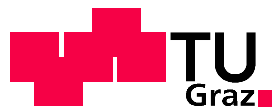
Gregor Pobegen, BSc

Fortgeschrittene elektrische Charakterisierung von durch NBTI erzeugten Steuerelektrodenoxiddefekten

MASTERARBEIT

zur Erlangung des akademischen Grades
Diplom-Ingenieur

Masterstudium Technische Physik



Technische Universität Graz

Betreuer:
Univ.-Prof. Ph.D. Peter Hadley
Institut für Festkörperphysik

Graz, Mai 2010

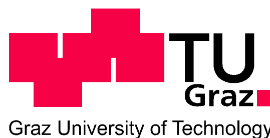
Gregor Pobegen, BSc

Advanced electrical characterization of NBTI induced gate oxide defects

MASTER THESIS

For obtaining the academic degree
“Diplom-Ingenieur”

Master Program Technical Physics



Graz University of Technology

Supervisor:
Univ.-Prof. Ph.D. Peter Hadley
Institute for Solid State Physics

Graz, May 2010

Deutsche Fassung:
Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008
Genehmigung des Senates am 1.12.2008

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am …………………………… ………………………………………………..
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

…………………………… ………………………………………………..
date (signature)

Kurzfassung

Die negative Spannung – Temperatur Instabilität ist ein wichtiger Degradationseffekt in Metall-Oxid-Halbleiter Feldeffekttransistoren. Das Thema ist eines der drängendsten Zuverlässigkeitsthemen der Halbleiterindustrie. Die vorliegende Arbeit beschäftigt sich mit dem Einfluss der Dicke des Steuerelektroden-Dielektrikums auf diesen Instabilitätseffekt.

Schon in den Siebzigern des vorherigen Jahrhunderts wurde die Instabilität von wichtigen Transistorparametern nach einer Belastungsperiode experimentell entdeckt. Die Belastung bestand aus einer großen negativen Spannung zwischen der Steuerelektrode und dem Rest des Transistors bei erhöhten Temperaturen. Bereits kurz nach der Entdeckung des Effekts wurden Modelle publiziert, welche den mikrophysikalischen Ursprung der Instabilität zu erklären versuchten. Die Modellentwicklung war aber weitgehend nicht erfolgreich, sodass bis heute kein vollständiges Modell existiert, welches den Großteil der empirisch erfassten Eigenschaften des Effekts erklären könnte. Genau aus diesem Grund ist es weiterhin wichtig die Abhängigkeit des Degradationseffekts von Bauteileigenschaften, wie der Dicke des Steuerelektroden-Dielektrikums, genau zu kennen um bestehende Modelle korrekt zu erweitern oder neue, innovative Modelle zu entwerfen.

Die vorliegende Masterarbeit fasst zu Beginn die theoretischen Grundlagen, die zu Bearbeitung des Themas nötig sind, zusammen. Danach werden experimentelle Methoden zur Bestimmung des elektrischen Zustandes des Transistors und Methoden für die Analyse der Auswirkungen des Effekts entwickelt. Mit diesen Mitteln wird schließlich die wissenschaftliche Fragestellung beantwortet, ob die Steuerelektroden-Oxiddicke ein relevanter Parameter für die Auswirkung des Effekts ist. Die Untersuchungen im Zuge dieser Arbeit brachten zu Tage, dass die negative Spannung – Temperatur Instabilität unabhängig von der Oxiddicke in dem untersuchten Bereich von 5 nm bis 30 nm ist. Außerdem ist es mithilfe von unterschiedlichen Oxiddicken möglich, eine Aussage über die räumliche Verteilung von Defekten, welche die Instabilität verursachen, zu treffen. Die Arbeit wurde an der KAI GmbH, ein Kompetenzzentrum für Automobil- und Industrieelektronik in Villach, verfasst. Die untersuchten Bauteile wurden von Infineon Technologies am Standort Villach gefertigt. Die Erkenntnisse der vorliegenden Arbeit wurden zur Präsentation bei der IRPS 2010, einer internationalen Zuverlässigkeitskonferenz, zusammengefasst und wurden im Konferenzbericht wissenschaftlich veröffentlicht.

Abstract

The negative bias temperature instability is an important degradation effect in metal oxide semiconductor field effect transistors and a major reliability concern in the semiconductor industry. This thesis elaborates the influence of the gate oxide thickness on the negative bias temperature instability.

Since the 1970s an instability of important electrical transistor parameters has been experimentally observed after subjecting silicon devices to a specific stress period. The stress was performed at elevated temperatures by applying a highly negative potential to the gate with respect to the other parts of the transistor. Soon after the experimental discovery, sophisticated physical models were developed which attempted to describe the microscopic origin of the instability. Numerous microscopic models have been developed up to the present date, however, a complete framework giving reasonable explanations for all of the experimentally observed characteristics of the effect is still missing. Consequently, exact knowledge about the dependence of the instability on device parameters like the oxide thickness is of great importance for the further development of existing models and the creation of new modeling.

This thesis starts with a summary of the theoretical fundamentals needed to introduce the topic. The following sections address the development of experimental techniques used to characterize the electrical situation within a transistor and to analyze the impact of the instability on the device performance. This allows answering the scientific question whether the gate oxide thickness is a relevant parameter for the instability effect or not.

It was found that the negative bias temperature instability is independent of the gate oxide thickness, at least for conventional stress conditions applied to a representative sample of devices having gate oxide thicknesses between 5 nm and 30 nm. This work further demonstrates that the use of different oxide thicknesses allows drawing conclusions on the spatial location of defects within the oxide which are responsible for the instability. The master thesis was developed at KAI (Kompetenzzentrum für Automobil- und Industrieelektronik GmbH) using hardware fabricated at Infineon Technologies in Villach. The main results of this thesis were presented at the International Reliability Physics Symposium 2010 and were published in the according proceedings.

Acknowledgment

I would like to express my sincere gratitude to a lot of persons which were partially involved into the work toward this thesis. I acknowledge the excellent work of Mag. Thomas Aichinger and thank for his generous support and kind guidance. I extend my thanks to Dr. Michael Nelhiebel for the supervision from Infineon Technologies Austria and Univ.-Prof. Ph.D. Peter Hadley from the Graz University of Technology for the academic supervision. I would also like to thank Ao. Univ.-Prof. DI Dr. Klaus-Tibor Grasser for the stimulating discussions throughout the thesis.

Contents

I Fundamentals **8**

1 The MOS structure 9
- 1.1 Composition of the MOS structure 9
- 1.2 Electrical behavior without external voltage 10
- 1.3 Applying external voltage 15
- 1.4 Discussion on interface traps and oxide charges 19
- 1.5 Impedance of the MOS structure 21
- 1.6 Carrier transport through the oxide 23

2 The MOS field effect transistor 26
- 2.1 Composition of a MOSFET 26
- 2.2 Mathematical modeling of the charge transport 27
- 2.3 Impedance of the MOSFET 28

3 The negative bias temperature instability 29
- 3.1 The Reaction–Diffusion model 29
- 3.2 The Grasser model 32

II Experimentals **37**

4 Experimental determination of transistor parameters 38
- 4.1 Gate oxide electric field 38
- 4.2 Gate poly doping concentration 46
- 4.3 Surface potential 48
- 4.4 Flat–band voltage 49
- 4.5 Acceptor/Donor doping concentration 51
- 4.6 Density of interface traps 53

5 Experimental methods for NBTI characterization 61
- 5.1 Experimental determination of the threshold voltage shift 61
- 5.2 Separating stress and recovery temperature 62

III Findings **66**

6 First order dependence of the NBTI on the gate oxide thickness . 67
- 6.1 Set-up of a comparative study 67
- 6.2 Dependence of the threshold voltage shift 73
- 6.3 Dependence of the interface trap creation 76
- 6.4 Dependence of the permanent degradation 77

7 Second order dependence of the NBTI on the gate oxide thickness 81
- 7.1 Voltage driven degradation mechanisms 81
- 7.2 Spatial defect distribution 86

Chapter I

Fundamentals

In this part of my thesis I will present all facts, formulas and models on which I will rely on in further argumentations. For a comprehensive treatment of semiconductor fundamentals and device physics I suggest the reader to read the excellent books [SN06] or [NB02].

I remark that in all subsequent explanations, as not clearly denoted otherwise, always the simplest physical model of a device is assumed. This means that all dimensions, defects, impurities or other parameters of a device are assumed to be such, that its performance is within the scope of the basic models.

1 The MOS structure

The thermally grown oxide of silicon has became the most important metal oxide semiconductor (MOS) structure in semiconductor industry. This is because of the straightforward fabrication technique and the fairly good electric properties. Figure 1 shows the chemical composition following thermal oxidation of a silicon substrate, covered by a metal layer above the oxide to form a MOS structure. Additionally, possible defects are drawn which will be addressed in more detail in the subsequent chapters.

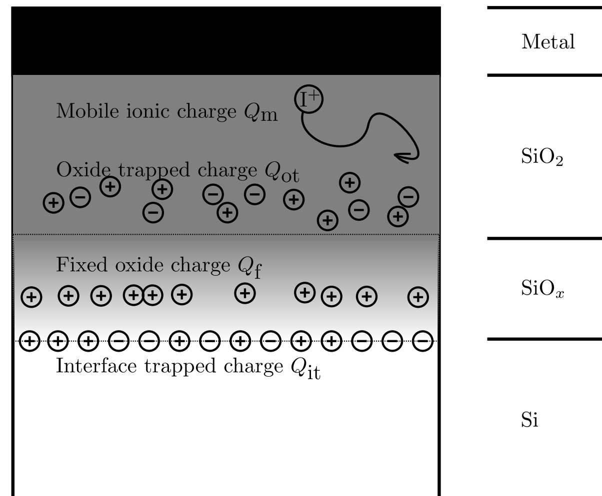


Fig. 1: Outline of various sources of defect charges possibly appearing in the MOS structure after [Dea80].

1.1 Composition of the MOS structure

Usually, semiconductor industry uses silicon (Si) crystals cut in such a way that the surface of the material is parallel to the {100} planes of the crystal. As a consequence, the surface consists of around 6.8×10^{14} atoms per square centimeter.

The bulk of pure silicon in the MOS structure might be doped by other materials. Approaching the interface, distortions of the periodic lattice structure will arise which may have interface traps as a consequence. At the Si–silicon dioxide (SiO_2) interface a mono–layer of incompletely oxidized Si forms [SN06].

The subsequent SiO_2 layer is amorphous as a consequence of thermal oxidation. On the other side of the SiO_2 today's semiconductor industry mainly uses highly doped polycrystalline silicon (poly).

1.2 Electrical behavior without external voltage

An important approach to describe semiconductor structures is a schematic drawing of energy over spatial dimension. It is called a *band diagram* because one draws the conduction band edge energy E_C and the valence band edge energy E_V together with other parameters as functions of a dimension parameter. In Fig. 2 the sequence poly–oxide–semiconductor without being in electrical contact is depicted. The energy position of the several bands are plotted using the electron

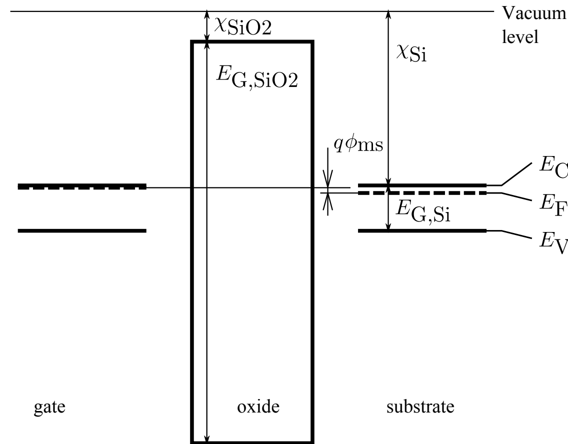


Fig. 2: Band diagram of a MOS structure without electric contact between the different materials. The gate is n^{++} doped poly and the substrate silicon is n doped.

affinity values $\chi_{Si} = 4.05\,eV$, $\chi_{SiO_2} = 0.9\,eV$ and the energy gap values $E_{G,Si} = 1.12\,eV$, $E_{G,SiO_2} = 9\,eV$. The values for the band gaps of Si and SiO_2 are given for room temperature. For arbitrary temperatures, the band gap of Si may be approximated as

$$E_G(T) = E_G(0) - \frac{\alpha T^2}{\beta + T}, \tag{1}$$

where $E_G(0) = 1.169\,eV$, $\alpha = 4.9 \times 10^{-4}\,eV/K$ and $\beta = 655\,K$ for Si [SN06].

Different doping of the materials on either side of the oxide leads to different FERMI level positions of the materials. A good overview to this is given in Fig. 3 for Si over a wide temperature range.

When the materials are capable of carrier exchange the FERMI levels of the materials align. So the resulting potential situation depends on the initial FERMI levels of the materials. In particular, the difference between the work functions of the poly gate and the semiconductor substrate ϕ_{ms} determines the band bending in the MOS structure provided no external bias applied. ϕ_{ms} is negative when the FERMI level of the gate material lies higher than the FERMI level of the silicon substrate. In this case the built–in potential drives electrons

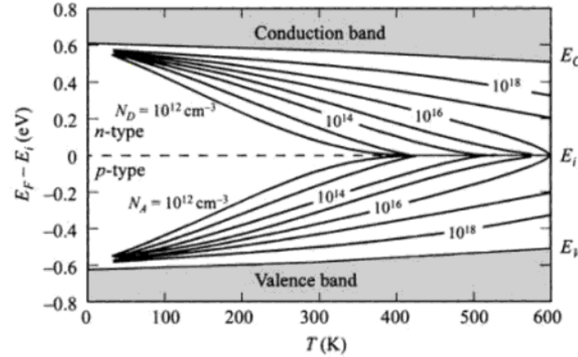


Fig. 3: FERMI level for doped Si as a function of temperature and impurity concentration. Illustration copied from [SN06, p. 27].

from the gate poly to the silicon substrate. Neglecting any oxide or interface charges, the initial difference in the work function of the materials will result in a voltage drop across the oxide V_{ox} and a band bending in the semiconductor ψ_{S}, $V_{\text{ox}} + \psi_{\text{S}} + \phi_{\text{ms}} = 0$, as depicted in Fig. 4.

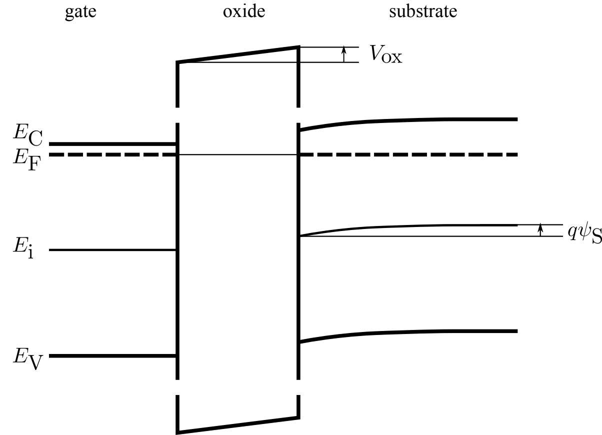


Fig. 4: Band diagram of a MOS structure with balanced FERMI levels throughout the whole solid state.

I remark that every potential ϕ in a semiconductor can be defined by

$$\phi(x) = \frac{E_{\text{F}} - E_{\text{i}}(x)}{q}, \tag{2}$$

the difference between the intrinsic energy (dependent on a spatial dimension x) and the FERMI energy divided by the elementary charge q. The zero of energy is chosen such that the potential is zero when the semiconductor is undoped, that is to say the FERMI level coincides with the intrinsic energy. Consequently, n doped materials have a positive potential, while the potential of p doped materials is negative. The unit of a potential is Volts (V). The intrinsic energy

level is the FERMI level of the undoped semiconductor. $\phi_{\mathrm{Bn}} > 0$ is the potential in the bulk of a n doped semiconductor and $\phi_{\mathrm{Bp}} < 0$ is the potential in the bulk of a p doped semiconductor, respectively. Consequently, the general case without any reference to the type of doping is ϕ_{B}. ϕ_{S} is the potential at the Si–SiO_2 interface.

A band bending ψ is defined by

$$\psi(x) = \phi(x) - \phi_{\mathrm{B}}. \tag{3}$$

Consequently, $\psi_{\mathrm{S}} = \phi_{\mathrm{S}} - \phi_{\mathrm{B}}$ is the total band bending within the semiconductor substrate. Note that ψ_{S} is often called the *surface potential* although it differs from the actual surface potential ϕ_{S} by the potential ϕ_{B}, due to the intensity of doping in the bulk of the substrate. ψ_{n} / ψ_{p} usually refers to the band bending of n doped / p doped materials at a certain position x.

Due to the excess of electrons in a n doped semiconductor, holes are called *minority carriers* while electrons are referred to as *majority carriers*. Naturally, holes are the majority carriers and electrons the minority carriers in p doped materials.

Mathematical solution for the charges in an ideal MOS structure

The band bending and the overall potential situation in a MOS structure can be calculated by solving the one–dimensional POISSON equation

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} = -\frac{\rho(x)}{\varepsilon_{\mathrm{Si}}}. \tag{4}$$

In equation (4) ρ is the charge density as a function of spatial dimension x and $\varepsilon_{\mathrm{Si}}$ is the permittivity of Si. A one–dimensional mathematical treatment neglects the influence of the fringing field at the border of the gate. The total charge density is composed of immobile ionized donors and acceptors and mobile holes and electrons,

$$\rho(x) = q\left(p(x) - n(x) + N_{\mathrm{D}}^{+} - N_{\mathrm{A}}^{-}\right), \tag{5}$$

where $n(x)/p(x)$ is the electron/hole density at the point x per cm^3. $N_{\mathrm{D}}^{+}/N_{\mathrm{A}}^{-}$ is the ionized donor/acceptor density per cm^3, assumed to be independent of x (uniform doping profile).

The carrier concentration in a semiconductor can be calculated by integrating the density of carriers with respect to energy weighted by an appropriate occupancy function. The density of free electrons in a semiconductor is

$$n = \int_{E_{\mathrm{C}}}^{\infty} N(E) f(E) \mathrm{d}E. \tag{6}$$

Electrons have half–integer spin and thus the PAULI exclusion principle holds. For arbitrary temperatures, the occupancy is given by the FERMI–DIRAC distribution function

$$f(E) = \frac{1}{1 + \exp\left(\frac{E - E_{\mathrm{F}}}{kT}\right)} \tag{7}$$

(k is the BOLTZMANN constant in eV/K). The density of states can be approximated with an effective density of states of electrons in the Si conduction band

of $N_{\mathrm{C}} = 2.8 \times 10^{19}\,\mathrm{cm}^{-3}$ [SN06](at room temperature) to

$$N(E) = \frac{2N_{\mathrm{C}}}{\sqrt{\pi}kT}\sqrt{\frac{E-E_{\mathrm{C}}}{kT}}. \tag{8}$$

The Fermi–Dirac integral is

$$F_{1/2}\left(\frac{E_{\mathrm{F}}-E_{\mathrm{C}}}{kT}\right) = \int_{E_{\mathrm{C}}}^{\infty} \frac{\left(\frac{E-E_{\mathrm{C}}}{kT}\right)^{1/2}}{1+\exp\left(\frac{E-E_{\mathrm{F}}}{kT}\right)}\frac{\mathrm{d}E}{kT} \tag{9}$$

and plotted in Fig. 5. The integral can be approximated fairly well with

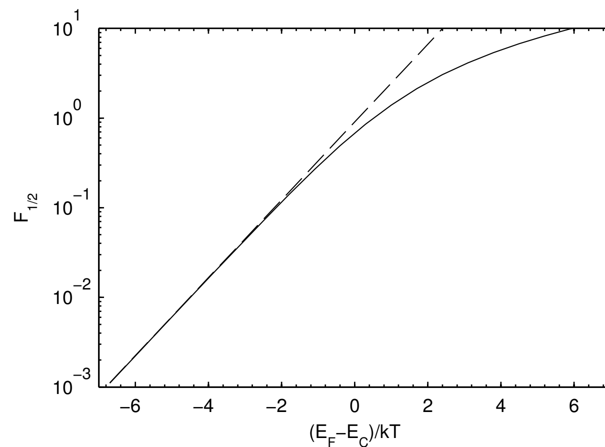


Fig. 5: Value of the Fermi–Dirac integral (solid line). Additionally plotted is the common approximation, $\frac{\sqrt{\pi}}{2}\exp\left(\frac{E_{\mathrm{F}}-E_{\mathrm{C}}}{kT}\right)$ (dashed line).

$\frac{\sqrt{\pi}}{2}\exp\left(\frac{E_{\mathrm{F}}-E_{\mathrm{C}}}{kT}\right)$ whenever $E_{\mathrm{C}} - E_{\mathrm{F}} > kT$. Consequently,

$$\begin{aligned} n &= \int_{E_{\mathrm{C}}}^{\infty} N(E)F(E)\mathrm{d}E \\ &= \frac{2N_{\mathrm{C}}}{\sqrt{\pi}}F_{1/2}\left(\frac{E_{\mathrm{F}}-E_{\mathrm{C}}}{kT}\right) \\ &\approx N_{\mathrm{C}}\mathrm{e}^{\frac{E_{\mathrm{F}}-E_{\mathrm{C}}}{kT}}, \end{aligned} \tag{10}$$

when the Fermi level lies at least three kT (around 78 mV at room temperature) below the conduction band. One can also refer the density of free electrons at a position x to the density in the bulk of the semiconductor (assumption $x \longrightarrow \infty$) $n(\infty)$,

$$n(x) = n(\infty)\mathrm{e}^{\frac{q\psi}{kT}}. \tag{11}$$

The approximation $F_{1/2} \approx \frac{\sqrt{\pi}}{2}\exp\left(\frac{E_{\mathrm{F}}-E_{\mathrm{C}}}{kT}\right)$ is a Boltzmann distribution of classical statistics. Whenever the electron concentration n is around a hundred times smaller than the effective density of states N_{C} the convenient Boltzmann distribution approximates fairly well and equation (10) may be used. That is to say, whether to use classical statistics or quantum mechanical statistics is

insignificant. A semiconductor which is doped such that the FERMI level lies at least one kT below E_C, i.e. the doping concentration is smaller than N_C, is called a nondegenerate semiconductor.

A very similar derivation yields

$$\begin{aligned} p &= \frac{2N_\mathrm{V}}{\sqrt{\pi}} F_{1/2}\left(\frac{E_\mathrm{V}-E_\mathrm{F}}{kT}\right) \\ &\approx N_\mathrm{V}\mathrm{e}^{-\frac{E_\mathrm{F}-E_\mathrm{V}}{kT}}, \end{aligned} \tag{12}$$

with $N_\mathrm{V} = 2.65 \times 10^{19}\,\mathrm{cm}^{-3}$ for Si [SN06] and consequently

$$p(x) = p(\infty)\mathrm{e}^{-\frac{q\psi}{kT}}. \tag{13}$$

By combining equations (10) and (12) the intrinsic (undoped semiconductor) energy level is

$$E_\mathrm{i} = \frac{E_\mathrm{C}+E_\mathrm{V}}{2} + \frac{kT}{2}\ln\left(\frac{N_\mathrm{V}}{N_\mathrm{C}}\right) \tag{14}$$

and thus near, but not exactly at, the middle of the bandgap. Because $kT \approx 26\,\mathrm{mV}$ at room temperature and $\ln(N_\mathrm{V}/N_\mathrm{C}) \approx -0.0551$ the intrinsic energy level is around 1 mV away from the middle of the bandgap at room temperature for Si.

