Charge carrier traffic at interfaces in nanoelectronic structures
Charge carrier traffic at interfaces in nanoelectronic structures

BAHMAN RAEISSI

Department of Microtechnology and Nanoscience - MC2
Physical Electronics Laboratory
CHALMERS UNIVERSITY OF TECHNOLOGY
GÖTEBORG, SWEDEN 2010
Charge carrier traffic at interfaces in nanoelectronic structures

BAHMAN RAEISSI

©Bahman Raeissi 2010
Doktorsavhandlingar vid Chalmers Tekniska Högskola
Ny serie nr 3056
ISSN 0346-718X

ISSN 1652–0769
Technical Report MC2-163

Department of Microtechnology and Nanoscience - MC2
Physical Electronics Laboratory
CHALMERS UNIVERSITY OF TECHNOLOGY
SE – 412 96 Göteborg, Sweden
Telephone: +46(0)31 772 10 00
http://www.chalmers.se

Cover:
Top left: Measured capacitance as a function of temperature and gate voltage for the Si/SiO$_x$/HfO$_2$ structure. Top right: Measured conductance as a function of temperature and gate voltage for the same structure. Bottom: TEM picture of the Si/SiO$_x$/HfO$_2$ structure.

Printed by Chalmers Reproservice
Göteborg, Sweden, February 2010
ABSTRACT

This thesis describes investigations in relation to the search for materials with high dielectric constant, $k$, for future CMOS transistors. The most elementary quantities to be considered are $k$-value and energy band offsets between the dielectric and the silicon crystal on which it is deposited. Empirical relations for these two quantities are presented demonstrating that only a few dielectrics investigated up to now have properties providing the basic demands in the development of CMOS technology.

Process development was done to deposit HfO$_2$, Pr$_2$O$_3$ and HiPrO on silicon by reactive sputtering in order to fabricate MOS capacitors. Electrical properties of these oxides were investigated by employing different techniques such as capacitance-voltage (C-V), current-voltage (I-V), capacitance frequency spectroscopy, stepped C-V, multiparameter admittance spectroscopy (MPAS) and thermally stimulated current (TSC).

A capacitance frequency spectroscopy technique was developed to investigate electron capture cross sections of the interface states at high-$k$/Si interface from experimental results. It is found that capture cross sections of electron states at this interface are thermally activated and exponentially depend on energy depth of interface states in the silicon bandgap. These processes indicate that the capture mechanism is governed by multiphonon.

MPAS a diagnostic tool developed from the conductance method to deliver more information regarding charge carrier states in semiconductor structures. Using this technique on HfO$_2$/Si interface, two different types of interface states with different capture mechanisms were found.

An interlayer of SiO$_x$ is found between the silicon crystal and HfO$_2$ in TEM pictures. A transition region, which is expected to have strong concentration gradients, exists between SiO$_x$ and HfO$_2$. This region shows an unstable atomic arrangement and contains charge carrier traps. These traps exchange electrons with the conduction band of the silicon crystal through a combined thermal-tunneling mechanism.

Keywords: High-$k$, HfO$_2$, interface states, oxide traps.
List of publications

Appended papers

The thesis is based on the following papers:

A. Navigation aids in the search for future high-\(k\) dielectrics: Physical and electrical trends

B. Vibronic nature of hafnium oxide/silicon interface states investigated by capacitance frequency spectroscopy
   O. Engström, B. Raeissi, and J. Piscator

C. High-\(k\)-oxide/silicon interfaces characterized by capacitance frequency spectroscopy

D. Characterization of traps in the transition region at the HfO\(_2\)/SiO\(_x\) interface by thermally stimulated currents
   B. Raeissi, J. Piscator, Y. Y. Chen, and O. Engström

E. The conductance method in a bottom-up approach applied on hafnium oxide/silicon interfaces
   J. Piscator, B. Raeissi, and O. Engström

F. Multiparameter admittance spectroscopy for metal-oxide-semiconductor systems
   J. Piscator, B. Raeissi, and O. Engström
G. Multiparameter admittance spectroscopy as a diagnostic tool for interface states at oxide/semiconductor interfaces
B. Raeissi, J. Piscator, and O. Engström

H. The role of mobile charge in oxygen plasma enhanced silicon-to-silicon wafer bonding
B. Raeissi, A. Sanz-Velasco, and O. Engström
Other contributions

Other papers not appended to the thesis:

- **Charging phenomena at the interface between high-\(k\) dielectrics and SiO\(_2\) interlayers**

- **Wafer bonding strength increased by mobile ions**
  B. Raeissi, A. Sanz-Velasco, O. Engström

- **Electron traps at HfO\(_2\)/SiO\(_x\) interfaces**

- **Multiphonon capture of electrons at high-k-silicon interfaces**
  O. Engström, B. Raeissi, J. Piscator

- **Interface Defects in HfO\(_2\), LaSiO\(_x\), and Gd\(_2\)O\(_3\) High-k/MetalU\(G\)ate Structures on Silicon**

- **Leakage current effects on C-V plots of high-\(k\) MOS capacitors**

- **A generalised methodology for oxide leakage current metric**
  *Proc. 9\(^{th}\) European Workshop on Ultimate Integration of Silicon (ULIS 2008)*, p. 167, March 12-14, Udine, Italy (2008).
• High-$k$-Oxide/Silicon Interfaces Characterized by Capacitance Frequency Spectroscopy

• Relation Between Electrical and Mechanical Characteristics of Low-Temperature Bonded Si/Si Interfaces
  B. Raeissi, A. Sanz-Velasco, and O. Engström

• Extracting the relative dielectric constant for "high-$k$ layers" from CV measurements - Errors and error propagation

• Navigation aids in the search for future high-$k$ dielectrics: physical and electrical trends
  O. Engström, B. Raeissi, S. Hall, O. Buiu, M. C. Lemme, H. Gottlob, P. K. Hurley, K. Cherkaoui

• Electrical properties of low-temperature bonded unipolar Si/Si junctions
  B. Raeissi, P. Amirfeiz and O. Engström
## Contents

Abstract

v

List of publications

vii

Contents

xi

List of Figures

xiii

1 Introduction

1

2 Charge carrier statistics

5
  2.1 Introduction ........................................ 5
  2.2 Recombination-generation processes ................. 5
  2.3 Emission rate and capture cross section ............ 7
  2.4 Statistical background .............................. 9

3 High-\(k\) dielectrics

19
  3.1 Inside dielectrics ................................. 19
  3.2 Requirements .................................... 21
  3.3 Interface states and bulk oxide traps ............. 24

4 Silicon/silicon bonded interfaces

27

5 Sample preparation

29
  5.1 MOS fabrication .................................. 29
  5.2 Oxygen plasma enhanced Si/Si bonding ........... 32

6 Measurement techniques

35
  6.1 Measurement setup .............................. 35
  6.2 Capacitance-voltage measurements ................. 36
  6.3 Stepped capacitance-voltage measurements ......... 39
  6.4 Capacitance frequency spectroscopy ............... 42
  6.5 Conductance method ............................. 44
6.6 Multiparameter admittance spectroscopy ...................... 45
6.7 Current-voltage measurements ................................. 45
6.8 Thermally stimulated current ................................. 48
6.9 Physical characterization .................................... 51

7 Summary ......................................................... 55

Acknowledgement .................................................. 57

References .......................................................... 59

Paper A ............................................................. 71
Paper B ............................................................. 79
Paper C ............................................................. 87
Paper D ............................................................. 95
Paper E ............................................................. 127
Paper F ............................................................. 133
Paper G ............................................................. 147
Paper H ............................................................. 165
List of Figures

2.1 Energy-band demonstrations of (a) band to band, (b) trap-assisted, and (c) Auger recombination processes. Filled and empty circles represent electron and hole, respectively. 7

2.2 (a) A system with a number of subsystems isolated from the rest of cosmos. (b) Subsystem i with energy $E_i$ and no interaction with other subsystems than thermal. (c) A subsystem which can exchange particle with other subsystems. 10

2.3 Schematic illustration of the relation between the energy eigenvalue $\Delta E_n$, the enthalpy $\Delta H_n$ and the free energy $\Delta G_n$ as a function of temperature. 15

2.4 Configurational coordinate diagram of a vibrating electron-trap system. The parabolas represent the vibrational energies and are separated in energy by the electronic energy needed to take the electron from the trap to the conduction band. The lower curve represents the sum of electronic and vibrational energies of the trap when the electron is captured to it. The upper curve gives the corresponding energy when the trap is empty and the electron is in the conduction band. Transition of an electron between the trap level and the conduction band occurs at the crossing point where the atomic positions coincide for empty and filled trap. The separation in lattice coordinate depends on the change in the symmetry of atomic binding force when an electron is transferred. 16

3.1 A simplified illustration of the polarization effect. (a) A parallel plate capacitor with charges $+Q$ and $-Q$ on the plates, and the electric field as a function of distance between the plates. (b) The capacitor when an atom is placed between the plates, and also the change in the electric field. 21
3.2 Different contributions to the polarizability. $E$ and $P$ are an applied electric field and polarization, respectively. In the case of dipolar polarizability, arrows represent the electric dipole moments of the constructing molecules of the material.  

5.1 Schematic structure of the fabricated MOS devices. High-$k$ oxides like HfO$_2$, Pr$_2$O$_3$ and HfPrO have been deposited by reactive sputtering.  

5.2 Schematic structure of the Si/Si bonded samples used for electrical measurements.  

6.1 Capacitance-voltage data of Al/HfO$_2$/Si gate stack at different frequencies in the range of 1 kHz - 1 MHz before PMA. $T = 300$ K.  

6.2 Capacitance-voltage measurements from the same sample as the one in Fig. 6.1 at the frequency of 1 kHz for various temperatures in the range of 50 - 300 K.  

6.3 Capacitance-voltage data of Al/HfO$_2$/Si gate stack at two different signal frequencies, 1 kHz and 1 MHz, before and after PMA.  

6.4 (a) Capacitance-voltage data of Al/HfO$_2$/Si gate stack at 100 kHz measured with 30 sec voltage sweeping cycle (two middle graphs) and by stepped C-V method (outer graphs). (b) Capacitance transient measured at 0.2 V, when waiting at $+4$ V and then stepping to 0.2 V. (c) Capacitance transient measured at 1.8 V, when waiting at $+4$ V (in accumulation) for 5 min and then stepping to 1.8 V. Taking the initial values of such transients for different gate voltages will result in stepped C-V data (outer graphs of (a)). $T = 300$ K.  

6.5 Capacitance data as a function of frequency from Fig. 6.1 at 0.2 V.  

6.6 Capacitance and conductance measured on a Al/HfPrO/SiO$_x$/Si MOS structure as a function of gate voltage and temperature for the frequency of 1 kHz. (a) and (b) show the capacitance in 3D and contour representations, respectively. (c) and (d) are the corresponding conductance measurements.  

6.7 Capacitance and conductance measured on a Al/HfO$_2$/SiO$_2$/Si MOS structure after PMA as a function of gate voltage and temperature for the frequency of 4 kHz. (a) and (b) show the capacitance in 3D and contour representations, respectively. (c) and (d) are the corresponding conductance measurements.  

6.8 Leakage current as a function of capacitance equivalent thickness (CET) of HfPrO and HfO$_2$ samples.
6.9 TSC measurements of the Al/HfO$_2$/SiO$_x$/Si MOS structure before PMA. The linear temperature rate is 10 K/min for all the curves and two different waiting times, 1 and 5 minutes, in inversion were used. The positive voltage values in the legends are charging biases and the negative ones are discharging biases.

6.10 TSC measurements of the Al/HfO$_2$/SiO$_x$/Si MOS structure. The sample is annealed (PMA) at 350 °C for 30 minutes in N$_2$ ambient. The linear temperature rate is 10 K/min for all the curves and two different waiting times, 1 and 5 minutes, in inversion were used. The positive voltage values in the legends are charging biases and the negative ones are discharging biases.

6.11 Comparison of TSC measurements of the Al/HfO$_2$/SiO$_x$/Si MOS structure before and after PMA. The linear temperature rates of 10 and 20 K/min and 1 minute waiting time in inversion were used. The positive voltage values in the legends are charging biases and the negative ones are discharging biases.

