uniformities in the deposited oxide and nitride layers, depth variations in the etched trenches and other technical difficulties coming from using university research level equipment.

4.4. Conclusions

As a summary, the functionality of both techniques for CMP planarity optimization is demonstrated. Oxide etchback and reverse nitride masking techniques are applicable in various CMP processes: STI, replacement gate technology, etc. Their major advantage to the commonly used dummy tile insertion is the absence of additional parasitic capacitances in the circuit and their major disadvantage is the increased process complexity. Global planarity of around 100nm is achieved using either technique, university level equipment and fixed support structure geometries. It is clear that state of the art equipment and adaptive support feature geometries would yield better results. Nevertheless, the level of planarity is more than sufficient for the completing the tasks at hand.

5. Structural and electrical investigation of high-*K* dielectrics

The basic underlying motivation for all the AFM work in this thesis is to find the right high-*K* dielectric material with the help of C-AFM and integrate it using a compatible processing scheme in order to produce functional MOSFET devices that demonstrate its applicability in future CMOS technology.

A lot of research and large amounts of resources are being invested in order to make the transition from silicon dioxide to high-*K* materials. This is due to the fact that silicon dioxide is a unique material in terms of physical and electrical properties. The novel gate dielectric must at least come close to, if not surpass, the properties of silicon dioxide, so that the benefits of its higher permittivity are not outweighed by integration and performance related problems.

Traditionally, evaluation of SiO_2 gate dielectrics is performed by means of electrical measurements on capacitor or MOSFET test structures using these dielectrics. Capacitance, conductance, breakdown, output characteristics, transfer characteristics, charge pumping, etc. measurements allow determining among other things the dielectric strength, thickness, quality and interface properties of the dielectrics. For studying high-*K* films, however, additional measurement techniques are needed. Degradation phenomena and defect generation mechanisms in high-*K* dielectrics could differ from those in SiO_2. For example, leakage paths along grain boundaries of the micro- or nano- crystalline structure of the materials are a novel phenomenon due to the excellent resistance of amorphous silicon dioxide to crystallization. Different leakage mechanisms in one and the same film are often superimposed and it is complicated, if not impossible, to separate each one of them from the rest by a macroscopic measurement where leakages are integrated over comparatively large areas ranging from several hundred to several tens of thousands square microns. Conductive atomic force microscopy on the other hand allows nanometer resolution electrical measurements in which separate leaky spots can be isolated and their electrical behavior can be studied separately. In case of C-AFM, the probe itself acts as a gate electrode of the measurement structure and therefore it is easy to measure any area of interest and also to create different MOS structures by simply using probes made of different materials.

5.1. SiO_2 reference – verification test

Initially, state-of-the-art 5nm thick SiO_2 layers are investigated with C-AFM as a reference dielectric system. The goal of this study is to obtain a first impression of the capabilities of the measurement technique.

In Fig. 5.1 a topography AFM image of the surface of a partially processed wafer is shown. The lower, darker region is the thin thermally grown SiO_2 gate dielectric layer and the brighter higher region is the LOCOS isolation area. It is easy to see the transition area mentioned in Chapter 4.1 that is a limiting factor to the applicability of LOCOS in ULSI manufacturing as discussed in Chapter 3.3.

Also in Fig. 5.1, the electric current AFM image obtained simultaneously with the topography image at sample-to-tip voltage of 4.2V is shown. The higher and brighter area represents tunneling currents between the tip and the substrate through the dielectric surface. The measured currents are very uniform which is a testament to the very high quality of the gate dielectric considering the fact that direct tunneling currents are exponentially proportional to the dielectric thickness. As the tip moves from the gate dielectric to the LOCOS area, the currents are reduced to zero – the darker lower region of the image.

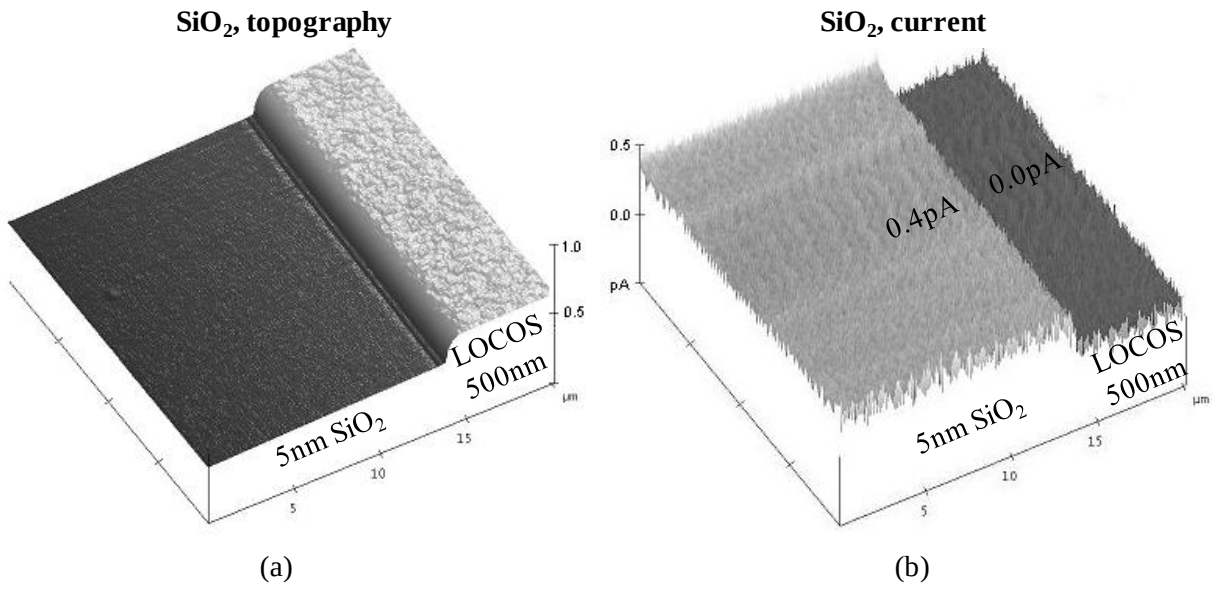


Fig. 5.1: (a) Topography and (b) electric current AFM images of the surface of a partiall processed wafer with SiO_2 gate dielectric and LOCOS.

After scanning over the sample surface at a constant voltage, the tip is locked at several random positions on the gate dielectric surface and a spectroscopic current-voltage measurement is performed there. As expected from the current images, the *I-V* measurements at different positions on the gate dielectric yield similar results. Fig. 5.2 shows a typical *I-V* characteristic. Several observations can be made from this graph. First of all, in the range of approx. -4.5V to 3.8V the tunneling currents can not be observed because they are lower than the noise level of 0.1pA. Second, the breakdown can not be observed either because of the upper limit of the amplifier of 10pA. It is clear that breakdown occurs for fields close to or above 10MV/cm.

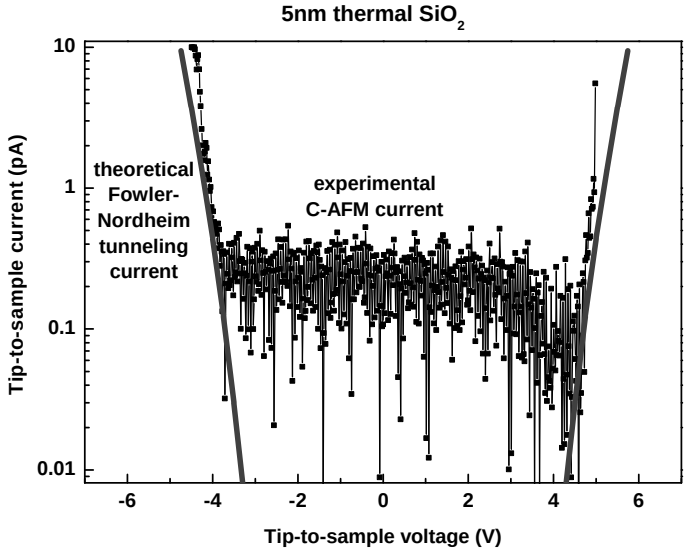


Fig. 5.2: Graph of experimental C-AFM current-voltage measurement on a 5nm thick thermal SiO_2 film and theoretical Fowler-Nordheim tunneling current.

The current densities are in agreement with calculations for F-N tunneling through a circular contact area with radius of 10nm between the tip and the dielectric. However, the measured currents increase

with the applied voltage more rapidly than the theoretical ones. This is due to the complex tip shape which results in additional tunneling from regions of the tip neighboring the contact area at higher electrical fields. The currents are not symmetric with respect to the voltage polarity. They are higher for electron injection from the tip than from the pSi substrate. Information about the doping of the used diamond-like carbon (DLC) coated AFM probe is not available, but the measurements suggest that the material has a lower work-function than that of the substrate.

The investigation of SiO_2 also helps discover some of the disadvantages of C-AFM. Because of the high current densities through the small contact area of the tip, the temperatures can increase significantly. This increase has two significant effects. On one hand, the conductive coating of the tip can be evaporated. This results in lower measurable currents because of the lower conductivity of the silicon base of the tip pyramid and in image distortions because of the increased tip radius. Coating evaporation is more pronounces for tips with a metal coating such as the PtIr coated tips used in the course of this work. Despite of their higher cost, diamond-like-carbon (DLC) coated tips are much more suitable for C-AFM measurements due to the higher melting point of their coating and the supperior mechanical wear resistance. For more information check Chapter 8.

The conclusion that can be drawn from this investigation is that C-AFM can yield information about the local electrical properties of the gate dielectric. As long as some precations are taken in order to preserve the tip and sample in their original state (see Appendix), the currents are consistent and can be used for evaluation the electrical integrity of the gate dielectric.

5.2. Application of C-AFM for investigation of dual-layer TiO_2-SiO_2 high-*K* dielectrics

This work is the first attempt for high-*K* gate dielectric investigation by C-AFM [60] within the framework of this thesis. It is the result of a collaboration between the ISTN, TU Darmstadt, the Department of Electrical Engineering, IIT Madras, India and the Thin Films Laboratory, Centre for Advanced Technology, India.

TiO_2 is interesting as an alternative dielectric because of its high permittivity and the acceptance of titanium in most modern CMOS fabrication facilities [61]. One of the difficulties in using simple dielectrics is that when the permittivity (K) increases, the bandgap (E_g) decreases [62] and thermionic emission is a serious concern, particularly if the band alignment is not favorable. TiO_2, however, is a lattice polarizable material, which has a high dielectric constant as well as an adequate bandgap. The reported values of the dielectric constant for TiO_2 are in the wide range of 4 to 86 [62; 63] depending on deposition conditions and film thickness, while the bandgap is 3-3.5eV [64]. Also, previous studies indicate good thermal stability of the rutile phase TiO_2 on Si [65]. Because of these attractive properties, TiO_2 has been selected as a suitable high-*K* dielectric for this work.

TiO_2 is deposited by Pulsed Laser Deposition (PLD) over thermally grown SiO_2. This stacked gate dielectric provides good interfacial quality with the underlying silicon substrate. PLD is a relatively simple deposition technique, which is compatible to VLSI technology [66]. It is quite easy to produce multi-layered films of different materials by sequential ablation of assorted targets. Besides, by controlling the number of laser pulses, a fine control of film thickness down to atomic monolayer can be achieved. The most important feature of PLD is that the stoichiometry of the target can be retained in the deposited films. The dielectric constant for PLD TiO_2 is around 26.5.

(100) p-Si silicon wafers with resistivity 4-11Ωcm are used as substrates. After RCA cleaning [67] of these substrates, they are oxidized at 1000°C to grow a thick field oxide (about 500nm). Subsequently, windows are opened in the oxide layer by photolithography. Gate oxidation is carried out at 800°C in diluted mixture of oxygen and nitrogen followed by annealing for 2min. in nitrogen. This thin SiO_2 layer is grown first in order to preserve the interface quality of the SiO_2-Si system. Pulsed laser

deposition of TiO_2 is carried out at a substrate temperature of 750°C in oxygen ambient at a pressure of 10^{-3}Torr for 2min. Prior to filling the deposition chamber with oxygen, it is evacuated to a base pressure of 1×10^{-6}Torr using a turbo-molecular pump. The third harmonic of a Q-switched Nd:YAG laser (λ=355nm) with a repetition rate of 10Hz and a pulse width of 6ns is used for the ablation of TiO_2 palletized target. The laser fluorescence on the target is maintained at 1.5J/cm^2 and target to substrate distance is 5cm. It is to be noted here that most of the deposition parameters have been optimized previously [68].

After the deposition, the samples are annealed at two different temperatures: 750°C and 950°C. Finally, aluminum is deposited on the samples by vacuum evaporation and patterned by photolithography to realize MOS capacitors.

C-AFM measurements are carried out on the TiO_2 surface near the Al-gate of the same samples to correlate the macroscopic electrical behavior with the microscopic surface structure and the local electrical characteristics. Initially, topography images in tapping mode are obtained in order to check surface roughness. The samples annealed at 750°C have a less smooth topography compared to those annealed at 950°C, as seen in Fig. 5.3. The root mean square (RMS) roughness of the samples is 0.96nm and 0.49nm respectively.

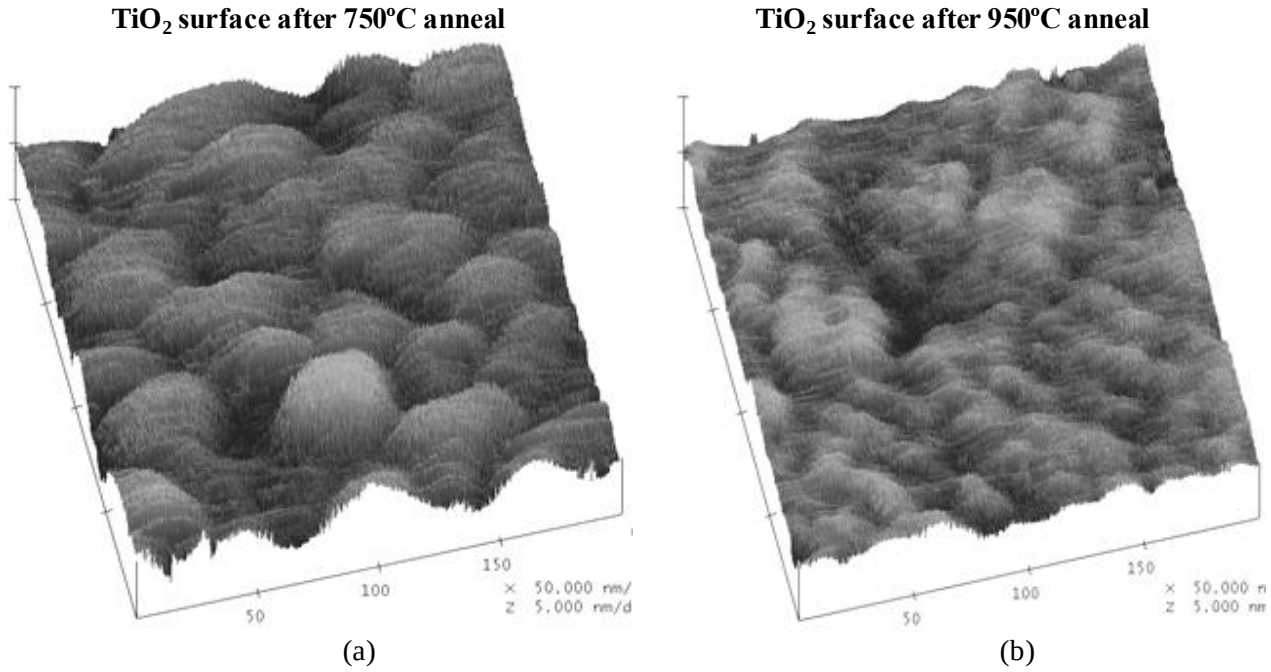


Fig. 5.3: Topography AFM images of the TiO_2 surface after annealing at (a) 750˚C and (b) 950˚C.

Macroscopic gate leakage characteristics are performed on two samples annealed at two different annealing temperatures. At V_G=2V, the capacitors annealed at 750°C exhibit gate leakages of more than five orders of magnitude lower than those of the capacitors annealed at 950°C ($2*10^{-6}$A/cm^2 and $5*10^{-1}$A/cm^2 respectively) (Fig. 5.4). The C-AFM measurements for samples annealed at 750°C indicate low currents in agreement with the macroscopic electrical measurements. However, contrary to the macroscopic current-voltage measurements, the C-AFM tunneling currents are found to be even smaller for the samples annealed at 950°C, as shown in Fig. 5.5.

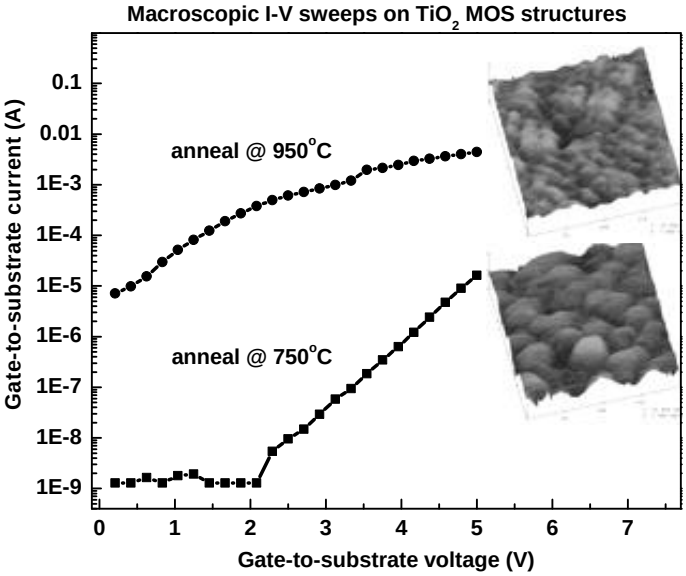


Fig. 5.4: Macroscopic gate current-voltage characteristics of TiO_2-SiO_2 gate dielectric capacitors (area $5.5*10^{-4}cm^{-2}$) for samples annealed at 750˚C and 950˚C.

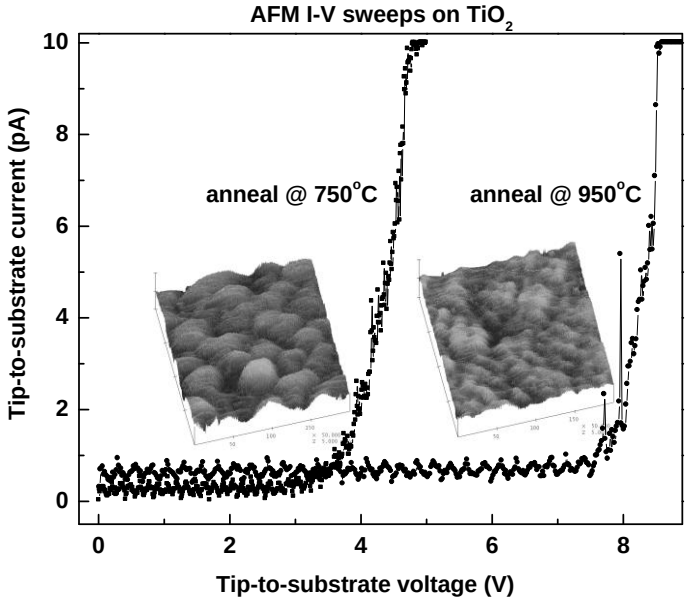


Fig. 5.5: C-AFM current-voltage characteristics on TiO_2 surface in point contact mode of samples annealed at 750˚C and 950˚C.

In order to understand this apparent contradiction, a detailed study is carried out. The leakage current measurements by C-AFM on a large number of sites indicate that for the samples annealed at 950°C, the current distribution is not uniform over the surface. Locally enhanced currents are found only through a number of small leaky spots. In the remaining area, the currents maintain the normal low value (Fig. 5.6).

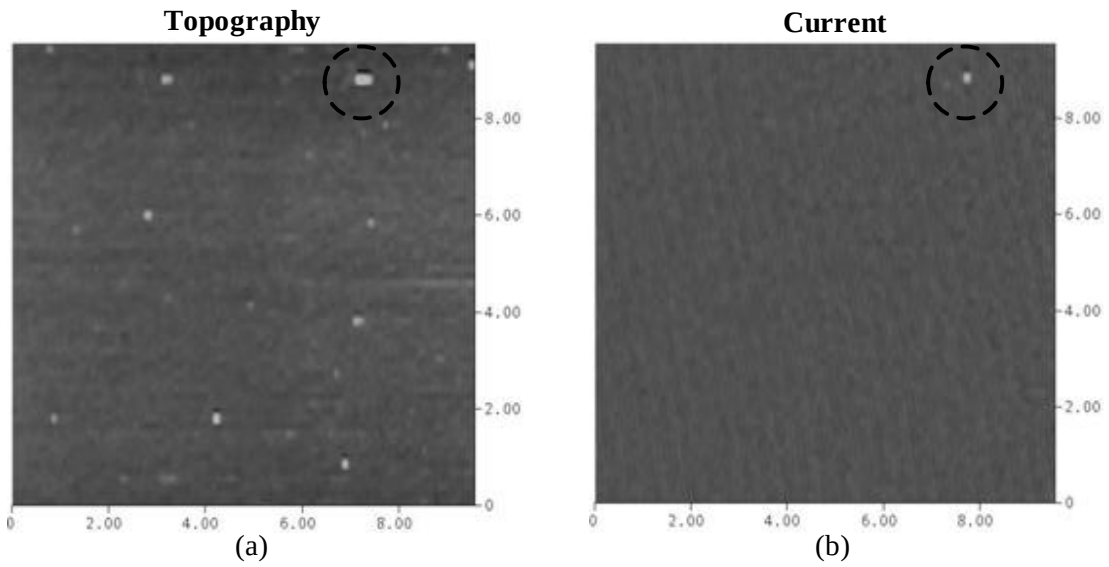


Fig. 5.6: (a) Topography and (b) electric current AFM images of the TiO_2 suface annealed at 950˚C. The bright area in the current image indicates one of the rare instances of leaky spots.

It is inferred that the macroscopic high leakage seen in the MOS devices annealed at 950°C is primarily due to these structural defects rather than a degraded intrinsic property of the dielectric film as a whole. The size of these leaky spots is up to a few hundred nanometers in diameter and the spot density is approximately $10^4 cm^{-2}$. The current-voltage dependence of these spots follows two distinct patterns. Some of these leaky spots have Schottky-like current-voltage characteristics (Fig. 5.7a). It is suspected that during the annealing step at 950°C, titanium may diffuse all the way to silicon at some spots, showing Schottky-like behavior (Fig. 5.7b). Other spots show nearly symmetric behavior (Fig. 5.8a). In some case, the Ti may diffuse sideways, resulting in a local thinning (Fig. 5.8b). The current observed in these cases is due to F-N tunneling through the SiO_2 layer and has a symmetric nature.

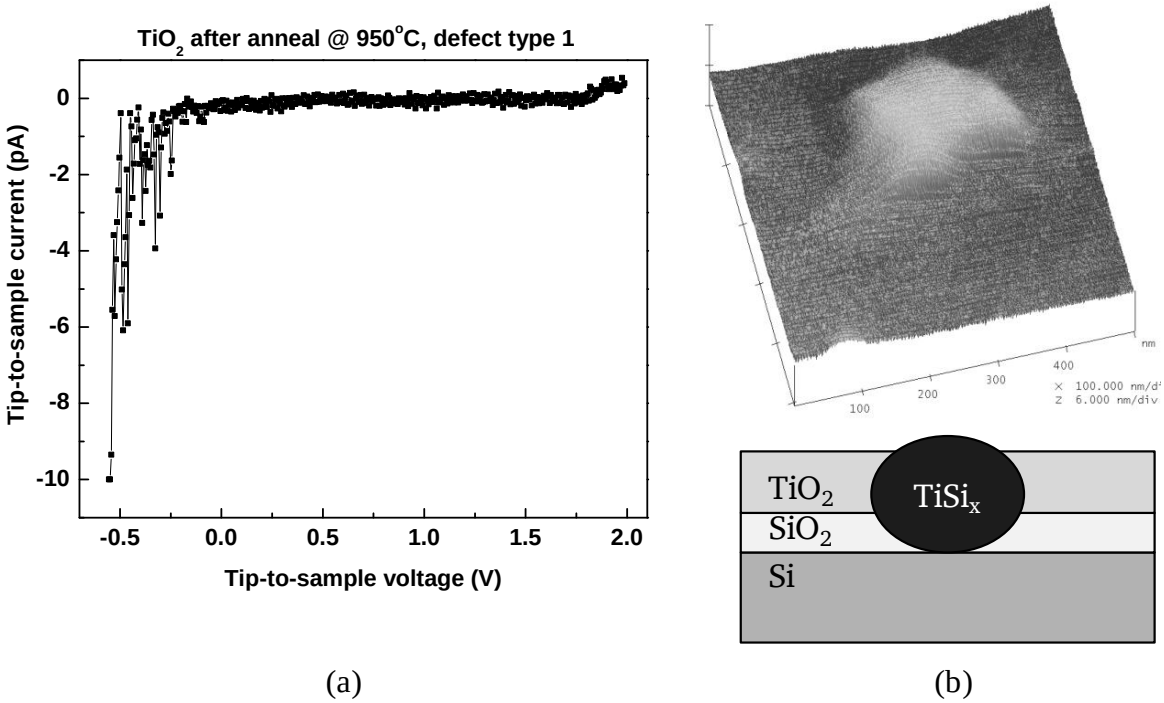
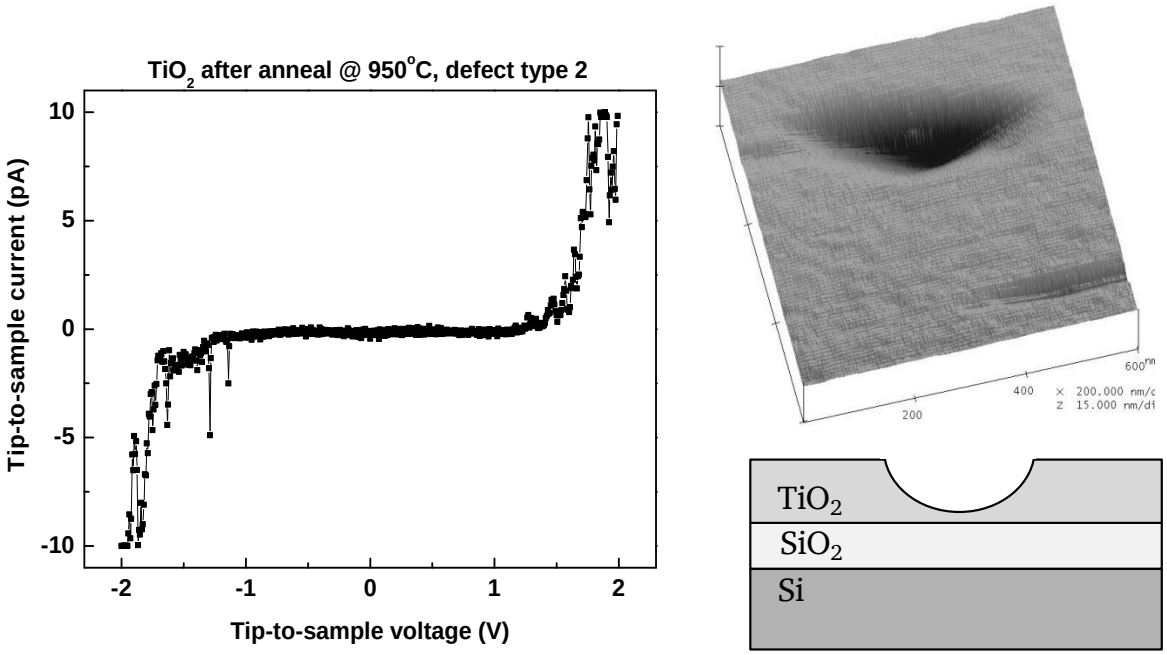


Fig. 5.7: (a) AFM current-voltage characteristic at a defect site on the TiO_2 surface after 950˚C anneal showing a Schottky-diode behavior and (b) topography AFM image and schematic of local silicide formation explaining the observed topography and electric behavior.