With all that I can continue solving the one dimensional POISSON equation (4). In the bulk (assumption $x \longrightarrow \infty$) of the semiconductor, charge neutrality $\rho(\infty) = 0$ must exist. Therefore, as a result of equation (5),

$$N_\mathrm{D}^+ - N_\mathrm{A}^- = n(\infty) - p(\infty) \tag{15}$$

holds and the POISSON equation (4), combined with equations (11) and (13), becomes

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} = -\frac{q}{\varepsilon_\mathrm{Si}}\left(p(\infty)\left(\mathrm{e}^{-\frac{q\psi}{kT}}-1\right) - n(\infty)\left(\mathrm{e}^{\frac{q\psi}{kT}}-1\right)\right). \tag{16}$$

This equation can be solved by a little trick, using the chain rule

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}\psi}{\mathrm{d}x}\right)^2 = 2\frac{\mathrm{d}\psi}{\mathrm{d}x}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}. \tag{17}$$

The electric field is given by $\mathcal{E} = -\mathrm{d}\psi/\mathrm{d}x$, so it follows that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}\psi}{\mathrm{d}x}\right)^2 &= -\frac{2q}{\epsilon_\mathrm{Si}}\frac{\mathrm{d}\psi}{\mathrm{d}x}\left(p(\infty)\left(\mathrm{e}^{-\frac{q\psi}{kT}}-1\right) - n(\infty)\left(\mathrm{e}^{\frac{q\psi}{kT}}-1\right)\right) \\ \int_0^{\mathcal{E}_\mathrm{S}} \mathrm{d}\mathcal{E}^2 &= -\frac{2q}{\epsilon_\mathrm{Si}}p(\infty)\int_0^{q\psi_\mathrm{S}}\left(\mathrm{e}^{-\frac{q\psi}{kT}} - 1 - \frac{n(\infty)}{p(\infty)}\left(\mathrm{e}^{\frac{q\psi}{kT}}-1\right)\right)\mathrm{d}\psi \\ \mathcal{E}_\mathrm{S}^2 &= \frac{2kT}{\epsilon_\mathrm{Si}}p(\infty)\underbrace{\left(\mathrm{e}^{-\frac{q\psi_\mathrm{S}}{kT}} + \frac{q\psi_\mathrm{S}}{kT} - 1 + \frac{n(\infty)}{p(\infty)}\left(\mathrm{e}^{\frac{q\psi_\mathrm{S}}{kT}} - \frac{q\psi_\mathrm{S}}{kT} - 1\right)\right)}_{\equiv F^2\left(\frac{q\psi_\mathrm{S}}{kT},\frac{n(\infty)}{p(\infty)}\right)}, \end{aligned} \tag{18}$$

where F is defined as

$$F\left(\frac{q\psi_{\mathrm{S}}}{kT},\frac{n(\infty)}{p(\infty)}\right)\equiv\operatorname{sgn}\left(\psi_{\mathrm{S}}\right)\sqrt{\mathrm{e}^{-\frac{q\psi_{\mathrm{S}}}{kT}}+\frac{q\psi_{\mathrm{S}}}{kT}-1+\frac{n(\infty)}{p(\infty)}\left(\mathrm{e}^{\frac{q\psi_{\mathrm{S}}}{kT}}-\frac{q\psi_{\mathrm{S}}}{kT}-1\right)}. \tag{19}$$

Sometimes the extrinsic DEBYE length for holes in the bulk of the semiconductor

$$L_{\mathrm{D}}=\sqrt{\frac{kT\epsilon_{\mathrm{Si}}}{p(\infty)q^2}} \tag{20}$$

is used to express the electric field at an arbitrary position x,

$$\mathcal{E}(x)=\frac{\sqrt{2}kT}{qL_{\mathrm{D}}}F\left(\frac{q\psi(x)}{kT},\frac{n(\infty)}{p(\infty)}\right). \tag{21}$$

The charge in the semiconductor per area Q_{S} is linked to the electric field by $Q_{\mathrm{S}}=-\varepsilon_{\mathrm{Si}}\mathcal{E}_{\mathrm{S}}$. So this derivation gives the relationship between the surface potential ψ_{S} and the semiconductor charge Q_{S}, which is depicted in Fig. 6.

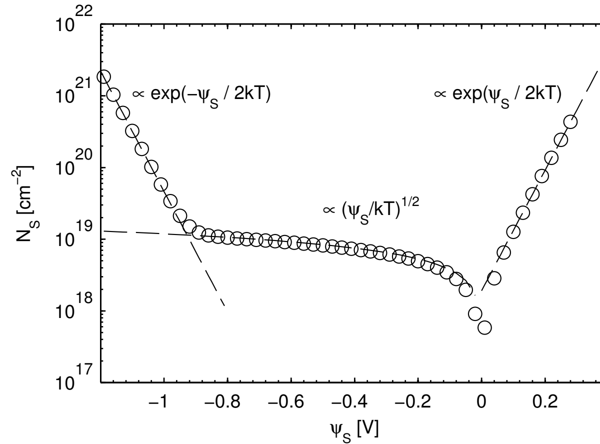


Fig. 6: Relationship between the semiconductor charge $N_{\mathrm{S}}=|Q_{\mathrm{S}}|/q$ and the surface potential ψ_{S} after equation (21) (circles). Assumed is a n type silicon with a fully ionized donor doping of $N_{\mathrm{D}}=N_{\mathrm{D}}^{+}=10^{17}\,\mathrm{cm}^{-3}$. The dashed lines indicate the values of convenient approximations in the different regions. Mind that an absolute number of charges is plotted. The sign of the according charge is positive where ψ_{S} is negative and vice versa.

The surface potential and the overall potential situation in a MOS structure can be changed by applying an external voltage to the gate with respect to the substrate.

1.3 Applying external voltage

As mentioned in chapter 1.2, a MOS structure might have internal voltage drops because of different work functions of the materials on either side of the oxide because of different doping levels. When the gate is a n^{++} highly phosphorous

doped poly and the substrate is lightly n doped silicon, the resulting work function difference ϕ_{ms} is negative (neglecting any oxide or interface charges). As a consequence, an external voltage V_G=V_{FB}= $\phi_{ms} < 0$ (flat–band voltage) has to be applied to the gate with respect to the substrate to nullify the internal voltage drops and flatten all bands again ($\psi_S = 0$). At flat–band condition, the carrier concentration at the interface is determined by the carrier concentration in the bulk of the semiconductor ($\phi_S = \phi_B$). When the substrate is n doped electrons will be the majority carriers at the interface. The flat–band condition in Fig. 6 is reached when the space charge density is minimal. Applying further negative or positive voltage to the gate will have a strong influence on the carrier concentration at the interface. Voltage applied to the gate will drop across the oxide and the semiconductor. The voltage applied to the gate can be written as

$$V_G = V_{FB} + V_{ox} + \psi_S. \tag{22}$$

To characterize the carrier situation in the MOS structure one differs between roughly three regions according to the voltage applied to the gate.

Accumulation

For a n^{++} gate / n substrate MOS structure the application of V_G greater than $V_{FB} < 0$ causes an accumulation of majority electrons at the interface. The energy bands in the Si substrate will experience further (positive) band bending downwards, as depicted in Fig. 7. The density of majority electrons n at the interface will greatly exceed the number of free electrons in the bulk of the substrate.

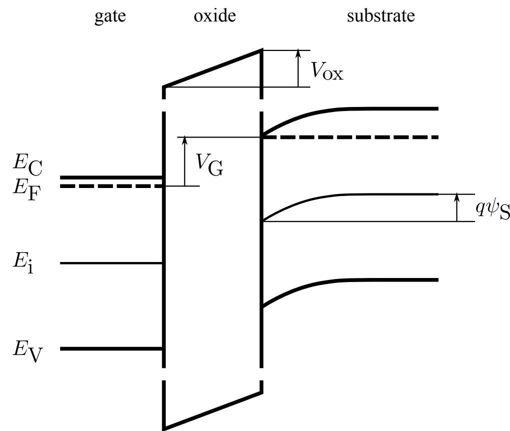


Fig. 7: Band diagram of a n^{++}/n MOS structure in accumulation. Accumulation happens for a positive gate voltage V_G.

Depletion

When V_G is less than $V_{FB} < 0$ for a n^{++}/n MOS structure majority electrons are repelled from the Si–SiO_2 interface and adjacent regions in the semiconductor. This regions are *depleted* of majority carriers. The bands in the semiconductor

will bend slightly upwards and the density of free electrons at the interface will be smaller than the density in the bulk of the semiconductor. However, there are still more free electrons than free holes at the interface. The band bending is still small such that the intrinsic energy E_i does not intersect the FERMI level E_F. Consider Fig. 8 for a sketch of a MOS structure in depletion. The borders which define the depletion region for the surface potential are the flat–band condition $\psi_\mathrm{S}= 0$ on one side and on the other side the value $-\phi_\mathrm{Bn}$, at which the density of electrons coincides with the density of holes.

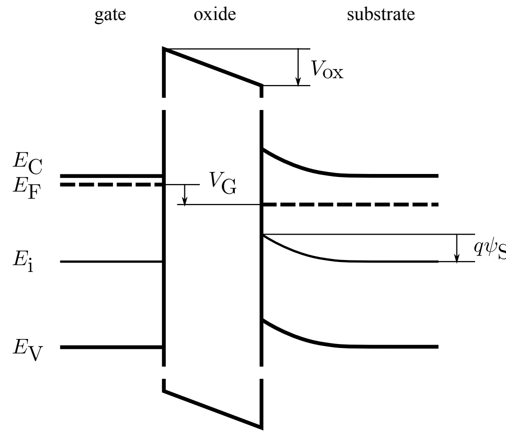


Fig. 8: Band diagram of a $\mathrm{n}^{++}/\mathrm{n}$ MOS structure in Depletion. A small negative gate voltage V_G has to be applied to obtain such a band bending situation.

Inversion

By further reducing the voltage V_G applied to the gate with respect to the silicon bulk, the number of free holes p will exceed the number of free electrons n. So the carrier situation at the interface is inverted compared to the carrier situation in the bulk of the Si. n type material is inverted to p type material and vice versa for a p substrate MOS structure and sufficient positive gate voltage. The potential situation in inversion is depicted in detail in Fig. 9. As long as the number of free holes p is smaller than the number of free electrons n in the bulk of the semiconductor one usually refers to *weak inversion.* The limit to a fully inverted interface, and thus the definition of threshold for inversion is $\psi_\mathrm{S}= -2\phi_\mathrm{Bn}$. Exceeding this point the carrier situation at the interface is at least the same as in the bulk of the semiconductor but with opposite sign. The potential difference at the gate needed to achieve this situation is denoted as V_TH, the *threshold voltage.* An analytical solution for the V_TH is derived using equation (22) and the fact that the voltage drop across the oxide is given by the total charge stored in the semiconductor $V_\mathrm{ox}= -Q_\mathrm{S}/C_\mathrm{ox}$. The value of Q_S at the threshold condition can be approximately determined by simplifying the function F in equation (21) of section 1.2 for weak inversion or depletion conditions to $F \approx -\sqrt{n(\infty)/p(\infty)}\sqrt{q|\psi_\mathrm{S}|/kT}$. Simple approximating functions for F in other regions are listed in Table 1. From that and from equation (20)

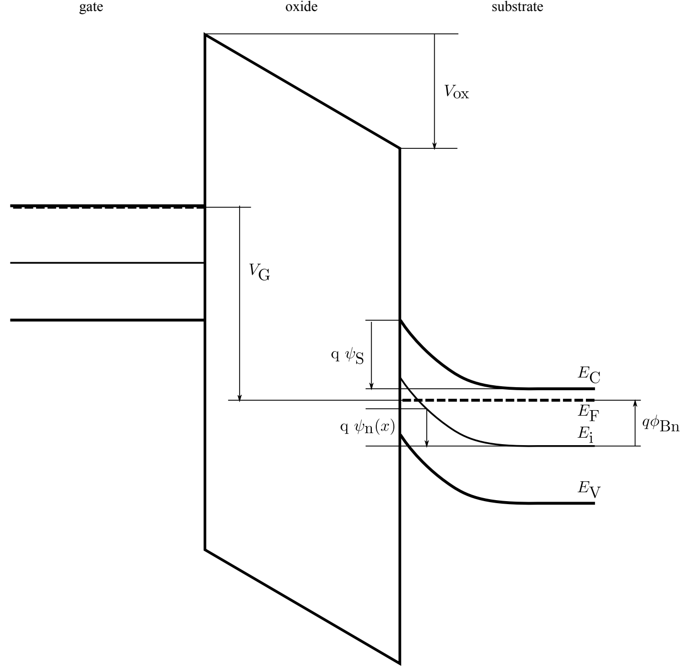


Fig. 9: Band diagram of a n^{++}/n MOS structure in inversion. The poly gate is n^{++} doped and therefore experiences negligible band bending in inversion. ψ_{S} is the total band bending in the semiconductor caused by the potential ϕ_{S} at the Si–$\mathrm{SiO_2}$ interface. ψ_{n} is the band bending of the n doped semiconductor at the position x with respect to the bulk. ϕ_{Bn} is the potential difference between the intrinsic energy and the FERMI energy in the bulk of the semiconductor. E_{C}, E_{F}, E_{i} and E_{V} are the conduction band, FERMI level, intrinsic and valence band energy, respectively. V_{G} is the potential difference between the gate and the bulk of the semiconductor and V_{ox} is the potential drop over the oxide. Downward arrows indicate negative values of the according variables.

it follows that

$$\begin{aligned} Q_{\mathrm{S}} &= -\varepsilon_{\mathrm{Si}}\mathcal{E}_{\mathrm{S}} \\ &\approx \varepsilon_{\mathrm{Si}}\frac{\sqrt{2}kT}{qL_{\mathrm{D}}}\sqrt{\frac{n(\infty)}{p(\infty)}}\sqrt{\frac{q|\psi_{\mathrm{S}}|}{kT}} \\ &= \sqrt{2\varepsilon_{\mathrm{Si}}q|\psi_{\mathrm{S}}|n(\infty)} \end{aligned} \tag{23}$$

and so for the threshold voltage

$$\begin{aligned} V_{\mathrm{TH}} &= V_{\mathrm{FB}} + V_{\mathrm{ox}} + \psi_{\mathrm{S}}(V_{\mathrm{TH}}) \\ &= V_{\mathrm{FB}} - \frac{\sqrt{2\varepsilon_{\mathrm{Si}}q(2\phi_{\mathrm{Bn}})n(\infty)}}{C_{\mathrm{ox}}} - 2\phi_{\mathrm{Bn}}, \end{aligned} \tag{24}$$

for the n^{++}/n MOS structure discussed in our example. $n(\infty)$ in equation (24) can usually be approximated by the ionized donor density N_{D}^{+}.

Table 1: Approximating functions for $F\left(\frac{q\psi(x)}{kT}, \frac{n(\infty)}{p(\infty)}\right)$.

region	F
accumulation	$\approx \operatorname{sgn}(\psi(x)) \sqrt{\frac{n(\infty)}{p(\infty)}} \mathrm{e}^{\frac{q\lvert\psi(x)\rvert}{2kT}}$
depletion or weak inversion	$\approx \operatorname{sgn}(\psi(x)) \sqrt{\frac{n(\infty)}{p(\infty)}} \sqrt{\frac{q\lvert\psi(x)\rvert}{kT}}$
inversion	$\approx \operatorname{sgn}(\psi(x))\, \mathrm{e}^{\frac{q\lvert\psi(x)\rvert}{2kT}}$

1.4 Discussion on interface traps and oxide charges

Charges resulting from defects are usually characterized by their effective charge per unit area Q. Sometimes, the effective number of charges per unit area N is used. The two variables are related by $N = |Q|/q$.

Interface traps

As mentioned in section 1.1, imperfections in the crystal structure of the Si near the Si–SiO_2 interface cause interface trapped charge. Interface traps are named acceptor– or donor–like, just as defects in the bulk of a semiconductor. This means a *donor* interface trap is charged positively when unoccupied and neutral when occupied (+/0). An *acceptor* interface trap is neutral when unoccupied and charged negatively when occupied (0/−). Occupancies of those energy levels change by electron emission/capture from the conduction band or hole emission/capture from the valence band. The occupancy states of interface traps are determined by a ground–state degeneracy weighted FERMI–DIRAC distribution function [SN06] which depends heavily on the position of the FERMI level with respect to the conduction and valence band edges. At room temperature it is justifiable to reduce the distribution to a step function. So interface traps with an energy level lower than the FERMI level are occupied and those higher are unoccupied. For every system of acceptor– and donor–like interface traps an energy E_0 can be found for which the total charge of the system Q_{IT} is positive when $E_{\mathrm{F}} < E_0$ and negative when $E_{\mathrm{F}} > E_0$. When E_0 lies near the middle of the Si band gap the interface trap system is called amphoteric [HP94].

From an atomic–scale point of view, interface traps are dangling bonds of Si atoms at the interface. The most important among them are the *Pb centers*, characterized extensively by electron spin resonance measurements [HP94]. On (100) oriented Si crystals, two forms of Pb centers occur. A schematic illustration of these two defects is given in Fig. 10. The Pb0 center has a donor–like energy level (+/0) around 0.25 eV above the valence band edge and an acceptor–like level (0/−) approximately 0.35 eV below the conduction band edge [Edw91]. Thus the Pb0 defect is of amphoteric nature. Information about the energy levels of the Pb1 center is limited. In 2001, it was found that the Pb1 center has two energy levels which are significantly more narrowly distributed and are predominantly in the lower part of the band gap [LMJFW01].

Usually interface traps are described by a density of states as a function of energy

$$D_{\mathrm{IT}}(E) = \frac{1}{q}\frac{\mathrm{d}Q_{\mathrm{IT}}}{\mathrm{d}E}. \tag{25}$$

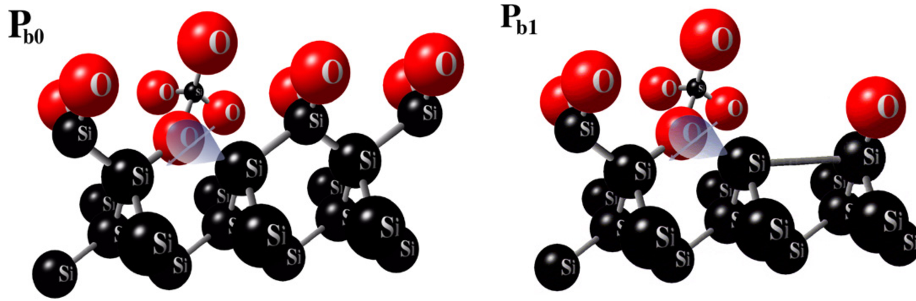


Fig. 10: Schematic illustration of Pb0 and Pb1 defects at the Si–SiO_2 interface, copied from [Len07, p. 892].

Charges trapped at the interface influence the potential situation in the MOS structure. A charge Q_{IT} located at the interface causes a voltage drop across the oxide of

$$V = -\frac{Q_{\mathrm{IT}}}{C_{\mathrm{ox}}}, \tag{26}$$

scaled by the oxide capacitance C_{ox} per unit area.

The dangling bonds of Si atoms may be passivated by hydrogen atoms. This causes a reduction of the density of interface traps. Because interface traps may have undesirable effects on device performance, the density of interface traps is reduced by exposing MOS structures to diatomic hydrogen (H_2) rich atmosphere at elevated temperatures.

Oxide charges

Defects and imperfections in the atomic structure of the bulk of the SiO_2 may cause *oxide charges*. One important difference between oxide charges and interface traps is that oxide charges are not in direct electric contact with the underlying Si. However, they may exchange carriers with the Si by tunneling processes. Fleetwood thus suggested that traps which are capable of charging/discharging processes should be referred to as *border traps* [Fle92].

Among several types of atomic defects in SiO_2, the E' centers are the most important defects in thin film SiO_2 [HP94]. An E' center consists of one Si atom bound to only three instead of four oxygen (O) atoms. The remaining fourth electron is a dangling bond. This structure alone is called an E'_{s} center. If another positively charged E'_{s} structure (with a positively ionized Si atom instead of the dangling bond) is present, one refers to an E'_{γ} center. Figure 11 gives a schematic illustration of the E'_{γ} center and its precursor, a neutral oxygen vacancy.

A charge distribution $\rho(x)$ within the oxide layer causes a voltage drop across the oxide according to Gauss' law of (a derivation can be found in [See, chapter 9.8.3])

$$V = -\frac{1}{C_{\mathrm{ox}}} \left(\frac{1}{T_{\mathrm{ox}}} \int_0^{T_{\mathrm{ox}}} x\rho(x)\mathrm{d}x \right), \tag{27}$$

where the oxide thickness is denoted as T_{ox}. The resulting voltage shift is often used to define an effective charge $Q = -V\,C_{\mathrm{ox}}$. Defect nomenclature

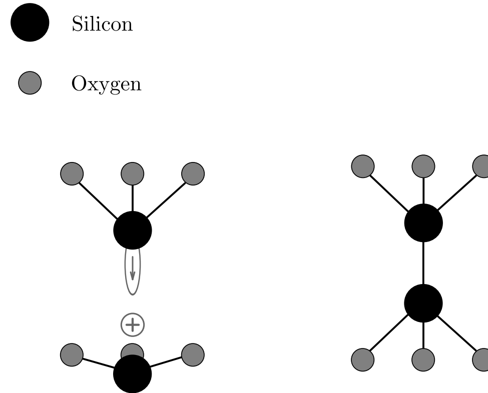


Fig. 11: Schematic illustration of a E'_γ center (left hand side). The small arrow symbolizes an electron in an dangling bond. The plus sign advises the reader of the fact, that the Si atom is positively ionized and thus acts as a fixed positive charge in the oxide. On the right hand side the precursor state for the E'_γ center, a neutral oxygen vacancy, is depicted.

differentiates between fixed oxide charges Q_{ox}, mobile ionic charge Q_{m} and oxide trapped charge Q_{ot} after [Dea80] (see Fig. 1). Assuming all of these charges to positive, they cause that a more negative voltage has to be applied to the gate to end up with the same band bending as without the charges. In particular, the flat–band voltage V_{FB} and the threshold voltage V_{TH} shift toward negative infinity. Mind that the oxide field is not longer zero at V_{FB} in the presence of oxide charges. Summarizing all effective oxide charges to Q, the voltage drop across the oxide at the new flat–band bias is

$$V_{\mathrm{ox}}\left(V_{\mathrm{G}}=V_{\mathrm{FB}}\right)=-2\frac{Q}{C_{\mathrm{ox}}}. \tag{28}$$

1.5 Impedance of the MOS structure

Because of the presence of the oxide as an isolating layer between two conducting layers, the MOS structure can be regarded as a capacitor. The dominant capacitance of the MOS structure is clearly the value of the oxide layer $C_{\mathrm{ox}}=\varepsilon_{\mathrm{SiO_2}}/T_{\mathrm{ox}}$. However, a voltage–dependent capacitance due to the space charge in the semiconductor adds to the oxide capacitance. The voltage V_{G} applied between the gate and the substrate of a MOS structure can be written as (see equation (22))

$$V_{\mathrm{G}}=V_{\mathrm{FB}}+V_{\mathrm{ox}}+\psi_{\mathrm{S}}. \tag{29}$$

Because the flat–band voltage can be regarded as constant, a differential small change in the gate voltage will result in

$$\mathrm{d}V_{\mathrm{G}}=\mathrm{d}V_{\mathrm{ox}}+\mathrm{d}\psi_{\mathrm{S}}. \tag{30}$$

The semiconductor charge is related to the charge in the gate electrode Q_{G} due to charge neutrality by $Q_{\mathrm{S}}+Q_{\mathrm{G}}=0$. Beginning with the reciprocal value of the

differential capacitance C of the MOS structure,

$$\begin{aligned}\frac{1}{C} &= \frac{1}{\frac{\mathrm{d}Q_{\mathrm{G}}}{\mathrm{d}V_{\mathrm{G}}}} \\ &= \frac{\mathrm{d}V_{\mathrm{ox}} + \mathrm{d}\psi_{\mathrm{S}}}{\mathrm{d}Q_{\mathrm{G}}} \\ &= \frac{\mathrm{d}V_{\mathrm{ox}}}{\mathrm{d}Q_{\mathrm{G}}} - \frac{\mathrm{d}\psi_{\mathrm{S}}}{\mathrm{d}Q_{\mathrm{S}}} \\ &= \frac{1}{C_{\mathrm{ox}}} + \frac{1}{C_{\mathrm{S}}},\end{aligned} \tag{31}$$

where C_{S} can be regarded as the capacitance of the semiconductor. The definition of $C_{\mathrm{S}} = -\frac{\mathrm{d}Q_{\mathrm{S}}}{\mathrm{d}\psi_{\mathrm{S}}}$ accounts for the different signs of the band bending and the semiconductor charge. A band bending downward ($\psi_{\mathrm{S}} > 0$) is linked to a negative semiconductor charge ($Q_{\mathrm{S}} < 0$). Thus, the capacitance of the MOS structure has a constant (C_{ox}) and a voltage dependent part (C_{S}), connected in series. The voltage dependence of C_{S} is originated in the voltage dependence of the surface potential ψ_{S}.