6.12 TEM picture of the HfO$_2$/SiO$_x$/Si structure.

6.13 TEM picture of the HfPrO/SiO$_x$/Si structure.

6.14 ERDA data from the sputtered HfPrO, where the TEM picture is depicted in Fig. 6.13 (with courtesy of Anders Hallen).
Chapter 1

INTRODUCTION

The most vital parts of CMOS based nanoelectronic circuits are found in the channels of the MOSFETs. There, the transistor currents are transported in intimate contact with an amorphous oxide on one side and a silicon crystal on the other. For the down-scaling of dimensions in CMOS technology, good control of the charge carriers running along this interface becomes more and more important and sets subtle demands on the channel properties. The complicated atomic geometries occurring at such a link between two disparate materials gives rise to electron and hole traps and thus to scattering centers for the channel current, leading to lower charge carrier mobilities. Together with similar trapping states in the bulk of the oxide, also instabilities occur in transistor threshold voltage. The main task of the work presented in this thesis has been to investigate and understand the physics of the charge carrier traffic in such structures for coming nanoelectronic generations.

Today, the MOSFET is the dominating device for amplifying and switching electronic signals in circuits for logic and computation. The basic principle of this device was first proposed by Julius Edgar Lilienfeld [1] in 1925 and later in 1960, Dawon Kahng and Martin Atalla developed a working device at Bell Labs [2, 3]. The MOSFET has achieved electronic domination and sustains the large scale integrated circuits underlying today’s information society.

The invention of integrated circuit (IC) by both Jack Kilby of Texas Instruments and Robert Noyce of Fairchild Semiconductor working independently of each other in 1958 has introduced an enormous improvement in electronics. The IC’s mass production capability, reliability, and building-block approach to circuit design ensured the rapid adoption of standardized ICs in place of designs using discrete transistors. However, Moore’s law, a forecast made by Intel co-founder Gordon Moore that the number of transistors on an IC will double every 18-24 months, has been the leading goal in the development of the ICs [4] and in order to follow this goal, device scaling is the key attribute. This relies on the possibility of down scaling the most
important electronic device combination of the complementary metal oxide semiconductor (CMOS) technology. For the gate function of the MOSFET, silicon dioxide, SiO₂, has been the ideal gate dielectric because of its amorphous structure, non-hygroscopicity, large band gap of 9 eV, band offsets more than 3 eV, good surface morphology, low density of interface states and process compatibility [5].

The thickness of the SiO₂ layer used as the gate dielectric has become so thin (below 2 nm) that the gate leakage current due to the direct tunneling of electrons through the SiO₂ exceeds 1 A/cm² at 1 V, such that the circuit power dissipation will increase to unacceptable values. Also, the reliability of SiO₂ films against electrical breakdown declines in thin films. Thus in the down-scaling of CMOS transistors, finding an alternative gate material with high dielectric constants, k, to replace SiO₂ for coming CMOS generations is one of the most challenging problems in the continued development of electronics [5,6].

Intensive research is underway to develop the high-k oxides such as HfO₂, ZrO₂, Gd₂O₃, Pr₂O₃, and La₂O₃ into new high quality electronic materials. Two of the most intensively investigated materials are HfO₂ and ZrO₂. In early 2007, Intel announced the deployment of hafnium-based high-k dielectrics in conjunction with a metallic gate for components built on 45 nanometer technologies, and has shipped it in the 2007 processor series codenamed Penryn [7,8].

The most important elementary requirements that need to be fulfilled by high-k candidates are both high enough k-value and band offset values between the conduction/valence bands of the oxide and the silicon crystal. These demands on physical properties of dielectrics are discussed in Paper A. For the low standby power (LSTP) 22 nm bulk node, it was found that only a few oxides could remain as possible candidates.

Another crucial attempt in finding potential high-k gate dielectrics is to attain a sufficiently high quality interface with the silicon channel, as close as possible to that of SiO₂. Finding any material to create a better interface than that of SiO₂ is improbable, since typical production SiO₂ gate dielectrics have a midgap interface state density, Dᵢ, in the range of 10¹⁶ cm⁻². Most of the high-k materials show Dᵢ in the range of 10¹² cm⁻² and furthermore exhibit a substantial flatband voltage, VFB, shift. It is important to understand the origin of the interface properties of any high-k gate dielectric, so that an optimum high-k/Si interface may be obtained [5,6]. In Papers B, C, D, E, F, and G such electrical properties of the interface between silicon and HfO₂ are investigated.

The interface between high-k dielectric and silicon plays a key role especially for all thin gate dielectrics and, in most cases, is the dominant factor in investigating the electrical properties of the gate stack. Most of the high-k
metal oxide structures investigated so far have unstable interfaces with silicon; somehow oxides react with silicon under equilibrium conditions to form an undesirable interfacial layer of $\text{SiO}_x$. Any ultrathin interfacial layer in the range of 2 nm will have the same quality, uniformity and reliability concerns as $\text{SiO}_2$ has in this thickness regime. Another issue is probable occurrence of a transition region between $\text{SiO}_x$ and high-$k$ oxide with undefined stoichiometry and possible structural instabilities [9]. Hence, it is important to understand the thermodynamics of these systems and thereby try to control the interface with silicon.

Existence of defect centers at the oxide/semiconductor interface and in the bulk of the oxide have caught a lot of attentions and investigations in CMOS technology. Characterizing the properties of charge carrier traps and interface states by employing physical and electrical methods has been demanded for the continuous scaling of transistors and become more important due to the use of novel high-$k$ materials for coming generations. The undesirable charge carrier traps are adversely affecting device performance and stability. The typical electrical techniques such as capacitance-voltage (C-V), conductance-voltage (G-V), current-voltage (I-V), thermally stimulated current (TSC) have been used to characterize the interface states and traps at High-$k$/Si interfaces. Other methods like capacitance frequency spectroscopy and multiparameter admittance spectroscopy (MPAS) as presented in Papers B, C, E, F and G, have been developed to extract more detailed information about charge carrier emission rates and capture cross sections of the interface states.

In the search for new dielectric, it is desirable to choose a material which remains in an amorphous phase throughout the processing treatments. Most of the gate dielectrics studied so far form polycrystalline grains during deposition or upon modest thermal treatments [6]. Polycrystalline gate dielectrics could be problematic due to grain boundaries, which serve as high leakage paths. Furthermore, grain size and orientation variations throughout a polycrystalline film may cause changes in $k$-value, leading to irreproducible characteristics. In such cases, mixing additional elements into the high-$k$ oxide may transfer the structure into a more glassy state.

Integrating any gate dielectric into mature CMOS technology requires compatibility with the gate material. Polysilicon had been used as the gate electrode with $\text{SiO}_2$ and had advantages such as being easily deposited, being refractory and compatible with $\text{SiO}_2$. In the case of High-$k$ gate dielectrics, both poly-silicon and metals are being evaluated as the gate electrodes [10,11]. A fundamental issue with polysilicon gate on high-$k$ oxides like hafnium based dielectrics is the Fermi-level pinning, which is consistent with defect states due to Hf-Si bonding. This effect decreases the modulation efficiency of the channel potential by the gate voltage and causes high
threshold voltage \[12\]. Even if process integration schemes of silicon based gates are well established, most of the gate dielectric candidates investigated so far require metal gates. This is expected because in the case of poly-silicon gate, the presence of SiO\(_x\) at both channel and gate interfaces limits device performance and enhances equivalent oxide thickness (EOT). Surface phonon scattering is a primary source of mobility degradation in high-\(k\) oxides, where using metal gate is an effective solution for screening phonon scattering and improving channel mobility \[13\]. Metal gates such as TiN have been used with most of the high-\(k\) gate dielectrics to prevent reaction at the gate interface. In addition, metal gate processing can lower the required thermal budget by eliminating the need for a dopant activation anneal, which is used in the polysilicon electrode \[6\].

There are two basic approaches toward achieving successful metal electrodes, either a single midgap metal or two separate metals for nMOS and pMOS. The main advantage of employing a midgap metal, such as TiN, arises from a symmetrical threshold voltage, \(V_T\), value for both nMOS and pMOS. Therefore, the same energy difference exists between the metal Fermi-level and the conduction and valence bands of the silicon substrate. This affords a simpler CMOS processing scheme, since only one mask and one metal would be required for the gate electrode. The second main approach toward metal electrodes involves two separate metals, one for pMOS and one for nMOS devices. Two metals could be chosen by their work functions such that their Fermi levels line up favorably with the conduction and the valence bands of silicon, respectively.

The deposition method in a fabrication process is another important factor to determine the final film quality. It must be compatible with current or expected CMOS processing, cost and throughput. Non-equilibrium conditions in deposition technique will certainly change the quality, properties and reproducibility of dielectric.

There are many challenges needed to overcome in order to continue scaling CMOS technology. Considering the possible technology roadmap, the device scaling is expected to continue through 15, 11 nm and beyond, while the challenge is fitting the device to smaller footprint. Silicon has the possibility to enable scaling to beyond 11 nm, while disruption in device architecture will be required. Fully depleted devices will be needed after the 22 nm node for the 15 and 11 nm nodes and beyond that silicon nanowires possibly could be used for the 8 and 5 nm nodes.
Chapter 2

Charge Carrier Statistics

2.1 Introduction

When a semiconductor is not perturbed for an extended period of time, it approaches thermal equilibrium condition. Disturbing this situation will typically cause a change in the number of charge carriers in the semiconductor such that excess carriers appear. Carrier lifetime is an elementary concept of paramount importance to characterize semiconductor materials and devices. This quantity is limited by the process of filling holes in the valance band by electrons in the conduction band resulting in annihilation of both carriers and a simultaneous emission of energy. The most typical appearances of this energy are heat or photon. However, thermal vibrations from lattice or photons can generate electron-hole pairs.

In this chapter, the statistics of the recombination and generation processes will be discussed. The concepts of emission rate and capture cross sections in relation to interface states and oxide traps will be also explained.

2.2 Recombination-generation processes

The statistics of the recombination of holes and electrons was first introduced by Shockley and Read in 1952 as often mentioned Shockley-Read statistics [14]. Carrier generation and recombination result from interaction between electrons and other carriers, either with the lattice of the material, or with photons. When the electron is transferred from one energy band to another, its lost or gained energy must appear in some other form. Depending on the form of energy, various types of generation and recombination processes can be distinguished such as Shockley-Read-Hall (SRH) processes and Auger recombination. However, these processes are generally sorted into two main categories without consideration of the energy exchange mechanism: direct transmission across the band gap and energy level assisted transmission.
through the band gap.

- **Band to band process**
  This recombination involves the direct annihilation of an electron in the conduction band and a hole in the valance band and eventually both carriers disappear. It is a typical radiative process in direct band gap semiconductors releasing the excess energy as a photon. This effect is the basis of light emitting diodes (LED) and lasers. When photons are present in the material, they can either be absorbed, generating a pair of free carriers, or they can stimulate a recombination event, producing a photon with the same properties as the one responsible for the event. Absorption is the active process in solar cells, photodiodes, and other semiconductor photodetectors, while stimulated emission is responsible for the action in laser diodes.

- **Shockley-Read-Hall (SRH) process**
  This process is also known as traps-assisted process in which the electron in transition between bands passes through an energy level (trap level) created within the band gap by an impurity in the lattice. The trap level has the possibility to exchange energy in the form of lattice vibration (phonon exchange thermal energy) with the material. Therefore, this process is the most dominant generation-recombination mechanism in silicon and other indirect band gap materials. It can also be valid for direct band gap materials with a very low carrier densities. The trap level could be in the midgap, introduced by impurity atoms.

- **Auger recombination**
  In this non-radiative process, a band to band recombination happens simultaneously with the collision between two carriers. The released energy is given to the survived carrier during collision which is excited to a higher energy level without moving to another energy band. This carrier normally loses its excess energy after the interaction to thermal vibrations. Since this process is a three-particle interaction, it is normally significant in non-equilibrium conditions at high carrier concentrations.

Any of these recombination mechanisms can be reversed leading to generate carriers. The critical issue in all cases of recombination-generation processes is the rate at which the various processes happen. The deep concern is about the dominant process, which is the one accomplishing at the fastest rate.

In semiconductors with indirect band gap like Si and Ge, not only the energy, but also the momentum of carrier changes during the transition. This
2.3. EMISSION RATE AND CAPTURE CROSS SECTION

means that lattice vibration (phonon) should also be involve in the process and consequently the transition occur at a slower rate compared to that of direct band gap semiconductors. As a comparison, the carrier lifetime of the band to band recombination for silicon is in the order of seconds, while GaAs as a direct band gap material has the corresponding quantity of about microseconds. The trap levels in the band gap increase the rate due to the extension of the carrier wave function in momentum space.