(a) (b)

Fig. 5.8: (a) AFM current-voltage characteristic at a defect site on the TiO_2 surface after 950˚C anneal showing a symmetric leakage and (b) topography AFM image and schematic of local TiO_2 thinning due to side diffusion explaining the observed topography and electrical behavior (lower right).

Summary

Electrical characterization at the nanometer scale reveals that the enhanced gate leakage of TiO_2-SiO_2 capacitors annealed at 950°C is caused by a superposition of microscopic current paths originating from extrinsic defects rather than a degraded intrinsic property of the dielectric film as a whole.

Conductive atomic force microscopy proves to be a useful addition to the wide spectrum of available macroscopic electrical measurement techniques. C-AFM provides for a better understanding of the microscopic causes of the observed macroscopic electrical behavior in high-*K* dielectric films.

5.3. Epitaxial rare-earth metal oxides

Because of the considerations in Chapter 3.4.3 and as a result of the unsatisfactory quality of the dual TiO_2-SiO_2 layers available, the investigation efforts are turned to epitaxial oxides. The epitaxial layers are grown at a partner institution – the Institute of Electronic Materials and Devices (IEMD), University of Hannover.

Rare-earth metals are the elements from the lanthanide and the actinide series (periods 6 and 7) in group IIIB of the periodic table.

Investigation of rare-earth metal oxides as potential candidates for CMOS gate dielectrics is initiated by Prof. Jörg Osten and Dr. Hans-Joachim Müssig at the Innovations for High Performance Microelectronics (IHP) in Frankfurt, Oder [51] and later continued by Prof. Osten and his group at the Institute of Electronic Materials and Devices (IEMD), University of Hannover [69].

All rare-earth epitaxial oxides in this work are grown by Molecular Beam Epitaxy (MBE). The MBE system at the IEMD is a DCA Instruments multi-chamber ultrahigh vacuum MBE system with in-vacuo XPS.

Pr_2O_3

Chronologically, the first investigated material is Pr_2O_3. Pr_2O_3 crystallizes in three different crystallographic structures: cubic PrO_2 calcium fluoride structure (C1-type with O_h^5 symmetry), manganese oxide Pr_2O_3 ($D5_3$ type with T_h^7 symmetry) and a hexagonal lanthanum oxide structure ($D5_2$ type with D_{3d}^3 symmetry). Pr_2O_3 layers are grown by MBE using a commercially available powder-packed ceramic Pr_6O_{11} source. The deposited crystalline layers show the Pr_2O_3 stoichiometry. Investigation is carried out with high resolution X-ray photoelectron spectroscopy (XPS), X-ray diffraction (XRD), X-ray reflectivity, TEM, selected area diffraction (SAD), etc. [69].

The results show that on hydrogen terminated Si(100) the Pr_2O_3 film grows in the manganese oxide structure in orthogonal (110) domains that follow the domains on the Si surface (Fig. 5.9). Such layers can not be epitaxially overgrown. RMS between the film and the top polysilicon is 0.5nm. On the other hand, on Si(111) a monocrystalline hexagonal structure grows and the films can be epitaxially overgrown [70]. AFM investigation shows a smooth top surface with RMS < 3 % [69]. The layers must not be exposed to air because ambient oxygen leads to the formation of an interfacial layer (SiO_2). A polysilicon capping layer of 50nm is sufficient for protection.

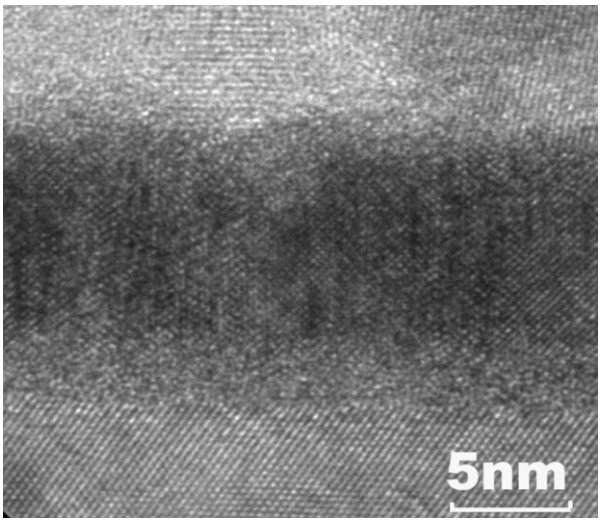


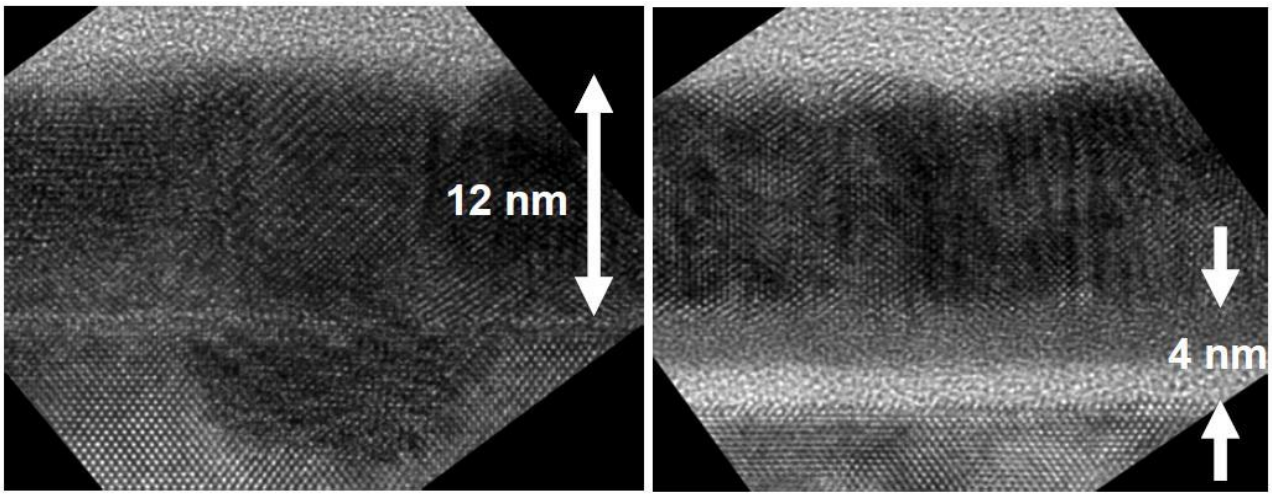
Fig. 5.9: TEM image of an epitaxial Pr_2O_3 film on a Si(100) substrate [69].

Pr_2O_3 layers on Si (100) have a dielectric constant of 30 [69]. The films show ultralow leakages of only 10^{-8} A/cm^2 for EOT 1.4nm. This is 10^4 better than the best published results with HfO_2 and is the result of heavy electron masses in the oxide. Breakdown occurs at 43MV/cm. Reversible breakdown is observed as well as excellent stress induced leakage current (SILC) characteristics. Pr_2O_3 exhibits symmetric band offsets of at least 1eV to Si. The bandgap is between 3 and 4eV. For annealed films, no significant hysteresis in the capacitance-voltage characteristic is observed. The layers survive anneals up to 1000°C for 15sec. without degradation [69].

Nd_2O_3

As an alternative to Pr_2O_3, Nd_2O_3 is investigated next. XPS measurements of MBE grown layers reveal sufficient bandgap and valence band edge offset: $E_g = 5.5 \pm 0.2$eV and $\varphi_{BV} = 2.65 \pm 0.2$eV. However, Nd_2O_3 exhibits lower stability in contact with Si than Pr_2O_3 leading to the formation of an interfacial silicide with

thickness up to several nanometers (Fig. 5.10a). The silicide can be turned into a silicate in a following anneal step (Fig. 5.10b) but this lowers the total dielectric capacitance of the dielectric stack significantly.



(a) (b)

Fig. 5.10: XTEM images of (a) as-grown and (b) annealed Nd_2O_3 layers grown on a Si(100) surface [71].

Tests with MOS structures with Al and Pt dots as gates are carried out. It is concluded that in the case of Al a reduction reaction takes place at room temperature (Eq. 5.1). It leads to the formation of an amorphous interracial layer that reduces K_{eff} and increases leakages through the gate dielectric. With Pt dots the results are better. The results obtained for a 6.8nm Nd_2O_3 layer are permittivity $K \approx 11$, and leakage current density $J=2*10^{-3}$A/cm^2@1V.

$$Nd_2O_3 + Al \rightarrow Al_2O_3 + Nd \quad (5.1)$$

Unfortunately, annealed samples are not delivered for MOSFET fabrication within the duration of the KrisMOS project. As later investigation shows, non-annealed Nd_2O_3 are not very suitable for gate dielectric material. Annealing for 10 minutes at 10^{-2}mbar and 600^{o}C results in the following characteristics for 10.5nm thick layers: $K \approx 13$, $J=10^{-4}$A/cm^2@1V.

<u>Gd_2O_3</u>

It is argued that better material stability and consecutively better device characteristics can be obtained if an oxide with a single valence state is used instead of an oxide with multiple valence states (Pr_2O_3). This is because multiple valence states bring the possibility of coexistence of phases with different oxygen content and transformations between them, as well as occurrence of mixed valence-state structures such as for example Pr_6O_{11} [72]. For this reasons Gd_2O_3, which exhibits only a valence state of +3 for Gd, is considered a potentially better candidate.

Gd_2O_3 films are grown on Si(100) pre-annealed (2x1) substrates using commercially available Gd_2O_3 source material (Fig. 5.11) [50]. Evaporation is carried out with an electron-beam evaporator and growth temperatures are in the range of 500-700˚C with growth rates of 0.01-0.05nm/s. The films exhibit the bixbyite structure in the (110) direction.

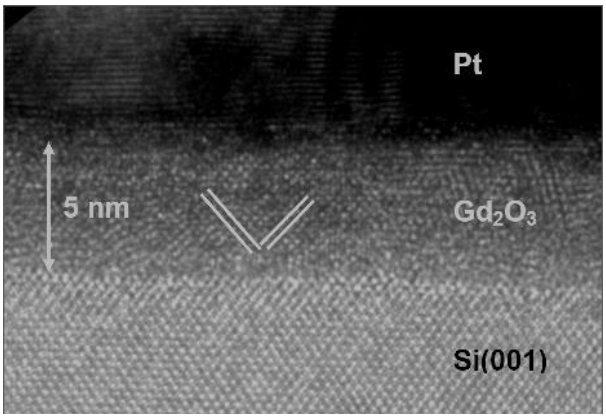


Fig. 5.11: TEM image of a Pt-Gd_2O_3-Si MOS structure [50].

The growth process is optimized by selecting the correct growth temperatures and the right oxygen supply. AFM is used for direct evaluation of the effects of different parameter values on the dielectric layer quality.

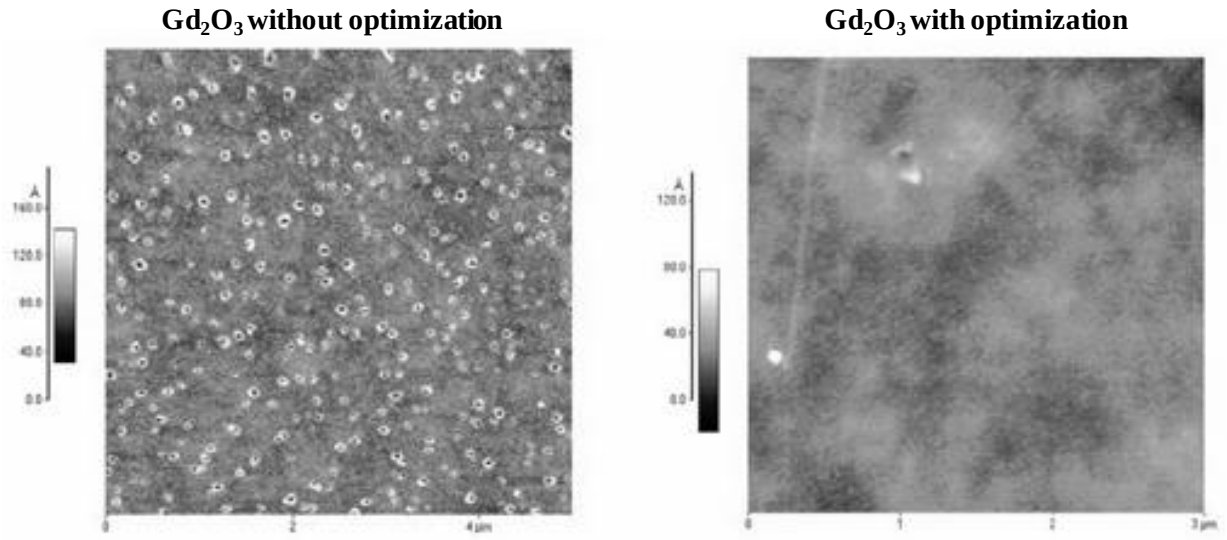


Fig. 5.12: AFM images of Gd_2O_3 layers grown (a) before and (b) after process optimization with additional oxygen supply [50].

Electrical measurements are carried out on MOS capacitors with in-situ evaporated Pt dot gates. With optimized growth conditions, gates stacks with EOT below 1nm yield leakages below 1mA/cm^2 at 1V. The dielectric constant is calculated to be 24±2 for layers of different thicknesses.

5.4. C-AFM characterization of Pr_2O_3 epitaxial films

Only polysilicon-coated samples of Pr_2O_3 are available for investigation. These are investigated in more detail in the following chapter. However, in view of the results in Chapter 6, the results in Chapter 5.3 can not be considered representative for the quality of the original material. Therefore, we rely on the obtained results by Osten and his group on capacitors evaporated Au gates through a shadow mask [69].

Measurements are performed on the Pr_2O_3 surface immediately after removing the polysilicon capping layer with SF_6/Cl_2 RIE. Surface topography and electrical current images are recorded simultaneously.

Usually smooth surface topologies exhibit low leakage currents as evident from (1) in Fig. 5.13. However, this is not always the case, as shown for site (2), where high leakage currents are detected in the current image despite the undistorted surface topography. Since no structural deficiency is observed, the excessive leakage is attributed to some kind of an electronic defect. Several additional high leakage spots seen in the current image can be clearly assigned to structural defects (3). Their size is up to a few hundred nanometers in diameter and their density is approximately $10^4 cm^{-2}$. In order to obtain further information on local electron transport mechanisms in epitaxial Pr_2O_3 gate oxides, measurements in the point-contact mode have been performed on structural defects and electronic defects as well as low leakage sites.

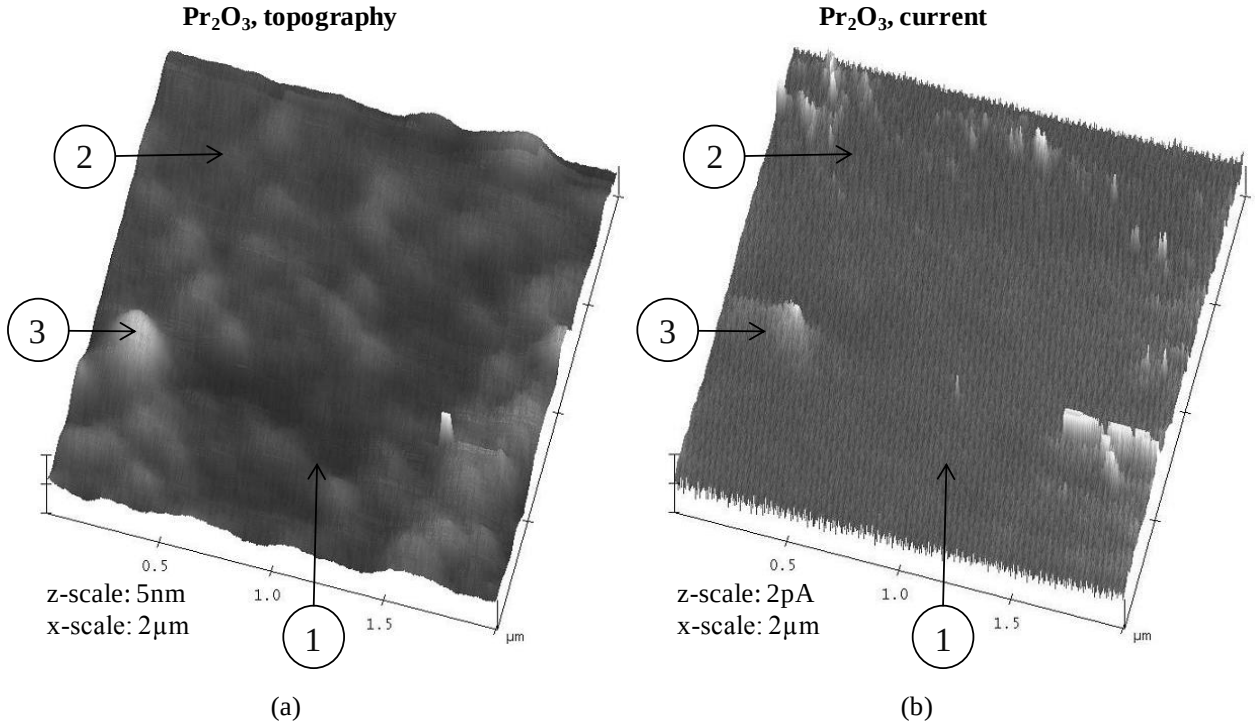


Fig. 5.13: (a) Topography and (b) electric current AFM images of an epitaxial Pr_2O_3 layer surface. Three different leakage mechanisms are identified (marked 1 to 3).

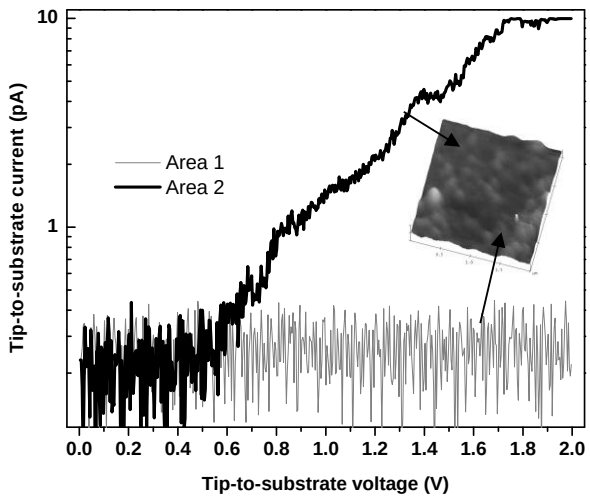


Fig. 5.14: Local current-voltage characteristics measured by C-AFM in the point-contact mode of two different sites with virtually identical surface topography. Area 1: non-leaky site, area 2: leaky site.

The corresponding current-voltage curves of an intact site (1) and a pure electronic defect (2) are shown in Fig. 5.14 for comparison. The enhanced current conduction through the electronic defect (2) may be due to Poole-Frenkel emission in the presence of trapped charge with high density. In contrast, for structural defects often Ohmic or Schottky-diode-like current-voltage characteristics are observed.

Further evidence that charge trapping occurs and will eventually lead to electronic defects is shown in Fig. 5.15. Scans are performed with an applied voltage of 7V. Both topography and electrical current data is collected. After performing three consecutive scans on the same area the creation of enhanced leakage sites is observed without a change in surface topography. This is a very important result. It suggests that positive charge is being trapped or negative charge is being detrapped at certain areas of the high-K material at high electric fields.

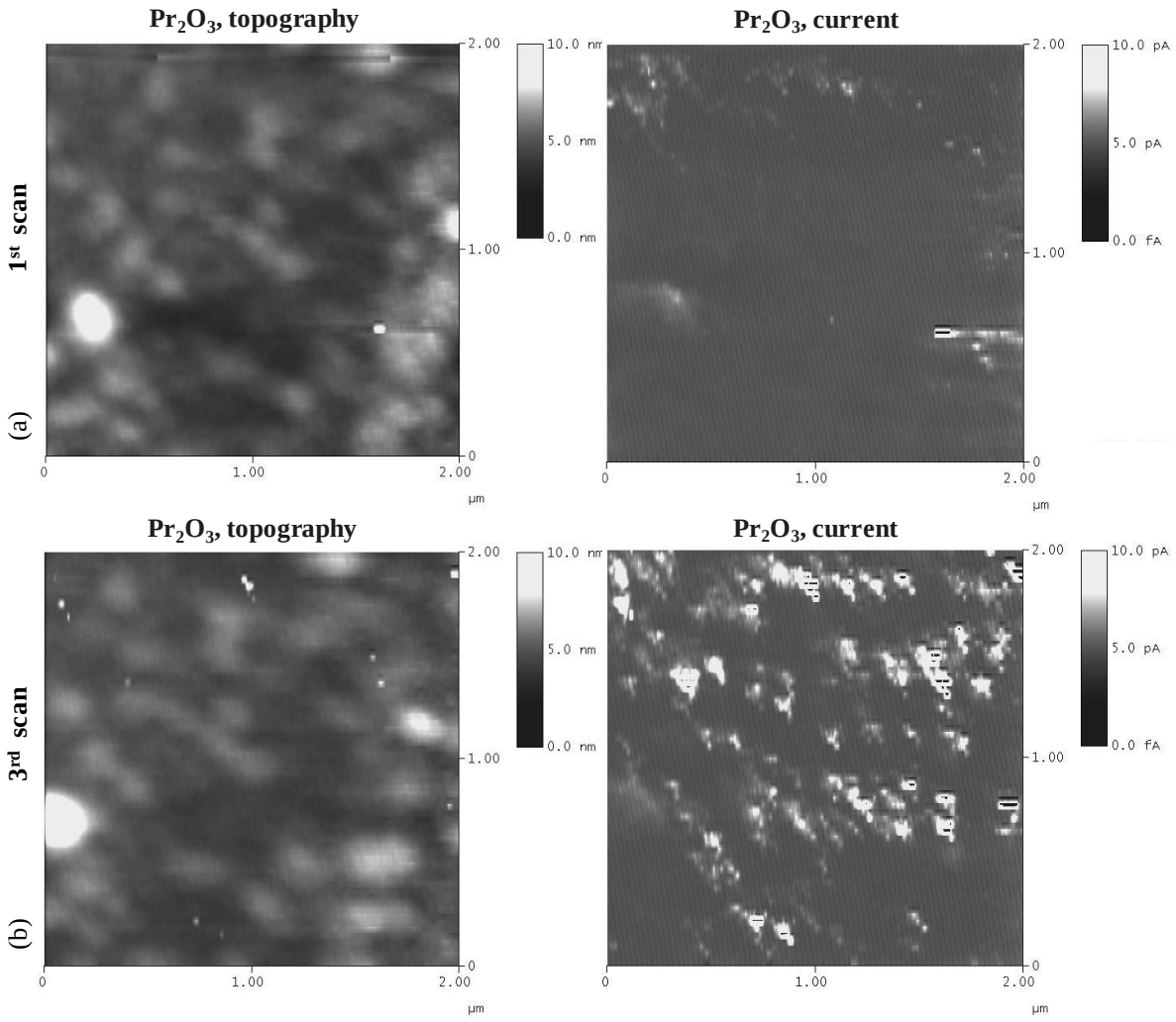


Fig. 5.15: Topography (left) and electric current (right) AFM images of the Pr_2O_3 surface of (a) fresh Pr_2O_3 film and (b) film with built-up trapping sites after 3 consecutive scans.

Different leakage mechanisms in Pr_2O_3 films are discovered with the help of C-AFM. Charge trapping is observed on the nanoscale for the first time. Based on the C-AFM results, predictions are made regarding device characteristics of conventionally integrated MOSFETs with Pr_2O_3 gate dielectric. It is expected that similar trapping effects will be observed in Pr_2O_3 MOSFETs. Gate leakages through extrinsic defects within the gate dielectrics are also likely.

5.5. C-AFM characterization of Nd_2O_3 and Gd_2O_3 epitaxial films

C-AFM is performed on the grown Gd_2O_3 layers without a polysilicon cap. Gd_2O_3 layers exhibit good topographic uniformity. Tunneling currents are also uniform (Fig. 5.16). Spectroscopy measurements (Fig. 5.17) suggest that breakdown occurs at fields above 5MV/cm.

From the C-AFM data it can be concluded that structurally and electrically the quality of the grown layers is very good and from this point of view they are good candidates for MOS integration.

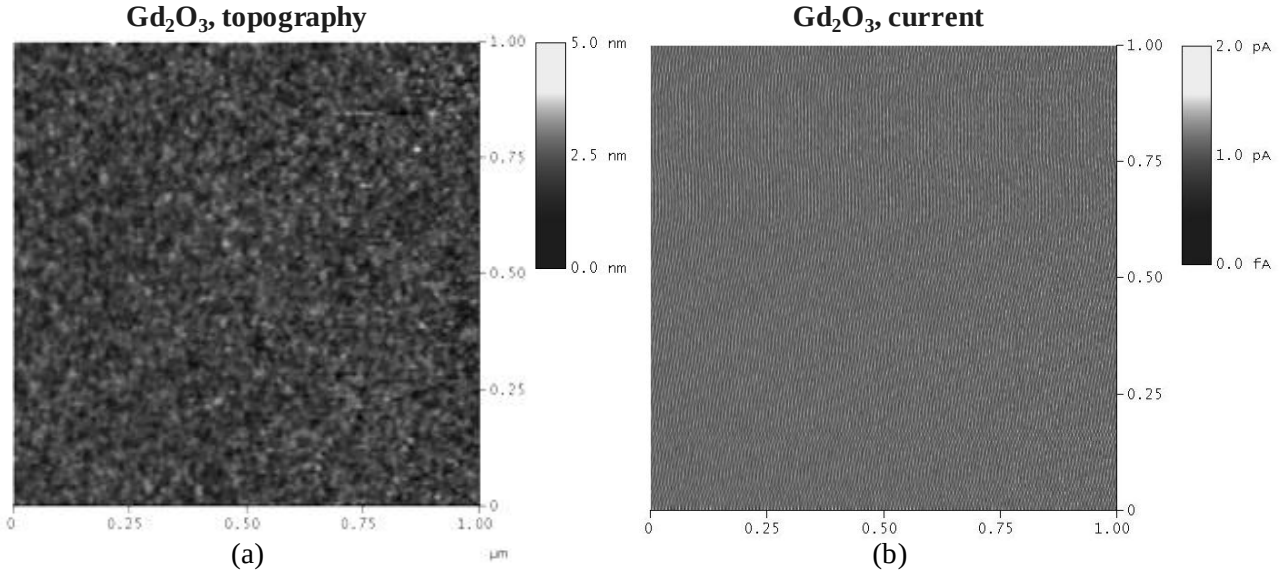


Fig. 5.16: (a) Topography and (b) electric current AFM images of the Gd_2O_3 surface at 6V sample-to-tip voltage.