The value of C_{S} as a function of the gate voltage can be calculated by finding the theoretical dependence of ψ_{S} on V_{G} and deducing $C_{\mathrm{S}} = -\frac{\mathrm{d}\psi_{\mathrm{S}}}{\mathrm{d}Q\mathrm{S}}$ [NB02; See; SN06]

$$\begin{aligned}C_{\mathrm{S}} &= -\frac{\mathrm{d}Q_{\mathrm{S}}}{\mathrm{d}\psi_{\mathrm{S}}} \\ &= \frac{\mathrm{d}}{\mathrm{d}\psi_{\mathrm{S}}} \frac{\sqrt{2}kT\varepsilon_{\mathrm{Si}}}{qL_{\mathrm{D}}} F\left(\frac{q\psi_{\mathrm{S}}}{kT}, \frac{n(\infty)}{p(\infty)}\right) \\ &= \mathrm{sgn}\left(\psi_{\mathrm{S}}\right) \frac{\sqrt{2}kT\varepsilon_{\mathrm{Si}}}{qL_{\mathrm{D}}} \frac{\mathrm{d}}{\mathrm{d}\psi_{\mathrm{S}}} \sqrt{\mathrm{e}^{-\frac{q\psi_{\mathrm{S}}}{kT}} + \frac{q\psi_{\mathrm{S}}}{kT} - 1 + \frac{n(\infty)}{p(\infty)}\left(\mathrm{e}^{\frac{q\psi_{\mathrm{S}}}{kT}} - \frac{q\psi_{\mathrm{S}}}{kT} - 1\right)} \\ &= \frac{\varepsilon_{\mathrm{Si}}}{\sqrt{2}L_{\mathrm{D}}} \frac{1 - \mathrm{e}^{-\frac{q\psi_{\mathrm{S}}}{kT}} - \frac{n(\infty)}{p(\infty)}\left(1 - \mathrm{e}^{\frac{q\psi_{\mathrm{S}}}{kT}}\right)}{F\left(\frac{q\psi_{\mathrm{S}}}{kT}, \frac{n(\infty)}{p(\infty)}\right)}.\end{aligned} \tag{32}$$

This derivation assumes that majority as well as minority carriers can follow instantaneously every change of the applied voltage. The theoretical capacitance voltage (CV) curve is depicted in Fig. 12. At sufficiently low frequencies and sufficiently high temperatures minority carriers are supplied by various mechanisms. Because of that, the characteristics in Fig. 12 are called low frequency capacitance voltage (LF CV) curves. At higher frequencies and under the absence of any minority carrier sources a high frequency capacitance voltage (HF CV) curve is obtained. The exact solution for the dependence of the MOS capacitance on the gate voltage at high frequencies is derived in [NB02]. It has a capacitance value around the minimum of the LF CV curve in inversion because the depletion layer width, responsible for the voltage dependent capacitance C_{S}, saturates at its maximum in strong inversion. The depletion width x_{D} can be calculated when assuming that the charge in the semiconductor in depletion is given by $Q_{\mathrm{S}} = qx_{\mathrm{D}}N_{\mathrm{D}}^{+}$ for a n doped MOS structure. In depletion or weak inversion Q_{S} can be approximated as given in equation (23) on page 18

$$Q_{\mathrm{S}} = \sqrt{2\varepsilon_{\mathrm{Si}} q |\psi_{\mathrm{S}}| N_{\mathrm{D}}^{+}}. \tag{33}$$

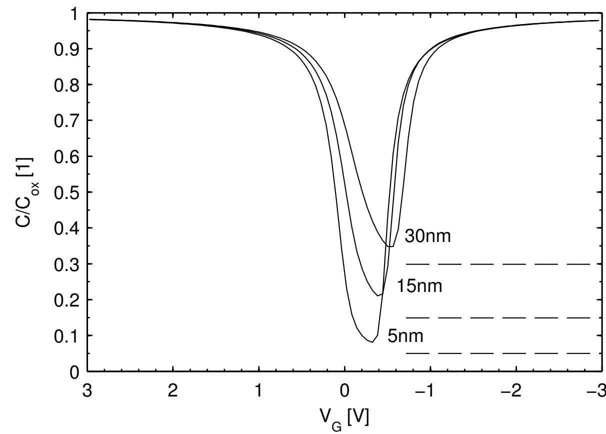


Fig. 12: Theoretical dependence of the capacitance of a MOS structure with n substrate on the gate voltage. Minority carriers are able to follow any change of the voltage applied (solid lines). Additional charges located in the oxide or at the interface, as well as a work function difference, are neglected. Depicted is the normalized capacitance for three different oxide thicknesses. The curve is calculated for a donor doping of $10^{16}\,\mathrm{cm}^{-3}$ at room temperature. The dashed lines correspond to the according high frequency minimum capacitance values in strong inversion.

The value of the HF CV curve in strong inversion can be approximated by the capacitance value at the onset of strong inversion, where $\psi_\mathrm{S} = 2kT\ln\left(N_\mathrm{D}^+/n_\mathrm{i}\right)$ by using

$$x_\mathrm{D} = \sqrt{\frac{2\varepsilon_\mathrm{Si}|\psi_\mathrm{S}|}{qN_\mathrm{D}^+}} \tag{34}$$

from the above equations and deriving

$$C_\mathrm{S} = \frac{\varepsilon_\mathrm{Si}}{x_\mathrm{D}} = \sqrt{\frac{q\varepsilon_\mathrm{Si}N_\mathrm{D}^+}{4kT\ln\left(\frac{N_\mathrm{D}^+}{n_\mathrm{i}}\right)}}. \tag{35}$$

The values of the high frequency capacitance in strong inversion is depicted in Fig. 12.

1.6 Carrier transport through the oxide

In most applications the oxide of the MOS structure can be regarded as a perfect insulator, thus no carrier transport through the oxide has to be considered. However, the thin oxides used in todays semiconductor industry require us to consider conducting mechanisms through the insulator as well. Among those two particular tunneling mechanisms derived from quantum mechanics are the most important ones. Depending on the oxide thickness and the applied electric field, tunneling through the full width of the oxide is referred as *direct tunneling* (see Fig. 14), whereas tunneling through only a part of the oxide (through a triangular shaped barrier) is referred as Fowler–Nordheim *tunneling* (see

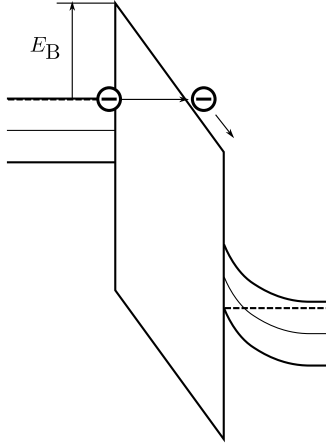


Fig. 13: Schematic illustration of the band situation in a n^{++}/nMOS structure during FOWLER–NORDHEIM tunneling of poly gate conduction band electrons toward the substrate. E_B denotes the height of the tunneling barrier for an electron in the conduction band in the poly.

Fig. 13). Reference [DVMMH95] provides a good derivation of analytical expressions for the two tunneling mechanisms. Accordingly, FOWLER–NORDHEIM tunneling current density can be calculated by

$$J_\mathrm{FN} = \frac{q^2}{16\pi^2\hbar E_\mathrm{B}}\mathcal{E}_\mathrm{ox}^2 \exp\left(-\frac{4\sqrt{2qm_{\mathrm{SiO}_2}}E_\mathrm{B}^{3/2}}{3\hbar\mathcal{E}_\mathrm{ox}}\right), \tag{36}$$

where E_B refers to the barrier height, $\hbar$ is the reduced PLANCK constant and m_{SiO_2} is the effective mass of an electron in the conduction band of the SiO_2. The value of m_{SiO_2} is around a third of the mass of an electron m_e [DVMMH95]. FOWLER–NORDHEIM tunneling can usually regarded to be fairly independent of temperature. Electrons tunneling from the conduction band from either the poly gate or the Si substrate represent usually the majority of tunneling carriers because the barrier height can be as small as $E_\mathrm{B} = \chi_\mathrm{Si} - \chi_{\mathrm{SiO}_2} \approx 3.15\,\mathrm{eV}$. By contrast, the barrier height for an hole in the valence band can not be smaller than $E_\mathrm{B} = (\chi_{\mathrm{SiO}_2} + E_{\mathrm{G,SiO}_2}) - (\chi_\mathrm{Si} + E_\mathrm{G,Si}) \approx 4.73\,\mathrm{eV}$ at room temperature. However, recent studies [SMPD98] demonstrate that the contribution of different carriers to the tunneling current can be separated. I may remark that it is important to realize that electrons are the only carrier species which actually tunnel. Holes are naturally not capable of tunneling because they are just a model to describe charge transport in the valence band. Consequently, when the energy level of the tunneling electron is within the valence band one denotes the tunneling event a hole–tunneling process.

Other conducting mechanisms are thermionic emission and tunneling processes involving defects within the bulk of the oxide. Because those mechanisms are usually not important regarding to negative bias temperature instability (NBTI), they are not explained in more detail here, but the reader may find an interesting discussion about those mechanisms in [SN06].

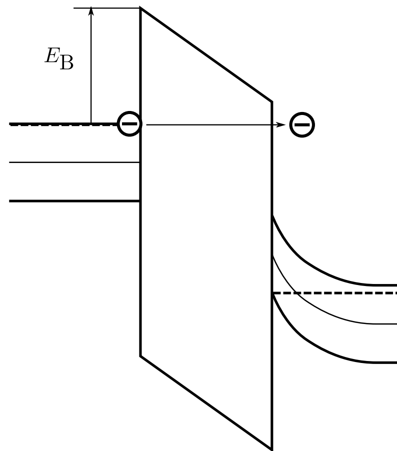


Fig. 14: Schematic illustration of the band situation in a n^{++}/nMOS structure during direct tunneling of poly gate conduction band electrons toward the substrate. E_B denotes the height of the tunneling barrier for an electron in the conduction band in the poly.

2 The MOS field effect transistor

In section 1 I have discussed that the potential situation at the surface of a semiconductor can be changed rigorously by applying a voltage to the gate with respect to the bulk of the semiconductor. By this effect a conduction path from one to the other side of the gate may be controlled by the gate potential. This principle was firstly discovered in 1925 [Lil30]. The name MOS field effect transistor (MOSFET) comes from the fact that the electric field originating from the gate causes the change in the potential situation in the semiconductor in a MOS structure.

2.1 Composition of a MOSFET

The MOS structure discussed in section 1 is extended with two additional implants to form a MOSFET. The two regions are doped oppositely to the substrate by ion implantation. In Fig. 15, a MOS structure with two p doped regions which form a p channel MOS field effect transistor (pMOS transistor) is depicted as an example. Through the application of a sufficient negative gate

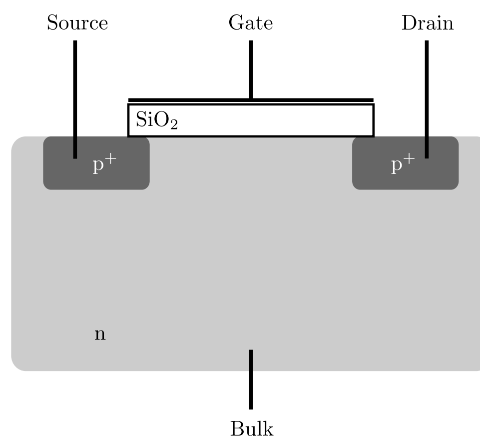


Fig. 15: Schematic sketch of a pMOS transistor. The source and drain regions are p^+ doped while the bulk of the semiconductor is n doped.

bias with respect to the substrate, the region of the semiconductor near the SiO_2 gets inverted and forms a conducting path between the source and the drain. The source usually acts as a supplier for the charge carriers. For a pMOS transistor the carriers are the positively charged holes. Thus, upon the application of a negative bias to the drain with respect to the substrate causes a drift of holes from the source toward the drain. The resulting current is measured as a negative current at the drain.

The counterpart to a pMOS transistor is the n channel MOS field effect transistor (nMOS transistor) where two n doped regions are connected with each other by inverting a p substrate MOS structure through the application of a sufficient positive gate voltage. However, the subsequent derivations are for a pMOS transistor.

2.2 Mathematical modeling of the charge transport from the source to the drain

As indicated in the previous section, the inversion charge in the semiconductor causes the electric contact between the source and the drain. Thus the resulting current at the drain I_{D} is determined by Q_{i} the inversion layer charge and the average velocity of the carriers along the channel v as

$$\int_0^L I_{\mathrm{D}}(y)\mathrm{d}y = I_{\mathrm{D}} = \frac{W}{L}\int_0^L |Q_{\mathrm{i}}(y)|v(y)\mathrm{d}y, \tag{37}$$

where L is the channel length and W the channel width. y denotes the spatial position from the source ($y = 0$) to the drain ($y = L$). The integral of the exact expression for the inversion layer charge has no analytical solution [SN06]. Thus simplifying assumptions have to be made to obtain analytical approximations for the drain current. Among several different approaches an idealization named *charge–sheet model* [Bre78] has become a very popular one. It assumes the inversion layer to be a charge sheet with zero thickness at the Si–$\mathrm{SiO_2}$ interface. Using this assumption, the inversion layer charge is according to GAUSS's law $\mathrm{div}D = \rho$ the root cause for the difference in the electric displacement fields ($D = \varepsilon\mathcal{E}$) in the oxide $\mathcal{E}_{\mathrm{ox}}$ and the semiconductor $\mathcal{E}_{\mathrm{S}}$,

$$Q_{\mathrm{i}} = \varepsilon_{\mathrm{Si}}\mathcal{E}_{\mathrm{S}} - \varepsilon_{\mathrm{SiO2}}\mathcal{E}_{\mathrm{ox}}. \tag{38}$$

The oxide field can be determined from equation (22)

$$\mathcal{E}_{\mathrm{ox}} = \frac{V_{\mathrm{G}} - V_{\mathrm{FB}} - \psi_{\mathrm{S}}}{T_{\mathrm{ox}}} \tag{39}$$

and the field in a n doped semiconductor in inversion from equation (23)

$$\mathcal{E}_{\mathrm{S}} = -\sqrt{\frac{2q|\psi_{\mathrm{S}}|n(\infty)}{\varepsilon_{\mathrm{Si}}}}. \tag{40}$$

The surface potential ψ_{S} is clearly dependent on the voltage applied to the drain V_{D} with respect to the source and thus also has a dependence on the spatial dimension y. The potential is usually approximated at the onset of strong inversion by

$$\begin{aligned}\psi_{\mathrm{S}}(y) &\approx -2\phi_{\mathrm{Bn}} + \frac{E_{\mathrm{i}}(y=0) - E_{\mathrm{i}}(y)}{q} \\ &= -2\phi_{\mathrm{Bn}} + \phi_{\mathrm{D}}(y),\end{aligned} \tag{41}$$

defining a lateral position dependent potential $\phi_{\mathrm{D}}(y)$. From that and $C_{\mathrm{ox}} = \varepsilon_{\mathrm{SiO2}}/T_{\mathrm{ox}}$, the inversion layer charge is

$$Q_{\mathrm{i}}(y) = -\sqrt{2q\varepsilon_{\mathrm{Si}}n(\infty)(-2\phi_{\mathrm{Bn}} + \phi_{\mathrm{D}})} - C_{\mathrm{ox}}\left(V_{\mathrm{G}} - V_{\mathrm{FB}} + 2\phi_{\mathrm{Bn}} - \phi_{\mathrm{D}}\right). \tag{42}$$

Assuming a constant mobility μ of the carriers within the channel, their velocity can be described as $v = \mu\mathcal{E}$ and equation (37) can be integrated

$$\begin{aligned} I_{\mathrm{D}} &= \mu\frac{W}{L}\int_0^L |Q_{\mathrm{i}}(y)|\mathcal{E}(y)\mathrm{d}y \\ &= \mu\frac{W}{L}\int_0^L |Q_{\mathrm{i}}(y)|\frac{\mathrm{d}\phi_{\mathrm{D}}}{\mathrm{d}y}\mathrm{d}y \\ &= \mu\frac{W}{L}\int_0^{V_{\mathrm{D}}} |Q_{\mathrm{i}}(\phi_{\mathrm{D}})|\mathrm{d}\phi_{\mathrm{D}} \\ &= \mu\frac{W}{L}C_{\mathrm{ox}}\Bigg(\left(V_{\mathrm{G}} - V_{\mathrm{FB}} - 2\phi_{\mathrm{B}} - \frac{V_{\mathrm{D}}}{2}\right)V_{\mathrm{D}} \\ &- \frac{2}{3}\frac{\sqrt{2q\varepsilon_{\mathrm{Si}}n(\infty)}}{C_{\mathrm{ox}}}\left((V_{\mathrm{D}} + 2\phi_{\mathrm{B}})^{3/2} - (2\phi_{\mathrm{B}})^{3/2}\right)\Bigg). \end{aligned} \tag{43}$$

This formula already successfully describes the commonly observed behavior of the drain current to be linearly dependent on the drain voltage as long as $V_{\mathrm{D}}<V_{\mathrm{D,sat}}$ and to saturate to some value for $V_{\mathrm{D}}>V_{\mathrm{D,sat}}$. The voltage where the transition to a constant (saturated) current happens is called saturation drain voltage $V_{\mathrm{D,sat}}$. The value of

$$V_{\mathrm{D,sat}} = \frac{V_{\mathrm{G}} - V_{\mathrm{TH}}}{1 + \sqrt{\frac{q\varepsilon_{\mathrm{Si}}n(\infty)}{4C_{\mathrm{ox}}^2\phi_{\mathrm{B}}}}} \tag{44}$$

is obtained by setting $\mathrm{d}I_{\mathrm{D}}/\mathrm{d}V_{\mathrm{D}} = 0$ or $Q_{\mathrm{i}}(L) = 0$. Consequently, the dependence of the drain current on the gate voltage is approximately

$$I_{\mathrm{D}} = \mu C_{\mathrm{ox}}\frac{W}{L}\left(V_{\mathrm{G}} - V_{\mathrm{TH}} - \frac{V_{\mathrm{D}}}{2}\right)V_{\mathrm{D}} \qquad \text{for } V_{\mathrm{G}} < V_{\mathrm{D,sat}}, \tag{45}$$

$$I_{\mathrm{D}} = \frac{1}{2}\mu C_{\mathrm{ox}}\frac{W}{L}\left(V_{\mathrm{G}} - V_{\mathrm{TH}}\right)^2 \qquad \text{for } V_{\mathrm{G}} > V_{\mathrm{D,sat}}. \tag{46}$$

The threshold voltage used in the upper equations was defined for a MOS structure in equation (24). The definition of the threshold voltage in a MOSFET is equivalent except for the fact that a voltage applied to the bulk of the semiconductor with respect to the source V_{B} enters the equation as

$$V_{\mathrm{TH}} = V_{\mathrm{FB}} - 2\phi_{\mathrm{Bn}} - \frac{\sqrt{2q\varepsilon_{\mathrm{Si}}n(\infty)(2\phi_{\mathrm{Bn}} + V_{\mathrm{B}})}}{C_{\mathrm{ox}}} \tag{47}$$

for a pMOS transistor.

2.3 Impedance of the MOSFET

The impedance of a MOSFET is quite similar to the one of a MOS structure. However, the presence of oppositely doped regions near the Si–SiO_2 interface changes the behavior of the measured impedance in inversion. This is due to the fact that minority carriers are not only supplied by thermal excitation but also move from the source/drain regions toward the interface. More precisely, by measuring the impedance of a MOSFET not only between the gate and the bulk, one can additionally connect the source and the drain with the bulk in order to obtain a LF CV curve also at high frequencies and low temperatures.

3 The negative bias temperature instability

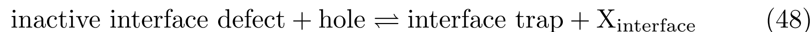
First observations [GN66; MM66] of a drift of the flat–band voltage of a MOS structure as a consequence of a negative bias applied to the gate with respect to the substrate at elevated temperatures date back to 1966. Soon after, the effect was referred as NBTI because also other transistor parameters like the threshold voltage, the transconductance or the channel carrier mobility were measured to be instable under the same stress conditions. Since that time, NBTI has been a reliability concern of integrated circuits and was the focus of much research during the past roughly 40 years. The effect is of greatest importance in pMOS transistor devices, not at least because they experience negative bias during normal circuit operation, particularly in complementary metal oxide semiconductor (CMOS) technology.

Nowadays it is almost universally accepted that the application of a negative bias to the gate of a MOS structure or a MOSFET leads to a build–up of positive charge in the oxide layer or at the Si–SiO_2 interface. This positive charge causes a change in the potential situation within the structure, as described in section 1.4, which leads to a degradation of the commonly observed transistor parameters.

3.1 The Reaction–Diffusion model

The earliest modeling attempts [JS77] proposed already a microscopic model which consists on a reaction causing defect creation and a subsequent diffusion of a mobile neutral hydrogen (H) species. This idea was later completed to the reaction–diffusion model (RD model).

The RD model only aims to explain the interface trap generation happening during negative bias temperature stress (NBTS). Generation of oxide charges is attributed to other effects like impact ionization or anode hole injection [AM05]. The RD model proposes the reaction



$$\text{inactive interface defect} + \text{hole} \rightleftharpoons \text{interface trap} + \mathrm{X}_{\text{interface}} \tag{48}$$

and the subsequent diffusion

$$\mathrm{X}_{\text{interface}} \rightleftharpoons \mathrm{X}_{\text{bulk}} \tag{49}$$

of the mobile species X away from the interface.

It is argued that the application of an electric field supported by elevated temperatures breaks hydrogen passivated silicon dangling bonds (Si–Hs) at the Si–SiO_2 interface. It is generally believed that a Si–H at the interface is the electric inactive interface defect precursor in equation (48) [SZ06]. Most probably the dissociation of a Si–H is initiated by the capture of a hole via field assisted tunneling [AM05]. The remaining unpassivated Si atom is trivalent, providing one dangling bond, which acts as an interface trap. It is still debated which specific configuration the released hydrogen forms. It may react with an other H atom to form a H_2 molecule or it may stay a mono–atomic H which may be electrical neutral or charged positively. It is assumed that any species diffuse away from the interface because of the high concentration of hydrogen after the bond breaking of several preexisting Si–Hs. The diffusion X can be described

by FICK's laws. The Si dangling bond will stay an electrically active interface defect until it is re–passivated by a hydrogen atom.

Once the stress bias is removed, it is assumed that hydrogen passivates the Si dangling bonds again. The concentration of hydrogen at the interface decreases with respect to adjacent areas and back–diffusion of hydrogen initiates.

Mathematical modeling

The RD model is described by the following set of differential equations.

$$\frac{\mathrm{d}N_{\mathrm{IT}}}{\mathrm{d}t} = k_{\mathrm{F}}\left(N_{\mathrm{SiH}} - N_{\mathrm{IT}}\right) - k_{\mathrm{R}} N_{\mathrm{X}} N_{\mathrm{IT}} \tag{50}$$

$$\begin{aligned} \frac{\mathrm{d}N_{\mathrm{IT}}}{\mathrm{d}t} &= -D_{\mathrm{X}} \frac{\mathrm{d}N_{\mathrm{X}}}{\mathrm{d}x} + \mu_{\mathrm{X}} \mathcal{E}_{\mathrm{ox}} N_{\mathrm{X}} \\ &= -D_{\mathrm{X}} \left(\frac{\mathrm{d}N_{\mathrm{X}}}{\mathrm{d}x} - \frac{q\mathcal{E}_{\mathrm{ox}}}{kT} N_{\mathrm{X}} \right) \end{aligned} \tag{51}$$

where N_{SiH} is the initial number of hydrogen passivated Si dangling bonds at the interface, N_{IT} the number of interface traps, N_{X} the number of diffusing species at the interface, $k_{\mathrm{F}}/k_{\mathrm{R}}$ the forward/reverse dissociation rate respectively, D_{X} the diffusion constant for the species X, μ_{X} the electric mobility of the charged species X and $\mathcal{E}_{\mathrm{ox}}$ the oxide field. For the final form of equation (51) the EINSTEIN relation for charged particles $D = \mu kT/q$ was used. The drift term (oxide field dependent term) of equation (51) is clearly only needed when the diffusing species X is charged. In the RD model it is generally assumed that the drift/diffusion is the limiting factor governing the creation rate of interface traps, that is to say it is assumed that Si–H dissociation happens in a time scale way shorter than the diffusion/drift time scale. Thus, the time dependence of NBTS induced degradation is strongly dependent on the diffusion/drift characteristics of the species X.