2.3 Emission rate and capture cross section

Quantifying the carrier traffic at a defect centers in the band gap requires knowledge about two statistical quantities called emission and capture rates.

- **Thermal emission rate**: $e_n, e_p$
  The quantity $e_n$ ($e_p$) is a measure of the number of emission events that an electron (hole) undergoes per unit time from an impurity energy level to the conduction band (valence band), if the impurity center is occupied by one electron. Unit: $[1/s]$.

- **Thermal capture rate**: $c_n, c_p$
  The quantity $c_n$ ($c_p$) is a measure of the number of capture events that an electron (hole) undergoes per unit time from the conduction band (valence band) to the impurity energy level, if there is one empty impurity per unit volume and the conduction band (valence band) contains one electron (hole) per unit volume. Unit: $[cm^3/s]$. 
Assuming that the concentrations of electrons in the conduction band and holes in the valence band are \( n_0 \) and \( p_0 \), respectively, and that \( n_T \) and \( p_T \) are the concentrations of occupied impurities by electrons and holes, respectively, one can express the electron traffic at thermal equilibrium as following:

\[
e_{n}n_T = c_{n}n_0p_T \tag{2.1}
\]

The standard expression used for \( n_0 \) is:

\[
n_0 = N_c \exp\left(-\frac{E_c - E_F}{k_B T}\right) \tag{2.2}
\]

where \( N_c \) is the effective density of states in the conduction band, \( E_c \) is the energy level of the conduction band edge, \( E_F \) is the Fermi-level, \( k_B \) is the Boltzmann constant, and \( T \) is absolute temperature. The total impurity concentration, \( N_T \), is the sum of occupied and empty impurities per unit volume.

\[
N_T = n_T + p_T \tag{2.3}
\]

The probability of filling an impurity level is determined by the Fermi-Dirac distribution

\[
\frac{n_T}{N_T} = \frac{1}{1 + \exp\left(\frac{E_T - E_F}{k_B T}\right)} \tag{2.4}
\]

where \( E_T \) is the energy level of the impurity. Combining Eqs. 2.1-2.4 will result in

\[
e_{n} = c_{n}N_c \exp\left(-\frac{E_c - E_T}{k_B T}\right) \tag{2.5}
\]

Using similar reasoning, the hole emission can be expressed as

\[
e_{p} = c_{p}N_v \exp\left(-\frac{E_T - E_v}{k_B T}\right) \tag{2.6}
\]

This treatment is a typical way mentioned in text books, but a more detailed derivation will be described in the next section.

When the equilibrium condition is perturbed, one can consider excess charge carriers of \( \Delta n \) and \( \Delta p \) in the conduction and valence bands, respectively. The net recombination rate, \( R_n \), for electrons from the conduction band to an energy level (trap level), \( E_T \), can be written as

\[
R_n = e_{n}n_T - e_{n}n_T = \frac{\Delta n}{\tau_n} \tag{2.7}
\]
where $\tau_n$ is defined as the electron life time, which is the average time one electron survives in the conduction band before being captured by the trap center. In a similar way for hole recombination from the valence band, one can write

$$R_p = c_{p}p_{nT} - c_{p}p_{T} = \frac{\Delta p}{\tau_p} \tag{2.8}$$

where $\tau_p$ is the hole life time. The recombination rate is positive if the recombination process is dominant and negative if generation process is dominant.

When analyzing a semiconductor device, mostly steady state or quasi steady state conditions are assumed. Under both steady state and equilibrium conditions, the average value of all macroscopic measurable parameters within a system are constant in time. For the latter, the static situation is maintained by the self-balancing of each fundamental process and its inverse. However for the former, the status is maintained by balancing between different processes. In a steady state situation, $R_n$ and $R_p$ have to be equal.

Quantities like capture and emission rates are statistical parameters based on average values taken across an ensemble of particles. The connection between statistical and quantum mechanics is done by the capture cross section. This is the area seen by a carrier which is captured into the recombination center. Assuming the thermal velocity (average velocity under equilibrium conditions) of $v_{th}$ for carriers and the capture cross section of $\sigma$ [cm$^2$], the capture rates are given by

$$c_{n,p} = \sigma_{n,p}v_{th} \tag{2.9}$$

where $\sigma_n$ and $\sigma_p$ are the electron and hole capture cross sections, respectively. This quantity is often used to define the effectiveness of the recombination-generation centers in capturing carriers.

### 2.4 Statistical background

Charge carrier statistics can be treated in a general way by fundamental thermodynamic terms to interpret experimental results. Different concepts such as free energy and entropy have been used in the literature to explain statistics of capture and emission of charge carriers at defects and impurities in semiconductor devices. In order to explain the background of this statistics, the canonical ensemble representing a probability distribution of microscopic states of the system is assumed.

The system as depicted in Fig. 2.2(a) has a number of subsystems and is isolated from the rest of the world. In addition, there is not any interaction...
between subsystems except thermal. The probability of finding subsystem \(i\) with energy of \(E_i\) (Fig. 2.2(b)) is given by the Boltzmann distribution

\[
P(E_i) = \frac{1}{Z} \exp\left(-\frac{E_i}{k_B T}\right) = \exp\left(-\frac{E_i - A}{k_B T}\right)
\]

(2.10)

where the quantities \(A\) and \(Z\) are constants for a particular ensemble, which ensure that \(\sum P_i\) is normalized to 1. Therefore, \(Z\), which is called the partition function of the canonical ensemble, is given by

\[
Z = \sum_i \exp\left(-\frac{E_i}{k_B T}\right) = \exp\left(-\frac{A}{k_B T}\right).
\]

(2.11)

Now a grand canonical ensemble is assumed which has the same isolation as before, but subsystems have the possibility to exchange particles, as shown in Fig. 2.2(c). Therefore, the partition function and the probability function can be written as

\[
Z = \sum_i \exp\left(-\frac{E_i}{k_B T}\right) \exp\left(\frac{\mu N_i}{k_B T}\right)
\]

(2.12)

and

\[
P(E_i, N_i) = \frac{1}{Z} \exp\left(-\frac{E_i}{k_B T}\right) \exp\left(\frac{\mu N_i}{k_B T}\right)
\]

(2.13)
respectively, where $\mu$ is the electrochemical potential, which is the free energy per particle, and $N_i$ is the number of particles in the subsystem $i$. It will be shown that $\mu$ will represent the Fermi-level in these treatments. In Eqs. 2.12 and 2.13, the term $\exp(-\frac{\mu}{k_BT})$ is called fugacity, which is the easiness of adding a new particle into the system. Considering an ensemble of identical subsystems where each can accept a maximum of $R$ particles distributed among $J$ possible states will result in the partition and probability functions as following

$$Z = \sum_{r=1}^{R} \sum_{i=1}^{J} \exp\left(\frac{E_{jr}}{k_BT}\right) \exp\left(\frac{\mu r}{k_BT}\right) = \sum_{r=1}^{R} Z_r \exp\left(\frac{\mu r}{k_BT}\right)$$

(2.14)

and

$$P(r) = \frac{Z_r \exp\left(\frac{\mu r}{k_BT}\right)}{\sum_{r=1}^{R} Z_r \exp\left(\frac{\mu r}{k_BT}\right)}$$

(2.15)

where similar to Eq. 2.11, $Z_r$ is

$$Z_r = \sum_{i} \exp\left(-\frac{E_{jr}}{k_BT}\right) = \exp\left(-\frac{A_r}{k_BT}\right)$$

(2.16)

Therefore, we have

$$Z = \sum_{r=0}^{R} \exp\left(-\frac{A_r - \mu r}{k_BT}\right).$$

(2.17)

Now, we can use this background and extend it to a simple defect system having energy states in a semiconductor. To start with, we can assume an energy level in the band gap of semiconductor which can exchange electron through conduction band. Using Eqs. 2.13 and 2.12 and knowing that for a single-electron trap $r$ is equal to 0 or 1, we get

$$P(1) = \frac{1}{1 + \exp\left(\frac{A_1 - A_0}{k_BT}\right) \exp\left(-\frac{\mu r}{k_BT}\right)}$$

(2.18)

We put $A_1 - A_0 = A_T$, thus

$$P(1) = \frac{1}{1 + \exp\left(\frac{A_T - \mu}{k_BT}\right)}$$

(2.19)

Eq. 2.19 is similar to the Fermi function and $\mu$ represents the Fermi-level. If the Fermi-level, $\mu$, is used as a probe to scan through the energy levels in the band gap of semiconductor, the probed energy is a free energy.
Let us suppose that energy levels are degenerate such that \( g_i \) states exist on each level. Thus using Eqs. 2.16 and 2.15 for \( i = 0, 1; r = 0, 1 \)

\[
P(1) = \frac{1}{1 + \frac{g_0}{g_1} \exp\left(\frac{E_{1,1} - E_{1,0} - \mu}{k_B T}\right)}
\]

(2.20)

Putting \( E_{1,1} - E_{1,0} = E_T \), Eq. 2.20 will end up in

\[
P(1) = \frac{1}{1 + \frac{g_0}{g_1} \exp\left(\frac{E_T - \mu}{k_B T}\right)}
\]

(2.21)

By comparing Eqs. 2.19 and 2.21, one can find that

\[
A_T = E_T - k_B T \ln\left(\frac{g_1}{g_0}\right)
\]

(2.22)

Relating the energy levels to the conduction band through \( \Delta E_n = E_c - E_T \) and \( \Delta A_n = E_c - A_T \), we can write

\[
\Delta A_n = \Delta E_n - k_B T \ln\left(\frac{g_0}{g_1}\right) = \Delta E_n - \Delta S_n T.
\]

(2.23)

where \( \Delta S_n \) is assumed to be equal to \( k_B \ln\left(\frac{g_0}{g_1}\right) \). Here we have a relationship similar to one between free energy and entropy. Up to this point, only the electronic part of the system has been involved which means that \( E_{j,r} \) and \( A_T \) are the energy eigenvalues and free energies of the pure electronic system, respectively.

Assuming an atom-electron trap system, one needs to take into account both atomic-vibrational and electronic parts. If the electron-vibrational coupling is weak such that Born-Oppenheimer and Franck-Condon approximations are valid, we have a decoupled system which can be divided into two independent systems, the electronic system and the atomic one. Fulfilling these two approximations allows the total wavefunction, \( \Psi_{\text{total}} \), to be broken into its electronic, \( \Psi_{\text{electronic}} \), and atomic (vibrational), \( \Psi_{\text{atom}} \), components.

\[
\Psi_{\text{total}} = \Psi_{\text{electronic}} \times \Psi_{\text{atom}}
\]

(2.24)

In addition, the energy eigenvalues, \( E_i \), can be expressed as the sum of the energy eigenvalues of the electronic system, \( E_c \), and the energy eigenvalues of the system of vibrating atoms, \( E_a \),

\[
E_i = E_c + E_a
\]

(2.25)

This shows that we have to take into account the vibrational eigenenergies of the trap, in addition to energy eigenvalues for the bare electronic system. One can consider the elastic energies involved in the trap vibration about
2.4. STATISTICAL BACKGROUND

a lattice coordinate (configuration coordinate) obtained as a transformation from the displacement coordinates. Modeling the vibrating center as a harmonic oscillator, its potential can be assumed parabolic as a function of the deviation of lattice coordinate from its equilibrium point. Then its energy eigenvalues become equally spaced.