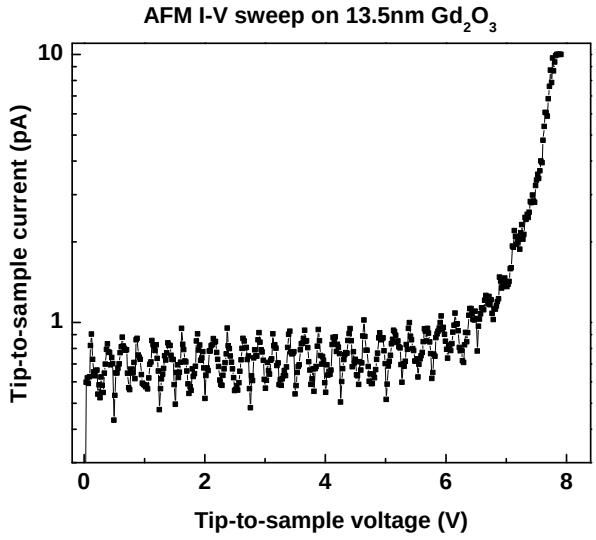


Fig. 5.17: Graph of typical AFM current-voltage characteristic performed in point contact mode on the surface of 13.5nm Gd_2O_3.

For Nd_2O_3 layer, the situation is different. The topography is as smooth as the one of Gd_2O_3 (Fig. 5.18a), however the electric image shows a very non-uniform distribution of current densities at high electric fields (Fig. 5.18b).

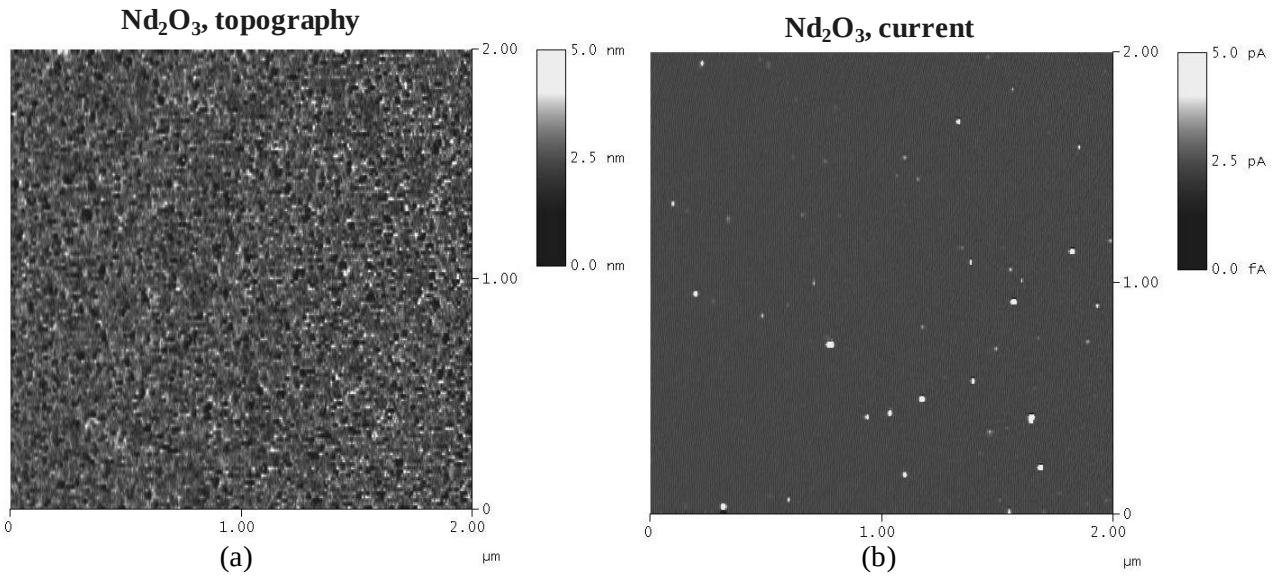


Fig. 5.18: (a) Topography and (b) current AFM images of the Nd_2O_3 surface at 6V sample-to-tip voltage.

5.6. Conclusions

C-AFM proves its applicability for investigation of thin films. With its help, individual investigation of separate leakage mechanisms on the high-*K* surface is possible. The information it brings can not be obtained by macroscopic device electrical characterization due to the superposition of these mechanisms (Fig. 5.19a). On the other hand, C-AFM allows instantaneous nanoscale MOS-structure formation using the conductive tip as a gate. In this way, different defects can be distinguished with very high special resolution (Fig. 5.19b). They can be studied separately by spectroscopic current-voltage measurements in point-contact mode at a position of interest.

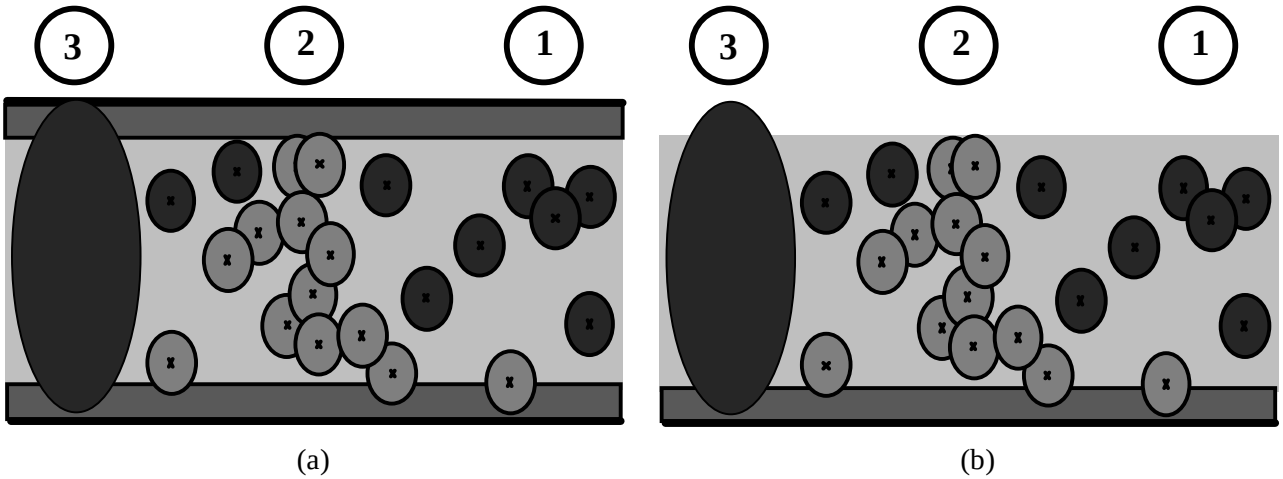


Fig. 5.19: Schematics of different leakage paths in high-*K* gate dielectrics. They can not be evaluated independently by macroscopic device evaluation (a), but only microscopically prior to gate electrode deposition (b).

Verification of the applicability of C-AFM for thin film investigation is carried out using state-of-the-art thin SiO_2 film. Next, a first generation dual-layer TiO_2SiO_2 dielectric stack is evaluated. Different defects are found that help understand the degradation mechanisms during high-temperature annealing. Finally, second generation epitaxial films are studied. It is concluded that high quality high-*K* layers can be obtained with MBE by selecting the proper material and performing the necessary

optimization of growth conditions. However, these films seem to be susceptible to charge trapping effects, observed for the first time on the nanoscale level directly on the high-*K* surface by C-AFM. The trapping susceptibility is likely due to RIE related damage. Arguably, C-AFM helps foreseeing macroscopic electrical behavior of devices based on the investigated dielectric layers. The charge trapping is very likely to be observed in complete conventionally integrated MOSFETs incorporating epitaxial oxides if RIE is used for gate structuring.

The epitaxial oxides are selected for further investigation. Pr_2O_3 and Gd_2O_3 show promise for successful MOS integration as very good homogeneous thin films can be grown on silicon. Despite the indications of a very high density of traps in Nd_2O_3 layers, investigation of this material is also carried on.

6. CMOS compatibility tests

In view of the theoretical advantages epitaxial rare-earth metal oxides offer and the experimental results which support them, the integration effort is placed on this type of high-*K* materials.

A complete CMOS fabrication process contains up to several hundred fabrication steps depending on its complexity. For successful integration, the high-*K* material must be compatible with all of them. Unfortunately, there is no previous information about the impact of different process steps on rare-earth metal oxides. Therefore, thorough investigation of the selected high-*K* dielectrics is performed. The compatibility of the materials with cleaning, etching and annealing steps in CMOS processing is tested. The tests are grouped in 4 categories: wet chemistry, photoresist removal, reactive ion etch and anneals and examined separately in the next four chapters. The effects of each processing step on the dielectrics are evaluated by AFM and elipsometry measurements.

The Institute for Semiconductor Technology and Nanoelectronics in Darmstadt University of Technology houses the only complete semiconductor fabrication technology line in the state of Hessen, Germany. Its equipment allows CMOS processing from the very initial stages – blanket silicon wafers to complete packaged chips. All process compatibility steps are performed in the ISTN cleanrooms.

6.1. Wet chemistry

The chemical reactivity of the investigated epitaxial rare-earth metal oxides materials is very similar due to the similar chemical properties of rare-earth metals. Only the results for Nd_2O_3 are shown as an example for those steps that affect all tested materials in the same way. Fig. 6.1a shows the fresh condition of a Nd_2O_3 sample before any intervention on it. RMS roughness is S_q=0.494nm and the surface is free from contaminations.

HF dip: diluted HF is used to remove native SiO_2 from the silicon surface. Usually, a 2% solution of HF in water is used and the wafer is dipped in the solution at room temperature for a short period of time: 1-2min. In Fig. 6.1b, an AFM image taken after performing a 1min. 0.5% HF dip is shown. Roughness is increased more than three times after the procedure to S_q=1.738nm. However, this increase is a result of particles attached on the wafer surface, most probably from the solution. The high-*K* surface beneath is practically unchanged. Elipsometry confirms the conclusions made by AFM – no change in film thickness is detected.

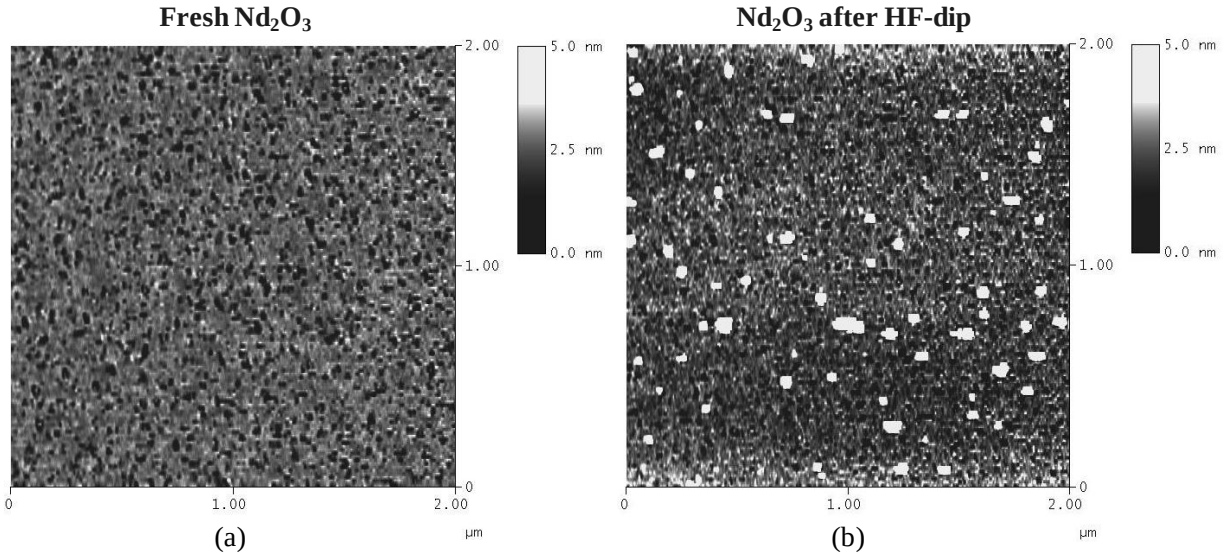


Fig. 6.1: Topography AFM images of the Nd_2O_3 surface (a) before and (b) after HF-dip.

Buffered HF is used for controllable removal of SiO_2 and Si_3N_4. A 6:1 volume ratio of 40% NH_4F and 49% HF in water is commonly used. The tests reveal that all examined rare-earth metal oxides are soluble in buffered HF. The AFM image after buffered HF followed by DI water cleaning (Fig. 6.2a) indicates complete dissolving of the Nd_2O_3 film as well as additional contaminations. Elipsometry is not applicable because of the rough surface.

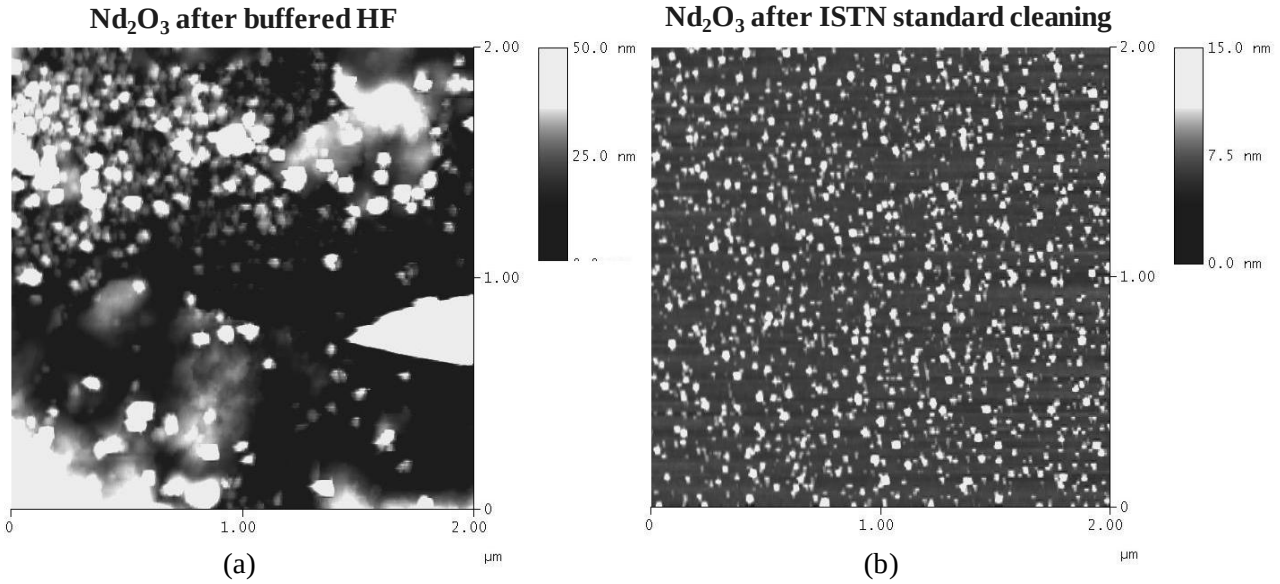


Fig. 6.2: Topography AFM images of the Nd_2O_3 surface (a) after buffered HF and (b) after ISTN standard cleaning.

'Standard cleaning' is a commonly used cleaning step at the ISTN for cleaning of fresh wafers, after photoresist removal, etc. The wafers are submerged in a 95% H_2SO_4:65%HNO_3 (10:1) solution initially at 95˚C for 10min and afterwards at room temperature for additional 5min. The effect of the standard cleaning on rare-earth metal oxides is illustrated in Fig. 6.2b. On first look it seems as if the high-*K* surface is contaminated by a large number of particles. However, the underlying surface is almost atomically flat, which leads to the conclusion that this is actually the silicon substrate and the particles are the remains of the dielectric which has been destroyed. The elipsometer indicates thickness of a couple of nanometers.

Next, a test is performed with the alkaline part of the RCA clean - the first step also known as SC-1 clean. The NH_4OH - H_2O_2 - H_2O solution is used for cleaning organic contamination. AFM reveals a very clean surface after a 0.25:1:5 mixture at 70˚C for 15min (Fig. 6.3a) that looks identical to the fresh surface.

Diluted HCl is used for removing metallic contaminations. A 1min. dip in a 0.5% water solution leads to similar results as the complete RCA clean (Fig. 6.3b). Most of the surface is fully stripped of the film. Elipsometry however, fails in this case. Due to the large amount of remains on the surface, the elipsometer gives an erroneous value indicating that half of the film thickness is still intact. The high-*K* materials withstand only very diluted solutions (1:10000) for a couple of minutes.

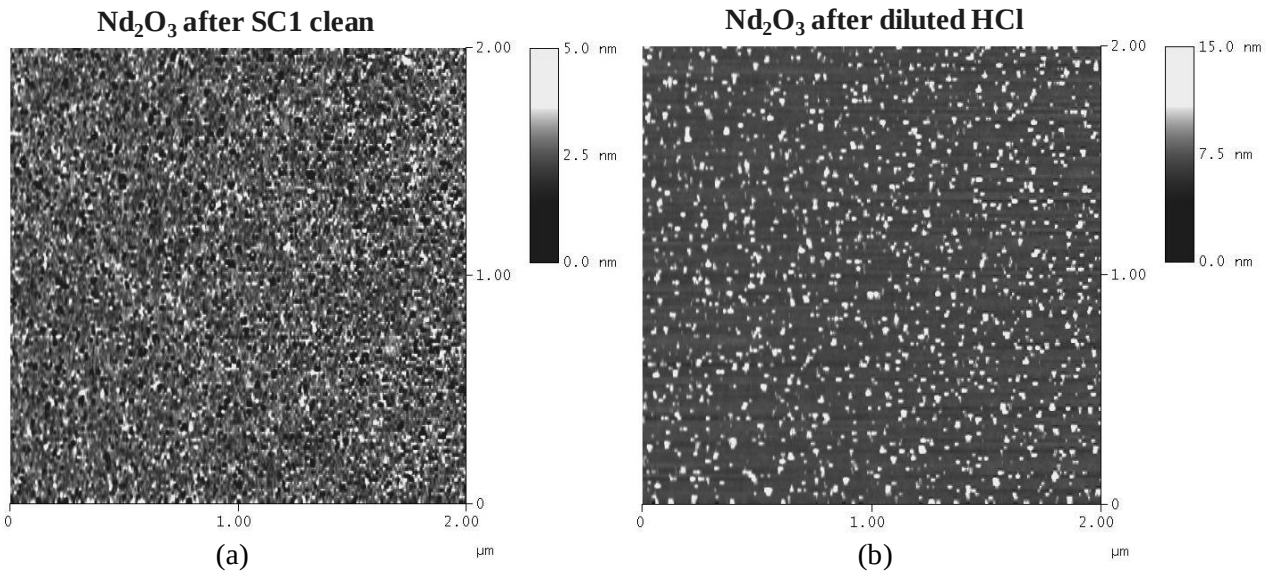


Fig. 6.3: Topography AFM images of the Nd_2O_3 surface (a) after SC1 clean and (b) after dilluted HCl.

Several other strong-acid solutions are used at the ISTN, the most common of which are:

1. Si_3N_4-etchant: 85% H_3PO_4 at 155˚C
2. Polysilicon-etchant: 72g HIO_3, 82gml 40% NH_4F, 18ml 50% HF and 3680ml H_2O
3. SP clean (sulfiric/peroxide): removal of organic contaminations: 1:4 30% H_2O_2 and 95-97% H_2SO_4 at 90˚C for 10-15min.

All of these solutions yield similar results – the high-*K* dielectric material is destroyed.

These results are crucial for the successful integration of the high-*K* gate dielectrics. Clearly, unlike SiO_2, rare-earth metal oxides are not stable in acid solutions. Some acids can be used for selective removal of the high-*K* from the silicon surface. The results are summarized in Table 3.

Procedure	**Effect on high-*K***
HF-dip, 0.5%, 1min, RT	Stable
Buffered HF	Soluble
ISTN standard cleaning	Soluble
SC1 clean	Stable
H_3PO_4, 85%, 155˚C	Soluble
HCl, 0.5%, 1min	Soluble
Deionized water	Stable

Table 3: Stability of rare-earth metal oxides against common CMOS wet etching and cleaning steps.

Comparison of the AFM results with those obtained by elipsometry [73] speaks in favor of the former. AFM gives more adequate and detailed information about the surface condition.

6.2. Photoresist removal

The next step is to test the dielectric stability against wet chemicals used for photoresist removal after photolithography. There are different methods for removing photoresists. The high-*K* dielectrics remain stable after cleaning with organic solvents. Boiling acetone (56.53˚C) for 15min has no noticeable effect on the dielectrics (Fig. 6.4a). AZ100 photoresist remover at 80˚C for 30min. also does not seem to attack them (Fig. 6.4b). The roughness values after the two cleaning steps are 0.611nm and 0.562nm respectively.

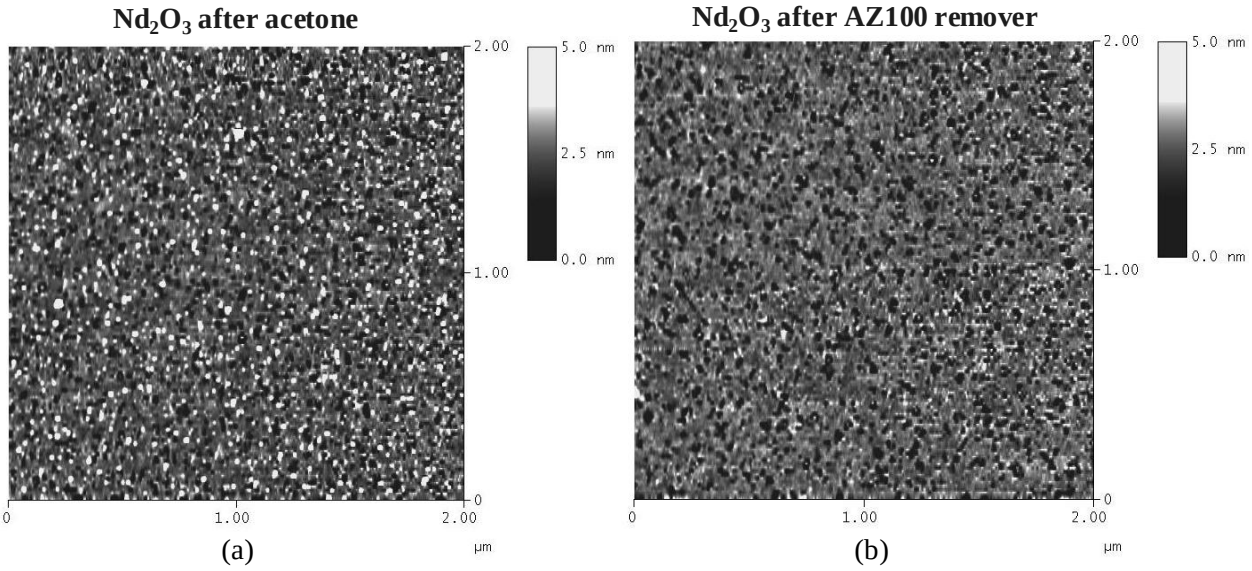


Fig. 6.4: Topography AFM images of the Nd_2O_3 surface (a) after acetone and (b) after AZ100 remover.

O_2 plasma ashing is another commonly used procedure for photoresist removal. AFM indicates only a small increase in roughness (S_q=1.461nm) after a 15+90min ashing for resist removal (Fig. 6.5) and practically no change after the 15min 'O_2 dip' used for cleaning non-polimerized organic remains on the photoresist. The increased roughness after a long exposure to O_2 plasma could indicate some structural change. Elipsometry indicates no thickness change.

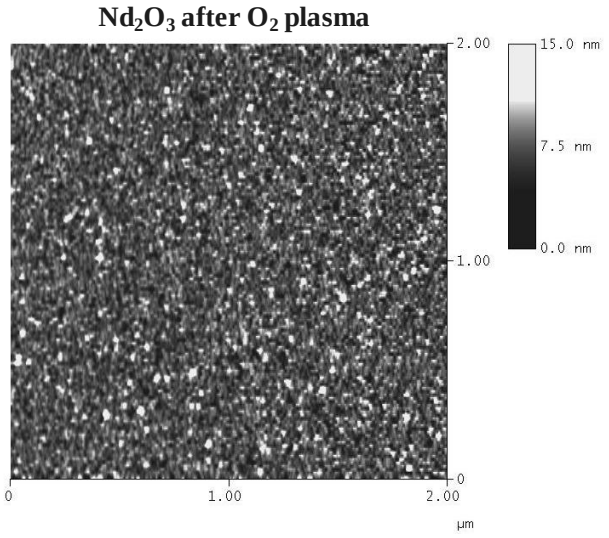


Fig. 6.5: Topography AFM image of the Nd_2O_3 surface after O_2 plasma ashing.

More aggressive wet chemistry photoresist removal solutions such as H_2SO_4/HO_2 and HNO_3 (100%) dissolve the dielectrics. The results are summarized in Table 4.

Procedure	**Effect on high-*K***
Acetone	Stable
H_2SO_4/HO_2	Soluble
HNO_3 (100%)	Soluble
AZ100 wet remover	Stable
O2-plasma ashing	Unstable

Table 4: Stability of rare-earth metal oxides against common CMOS photoresist removal steps.

6.3. Ion implantation

Through source and drain ion implantation dopants are introduced into the silicon substrate in the source and drain areas in order to change their conductivity. The high kinetic energy phosphorus and boron ions in the case of respectively n+ and p+ implantation are expected to have an impact on the structure of the high-*K* material.

The standard used dose for source/drain implantation of $10^{15}cm^{-2}$ at 50keV results in a clear change of the surface (Fig. 6.7). The roughness is increased to S_q=0.879nm and there is a clear directional pattern. Unfortunately, using a standard gate-first integration scheme, there is a no possibility to avoid this damage. On the good side, damage is expected to concern only a very narrow region at the periphery of the gate stack.

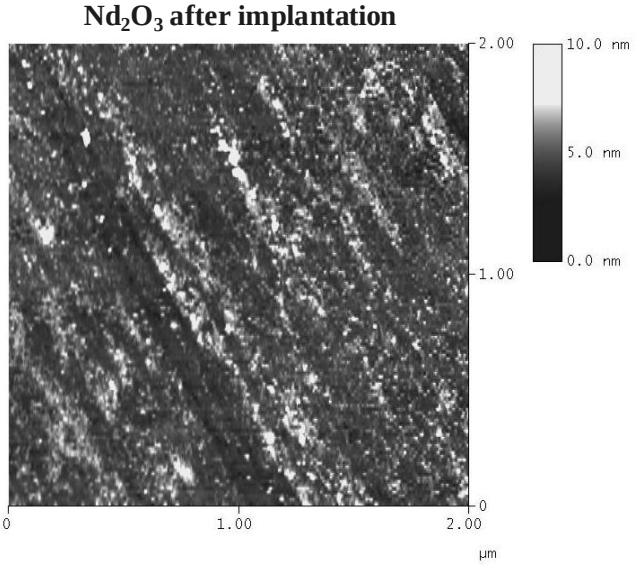


Fig. 6.6: Topography AFM image of the Nd_2O_3 surface after ion implantation.

In order to reduce the damage and distribute it along the gate periphery uniformly instead of concentrating it in one area, the implantation is done in four steps at one fourth of the dose with rotation at 90° after each implantation.

6.4. Anneals

High temperature anneals are needed after source and drain ion implantation for healing structural damage in the silicon substrate and for dopant activation. High temperatures can lead to crystallization and/or decomposition of the dielectric combined with interfacial silicide formation. As a matter of fact, four-digit temperatures are often lethal for high-*K* materials [74].

The epitaxial rare earth-metal oxides: Nd_2O_3, Pr_2O_3 and Gd_2O_3 appear stable after 10sec RTA at 900˚C. Pr_2O_3 and Gd_2O_3 show similar results to Nd_2O_3 (Fig. 6.7).