A derivation of the time dependence of the degradation using the above equations was already published in 1977 (see appendix of [JS77]). It was done with a few simplifying assumptions. Among them the most important one is an infinitely thick oxide, because otherwise the solution would be affected when the diffusion front reaches the gate poly. Another assumption is that neutral atomic H is the diffusing species. This analytical treatment has a time dependence of the number of interface traps N_{IT} of $t^{0.25}$ as a result. This means in a plot of $N_{\mathrm{IT}}(t)$ with logarithmic scales for both axes the growth of the number of interface traps is proposed to have a slope of 0.25, so every forth time decade the number of interface traps should grow by one decade. This power law dependence is described with the exponent α. The derivation of a power law exponent α= 0.25 is a key achievement of the RD model, because a majority of experiments confirm this power law dependence. Reference [AM05] gives a more complete derivation of this fact.

Predictions of the RD model

The power law exponent of α= 0.25 in the previous section based on the assumptions that the Si–H dissociation happens instantaneously, the oxide is infinitively thick and there are always enough Si–Hs. In a more general model five different regimes, determined by different power law exponents, are characteristic.

The five regimes are depicted in Fig. 16. I remark that it is assumed that the

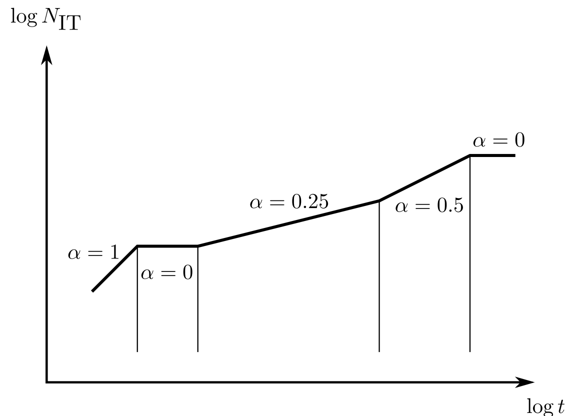


Fig. 16: Evolution of the power law exponent α as a function of stress time. The existence of five regimes, determined by different power law exponents, is an important prediction of the RD model.

diffusing species X is electricly neutral atomic H.

The first regime is characterized by the power law exponent $\alpha = 1$, where it is assumed that the virgin interface trap density as well as the hydrogen density N_{X} is negligible small. Especially the number of Si–Hs is much higher than N_{IT}. Solving equation (50) with this assumption gives $N_{\mathrm{IT}} \approx k_{\mathrm{F}} N_{\mathrm{SiH}} t$, which has a power law exponent of $\alpha = 1$. This means the creation of interface traps is fully determined by the forward reaction rate k_{F} in this early phase.

The second regime has a power law exponent $\alpha = 0$, because it is assumed, that the diffusion of X has not yet initiated. Thus the number of interface traps N_{IT} equals the number of released hydrogen N_{X} and neither increases nor decreases. From equation (50) this means $k_{\mathrm{F}} N_{\mathrm{SiH}} \approx k_{\mathrm{R}} N_{\mathrm{IT}}^2$ and so $N_{\mathrm{IT}} \approx \sqrt{k_{\mathrm{F}} N_{\mathrm{SiH}} / k_{\mathrm{R}}}$. It is argued that both phases last up to a few microseconds in modern MOSFETs [AM05] and are thus usually not observed in experiments.

The third regime is the most important one for the experimental observation of NBTS induced degradation. It begins when the diffusion of X controls the dynamic of the interface trap generation. Consequently the exact type of the diffusing species X determines the power law exponent in this regime. Numerical simulations assuming various different species were made [CKRMK04] and showed that any power law exponent α between roughly 0.165 and 0.5 can occur depending on the type of diffusing species. Charged species are believed to have a somewhat higher power law exponent $\alpha = 0.25$ to 0.5, whereas electricly neutral atomic H or H_2 have a power law exponent of $\alpha = 0.165$ to 0.25. Since most experimental data have a power law dependence of $\alpha = 0.25$ or slightly smaller, most researchers reject the existence of charged diffusers during NBTS.

The fourth regime begins when the diffusing species front reaches the other side of the gate oxide, that is to say the gate poly. It is assumed, that at the oxide–poly interface and/or at the grain boundaries in the poly, a number of dangling bonds exist which may passivate by the diffusing hydrogen species.

The poly thus acts as an absorber for the diffusing species and the equation

$$D_\mathrm{X} \frac{\mathrm{d}N_\mathrm{X}}{\mathrm{d}x} = k_\mathrm{P} N_\mathrm{X} \tag{52}$$

has to be added to the set of differential equations (50) and (51). The consequence is a power law exponent of up to 0.5 and the postulated existence of a transition point, where the power law exponent changes. This transition point should occur early in thin oxide devices and later in thick oxide devices assuming equivalent velocity of the diffusion front.

The fifth regime with a power law exponent $\alpha = 0$ should occur when all Si–Hs at the interface are dissociated and $N_\mathrm{IT} \approx N_\mathrm{SiH}$. At this point there are simply no electric inactive interface defect precursors any more. It is stated that this region may not be experimentally observable, because other failure modes are likely to occur in the meantime and cause oxide breakdown.

Another important prediction of the RD model is the existence of interface trap annealing due to back diffusion of hydrogen. It is assumed that after the termination of the stress bias the forward diffusion (toward the gate poly) stops because no more additional Si–Hs break up. The existing hydrogen species at the interface anneal the Si dangling bonds and thus the hydrogen concentration at the interface N_X decreases and backward diffusion initiates. Straightforward application of diffusion equations leads to a power law exponent of $\alpha = 0.5$ for the recovery of the NBTI induced degradation.

3.2 The Grasser model

Several alternative microscopic models and extensions to the standard RD model have been proposed by various researchers in the past [HDP06; AM05; HAGGHS05; GKGAHN09]. A very promising one was published in 2009 [GKGAHN09] and will be addressed in more detail in the following.

The model proposes the activation of E'_γ centers in SiO_2 as well as the generation of interface traps. It assumes that the degradation consists of two tightly coupled microscopic mechanisms. Those two mechanisms may result in defects which can be separated by their time dependence during recovery. That is to say, the model assumes NBTS induced degradation to consist of two different parts: a fast recovering and a slowly recovering, nearly permanent part.

Micro–physical background

The model assumes an electrically activated E'_γ center to be a recoverable, positively charged defect. The activation of E'_γ centers during NBTS was already proposed earlier by others [CLKK06]. The process is schematically depicted in Fig. 17.

It is assumed, that the precursor state (state 1 in Fig. 17) has its energy level roughly 1 eV below the Si valence band edge. Thus a highly energetic hole is needed to tunnel through a part of the oxide layer toward the precursor state in the SiO_2 bulk. I want to remark, but not fully explain, that this hole capture process is assumed to be a multiphonon–field–assisted–tunneling process. This particular process has the required properties in order to explain experimentally observed characteristics, such as a dependence on temperature and on the electric field. The hole leads to the breaking of the weak bond between

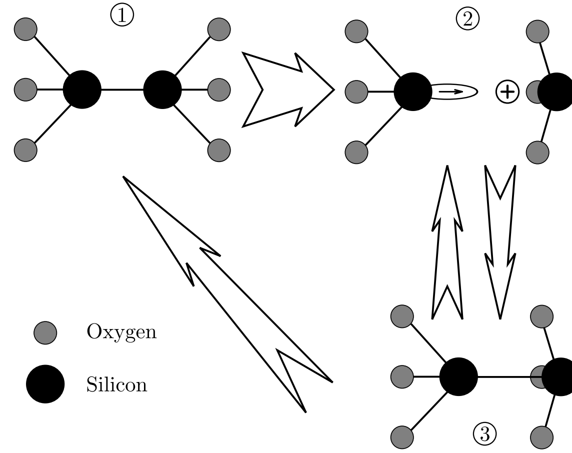


Fig. 17: Schematically depicted are the transitions which lead to the recoverable part of the degradation in the GRASSER model. State 1 is the precursor state for an E'_γ center, a neutral oxygen vacancy in SiO_2. State 2 is the E'_γ center, consisting of a dangling bond and a positively ionized Si atom. The transition from state 1 to state 2 is triggered by a hole capture event and leads to a structural relaxation of the adjacent SiO_2 areas. Hole emission/capture leads to transitions between state 2 and 3/3 and 2. State 3 is considered to be a neutral defect which may capture a hole more likely then the same defect in state 1. As the structure around the defect relaxes to its previous situation the defect is considered to recover fully, back to state 1.

the two Si atoms and thus causes the ionization of one of the Si atoms and a dangling bond at the other Si atom. This configuration of the Si and O atoms has a new equilibrium position which requires the adjacent areas of the SiO_2 to reposition. In this state, the E'_γ center can be repeatedly charged by hole emission or capture processes. The behavior of the resulting E'_γ centers as switching traps was already suggested earlier as the HARRY–DIAMOND–Laboratories (HDL) model [LO94]. The according energy levels are assumed to lie within the band gap of Si. The amorphous nature of the Si–SiO_2 interface causes the given energy levels to spread out. Thus, it is justifiable to argue that the energy levels of the E'_γ center lie homogeneously distributed within the bandgap of Si. Mind that the E'_γ defect is considered to be only donor–like (+/0).

The second part of the model, the slowly recovering or even permanent part of the NBTI degradation, is explained using a hydrogen transition and an interface trap. The microscopic model is depicted in Fig. 18. The starting point for this part is the charged E'_γ center (state 2 in Fig. 17). It is assumed that a hydrogen atom, which formerly passivated a Si dangling bond at the interface, migrates toward the E'_γ center and passivates the dangling bond there, thus leaving an interface trap behind. This H transition is modeled as a field–dependent transition over a barrier. The back transition of the H atom is unlikely, but of course not impossible. Thus it is believed that interface traps are the root cause of a nearly permanent degradation.

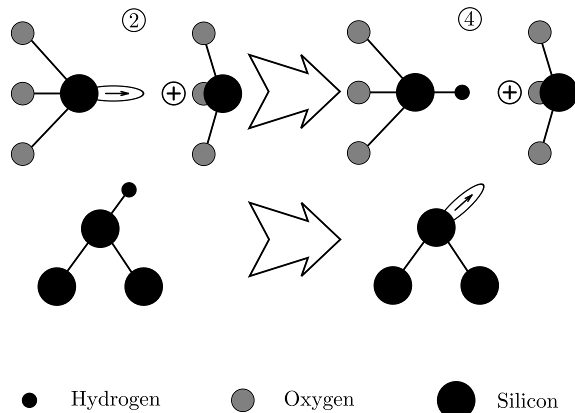


Fig. 18: Schematically depicted are the transitions which lead to the slowly recovering or even permanent part of the degradation in the GRASSER model. State 2 is identical with state 2 in Fig. 17, an E'_γ center. For the transition from 2 to 4, a hydrogen atom must move from a passivated Si dangling bond toward the E'_γ center. It is assumed that this transition is a field–dependent thermal transition over a barrier. The backward transition 4 to 2 is stated to be unlikely (thus the defect is considered to be of nearly permanent type) but not impossible.

Mathematical modeling

The mathematical ansatz to model the assumptions described in the previous sections is to write down a set of rate equations

$$\frac{\partial f_1}{\partial t} = -f_1 k_{12} + f_3 k_{31}, \tag{53}$$

$$\frac{\partial f_2}{\partial t} = +f_1 k_{12} - f_2 k_{23} + f_3 k_{32} - f_2 k_{24} + f_4 k_{42}, \tag{54}$$

$$\frac{\partial f_3}{\partial t} = +f_2 k_{23} - f_3 k_{32} - f_3 k_{31}, \tag{55}$$

$$\frac{\partial f_4}{\partial t} = +f_2 k_{24} - f_4 k_{42}, \tag{56}$$

where the probability of a defect being in state i is given by f_i and the transition rate from state i to state j is given by k_{ij}. Naturally, a defect has to be in one of the four states, so

$$1 = f_1 + f_2 + f_3 + f_4 \tag{57}$$

has to hold. The individual transition rates, where a charge exchange is involved, are calculated from capture/emission process rates for holes and electrons

$$k_{12} = k_{\mathrm{p}}^{\mathrm{c}} + k_{\mathrm{n}}^{\mathrm{e}}, \tag{58}$$

$$k_{23} = k_{\mathrm{p}}^{\mathrm{e}} + k_{\mathrm{n}}^{\mathrm{c}}, \tag{59}$$

$$k_{32} = k_{\mathrm{p}}^{\mathrm{c}} + k_{\mathrm{n}}^{\mathrm{e}}, \tag{60}$$

which are calculated using assumptions of the multiphonon–field–assisted–tunneling process. The capture/emission rates for holes or electrons are thus dependent on the energy level of the trap, as well as on the barrier height and the electric field for the multiphonon–field–assisted–tunneling process. The transition rate where the structural relaxation happens (k_{31}) is described by a simple BOLTZMANN distribution assuming an activation energy. The hydrogen transitions (k_{24} and k_{42}) also use a BOLTZMANN distribution but it is assumed that the activation energy is dependent on the electric field.

The number of oxide charges can than be calculated by suitably averaging the probability of defects in either the E'_γ state 2 or the locked state 4 and multiplying with a total number of defects N

$$N_{\mathrm{ox}} = N\langle f_2 + f_4 \rangle. \tag{61}$$

This equation assumes that all created defects can be regarded to be at the interface, that is to say T_{ox} is very big compared to the distance of the defect from the Si–SiO_2 interface. The number of interface traps created through NBTS is then dedicated as

$$N_{\mathrm{IT}} = N\langle f_4 \rangle. \tag{62}$$

Predictions of the Grasser model

The GRASSER model is capable of explaining various aspects of experimentally observed NBTI degradation. One of the most important is the asymmetry of the degradation behavior during stress and recovery. The GRASSER model successfully predicts that the degradation happens somewhat faster than the recovery under comparable conditions. In contrast to the previous described RD model, the GRASSER model predicts that the recovery after NBTS is due to charging/discharging of bulk oxide charges and possible subsequent relaxation the SiO_2 lattice, while interface traps are considered to be nearly permanent. Due to the capability of the switching traps to be charged and uncharged, the occupancy state of the E'_γ center will depend on the position of the FERMI level at the Si–SiO_2 interface. The structural relaxation of the E'_γ center can only happen when the defect does not lack an electron, that is to say is in state 3. So the time dynamics of recovery is considered to be dependent on the FERMI level position at the Si–SiO_2 interface during recovery. But not only the time dynamics, also the total active charge will be dependent on the FERMI level position at the interface, because of the occupancy state of the oxide charges and the amphoteric nature of the interface traps.

A schematic drawing of the full model can be found in Fig. 19.

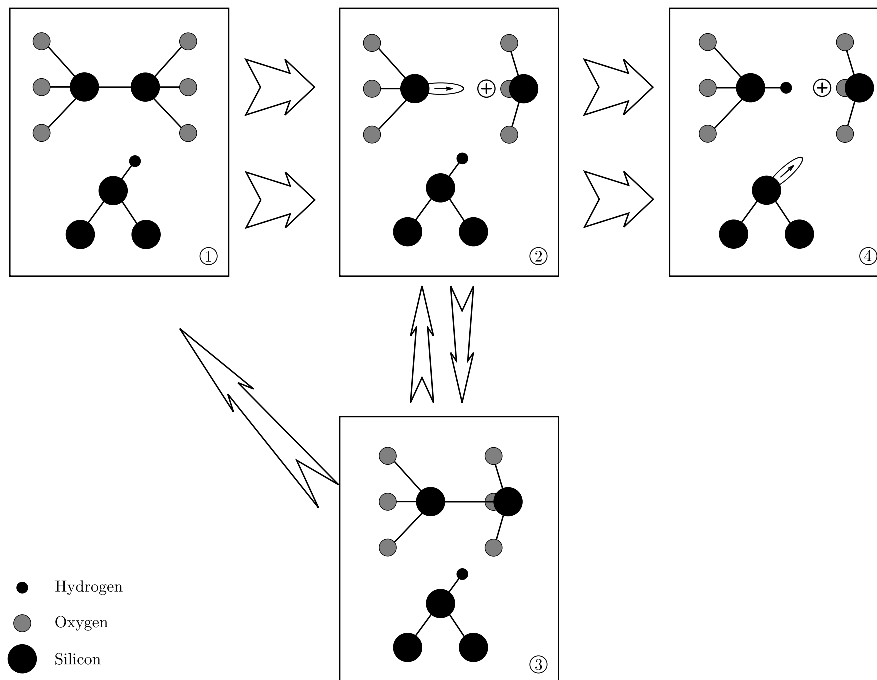


Fig. 19: Schematically depicted are the transitions of the GRASSER model. State 1 is the precursor state, the neutral oxygen vacancy. State 2 is an E'_γ center. State 3 is the electrically neutral but not fully relaxed state of the E'_γ center. In state 4, the hydrogen which moved away from the interface dangling bond, effectively locked the defect, resulting in a nearly permanent degradation.

Chapter II

Experimentals

This second part of the thesis is dedicated to a discussion about various different experimental methods used to characterize a MOS structure or a MOSFET and its degradation and recovery dynamics due to NBTI. All measurement techniques presented here are important tools for the findings obtained during the work on the thesis.

Because of the limitations of the measurement equipment used for the experiments in this thesis, all presented techniques are *electrical* characterizations of a MOS structure or a MOSFET.

4 Experimental determination of transistor parameters

The fabrication of a transistor on a Si wafer consists of several subsequent process steps. The most important steps are the doping of the Si bulk, the oxidation of the gate oxide, the deposition of the gate poly and the ion implantation of the source/drain regions. Due to fabrication variances and the large number of different process steps, it is difficult to obtain characteristic MOSFET or MOS structure parameters during fabrication. Thus, measurement techniques are needed to determine important parameters of the completely fabricated device.

4.1 Gate oxide electric field

The value of the oxide field during NBTS is reported to be an important parameter for NBTI induced degradation [HDP06]. The goal of this section is to derive an expression for the value of the oxide field as a function of the voltage applied to the gate $\mathcal{E}_{\mathrm{ox}}(V_{\mathrm{G}})$.

Theoretical calculation of $\mathcal{E}_{\mathrm{ox}}$

From a theoretical point of view, $\mathcal{E}_{\mathrm{ox}}$ can be calculated from equation (21) on page 15 by using GAUSS's law and neglecting any interface or oxide charges as

$$\begin{aligned}
\mathcal{E}_{\mathrm{ox}} &= \frac{Q_{\mathrm{G}}}{\varepsilon_{\mathrm{SiO2}}} \\
&= -\frac{Q_{\mathrm{S}}}{\varepsilon_{\mathrm{SiO2}}} \\
&= \frac{\varepsilon_{\mathrm{Si}}\mathcal{E}_{\mathrm{S}}}{\varepsilon_{\mathrm{SiO2}}} \\
&= \frac{\varepsilon_{\mathrm{Si}}}{\varepsilon_{\mathrm{SiO2}}} \frac{\sqrt{2}kT}{qL_{\mathrm{D}}} F\left(\frac{q\psi_{\mathrm{S}}}{kT}, \frac{n(\infty)}{p(\infty)}\right). \qquad (63)
\end{aligned}$$

This expression gives $\mathcal{E}_{\mathrm{ox}}$ as a function of the surface potential ψ_{S}. Now there is need to acquire an equation for the surface potential as a function of the gate voltage. V_{ox} in equation (22) on page 16 can be expressed as $V_{\mathrm{ox}}=\mathcal{E}_{\mathrm{ox}}T_{\mathrm{ox}}$ and so it follows with $\mathcal{E}_{\mathrm{ox}}$ from the above equation that

$$V_{\mathrm{G}} = V_{\mathrm{FB}} + \frac{\varepsilon_{\mathrm{Si}}}{C_{\mathrm{ox}}} \frac{\sqrt{2}kT}{qL_{\mathrm{D}}} F\left(\frac{q\psi_{\mathrm{S}}}{kT}, \frac{n(\infty)}{p(\infty)}\right) + \psi_{\mathrm{S}}, \qquad (64)$$

neglecting any oxide or interface charge. The inverse function of this equation would give $\psi_S(V_G)$. There is no analytical inverse but for the most applications it is sufficient to calculate the inverse numerically. Figure 20 shows the theoretical dependence of the electric gate oxide field on the gate voltage.

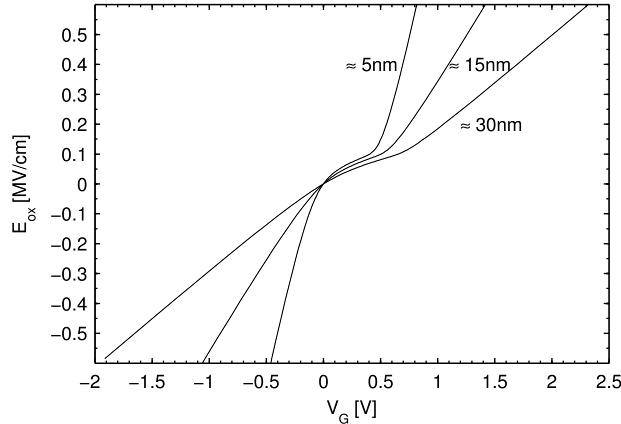


Fig. 20: Theoretical dependence of the gate oxide electric field on the gate voltage for a p doped MOS structure with n^{++} gate poly. Depicted are the curves for an effective donor doping $N_D = 10^{16}\,\mathrm{cm}^{-3}$ at room temperature for three different oxide thicknesses. $V_{FB} = \phi_{ms}$ is considered to be zero. The kinks in the characteristics are due to charges stored in the semiconductor. The left kink at 0 V occurs at the flat–band voltage, while the right kink is roughly at the threshold voltage.

This theoretical derivation neglects for example a possible voltage drop across the gate poly. A calculation of the oxide field which takes the voltage drop across the poly into account can be found in section 4.2.

Theoretical approximations of $\mathcal{E}_{ox}$

The theoretical relation between the oxide field and the gate voltage of equation (63) and (64) can be approximated by a linear relationship in either strong inversion or strong accumulation.

From Table 1 on page 19, $F \approx \mathrm{sgn}(\psi_S)\sqrt{n(\infty)/p(\infty)}\exp(q|\psi_S|/2kT)$ for strong accumulation and the F term in equation (64) will dominate over ψ_S. It follows that

$$F\left(\frac{q\psi_S}{kT}, \frac{n(\infty)}{p(\infty)}\right) \approx \frac{V_G - V_{FB}}{\frac{\varepsilon_{Si}}{C_{ox}} \frac{\sqrt{2}kT}{L_D}}, \tag{65}$$

which can be used in equation (63) to derive

$$\mathcal{E}_{ox}(V_G) \approx \frac{V_G - V_{FB}}{T_{ox}} \tag{66}$$

for strong accumulation. Equation (66) also holds for strong inversion. This derivation neglects any oxide or interface charge, thus $V_{FB} = \phi_{ms}$.

Equation (66) gives a very rough estimation of the oxide field which may be appropriate for devices with very thick oxides or for very high oxide fields.

Table 2: Approximated values for ψ_{S} for different substrate types in a MOS structure in strong inversion/accumulation. ($\phi_{\mathrm{Bn}} > 0$ and $\phi_{\mathrm{Bp}} < 0$) The column *region* refers to whether the semiconductor is driven into inversion (inv) or accumulation (acc) with the according polarity of V_{G}. It is assumed, that the intrinsic energy level lies exactly in the middle of the band gap.

substrate	V_{G}	region	ψ_{S}
n	< 0	inv	$-\left(\frac{E_{\mathrm{G}}}{2q} + \phi_{\mathrm{Bn}}\right)$
n	> 0	acc	$+\left(\frac{E_{\mathrm{G}}}{2q} - \phi_{\mathrm{Bn}}\right)$
p	< 0	acc	$-\left(\frac{E_{\mathrm{G}}}{2q} + \phi_{\mathrm{Bp}}\right)$
p	> 0	inv	$+\left(\frac{E_{\mathrm{G}}}{2q} - \phi_{\mathrm{Bp}}\right)$

However, for common oxide fields applied during bias temperature stress (BTS), this expression is usually too inaccurate. The approximation can be improved by taking ψ_{S} in equation (64) into account. This gives

$$\mathcal{E}_{\mathrm{ox}}\left(V_{\mathrm{G}}\right) \approx \frac{V_{\mathrm{G}} - V_{\mathrm{FB}} - \psi_{S}\left(V_{\mathrm{G}}\right)}{T_{\mathrm{ox}}}. \tag{67}$$

To express ψ_{S} in strong inversion or accumulation, it is justifiable to argue, that the bands in the Si substrate bend maximally in strong inversion or accumulation. Neither the conduction nor the valence band energy may intersect with the FERMI level. The maximum values for ψ_{S} for n and p doped substrates and for a positive or negative gate voltage are given in Table 2.