Under these conditions, the atom and the electron belong to the canonical ensemble and the grand canonical ensemble, respectively. Therefore, the system is represented by two different partition functions, one for the atomic part, \( Z_a \), and one for the electronic part, \( Z_e \). The total partition function, \( Z_t \), can be written as

\[
Z_t = Z_a Z_e.
\]  

(2.26)

According to the previous discussions and using Eqs. 2.11 and 2.17, \( Z_a \) and \( Z_e \) can be written as

\[
Z_a = \exp\left(-\frac{G_a}{k_B T}\right)
\]  

(2.27)

and

\[
Z_e = \sum_{r=0}^{R} \exp\left(-\frac{G_{e,r} - \mu r}{k_B T}\right)
\]  

(2.28)

where \( G_e \) and \( G_a \) are the free energies of electron and atom, respectively. Thus,

\[
Z_t = \sum_{r=0}^{R} \exp\left(-\frac{G_t,r - \mu r}{k_B T}\right)
\]  

(2.29)

where \( G_t \) is equal to \( G_e + G_a \). Considering a bare electron system in which an electron trap is capable of capturing an electron, similar expression to Eq. 2.19 - 2.23 can be written. Hence, using Eqs. 2.5 and 2.23, the electron emission rate can be expressed as

\[
e_n = c_n N_e \exp\left(-\frac{\Delta G_n}{k_B T}\right)
\]  

(2.30)

where \( \Delta G_n \) is the free energy needed to take an electron from the trap and put it into the conduction band. As \( \Delta G_n \) is a free energy, it is expected to be temperature dependent. From thermodynamics, the entropy \( \Delta S_n \) is defined as

\[
\Delta S_n = -\frac{\partial}{\partial T} \Delta G_n.
\]  

(2.31)
Thus the defect can be associated with an enthalpy $\Delta H_n$ and we can use the following thermodynamic relation

$$\Delta H_n = \Delta G_n + \Delta S_n T.$$  \hfill (2.32)

This will result in an alternative expression for Eq. 2.30 as

$$e_n = c_n N_c \exp\left(\frac{\Delta S_n}{k_B T}\right) \exp\left(-\frac{\Delta H_n}{k_B T}\right)$$  \hfill (2.33)

Using Eq. 2.29, the change in entropy can be divided into two distinct parts

$$\Delta S_n = \Delta S_{na} + \Delta S_{ne}$$  \hfill (2.34)

where $\Delta S_{ne}$ comes from a change in electronic degeneracy and $\Delta S_{na}$ is the change in entropy because of the change in vibrational frequency of the atom when an electron is captured or emitted. $\Delta S_{ne}$ is

$$\Delta S_{ne} = k_B \ln \frac{g_0}{g_1}$$  \hfill (2.35)

where $g_0$ and $g_1$ are the electronic degeneracies of the trap when emptied and filled with an electron, respectively. This kind of entropy splitting can be done if the free energy can be distinctly divided into electronic and vibrational parts, which is valid if the energy eigenvalues can be partitioned in the way mentioned above [15].

For traps in semiconductors, the entropy $\Delta S_n$ is often temperature dependent for lower temperatures [16]. In practice for temperatures higher than about 100 K, it is often possible to assume a constant $\Delta S_n$, which means that $\Delta G_n$ is linearly dependent on temperature as realized from Eq. 2.31. The relation between the energy eigenvalue $\Delta E_n$, the enthalpy $\Delta H_n$ and the free energy $\Delta G_n$ is illustrated in Fig. 2.3. At $T = 0$, the free energy and energy eigenvalue coincide.

From Eq. 2.33 it can be concluded that in an Arrhenius plot of the emission rate for different temperatures, if the temperature dependencies of the pre-exponential factors are known, the slope of the activation graph gives the enthalpy value.

For the electron-trap system with the atomic energy eigenvalues given by a harmonic oscillator, the total energy (Eq. 2.25) can be represented by two parabolas in a configuration coordinate as depicted in Fig. 2.4. The lower parabola represents the energy of the system when the electron is captured into the trap, while the upper one represents the energy when the electron is emitted to the conduction band and the trap is empty. This means that the energy difference between the minima of the two oscillators is determined by
the electronic energy difference between the conduction band edge and the ground state of the trap center. Emitting an electron from the trap to the conduction band will consequently decrease the binding forces with the neighboring lattice atoms followed by a decrease in the curvature of the potential curve and thus in the vibrational frequency of the trap. Furthermore the vibrational energy levels will have smaller energy separation, i.e. with a higher density. If the change in binding forces is non-symmetric, the equilibrium point of the oscillator may change when the electron is emitted. This gives rise to a crossing point between the two atomic potential curves, where the atomic wave functions have a large overlap. This point has to be reached by the atomic part of the trap to achieve an electron transfer between the trap and the conduction band (Fig. 2.4). This in turn gives rise to the change in $\Delta S_{na}$ as discussed above.

There is a strong overlap between wave functions of the two oscillators at the intersection point and thus a transition may occur without exchanging en-
Figure 2.4 Configurational coordinate diagram of a vibrating electron-trap system. The parabolas represent the vibrational energies and are separated in energy by the electronic energy needed to take the electron from the trap to the conduction band. The lower curve represents the sum of electronic and vibrational energies of the trap when the electron is captured to it. The upper curve gives the corresponding energy when the trap is empty and the electron is in the conduction band. Transition of an electron between the trap level and the conduction band occurs at the crossing point where the atomic positions coincide for empty and filled trap. The separation in lattice coordinate depends on the change in the symmetry of atomic binding force when an electron is transferred.

energy depending on the probability of finding the trap center with the vibronic energy at the level of that point. In this model, the emission and capture processes occur due to moving up and down the trap oscillator through the capture and emission of phonons among its vibronic states. Therefore these transitions are mentioned as multi-phonon processes. The capture of charge carrier from conduction/valence band into the trap may be thermally activated with an activation energy determined by the position of the intersection point.

A basic treatment of the multi-phonon mechanism was presented by Henry and Lang to explain the nonradiative recombination processes at deep levels in GaP and GaAs [17]. It has been observed in most semiconductors and also for interface states at the SiO₂/Si interfaces [18–21]. In this work, it was found for interface states at the interface between silicon crystal and
high-\textit{k} oxides such as HfO\textsubscript{2}, Gd\textsubscript{2}O\textsubscript{3} and HfPrO as discussed in Papers B, C, D, E, F and G.

Another nonradiative process called \textit{phonon cascades} mechanism was introduced by Lax [22]. It is based on pure electronic potentials where the excited states are tightly separated in energy such that single phonons can be emitted in a cascade of electronic transitions. The probability of absorbing phonons in the cascade mechanism increases by raising temperature and follows a $T^{-n}$ function ($1 \leq n \leq 4$). A set of interface states close to the conduction band edge was found such that this mechanism is governing the charge carrier capture processes as presented in Papers E, F and G.
Chapter 3
HIGH-k DIELECTRICS

Tunneling currents decrease exponentially with increasing insulator thickness, which motivated the introduction of gate materials with dielectric constant, $k$, higher than 3.9 of SiO$_2$, known as high-$k$ materials [5, 23–26]. Replacing SiO$_2$ with a physically thicker layer of high-$k$ dielectric, the same capacitance but lower leakage can be achieved. The problem of the gate leakage current has been clear since the late 1990s [27], but the criteria to select an oxide were unknown. In late 2001, the choice narrowed to HfO$_2$ as the best candidate, but processing challenges raised. Finally for the first time in CMOS production, high-$k$ dielectric was employed in 2008 for the 45 nm ITRS node by using a "hafnium based" material [28].

This chapter deals with the concept of dielectric constant, $k$, and the most important requirements of a high quality high-$k$ dielectric. Furthermore, charge carrier traps and interface states for these oxides will be discussed.

3.1 Inside dielectrics

Using a simple electroscope and a parallel plate capacitor, Faraday discovered that the capacitance of such a capacitor is increased when an insulator is put between the plates. If the insulator fills the space between the plates, the capacitance is increased by a factor $k$ which depends only on the nature of the insulating material. This factor is then a property of the insulator and is called relative static permittivity or known as the dielectric constant. It is described as the permittivity of the insulator relative to the permittivity of vacuum, $\varepsilon_0$. Permittivity is determined by the ability of a material to polarize in response to an applied electric field. Thus, permittivity relates to a material’s ability to transmit or permit an electric field.

When an electric field $E$ is applied to a dielectric, it induces a dipole moment in the atoms. Assuming an average dipole moment per unit volume $P$, then the dielectric constant, $k$, is given by
\[ k = 1 + \frac{P}{\epsilon_0 E} \]  

(3.1)

If there are \( N \) atoms in a unit volume, the polarization \( P \) is expressed by

\[ P = N\alpha\epsilon_0 E \]  

(3.2)

where \( \alpha \) is the polarizability of the atom, which is a measure of how easy it is to induce a dipole moment in a molecule by an electric field.

In order to get an intuitive cognition of the polarization effect, a simplified model is demonstrated in Fig. 3.1. Consider a parallel plate capacitor in vacuum with a voltage \( V \) applied such that charges \( +Q \) and \( -Q \) occur on the surfaces of the conductors as in Fig. 3.1(a). Suppose that the spacing between the plates is \( d \) and the area of each plate is \( A \). This gives rise to the capacitance \( C = \epsilon_0 A/d \) and a constant electric field, \( E = V/d \), between the plates. Now assume an atom placed between the plates. Due to the electric field, its distribution of electrons \( -q' \) displaces with respect to the nucleus \( +q' \) as shown in Fig. 3.1 (b). This effect will decrease the electric field in the interval between \( -q' \) and \( +q' \). Upon constant applied voltage, the average electric field between the plates is still \( V/d \), therefore, the field outside the atom increases in order to compensate for the loss inside. This results in an increase of surface charges on the plates from \( Q = \epsilon_0 E \) to \( Q' = \epsilon_0 E' \) as shown in Fig. 3.1(b). Here \( E' = kE \) and \( k > 1 \), where the \( k \)-value depends on the magnitude of dipole charges, \( q' \), and the distance between them. This quantity represents the dielectric constant for this simplified case.

The situation of a dielectric is expected to be more complicated than what is mentioned above. There are three different mechanisms responsible for polarizability: [29] electronic, ionic, orientational (dipolar) polarizations as shown in Fig. 3.2. The electronic part comes from the displacement of the center of gravity of the negative charge in an atom relative to the nucleus which is from the deformation of electron shell surrounding the nucleus. The ionic contribution arises from the displacement and deformation of charged ions. The orientational or dipolar polarizability occurs when the material is built up of molecules possessing a permanent electric dipole moment which may be more or less free to change orientation in an applied electric field.

The polarization of typical high-\( k \) candidates for gate stacks mainly originates from electronic and ionic mechanisms. In case of transition and rare earth metal oxides, the ionic polarization is more dominant. Hence, the dielectric constant, \( k \), of these oxides depends not only on the oxide elements, but also on their molecular structure, whether crystalline or amorphous, ionicity and the shape of electron shells. Therefore, the \( k \)-value would be expected to be higher for materials with soft bonds. The relation between polarizability
and dielectric constant and also material structure are further discussed in Paper A.

### 3.2 Requirements

The main reason that microelectronics has used silicon among all semiconductors has been the innate excellent oxide, SiO$_2$. The advantage is that it can be made from silicon by thermal oxidation, while other semiconductors like Ge, GaAs, GaN, SiC, and etc have poor native oxides. Having amorphous structure, low density of electronic defects and an exceptional interface with silicon are great advantages of SiO$_2$. However, as mentioned above these advantages were abandoned to find a new material when it became so thin such that tunneling occur. A high-$k$ candidate for integration into MOSFET gate stacks should fulfill the following requirements in a time frame of 10 - 15 years to follow:
It is necessary to have a high enough $k$-value, preferably 25 - 30, to be reasonable as a replacement for $\text{SiO}_2$ for a number of years of scaling.

In order to limit the gate leakage current, a sufficiently large energy band offsets (over 1 eV) is demanded.

It must have a low density of states at high-$k$/silicon interface. For $\text{SiO}_2$/Si interface, it has been in the range of $10^{10}$ cm$^{-2}$, while $10^{12}$ - $10^{13}$ cm$^{-2}$ has been often reported for investigated High-$k$/Si interfaces so far. However, in the case of hafnium based dielectrics, densities less than $10^{11}$ cm$^{-2}$ have been achieved.

Low density of charge carrier traps in the bulk.

Thermodynamically stability in its direct contact with silicon channel and the gate metal.

Process compatibility to 1000 °C for dopant activation annealing must be fulfilled.
$k$-value and energy band offset are the first fundamental requirements for a potential candidate (Paper A). However, there is trade-off between these two parameters, because oxides with lower $k$-value often have higher offset values. In order to describe oxide-silicon combinations, similar reasoning as that for metal-semiconductor (MS) structures has been used in the literature [5] and is described in Paper A. The MS barrier has been found to be a function of electronegativity difference between the elements building up semiconductors [30]. Using the same reasoning for different oxide/Si structures, similar relations can be found. According to Pauling, the heat of molecular formation is linearly related to the square of the electronegativity difference between the molecular elements [31]. In Paper A, we found an approximately linear relationship between conduction band offset value of different high-$k$/Si structures and this quantity. One would expect that oxides with lower heat of formation would be mechanically softer and hence exhibit higher $k$-values. Plotting energy band offset values versus $k$-values as shown in Paper A confirms this conclusion. One finds that generally most of the transition metal oxides have a high $k$-value and low offset values, while the opposite is for the rare-earth oxides (Paper A). Considering the specifications defined by the International Technology Roadmap for Semiconductors (ITRS) [32], one finds that even if many oxides have a high-$k$ value, few can fulfill the preliminary requirements for the next generations of CMOS technology.