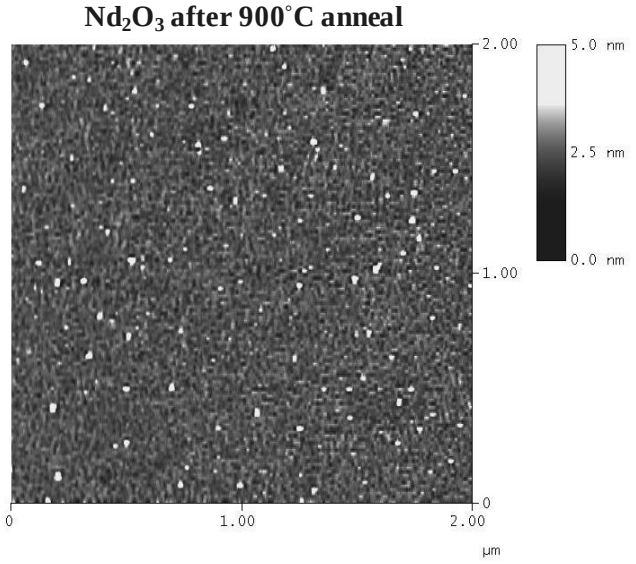


Fig. 6.7: Topography AFM images of the Nd_2O_3 surface after a 900˚C anneal.

Other investigated MBE oxides grown without sufficient growth optimization show inferior stability to high temperatures. In Fig. 6.8, the results of the annealing tests performed on a mixed rare-earth metal oxide are shown. This material is an MBE-grown epitaxial layer comprising a mixture of two rare-earth metal oxides. Spots in some areas of the mixed oxide are evident after 850˚C anneal (Fig. 6.8a). After 900˚C the deterioration is spread over the entire surface (Fig. 6.8b).

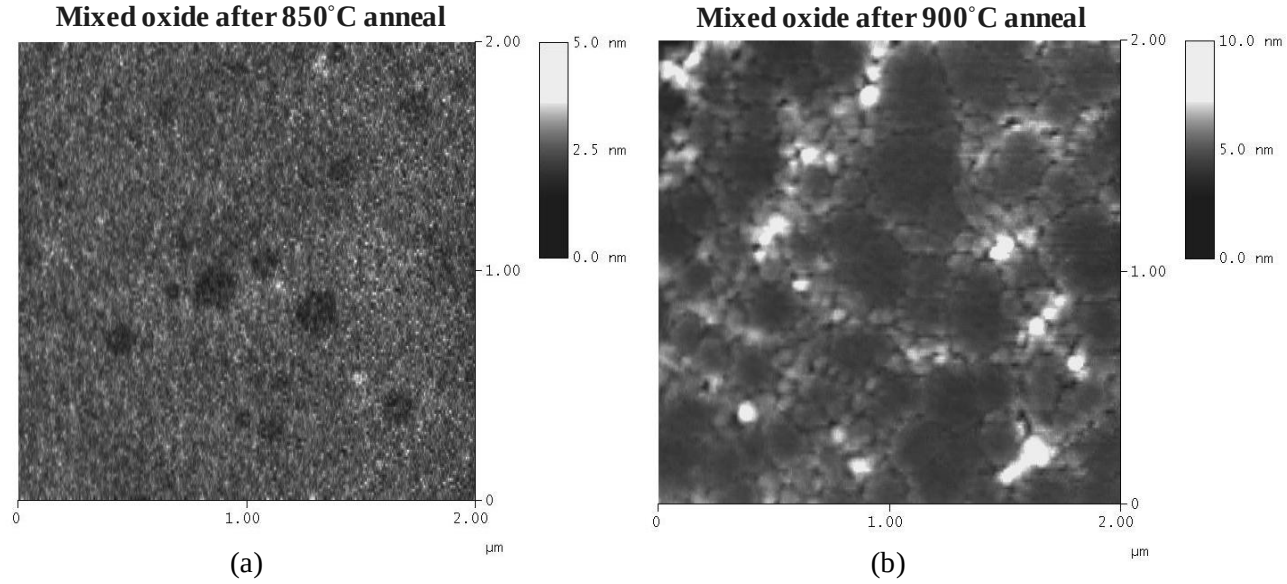


Fig. 6.8: Topography AFM images of mixed rare-earth metal oxide (a) after 850˚C anneal and (b) after 900˚C anneal.

In spite of their higher stability compared to amorphous high-*K*, epitaxial oxides are also not as stable against annealing as SiO_2. However, if grown at optimizated conditions, most of them are stable enough to survive sufficient dopant activation.

6.5. Reactive ion etch

There are several reactive ion etch steps during the MOSFET manufacturing process. The most critical of them by far is the RIE of polysilicon for gate formation. The development of a highly selective RIE processes is another major issue for a successful process integration of praseodymium oxide in a polysilicon gate CMOS process. The polysilicon gate etch has to stop on the dielectric. It is critical that this process does not attack the high-*K* material for two reasons. Not only can the gate dielectric beneath the gate be damaged, but also silicon substrate around the gate could be etched, changing the MOS device geometry and degrading its characteristics additionally.

The two used polysilicon RIE processes at the ISTN cleanroom are RIE with $SiCl_4$ and RIE with SF_6/Cl_2. RIE with $SiCl_4$ is tested first. The stability of the high-*K* material to this process is very low for all tested materials. After a 2min of etch, the layers are destroyed. The damage is visible even with low magnification optical microscopy. However, the defects look like holes. The C-AFM images (Fig. 6.9) indicate that the spots are actually hillocks of remaining high-*K* material on the silicon surface.

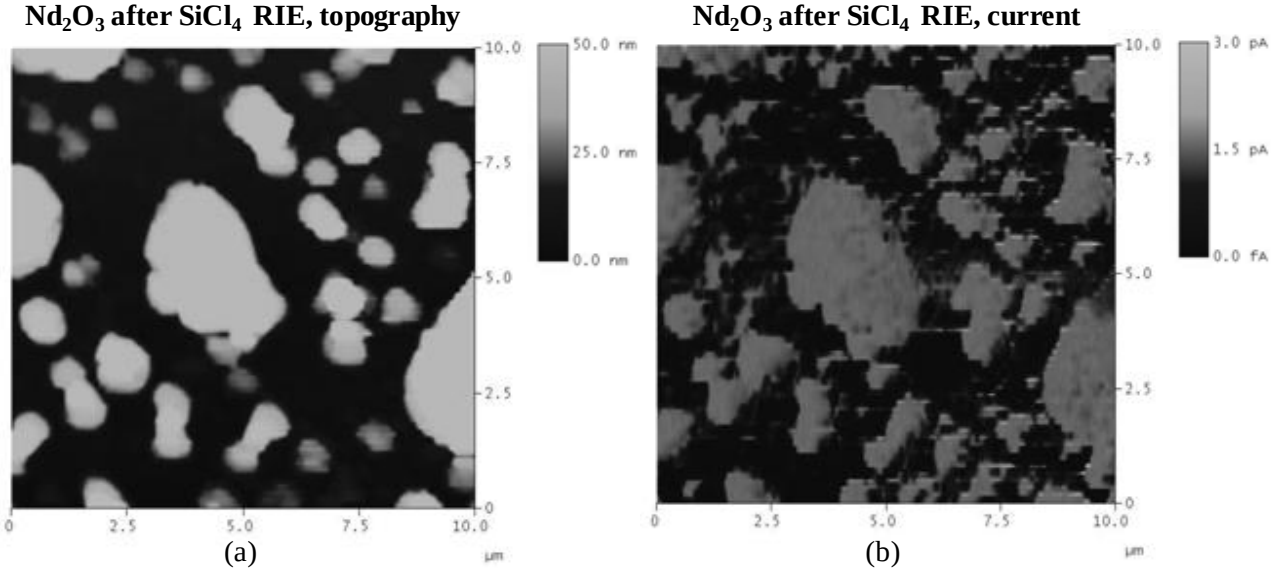


Fig. 6.9: (a) Topography and (b) electric current AFM images of the Nd_2O_3 surface after $SiCl_4$ reactive ion etch.

Next, SF_6/Cl_2 is tested. RIE for 2min is performed. The results vary for the different materials. Pr_2O_3 and Gd_2O_3 seem reasonably stable. Although the roughness of the layers is increased to S_q=1.26nm (Fig. 6.10), they are completely intact. The results show selectivity above 300 between high-*K* and polysilicon. Results on Pr_2O_3 are presented in Chapter 5.4.

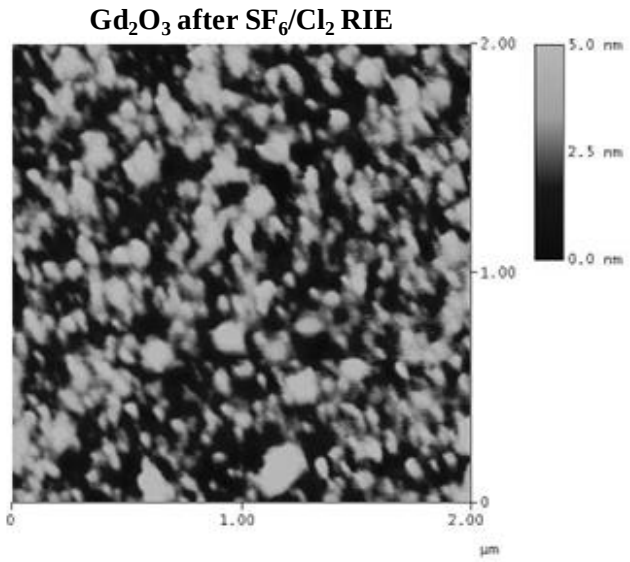


Fig. 6.10: Topography AFM image of the Gd_2O_3 surface after SF_6/Cl_2 RIE.

For Nd_2O_3 defects are observed on the surface. A more detailed investigation is carried out. The surface is scanned in contact mode (Fig. 6.11). The electric current image shows no currents anywhere on the surface, confirming that the planar area around the defects is indeed the high-*K* surface and not the silicon substrate. Elipsometry indicates insignificant change in the layer thickness compared to its original state. The next measurement is performed on the exact same area but with much larger force between the tip and the surface. The force is estimated to approx. 50pN for the second scan versus approx. 1pN for the first scan. The results are shown in Fig. 6.12. The large bumps have turned into holes. The electric current image indicates some conductivity in the holes. Another scan of the same area is performed with the force reduced back to 1pN. The topography and the currents do not change significantly.

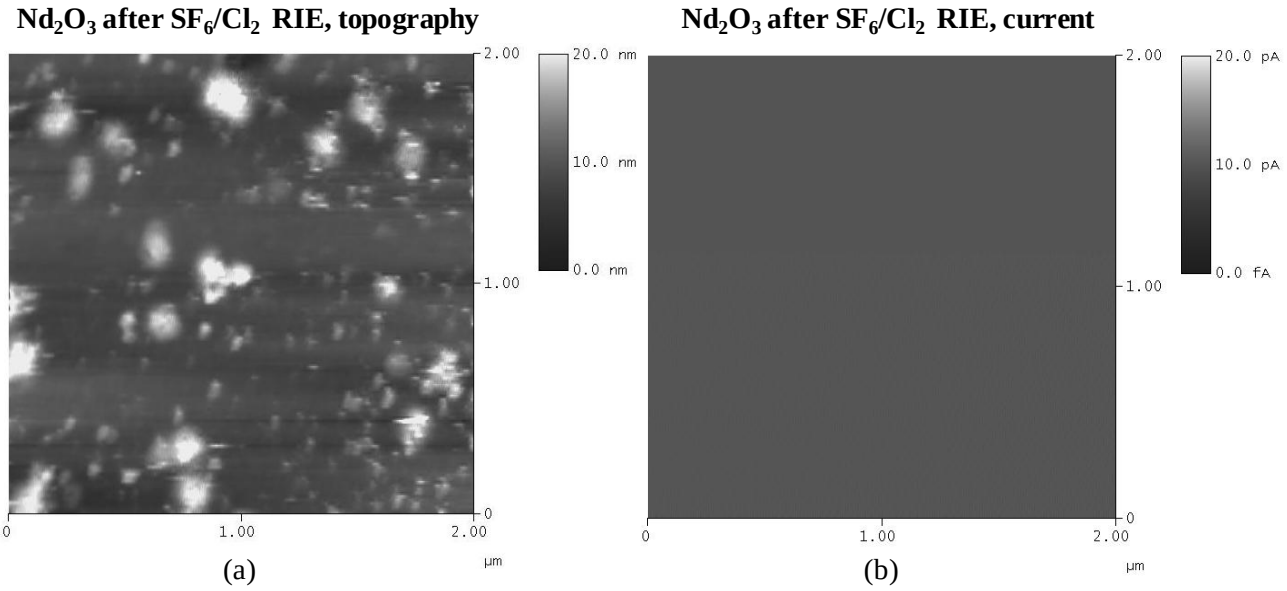


Fig. 6.11: (a) Topography and (b) electric current AFM images of the Nd_2O_3 surface after SF_6/Cl_2 reactive ion etch. Images obtained with low tip-sample interaction force.

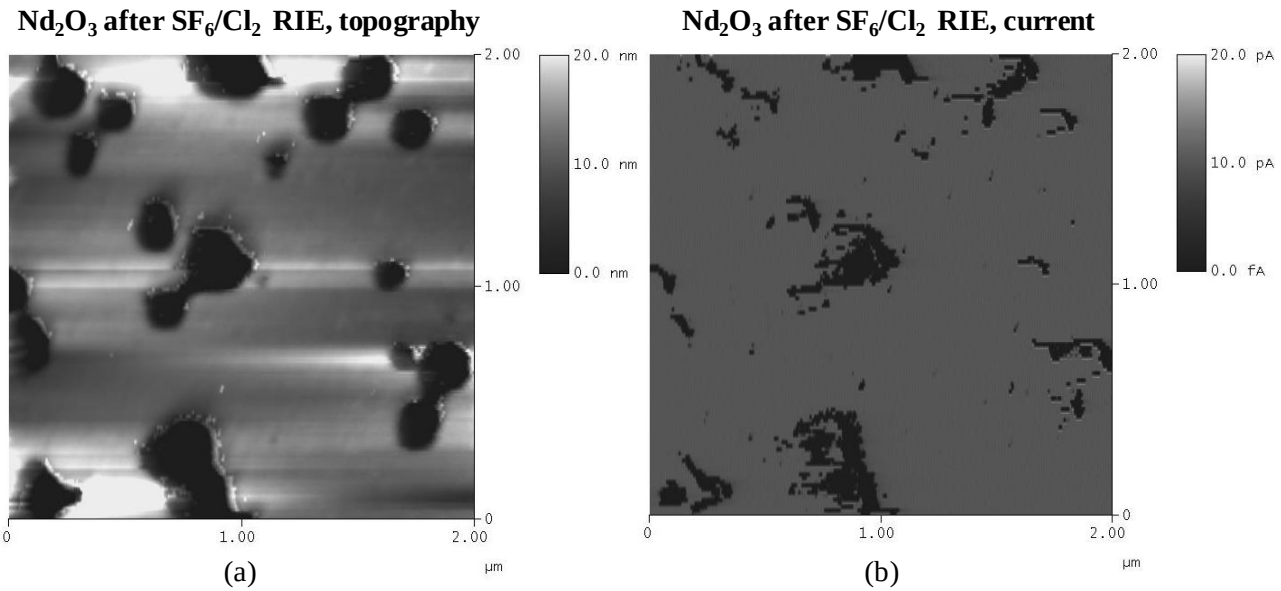


Fig. 6.12: (a) Topography and (b) electric current AFM images of the Nd_2O_3 surface after SF_6/Cl_2 reactive ion etch. Images obtained with high tip-sample interaction force.

It is concluded from these measurements that the defects on the surface are hollow bulbs. The RIE ions have been able to penetrate below the high-*K* and etch pits in the silicon substrate. When the second scan is performed, the upper surface of the surface of the bulbs collapses inward, allowing the tip to penetrate the empty space and get in electrical contact with the conductive substrate in the bottom. The currents are not uniform inside the holes. This can be explained with the inability of the tip to follow the edge because of the high aspect ratio of the holes. One additional factor is the debris from the collapsed high-*K* material inside. One might wonder how the solid high-*K* film has bent up to form a curved surface without breaking. The explanation lies in the different scales in the vertical and horizontal directions. The largest hills are actually less than 20nm high over a horizontal distance of more than 200nm.

6.6. Conclusions

With the help of AFM it is determined which of the standard processing steps are compatible with the rare-earth metal oxides and to what extent. It is determined that some of the novel materials can be integrated into a standard gate-first process as long as the appropriate etching, cleaning and annealing procedures are kept. However, it is clear that the stability of the materials is not as high as that of SiO_2 and for obtaining high-performance devices, a modification of the integration process as a whole is necessary. Three key steps cause problems to the rare-earth metal oxides, namely the polysilicon RIE, the source/drain implantation and the high temperature source/drain activation anneal.

Gate structuring is performed by RIE. During this type of etch, high-energy gamma radiation is present in the processing chamber. This radiation can create large amounts of trapping sites. Additionally, RIE causes direct damage to the sidewalls of the gate stack by bombarding it with ions (Fig. 6.13a). Moreover, long term defects are created due to radical residues on the wafer.

Similarly to RIE, ion implantation causes sidewall damage by direct bombardment with ions. Since ions trajectories are highly directional and they are directed at a small angle to the normal of the wafer surface, certain part of the high-*K* dielectric at the periphery of the gate stack is going to be affected more than the rest (Fig. 6.13b)

Finally, high-temperature source/drain activation anneals also have a negative structural effect on the integrity of the dielectrics (Fig. 6.13c) and this may affect their electrical properties. None of the investigated materials can survive standard annealing CMOS conditions with anneal temperatures of 1000˚C and above. Some of the materials survive 900˚C for 10sec which is enough for a reasonable level of dopant activation.

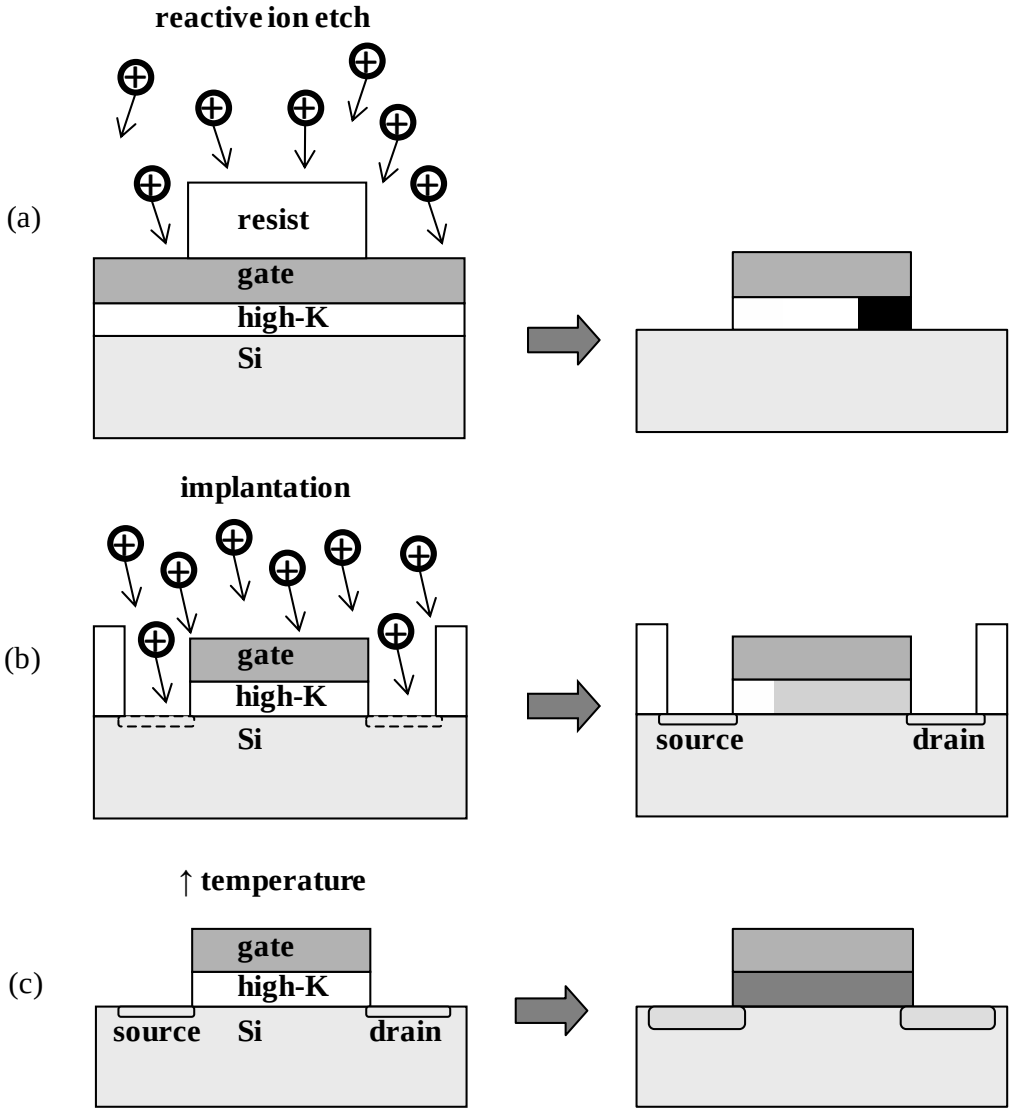


Fig. 6.13: Schematics of critical process steps (left) andtheir effect on the gate dielectric (right): (a) reactive ion etch, (b) ion implantation and (c) high-temperature anneal.

The epitaxial Pr_2O_3, Nd_2O_3 and Gd_2O_3 dielectrics are chosen for integration into a modified gate-first process. The evaluation on device level will show to what extent the observed microscopic structural defects effect MOSFET characteristics.

7. Gate-first integration of nMOSFETs with epitaxial gate dielectrics

A modified gate-first fabrication process is developed as a result of the AFM investigation on the effect of different processing procedures on the epitaxial high-*K* dielectric materials. Having in mind the considerations in Chapter 3.4.2, for simplicity only nMOSFETs are manufactured in this first integration attempt. Pr_2O_3 and Nd_2O_3 are integrated at the ISTN and Gd_2O_3 is integrated by our partners at AMO GmbH Aachen within the framework of the KrisMOS project [75].

7.1. Pr_2O_3 gate dielectric nMOSFETs

3" Si(100) wafers are used. Some of them receive LOCOS device isolation and the rest will be self-isolated. 17nm Pr_2O_3 layers (EOT=1.8nm) are grown by Prof. Osten's group. The layers are subsequently covered in-situ with 50nm undoped polysilicon as the unprotected layers are not stable against air. After high-*K* dielectric growth, the wafers are transported to the ISTN fabrication facility, where the rest of the fabrication process is carried out. After SC1 clean and a brief HF-Dip a 250nm in situ phosphorous-doped polysilicon layer is deposited at approximately 750°C. The n+ polysilicon gate stack is defined by optical lithography and SF_6 /Cl_2 RIE.

Subsequently, n+ source/drain phosphorus implantation with a dose of $10^{15}cm^{-2}$ at 50keV is performed and a brief RTA anneal at 900˚C for 10s in an inert ambient is carried out to activate the implanted dopants. An additional lithography is used to define the implantation areas on the wafers with self-isolation. The wafers with LOCOS are implanted directly because the isolation areas define the active areas. Low temperature oxide is deposited to protect the structures and contact holes are etched. Conventional Al/Ti back-end metallization is carried out and a forming gas anneal at 425˚C for 10 min. completes the process.

nMOSFETs with gate lengths from 100µm down to 4µm are fabricated. The gate width is kept constant at 100µm. Fig. 7.1 shows part of the test chip of a completed wafer with Pr_2O_3 nMOS devices of different gate lengths.

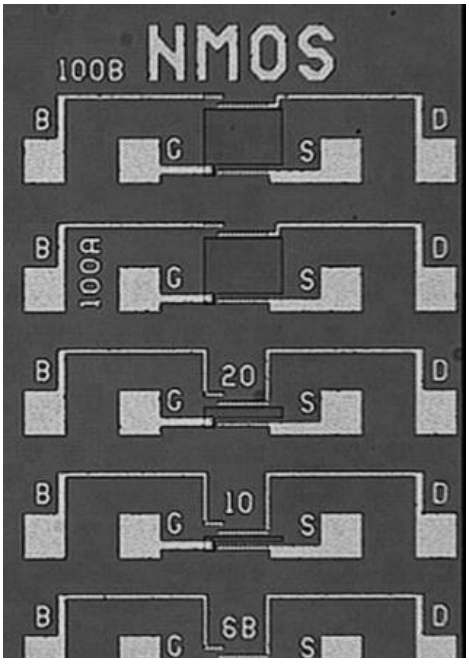
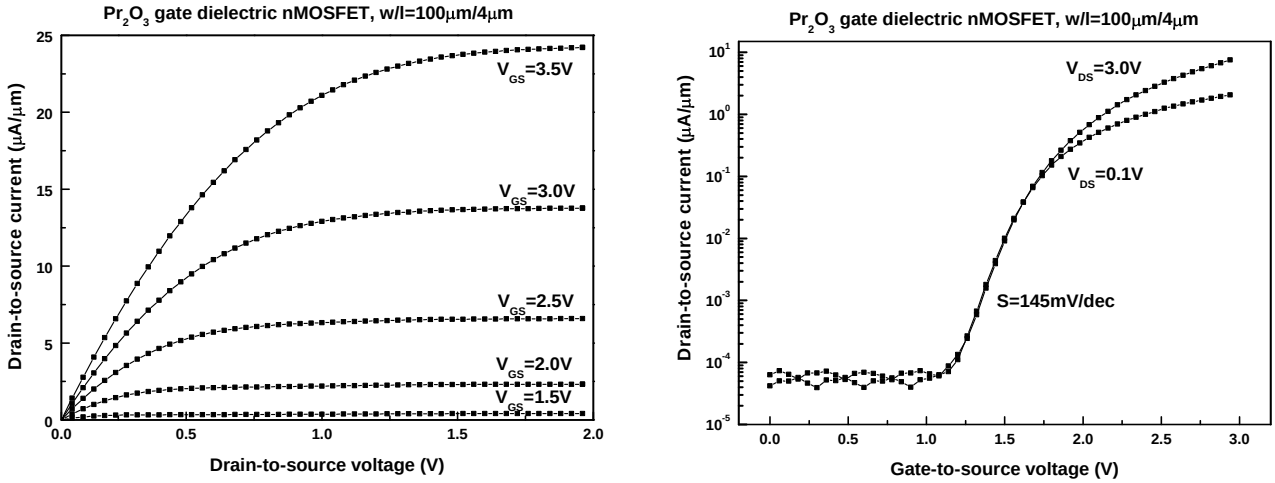


Fig. 7.1: Micrograph of the first functional Pr_2O_3 gate dielectric MOSFETs, manufactured at the semiconductor processing facility of ISTN, TU Darmstadt.