When assuming that the FERMI level energy in the gate poly is either E_{C} for a n^{++} doped gate poly or E_{V} for a p^{++} doped gate poly, insertion of the expressions of Table 2 into equation (67) gives practical approximations for the electrical gate oxide field in strong inversion or accumulation. I will calculate the approximation for a n^{++}/n MOS structure in strong inversion ($V_{\mathrm{G}} < 0$) as an example. The band bending situation is depicted in Fig. 21. The work function difference ϕ_{ms} and thus the flat–band voltage V_{FB} is

$$V_{\mathrm{FB}} = -\left(\frac{E_{\mathrm{G}}}{2q} - \phi_{\mathrm{Bn}}\right) \tag{68}$$

in a n^{++}/n MOS structure when assuming the FERMI level in the gate poly to be close to the conduction band energy and the intrinsic energy to be in the middle of the band gap. Inserting the flat–band voltage and the appropriate entry of Table 2 into equation (67) gives

$$\begin{aligned} \mathcal{E}_{\mathrm{ox}}\left(V_{\mathrm{G}}\right) &\approx \frac{V_{\mathrm{G}} + \left(\frac{E_{\mathrm{G}}}{2q} - \phi_{\mathrm{Bn}}\right) + \left(\frac{E_{\mathrm{G}}}{2q} + \phi_{\mathrm{Bn}}\right)}{T_{\mathrm{ox}}} \\ &\approx \frac{V_{\mathrm{G}} + \frac{E_{\mathrm{G}}}{q}}{T_{\mathrm{ox}}}, \end{aligned} \tag{69}$$

for a negative gate voltage and thus a negative electrical oxide field. The value of the flat–band voltage for the other combinations of gate poly doping and substrate doping types are listed in Table 3. The derivations of the approximating

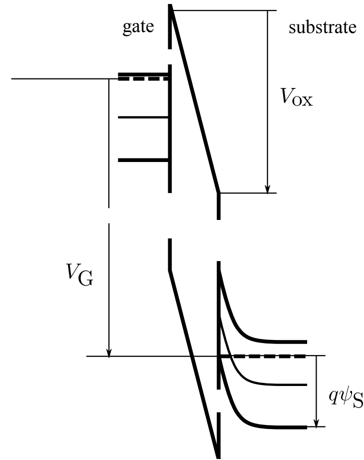


Fig. 21: Band diagram of a n^{++}/n MOS structure in strong inversion. The bands in the Si substrate are bent so that the valence band is pinned to the FERMI level.

Table 3: Value of the flat–band voltage V_{FB} for all possible combinations of gate poly doping and substrate doping types. Any additional interface or oxide charges which would affect the value of the flat–band voltage are neglected. It is assumed that the FERMI level energy of the gate poly is either E_{C} for n^{++} doped gate poly or E_{V} for p^{++} doped gate poly. The intrinsic energy level is assumed to be exactly in the middle of the band gap.

poly	substrate	V_{FB}
n^{++}	n	$-\left(\frac{E_{\text{G}}}{2q} - \phi_{\text{Bn}}\right)$
n^{++}	p	$-\left(\frac{E_{\text{G}}}{2q} - \phi_{\text{Bp}}\right)$
p^{++}	n	$+\left(\frac{E_{\text{G}}}{2q} + \phi_{\text{Bn}}\right)$
p^{++}	p	$+\left(\frac{E_{\text{G}}}{2q} + \phi_{\text{Bp}}\right)$

function for other combinations of gate poly doping and substrate doping types are very similar and the results are listed in Table 4.

Experimental determination of $\mathcal{E}_{\text{ox}}$ from tunneling characteristics

The FOWLER–NORDHEIM tunneling current (equation (36) on page 24) gives a direct relationship between the gate oxide electric field and the tunneling current. The FOWLER–NORDHEIM tunneling current is only dependent on the gate oxide field $\mathcal{E}_{\text{ox}}$, the tunneling barrier E_{B} and the effective mass of the electron in SiO_2 m_{SiO_2}. The tunneling barrier height depends on the type of carriers which form the tunneling current. The carrier type itself depends highly on the number of carriers in the energy bands on either side of the gate oxide. The carrier type tunneling can be experimentally determined by carrier separation experiments as described in [SMPD98].

Table 4: List of the value used in an approximating function for the electrical oxide field in strong inversion or accumulation. V in the Table is the value which has to be inserted into the equation $\mathcal{E}_{\mathrm{ox}}(V_{\mathrm{G}}) \approx (V_{\mathrm{G}} - V)/T_{\mathrm{ox}}$ in order to calculate the electrical oxide field. The column *region* refers to whether the semiconductor is driven into inversion (inv) or accumulation (acc) with the according polarity of V_{G}.

poly	substrate	V_{G}	region	V
n^{++}	n	< 0	inv	$-E_{\mathrm{G}}/q$
n^{++}	n	> 0	acc	0
n^{++}	p	< 0	acc	$-E_{\mathrm{G}}/q$
n^{++}	p	> 0	inv	0
p^{++}	n	< 0	inv	0
p^{++}	n	> 0	acc	E_{G}/q
p^{++}	p	< 0	acc	0
p^{++}	p	> 0	inv	E_{G}/q

The Fowler–Nordheim tunneling current is a reliable tool to determine the oxide field during BTS. Unfortunately, the tunneling current density is very low, making it difficult to reasonably resolve the current signal under conventional BTS fields. A current density in the $\mathrm{pA/cm^2}$ regime is typical for a BTS oxide fields $< 7\,\mathrm{MV/cm}$ [RVWHMAGGS08]. If currents down to $1\,\mathrm{nA}$ can be measured, the device area would have to be in the $\mathrm{cm^2}$ range in order to be able to resolve the tunneling current during stress. For common devices in the $\mu\mathrm{m}^2$ to mm^2 area regime, the tunneling current during BTS is usually too small to be measured directly.

Another way to use the Fowler–Nordheim current to characterize the oxide field during BTS is to measure a Fowler–Nordheim characteristic at higher fields and extrapolate the characteristic to lower voltages or oxide fields. Because the oxide field is not known a priori, the Fowler–Nordheim current has to be extrapolated in a $J_{\mathrm{FN}}(V_{\mathrm{G}})$ plot, assuming a linear relationship between the oxide field and the gate voltage. As described before, this assumption is reliable at high voltages in strong inversion or accumulation. In Fig. 22, such an extrapolation was performed for three equivalent devices having different gate oxide thicknesses.

It is possible to avoid an assumption regarding the relationship between the oxide field and the gate voltage when calculating the inverse of the Fowler–Nordheim expression equation (36) on page 24. There is no analytical solution for the inverse but it can be calculated numerically. With that, it is possible to draw a $\mathcal{E}_{\mathrm{ox}}(J_{\mathrm{FN}}(V_{\mathrm{G}}))$ plot. To extrapolate the data to common BTS fields there is still need to make an assumption on the $\mathcal{E}_{\mathrm{ox}}(V_{\mathrm{G}})$ relationship for low fields.

The extrapolation of experimentally gathered data may be uncertain and unreliable, especially when the target value is far away from the experimental data. Without any prior assumption of the relationship between the oxide field and the gate voltage, it is not possible to include a theoretical Fowler–Nordheim current characteristic into a $J_{\mathrm{FN}}(V_{\mathrm{G}})$ plot. Thus, it is not possible to extract $\mathcal{E}_{\mathrm{ox}}(V_{\mathrm{G}})$ unambiguously from the Fowler–Nordheim plot. However, a certain

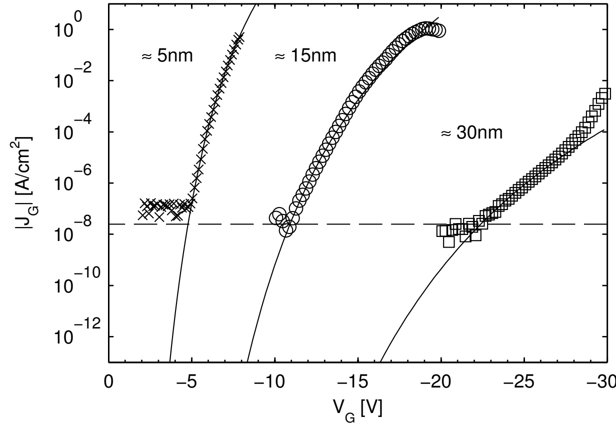


Fig. 22: Gate oxide tunneling current density versus voltage applied to the gate for three equivalent n^{++}/n MOS structures with different gate oxide thicknesses (symbols). The lines are least square fits of the function $J(V_G) = \alpha V_G^2 \exp(-\beta/V_G)$ to a part of the experimental data. The data of the 5 nm oxide thickness device agrees best with the fit. The data for the devices with roughly 15 nm and 30 nm oxide thickness deviate from the ideal Fowler–Nordheim relationship at higher voltages because of oxide damaging effects. These regions were excluded for calculating the data fit.

current density value corresponds reliably to a certain electric field, provided the type of tunneling carriers is known. Using the Fowler–Nordheim current is highly fruitful for conceptual studies, because it reflects the physical gate oxide electric field directly. However, a disadvantage is, that the recording of a tunneling characteristic usually incorporates charge into the gate oxide which finally leads to a full break–down of the gate oxide due to a conducting path of incorporated charge. Thus, the device is destroyed. So the $\mathcal{E}_{ox}(V_G)$ relationship has to be transferred to the device under test, assuming the destroyed device and the device under test to be equivalent.

For devices in the μm^2 to mm^2 area regime, the Fowler–Nordheim current characteristic is only observed for oxide thicknesses in the range of roughly 5 nm to 30 nm. In a device with a gate oxide thinner than roughly 5 nm, direct tunneling usually dominates, while devices with a gate oxide thicker than roughly 30 nm suffer from gate oxide damage already at low electric fields because of the high voltage drop across the oxide. Voltage driven degradation mechanisms will be explained in more detail in section 7.1.

Experimental determination of $\mathcal{E}_{ox}$ from CV measurements

An impedance meter measures the electrical impedance of a device under test by subjecting the device to a small high frequent alternating current (AC) sinusoidal oscillation. A MOS capacitor is usually regarded to be a capacitor parallel to a resistor. The capacitance value is calculated from the measured

electrical impedance $Z = |Z|e^{j\theta}$ by

$$C = \frac{|\sin\theta|}{2\pi f|Z|}, \theta \neq \frac{\pi}{2} + k\pi, k = 0, 1, 2, 3, \ldots, \tag{70}$$

where f is the AC signal frequency. The parallel resistor is usually described by the electrical conductance

$$G = \frac{|\cos\theta|}{|Z|}, \theta \neq \frac{\pi}{2} + k\pi, k = 0, 1, 2, 3, \ldots. \tag{71}$$

The ideal physical capacity at a certain gate voltage V_G, when measuring at the gate of a MOS capacitor, is

$$C(V_G) = \frac{dQ_G}{dV_G}(V_G), \tag{72}$$

the differential change of the charge in the gate related to the differential change of the voltage applied to the gate. The idea is to integrate an experimentally measured CV curve to measure the charge on the gate and thus the electric oxide field.

GAUSS's law in its integral form is

$$\frac{Q}{\varepsilon} = \frac{1}{|S|}\oint_S \vec{\mathcal{E}} \cdot \vec{u}\, da, \tag{73}$$

where ε is the permittivity, S is a closed surface and $|S|$ its area, $\vec{\mathcal{E}}$ is the electric field as a vector, $\vec{u}$ is a unit vector pointing outward perpendicular to the surface S and da is a differential small area of the surface S. The closed surface S in a MOS structure is roughly as depicted in Fig. 23. The electric field in the

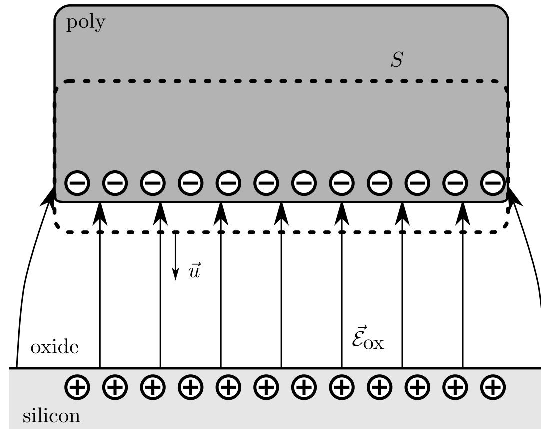


Fig. 23: Schematic drawing of a MOS structure and a possible closed surface S for integrating GAUSS's law. It is assumed that a negative voltage is applied to the gate so that negative carriers accumulate at the interface poly–SiO_2 ($Q_G < 0$) and positive carriers accumulate at the interface SiO_2–Si. The resulting electric oxide field is negative.

gate poly can be considered to be zero because of the poly being a fairly good conductor. The scalar product gives $-\mathcal{E}_{\mathrm{ox}}$ in the region between the poly and the Si and is approximately zero everywhere else. Integrating equation (73) yields

$$\mathcal{E}_{\mathrm{ox}} = -\frac{Q_{\mathrm{G}}}{\varepsilon_{\mathrm{SiO2}}}. \tag{74}$$

Q_{G} is the charge in the gate related to the device area in $\mathrm{C/cm^2}$. Q_{G} can be derived experimentally by integrating a LF CV curve.

$$\begin{aligned}\int C\left(V_{\mathrm{G}}\right) \mathrm{d}V_{\mathrm{G}} &= \int \frac{\mathrm{d}Q_{\mathrm{G}}}{\mathrm{d}V_{\mathrm{G}}}\left(V_{\mathrm{G}}\right) \mathrm{d}V_{\mathrm{G}} \\ &= Q_{\mathrm{G}}\left(V_{\mathrm{G}}\right) + \mathrm{const.}\end{aligned} \tag{75}$$

It is important to integrate a LF CV curve and not a HF CV curve because all charges have to be able to follow every change of the AC signal to correctly determine the charge density in the gate. The determination of the constant in equation (75) is not straight forward. One possible way is to find a gate voltage for which the charge in the gate is zero. As described in section 1.3, when neglecting any additional oxide or interface charges, this clearly happens at the flat–band voltage. Consequently, the electric oxide field is given by [RVWHMAGGS08]

$$\mathcal{E}_{\mathrm{ox}}\left(V_{\mathrm{G}}\right) = \frac{\int_{V_{\mathrm{FB}}}^{V_{\mathrm{G}}} C(V_{\mathrm{G}})\mathrm{d}V_{\mathrm{G}}}{\varepsilon_{\mathrm{SiO2}}}. \tag{76}$$

For practical reasons it might be more convenient to measure a LF CV characteristic from an arbitrary voltage V_1, over the flat–band voltage V_{FB}, to the target voltage V_{G} and subtracting the electric field at V_{FB} from the characteristic. That is to say, V_1 has to hold the requirement $\mathrm{sgn}\left(V_{\mathrm{G}} - V_{\mathrm{FB}}\right) = \mathrm{sgn}\left(V_{\mathrm{FB}} - V_1\right) \neq 0$. Figure 24 shows the dependence of the oxide field on the gate voltage for a pMOS transistor with n^{++} gate poly with roughly 30 nm gate oxide thickness.

This method might be the most exact method to determine the electric oxide field. It is directly related to the charge on the gate and the underlying theory relies on a small number of assumptions. The direct measurement of the capacitance implicitly contains for example the depletion of the gate poly or quantum mechanical effects like the quantization of the energy levels in the inversion layer. Although, it might become inaccurate when the gate area of the device becomes as small that the parasitic oxide field at the edges of the gate poly becomes important. Another disadvantage is, that oxide charges have to be neglected in order to argue that the oxide field is zero at the flat–band voltage. This assumption is justifiable for modern high–quality gate oxides, but it is not yet totally clear, which impact charges at the interface have on the oxide field. This is of particular importance for devices of CMOS technology, where often a slight ion implantation at the Si–$\mathrm{SiO_2}$ interface is used to adjust the value of the threshold voltage.

In order to determine the value of the gate voltage needed for a certain NBTS oxide field, the LF CV characteristic has to be measured from roughly the flat–band voltage toward negative voltages. The gate voltage value to where the LF CV characteristics has to be measured is not known a priori, but can be estimated by the approximations described in the previous sections. If the LF CV curve is measured at high temperatures, the measurement itself has to be

regarded as stress. The stress does not influence the LF CV curve measured, but the device is not virgin any more. Subsequent NBTI experiments on this device may be erroneous because of the unknown influence of the pre-stress through the LF CV measurement. The LF CV curve can be measured on one device and the result can be transfered to a neighboring device. This approach introduces errors because of unavoidable variations of the oxide thickness on a wafer. To reduce the temperature for the measurement of the LF CV curve, is a second approach, which has been used for the results obtained in this master thesis. It is possible to measure the electric oxide field also at low temperatures because the influence of the temperature on the LF CV characteristic is negligible when measuring on a MOSFET, where the minority carriers are supplied by the source/drain regions. The temperature is important when measuring a LF CV curve on a MOS structure, because the minority carrier generation is highly dependent on the temperature [Sch06].

4.2 Gate poly doping concentration

The work function difference is determined by the doping concentration of the materials on either side of the gate oxide. In order to correctly determine the work function difference, a measurement technique to obtain the doping concentration of the gate poly is therefore needed.

A method to determine electrically the effective impurity concentration in the gate poly N_{p} from FOWLER–NORDHEIM tunneling characteristics was proposed in 1997 [GFSLWCH97]. The key to determine N_{p} is to evaluate the voltage drop V_{p} across a layer in the gate poly due to the depletion of the majority carriers in the poly.

In order to measure N_{p}, a connection between the voltage drop V_{p} and the poly impurity concentration N_{p}, as well as the exact knowledge of other voltage drops in the MOS structure are needed. N_{p} can then be obtained by fitting the analytical expression for the tunneling current as a function of V_{G} to the experimental data.

Calculation of the poly depletion voltage drop

The band bending in the gate poly can be calculated in a similar way as the band bending in any other semiconductor using equation (21) on page 15. The gate poly is usually either highly n doped (n^{++}) or highly p doped (p^{++}) and so the corresponding FERMI levels are either near the conduction or near the valence band energy of Si. The maximum band bending is determined by the limitation, that the FERMI level may not intersect with E_{C} or E_{V} (pinning of the bands to the FERMI level). Consequently, positive (downward) band bending in a n^{++} doped poly and negative (upward) band bending in a p^{++} doped poly can be neglected. A voltage drop across the gate poly will thus only happen when the voltage applied to the gate is such, that the poly is driven into depletion or inversion. For a n^{++} doped gate poly this is the case when applying a positive voltage to the gate and for a p^{++} doped gate poly this happens for a negative gate voltage. For the depletion region of a doped semiconductor the approximation $F \approx \sqrt{q\psi_{\mathrm{S}}/kT}$ is sufficiently accurate. From equation (23) on page 18 follows

$$\varepsilon_{\mathrm{Si}}\mathcal{E}_{\mathrm{S}} = \sqrt{2\varepsilon_{\mathrm{Si}} q V_{\mathrm{p}} N_{\mathrm{p}}} \tag{77}$$

where the band bending ψ_S is the voltage drop across the poly V_p and the impurity concentration $p(\infty)$ is N_p in this case. Assuming no additional charges at the interface poly–SiO_2, the electric displacement field in the poly and in the SiO_2 must be equivalent

$$\varepsilon_{Si}\mathcal{E}_S = \varepsilon_{SiO2}\mathcal{E}_{ox}. \tag{78}$$

Rearranging gives

$$V_p = \frac{\varepsilon_{SiO2}^2\mathcal{E}_{ox}^2}{2q\varepsilon_{Si}N_p}. \tag{79}$$

This derivation is based on BOLTZMANN statistics and thus might be inappropriate when the gate poly is degenerately doped. Mind that the voltage drop across the poly has an upper limit of roughly the band gap E_G/q, because of the pinning of the FERMI level at the valence or conduction band edge energies.

Calculating the oxide field considering gate poly depletion

The FOWLER–NORDHEIM current density (equation (36), page 24) is a function of the electric oxide field $\mathcal{E}_{ox}$, the effective mass in the oxide m_{SiO_2} and the tunneling barrier height E_B. The effective mass and the tunneling barrier height depend on the type of tunneling carriers. For electrons tunneling from the Si conduction band the barrier height is $E_B = \chi_{Si} - \chi_{SiO2} \approx 3.15\,\text{eV}$ at room temperature. The electron ends up in the conduction band of the SiO_2, where its effective mass is around a third of the mass of a free electron [DVMMH95]. The oxide field as a function of the voltage applied to the gate can be determined as described in section 4.1, but the voltage drop due to gate poly depletion has to be considered. Adding V_p to equation (22) on page 16 gives

$$\begin{aligned}
V_G &= V_{ox} + V_{FB} + \psi_S + V_p \\
\mathcal{E}_{ox}T_{ox} &= V_G - V_{FB} - \psi_S - V_p \\
0 &= \mathcal{E}_{ox}^2\frac{\varepsilon_{SiO2}^2}{2q\varepsilon_{Si}N_p} + \mathcal{E}_{ox}T_{ox} - (V_G - V_{FB} - \psi_S) \\
\mathcal{E}_{ox} &= -\frac{q\varepsilon_{Si}N_p}{\varepsilon_{SiO2}^2}T_{ox} \pm \sqrt{\left(\frac{q\varepsilon_{Si}N_p}{\varepsilon_{SiO2}^2}\right)^2 T_{ox}^2 + \frac{2q\varepsilon_{Si}N_p}{\varepsilon_{SiO2}^2}(V_G - V_{FB} - \psi_S)}.
\end{aligned} \tag{80}$$

The oxide field is now only dependent on N_p, T_{ox}, V_G, V_{FB} and ψ_S. V_{FB} is assumed to be equal to ϕ_{ms}, that is to say oxide charges or interface traps are neglected. ψ_S might be determined for different gate biases by some means or other. However, it is justifiable to approximate ψ_S with the maximum possible band bending to either the valence or the conduction band energy. The reason for this is, that usually the MOS structure has to be driven to either strong accumulation or strong inversion to measure a tunneling current above the detection limit of common measurement equipments in devices in the mm^2 or μm^2 area regime. The values for ψ_S for different substrate types and gate biases are listed in Table 2.

The curve fitting function

The analytical expression for the FOWLER–NORDHEIM tunneling current is (see section 1.6)

$$J_{\mathrm{FN}} = \frac{q^2}{16\pi^2\hbar E_{\mathrm{B}}}\mathcal{E}_{\mathrm{ox}}^2 \exp\left(-\frac{4\sqrt{2qm_{\mathrm{SiO_2}}}E_{\mathrm{B}}^{3/2}}{3\hbar\mathcal{E}_{\mathrm{ox}}}\right) \quad (81)$$

where

$$\mathcal{E}_{\mathrm{ox}} = -\frac{q\varepsilon_{\mathrm{Si}}N_{\mathrm{p}}}{\varepsilon_{\mathrm{SiO2}}^2}T_{\mathrm{ox}} \pm \sqrt{\left(\frac{q\varepsilon_{\mathrm{Si}}N_{\mathrm{p}}}{\varepsilon_{\mathrm{SiO2}}^2}\right)^2 T_{\mathrm{ox}}^2 + \frac{2q\varepsilon_{\mathrm{Si}}N_{\mathrm{p}}}{\varepsilon_{\mathrm{SiO2}}^2}\left(V_{\mathrm{G}} - V_{\mathrm{FB}} - \psi_{\mathrm{S}}\right)}. \quad (82)$$

The resulting expression is fairly bulky, but can be implemented into a numerical computer program in order to extract the two parameters N_{p} and T_{ox} by fitting the expression to the experimentally measured data. I remark that, due to the exponential function in the expression of the FOWLER–NORDHEIM current, the numerical fitting problem is reduced considerably in its computing power demand when using the logarithm of the expression. Unfortunately, the optimization problem is still badly conditioned and good knowledge in numerical techniques is needed to bring fits of experimental data to a rapid convergence. The numerical problem is greatly reduced in its complexity when the oxide thickness is determined by other methods, leaving the poly doping concentration as the sole fitting parameter.