In order to have a good kinetic stability and withstand the processing temperatures in the range of 900-100 °C common in CMOS technology, an oxide with monocrystalline or amorphous structure is desired. The latter phase is preferred due to probable occurrence of lattice mismatch between the silicon crystal and most crystalline oxides. However, transition metal oxides and rare earth oxides are most often unstable and crystallize during common high temperature processes. Consequently a polycrystalline structure with grain boundaries acting as current leakage paths would occur [33]. The crystallization problem can be circumvented by alloying the desired oxide with an element tending to transfer the structure into a more glassy state [34–36]. Amorphous HfO$_2$, which turns to polycrystalline phase already at about 500 °C, retains stability against crystallization by adding Al or Si. The addition of nitrogen has been found to be beneficial in the case of HfO$_2$ [37,38]. However, adding an extra element into the oxide will change the $k$-value and offset value closer to the oxide of the added element [39]. Therefore, as hafnium based oxides like HfSiO or HfSiNO have already shown great potential for production, it is still an open question that how long they can be used on the ITRS road.

Beside HfO$_2$ and ZrO$_2$ as the most investigated candidates, La$_2$O$_3$ has attracted a lot of attentions due to its $k$-value of about 30 and offset en-
ergies higher than 2 eV. However, the drawback with this oxide is its high hygroscopicity. Surprisingly, when mixing Lu into La$_2$O$_3$, most of its hygroscopicity disappears and the $k$-value is still in the range of 30 [40], [41], [42], explained by a change in atomic structure [43].

Another challenge is the reactivity between high-$k$ oxides and silicon. A thin layer of SiO$_2$ between the oxide and the silicon crystal has commonly observed for both transition and rear earth metal oxides. This tendency depends on the heat of formation of the high-$k$ oxide with respect to that of the reaction product, which may be a silicate, a silicide or an oxide. The formation of SiO$_x$ layer in HfO$_2$ deposition has been avoided by high temperature growth [44], while a similar procedure for Y$_2$O$_3$ results in a silicate layer [45]. As discussed in Paper A, the heat of formation of rare earth oxides is larger as compared to that of transition metal oxides. Therefore, at high temperatures the balance between the formation of rare earth oxide and SiO$_2$ may result in a silicate, thus eliminating SiO$_2$ interlayer and enhancing the effective $k$-value of the stack. Such tricks have been used to avoid the SiO$_x$ interlayer and achieve a thin high quality dielectric [46–51]. The existence of a thin SiO$_2$ interfacial layer makes the interface properties very close to the well known Si/SiO$_2$ interface [52]. Fabrication of a high quality high-$k$/Si stack without employing an interlayer with the properties of SiO$_2$ is still a challenge for future CMOS scaling.

3.3 Interface states and bulk oxide traps

One of the key challenges for the successful implementation of a high-$k$ dielectric into the gate stack of MOSFETs is understanding, control and passivation of charge carrier traps. It has been shown that fabricating metal-gate/HfO$_2$/SiO$_x$/Si(100) stack including post deposition annealing (700-800 °C) and forming gas H$_2$/N$_2$ annealing (FGA) would result in interface state densities in the range of $2\times 10^{10}$ - $1\times 10^{11}$ cm$^{-2}$ as measured by charge pumping [53–55]. However, in order to investigate the origin and energy distribution of interface states in High-$k$/Si structure, the interface should be characterized before hydrogen passivation. Without FGA and high temperature post deposition annealing higher densities (about $1\times 10^{12}$ cm$^{-2}$) are generally reported, where a frequency dependent distortion in the capacitance-voltage (C-V) measurements and peaks in conductance-voltage (G-V) characteristics can be observed due to the interface states. Such results for sputtered HfO$_2$ and HfPrO are discussed in Chapter 6 and further investigations are presented in Papers B - G. Interface states in both the lower and upper portions of the silicon band gap are observed for a range of high-$k$ dielectrics such as:
3.3. Interface States and Bulk Oxide Traps

HfO$_2$, ZrO$_2$, La$_2$O$_3$, LaAlO$_3$, GdSiO$_x$, Lu$_2$O$_3$, Dy$_2$O$_3$, etc [48, 56–65]. The profile of interface state density across the band gap can provide further evidence about the origin of these states. One of the dominant sources is found to be silicon dangling bond defects ($P_{b0}/P_{b1}$) similar to what is observed for SiO$_2$/Si interface following hydrogen detachment [66]. This has been confirmed by electron spin resonance (ESR) analysis [67, 68]. A quantitative agreement between C-V analysis and density of $P_{b0}$ centers in the HfO$_2$/SiO$_x$/Si(100) structure has also been obtained [69]. This conclusion does not exclude other origins for high density of interface states such as Hf-Si bond at oxide/Si interface, isolated Hf atoms in the SiO$_x$ interfacial layer [70] or process contamination like C or Cl.

The frequency dispersion of capacitance as will be discussed in Chapter 6 has been used to investigate capture cross sections for charge carriers of the $P_b$ centers at the interface of HfO$_2$/SiO$_x$/Si, and Gd$_2$O$_3$/Si structures, as presented in Papers B and C. For these oxides, this quantity has been found to be thermally activated and increase exponentially with energy depth from the silicon conduction band edge. In the past 30 years, similar properties have been reported for SiO$_2$/Si interface without any explanation [71]. We understood these phenomena as governed by multiphonon mechanism, as discussed in Chapter 2.

Multiparameter admittance spectroscopy (MPAS), as explained in Chapter 6, has been developed and used on HfO$_2$/SiO$_x$/Si structures. Those results revealed a second set of interface states closer to the conduction band edge, beside the $P_b$ centers discussed above (Papers E - G). The capture cross sections of these states are found to have opposite dependencies on temperature and energy position compared to those of the $P_b$ centers. This indicates that a phonon cascade mechanism is governing the capture processes as further explained in Papers E - G.

The interface between the high-$k$ oxide and the SiO$_x$ layers cannot be considered as an abrupt interface as demonstrated in literature [39, 48, 72–75]. A transition region with undefined stoichiometry and possible structural instabilities exists. Therefore, this region is expected to be a source of charge carrier defects [76, 77]. These defects were characterized for HfO$_2$/SiO$_x$/Si systems by thermally stimulated currents, as explained in Chapter 6, where charge carrier traffic found to follow a combined thermal-tunneling process as presented in Paper D. Extremely small capture cross sections have been observed for these states [78, 79].

Compared to SiO$_2$, high-$k$ oxides are often poor glass formers and crystallize at high temperatures, which lead to high levels of charge carrier bulk traps. The most likely candidate for an intrinsic defect causing charge carrier traps in the bulk of transition metal oxides is commonly considered to be oxygen vacancy, which has similar properties as the E' center in SiO$_2$ [74–77].
In the case of HfO$_2$, it has been predicted to be an amphoteric center with four [76] or five [77] charge states ranging from double donor to double acceptor behavior with the latter states positioned in the range of 1-2 eV from the conduction band edge of HfO$_2$. Furthermore, the double acceptor level is argued to be connected with large lattice relaxation and strong electron-lattice interaction [77]. It has even been proposed that the oxygen vacancy has a negative-U center property [80].
Chapter 4
Silicon/Silicon Bonded Interfaces

Wafer bonding generally refers to a process by which two mirror-polished wafers adhere to each other and make an intimate contact at room temperature without the application of any macroscopic gluing layer or external force. They stick to each other via intermolecular forces such as van der Waals forces or hydrogen bonding. Wafer bonding is known as "direct bonding" and "fusion bonding", which was first presented in 1985 by Lasky [81]. By temperature treatment or introducing some monolayers of appropriate molecules into the interlayer, the bonding strength can be increased in such a way that it is comparable with the chemical bond and acts like a welding. Annealing the bonded wafers increases the bond strength up to the cohesive strength of the materials involved. However, in order to keep a low thermal budget, low temperature techniques (temperatures less than 400 °C) have been investigated to achieve similar bond strength.

This joining technique has shown a great potential in microelectronics and microsystems technologies. It has been utilized in the industrial production of sensors, light emitting diodes and semiconductor-on-insulator (SOI) substrates and is required to fabricate many heterostructures. The most significant limitation in prevalent bonding processes pertains to the high temperature required, limiting choices of materials and components. As a result of this problem, researchers have turned to bonding at low temperature enabling the formation of novel semiconductor structures for mechanical as well as electrical applications. Unfortunately, lack of fundamental knowledge of low temperature bonding procedures and resulting bond interfaces limits the range of applications and a detailed model for the bonding mechanism is still not available. The understanding of the bonding mechanism is demanded to control and optimize the low temperature bonding step.

Wafer bonding using plasma pre-treatments has attracted notice as a potential method for low temperature (LT) and room temperature (RT) wafer bonding [82–85]. The inherent advantages of LT plasma-assisted-wafer-bonding include: a low thermal budget during the whole bonding process, no
need for intermediate layers, good hermetic sealing [86], and a high bond strength. Bond strengths achieved at RT are comparable to those obtained for wet chemical cleaned bonded wafer pairs annealed at 600-800 °C and above [82,84,87–89].

Even if detailed models are still lacking for plasma bonded interfaces which are hydrophilic, a number of treatments can be found in the literature for different stages of the bonding between hydrophilic silicon surfaces [90,91]. For this case, water seems to play an important role in the initial bonding phase at RT, resulting in a low bonding energy governed by hydrogen bonds to OH-groups, which in turn are bonded to the native oxide of the silicon crystal. After high temperature treatment, hydrogen is assumed to diffuse out of the structure or into the silicon bulk leaving oxygen, which gives rise to high bonding energy by covalent bonds to the oxide. It is reasonable to assume that a plasma bonded interface has properties between these two extremes of the hydrophilic bonded structure.

Additional observations to enhance the understanding of the oxygen plasma enhanced silicon/silicon bonded structure at room temperature have been performed in Paper H of the thesis. The samples have a complex interface and detailed models are not valid to interpret the behavior of current-voltage (I-V) and capacitance-voltage (C-V) characteristics. Earlier models [92], [93] for describing the electrical properties of high-temperature bonded silicon interfaces cannot be used for the present samples. Hysteresis effect in the I-V and C-V characteristics, satellite peaks in the C-V characteristic and a pronounced kink in the I-V characteristic, which was not reported for high temperature bonded samples, are revealed. These effects probably are related to ion transport in the interfacial layer.

The semiconductor-insulator-semiconductor combination can be considered as two MOS structures connected gate to gate. Hence, similar treatments as those used for MOS structure were employed for interpretation of the interface state properties of the bonded samples. By comparing the evolution of mechanical bonding strength and interface electron state concentration during the first 48 hours after bonding, we have demonstrated a relation between the two quantities for room temperature plasma enhanced bonded silicon surfaces. Further analysis and discussions are presented in Paper H.
Chapter 5
SAMPLE PREPARATION

Two types of samples have been used in the investigations discussed in Papers B - H, metal oxide semiconductor (MOS) and silicon/silicon bonded structures. Here, an overview of the fabrication processes is given.

5.1 MOS fabrication

MOS structures with HfO$_2$, Pr$_2$O$_3$ or HfPrO as high-$k$ dielectrics were fabricated to investigate the electrical properties of these materials. Reactive sputtering technique was employed for deposition of the oxides. The process was developed to calibrate different parameters such as power, gas flows and pressure involved in the sputtering method to deposit a high quality oxide.

Silicon wafers ($100$), P-type and N-type with diameter of three and four inches were used. Some of the wafers endured diffusion on their backsides (N$^+$ for N-type and P$^+$ for P-type) to achieve good ohmic back contacts. Various deposition times were used to deposit oxides with different thicknesses for further investigations. Aluminium was sputtered as the metal gate and also back contact. In order to pattern gate contacts, photolithography was employed. The process steps used to prepare the samples are described in the following.

Wafer cleaning

The surface properties of the wafers have a direct effect on performance and reliability of the semiconductor devices. The cleaning techniques employed prior to fabrication processes on the wafer must be able to remove all contaminations on the surfaces without degrading surface roughness. Therefore, the wet-chemistry wafer cleaning was used.