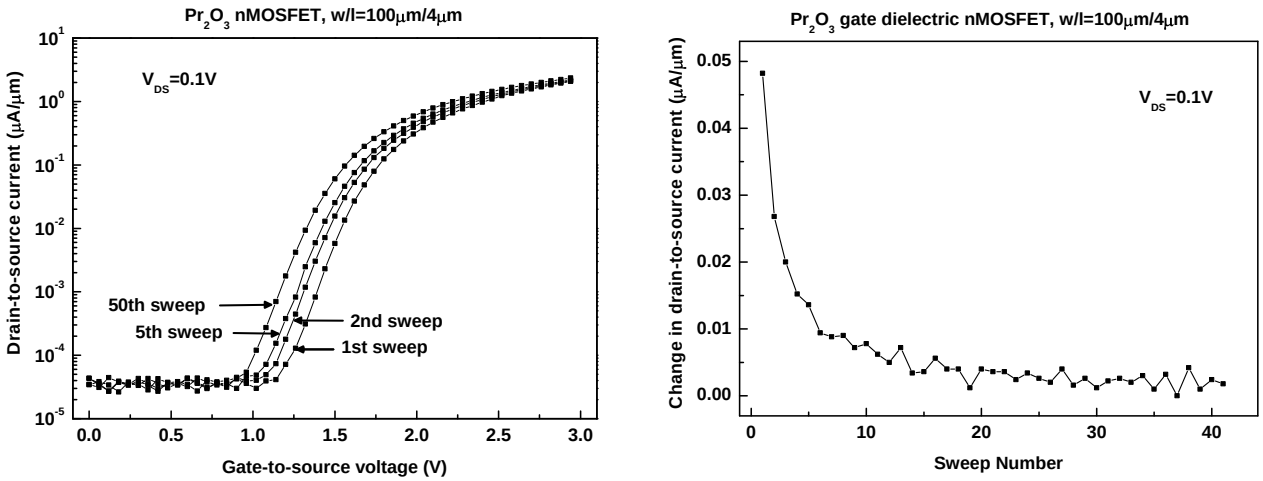
The Pr_2O_3 nMOSFETs are tested electrically. The Pr_2O_3 MOSFETs with $l<10\mu$m are fully functional as seen from their output and transfer characteristics (Fig. 7.2). When the output characteristic is compared to what is expected from simulation using SiO_2 of the same EOT, a reduction in the effective mobility by approximately 60% is observed. Part of this performance degradation is related to a high density of interface charges and oxide charges.



(a) (b)

Fig. 7.2: Graphs of (a) output and (b) transfer characteristics of a Pr_2O_3 gate dielectric gate-first nMOSFET, gate width/ gate length (w/l)=100μm/4μm.

Other evidences for the susceptibility of Pr_2O_3 MOSFETs to charge trapping effects are the increased subthreshold swing (S=145mV/dec) and the threshold voltage shifts to more negative voltages during measurements with increasing sweep number (Fig. 7.3a). The observed negative V_T shift may be due to trapping of positive charges or de-trapping of negative oxide charges. The magnitude of the shift decreases with increasing number of sweeps, as shown in Fig. 7.3b. The total shift is 250mV and corresponds to a trap density of approximately $2*10^{12}cm^{-2}$. Full recovery and even overcompensation is observed after applying negative gate voltages. Charge trapping is found to be partially recoverable with time at room temperature.



(a) (b)

Fig. 7.3: Graphs of (a) consecutive transfer sweeps indicating a threshold voltage shift and (b) drain-to-source current shift as a function of sweep number of a Pr_2O_3 gate dielectric gate-first nMOSFET, w/l=100μm/4μm.

Two important points must be noted. First, the observed device level charge trapping is in good agreement with the nanoscale trapping observed by C-AFM directly on the Pr_2O_3 surface after polysilicon RIE (Chapter 5.4). This fact proves the usefulness of the C-AFM technique in high-K dielectric evaluation. Second, no V_T instability is observed by capacitance-voltage measurements on as-grown capacitors with evaporated Au gates after a short furnace anneal in N_2 [69]. This is an indication of the negative impact of CMOS processing on the high-*K* dielectric quality.

Continuing the investigation of the MOSFET devices, gate leakage measurements on gates of various sizes are performed across the wafers in order to check for the integrity of the Pr_2O_3 gate dielectric after processing. Gate leakages of Pr_2O_3 transistors with gate lengths $l<10\mu$m (area$<10^3\mu m^2$) are well below 10pA (detection limit of the measurement system), corresponding to current density of less than 10^{-6}A/cm^2. Due to a high density of extrinsic defects, however, larger gates ($l>10\mu$m) are often found to be leaky as shown in Fig. 7.4. The weak spots may be due to pure electronic effects or may result from structural effects (e.g. silicide formation) or even both. Once more, the obtained device level results are in agreement with the C-AFM results (Chapter 5.4).

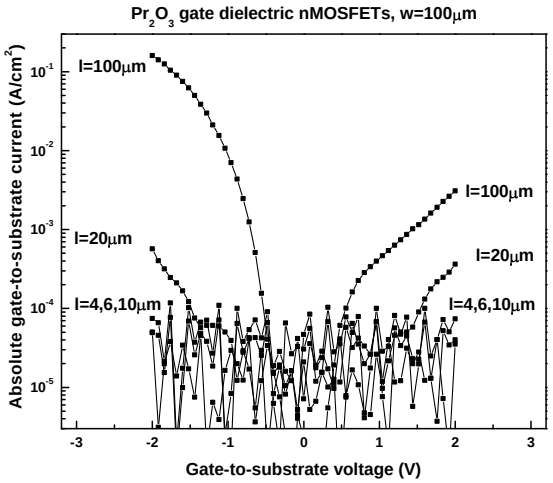
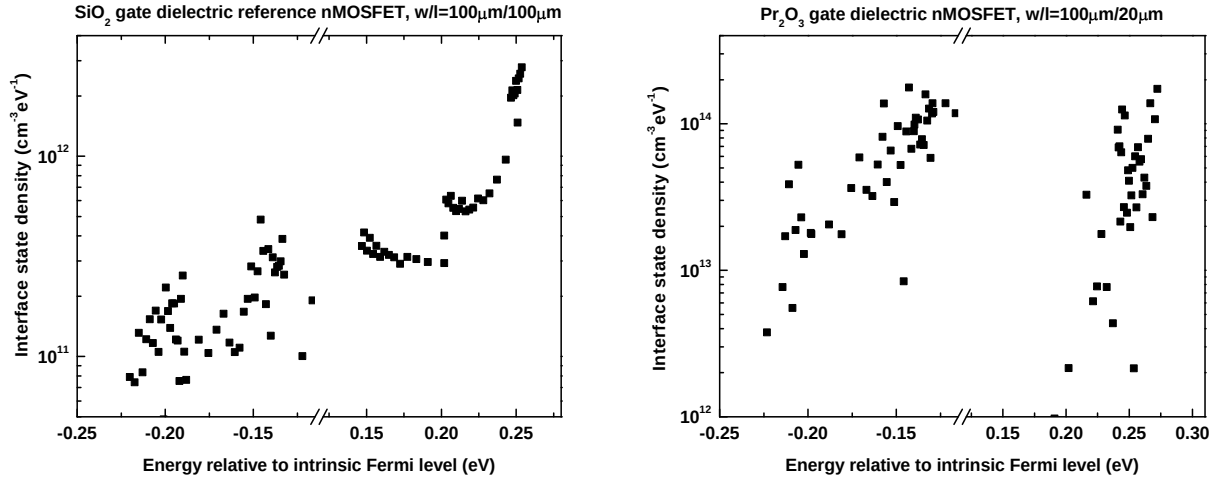


Fig. 7.4: Graph of the size dependence of gate leakage in Pr_2O_3 gate dielectric gate-first nMOSFETs, w=100 μm.

For more detailed investigation of the high amounts of traps present within the high-*K* and its interface with Si, charge pumping measurements are performed on the fabricated Pr_2O_3 gate dielectric nMOSFETs as well as on reference 5nm nitrided SiO_2 gate dielectric nMOSFETs fabricated in the same fabrication scheme later together with the Nd_2O_3 nMOSFETs (Chapter 7.2). The measurement setup is described in Ref. [76].

Initially, the average interface trap densities are measured. Typical interface trap densities are in the order of $6*10^{12}$cm^{-2}eV^{-1} for Pr_2O_3 nMOSFETs and $2*10^{11}$ cm^{-2}eV^{-1} for SiO_2 nMOSFETs.

Next, energy resolved CP measurements of both the SiO_2 and the Pr_2O_3 nMOSFETs are performed. The measurements indicate that most of the interface states at the SiO_2-Si interface are located in the upper half of the bandgap (Fig. 7.5a), consistent with previous studies [20]. The distribution of the traps in the high-*K* devices is different with roughly equal densities in the measurable regions in the upper and the lower part of the bandgap (Fig. 7.5b). It can be speculated that this distribution is either one more indication of the process induced damage or an intrinsic property of this particular interface.

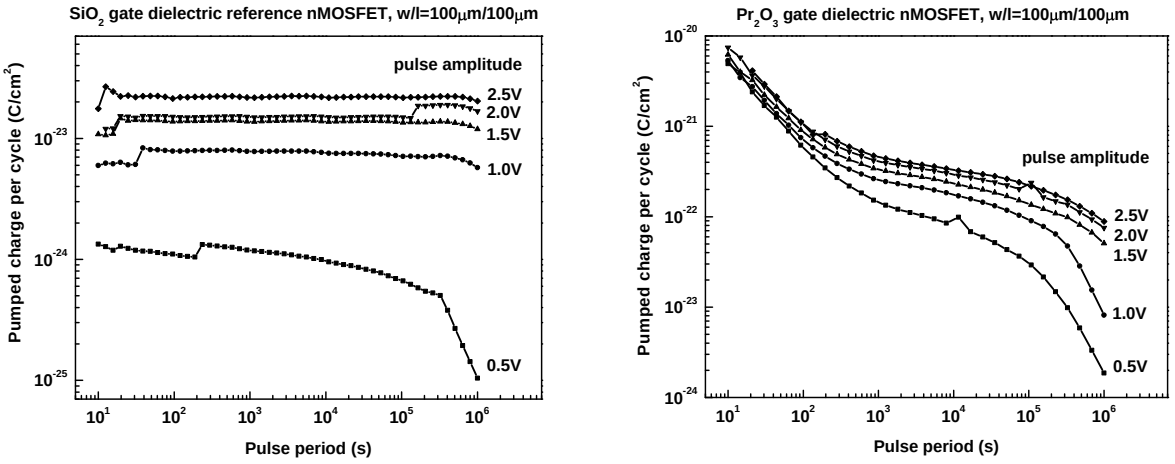


(a) (b)

Fig. 7.5: Graphs of interface state density distribution in (a) a reference SiO_2 gate dielectric nMOSFET, *w*/*l*=100μm/100μm and (b) a Pr_2O_3 gate dielectric nMOSFET, *w*/*l*=100μm/20μm.

Finally, CP evaluation of the distribution of traps across the thickness of the gate dielectric is performed. By applying different gate pulse amplitudes and different pulse frequencies traps at different distances away from the Si-dielectric interface are allowed to change their states. The traps that are situated at the very interface between the substrate and the gate dielectric can respond to very short pulses and low amplitudes by switching their occupancy state. However, the traps situated deeper in the gate dielectric need longer time and more energy to capture or emit an electron and are therefore active only at lower frequencies and higher amplitudes.

The pumped charge per cycle for the SiO_2 devices is relatively independent of the frequency of the gate pulses (Fig. 7.6a) indicating that most of the trapped charge is situated directly at the Si-SiO_2 interface. This is not the case for Pr_2O_3 gate dielectric MOSFETs (Fig. 7.6b). For high frequencies the pumped charge per cycle decreases very rapidly. The amplitude dispersion is also larger. The observed dependencies suggest a much broader distribution of the defects in the depth of the high-*K* material compared to SiO_2.



(a) (b)

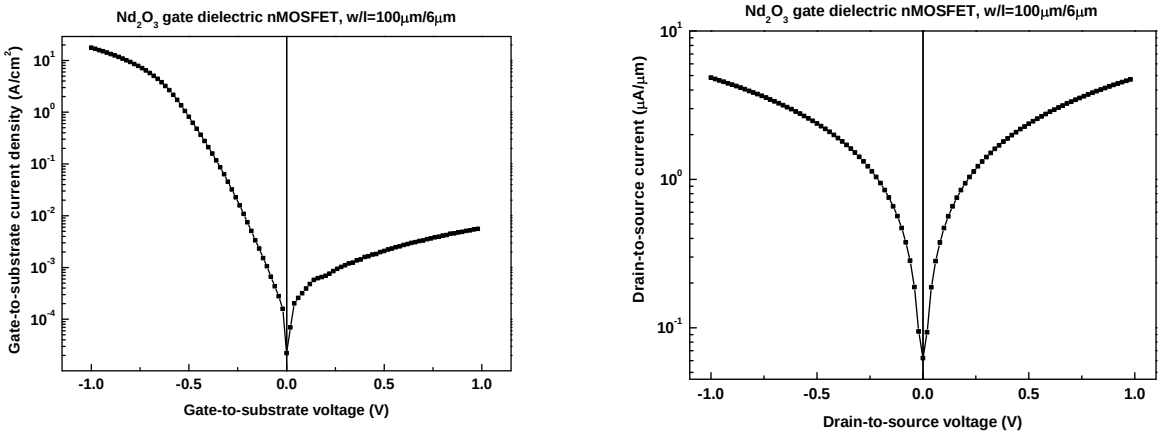
Fig. 7.6: Graphs of pumped charge as a function of pulse period and pulse amplitude as an indication of in-depth interface-state density distribution of (a) a SiO_2 gate dielectric MOSFET, *w*/*l*=100μm/100μm and (b) a Pr_2O_3 gate dielectric nMOSFET, *w*/*l*=100μm/20μm.

7.2. Nd_2O_3 gate dielectric nMOSFETs

4" Si(100) wafers are used for fabrication of the Nd_2O_3 gate dielectric devices. The isolation schemes utilized here are self isolation and STI. STI is fabricated according to the optimized STI processes with oxide etchback and reverse nitride masking described in Chapter 4.3.

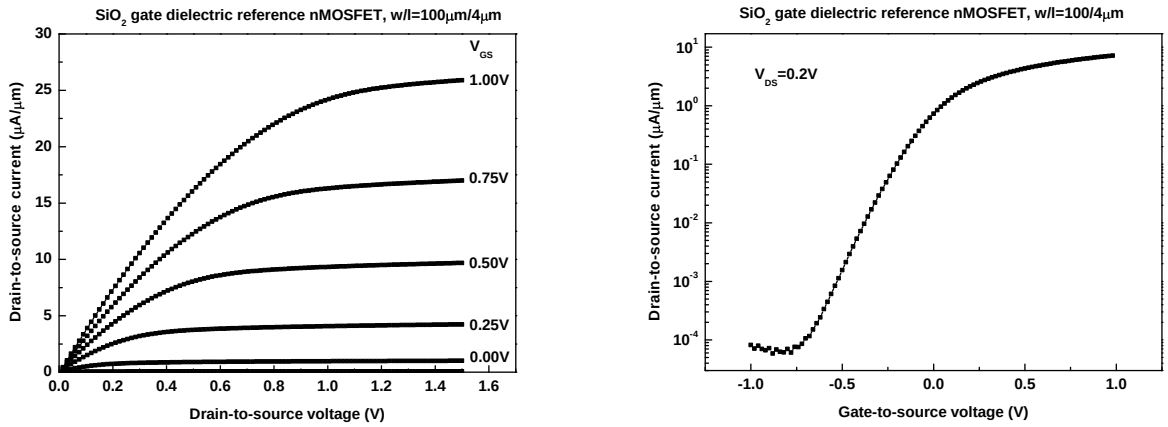
10.5nm Nd_2O_3 layers (EOT =1.5nm) are integrated into the same fabrication scheme as the Pr_2O_3 (Chapter 7.2) [77]. In addition to the wafers with high-*K* gate dielectrics, several wafers with thermally grown 5nm nitrided SiO_2 and 20nm pure SiO_2 are processed in parallel as reference wafers. Gate oxide growth is carried out by RTA at the ISTN. The SiO_2 wafers undergo the same processing steps as the high-*K* wafers.

Unfortunately, the Nd_2O_3 MOSFETs suffer from severe leakages, rendering them non-functional. The Shottky-type gate-to-substrate leakage shown in Fig. 7.7a is most likely the result of a metal-semiconductor junction formed between the metallic gate and the semiconducting substrate through a badly damaged and even possibly non-existent gate dielectric layer. The ohmic leakage from the source to the drain terminal shown in Fig. 7.7b is caused by a shorcircuit between the source and the gate on one side and between the gate and the drain on the other side, again suggesting the demise of the gate dielectric. Similar results are obtained for the wafers with self isolation and for those with STI. On the other hand, the reference SiO_2 gate dielectric MOSFETs are fully functional (Fig. 7.8), eliminating the possibility of fatal errors during processing. Clearly, as C-AFM has already shown, Nd_2O_3 is much less stable than Pr_2O_3 and the aggressive gate-first processing has destroyed the thin dielectrics layers.



(a) (b)

Fig. 7.7: Graphs of (a) gate-to-substrate and (b) drain-to-source current-voltage measurements of a Nd_2O_3 gate dielectric nMOSFET, *w/l*=100μm/6μm.



(a) (b)

Fig. 7.8: Graphs of (a) output and (b) transfer characteristics of a reference SiO_2 gate dielectric nMOSFET, *w*/*l*=100μm/100μm.

7.3. Gd_2O_3 gate dielectric nMOSFETs

Gd_2O_3 gate dielectric nMOSFETs with polysilicon, TiN and SiNi gates are integrated into a gate-first process on SOI wafers by our project partners from AMO, Aachen within the framework of the KrisMOS project (Fig. 7.9). Details on integration can be found in Ref. [75].

The electrical evaluation results indicate functional nMOSFETs with TiN gates (Fig. 7.10). However, the low slope of the output characteristics in the linear region suggests very high source/drain resistance. This resistance is the result of insufficient dopant activation after the 800˚C anneal. TEM investigation (Fig. 7.11) and decreased gate capacitance (Fig. 7.12) both prove that even at this temperature interface layer formation occurs. The high-temperature RTA needed for source/drain activation causes dielectric degradation. In the case of polysilicon and SiNi gate electrodes, temperatures as low as 800°C result in destruction of the gate dielectric and excessive leakages. Rapid thermal anneal at 930°C destroys the Gd_2O_3 layer even in the case of TiN gate electrodes [74].

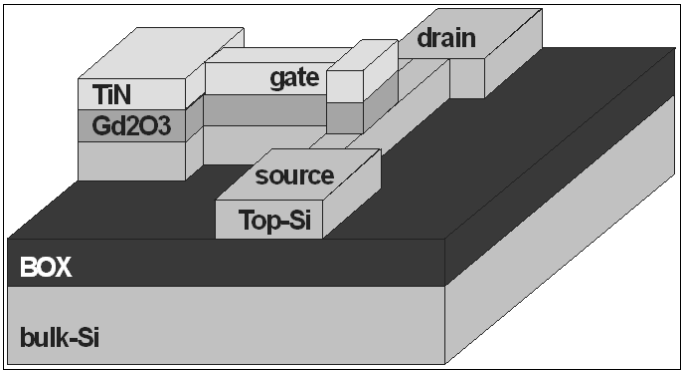


Fig. 7.9: Schematic structure of of a Gd_2O_3 gate dielectric nMOSFET with TiN gate on SOI substrate produced by AMO Aachen.

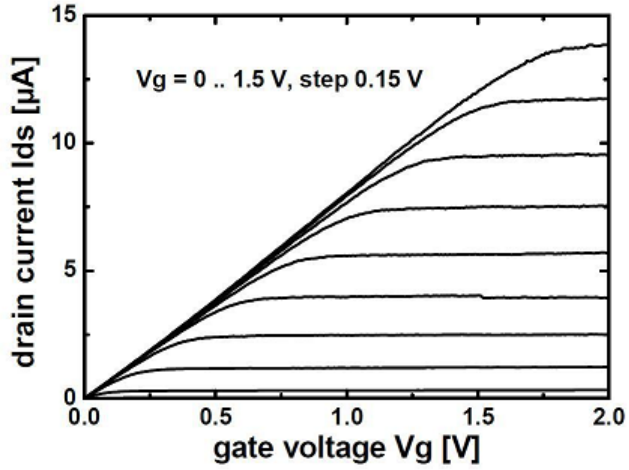
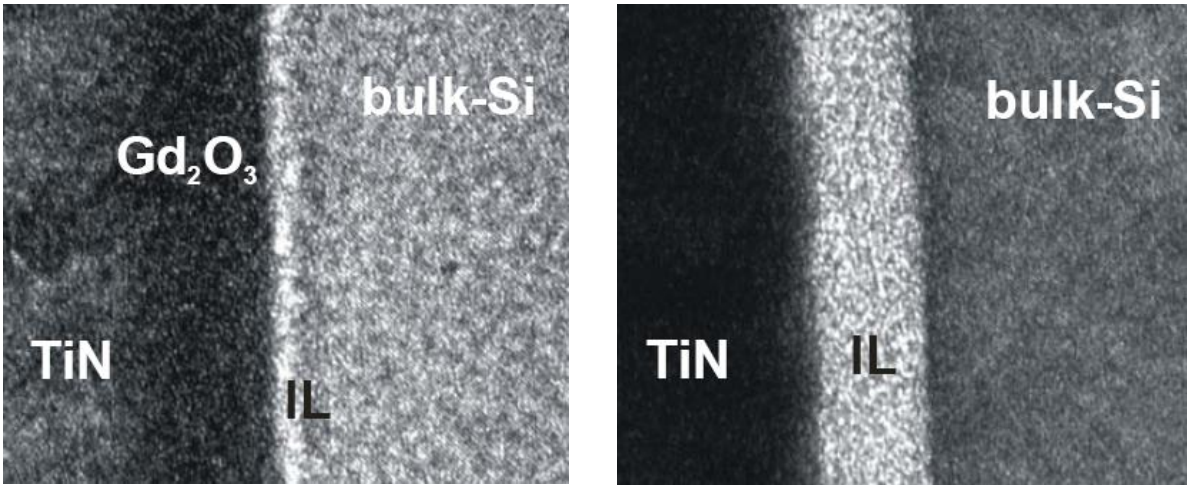


Fig. 7.10: Output characteristics of a Gd_2O_3 gate dielectric nMOSFET with TiN gate, *w*/*l*=20µm/10µm [74].



(a) **(b)**

Fig. 7.11: TEM images of TiN-Gd_2O_3-Si stacks revealing interface-layer formation after a RTA at (a) 800˚C and (b) 930˚C for 10s [75].

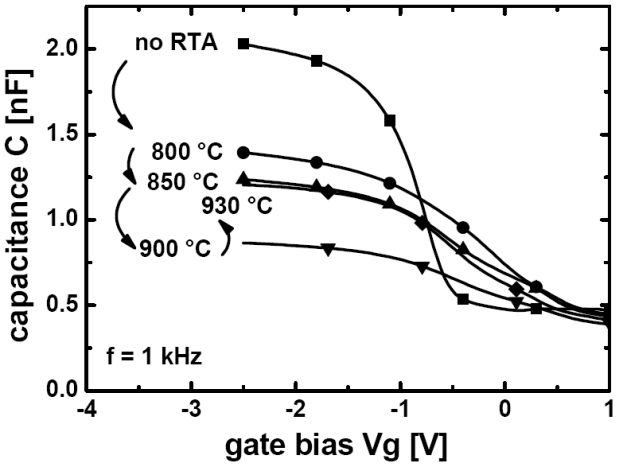


Fig. 7.12: Graph of capacitance-voltage characteristics of Gd_2O_3 gate dielectric pMOS capacitors showing the effects of high-temperature annealing [75].

The same conclusions as in Chapter 7.1 can be drawn from AMO's results: it is possible to integrate rare-earth metal oxides into a gate-first process and functional devices can be produced, but their performance is not optimal.

7.4. Analysis and conceptual solution – damage free integration

MOSFETs with three different epitaxially grown gate dielectrics are produced. Pr_2O_3 devices are functional but suffer from severe charge trapping. Large gate area devices are leaky. Nd_2O_3 devices are not functional due to severe dielectric degradation although the SiO_2 reference devices produced in parallel with them exhibit good device characteristics. The Gd_2O_3 devices are functional but exhibit degraded performance as well.

The results demonstrate that by using optimized conventional gate-first fabrication, functional devices with epitaxial high-*K* gate dielectrics can be realized. However, the characteristics of these devices are not optimal. The modified gate-first process with gate, source and drain silicidation also does not solve the problems.

The high-*K* device characterization yields results that are predicted by the AFM measurements. The leaky gates and the threshold voltage instability of Pr_2O_3 MOSFETs are expected after the microscopic leakages and microscopic charge trapping observed directly on the dielectric surface by C-AFM. C-AFM results actually give more detailed information about the leakages than the macroscopic results. The former indicate that the observed macroscopic gate leakage is a superposition of microscopic currents paths originating from extrinsic defects rather than due to a degraded intrinsic property of the dielectric film as a whole. Considering the current-voltage characteristics it is speculated that mainly structural defects are responsible for the excessive size dependent gate leakage observed in large gate area devices.

AFM and device level characterization demonstrate that at least a significant part of the degraded device characteristics is caused by processing. High-*K* dielectrics are not well compatible with gate-first processing. The three processes: RIE gate formation, implantation and high-temperature annealing are particularly destructive. Since they are key processing steps, they can not be avoided if the integration scheme is kept unchanged. A solution to this problem can be searched for in two major directions – a different material or a different integration process.

Since the investigated rare-earth metals hold a good potential as materials, it is decided to develop an alternative way for their integration. In order to eliminate the effect of the high-temperature step, it is possible to shift the gate stack formation module after the source/drain anneal. To keep the alignment of the source and drain to the gate stack, a place holder or 'dummy' gate stack can be used. With the help of CMP, the dummy gate is later replaced after source and drain formation by the final gate stack without the need for RIE. This is the idea behind the replacement gate technology which is described in more detail in Chapter 8.2. This approach is analyzed, modified to suit the specific requirements of the present task and verified experimentally by fabrication of epitaxial gate dielectric MOSFETs.

8. Towards Replacement Gate Technology

As the analysis of gate-first devices with epitaxial high-*K* gate dielectrics shows, the degradation of the high-*K* is at least in part caused by the incompatibility of the high-*K* dielectric material with the reactive ion etch, ion implantation and high-temperature steps used in the conventional processing scheme. Therefore, the effort is concentrated on development of 'gentle' damage free integration based on the replacement gate technology. The concept of RGT is first described in 1997 for integration of metal gate electrodes on SiO_2-based gate dielectrics [78]. The RGT is modified according to the requirements imposed by the used gate dielectric materials. Additional process steps are introduced and existing steps are modified or omitted. CMP is used not only for oxide patterning but also for metal gate patterning.

The details on the development and optimization of the RGT integration process are a subject of future work. Here, some initial results are presented as a proof of the concept of damage free integration.

8.1. Metal gate MOS capacitor formation with CMP

Before MOSFET RGT integration some simple capacitor structures are fabricated with CMP and electrically evaluated [79]. The batch uses Gd_2O_3 gate dielectric and SiO_2 for reference. With the help of these capacitors several important tests are performed: metal gate formation with CMP, Gd_2O_3 growth on pre-structured wafers and finally Gd_2O_3 stability against aluminum (Al) and tungsten (W) used as metal gate electrodes.

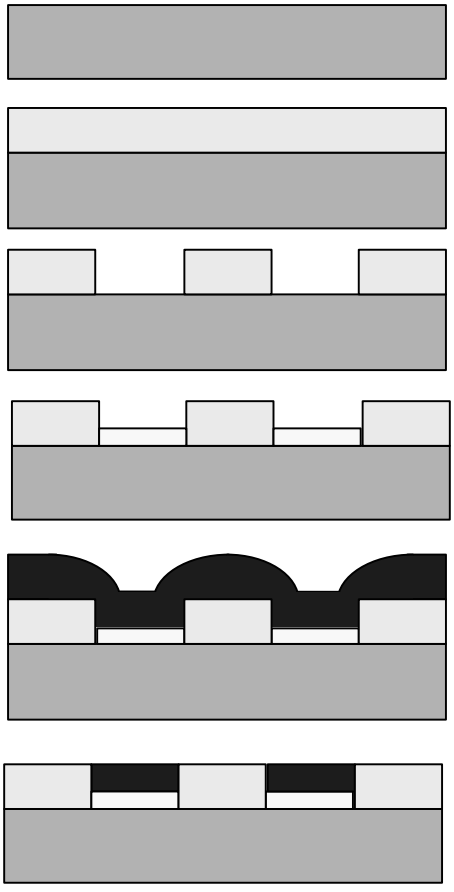


Fig. 8.1: Schematics of capacitor fabrication through CMP.