Experimental verification and discussion

The method to determine the gate poly doping concentration proposed by [GF-SLWCH97] relies on the rigorously studied quantum mechanical tunneling effect. Also the expression for the gate poly voltage drop is derived straightforwardly providing a non–degenerately doped gate poly. I remark that this assumption is a weak point of this method because the gate poly is often degenerately doped. Another disadvantage of the method might be that the acquisition of the experimental data through a gate voltage sweep usually destroys the device under test because of the charge incorporated in the gate oxide. Using thinner oxide devices allows the measurement of the tunneling current at low voltages and thus the probability for oxide damaging effects to occur is reduced. Figure 25 shows the method applied to a measurement of three equivalent devices with three different oxide thicknesses. Provided that the oxide thicknesses are known, one single value for N_{p} fits all three data sets accurately in the region where the measured data can be considered to be only the tunneling current.

4.3 Surface potential

The value of the surface potential determines a lot of other important parameters of a MOS structure or a MOSFET. So is for example the occupancy state of interface traps determined by the position of the FERMI level with respect to the band edges. The position of the FERMI level is reflected by the value of the surface potential. Because of the creation of interface traps during NBTS, it is highly beneficial for NBTI related research to know the value of the surface potential with respect to the gate voltage.

Berglund proposed an experimental method to obtain the surface potential ψ_S as a function of the applied gate voltage V_G [Ber66; SN06]. In an ideal MOS structure the applied gate voltage will drop partly across the oxide and partly across the semiconductor

$$\mathrm{d}V_\mathrm{G} = \mathrm{d}V_\mathrm{ox} + \mathrm{d}\psi_\mathrm{S}. \tag{83}$$

This treatment neglects for instance a voltage drop due to a possible depletion of the gate poly. This is justifiable because of the significantly higher doping concentration of the gate poly compared to the substrate. The band bending of a semiconductor decreases with increasing doping concentration, as described in section 1.3. So in a region around the flat–band voltage the potential of the substrate surface will change strongly compared to the negligible band banding of the gate poly.

One measures the capacitance

$$C\left(V_\mathrm{G}\right) = \frac{\mathrm{d}Q_\mathrm{G}}{\mathrm{d}V_\mathrm{G}} \tag{84}$$

of a MOS structure by recording a LF CV curve. The incremental charge can be expressed as $\mathrm{d}Q_\mathrm{G} = C_\mathrm{ox}\mathrm{d}V_\mathrm{ox}$. Combining this relation with the last two equations (83) and (84) one obtains

$$\begin{aligned} C\left(V_\mathrm{G}\right) &= \frac{\mathrm{d}Q_\mathrm{G}}{\mathrm{d}V_\mathrm{G}} \\ &= C_\mathrm{ox}\frac{\mathrm{d}V_\mathrm{G} - \mathrm{d}\psi_\mathrm{S}}{\mathrm{d}V_\mathrm{G}} \\ &= C_\mathrm{ox}\left(1 - \frac{\mathrm{d}\psi_\mathrm{S}}{\mathrm{d}V_\mathrm{G}}\right) \end{aligned} \tag{85}$$

and thus

$$\mathrm{d}\psi_\mathrm{S} = \left(1 - \frac{C\left(V_\mathrm{G}\right)}{C_\mathrm{ox}}\right)\mathrm{d}V_\mathrm{G}. \tag{86}$$

By integrating (86) one obtains the surface potential as a function of the gate voltage

$$\psi_\mathrm{S}(V_1) - \psi_\mathrm{S}(V_2) = \int_{V_1}^{V_2}\left(1 - \frac{C\left(V_\mathrm{G}\right)}{C_\mathrm{ox}}\right)\mathrm{d}V_\mathrm{G} + \mathrm{const.} \tag{87}$$

The integration constant can be determined considering that the surface potential is close to zero at the flat–band voltage V_FB. Figure 26 shows the $\psi_\mathrm{S}V_\mathrm{G}$ relations for equivalent devices with different oxide thicknesses. The validity of the resulting $\psi_\mathrm{S}V_\mathrm{G}$ characteristic can easily be proved by comparing the maximum values of the curve with the energy band gap. The value of the energy band gap of Si for arbitrary temperatures is given by equation (1) on page 10. Mind that the surface potential approaches the valence and conduction band edges slowly and it may not reach the band edges within an experimentally observable region.

4.4 Flat–band voltage

As described in section 1.3, neglecting any interface or oxide charge, the flat–band voltage V_FB is the work function difference between the gate material and

the substrate ϕ_{ms}. The doping concentrations and thus the FERMI levels of the gate and the substrate materials can be determined experimentally as described in section 4.2 and 4.5. Consequently, there is no need for a measurement technique to obtain the flat–band voltage when neglecting additional charges.

However, considering a real device, oxide and interface charges have to be taken into account. The flat–band voltage is then

$$V_{\mathrm{FB}} = \phi_{\mathrm{ms}} - \frac{Q}{C_{\mathrm{ox}}} - \frac{Q_{\mathrm{IT}}\left(\psi_{\mathrm{S}}\right)}{C_{\mathrm{ox}}}, \tag{88}$$

where Q is an effective oxide charge defined as described in section 1.4. The signs in equation (88) mean that positive charges cause the flat–band voltage to be more negative.

Interface traps may change their occupancy state with gate bias in a very small time scale. In particular, the total effective interface charge is dependent on the surface potential ψ_{S}. In this section, oxide charges are assumed to be not capable of charging or discharging processes. The existence of border traps as defined in section 1.4 is therefore neglected in this section. If the interface charge can be determined by a different measurement, it is in principle possible to calculate the oxide charge by the experimental determination of the flat–band voltage. This is beneficial for NBTI related research because some researchers suggest that oxide charge is created during NBTS [GKGAHN09; HDP06] while others either neglect the build-up of oxide charge or attribute its creation to not NBTI related effects [AM05; MAVIGMSA07].

Experimental determination of V_{FB} from a CV characteristic

The most common method to determine the flat–band voltage is to compare a measured CV curve with a theoretically calculated CV curve. The theoretical LF CV characteristic is calculated in section 1.5, equation (32) on page 22. The derivation assumes an uniform doping profile in the substrate, a zero work–function difference and no additional charges. The gate oxide thickness, the substrate doping concentration and the temperature must be known to calculate the theoretical CV curve. The flat–band voltage is then obtained by aligning the experimental and the theoretical CV characteristics in the same plot. Thereby, the theoretical curve is shifted toward positive or negative voltages until the two curves coincide best. The voltage value used to shift the theoretical curve is the flat–band voltage of the device.

Figure 27 shows a comparison between the theoretical and the experimental LF CV curve after aligning.

Experimental determination of V_{FB} from the flat–band capacitance

The capacitance of the semiconductor at the flat–band voltage

$$C_{\mathrm{FB}} = \frac{1}{\frac{1}{C_{\mathrm{ox}}} + \frac{1}{C_{\mathrm{S}}(V_{\mathrm{FB}})}} \tag{89}$$

has to be determined from equation (32) on page 22 assuming, that the surface potential is zero at the flat–band voltage. Direct insertion of $\psi_{\mathrm{S}}= 0$ into equation 32 gives an indefinite result. So the exponential terms have to be expanded

according to $e^x \approx 1 + x + x^2/2$ to derive

$$C_S = \frac{\varepsilon_{Si}}{\sqrt{2}L_D} \frac{1 - e^{-\frac{q\psi_S}{kT}} - \frac{n(\infty)}{p(\infty)}\left(1 - e^{\frac{q\psi_S}{kT}}\right)}{F\left(\frac{q\psi_S}{kT}, \frac{n(\infty)}{p(\infty)}\right)} \approx \frac{\varepsilon_{Si}}{L_D} \tag{90}$$

In order to determine the flat–band voltage of a device, a HF CV characteristic is measured and compared with the theoretical value of the capacitance. The flat–band voltage of the device is where the HF CV equals C_{FB}.

Experimental determination of V_{FB} from the knee in a CV curve

Another way to experimentally determine the flat–band voltage of a device is to look for transitions in a plot of $1/C^2$ of a HF CV curve. The lower knee of this curve corresponds to the flat–band voltage [Sch06]. This knee can be found by searching for the strongest curvature of the $1/C^2$ plot. The strongest curvature is where the change of the derivative of the $1/C^2$ plot is the largest. Accordingly, the maximum of a plot

$$\frac{d^2 \frac{1}{C^2}}{dV_G^2} \tag{91}$$

gives the flat–band voltage of the device.

4.5 Acceptor/Donor doping concentration

The ionized acceptor/donor doping concentration in the Si substrate determines the FERMI level position in the bulk of the semiconductor. Consequently, the work function difference and thus the flat–band voltage of a device depends highly on the doping concentration of the substrate. The exact knowledge of the doping concentration in the bulk of the semiconductor is thus important to correctly analyze a device in the context of NBTI related research.

Already at medium temperatures (around the room temperature), most of the acceptor/donor atoms are ionized and thus $N_A \approx N_A^- / N_D \approx N_D^+$ can be used. The doping concentration is for example important for the determination of the work function difference ϕ_{ms}. The doping concentration can be determined electrically either by using a CV characteristic or from the threshold voltage of a MOSFET.

Experimental determination of N_A/N_D from the MOSFET threshold voltage

The definition of the threshold voltage of a MOSFET (equation (47) on page 28) is

$$V_{TH} = V_{FB} - 2\phi_{Bn} - \frac{\sqrt{2q\varepsilon_{Si}N_D(2\phi_{Bn} + V_B)}}{C_{ox}} \tag{92}$$

for a pMOS transistor device. The appearance of the voltage applied to the bulk with respect to the source V_B in the expression for the threshold voltage

suggests to measure the threshold voltage for different bulk biases. The ansatz is to calculate

$$\begin{aligned}\gamma &\equiv \frac{\mathrm{d}V_{\mathrm{TH}}}{\mathrm{d}\sqrt{2\phi_{\mathrm{Bn}}+V_{\mathrm{B}}}} \\ &= -\frac{\sqrt{2q\varepsilon_{\mathrm{Si}}N_{\mathrm{D}}}}{C_{\mathrm{ox}}} \\ N_{\mathrm{D}} &= \frac{\gamma^2 C_{\mathrm{ox}}^2}{2q\varepsilon_{\mathrm{Si}}}. \end{aligned} \tag{93}$$

The doping concentration can be determined from the slope γ of a plot of V_{TH} against $\sqrt{2\phi_{\mathrm{Bn}}+V_{\mathrm{B}}}$. Figure 28 shows an example for such a plot of a pMOS transistor. Since ϕ_{Bn} depends on the donor concentration ($\phi_{\mathrm{Bn}}= kT/q\ln\left(N_{\mathrm{D}}/n_{\mathrm{i}}\right)$) it is suitable to start with a guess for the bulk potential $0<\phi_{\mathrm{Bn}}<E_{\mathrm{G}}/(2q)$. With the resulting value for the donor concentration a new bulk potential is calculated and reinserted into the calculation. Usually one or two iterations are sufficient to calculate ϕ_{Bn} accurately to a few mV [Sch06]. It is recommendable to determine the threshold voltage by a current criterion.

Using the equation

$$x=\sqrt{\frac{2\varepsilon_{\mathrm{Si}}\left(2\phi_{\mathrm{Bn}}+V_{\mathrm{B}}\right)}{qN_{\mathrm{D}}}} \tag{94}$$

it is possible to calculate a doping concentration profile along the spatial depth x from the Si–SiO_2 interface into the bulk of the semiconductor. In Fig. 29 the profile obtained with this method is compared to other methods which are explained in the following sections.

Experimental determination of $N_{\mathrm{A}}/N_{\mathrm{D}}$ from CV measurement

To derive an expression for the doping concentration from a CV curve the charge situation in a MOS structure is approximated by a step function. That is to say it is assumed that the charge in the semiconductor is uniformly distributed in a charge layer with a thickness x and zero deeper in the bulk of the semiconductor. Under these assumptions the charge in the semiconductor can then be calculated as

$$Q_{\mathrm{S}}=q\int_0^x\left(p-n+N_{\mathrm{D}}-N_{\mathrm{A}}\right)\mathrm{d}x'\approx q\int_0^x N_{\mathrm{D}}\mathrm{d}x' \tag{95}$$

and so the capacitance is

$$C=-\frac{\mathrm{d}Q_{\mathrm{S}}}{\mathrm{d}V}\approx -qN_{\mathrm{D}}(x)\frac{\mathrm{d}x}{\mathrm{d}V} \tag{96}$$

when neglecting the term $\mathrm{d}N_{\mathrm{D}}(x)/\mathrm{d}V$. The capacitance of an ideal parallel plate capacitor with thickness x is $C=\varepsilon_{\mathrm{Si}}/x$, which is differentiated with respect to V to derive

$$\frac{\mathrm{d}C}{\mathrm{d}V}=-\frac{C^2}{\varepsilon_{\mathrm{Si}}}\frac{\mathrm{d}x}{\mathrm{d}V}. \tag{97}$$

Using this equation to express the derivative $\mathrm{d}x/\mathrm{d}V$ gives for the donor concentration from equation (96)

$$N_{\mathrm{D}}=\frac{C^3}{q\varepsilon_{\mathrm{Si}}\frac{\mathrm{d}C}{\mathrm{d}V}}. \tag{98}$$

Using the idenity $\mathrm{d}(1/C^2)/\mathrm{d}V = -2/C^3 \mathrm{d}C/\mathrm{d}V$, the doping concentration can be calculated by [Sch06]

$$N_\mathrm{D}(x) = -\frac{2}{q\varepsilon_\mathrm{Si}\frac{\mathrm{d}1/C^2}{\mathrm{d}V}} \tag{99}$$

at the depth

$$x = \frac{\varepsilon_\mathrm{Si}}{C}. \tag{100}$$

Figure 29 shows the donor doping concentration for a pMOS transistor extracted from a HF CV characteristic, a LF CV characteristic and by using the MOSFET threshold voltage.

4.6 Density of interface traps

It is broadly accepted that NBTS leads to a build-up of the number of interface traps at the interface Si-SiO_2. Therefore, interface trap density sensitive measurement techniques are needed for NBTI related research.

A lot of different measurement techniques have been developed in the past to determine the density of interface traps [SN06; NB02; Sch06]. Among them a method called charge pumping (CP) is an advantageous and highly sensitive measurement technique, which has been used extensively in this master thesis. The idea behind CP is to create a net current from the source/drain regions of a MOSFET to the substrate over interface traps. The resulting current is then dependent on the total number of interface traps. This measurement technique was already developed in 1969 [BJ69].

Short description of the CP measurement principle

By applying a voltage around the V_TH to the gate of a MOSFET, the Si–SiO_2 interface becomes inverted and thus interface traps are filled with minority carriers from the source/drain regions. When the interface is than switched back to accumulation, some minority carriers may remain trapped because of the finite emission time constants of interface traps. During accumulation, the trapped minority carriers recombine with majority carriers provided by the substrate. Consequently, a charge transport has happened from the source/drain regions to the substrate because some charge was stored in interface traps during the transition period from inversion to accumulation. If this filling–storage–emission sequence is repeated by continuously pulsing the gate, the amount of charge transmitted with every cycle adds up to a measurable current, the CP current I_CP.

Voltage switches at the gate

To effectively pump charge from the source/drain regions to the substrate of a MOSFET, the gate bias is repeatedly switched from accumulation to inversion and vice versa. In practice, a pulse generator unit is used to generate a trapezoidal gate pulse. The trapezoidal pulse shape is sketched in Fig. 30. The high/low level $V_\mathrm{H}/V_\mathrm{L}$ is kept constant for a duration $t_\mathrm{H}/t_\mathrm{L}$. The transitions in between happen during t_r and t_f. In order to measure a CP current, the high level of the trapezoidal pulse must be above the threshold voltage and the low

level below the flat–band voltage of the MOSFET in order to sufficiently invert or accumulate the interface respectively.

Charge storage in interface traps

The key to the charge transport during CP is the storage of charge in interface traps during the transition from inversion to accumulation of the Si–SiO_2 interface. The energetic transitions of a carrier from the conduction band to the valence band or vice versa via traps within the band gap of the semiconductor is described with great accuracy by SHOCKLEY–READ–HALL (SRH) statistics [SN06]. The emission/capture time constants τ of interface traps are, according to the SRH theory, inversely proportional,

$$\tau = \frac{1}{\sigma v_{\mathrm{th}} N}, \tag{101}$$

to the capture/emission cross section σ, the thermal velocity v_{th} and the density of carriers N. The densities of carriers can be described with BOLTZMANN statistics as introduced in section 1.2. The density of carriers possibly contributing to a capture event of an interface trap is given in equation (10) and (12) on page 13

$$n \approx N_{\mathrm{C}} \mathrm{e}^{\frac{E_{\mathrm{F}} - E_{\mathrm{C}}}{kT}}, \text{ for electrons,} \tag{102}$$

$$p \approx N_{\mathrm{V}} \mathrm{e}^{-\frac{E_{\mathrm{F}} - E_{\mathrm{V}}}{kT}}, \text{ for holes.} \tag{103}$$

Thus, the capture time constants are dependent on the position of the FERMI level with respect to the valence and conduction band edges as

$$\tau_{\mathrm{c}n} \approx \frac{1}{\sigma v_{\mathrm{th}} N_{\mathrm{C}}} \mathrm{e}^{\frac{E_{\mathrm{C}} - E_{\mathrm{F}}}{kT}}, \text{ for electrons,} \tag{104}$$

$$\tau_{\mathrm{c}p} \approx \frac{1}{\sigma v_{\mathrm{th}} N_{\mathrm{V}}} \mathrm{e}^{\frac{E_{\mathrm{F}} - E_{\mathrm{V}}}{kT}}, \text{ for holes.} \tag{105}$$

Hence, the capture time constants for *electrons* decrease when the difference between the FERMI level and the *conduction band* decreases. Respectively, the capture time constants for *holes* decrease when the FERMI level approaches the *valence band*. With an increase of the temperature T the time constants always decrease. For the emission time constants the energy level of the interface trap E_{T} is important and the according time constants are [Aic07]

$$\tau_{\mathrm{e}n} \approx \frac{1}{\sigma v_{\mathrm{th}} N_{\mathrm{C}}} \mathrm{e}^{\frac{E_{\mathrm{C}} - E_{\mathrm{T}}}{kT}}, \text{ for electrons,} \tag{106}$$

$$\tau_{\mathrm{e}p} \approx \frac{1}{\sigma v_{\mathrm{th}} N_{\mathrm{V}}} \mathrm{e}^{\frac{E_{\mathrm{T}} - E_{\mathrm{V}}}{kT}}, \text{ for holes.} \tag{107}$$

Also these time constants decrease with increasing temperature.

The CP current in a pMOS

When a pMOS transistor is driven to inversion by applying a negative gate voltage (FERMI level position close to the valence band), holes are injected from the source/drain regions of the MOSFET to the Si–SiO_2 interface. Most interface

traps which provide hole capture time constants smaller than the high time of the trapezoidal pulse will capture holes. According to equation (104) and (105), the closer the FERMI level is to the valence band, the more interface traps will capture a hole, because in thermal equilibrium, traps with energy values below E_F become occupied by an electron. If an interface trap has a hole emission time constant which is smaller than its hole capture time constant, it is likely that the interface trap emits the hole again. According to the equations (106) and (107), interface traps located energetically close to the valence band have smaller emission time constants. The contribution of such traps to the CP current may be neglected safely. Consequently, the electrically active energy region where interface traps carry charges for CP is decreased, away from the valence band. So there is a theoretical limit for the lower part of the electrically active energy region measurable with CP because of the small emission times of traps near the band edges. It is assumed that the transition of the gate voltage from the high to the low level happens instantaneously, that is to say $t_\mathrm{f}= 0\,\mathrm{s}$. Interface traps which captured and kept holes during the inversion phase of the gate pulse may now loose the carrier when the emission happens during t_L. Those holes recombine with electrons of the substrate and thus charge has been transferred from the source/drain regions to the substrate. In accumulation (FERMI level close to the conduction band), traps energetically located close to the valence band emit their holes first because of their short emission time constants. Some interface traps energetically located closer to the conduction band may not emit their hole until t_L has elapsed. This fact causes another decrease of the active energy region for CP in the upper part of the band gap. The remaining active energy region is located within the band gap, roughly symmetrical around E_i. Later the interface is switched from accumulation back to inversion and the sequence starts all over again.

Extraction of the interface trap density

Finite rise and fall times cause the active energy interval for CP to decrease further [Aic07]. The active energy region, approximately centered within the band gap for a symmetrical pulse shape, can be calculated by [AN08; GMBDK84]

$$\Delta E_\mathrm{CP} = -kT \ln \left(v_\mathrm{th} \sigma n_\mathrm{i} \frac{|V_\mathrm{TH} - V_\mathrm{FB}|}{|\Delta V_\mathrm{G}|} \sqrt{t_\mathrm{r} t_\mathrm{f}} \right). \tag{108}$$

$|\Delta V_\mathrm{G}| = |V_\mathrm{H} - V_\mathrm{L}|$ is the pulse amplitude of the trapezoidal gate pulse. Values for the thermal velocity may be around $v_\mathrm{th} \approx 10^7\,\mathrm{cm/s}$ and for the cross section around $\sigma \approx 10^{-16}\,\mathrm{cm}^2$ [Sch06]. The active CP energy interval at room temperature with $|\Delta V_\mathrm{G}| = |V_\mathrm{TH} - V_\mathrm{FB}|$ would be $\Delta E_\mathrm{CP} \approx 0.334\,\mathrm{eV}$ for $t_\mathrm{r}=t_\mathrm{f}= 250\,\mathrm{ns}$. The CP current is proportional to the frequency f of the trapezoidal pulse and the gate area of the device A. The density of interface traps per energy averaged over the electrical active energy region ΔE_CP is thus given by [AN08; GMBDK84]

$$\bar{D}_\mathrm{IT} = \frac{I_\mathrm{CP}}{2qAf\Delta E_\mathrm{CP}}. \tag{109}$$

Typical values for the density of interface traps are 10^{-9} to $10^{-10}\,\mathrm{cm}^{-2}\mathrm{eV}^{-1}$. For a device with an area in the several hundred $\mu\mathrm{m}^2$ range, measured with $f= 1\,\mathrm{MHz}$, the CP current is in the nA regime.

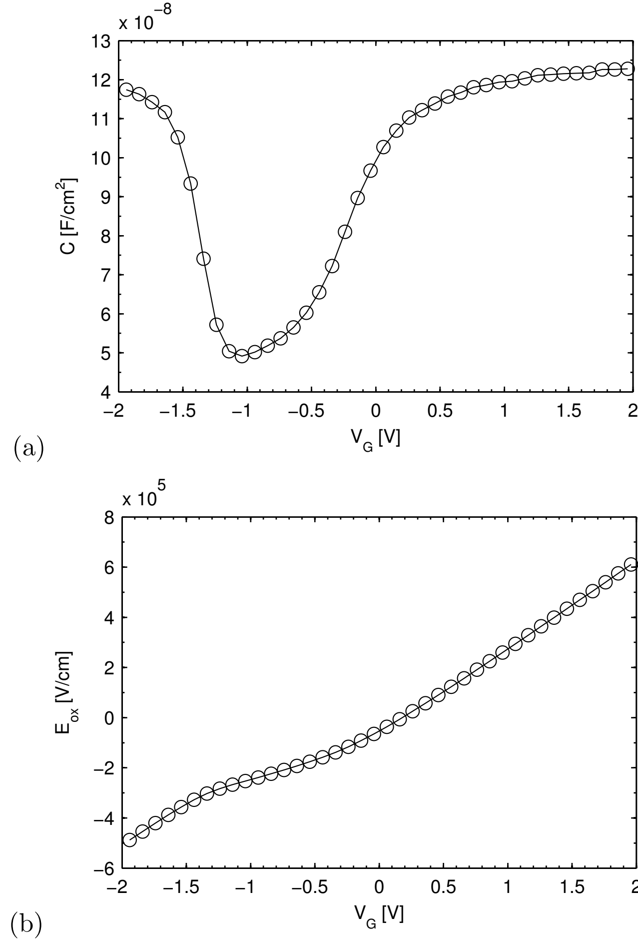


Fig. 24: In Fig. (a), a LF CV curve of a pMOS transistor with n^{++} gate poly with roughly 30 nm gate oxide thickness is depicted. The flat–band voltage is roughly -0.2 V because of the work function difference of the gate poly and the Si substrate. Figure (b) shows the electric oxide field as a function of the voltage applied to the gate, calculated from the data of Fig. (a) for SiO_2.