The first treatment step is RCA standard clean 1 known as SC1. In this step, the wafer is exposed to a hot mixture of de-ionized-water (DIW) diluted hydrogen peroxide and ammonium hydroxide (H$_2$O:H$_2$O$_2$:NH$_4$OH,
5:1:1) at 80 °C for 10 minutes. This procedure is designed to remove organic surface films by oxidative breakdown and dissolution to expose the silicon or oxide surface for concurrent or subsequent decontamination reactions. Group IB and IIB metals and several other metals, including gold, silver, copper, nickel, cadmium, zinc, cobalt, and chromium, are dissolved and removed by the complexing effectiveness of ammonium hydroxide [94, 95].

The second treatment step is rinsing the wafer with DIW, dipping in 2% hydrofluoric acid (HF) for 30 seconds, and then again rinsing with DIW.

The third treatment step is RCA standard clean 2 so called SC2. In this step, the rinsed wafer is exposed to a hot mixture of DIW-diluted hydrogen peroxide and hydrochloric acid (H₂O₂:H₂O:HCl, 5:1:1) at 80 °C for 10 minutes. This procedure removes alkali ions, and cations such as Al⁺², Fe⁺³ and Mg⁺² that form NH₄OH-insoluble hydroxides in basic solutions. It also eliminates metallic contaminants that were not entirely removed by the first treatment, such as gold. Electrochemical displacement re-plating of heavy metals from the solution is prevented by formation of soluble complexes with the dissolved metal ions. [94, 96]

Finally the wafer is dipped in 2% HF, rinsed with DIW, dried and immediately loaded into the vacuum chamber of sputtering machine.

**Reactive sputtering deposition**

Sputtering is a well established technique to deposit thin films. It was invented in 1852 by Grove [97] and was developed by Langmuir in the 1920’s [98]. It has better step coverage than evaporation. Also in the magnetron variant, it gives rise to far less radiation damage than electron beam evaporation, and is much better at producing layers of compound materials and alloys. These advantages have made sputtering the standard metal film deposition technique in microelectronic fabrication.

Sputtering includes striking the surface of a target material with energetic ions (energy more than 100 eV) in order to create a cascade of collisions in the surface of target such that multiple collisions eject or sputter atoms from the surface. It is often used to deposit metal thin films by running the process in an inert gas (usually Argon). To grow a compound layer when using a metallic target, a reactive gas (e.g. oxygen or nitrogen) is added to the plasma. The reaction between the reactive gas and sputtered atoms on the substrate results in the formation of the compound layer. Compound formation can also occur on the target which changes the overall deposition process and results in complex behavior of the deposition parameters as a function of the reactive gas flow.

The FHR MS 150 machine, used in the present work, is a computer controlled sputter down tool for substrates up to 150 mm, equipped with load-lock and five 200 mm magnetron sputter cathodes, has been used for
deposition of HfO$_2$, Pr$_2$O$_3$ and HfPrO. Three of the magnetrons can be used for DC sputtering and two for RF sputtering. The substrate carriers can be cooled or heated as well as RF-biased. The tool is equipped with optical spectrometer and an advanced gas flow control system for reactive sputtering control. Process gases are Argon, Nitrogen and Oxygen. Pumping is done by turbomolecular pumps in both main and load-lock chambers, the base pressure of the main chamber is $1 \times 10^{-7}$ mbar.

In order to sputter HfO$_2$, Pr$_2$O$_3$ and HfPrO, metallic targets of hafnium (Hf) and/or praseodymium (Pr) were used. Furthermore, the process pressure of $1 \times 10^{-2}$ mbar, different powers in the range of 0.2 to 1 kW, gas flows of 50 and 10 sccm of Ar and O$_2$, respectively, and different deposition times were applied. In the case of HfPrO, multi-layers of thin HfO$_2$/Pr$_2$O$_3$ films were sputtered and annealed afterwards to mix HfO$_2$ and Pr$_2$O$_3$.

**Post oxidation annealing**

Gate dielectric properties strongly depend on the processing condition and the post-oxidation annealing. Although it is known that this annealing step in general improves the breakdown strength. However prolonged high-temperature annealing causes degradation in oxide and oxynitride film properties, such as enhancement in charge carrier trapping, increase in fixed oxide and interface trap charge density and decrease in radiation hardness.

A rapid thermal annealing (RTP) tool, STEAG SHS 100M, was used for annealing the samples. In the present studies, most of the sputtered High-$k$ oxides were annealed in N$_2$ ambient directly after deposition at the temperatures of 500 °C and 800 °C for 10 minutes.

**Metal gate fabrication and patterning**

For the HfO$_2$ MOS capacitors investigated in Papers B-G, an aluminum (Al) layer of about 0.6 µm was always sputtered on the oxides as the metal gate. Then, a photolithography process followed by Al-wet etching was employed to pattern metal gate contacts. Regarding photolithography, the resist AZ1512 at 3000 rpm for 30 seconds was spinined on the Si wafer and soft-baked for 1 minute at 100 °C. Then, a positive mask was loaded to the Suss MicroTec mask aligner KS MA/BA 6 to expose the resist for 4 seconds in a hard contact mode. After exposure, the resist was developed in 50% MF312 (or MF322) for 1 minute followed by hard baking in oven at 120 °C for 30 minutes. Afterwards, the wafer was treated with wet chemical Al-etcher to transfer the pattern. Finally, the resist was removed by aceton followed by DIW rinsing and drying.

**Back contact**

In most of the cases, Si wafers with diffused back side were used in order
Figure 5.1  Schematic structure of the fabricated MOS devices. High-k oxides like HfO$_2$, Pr$_2$O$_3$ and HfPrO have been deposited by reactive sputtering.

to have a good ohmic back contact. In addition, an aluminum layer of about 0.6 µm was always sputtered on silicon as the back contact.

**Post metallization annealing (PMA)**

Some of the samples endured PMA after gate electrodes were prepared. A Centrotherm furnace was used to anneal the samples in the range of 350-450 °C in N$_2$ ambient for 30 minutes. This step has been used in order to decrease density of interface state at the oxide/Si interface. In the PMA process, some of the hydrogen atoms in aluminum gate diffuse to the oxide/silicon interface and chemically reacts with the interface states, making them electrically neutralized.

These samples were investigated by different electrical techniques and also transmission electron microscopy (TEM) as will be discussed in the next chapter. A schematic view of the sample geometry is shown in Fig. 5.1.

### 5.2 Oxygen plasma enhanced Si/Si bonding

Si/Si bonded structures were fabricated by using N-type, six inch (100) silicon wafers of 1-10 Ω-cm resistivity. The wafers were cleaned in SC1 solution, as described above, and rinsed in de-ionized water (DIW). Then, they were dipped in 2% HF solution for 20 s, rinsed in DIW and spin-dried before oxygen plasma exposure in an STS Multiplex ICP Reactive Ion Etcher using a coil power of 800 W, a platen power of 20 W and a chamber pressure of 40 mTorr. This plasma treatment enhances the chemical reactivity of the silicon surfaces and increase the surface energy of the bonded structure. After
plasma exposure, the wafers were dipped in DIW and spin-dried. Then, they were manually placed on top of each other in clean room environment (Class 100). A slight touch on the center of the wafer pair initiates spontaneous bonding resulting in two bonded wafer pairs.

For electrical characterization of the bonded interface, one quarter of each wafer pair was cut into $5 \times 5 \text{ mm}^2$ pieces immediately after bonding. Then they were treated in a polishing etch $(\text{CH}_3\text{COOH}(100\%):\text{HNO}_3(69\%):\text{HF}(49\%))$, 4:10:1) for 3 minutes to avoid surface roughnesses and hence leakage currents. Finally, Al/Ga ohmic contacts applied to the samples. The final schematic structure of the samples used for electrical measurements is shown in Fig. 5.2.

![Schematic structure of the Si/Si bonded samples used for electrical measurements.](image)

For mechanical characterization, the remaining three quarters of the bonded wafer pairs were cut into stripes, 20 mm wide and 55-75 mm long. These samples were used to determine the surface energy by a razer blade test set-up as described by Maszara et al [99].
Chapter 6
MEASUREMENT TECHNIQUES

Here, different methods to characterize the physical and electrical properties of metal oxide semiconductor (MOS) capacitors with high-\textit{k} dielectrics and also Si/Si bonded structures are presented. New electrical methods in addition to the traditional ones have been developed for detailed investigations on the MOS structures and in particular on the oxide/semiconductor interface. The traditional capacitance, conductance and current measurements were employed as the basis of these methods but used in various conditions to get more comprehensive information about the electron states. In this chapter, the applied techniques are described and some of the results are presented.

6.1 Measurement setup

An HP4284A precision LCR meter was used for capacitance and conductance measurements of MOS structures. For current measurements, an HP4140B pA meter was mainly employed. In order to measure low-noise leakage currents, a Keithley 4200 semiconductor characterization system together with a Keithley 4200-PA pre-amplifier connected to a Cascade shielded probe-station was used. Current, capacitance and conductance were measured in a Leybold He-recycled cryostat with the possibility to regulate sample temperature in the range of 30 to 300 K. Automated data collection was arranged by using Labview programs.

Regarding current-voltage (I-V) measurements on the Si/Si bonded samples, a data sampler, SDS 200, in combination with a Keithley 427 current amplifier and a function generator in order to create linear voltage ramps of different rates were employed. Capacitance-voltage (C-V) measurements were performed by using an HP4284A precision LCR meter.
6.2 Capacitance-voltage measurements

Capacitance-Voltage (C-V) method is the most common technique to characterize various charges in MOS capacitors. It is a diagnostic tool that gives a fast overall impression of oxide/semiconductor interface quality. There are two different ways of measuring capacitance, small signal and quasi static. During a small signal C-V measurement, a small AC voltage in the range of $10 \text{ mV}$ superimposed on a DC voltage is applied to the structure at typical frequency range of $1 \text{ kHz} - 1 \text{ MHz}$. The resulting charge variation gives rise to a differential capacitance, which is the measured quantity. This type of measurement is the first traditional investigation performed on a MOS structure as typically seen in the literature. The second one is based on measuring current at constant ramp-rate of the applied voltage. Therefore, the measured current is proportional to the capacitance of the structure. It is often used for low frequency measurements when noise level is high in small signal method.

Fig. 6.1 demonstrates C-V data of Al/HfO$_2$/Si gate stack collected at different signal frequencies and room temperature. All the graphs are measured
from inversion to accumulation with a voltage scan lasting about 20 sec for the voltage range -1 to 2.5 V. Furthermore, these data were collected before forming gas annealing (FGA) or post metallization annealing (PMA) of the sample. The variation of the graphs by changing frequency, which typically appears as a shoulder, is a vertical shift and depends on the relation between frequency and emission rate of interface states at the oxide/Si interface (paper C). This shift is not a result of filling bulk oxide traps during subsequent C-V sweeps, because the Fermi-level is at the same energy position for a given voltage for all curves in the set and they have the same amount of horizontal shift.

When measuring at room temperature, a limited energy range of the band gap can be investigated. As shown in Fig. 6.2, keeping the signal frequency from the capacitance meter constant and changing the temperature for different voltage scans would result in a similar behavior, as seen in Fig. 6.1 the graph obtained when changing the signal frequency. This was found to be due to the thermal activation of the emission rate of the interface states as explained in details in papers B and C. Fig. 6.2 shows typical data
collected at 1 kHz for the temperature range of 300 K down to 50 K from the Al/HfO$_2$/Si gate stack. For this temperature range, the position of the Fermi-level varies about 0.25 eV in the silicon bulk of the present samples, with the resistivity of 1 - 10 $\Omega$-cm.

Some of the samples were annealed, through PMA process explained in the previous chapter, in order to decrease the density of interface states and improve the quality of gate stack. The C-V results are presented in Fig. 6.3 before and after PMA for low and high frequencies, 1 kHz and 1 MHz. The graphs before PMA have the typical frequency dependent shoulder, while it has disappeared for the graphs after PMA. One notices that low and high frequency graphs after PMA are almost on top of each other and are steeper compared to the data before PMA. This demonstrates a decrease in the density of the interface states at the HfO$_2$/Si interface due to annealing. The shape of these curves are very close to that of an ideal MOS structure.

According to a range of experimental investigations on gate stacks with high-$k$ dielectric, the most probable physical origin of interface states is the $P_b$-center (silicon dangling bond) [5, 52, 69], which is a center characteristic of the SiO$_2$/Si interface. It has been found that proper thermal annealing passivates these states [56]. When using aluminum as metal gate and performing
PMA, the hydrogen passivation mechanism is driven by the hydrogen present in the Al-film. More investigations on the properties of interface states and their capture mechanisms have been done in this work as presented in the attached papers.