The major steps of the integration process flow are illustrated in Fig. 8.1. 4" pSi(100) wafers are used. First, a thermal oxide with a thickness of 550nm is grown by dry oxidation. Next, the SiO_2 layer is structured by lithography and RIE with CHF_3/O_2. Next, the wafers are sent to the IEMD, University of Hannover, where 13.5nm Gd_2O_3 is grown on each of them. Two of the wafers are covered in situ with 100nm tungsten for protection, one – with 100nm aluminum and two remain without a capping layer. The wafer without a capping layer is intended for investigation of the effect of exposure to air to electrical properties of the high-*K* gate dielectric. The wafers are then transported back to the ISTN. They are further processed together with several wafers with SiO_2 between 5 and 25nm as reference. The one with Al receives additional 800nm Al and the rest – tungsten-titanium alloy (WTi) for a total metal thickness of 900nm. Finally, the metals are indirectly patterned through CMP, leaving only the metal gates on top of the thin gate oxide. EP-W6300 slurry is used for WTi CMP and EP-A5680 – for Al CMP.

The obtained CMP results are summarized below. Polish rates for both metals are approximately 100nm/min. Selectivity of metal to oxide polish rates are above 200:1. The surface roughness is reduced approximately 10.5 times for Al and 6.6 for WTi respectively (Fig. 8.2). The dishing is very low for Al. It is more pronounced for WTi but still acceptable at device scales (Fig. 8.3) and does not require immediate optimization.

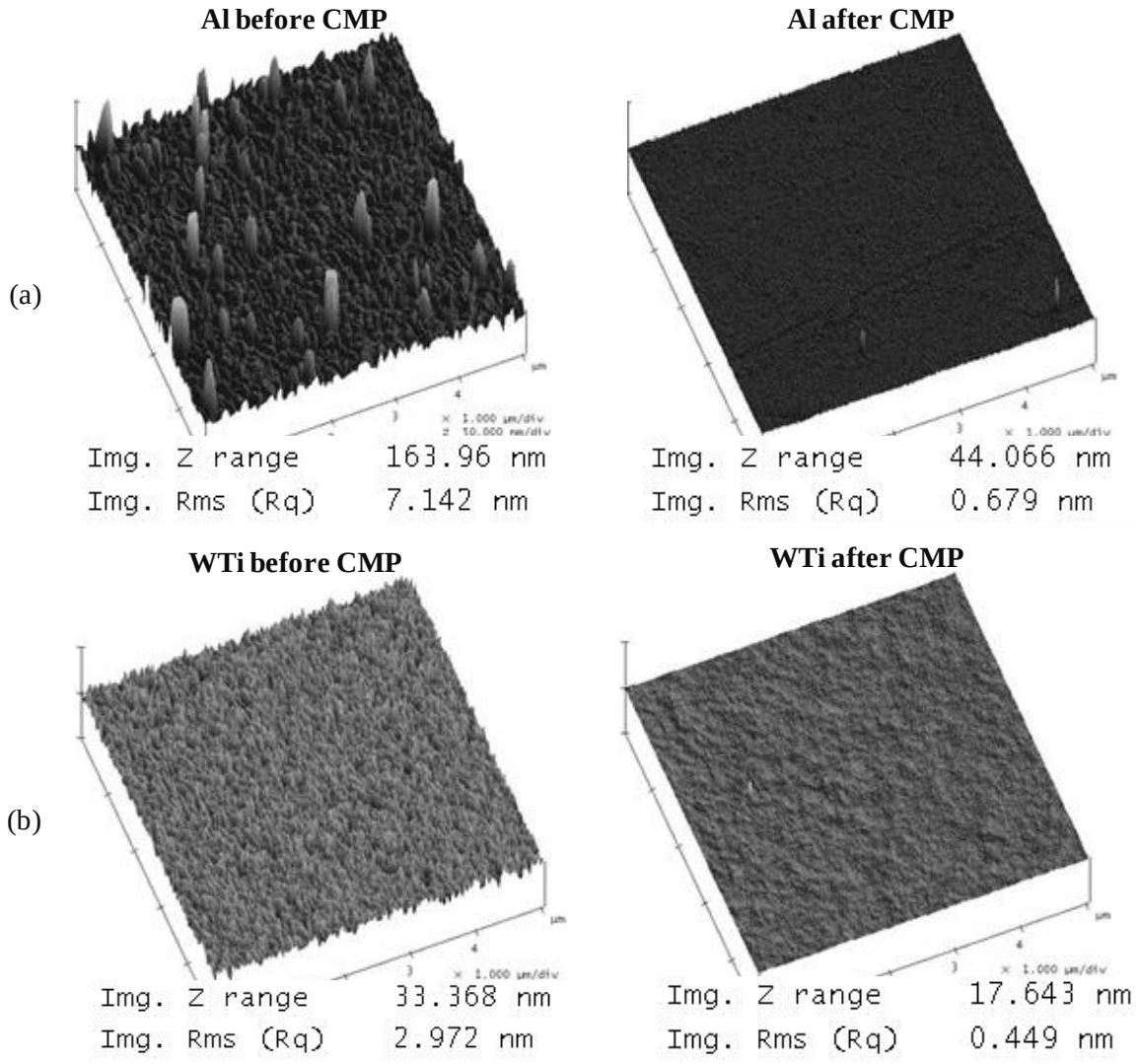


Fig. 8.2: AFM roughness measurements of (a) Al and (b) WTi surfaces before and after CMP.

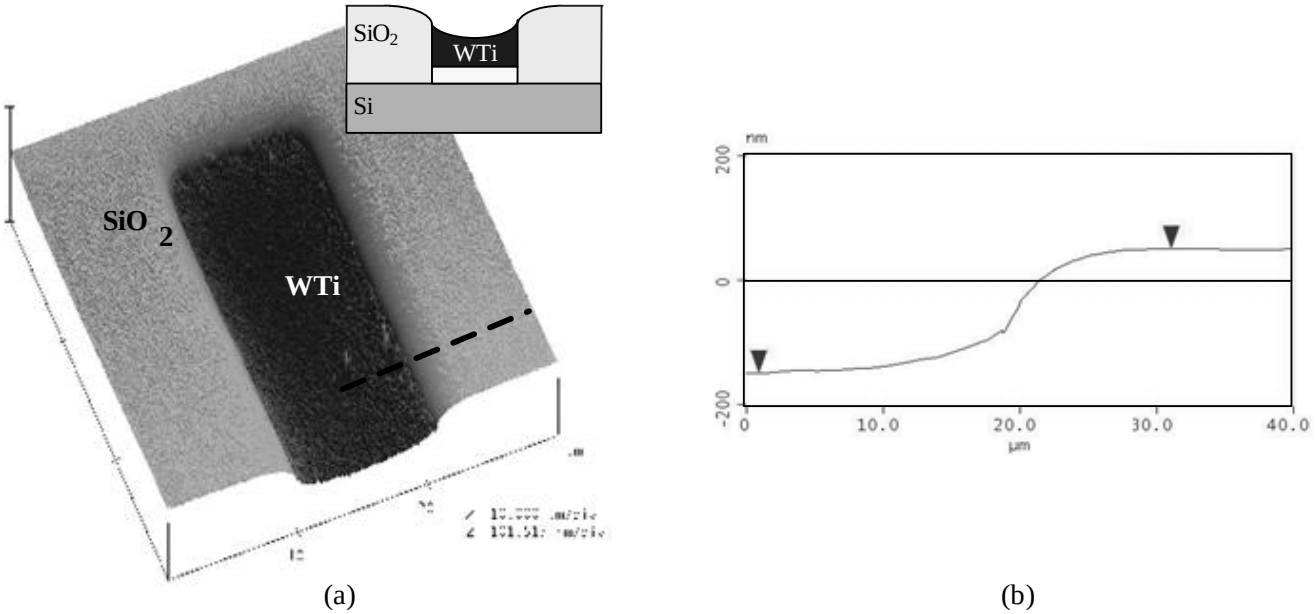
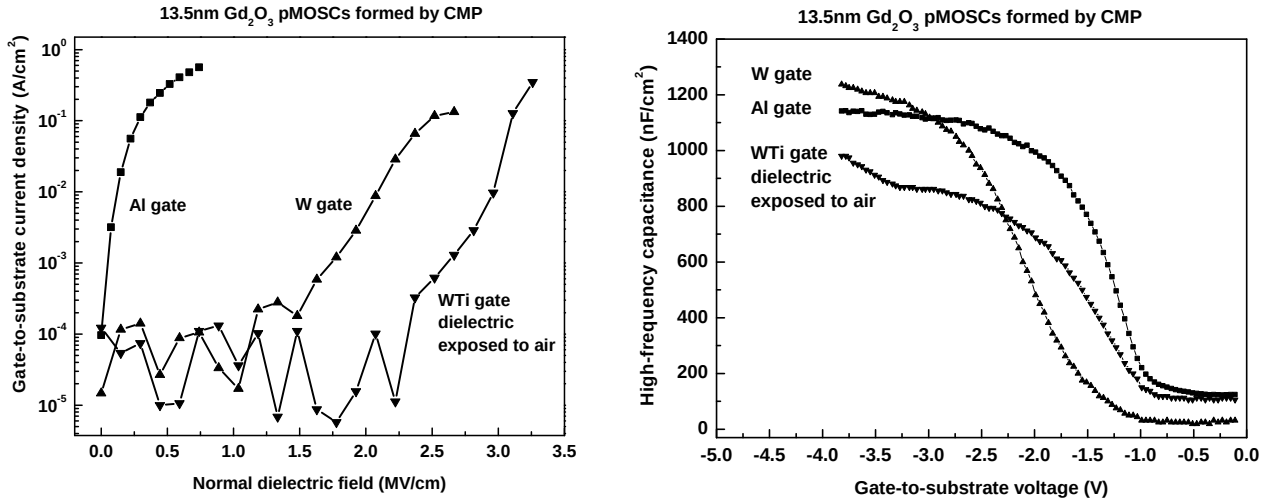


Fig. 8.3: (a) AFM image of a part of a WTi replacement gate pMOS capacitor and schematic cross section, and (b) measured profile of the transition between the oxide and the gate along the dashed line in (a).

The electrical characterization results reveal that Al is not suitable for gate material in combination with a Gd_2O_3 gate dielectric. The gate leakages are extremely high (Fig. 8.4), indicating short-circuited gate and substrate. The most probable reason for this result is a chemical reaction between Al and the gate dielectric (Eq. 8.1). This results in free Gd atoms that enhance the leakage currents.

$$Gd_2O_3 + Al \rightarrow Al_2O_3 + Gd \tag{8.1}$$

The dielectric of W and WTi gate capacitors is intact. WTi gate capacitors exhibit lower leakages than the W ones indicating interfacial layer formation in the period between dielectric growth and metal deposition because of the lack of a capping layer. This is also proved by the capacitance-voltage characteristics (Fig. 8.4). The WTi gate capacitors exhibit lower dielectric capacitance as a result of the dual dielectric layer and higher defect densities (shifted *C-V* curve and lower slope) than the W gate capacitors. The dielectric constant extracted from the *C-V* measurements is $K=10.4$ corresponding to EOT=5.1 nm.



(a) (b)

Fig. 8.4: Graphs of (a) gate current-voltage and (b) capacitance-voltage characteristics of Gd_2O_3 gate dielectric pMOS capacitors with different metal gate electrodes formed by CMP.

These preliminary tests indicate that W and WTi are suitable for gate materials from a fabrication point of view. Therefore, these two materials are selected for integration of the Gd_2O_3 gate dielectric RGT MOSFETs.

8.2. RGT nMOSFETs with Gd_2O_3 gate dielectric and Ti gate electrodes

The replacement gate process shifts the gate dielectric growth and the gate metal deposition after the source/drain implantation and the high temperature anneal. Additionally, it relies on CMP for indirect patterning thus eliminating the dangerous RIE for gate formation. Since these three processing steps are identified as particularly damaging to high-*K* gate dielectrics, replacement gate technology provides the necessary conditions for gentle integration and improved high-*K* device performance.

For simplicity, only the fabrication steps for creating nMOSFETs are shown.

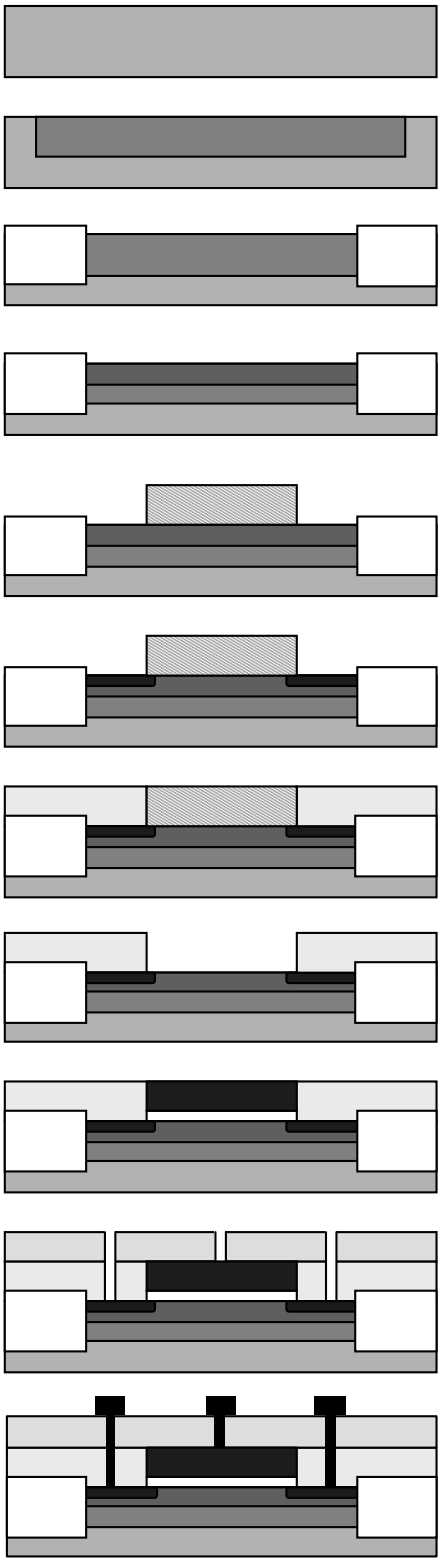


Fig. 8.5: Schematics of RGT fabrication modules.

The process starts with blanket wafers.

1. P-well formation
Similarly to the conventional process, the replacement gate process starts with p-well formation.

2. STI formation
The module is a multi-step process that is described in Chapter 4.

3. Threshold voltage adjustment
The threshold voltage adjustment implantation is performed.

4. Dummy gate formation
The dummy gate stack is deposited and structured with RIE. Commonly, multilayer stacks are used. Examples are oxide-polysilicon-nitride and nitride-polysilicon-nitride.

5. Source/drain formation
Source and drain are formed in the conventional way by ion implantation and a high-temperature anneal for dopant activation.

6. Negative gate-stack oxide layer formation
Oxide deposition, anneal and indirect patterning with CMP are carried out. The formed layer is the negative imprint of the gate stack layer that 'memorizes' the position of the gate stack.

7. Dummy gate removal
Dummy gate is removed by a wet etch in order to avoid damage to the substrate.

8. Gate stack formation
The gate dielectric is grown or deposited, then the gate material is deposited and the gates are formed by indirect patterning with CMP.

9. ILD deposition and structuring
The ILD oxide is deposited and structured to form the contact holes. Here RIE is used but damage to the high-*K* dielectric is minimized since the latter is covered completely by the metal gate and the ILD oxide.

10. Interconnects/contacts formation
The contact-metal is deposited and structured.

11. Passivation
A final forming gas anneal and nitride passivation protect the devices from contamination.

The benefits of this method of fabrication have already been outlined. They come at the expense of increased process complexity compared to conventional gate-first processing.

Although STI is developed and optimized, the module is omitted in this first integration attempt in order to simplify and speed up processing. Furthermore, this allows a more direct comparison between gate-first and RGT devices produced at the ISTN, since the functional gate-first devices are also fabricated with self isolation.

The integration tests performed on Gd_2O_3 reveal similar results to those on Pr_2O_3 and Nd_2O_3. Therefore, the wet chemical, cleaning and resist removal steps are not modified. The major process steps of the Gd_2O_3 RGT nMOSFET fabrication are described below [79].

4" Si(100) wafers with resistivity of 10Ωcm are used. The wafers are oxidized in a furnace at 900°C for 75min in order to form a 20nm thick scattering SiO_2 layer. Boron is implanted through the SiO_2 for threshold voltage adjustment in two steps. First step is with a dose of $2*10^{13}cm^{-2}$ at 45keV and the second – with the same dose at 180keV. An ICECREM simulation of the dopant distribution yields peak concentration of $9*10^{17}cm^{-2}$ at a depth of 170nm beneath the silicon surface after all high-temperature steps. Threshold voltages for oxides in the 5-20nm range and an aluminum gate are close to 0V. For tungsten, whose work-function is higher, V_T is expected to be around 0.5V.

Next, dummy gate stacks are deposited and structured. A part of the wafers receives thermal SiO_2 - polysilicon - PECVD Si_3N_4 stacks and the rest – APCVD Si_3N_4 – polysilicon – PECVD Si_3N_4 stacks. The layers are deposited with different thicknesses – SiO_2 50-70nm, APCVD Si_3N_4 – 150-200nm, polysilicon 100-300nm, PECVD Si_3N_4 150-300nm. Gate lithography is performed. Si_3N_4 is etched with CHF_3-O_2 RIE and polysilicon – with SF_6-Cl_2 RIE. SiO_2 is kept as a scattering oxide for the following source/drain implantation.

The wafers with a Si_3N_4-polysilicon-Si_3N_4 dummy stack receive an 80-100nm thick scattering PECVD SiO_2. Source/drain lithography is performed and phosphorus implantation with a dose of $10^{15}cm^{-2}$ at 90keV forms the source and drain regions. Simulation shows a peak concentration of $9*10^{19}cm^{-3}$ and p-n junction depth of 200nm.

The negative gate imprint oxide is formed next. PECVD oxide is deposited and annealed at 900°C for 30min for dopant activation and implantation damage cure. Additionally, the anneal serves to increase the CMP resistance of the PECVD Si_3N_4. Tests indicate that this increase is higher than the resistance increase of the PECVD SiO_2. As a result, the selectivity of the following polishing process increases.

CMP follows. Polish control is carried out by means of AFM and elipsometry. The CMP+control loop is repeated in a multistep process until the PECVD SiO_2 on top of the dummy gates is completely removed. CMP starts with low selectivity slurry (Levasil CK-30-862 by H. C. Starck) and continues with high selectivity slurry (HS-STI-D11 by H. C. Starck) once the dummy gates are close beneath the SiO_2 surface. The polish times are in the range 15-30min and the polish parameters are equal or very close to 6psi downforce, 50rpm table rotation speed, 40rpm chuck rotations peed, 5psi back pressure, 2psi ring force and 125mL slurry flow rate.

Now that the negative imprint of the dummies is in place, the dummy gates are removed. 85% phosphoric acid (H_3PO_4) at 155°C is used for stripping the Si_3N_4 layers, 50% HNO_3 / 1% HF - for the polysilicon layers and 12.5% HF / 85% ammonium fluoride (NH_4F) - for the SiO_2 layers. AFM and elipsometry measurements after dummy gate removal indicate that the negative gate imprint SiO_2 layer is approximately 300nm thick.

The final gate stack is formed next. The wafers planned as reference wafers receive a 5nm RTA nitrided SiO_2 or a 25-31nm RTA SiO_2 gate dielectric. A 900nm thick layer of W-Ti is then sputtered on top. The rest of the wafers are transported to the University of Hannover for molecular beam epitaxy of Gd_2O_3. 13.5nm layers are grown and then covered in situ by a 100nm W layer in order to protect the

high-K material from the ambient atmosphere. Details of Gd_2O_3 MBE growth are published previously [50]. The wafers are then transported back to the ISTN and additional 800nm of W-Ti are deposited.

The W and WTi layers are indirectly patterned with CMP using an AP-W6300 slurry from H. C. Starck. The polish parameters are the same as those of the previous CMP step. The polish times are in the range 6-11min.

500nm of PECVD SiO_2 is deposited as interlayer dielectric. A third lithographic step is performed and contact holes are etched into the oxide using CHF_3, O_2 and Ar reactive ion etch in a multistep process.

Metal contacts are formed by depositing and structuring 150nm of Ti as a diffusion barrier and 800nm of Al. After the fourth and final lithography and the following $SiCl_4$ RIE of Al, a forming gas anneal at 425°C for 15min completes the fabrication process.

AFM and C-AFM are used at several points during fabrication for process control. For example, C-AFM is used for end-point detection after contact hole etch (Fig. 8.6).

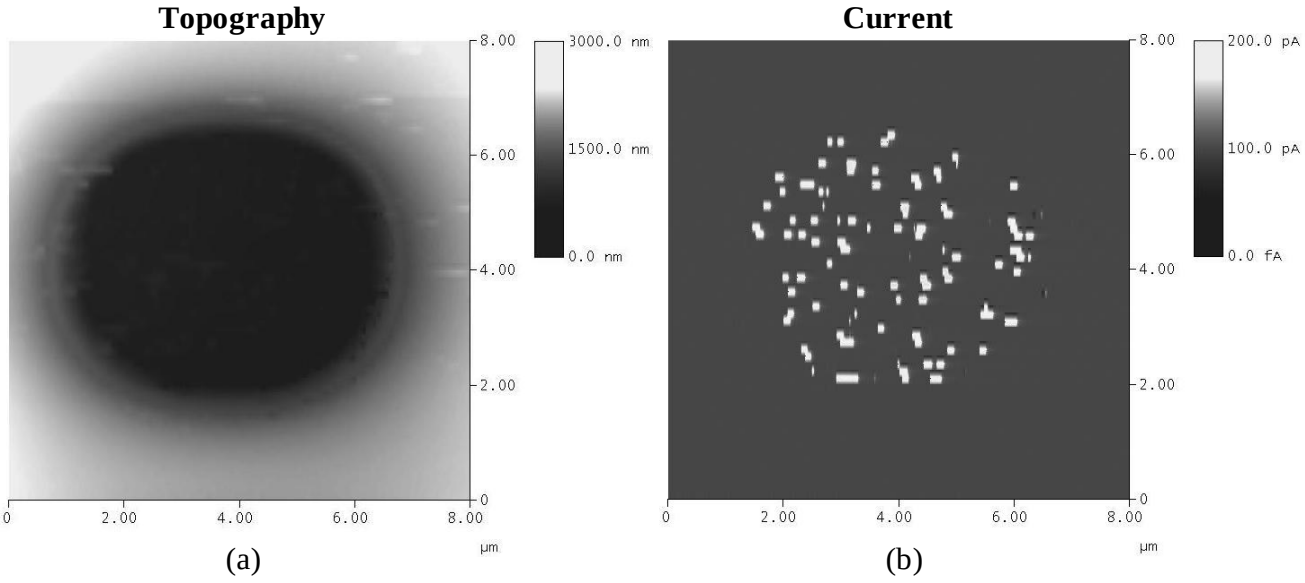
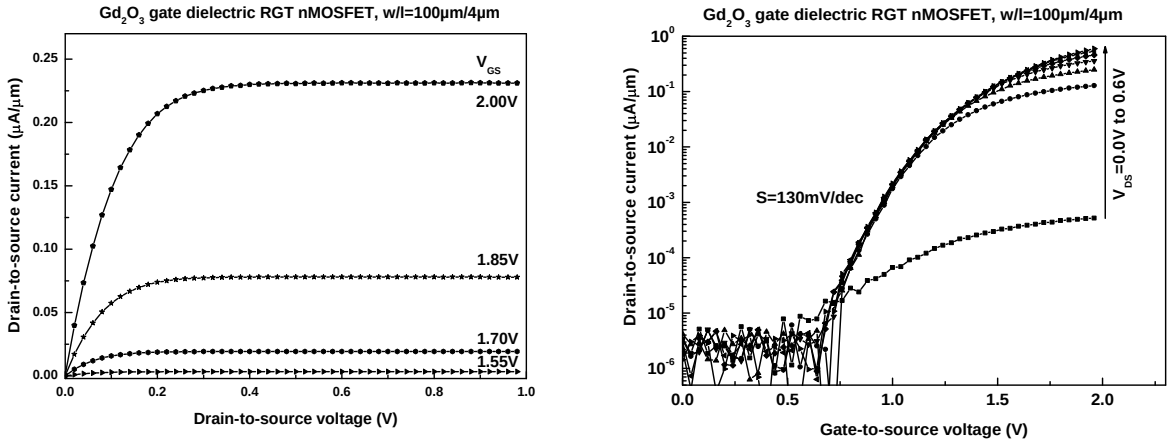


Fig. 8.6: (a) Topography and (b) electric current AFM images obtained simultaneously across an etched hole in the interlayer dielectric and the Pr_2O_3 layer. The high current areas are the areas where the conductive AFM tip is in electrical contact with the silicon substrate. The contact hole check must be carried out prior to resist removal so that the etch can continue if needed. The non-uniform currents in the etched contact hole are a result of contamination of the conductive tip with the non-conductive photoresist.

The fabricated devices with Gd_2O_3 gate dielectric devices are fully functional [79]. Leakages are below 10^{-3}A/cm^2 at V_{GS}=1V, fulfilling leakage requirements set by the ITRS. The Gd_2O_3 gate dielectric nMOSFETs show proper transistor behavior (Fig. 8.7) with $I_{on}=5*10^{-4}$mA/μm and $I_{off}<5*10^{-9}$mA/μm. The threshold voltage of 1.2V is higher by 1V than the one calculated by device simulations and can be attributed to a high density of fixed charge in the order of 10^{13}cm^{-2}.



(a) (b)

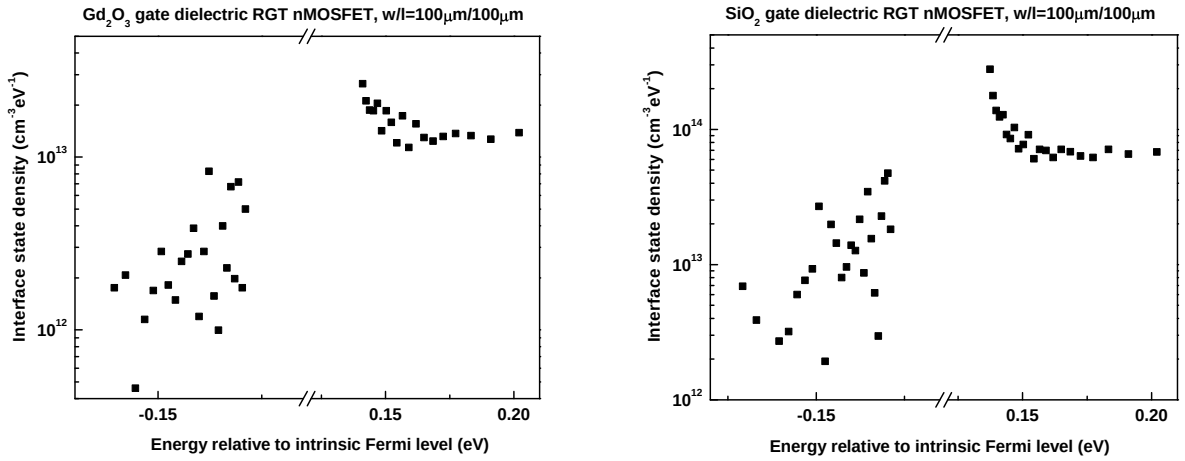
Fig. 8.7: Graphs of (a) output and (b) transfer characteristics of a 13.5nm Gd_2O_3 gate dielectric replacement gate technology nMOSFET with a tungsten gate electrode, *w*/*l*=100µm/4µm.