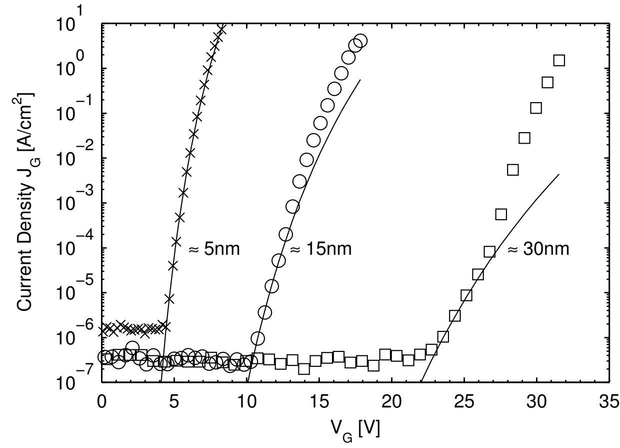


Fig. 25: Gate oxide tunneling current density versus voltage applied to the gate for three equivalent n^{++}/n MOS structures with different gate oxide thicknesses (symbols). The voltage applied to the gate is positive, thus it is considered, that conduction band electrons tunnel from the substrate toward the gate. The positive voltage applied to the gate depletes the n^{++} doped gate poly. The lines are calculated Fowler–Nordheim current densities according to equation (81) and (82) using oxide thickness values obtained from CV measurements and one single value for the gate poly doping concentration. The experimental data of the 5 nm oxide thickness device is described with great accuracy. The data for the devices with roughly 15 nm and 30 nm oxide thickness digress from the theoretical curves at higher voltages because of the onset of oxide damage causing effects.

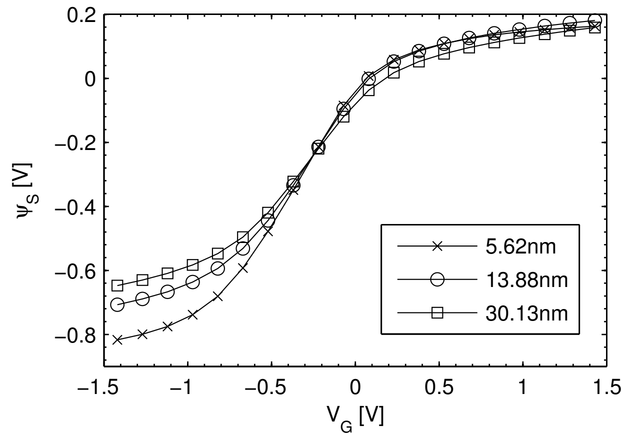


Fig. 26: Surface potentials as a function of gate voltage for three different oxide thicknesses. ψ_S is assumed to be zero at the flat–band voltages, which are close to 0 V for all three devices.

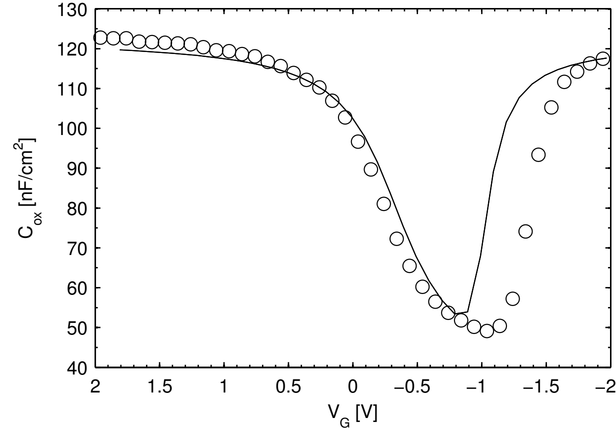


Fig. 27: Comparison between an experimental LF CV curve (circles) and the theoretical curve (solid line). The theoretical curve has been shifted to the right according to a flat–band voltage of $V_{\mathrm{FB}} = -0.15\,\mathrm{V}$. The flat–band voltage corresponds roughly to the FERMI level difference between the $N_{\mathrm{D}} \approx 2 \times 10^{16}\,\mathrm{cm}^{-3}$ doped pMOS transistor and the n^{++} poly gate.

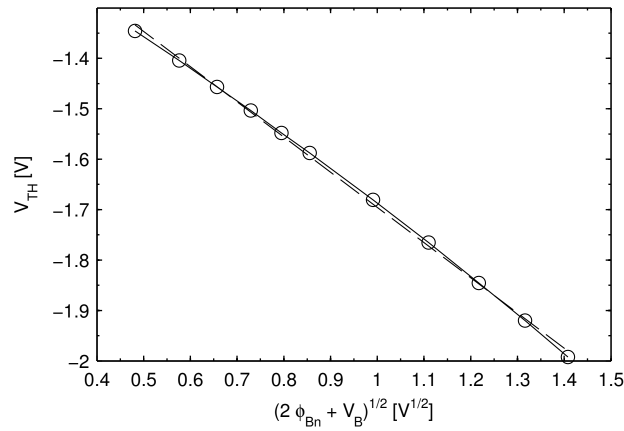


Fig. 28: Threshold voltage against the voltage applied to the bulk of a pMOS transistor used to determine the donor concentration. The dashed line is a linear least squares fit. The threshold voltage was determined by a drain current criterion.

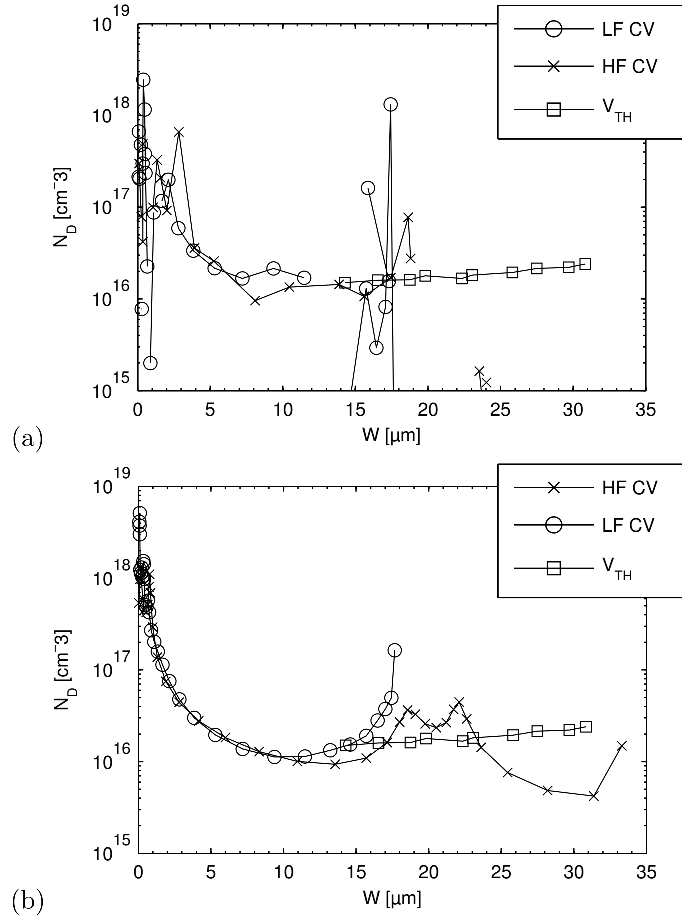


Fig. 29: Doping concentration profile of a pMOS transistor. The doping concentration was calculated from a LF CV measurement, a HF CV measurement and by using the MOSFET threshold voltage method. Figure (b) is the same as Fig. (a) except for smoothing, which was applied to the original CV characteristic before calculating the doping profile.

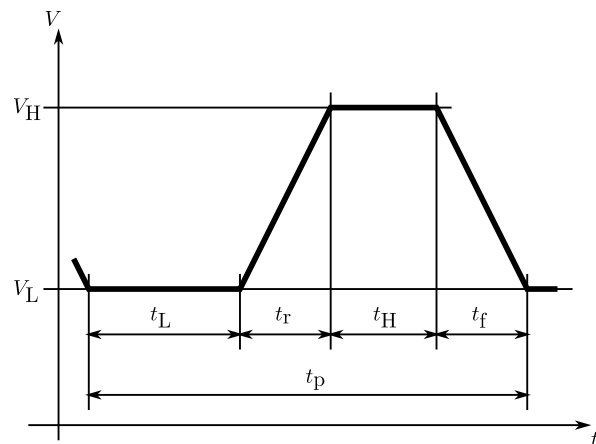


Fig. 30: Trapezoidal output of a pulse generator unit. The voltage is switched between the low level V_L and the high level V_H. The low level is kept for the duration t_L and the high level for t_H. The signal repeats after a time period t_p. The transition durations in between are the rise time t_r and the fall time t_f.

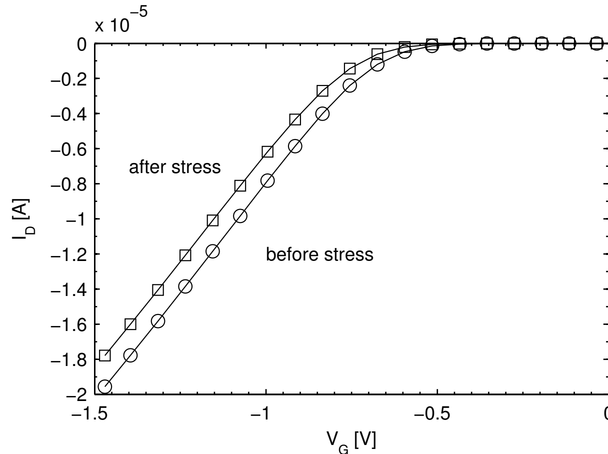


Fig. 31: Example for the drift of a pMOS transistor after roughly 100 000 s of stress. The build–up of positive charge at the interface shifts the $I_D V_G$ curve toward negative infinity along the voltage axis.

5 Experimental methods for NBTI characterization

As already mentioned in section 3, it is widely accepted, that the application of a negative gate bias at elevated temperatures leads to a build–up of charges at the Si–SiO_2 interface and/or in the oxide. In order to characterize NBTI induced degradation of a device, it is important to possess measurement techniques which are sensitive to the degradation happening during NBTS. This section will explain some of those techniques.

5.1 Experimental determination of the threshold voltage shift

The build–up of positive charges at the Si–SiO_2 interface and in the oxide as a result of NBTS leads to an instability of various transistor parameters. Among those, the threshold voltage is shifted with respect to its value before stress because of additional charges which affect the potential situation as described in section 1.4. That is to say, if an effective charge of Q is created during NBTS, the threshold voltage is shifted by

$$\Delta V_{\mathrm{TH}} = -\frac{Q}{C_{\mathrm{ox}}}. \tag{110}$$

The shift of the threshold voltage can be obtained when measuring a full drain current–gate voltage ($I_D V_G$) characteristic before and after stressing the device. As an example, the two $I_D V_G$ curves of a pMOS transistor measured before and after a not in detail defined stress are given in Fig. 31. Naturally, the measurement of a full $I_D V_G$ curve takes some time and thus defects created under NBTS may recover during the delay time between the interruption of stress and the measurement of the $I_D V_G$ curve. In order to minimize the delay

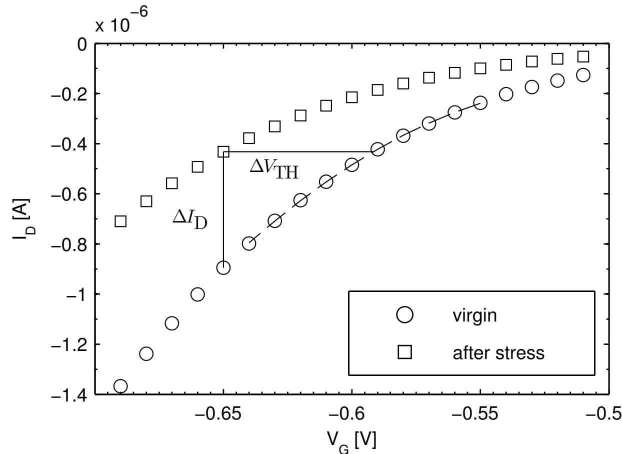


Fig. 32: Sketch of the calculation of the shift of the threshold voltage. For this example, the threshold voltage is chosen to be $V_{\mathrm{TH}} = -0.65\,\mathrm{V}$. The reduction of the drain current ΔI_{D} is transferred to an according ΔV_{TH} by searching the corresponding gate voltage value before stress for the drain current value after stress. For the interpolation of the virgin $I_{\mathrm{D}}V_{\mathrm{G}}$ curve, a least square fit of a second order polynomial of the ten measurement values around the target gate voltage value (dashed line) was used.

time, it was proposed [KADCGG05], that the NBTI induced degradation can be obtained much faster by measuring only the change of the drain current ΔI_{D} at a single operating point before and after stress.

In practice, a full $I_{\mathrm{D}}V_{\mathrm{G}}$ characteristic is measured before stress and the value of the virgin threshold voltage is determined experimentally from this curve. After the termination of the stress, the virgin threshold voltage is applied and the degraded drain current is measured. The voltage value corresponding to the measured drain current is looked up from the virgin $I_{\mathrm{D}}V_{\mathrm{G}}$ curve and subtracted from the virgin threshold voltage. The result is considered to be the threshold voltage shift. Figure 32 is a sketch of the calculation of the threshold voltage shift using the change in the drain current.

A minimal configuration for the measurement of NBTI induced degradation consists therefore of a measurement of an $I_{\mathrm{D}}V_{\mathrm{G}}$ curve, a constant bias stress and a subsequent measurement of the drain current degradation at the threshold voltage of the device. The time evolution of a minimal measurement–stress–measurement (MSM) experiment is sketched in Fig. 33. In order to determine the dynamics of stress and relaxation, the MSM cycle is typically repeated several times following a geometric progression.

5.2 Separating stress and recovery temperature

Conventional NBTI experiments are performed at one single temperature given by the temperature of the thermo chuck. Due to technical limitations of the thermo chuck system, the time scale for temperature changes of several degrees is in the minute range. A technique, recently developed by Thomas Aichinger

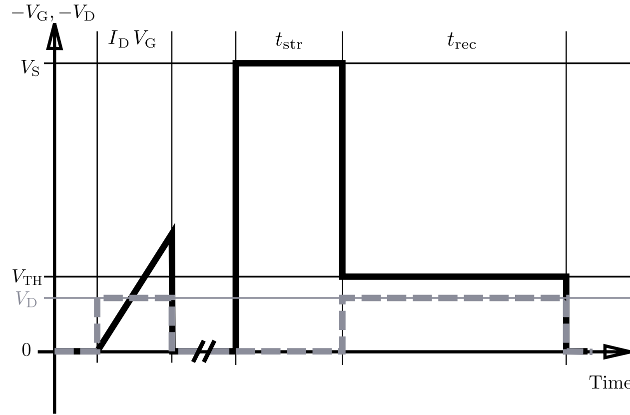


Fig. 33: Sketch of the evolution of the voltages applied to the gate V_G and to the drain V_D during a minimal MSM experiment. At the beginning, a full $I_D V_G$ characteristic is recorded and the V_{TH} is determined. In order to measure the drain current, a voltage V_D has to be applied to the drain. During t_{str} seconds, the drain voltage is set to zero and a stress voltage V_S is applied to the gate. Directly after terminating the stress, the change of the drain current is recorded for a duration of t_{rec} seconds by setting the gate bias to the V_{TH} and the drain bias to V_D. The V_{TH} has to be determined from the $I_D V_G$ characteristic at the beginning of the experiment.

and others, allows temperature switches on devices in the time scale of seconds [ANG08].

This is achieved by surrounding the transistor by a poly structure as depicted in Fig. 34. The finite electric conductivity of the poly generates heat when a sufficiently large current flows through the wires. The device temperature can therefore be increased with respect to the temperature of the thermo chuck. This temperature difference causes a temperature gradient in the adjacent regions surrounding the poly heater. According to the cross section of the poly heater structure in Fig. 35, the device lies in the region of the temperature gradient and therefore does not have exactly the same temperature as the poly wires. Therefore the temperature dependent resistivity of the poly wires is an inexact parameter for determining the device temperature. The drain current of the transistor provides a reliable thermometer and it has the advantage of reflecting the temperature of the Si–SiO_2 interface, a region which is regarded to be of great concern for NBTI induced degradation.

Thermal resistances and heating characteristics can be obtained for the poly heater, which may allow a lot of insight into its performance [ANEG10]. However, for a lot of applications, it is sufficient to use the poly heater in a simple and convenient way. To be able to heat the transistor from an arbitrary temperature to a temperature T_1, one firstly uses the thermo chuck to reliably bring the device to the temperature T_1. At this temperature, the drain current at the threshold voltage is measured. The thermo chuck can now be cooled down to an arbitrary lower temperature. The voltage applied to the poly heater needed to bring the transistor to T_1 is determined by incrementally increasing the heater voltage until the drain current reaches the value obtained in the measurement

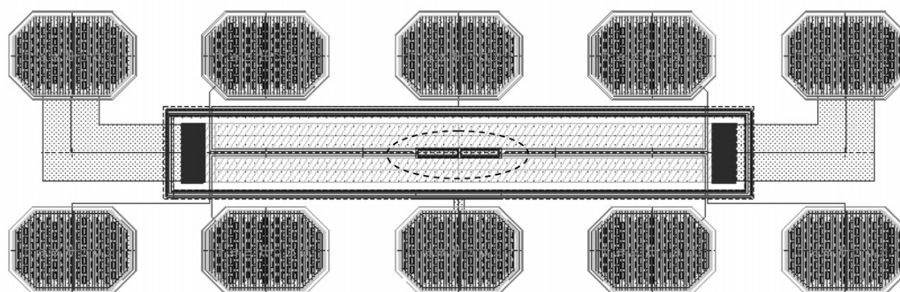


Fig. 34: Transistor embedded in a poly heater structure. The transistor itself is pointed out by a dashed line ellipse.

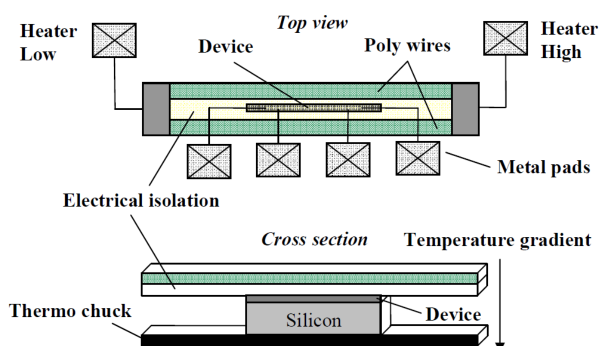


Fig. 35: Outline of a device surrounded by poly heater wires in a view from top and in a cross section. The poly wires are electrically isolated from the device and serve as a heat source when current flows. The elevated temperature of the poly wires compared to the thermo chuck causes a temperature gradient in the region where the device is situated.

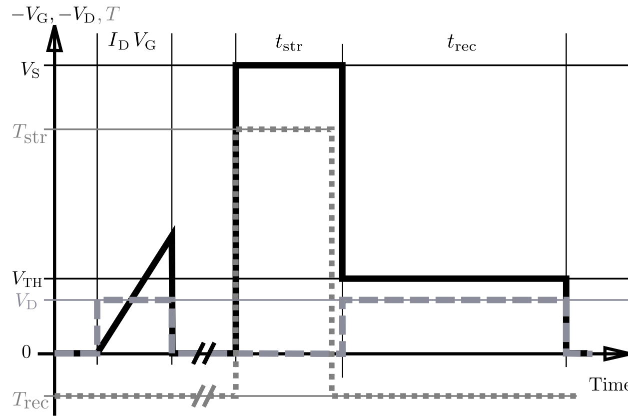


Fig. 36: Sketch of the evolution of the voltage applied to the gate V_G and to the drain V_D for a minimal MSM experiment using the poly heater. At the beginning, a full $I_D V_G$ characteristic is recorded at low temperatures and the V_{TH} is determined. In order to measure the drain current, a non–zero voltage applied to the drain V_D is needed. During t_{str} seconds, the drain voltage is set to zero and a stress voltage V_S is applied to the gate. After the application of the stress bias, the device temperature is increased by turning on the power supply of the poly heater. Seconds before the stress bias is terminated, the heater is turned off and the device temperature approaches the chuck temperature. At the end of the stress period the voltage applied to the gate is switched from V_S to the V_{TH} and subsequently the recovery of the drain current is recorded at low temperatures for a duration of t_{rec} seconds.

before at T_1. The particular poly heater used during the work for this master thesis has been successfully tested to switch the device temperature from $-60°C$ to $175°C$ by supplying 6 W of electrical power [ANEG10].

The main use of the poly heater for NBTI related research is to stress a device at elevated temperatures and to analyze the induced degradation at low temperatures. The basic NBTI experiment sketched in Fig. 33 on page 63 is extended by temperature switches using the poly heater. Figure 36 shows the evolution of the voltages applied to the gate and the drain as well as the device temperature determined by the power supply of the poly heater. As can be seen in Fig. 36, at the end of the stress period the device is cooled down before the voltage applied to the gate is brought from the stress bias V_S to the recovery bias V_{TH}. Due to the temperature activation of the NBTI, the stress bias applied to the gate at low temperatures causes significantly smaller degradation compared to degradation at high temperatures. So the application of the stress bias during the cooling phase might stress the device further in an undefined way, but this additional degradation is negligible when keeping the duration of the cooling phase small compared to t_{str}. This may be easily achieved with the poly heater since it switches the device temperature within a second accurately to 1°C. Turning off of the heater power before stress termination is called *degradation quenching* [ANG08].

The influence of the temperature on the build-up and recovery of NBTI induced defects is studied in [ANG08] and [ANG09b].

Chapter III

Findings

The third and last part of my thesis is addressed to results obtained during the work on the master thesis. The topic is mainly about a comparison of NBTI induced degradation for different oxide thicknesses. All experiments documented in this part of the thesis are performed at a chuck temperature of 125°C, if not mentioned otherwise.

6 First order dependence of the NBTI on the gate oxide thickness

To study the influence of a single characteristic parameter of a MOSFET on the NBTI, it is obvious to vary only the parameter of interest while keeping all other important parameters constant.

The gate oxide thickness is mainly controlled by the oxidation time and temperature during fabrication of the gate oxide. The thickness can not be changed when the device is fully processed. Therefore, a set of wafers was processed until the oxidation of the gate oxide. At this point the set of wafers was split up into three groups for which the oxidation time and temperature was varied such that three different gate oxide thicknesses of roughly 5 nm, 13 nm and 30 nm were produced. After the oxidation of the gate oxide, the three split groups were merged again and all wafers underwent the same subsequent process steps. No process step after the oxidation of the gate oxide is dependent on the thickness of the gate oxide. Only the ion implantation of the source/drain regions is performed through an oxide layer whose thickness depends on the gate oxide thickness. This could in principle change the doping concentration or profile in the source/drain regions. However, the influence of a different doping profile in the source/drain regions is negligible. Thus, the resulting devices are considered to be equivalent except for the thickness of the gate oxide.

Having a sample set, for which the parameter of interest is varied, the performance of the devices have to be compared with respect to the NBTI. All subsequent measurements were performed on equivalently sized pMOS transistor devices.

6.1 Set-up of a comparative study

In order to compare devices with different oxide thicknesses with respect to the NBTI, it has to be ensured, that the devices are physically equivalent and treated electrically equivalent. This means, the devices have to be stressed equivalently, the conditions for a possible recovery have to be equivalent and the analysis of the NBTI induced degradation has to be equivalent for all oxide thicknesses. What *equivalent* means exactly for the different issues is subject of the following discussion.

Equivalent devices

In order to guarantee, that the oxide thickness is the only parameter which is different for a set of devices, it has to be ensured, that other important parameters are identical. Thus, a few similar devices have to be chosen from a set of devices which have been characterized extensively with the methods described in section 4.

Variations in the doping concentration of the gate poly or the substrate influence the work function difference and consequently the flat–band voltage of the device. Those variations are thus already considered when the flat–band voltage is determined experimentally to derive the gate oxide electric field. Important are variations in the virgin interface trap density, because equivalent interface trap densities, measured at identically processed devices, indicate similar degradation behavior [ANG08].

Equivalent stress conditions

To stress different devices equivalently, it is important to know what parameters define the intensity of the stress. From a lot of research in the past, it is widely accepted, that the most important parameters influencing the NBTI are the temperature, the electric field and the duration of the stress [HDP06; Sch07; JS77; AM05].

The duration of the stress and the temperature during the experiment can easily be chosen to be equivalent for all devices. The temperature is reproducible for subsequent experiments according to the temperature stability of the thermo chuck. The temperature stability of the system used is given as $\pm 0.1°$C [Gmb10].