6.3 Stepped capacitance-voltage measurements

Most of the high-$k$ dielectric candidates used so far appear to have a substantial amount of undesirable fixed charge. If it has a large density difficult to minimize, it will be a significant issue to obtain the desired device performance. The density of fixed charge in the oxide layer of a MOS structure is typically estimated by fitting an ideal theoretical C-V function to experimental data and then comparing the shift along the voltage axis. The voltage shift of a double voltage sweep C-V measurement can also provide information about the oxide traps available in the MOS system. The data collection is most often performed by a rather fast voltage ramp from inversion to accumulation and back to inversion without any waiting time at the end points as depicted by the two middle graphs in Fig. 6.4. A hysteresis appearing as a horizontal shift is commonly observed due to oxide trap charging during a measurement cycle. From this hysteresis, one can not accurately evaluate charge carrier exchange between oxide traps and the silicon band edges as often used in literature. The reason is that the time constants of charge carrier exchange processes for oxide traps is often longer than that of interface states. Therefore, depending on the waiting times chosen at each end-points of the sweeping cycles and also on the scanning rate, different magnitudes of hysteresis may observed. Slow surface states, oxide traps in a close neighborhood to the oxide/Si interface or/and creating new traps due to voltage stress may contribute to this effect.

When an N-type MOS sample is biased into inversion, the energy level of electron traps will be above the Fermi-level and become empty. After sweeping the bias to accumulation region, the trap level sits below the Fermi-level, and will become filled. If the emission rates of the filled traps are smaller than the sweeping rate, the captured negative charge behaves as a fixed charge and shifts the whole C-V graph to a higher positive voltage when sweeping back from accumulation to inversion. However, for interface states and voltage ramping times in the range of 30 s, this happens only for lower temperatures. Traps deep in the band gap have very small emission rates, which limits the electron traffic. In the case of gate stacks with high-$k$ dielectrics, a transition region with undefined stoichiometry and possible structural instabilities between the silicon crystal and the oxide has been
found to be additional source for oxide traps. The charge carrier traffic of these traps has been found to be a combined thermal-tunneling process as described in Paper D.

Figure 6.4 (a) Capacitance-voltage data of Al/HfO\(_2\)/Si gate stack at 100 kHz measured with 30 sec voltage sweeping cycle (two middle graphs) and by stepped C-V method (outer graphs). (b) Capacitance transient measured at 0.2 V, when waiting at +4 V and then stepping to 0.2 V. (c) Capacitance transient measured at 1.8 V, when waiting at 4 V (in accumulation) for 5 min and then stepping to 1.8 V. Taking the initial values of such transients for different gate voltages will result in stepped C-V data (outer graphs of (a)). T = 300 K

In order to identify a more accurate estimation of the oxide charge than
what we get from fast sweeping method the stepped C-V technique has been introduced. The idea is to measure the C-V data when as large portion as possible of oxide traps are empty or filled [100]. To approach this aim, one needs first to perform a quick C-V measurement on the sample in order to find out high enough accumulation and inversion voltages required for the next step. The two middle graphs in Fig. 6.4 were measured on a Al/HfO$_2$/Si structure by sweeping the gate voltage from -3 V (inversion) to +3 V (accumulation) and back to -3 V without any waiting time at the end points and with a constant rate such that the whole measurement took 30 seconds. The hysteresis between the two middle graphs is about 30 mV. It is clear that choosing -4 V and +4 V for this sample will satisfy large enough inversion and accumulation voltages, respectively.

Second step is to establish a sufficient waiting time in inversion. To do so, an arbitrary time is chosen for waiting in inversion, followed by a step to another applied voltage of the C-V curve preferably in the raising part of the graph close to accumulation, and measure the capacitance transient which is occurred with a fast decay due to charge emission from oxide/silicon interface states and a considerably slower decay due to emission from oxide traps. The initial value of the transient should be picked up. Then the same procedure with longer waiting times needs to be repeated. When the initial values are almost similar even if the waiting time has increased, one can assume that it is long enough for filling and emptying a large portion of traps.

The next step is applying the inversion voltage and keeping the situation for the chosen waiting time, then switching the voltage to a higher value and collect the capacitance transient. This routine needs to be repeated for each data point until it reaching the accumulation voltage. An example is shown in Fig. 6.4, where the outer left curve was measured at 100 kHz by assuming an inversion voltage of -4 V, a waiting time of 5 minutes and measuring the starting points of the capacitance transients. 2 min of a typical capacitance transient measured this way at 0.2 V is depicted in Fig. 6.4(b). The outer right curve in Fig. 6.4 was measured in a similar way, but this time by waiting at accumulation (+4 V) and stepping to different lower gate voltages. This would result in measuring capacitance while most of the oxide traps are filled during the whole procedure. 2 min of the capacitance transient measured at 1.8 V is shown in Fig. 6.4(c). The shift between the two outer C-V graphs in Fig. 6.4 gives a larger and more reliable value of the oxide charge density than that by the fast sweeping method (the two middle graphs) most often used in the literature. The concentration of elementary oxide charge estimated this way was in the range of $3 \times 10^{12}$ cm$^{-2}$.

The outer left curve is closer to the fast sweep graphs in the middle compare to the outer right one. This indicates that the rate of electron capture is larger than the corresponding rate for electron emission. In Paper D more
detailed analysis about results of this method and the capture mechanism of traps are presented.

6.4 Capacitance frequency spectroscopy

This technique is based on the frequency and temperature dependence of capacitance (as depicted in Figs. 6.1 and 6.2) to measure capture cross sections of interface states. The charge carrier traffic of the interface states at the oxide/silicon interface can be described by

\[
\frac{dn_T}{dt} = -e_n n_T + v_{th} \sigma_n (N_T - n_T) n_s
\]

where \( t \) is time, \( n_T \) is the concentration of captured electrons, \( e_n \) is the thermal emission rate of electrons from the interface state to the conduction band, \( \sigma_n \) is the capture cross section for electrons at the interface state, \( v_{th} \) is the average thermal velocity of electrons in the conduction band, \( N_T \) is the total concentration of the interface states and \( n_s \) is the concentration of electrons in the semiconductor conduction band at the interface. Further derivation of Eq. 6.1 as described in Papers B and C gives the dependence of the capacitance of the interface states, \( C_{it} \), on the thermal emission rate, which is a function of temperature (Eq. 2.33), and frequency through \( C_{it} \propto e_n/\sqrt{4e_n^2 + \omega^2} \). Here \( \omega \) is the angular frequency of the AC signal from the capacitance meter. However, \( C_{it} \) of the MOS structure can be considered in parallel with the semiconductor capacitance, \( C_s \), and in series with the oxide capacitance, \( C_{ox} \). This means that the total capacitance of the structure can be written as

\[
C = \frac{C_{ox}(C_s + C_{it})}{C_{ox} + C_s + C_{it}}
\]

The extraction of capture cross sections from frequency dependent C-V curves can be done by collecting the capacitance data at an specific applied voltage where the C-V shoulder has the highest vertical variation when changing frequency. Plotting these values versus frequency a damping effect is found as shown in Fig. 6.5. These data are collected from Fig. 6.1 at 0.2 V. At a high enough frequency, capture and emission of carriers from the interface states cannot follow the frequency, opposite to the situation at low frequencies where almost all of the interface states can contribute to the capacitance. Fitting theory to measured data, one needs to find the interface state concentration by fitting the height of the C-V shoulder at the lowest frequency. Hence, the capture cross sections can be extracted by fitting the
magnitude of the thermal emission rate, $e_n$. A detailed description of the method is given in Papers B and C.

This analysis is the same as the corresponding fitting procedure for the conductance method. However, capacitance measurement generally has the advantage of less sensitivity to leakage, which could be beneficial for investigations of non-ideal dielectrics in the development of high-$k$ oxides.

Experimental results for the capture cross sections of the electron states at the interface between the silicon crystal and high-$k$ dielectrics such as HfO$_2$ prepared by reactive sputtering, Gd$_2$O$_3$ prepared by molecular beam epitaxy (MBE) and Gd$_2$O$_3$ prepared by atomic layer deposition (ALD) are presented in papers B and C. The cross sections follow an exponential dependence on the energy position of the electron states in a similar way to the case of Si/SiO$_2$ interfaces. This qualitative similarity confirms that the physical properties of the interface states are similar for these oxides. Moreover, measuring the C-V data at low frequency for various temperatures, as shown in Fig. 6.2, and employing this technique, the electron exchange process is found to be thermally activated. These effects are based on the vibronic properties of the interfacial defect centers as described in Chapter 2.
6.5 Conductance method

The conductance method, developed by Nicollian and Goetzberger in 1967, is considered to be one of the most accurate and sensitive electrical techniques to investigate localized states at oxide/semiconductor interfaces [101]. It is a method often used to measure the density of interface states in the depletion and weak inversion region of the band gap and also the capture cross section of interface states. Similar to the C-V method, it is based on charge carrier capture and emission from the interface states to the semiconductor bands in which the equivalent parallel conductance of an MOS capacitor as a function of applied voltage and frequency is measured. The measured conductance is directly related to the density of interface states which is an advantage compared with C-V method.

In traditional conductance investigations, it was found that experimental data had a broader conductance peak compared to a theoretical derivation originally performed by Lehovec [102]. Among various interpretations proposed to explain the broadening, [103] the idea of local fluctuations of surface potential was motivated to fit theory to the experimental data. However, that treatment did not cover the energy variation in thermal emission, density of states and capture cross section within the thermally accessible energy region around the Fermi-level.

In Paper E, an improved theory based on Shockley-Read-Hall statistics [14, 15, 104], is introduced including thermal broadening on all energy dependent quantities. In addition, the energy dependence of both capture cross sections of charge carriers and density of interface states [15] are taken into account. For the oxide/semiconductor interface of Al/HfO₂/SiOₓ/Si investigated in this study, no assumption about varying lateral surface potentials was needed for an interpretation using the improved theory. A brief description of this reasoning, starting from the dynamics of charge carrier capture and emission in a bottom-up approach, is given in Paper E and a comprehensive theoretical background is derived in Paper F.

In order to assure that a majority of the interface states are taking part in capture and emission processes, the capacitance and conductance of a MOS structure are measured at room temperature and low frequency. Then the measured capacitance data are fitted by assuming an energy dependent distribution of interface states. Here a capture cross section is assumed which is large enough to fulfill capture and emission processes of all interface states. Furthermore, the obtained interface state distribution is used for fitting theory to measured conductance data in order to extract the energy dependent capture cross section of the interface states. Two different capture mechanisms, phonon cascade [22] and multiphonon [17, 18] processes, were found to govern charge carrier traffic in shallow and deep regions of the band gap,
6.6 Multiparameter admittance spectroscopy

Multiparameter admittance spectroscopy (MPAS) technique is a diagnostic tool developed from the conductance method to deliver more information regarding charge carrier states in the semiconductor structures. For MPAS measurements, capacitance and conductance are measured as a function of bias voltage for a large number of frequencies and temperatures. In Paper F, the full theoretical background is described together with comparisons between experimental data and theoretical conductance. Furthermore, in Paper G experimental results from Al/HfO$_2$/SiO$_x$/Si MOS structures, before and after annealing (PMA), are evaluated by MPAS. This method gives more detailed identification possibilities for interface states compared with the most common capacitance and conductance techniques.

In addition to what is presented in Papers F and G on Al/HfO$_2$/SiO$_x$/Si MOS capacitors, more experimental results done by MPAS are shown in Figs. 6.6 and 6.7. Capacitance and conductance data of Al/HfPrO/SiO$_x$/Si MOS structure as a function of bias voltage and temperature at the frequency of 1 kHz are depicted in Fig. 6.6. Both types of 3D and contour plots are shown such that the former makes the results more illustrative and latter is more convenient in order to investigate the temperature dependence.

Similar investigations on MOS structure of Al/HfO$_2$/SiO$_x$/Si which has endured PMA are depicted in Fig. 6.7. The measurements are performed at the signal frequency of 4 kHz from the capacitance meter. The effect of annealing is further investigated in Paper G for these type of samples. As discussed in Papers E, F and G, two different mechanisms are responsible for the charge carrier traffic at oxide/silicon interface.