Especially important is the extremely low hysteresis of less than 30mV observed in the subthreshold characteristics. This is a substantial improvement when compared to nMOSFETs with conventionally integrated high-*K* dielectrics [80]. In the case of the gate-first devices with Pr_2O_3 gate dielectric, large hysteresis effects with threshold voltage shifts of more than 250mV are observed (Chapter 7.1).

The subthreshold swing of approximately 130mV/dec indicates high interface states densities. Effective mobility of $130cm^2/Vs$ is measured for the Gd_2O_3 MOSFETs. Compared to SiO_2 references this corresponds to a reduction of approximately 40% at the same effective electric field. It is suspected that the acceptor-type interface states significantly degrade mobility due to Coulomb scattering.

Charge pumping (CP) measurements reveal average trap densities of $2*10^{13}eV^{-1}cm^{-2}$, which are consistent with the degraded subthreshold swing. However, high values of $2*10^{12}eV^{-1}cm^{-2}$ are obtained for the SiO_2 reference devices as well which puts in question the effectiveness of the forming gas anneal when using tungsten gate electrodes. This means that if a different gate material is used and the forming gas anneal is successful, the trap density for the high-K devices might be significantly reduced.

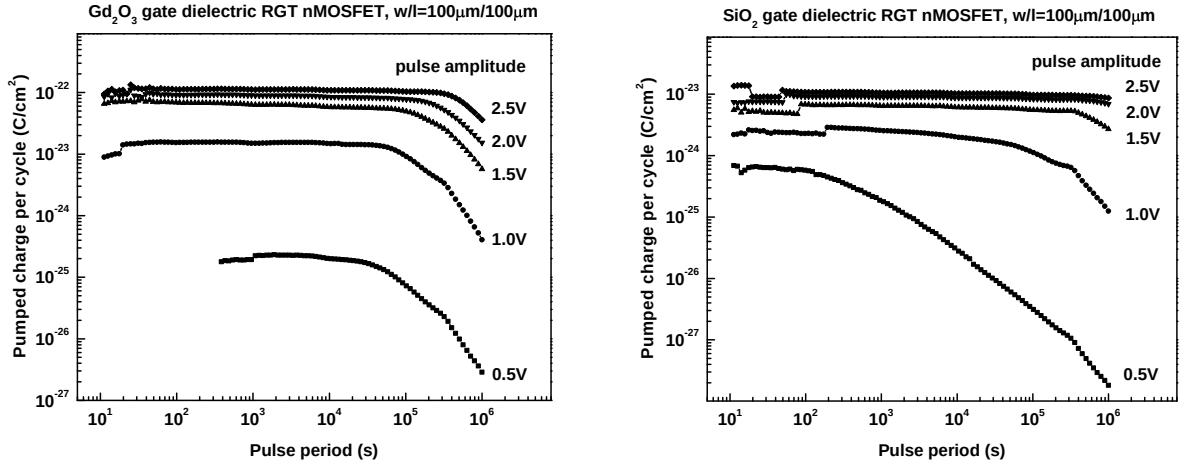
The energy resolved measurements indicate dominance of acceptor states (Fig. 8.8) for both gate dielectrics. As a matter of fact, the distribution of traps in the measurable regions of the bandgap looks very similar for SiO_2 and Gd_2O_3 gate dielectric nMOSFETs. Considering the CP results obtained on gate-first devices, for which the trap distributions of high-*K* nMOSFETs and the reference SiO_2 nMOSFETs differ significantly, one can conclude that the change in integration technology brings the high-*K* devices closer to SiO_2 devices in terms of quality.



(a) (b)

Fig. 8.8: Interface state density distribution of (a) Gd_2O_3 and (b) SiO_2 gate dielectric replacement gate technology nMOSFETs, *w*/*l*=100μm/100μm.

Further evidence of this is found in the depth distribution of interface states. Gd_2O_3 devices actually show a slower decline in the pumped charge per cycle for low pulse amplitudes than the reference devices (Fig. 8.9). This suggests a faster decline in trap density with distance immediately after the interface for the high-*K* devices.



(a) (b)

Fig. 8.9: Graphs of pumped charge as a function of pulse period and pulse amplitude of (a) Gd_2O_3 and (b) SiO_2 gate dielectric replacement gate technology nMOSFETs, *w*/*l*=100μm/100μm.

To summarize, functional replacement gate pMOS and nMOSFET devices with Gd_2O_3 gate dielectric are produced and evaluated. They show an improvement in characteristics compared to the high-*K* devices manufactured using a conventional gate-first process. However, both high-*K* and reference SiO_2 devices indicate large amounts of fixed and interface states. It is expected that they are a result of unsuccessful forming gas anneal in the case of W gate material. A proper gate material must be selected in order for the forming gas anneal to be successful or the process flow must be modified.

9. Summary and outlook

9.1. Summary

The work contributes to state-of-the-art CMOS technology in two aspects. First, it presents the extensive application of Atomic force microscopy in semiconductor technology. Critical high-resolution measurements are performed by AFM which assist in the selection, integration and process optimization of a suitable high-*K* gate dielectric material to replace SiO_2 in advanced CMOS ICs. Development of the STI device isolation module is carried out exclusively under AFM control. A novel technique for reliable nitride stop-layer erosion detection with the help of C-AFM is implemented for the first time. Structural and electrical evaluation of potential high-*K* dielectric candidates is performed directly on material level i.e. without having to produce full MOS devices. Different types of carrier transport mechanisms in high-*K* materials are distinguished by C-AFM, which proves to be a very valuable characterization technique in particular for pilot processes when there are not enough devices for statistical device evaluation. Charge trapping on the nanoscale is observed for the first time. It is in agreement with device characteristics measured later on fabricated devices, demonstrating the ability of C-AFM to help in predicting macroscopic device behavior by measurements on simple non-structured dielectric films, thus opening the road for accelerated and simplified high-*K* material investigation. What is more, AFM supports the investigation of the compatibility of high-*K* dielectrics with critical CMOS processing steps. The manufacturing process flow is modified on the basis of the AFM results. When an alternative process flow is implemented, namely the replacement gate technology, AFM is used again to support its development and implementation.

Second, this work is a demonstration of the potential of epitaxial rare-earth metal oxides as gate dielectrics in CMOS technology. The first integration attempts with Pr_2O_3 and gate-first processing prove that ultralow leakage MOSFET devices with epitaxial oxides can be realized. However, their performance is not optimal. There is a high density of defects at the Si-high-*K* interface and within the volume of the high-*K* films. These defects lead to threshold voltage instability, increased subthreshold swing, enhanced gate leakage for large area devices and reduced carrier mobility. To a large extent, this degradation of device characteristics is attributed to process-induced damage. Clearly, the high-*K* materials are more susceptible than SiO_2 to damage coming from several particular processing steps. The three most critical processing steps are identified as the gate structuring with reactive ion etch, the source/drain ion implantation and the source/drain high-temperature activation anneal. Since these processing steps are inseparable from gate-first processing, a simple modification of the standard gate-first processing is not sufficient to guarantee preservation of optimal gate dielectric quality throughout the manufacturing process. Therefore, a novel integration scheme is considered for epitaxial oxide integration using a gate-last approach. Replacement gate technology relies on indirect patterning with chemical mechanical planarization for gate stack formation, thus eliminating RIE. The source/drain implantation and the consecutive activation anneal are performed before gate dielectric growth, while still keeping the source and drain alignment through the use of a dummy gate. In this way, the gate stack is protected from all potentially damaging process steps. With the very first integration attempt of epitaxial oxide – Gd_2O_3 with RGT functional devices are produced and their characteristics are indeed better when compared to those of Pr_2O_3 gate-first devices. Threshold voltage instability is significantly reduced. The density of interface states and their energetic distribution is brought to the levels of the SiO_2 gate dielectric reference devices.

9.2. Outlook

A couple of points should be addressed in order to improve the capabilities of the C-AFM technique. First, optimization of the electric current module pre-amplifier is needed. It should have a larger range of measurable currents in order to allow evaluation and identification of leakage mechanisms through dielectric films more accurately. Second, the contact area between the tip and the surface depends on the slope of the surface. The contact area is proportional to the currents. Therefore, the measured currents are not directly proportional to the local resistivity at the point of contact. A software correction can allow more accurate resistivity mapping of the sample surface.

With respect to the high-*K* dielectric integration, development and optimization of the replacement gate technology with epitaxial gate dielectrics is required in order to improve device performance. The performance of RGT epitaxial oxide devices is inferior to that of state-of-the-art SiO_2 devices. However, it must be kept in mind that SiO_2 has a four-decade optimization history behind its back. Epitaxial high-*K* gate dielectrics show a potential and maybe one of them might be the new SiO_2 a decade from now.

Appendix – Challenges for AFM

This thesis demonstrates clearly the applications of AFM in semiconductor technology. It also demonstrates several advantages of AFM over other measurement techniques. However, some considerations must be kept in mind in order to make the most out of this technique. Otherwise, it is easy to be misled when interpreting AFM images and come to erroneous conclusions.

The AFM tip is the sensory element through which the AFM system gathers information about the surface properties. Therefore, scan parameters must be carefully adjusted in order to keep the interaction between the tip and the sample constant and within the region of non-destructive interaction. Otherwise, either the tip shape or the sample surface or both can be adversely affected. Image analysis is also critical. AFM images are only a representation of the sample surface, limited by all consequences resulting from tip shape and the nature of tip movement over the sample surface.

Below are several examples of how AFM images that can be misinterpreted due to different tip related reasons.

Example 1 – tip degradation

The AFM images in Fig. 0.1 illustrate the deterioration of the tip within a single scan. The currents become smaller as the tip scans from the top down. One might come to the false conclusion that the conductivity of the silicon area decreases in the lower part of the scanned area. However, in reality it is the tip apex conductivity that changes due erosion of the conductive coating as a result of too high tip-sample interaction forces. A second scan of the same surface proves that. The currents are smaller than the ones during the first scan.

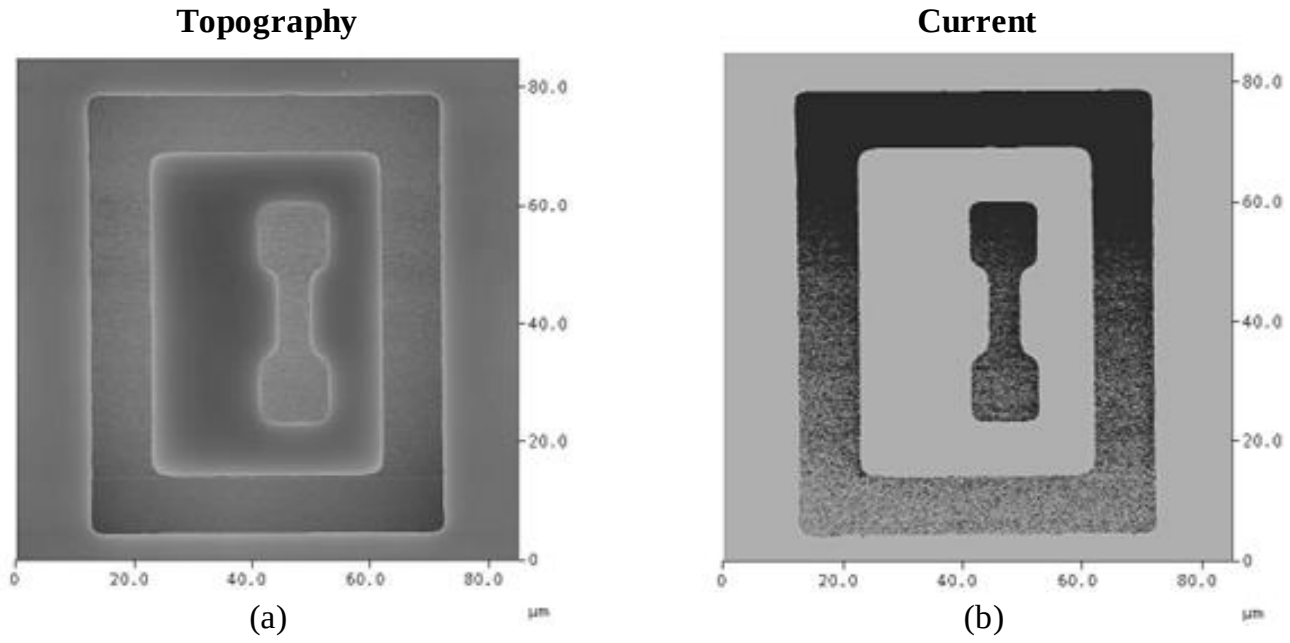


Fig. 0.1: (a) Topography and (b) electric current images of the surface of a partially processed wafer. The rectangular and the dumbel areas are silicon while the rest of the area is a thick SiO_2 film.

Example 2 – hidden features because of damaged tip

Fig. 0.3 shows AFM images of a partially processed wafer surface in an area with nitride erosion (see Chapter 4.2). The electric current image indicates erosion only on the left and lower sides along the periphery. However, after the sample is rotated at 180˚ and scanned again, the electric current image again shows erosion on the left and down (Fig. 0.4). This is because erosion is actually present all along the square structure periphery, but the worn out conductive coating in the left and lower sectors of the tip apex prevents C-AFM from detecting the erosion in the right and upper side of the square periphery where the slope of the eroded area is unfavorable.

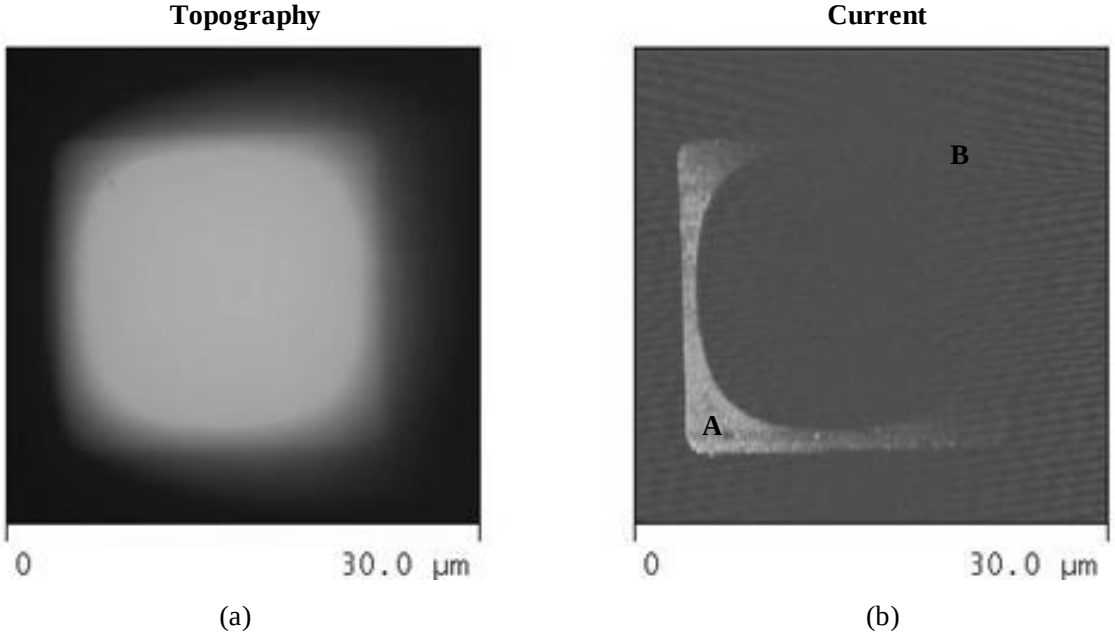


Fig. 0.2: (a) Topography and (b) electric current AFM images of a partially processed wafer with nitride erosion.

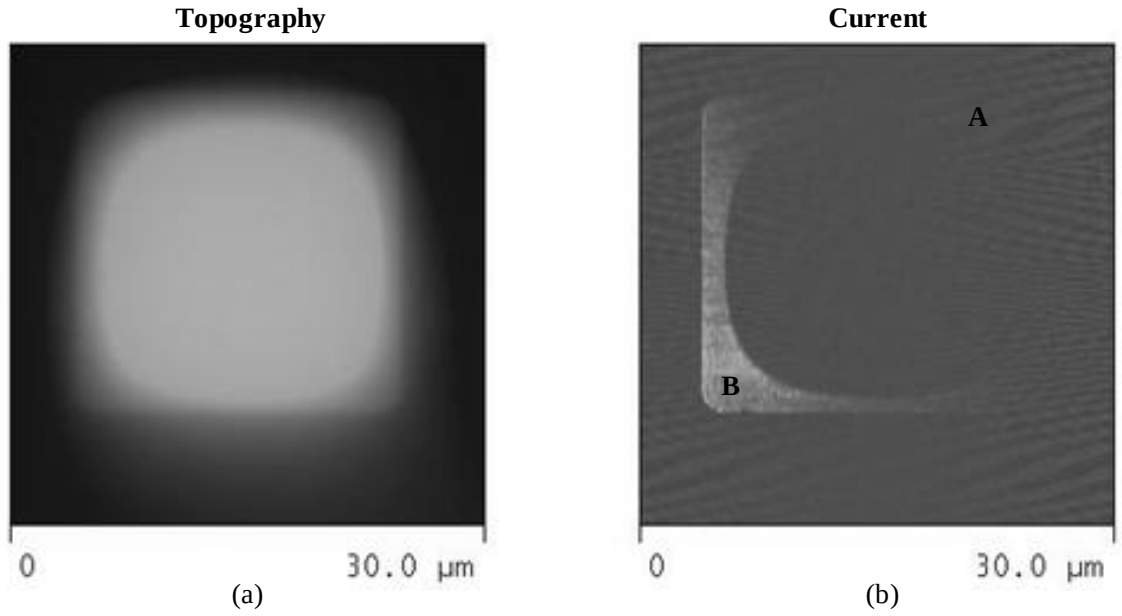
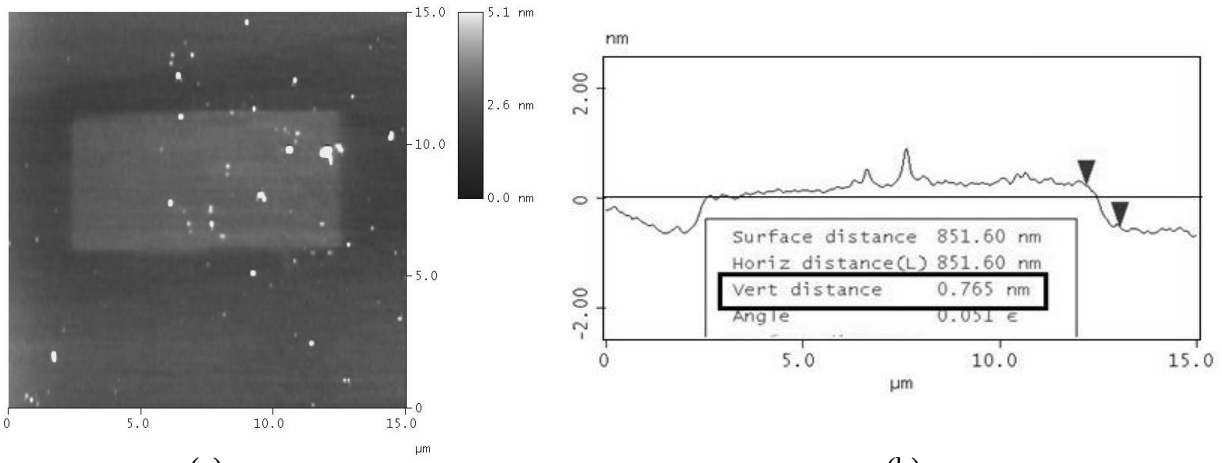


Fig. 0.3: (a) Topography and (b) electric current AFM images of a partially processed wafer with nitride erosion. The same area as the one in Fig. 0.3 is scanned after rotating the wafer 180˚. There is an apparent contradiction in the two electric current images.

Example 3 – electrically induced oxide growth

High temperatures can affect the measured sample. In the case of SiO_2 they result in local oxidation of the Si substrate below the point of contact i.e. electrically stimulated oxide thickening. Fig. 0.5 illustrates this phenomenon. Several scans are performed in a rectangular area with a high tip voltage of -10V. A larger area scan at zero bias (Fig. 0.5a) reveals that the area scanned at -10V has been oxidized. The oxidized area is elevated by 0.77nm above the baseline (Fig. 0.5b) indicating 1.5nm oxide growth.

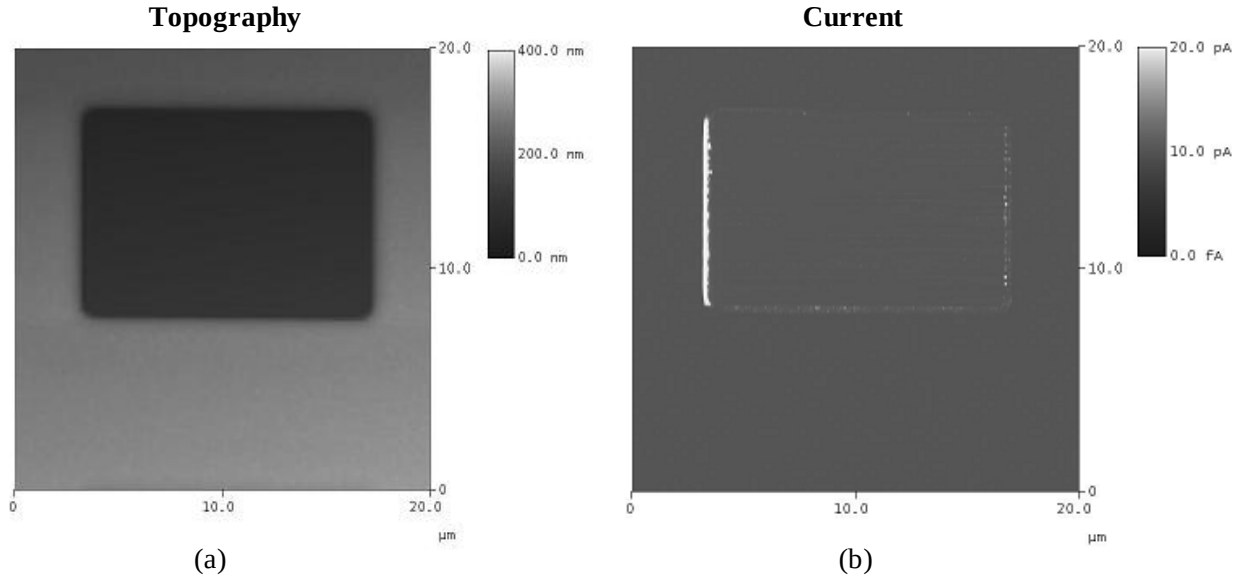


(a) (b)

Fig. 0.4: (a) Topography AFM image and (b) cross section indicating oxide growth after a previous AFM scan with a high tip-to-substrate voltage of 10V.

Example 4 – current artefacts

Artifacts appear when measuring currents over large steps. The reasons for this are two: first – the increased interaction force when the tip climbs up a hill due to slow time for feedback reaction (hence more currents on the left sidewall where the tip climbs up) and second – the larger contact area with steep slopes due to the apex shape.

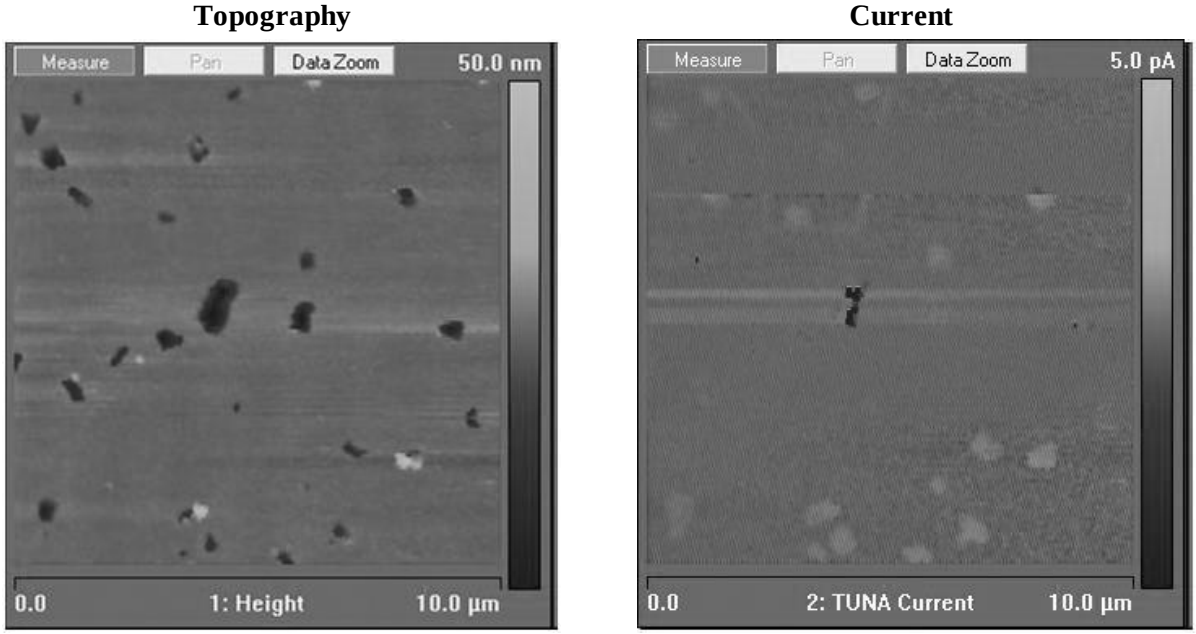


(a) (b)

Fig. 0.5: (a) Topography and (b) electric current AFM images of a partially processed wafer. The dark are on the topography scan is a region with a 5nm SiO_2 while the bright area is the surrounding thick LOCOS oxide.

Example 5 – insufficient tip aspect ratio

Fig. 0.2 shows topography and electric current images of a high-*K* surface after reactive ion etch. There are several holes on the topography image, but current flows only through one of them – the largest one. It could be concluded that the other holes in the high-*K* material are not deep enough and do not reach the silicon substrate. However, depth analysis of the holes indicates that they are deeper than the dielectric layer thickness. What this means is that the absence of currents is due to the insufficiently high aspect ratio of the tip that is unable to reach the bottom surface of the narrower holes, only on the large one. Clearly, a high aspect-ratio tip is needed for this type of measurements.



(a) (b)

Fig. 0.6: (a) Topography and (b) electric current images of a high-*K* dielectric surface after reactive ion etch.