6.7 Current-voltage measurements

The most important limiting element in scaling of SiO$_2$ was the leakage current through the oxide. Replacing SiO$_2$ with a high-$k$ dielectric is required to gain larger drive current compared to leakage current in transistors. Current-voltage (I-V) measurements may give information about the transport mechanism through the oxide layer and absolute leakage level. Common conduction mechanisms often observed in dielectrics are Shottky emission,
Poole-Frenkel emission, Fowler-Nordheim emission and/or direct tunneling. The main quantities determining the leakage mechanism are oxide thickness, energy band offsets oxide trap concentration and applied electric field across oxide. When comparing the leakage level of different dielectrics, the flat band voltage, $V_{FB}$, of the MOS structure is commonly used as a reference point when choosing the voltage at which the leakage currents are taken. The standard rule is to use $V_{FB} + 1$ for n-type and $V_{FB} - 1$ for p-type substrates. However, such a rule is not an accurate way to compare leakage data. The reason is that the varying relations between semiconductor surface charge
6.7. CURRENT-VOLTAGE MEASUREMENTS

Figure 6.7 Capacitance and conductance measured on a Al/HfO$_2$/SiO$_x$/Si MOS structure after PMA as a function of gate voltage and temperature for the frequency of 4 kHz. (a) and (b) show the capacitance in 3D and contour representations, respectively. (c) and (d) are the corresponding conductance measurements.

and applied voltage across oxide are not taken into account. At a given gate voltage for different dielectrics, applied voltage across oxide changes with $V_{FB}$ due to varying concentrations of oxide charge, interface state densities or metal-semiconductor work function differences. In Ref. [105], the following standard procedure is described in details which would give leakage data for better comparison between samples.

- Substrate resistivity of 1-10 Ωcm.
- Equivalent oxide thickness (EOT) of 0.7-1.5 nm.
• Density of interface states \( (D_{it}) < 5 \times 10^{13} \).
• \(-0.5 \text{ V} < V_{FB} < 0.5 \text{ V}\).
• For samples fulfilling the above four conditions, use gate voltage, \( V_G \), of \( 1.5 \text{ V} \) for n-type and \(-1.5 \text{ V} \) for p-type silicon substrate.

Here, EOT is the thickness of the high-\( k \) oxide in terms of its equivalent SiO\(_2\) thickness, as \( \text{EOT} = (3.9/k) \times \text{(physical thickness of High-} k) \). However, as the \( k \)-value of an oxide is often unknown, there is another expression for the thickness called capacitance equivalent thickness (CET). It is calculated from the accumulation capacitance of MOS structure by assuming an equivalent SiO\(_2\) layer for the dielectric.

The leakage current measurements based on above criteria for HfO\(_2\) and HfPrO are presented in Fig. 6.8 for MOS samples with different oxide thicknesses.

6.8 Thermally stimulated current

Thermally stimulated current (TSC) measurement is one of the earliest techniques used to study deep states in semiconductors [106, 107] and was later
Figure 6.9  TSC measurements of the Al/HfO$_2$/SiO$_x$/Si MOS structure before PMA. The linear temperature rate is 10 K/min for all the curves and two different waiting times, 1 and 5 minutes, in inversion were used. The positive voltage values in the legends are charging biases and the negative ones are discharging biases.

also developed for investigating MOS systems. In this method, for a MOS structure with N-type semiconductor a bias is applied bringing the sample into accumulation at room temperature such that all traps below the Fermi-level are filled. This is followed by a constant bias in accumulation, which is a charging bias, while cooling down to lower temperatures. Then, at a low temperature in the range of 50 K, the bias is switched to deep depletion, which is a discharging bias, and the temperature is increased linearly with time. Upon heating, the trapped carriers are released to the silicon conduction band and give rise to a current in the external circuit.

The amount of trapped charge is determined by the concentration of traps, and as the temperature is increased, all the trapped carriers are eventually released. This results in a current peak as a function of temperature. The temperature at which this peak occurs is related to the energy level of the trap while the area under the peak is related to the trap concentration. The theoretical background of current resulting from carrier emission at oxide traps and interface states of a MOS structure is described in paper D. The measured current can be used to extract activation energy and apparent
Figure 6.10  TSC measurements of the Al/HfO$_2$/SiO$_x$/Si MOS structure. The sample is annealed (PMA) at 350 °C for 30 minutes in N$_2$ ambient. The linear temperature rate is 10 K/min for all the curves and two different waiting times, 1 and 5 minutes, in inversion were used. The positive voltage values in the legends are charging biases and the negative ones are discharging biases.

capture cross sections of carriers.

Some of the TSC measurements of Al/HfO$_2$/SiO$_x$/Si capacitors are demonstrated in Figs. 6.9, 6.10 and 6.11. In these measurements, different waiting times in inversion at 50 K, charging and discharging biases and temperature ramping rates were used. In Fig. 6.9, the graphs representing one minute waiting time show one TSC peak at about 210 K. However, the other data measured after 5 minutes waiting time have not only the same peak at about 210 K with a lower amplitude, but also a second asymmetric peak with higher amplitude at about 130 K. Using the same waiting time, charging voltage and temperature ramp rate, but increasing the absolute discharging voltage will result in higher amplitude of TSC peaks showing that the portion of released carriers has enhanced. This phenomenon is due to moving more electron states above the Fermi level. Detailed discussions and interpretations are presented in paper D.

The same type of TSC measurements as in Fig. 6.9 are performed on a similar sample which endured PMA at 350 °C for 30 minutes in N$_2$ ambient, as presented in Fig. 6.10. Increasing the waiting time in inversion from one
to five minutes does not result in a second TSC peak. One single peak occurs in the range of 200-250 K. This demonstrates stabilization of charge carrier traps by PMA.

Fig. 6.11 shows TSC measurements at the same charging and discharging voltages for the similar Al/HfO$_2$/SiO$_{x}$/Si capacitors but before and after PMA. One notices that for the same measurement conditions, the amplitude of the current has decrease for the annealed sample. Since the area under the TSC curve is related to the trap concentration, annealing has lowered this concentration.

### 6.9 Physical characterization

Transmission electron microscopy (TEM) technique was used to investigate the interfaces between the silicon crystal and high-$k$ oxides. Fig. 6.12 demonstrates a cross section of the HfO$_2$/SiO$_{x}$/Si structure of the sputtered HfO$_2$ on the silicon wafer. The TEM picture shows that a SiO$_x$ interlayer of about...
1 nm occurs as a bright band on top of the silicon crystal. This layer is predominantly existing in HfO\textsubscript{2} gate stacks either due to the growth conditions of HfO\textsubscript{2} or to an intentional growth before HfO\textsubscript{2} deposition. A region consisting of about 2 - 3 broken planes of the silicon crystal is observed in the vicinity of the SiO\textsubscript{x} interlayer that contributes to the occurrence of interface states in the silicon band gap. This irregularity could be caused by the sputtering process and/or wet chemical cleaning processes. The electrical properties of these interface states are investigated in papers B - G.

Another domain, the "transition region", is noticed between the SiO\textsubscript{x} and the HfO\textsubscript{2} layers including an irregular atomic arrangement and a large concentration gradient of the mixed elements. This interface region is an expected source of unstable charge carrier traps where captured electrons interact with energy band states. These oxide traps are investigated by C-V and TSC techniques in paper D. The dark region beyond the transition region is HfO\textsubscript{2}.

Fig. 6.13 is a TEM picture of the HfPrO/SiO\textsubscript{x}/Si sample. This sample is annealed at 500 °C for 10 minutes in N\textsubscript{2} ambient directly after sputtering thin layers of HfO\textsubscript{2} and Pr\textsubscript{2}O\textsubscript{3} on silicon. The transition region and the SiO\textsubscript{x} interlayer are also observed for the HfPrO samples.
Elastic recoil detection analysis (ERDA) is a method used to obtain elemental concentration depth profiles in thin films. The sample to be investigated is exposed by an energetic ion beam experiencing an elastic nuclear interaction with the atoms of the sample as in Rutherford backscattering. The incident energetic ions typically have energy of MeV enough to kick out (recoil) the atoms being struck. An appropriate detector is needed to detect these recoiled atoms. The great advantage of ERDA is that all the atoms of the sample can be recoiled if a heavy enough incident beam is applied, thus a complete analysis of the sample is immediately available.

The results of employing ERDA for the sputtered HfPrO film, where the TEM picture is depicted in Fig. 6.13, is shown in Fig. 6.14. One can realize the confirmation of the transition region between the silicon crystal and HfPrO as found by TEM.
**Figure 6.14** ERDA data from the sputtered HfPrO, where the TEM picture is depicted in Fig. 6.13 (with courtesy of Anders Hallen).
Chapter 7

SUMMARY

In the search for an alternative high-\(k\) gate dielectric, the basic requirements such as \(k\)-value and energy band offsets to silicon were investigated. From physical perspective, \(k\)-value is determined by polarizability, \(\alpha\), and molecular volume, \(V_m\), in dielectric, which are related to the material structure. A strong relationship between \(k\)-value and mean atomic number, \(< Z >\), for the rare earth oxides and also for neighbors in period 5 and 6 of periodic table were found. This effect is related to \(V_m\) and \(\alpha\) factors. Considering the demands for \(k\)-value and energy band offsets, a few oxides were found to fulfills the requirements of bulk CMOS technology beyond the 22 nm node.

The Si/HfO\(_2\)/Al MOS capacitors prepared by reactive sputtering were characterized by capacitance frequency spectroscopy. The capture cross sections of the electron states at the Si/HfO\(_2\) interfaces have an exponential energy dependence and are also thermally activated similar to those for Si/SiO\(_2\) interfaces. These properties have the same physical origin and are based on a multiphonon mechanism. The same qualitative tendency was found for electron capture cross sections of Gd\(_2\)O\(_3\) prepared by ALD and MBE. The interface states seem to have \(P_b\) properties.

By developing and using a new measurements methodology, labeled as multiparameter admittance spectroscopy (MPAS), an additional quality of electron states at Si/HfO\(_2\) interfaces was found. Beside the \(P_b\) like centers with energy levels in the range of 0.3 eV, an independent set of states were found close to the silicon conduction band. The capture of electrons into these states was found to be governed by a phonon cascade mechanism, giving rise to opposite energy and temperature dependencies as compared with those of the \(P_b\) like centers. Also, the advantage of using MPAS as a diagnostic tool for MOS interfaces was demonstrated.

The complicated stoichiometry of high-\(k\) oxides close to the SiO\(_x\) interlayer was found to include a high concentration of electron traps, releasing electrons by a combined thermal and tunneling process. A model for extracting physical parameters of such defects was developed. The traps were found
unstable with extremely low effective electron capture cross sections.

Interfaces between plasma bonded silicon surfaces was found to include a thin insulating material, which motivated a model corresponding to two MOS structures connected "face-to-face". By using measurement techniques adopted from the MOS field, it was possible to conclude that drifting ionic charge bonding to interface states is connected with the increasing bonding energy as a function of storage time.
ACKNOWLEDGEMENT

First of all, I would like to thank my supervisor Professor Olof Engström for his stimulating support, encouragement, believing in me and for being enthusiastic. He has always inspired me in learning and his extensive knowledge has been a privilege to do research.

I would like to thank Johan Piscator for good collaboration in our group work. I also want to express my gratitude to all those who gave me the possibility to complete this thesis. The personnel in the Nanofabrication Laboratory have provided great assistance in various ways. In particular, I would like to thank Henrik Frederiksen for his elaborate help with sputtering process and being always supportive. Many thanks to Zonghe Lai for his excellent and qualitative TEM analysis. Thank you Göran Petersson, Johan Andersson, Örjan Arthunsson, Mahdad Sadeghi and Ulf Södervall for helping and teaching me the processing steps. I like to appreciate Einar Sveinbjörnsson for sharing his knowledge in processing and characterization of MOS capacitors.

I also want to thank all the people in our national and European projects for fruitful collaborations. I would like to appreciate all my colleagues and friends in Physical Electronics Laboratory, previous Solid State Electronics Laboratory, BioNano Systems Laboratory, and Microwave Electronics Laboratory for contributing to a nice working atmosphere.

Financial supports from the following projects is gratefully acknowledged. The European Networks of Excellence NANOSIL and SiNANO, the European Integrated Project PULLNANO, the SSF project NEMO and the MC2 project MC2SOI.

I would like to thank my family for their support and encouragement. Finally and in particular, I want to express my gratitude to my wonderful wife, Bahareh, for all inspiration, support and happiness that she has provided. I would like to dedicate my thesis to my wife and my parents.

Bahman Raeissi
Göteborg, Feb. 15, 2010
References


59