List of publications

Periodicals and conferences

1. **Endres, R., Stefanov, Y., Wessely, F., Zaunert, F. and Schwalke, U.** Process damage-free damascene metal gate technology for gentle integration of epitaxially grown high-k gate dielectrics, Microelectronic Engineering, 85, pp. 15-19 (2008).
2. **Zaunert, F., Endres, R., Stefanov, Y., and Schwalke, U.** Evaluation of MOSFETs with crystalline high-k gate-dielectrics: device simulation and experimental data, Journal of Telecommunications and Technology, Vol.2, p. 78, 2007.
3. **Endres, R., Stefanov, Y. and Schwalke, U.** Electrical Characterization of Crystalline Gd_2O_3 Gate Dielectric MOSFETs Fabricated by Damascene Metal Gate Technology, Microelectronics Reliability, **47**, pp. 528-531, (2007).
4. **Endres, R., Stefanov, Y. and Schwalke, U.** Electrical Performance of Damascene Metal Gate MOSFETs with Crystalline Gd2O3 Gate Dielectric, 37th IEEE Semiconductor Interface Specialists Conference (SISC2006), December 7-9, 2006, San Diego, CA, USA.
5. **Endres, R., Stefanov, Y. and Schwalke, U.** Damascene Metal Gate Technology for Gentle Integration of Crystalline High-K-Gate Dielectrics, ECS Transactions, **3 (2)**, pp. 297-301, (2006).
6. **Rispal, L., Ruland, T., Stefanov, Y., Wessely, F. and Schwalke, U.** Conductive AFM Measurements on Carbon Nanotubes and Application for CNTFET Characterization, ECS Transactions, **3 (2)**, pp. 441-448, (2006)
7. **Endres, R., Stefanov, Y. and Schwalke, U.** Damascene Metal Gate Technology for Gentle Integration of Crystalline High-K-Gate Dielectrics, 210th Electrochemical Society Meeting, Cancun, Mexico, Oct. 29 – Nov. 3, 2006.
8. **Rispal, L., Ruland, T., Stefanov, Y., Wessely, F. and Schwalke, U.** Conductive AFM Measurements on Carbon Nanotubes and Application for CNTFET Characterization, 210th Electrochemical Society Meeting, Cancun, Mexico, Oct. 29 – Nov. 3, 2006.
9. **Endres, R., Stefanov, Y., Wessely, F., Zaunert, F. and Schwalke, U.** Process Damage-Free Damascene Metal Gate Technology for Gentle Integration of Epitaxially Grown High-K Gate Dielectrics, IEEE EDS 3rd International Symposium on Advanced Gate Stack Technology (ISAGST) September 27- 29, 2006, Austin, Texas, USA.
10. **Gottlob, H. D. B., Echtermeyer, T., Mollenhauer, T., Schmidt, M., Efavi, J.K., Wahlbrink, T., Lemme, M. C., Kurz, H., Endres, R., Stefanov, Y., Schwalke, U., Czernohorsky, M., Bugiel, E., Fissel, A. and Osten, H. J.** Approaches to CMOS integration of epitaxial gadolinium oxide high-k dielectrics, European Solid State Device Research Conference (ESSDERC), Montreux, Switzerland, Sept. 18 – 22, 2006.
11. **Endres, R., Stefanov, Y. and Schwalke, U.** Epitaxial High-K Oxide Metal Gate MOSFETs: Damascene CMP Process Integration and Electrical Results, International Conference on Solid State Devices and Materials (SSDM), September 13-15, 2006, Yokohama, Japan.
12. **Endres, R., Stefanov, Y. and Schwalke, U.** Electrical Characterization of Crystalline Gd2O3 Gate Dielectric MOSFETs Fabricated by Damascene Metal Gate Technology, 14th Workshop on Dielectrics in Microelectronic (WoDiM), Catania, Italy, June 26 – 28, 2006.
13. **Zaunert, F., Endres, R., Stefanov, Y. and Schwalke, U.** Evaluation of MOSFETs with Crystalline High-k Gate-Dielectrics: Device Simulation and Experimental Data, 7th Symposium Diagnostics & Yield, Advanced Silicon Devices and Technologies for ULSI Era, June 25-28, 2006, Warszawa, Poland (Invited).
14. **Rispal, L., Stefanov, Y., Wessely, F. and Schwalke, U.** Carbon Nanotube Transistor Fabrication Assisted by Topographical and Conductive Atomic Force Microscopy, Japanese Journal of Applied Physics, pp. 3672-3679, 2006.
15. **Rispal, L., Wessely, F., Stefanov, Y., Ruland, T. and Schwalke, U.** CMOS-compatible Fabrication Process of Carbon-Nanotube-Field-Effect Transistors, IEEE EDS Workshop on

Advanced Electron Devices (Micro- and Nano Devices), June 13 – 14, 2006, Fraunhofer-Institute IMS, Duisburg, Germany.
16. **Stefanov, Y., Cilek, F., Endres, R. and Schwalke, U.** Alternative Optimization Techniques For Shallow Trench Isolation And Replacement Gate Technology CMP, The 2nd Pacific-Rim International Conference on Planarization CMP and Its Application Technology (PacRim-CMP 2005) Seoul, Korea, November 17 - 19, 2005.
17. **Stefanov, Y., Singh, R., DasGupta, N., Misra, P. and Schwalke, U.** Conductive Atomic Force Microscopy Study of Leakage Currents through Microscopic Structural Defects in High-K Gate Dielectrics, The Electrochemical Society Conference on Crystalline Defects and Contamination, (ECS-DECON 2005), Grenoble, France, Sept. 12 – 16, 2005.
18. **Rispal, L., Stefanov, Y., Heller, R., Tzschöckel, G., Hess, G., Haberle, K. and Schwalke, U.** Topographic and Conductive AFM Measurements on Carbon Nanotube Field-Effect Transistors Fabricated by In-situ Chemical Vapor Deposition,
International Conference on Solid State Devices and Materials (SSDM), Kobe, Japan, September 12 – 15, 2005.
19. **Stefanov, Y., Rispal, L. and Schwalke, U.** Nanoscale Electrical Characterization of Crystalline Praseodymium Oxide High-k Gate Dielectric MOSFETs with Conductive Atomic Force Microscopy, Seeing at the Nanoscale III, Santa Barbara, California, USA, August 13-16, 2005.
20. **Rispal, L., Stefanov, Y. and Schwalke, U.** Atomic Force Microscopy (AFM) and Electrical Characterization of Carbon Nanotube (CNT) Devices Fabricated by Chemical Vapour Deposition, Seeing at the Nanoscale III, Santa Barbara, California, USA, August 13-16, 2005.
21. **Stefanov, Y., Komaragiri, R. and Schwalke, U.** Technological Advances for Memory Applications: Crystalline High-K Gate Dielectrics and Alternatively Doped Gates, Proceedings of the International Conference on Memory Technology and Design (ICMDT 2005), p. 167, Giens, France, May 21 – 24, 2005.
22. **Ruland, T., Stefanov, Y., Rispal L. and Schwalke, U.** Application of Atomic Force Microscopy in Resist Structure Evaluation, VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik, VDI-Berichte Nr. 1860, 2004.
23. **Schwalke, U. and Stefanov, Y.** Process integration and nanometer-scale electrical characterization of crystalline high-k gate dielectrics, Microelectronics Reliability, 45, pp. 790 - 793, 2005.
24. **Gottlob, H. D. B., Lemme, M. C., Mollenhauer, T., Wahlbrink, T., Efavi, J. K., Kurz, H., Stefanov, Y., Haberle, K., Komaragiri, R., Ruland, T., Zaunert, F. and Schwalke, U.,** Introduction of Crystalline High-K Gate Dielectrics in a CMOS Process
Journal of non-Crystalline Solids, Volume 351, Issues 21-23 , Pages 1885-1889, 2005.
25. **Stefanov, Y., Ruland, T. and Schwalke, U.** Electrical AFM Measurements for Evaluation of Nitride Erosion in ShallowTrench Isolation Chemical Mechanical Planarization, Materials Research Society, 2004 MRS Fall Meeting, Symposium on Scanning Probe and Other Novel Microscopies of Local Phenomena in Nanostructured Materials, Boston, USA Nov. 29 – Dec 3, 2004.
26. **Stefanov, Y., Komaragiri, R. and Schwalke, U.** Device Level and Nanoscale Electrical Characterization of Crystalline Praseodymium Oxide High-k Gate Dielectric MOSFETs, SEMATECH International Workshop on Electrical Characterization and Reliability for High-K Devices, Austin, Texas, USA, Nov. 4-5, 2004 .
27. **Stefanov, Y., Komaragiri, R., Ruland, T. and Schwalke, U.** Electrical AFM Measurements for STI CMP Erosion Evaluation, Seeing at the NanoScale II International Conference Abstracts p. 97, Grenoble, France 13th – 15th October, 2004.
28. **Ruland, T., Stefanov, Y., Rispal, L. and Schwalke, U.** Application of Atomic Force Microscopy in Resist Structure Evaluation, 8th International Symoposium on Metrology and Quality Control (ISMQC), October 12 – 15, 2004, Erlangen, Germany.
29. **Stefanov, Y., Hess, G., Tzschöckel, G. and Schwalke, U.** Global and Local Electrical Behavior of Crystalline Praseodymium Oxide High-k Gate Dielectric MOSFETs, Proceedings of the

Electrochemical Society International Semiconductor Technology Conference (ECS-ISTC) Shanghai China, September, 15 - 17, 2004.

30. **Schwalke, U. and Stefanov, Y.** Process Integration and Electrical Characterization of Crystalline High-K Gate Dielectrics, Workshop on Dielectrics in Microelectronics (WoDiM), Cork, Ireland, June 28 – 30, 2004.
31. **Gottlob, H.D.B., Lemme, M.C., Mollenhauer, T., Wahlbrink, T., Efavi, J.K., Kurz, H., Stefanov, Y., Haberle, K., Komaragiri, R., Ruland, T., Zaunert, Fl. and Schwalke, U.** Introduction of Crystalline High-K Gate Dielectrics in a CMOS Process, SiO2, Advanced Dielectrics & Related Devices, Chamonix Mont-Blanc, France, on June 21-23, 2004.
32. **Komaragiri, R., Schwalke, U., Stefanov, Y. and Ruland, T.** Comparison of praseodymium oxide gate MOSFETs with conventional SiO2 MOSFETs: A simulation study, Proceedings of the International Workshop on Physics of Semiconductor Devices, (IWSPD), pp. 462 - 464, Madras, India, 2003.
33. **Schwalke, U., Stefanov, Y., Komaragiri, R. and Ruland, T.** Electrical Characterisation of Crystalline Praseodymium Oxide High-k Gate Dielectric MOSFETs, Proceedings of 33rd European Solid State Device Research Conference (ESSDERC), pp. 247-250, 2003.

Book chapters

1. **Stefanov, Y. and Schwalke, U.** Shallow Trench Isolation Chemical Mechanical Planarization, Microelectronic Applications of Chemical Mechanical Planarization, Edited by Yuzhuo Li, Wiley & Sons, Inc. NY, 2007, ISBN: 978-0-471-71919-9.

List of supervised work projects and diploma theses

1. **Rispal, L.** Evaluation and Optimization of Chemical Mechanical Polishing, S237, 15.05.2003-31.10.2003.
2. **Brückner, C.** Evaluation and Optimization of Local and Global Planarity on Structured Silicon Wafers via Chemical Mechanical Polishing, S239, 20.10.2003-24.05.2004.
3. **Kunz**, **F.** Aufbau einer Messanordnung zur Grenzflaechen-Charakterisierung von MOSFETs mittels Ladungspumpen, S242, 15.30.2004-29.07.2004.
4. **Cilek, F.** Umsetzung von alternativen Verfahren zur globalen STI CMP Planarisierung, D228, 08.03.2004-11.10.2004.
5. **Wachter, C**. Prozessintegration und elektrische Bewertung von MOSFETs mit kristallinen high-K Gatedielektrika, S249, 01.11.2004-24.03.2005.
6. **Endres, R.** Entwicklung einer Replacement-Gate-Technologie für Metall-Gate-MOSFETs mit High-k-Dielektrika, D238, 09.05.2005-05.12.2005.

Bibliography

1. **Moore, G.** Cramming more components onto integrated circuits. *Electronics Magazine.* 19 April, 1965. Vol. 38, 8.
2. **Schmaltz, G.** Über Glätte und Ebenheit als physikalisches und physiologishes Problem. *Zeitschrift des Vereines deutscher Ingenieure.* Oct 12, 1929. pp. 1461-1467.
3. U. S. Patent 2, 7288,222.
4. **Young, R., Ward, J. and Scire, F.** The Topografiner: An Instrument for Measuring Surface Microtopography. *Review of Scientific Instruments.* 1972. Vol. 43, pp. 999-1011.
5. **Binnig, G. and Rohrer, H.** Scanning tunneling microscopy. *IBM Journal of Research and Development.* 1986. Vol. 30, 4.
6. **Binnig, G., et al.** Surface Studies by Scanning Tunneling Microscopy. *Physical Review Letters.* 1982. Vol. 49, 1, p. 57.
7. **Binnig, G., Quate, C. F. and Gerber, C.** Atomic Force Microscope. *Physical Review Letters.* 1986. Vol. 56, 9, p. 930.
8. **Checco, A., et al.** High resolution non-contact AFM imaging of liquids condensed onto chemically nanopatterned surfaces. *Ultramicroscopy.* June-July 2006. 106(8-9):703-8.
9. **Binnig, G., et al.** Atomic resolution with atomic force microscope. *Europhysics Letters.* Vol. 3, p. 1281.
10. **Lennard-Jones, J. E.** On the Determination of Molecular Fields. *Proceedings of the Royal Society of London.* Vol. A 106 (738), pp. 463–477.
11. **Martin, Y., Williams, C. C. and Wickramasinghe, H. K.** Atomic force microscope-force mapping and profiling on a sub 100-Å scale. *Journal of Applied Physics.* 1987. Vol. 61, pp. 4723–4729.
12. **Zhong, Q., et al.** Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy. *Surface Science Letters.* June 1993. Vol. 290, 1-2, pp. L688-L692.
13. **Hansma, P. K., et al.** Tapping Mode Atomic Force Microscopy in Liquids. *Applied Physics Letters.* 1994. Vol. 64, p. 1738.
14. **Serry, F. M., et al.** Electric Force Microscopy, Surface Potential Imaging, and Surface Electric Modification with the Atomic Force Microscope (AFM). *Application Notes.* s.l. : Veeco. http://www.veeco.com/pdfs/appnotes/AN27_EFMSurfPot_260.pdf.
15. **Maivald, P., et al.** Using force modulation to image surface elasticities with the atomic force microscope. *Nanotechnology.* 1991. Vol. 2, p. 102.
16. **Schmitt, C. C., Elings, J. R. and Serry, M.** Nanoindenting, Scratching, and Wear Testing with the Atomic Force Microscope. *Application Notes.* s.l. : Veeco. http://www.veeco.com/pdfs/appnotes/AN13_NanoIndent_081804_RevA1_6.pdf.
17. **DeWolf, P., et al.** Electrical Characterization with Scanning Probe Microscopes. *Application Notes.* s.l. : Veeco. http://www.veeco.com/pdfs/appnotes/AN079-RevA1-Electrical_Characterization_with_Scanning_Probe_Microscopes-AppNote.pdf.
18. **Hori, T.** Gate Dielectrics and MOS ULSIs. Berlin and New York : Springer, 1997. Vol. 34. ISBN 3540631828.
19. **Sabnis, A. and Clemens, J.** Characterization of the electron mobility in the inverted <100> Si surface. *IEDM Technical Digest.* December 1979. pp. 18-21.
20. **Gröseneken, G., et al.** A Reliable Approach to Charge-Pumping Measurements in MOS Transistors. *IEEE Transactions on Electron Devices.* January 1984. Vols. ED-31, No. 1, pp. 42-53.
21. **Maneglia, Y., Bauza, D. and Ouisse, T.** Extraction of the slow oxide trap density in MOS transistors using the charge pumping method. *Semiconductor Conference, 1995. CAS'95 Proceedings., 1995 International.* 11-14 Oct. 1995. pp. 155 - 158.
22. **Wolf, S.** Silicon Processing for the VLSI Era: Process Integration. Sunset Beach, CA : Lattice Press, 1990. Vol. 2. ISBN: 0-9616721-4-5.
23. **Atalla, M. M., Tannenbaum, E. and Scheibner, E. J.** Stabilization of silicon surfaces by thermally grown oxide. *Bell System Technical Journal.* 1959. Vol. 38, p. 749.

24. **Morandi, F.** The MOS planox process. *IEDM Abstracts.* 1969. p. 126.
25. **Appels, J. A., et al.** Local Oxidation of Silicon and Its Application in Semiconductor-Device technology. *Philips Research Reports.* 1970. Vol. 25, pp. 118-132.
26. **Chatterjee, A., et al.** Integration of unit processes in a shallow trench isolation module for a 0.25 μm complementary metal-oxide semiconductor technology. *Journal of Vacuum Science and Technology.* November/December 1997. B 15(6), pp. 1936-1942.
27. **Oldham, W.** Isolation technology for scaled MOS VLSI. *IEDM Technical Digest.* December 1982. pp. 216-219.
28. **Bryant, A., Haensch, W. and Mii, T.** Characteristics of CMOS Device Isolation for the ULSI Age. *IEDM Technical Digest.* December 1994. pp. 671-674.
29. **Fazan, P. C. and Mathews, V. K.** A Highly Manufacturable Trench Isolation Process for Deep Submicron DRAMs. *IEDM Technical Digest.* December 1993. pp. 57-60.
30. **Han, Y. and Ma, B.** Isolation process using buffer layer for scaled MOS/VLSI. *ECS Extended Abstracts.* 1984. Vols. 84-1, p. 98.
31. **Shimuzu, N., et al.** A Poly-Buffer Recessed LOCOS Process for 256Mbit Dram Cells. *IEDM Technical Digest.* December 1992. pp. 279-282.
32. **Nandakumar, M., et al.** Shallow Trench Isolation for Advanced ULSI CMOS Technologies. *IEDM Technical Digest.* December 1998. pp. 133-136.
33. **Pyi, S.-H., et al.** Roles of Sedewall Oxidation in the Devices with Shallow Trench Isolation. *Electron Device Letters.* August 1999. Vol. 20, pp. 384-386.
34. **Chang, C. P., et al.** A Highly Manufacturable Corner Rounding Solution for 0.18μm Shallow Trench Isolation. *IEDM Technical Digest.* December 1997. pp. 661-664.
35. **Nag, S., et al.** Comparative Evaluation of Gap-Fill Dielectrics in Shallow Trench Isolation for Sub-0.25μm Technologies. *IEDM Technical Digest.* December 1996. pp. 841-844.
36. **Curtis, T. O., Pye, J. T. and Poreda, J. T.** APCVD TEOS:O3 Advanced Isolation Applications. 9th Edition *Semiconductor Fabtech.* 1999.
37. **Lee, H. S., et al.** An Optimized Densification of the Filled Oxide for Quarter Micron Shallow Trench Isolation (STI). *Symposium on VLSI Technology, Technical Digest.* 1996. pp. 158-159.
38. **Neureither, B., et al.** Resist Etch Back as a Manufacturable Low Cost Alternative to CMP. *Proceedings of the Eleventh International VLSI Multilevel Interconnect Conference.* 1994. pp. 151-157.
39. **Davari, B., et al.** A New Planarization Technique, Using a Combination of RIE and Chemical Mechanical Polish (CMP). *IEDM Technical Digest.* December 1989. pp. 61-64.
40. **Roy, S. R., et al.** Post chemical-mechanical planarization cleanup process for interlayer dielectric films. *Journal of the ECS.* 1995. Vols. 312, No. 1, pp. 216-226.
41. **Chatterjee, A., et al.** Chemical Mechanical Planarization (CMP) Process Windows in Shallow Trench Isolation for Advanced CMOS. *Electrochemical Society Proceedings.* 1996. Vols. 96-22, pp. 219-227.
42. **Smekalin, K.** CMP dishing effects in shallow trench isolation. *Solid State Technology.* July 1997. Vol. 40, pp. 187-194.
43. **Boning, D. and Lee, B.** Nanotopography Issues in Shallow Trench Isolation CMP. *MRS Bulletin.* October 2002. pp. 761-765.
44. **Mohapatra, N. R., et al.** The effect of high-K gate dielectrics on deep submicrometer CMOS device and circuit performance. *IEEE Transactions on Electron Devices.* May 2002. Vol. 49, pp. 826-831.
45. **Gusev, E. P., Buchanan, D. A. and Cartier, E.** Ultrathin high-K gate stacks for advanced CMOS devices. *IEDM Technical Digest.* December 2001. pp. 451-454.
46. **Robertson, J.** Band offsets of wide-band-gap oxides and implications for future electronic devices. *Journal of Vacuum Science and Technology B.* 2000. Vol. 18, no. 3, pp. 1785-1791.
47. **Chen, P. J., et al.** Thermal stability and scalability of Zr-aluminate-based high-k gate stacks. *VLSI Symposium on technology.* 2002. pp. 192- 193.
48. **Jones, J. T., et al.** Epitaxial silicon grown on CeO2/Si(111) structure by molecular beam epitaxy. *Journal of Vacuum Science and Technology.* 1998. Vol. B16, pp. 2686-2689.

49. **Fork, D. K., Fenner, D. B. and Geballe, T. H.** Growth of epitaxial PrO2 thin films on hydrogen terminated Si(111) by pulsed laser deposition. *Journal of Applied Physics.* 1990. Vol. 68, pp. 4316-4318.
50. **Czernohorsky, M., et al.** Impact of oxygen supply during growth on the electrical properties of crystalline Gd2O3 thin films on Si(001). *Applied Physics Letters 88.* 2006. 152905-1/3.
51. **Osten, H. J., et al.** High-k Gate Deielctrics with Ultra-low Leakage Current Based on Praseodymium Oxide. *Technical digest IEDM.* 2000. p. 653.
52. **Rispal, L.** Evaluation and Optimization of Chemical Mechanical Polishing. *S237, ISTN, TU Darmstadt, 2003.*
53. **Brückner, C.** Evaluation and Optimization of Local and Global Planarity on Structured Silicon Wafers via Chemical Mechanical Polishing. *S239, ISTN, TU Darmstadt, 2003-2004.*
54. **Heaton, M., Ge, L. and Serry, F.** Atomic Force Profilometry for Chemical Mechanical Polishing Metrology. *Semiconductor FabTech.* 11th Edition.
55. **Steigerwald, J. M., Murarka, S. P. and Gutmann, R. J.** Chemical Mechanical Planarization of Microelectronic Materials. s.l. : John Wiley & Sons, Inc., 1997. ISBN: 9783527617753.
56. **Stefanov, Y., et al.** Alternative Optimization Techniques for Shallow Trench Isolation and Replacement Gate Technology CMP. *Proceedings of the PacRim-CMP.* 2005. pp. 51-57.
57. **Chen, Y., Gupta, P. and Kahng, A. P.** Performance-impact limited area fill synthesis. *Design Automation Conference, 2003. Proceedings.* 2-6 June 2003. pp. 22 - 27.
58. **Sarlet, G., Morthler, G. and Baets, R.** Dummy-free STI concept for mixed-signal applications. *IMEC Newsletter.* October 1999. No. 25, pp. 3-4.
59. **Cilek, F.** Umsetzung von alternativen Verfahren zur globalen STI CMP Planarisierung. *D228, ISTN, TU Darmstadt, 2004.*
60. **Stefanov, Y., et al.** Conductive Atomic Force Microscopy Study of Leakage Currents through Microscopic Structural Defects in High-K gate Dielectrics. *The Electrochemical Society Conference on Crystalline Defects and Contaminations (ECS-DECON 2005).* Grenoble, France : s.n., September 12-16, 2005.
61. **He, B., et al.** A 1.1 nm Oxide equivalent gate insulator formed using TiO2 on nitrided silicon. *Electron Devices Meeting International, IEDM Technical Digest.* 1998. pp. 1038-1040.
62. **Campbell, S. A., et al.** TiO2 Based Gate Insulators. *IBM Journal of research and Development.* 1999. 43, pp. 383-392.
63. **Kadoshima, M., et al.** Rutile-type TiO2 thin film for high-k gate insulator. *Thin Solid Films.* 2002. 424, pp. 224-228.
64. **Piscual, J., Camassel, J. and Mathieu, H.** Fine structure in the intrinsic absorption edge of titanium dioxide. *Physical Review B.* 1978. Vol. 18, pp. 5606-5614.
65. **Hubbard, K. J. and Schlom, D. G.** Thermodynamic Stability of Binary Oxides in Contact with Silicon. *Journal of Material Research.* 1996. 11, pp. 2757-2776.
66. **Singh, R. K. and Narayan, J.** Laser Technique for Deposition of Thin Films: Physics and Model. *Physical review B.* 1990. 41, pp. 8843-8861.
67. **Kern, W. and Puontinen, D. A.** Cleaning solutions based on hydrogen peroxide for use in silicon semiconductor technology. *Physical review B.* 1970. 31, pp. 187-206.
68. **Paily, R., et al.** Effect of Oxygen Pressure During Pulsed Laser Deposition of TiO2 on the MTOS (Metal-TiO2-SiO2-Si) Capacitor Characteristics. *Thin Solid Films.* 2004. Vols. 462-463C, pp. 57-62.
69. **Osten, H. J., et al.** Epitaxial Praseodymium Oxide: A New High-K Delectric. *International Workshop on Gate Dielectrics.* Tokyo : s.n., November 2001. pp. 100-106.
70. **Liu, J. P., et al.** Epitaxial growth of Pr2O3 on Si(111) and the observation of a hexagonal to cubic phase transition during postgrowth N2 annealing. *Applied Physics Letters.* July 30, 2001. Vol. 79, No. 5, pp. 671-673.
71. Fuenfter Zwischenbericht zum Teilprojekt 'Molekularstrahlepitaxie von SE-Oxiden fuer sun-100 nm CMOS Transistoren. *Kristalline Gate-Stacks fuer sub 100nm CMOS-Transistoren aud FD-SOI.* 2005.
72. **Delugas, P. and Fiorentini, V.** Dielectric properties of two phases of crystalline lutetium oxide. *Microelectronics Reliability 45(56).* 2005. pp. 831-833.

73. **Schwalke, U., et al.** Process Integration of Crystalline Pr2O3 High-k Gate Dielectrics. *Proceedings of the 32nd European Solid-State Device Research Conference.* September 2002. p. 407.
74. **Gottlob, H. D. B., et al.** Approaches to CMOS integration of epitaxial gadolinium oxide high-k dielectrics. *European Solid State Device Research Conference.* Montreux, Switzerland : s.n., September 18-22, 2006.
75. **Gottlob, H. D. B., et al.** CMOS Integration of Epitaxial Gd2O3 High-K Gate Dielectrics. *Solid State Electronics.* June 2006. Issue 6, pp. 979-985.
76. **Kunz, S.** Aufbau einer Messanordnung zur Grenzenflaechen-Charakterizierung von MOSFETs mittels Ladungspumpen. *S242, IHTN, TU-Darmstadt, 2004.*
77. **Wachter, C.** Prozessintegration und elektrische Bewertung von MOSFETs mit kristallinen high-K Gatedielektrika'. *S239, ISTN, TU-Darmstadt, 2003-2004.*
78. **Chatterjee, A., et al.** Sub-100nm Gate Length Metal Gate NMOS Transistors Fabricated by a Replacement Gate Process. *IEDM Technical Digest.* 1997. pp. 821-824.
79. **Endres, R.** Entwickling einer Replacement-Gate-Technologie für Metall-Gate-MOSFETs mit High-k-Dielektrika. *D238, ISTN, TU Darmstadt, 2005.*
80. **Schwalke, U. and Stefanov, Y.** Process integration and nanometer-scale electrical characterization of crystalline high-k gate dielectrics. *Microelectronics Reliability.* 2005. Vol. 45, pp. 790-793.

Curriculum vitae

Name	Yordan Stefanov
Date of birth	28.12.1978
Place of birth	Sofia, Bulgaria
1992-1997	American College of Sofia *High-school degree*
1997-2002	Sofia University, Engineering Physics Department *Master's degree*
2003-2011	Darmstadt University of Technology, Institute for Semiconductor Technology and Nanoelectronics *Doctoral degree*
2006-present	Innovative Solutions Bulgaria *R&D scientist*