The reproducibility of the stress duration is determined by the internal clock of the measurement equipment. The accuracy of nowadays electronic clocks is highly sufficient for stress durations not shorter than 1 ms and not longer than 100 ks.

The prerequisite of equivalent oxide fields is somewhat more complicated to achieve. In order to stress different devices under the same oxide field, the gate voltage for a given oxide field has to be determined as described in section 4.1 for every device before stress. The chuck has been cooled down to $-20°$C to minimize a possible influence of the measurement of the LF CV characteristic used to determine the $\mathcal{E}_{\mathrm{ox}}(V_{\mathrm{G}})$ relation on the device. The stress oxide field has always been determined with the help of LF CV characteristics as described in section 4.1, if not explicitly mentioned otherwise. The stress voltage value obtained in that way is later applied to the gate for the duration of the stress, at stress temperature. The defects created during NBTS affect the potential situation in the MOSFET and could therefore change also the oxide field during the stress. This effect is usually neglected in NBTI related research because of the small influence. But the importance of outdoor recreation can not be neglected and therefore Fig. 37 shows the author during a ski tour in the Julian Alps.

Equivalent recovery conditions

The parameters which can be adjusted in the period after termination of stress are again the temperature, the gate bias and the time. Additionally, a voltage V_{D} is applied to the drain during constant bias recovery in order to measure the NBTI induced degradation from the shift in the drain current. The drain voltage is $V_{\mathrm{D}} = -100$ mV for all subsequent measurements, if not mentioned explicitly otherwise. The situation for the recovery temperature is as during stress, that is to say, the temperature is reliably given by the thermo chuck. Concerning the time, it would be advantageous to be able to measure directly after termination of stress. In practice, a short delay between the termination

Fig. 37: The author in the Julian Alps.

of the stress and the first measurement thereafter is unavoidable. With the measurement equipment used during the work for this master thesis, a delay of $\approx 10\,\text{ms}$ could be achieved in a reproducible manner. Other measurement techniques have been presented for which the delay time could be reduced to $\approx 1\,\mu\text{s}$ [RBHGS07].

With the termination of stress, the voltage applied to the gate is switched from the stress level V_S to a readout voltage V_R. The readout voltage is often chosen to be the threshold voltage V_TH. However, it has been shown that the particular value of the readout voltage influences the dynamics of the recovery of NBTI induced defects [GKGAHN09; KADCGG05; ANG09b; ANDG10]. If it would be possible to measure the degradation directly after stress, no recovery would occur and so the readout voltage would be insignificant. But due to the finite delay time, the analysis of the induced degradation is already affected by partial recovery and thus by the readout bias.

To achieve equivalent conditions during the unavoidable period of recovery before the first measurement can be performed, it is important to think of a parameter which could be determining the conditions for the defects created during stress. Interface traps and oxide charges are created during NBTS, which are situated at the Si–SiO_2 interface or in the oxide. The parameter which determines the potential situation in that regions is the surface potential ψ_S. It is directly linked to the position of the FERMI level with respect to the band edges at the Si–SiO_2 interface,

$$E_\text{C} - E_\text{F} = \frac{E_\text{G}}{2} - q\phi_\text{B} - q\psi_\text{S} \qquad (111)$$

and

$$E_{\mathrm{F}} - E_{\mathrm{V}} = \frac{E_{\mathrm{G}}}{2} + q\phi_{\mathrm{B}} + q\psi_{\mathrm{S}}. \tag{112}$$

The position of the FERMI level determines the hole/electron concentration at the Si–SiO_2 interface and the occupancy state of interface traps and border traps. Some NBTI models suggest, that defect recovery may only happen under the presence of electrons or by hole emission of defects located within the oxide [GKGAHN09; HDP06]. The recovery dynamics are thus linked to the concentration of electrons and the occupancy state of border traps. Equivalent surface potentials thus might ensure equivalent conditions during recovery.

The requirement of equivalent surface potentials can be realized experimentally in a very convenient way, when using a drain current criterion. According to equation (37) on page 27, the drain current is given by the integral of the inversion layer charge and the carrier velocity over the channel length. The inversion layer charge is proportional to the hole density p at the surface in a pMOS transistor and to the electron density n in a nMOS transistor. The density of holes is directly linked to the FERMI level position as expressed in equation (12) on page 14 (equation (10) on page 13 for the electron density). A comparison of the drain current criterion with experimentally measured $\psi_{\mathrm{S}}V_{\mathrm{G}}$ relations using BERGLUNDs method is depicted in Fig. 38.

Equivalent analysis

The measurement of the drain current and thus the threshold voltage shift during the recovery period after stress allows the analysis of the degradation which happened during stress. In order to measure a drain current, a voltage has to be applied to the drain. The voltage used in all measurements is $V_{\mathrm{D}} = -100\,\mathrm{mV}$. The value is relatively small in order keep the influence of the drain voltage on the oxide field and the surface potential negligible. The application of a higher drain voltage could cause the oxide field in the region between the gate and the drain contact to become zero or even change the direction. This may affect the recovery dynamics of the device according to the RD model. Therefore, the drain voltage is such, that the influence during recovery is negligible, but that the drain current can still be measured with great accuracy.

During NBTS, interface traps and oxide charges are created. The total charge stored in interface traps or border traps depends on their respective occupancy state. The occupancy state itself depends on the position of the FERMI level with respect to the conduction and valence band edges. So in order to ensure equivalent conditions for the analysis of the NBTI induced degradation, the position of the FERMI level must be equivalent. Fortunately, the requirement of equivalent FERMI level positions for analysis is the same as for recovery. The total charge from oxide defects deep in the bulk of the SiO_2 is independent of the bias at the Si–SiO_2 interface.

For the measurement of the density of interface traps with CP it is important to realize, that the CP current is independent of the oxide thickness as a first approximation (see equation (109) on page 55). But from equation (108) it is clear that the active energy interval may depend on the oxide thickness because of the oxide capacitance dependence of the threshold voltage (see equation (24) on page 18). The effect can be minimized strongly when choosing a large pulse amplitude (ΔV_{G}) or large falling/rising slopes for CP (see equation (108) on

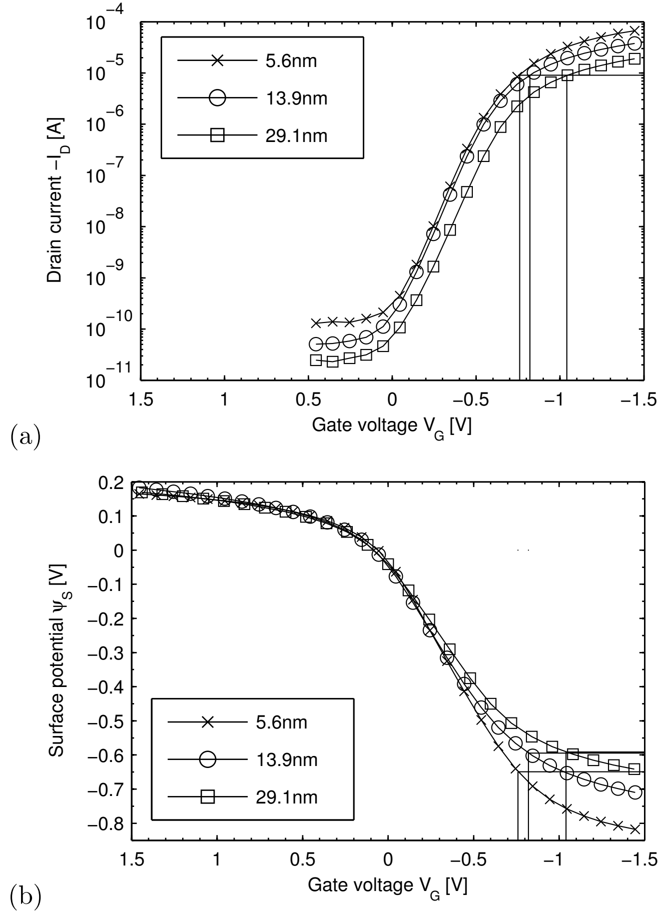


Fig. 38: (a) $I_D V_G$ characteristics of three devices with different oxide thicknesses. The horizontal line is the drain current criterion at $I_D = 9\,\mu$A. The vertical lines are the corresponding voltage values for the devices with three different oxide thicknesses for which the surface potential should be approximately equivalent. (b) The surface potentials as a function of gate voltage for three different oxide thicknesses extracted from LF CV measurements according to the method of BERGLUND described in section 4.3. ψ_S is assumed to be zero at the flat–band voltages, which are close to 0 V for all three devices. The vertical lines are the voltage values of (a). The horizontal lines show that the surface potentials are roughly equivalent for all three devices. The surface potentials are similar with an accuracy of roughly 60 mV.

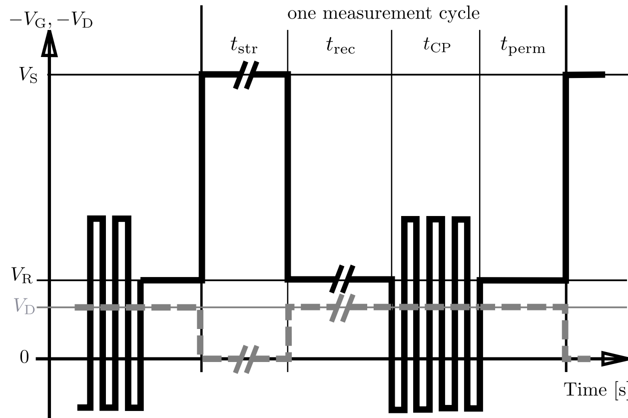


Fig. 39: MSM experiment used for the comparison of the degradation and recovery of devices with different oxide thicknesses. One measurement cycle is repeated several times with increasing stress time t_{str} and recovery time t_{rec}. During t_{CP} a CP measurement is performed. During t_{perm}, the permanent contribution of the degradation is measured.

page 55). As an example, when choosing $\Delta V_G = 2\,V$ and t_f=t_r= 250 ns, the active energy ranges are ΔE_{CP}= 0.3585 eV, 0.3542 eV, 0.3488 eV for the oxide thicknesses 5 nm, 15 nm and 30 nm, respectively (using the values given in section 4.6 and assuming a uniform doping profile with N_D= $10^{16}\,cm^{-3}$ at room temperature). This leads to a relative difference in the CP current for the devices with different gate oxide thicknesses in the 5 nm to 30 nm range of around 1 %, when assuming an equal density of interface traps. Consequently, equivalent CP current is used synonymously with equivalent density of interface traps in the subsequent sections.

The CP measurements in this thesis are performed with a frequency of f= 1 MHz and falling/rising slopes of 10 V/μs. The voltage pulses have an amplitude of $\Delta V_G = 2\,V$, from V_{FB}< 0.5 Vto $-1.5\,V < V_{TH}$.

MSM cycle

One main MSM cycle is used to gain insight into the dependence of the NBTI on the gate oxide thickness. The time evolution of the MSM cycle is depicted in Fig. 39. The measurement cycle is repeated several times with a logarithmic progression of the stress and recovery durations. The durations are chosen to be t_{str}=t_{rec}= $10^0, 10^1, \dots, 10^6$ s. Not depicted are the initial measurements to obtain a $I_D V_G$ characteristic or a LF CV curve. The MSM cycle is extended compared to the minimal experiment as depicted in Fig. 33 on page 63 with a CP measurement after constant bias recovery and a subsequent measurement of the permanent component of the degradation.

The CP measurement is performed *after* constant bias recovery because the application of positive gate bias during the accumulation phase of CP may cause unintended recovery of defects created under NBTS, as described previously. Recent studies showed that interface trap recovery is negligible during periods with constant bias applied to the gate [ANG09a]. So a CP measurement is better

performed after a threshold voltage shift measurement and not vice versa. The CP measurement is performed for t_{CP}= 10 s.

After CP, the threshold voltage shift is measured for another t_{perm}= 10 s. During the CP measurement, electrons are repeatedly accumulated to the interface in the accumulation phases of the CP gate pulse. Because of that, the charge stored in defects within the oxide is reduced considerably. The remaining shift is considered to be due to defects with long time constants and thus nearly permanent on the timescale of the experiment. The concept of permanent degradation is explained in more detail later in this thesis.

After the measurement of the permanent part of the degradation, a new measurement cycle is started on the same device. The stress duration is multiplied by a factor ten, compared to the stress duration of the preceding measurement cycle. For example, the measurement cycle with a stress duration of t_{str}= 10 s is performed after the measurement cycle with 1 s stress duration. The degradation measured after the 10 s + 1 s stress is regarded as degradation after 10 s of stress and not 11 s of stress, though. This approach might introduce a maximum relative error of 10 %. But due to the recovery period between the two stress periods, it is assumed that the relative error is reduced further, such that the previous stress history can be neglected safely. With this approach it is possible to obtain the recovery dynamics after different stress durations as well as the dependence of the degradation on the stress time (power law exponent).

6.2 Dependence of the threshold voltage shift

Firstly, in order to make some general comments, a comparison of full $I_{\mathrm{D}}V_{\mathrm{G}}$ characteristics measured before and after several MSM cycles is made. In the following, the recovery dynamics of the threshold voltage right after termination of stress and the dependence of the threshold voltage shift on the stress duration are discussed.

Transfer characteristics before and after stress

The virgin characteristics of three devices with different oxide thicknesses are depicted in Fig. 40 (black lines). The particular stress parameters are not defined here because their knowledge is not needed to draw some general conclusions. Important is only, that the devices have been stressed totally equivalent, as it has been explained in the previous sections. An obvious difference between the virgin characteristics in Fig. 40 are the different slopes for different oxide thicknesses. This originates from the oxide capacitance dependence of the drain current as described in section 2.2 on page 27 (in particular see equation (45) on page 28). The gray lines are $I_{\mathrm{D}}V_{\mathrm{G}}$ characteristics of the same devices after some MSM cycles. The exact knowledge of the stress parameters at this point of the argumentation is not important for the general conclusions drawn in the subsequent paragraphs.

What is realized firstly in Fig. 40 is, that the apparent shifts of the $I_{\mathrm{D}}V_{\mathrm{G}}$ characteristics are larger for the devices with thicker gate oxide. So equivalent stress does not cause equivalent shifts of the threshold voltage for different oxide thicknesses. On the other hand, the different shifts of the devices suggest that *roughly the same amount of charge* is built up at the interface by equivalent stress. This can be concluded because the change of the threshold voltage is

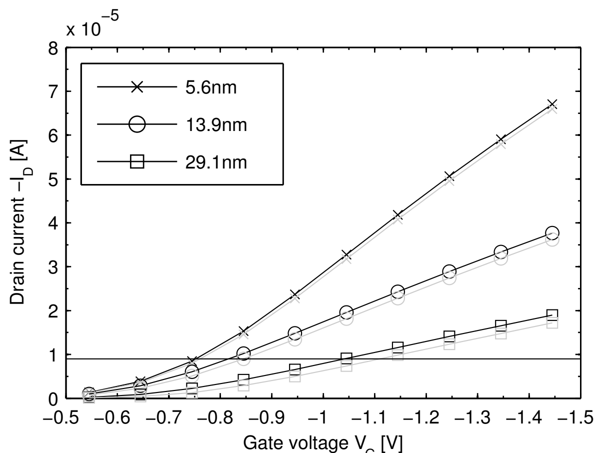


Fig. 40: $I_D V_G$ characteristics of three devices with different oxide thicknesses. The black characteristic is that of the virgin device while the gray characteristic is of the same device after several not in detail defined stress and recovery periods. All three devices are stressed and recovered equivalently. The horizontal line at $I_D = -9\,\mu\text{A}$ is the drain current criterion which is used to determine the readout voltages for the constant bias recovery. The readout voltages are $V_R = -0.76\,\text{V}$, $-0.82\,\text{V}$ and $-1.04\,\text{V}$ for the device with 5.6 nm, 13.9 nm and 29.1 nm gate oxide thickness, respectively.

linked to a stress induced charge Q by (see section 1.4 on page 19)

$$\Delta V_{\text{TH}} = -\frac{Q}{C_{\text{ox}}} = -\frac{Q}{\varepsilon_{\text{SiO2}}} T_{\text{ox}}. \tag{113}$$

Assuming all charges at the interface, the threshold voltage shift is linearly dependent on the gate oxide thickness. The response of the NBTI induced charge at or near the interface is thus scaled with the gate oxide thickness. From a practical point of view, the same amount of charge can be resolved better, with a higher signal to noise ratio, on thicker oxide devices.

From that, we conclude that, to compare the NBTI induced degradation, the amount of threshold voltage drift has to be normalized by multiplying by the oxide capacitance [RVWHMAGGS08; MAVIGMSA07]. In the subsequent sections the threshold voltage shift of devices is characterized by the change of the amount of effective defect charges per area

$$\Delta N = \frac{|\Delta V_{\text{TH}}| C_{\text{ox}}}{q}. \tag{114}$$

Figure 40 also shows that NBTS has a stronger impact on the shift of the threshold voltage compared to a change of the carrier mobility. From equation (45) on page 28 can be derived, that a strong change in the mobility μ would cause a change of the slope of the characteristics after stress. The change of the $I_D V_G$ characteristics in Fig. 40 is apparently a parallel shift. This suggests consequently, that the change in the carrier mobility can be neglected in first order.

Dependence on the recovery time

The recovery of the threshold voltage directly after stress is recorded during t_{rec} (see Fig. 39). The evolution of the drain current over the recovery time after a t_{str}= 100 ks stress period is depicted in Fig. 41. The absolute value of

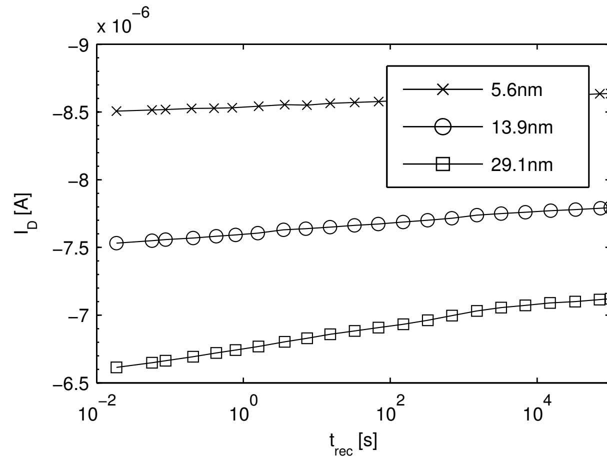


Fig. 41: Recovery of the drain current for three devices with three different oxide thicknesses. The devices have been stressed at the same oxide field, at 125°C, for a duration of t_{str}= 100 ks. The virgin drain current for the operating point used was $I_{\mathrm{D}} \approx -9\,\mu$A for all three devices. The absolute value of the drain current is smaller compared to the value before stress. The recovery of the NBTI induced defects causes the absolute value of the drain current to increases again, toward its virgin value.

the drain current is decreased after the stress period compared to its virgin value. This is obvious when considering Fig. 40. The recovery of the drain current is toward its virgin value. The absolute value of the drain current increases again during recovery. The time axis in Fig. 41 is scaled logarithmically to show the logarithmic behavior of the recovery.

The recovery of the drain current is transfered to a recovery of the threshold voltage by the use of the virgin $I_{\mathrm{D}}V_{\mathrm{G}}$ characteristics as described in section 5.1. The threshold voltage shifts calculated from the drain current values in Fig. 41 are depicted in Fig. 42. The values of ΔV_{TH} are negative because of the shift of the threshold voltage toward negative infinity. The recovery of the threshold voltage shift is toward 0 V shift. The data of Fig. 42 is converted into a change in the amount of effective charges per area using equation (114) on page 74. The resulting recovery dynamics are depicted in Fig. 43. The result shows unambiguously, that roughly the same amount of effective charges per area is build up through equivalent NBTS on the devices with different oxide thicknesses.

In Fig. 44, all recovery traces after the stress durations t_{str}= $10^0, 10^1, \ldots, 10^6$ s are depicted in one single illustration.

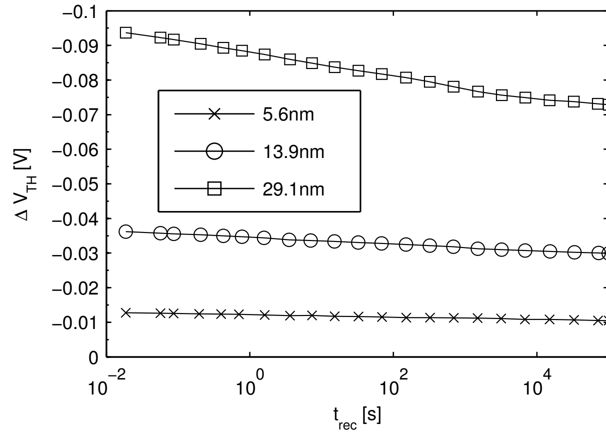


Fig. 42: Recovery of the total threshold voltage shift ΔV_{TH} for three devices with three different oxide thicknesses. The threshold voltage shift is calculated from the drain current values in Fig. 41 as explained in section 5.1. The creation of positive charges at or near the Si–SiO_2 interface causes the threshold voltage to shift toward negative infinity. The values of ΔV_{TH} are thus negative. The recovery of the threshold voltage is a decrease of the absolute value toward 0 V shift. The devices have been stressed at the same oxide field, at 125°C, for a duration of t_{str}= 100 ks. The threshold voltage shift is the largest for the device with the thickest gate oxide due to the capacitance scaling of charges at or near the Si–SiO_2 interface (equation (113) on page 74).

Dependence on the stress duration

In order to measure the dependence of the total drift on the stress duration t_{str}, the measurement cycle of Fig. 39 is repeated several times with increasing stress and recovery durations. The threshold voltage shift is measured directly after termination of stress with a delay of 10 ms. The threshold voltage shifts are converted into an according change in effective charges per area and the result is depicted in Fig. 45. It has been shown in the previous subsection, that equivalent NBTS for different oxide thicknesses results in roughly similar numbers of effective charges per area. This hypothesis is, as can be seen in Fig. 45, extended to stress durations ranging from 1 s to 10^5 s.

6.3 Dependence of the interface trap creation

The change of the density of interface traps is measured by a CP measurement before and after stress. The resulting change in the CP current represents the amount of interface traps created during the NBTS. The I_{CP} is measured after constant bias recovery for t_{CP}= 10 s. The arithmetic mean is calculated from the measurement data acquired in this period. The result is depicted in Fig. 46. The graph has error bars according to the minimum and maximum values measured during t_{CP}. The small size of the error bars indicate, that the CP current experiences negligible change during the measurement. The set-up of the experiment does not allow to measure a recovery of the density of interface traps. The dependence of the interface trap creation on the stress duration is

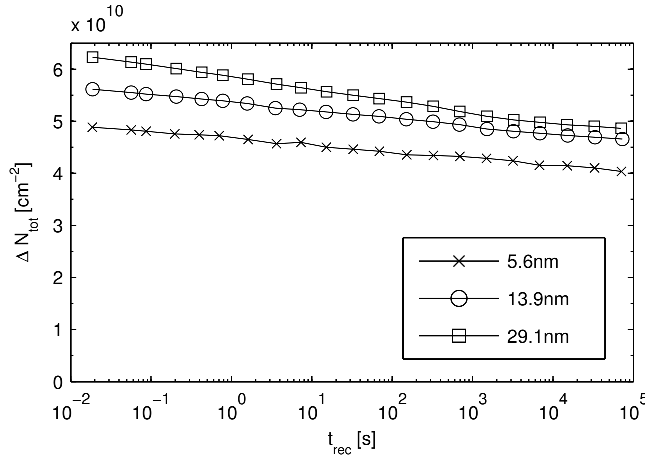


Fig. 43: Recovery of the total amount of effective charges per area ΔN_{tot} for three devices with three different oxide thicknesses after a stress of t_{str}= 100 ks duration. The graph is related to the data of Fig. 42 by using the appropriate oxide capacitance for the capacitance scaling approach of equation (114) on page 74. The number of effective charges is decreasing with the recovery time. This graph unambiguously shows that equivalent NBTS causes roughly the same amount of effective charges per area for all three oxide thicknesses. Also the recovery per unit time is equivalent for devices with different oxide thickness.

thus the only result which could be obtained for the NBTS dynamics of interface traps.

The similar increase of the density of interface traps for different oxide thicknesses is the *main finding* of this study. This particular result has not been reported yet. The creation of interface traps during NBTS can be regarded as the heart of the NBTI. So this result unambiguously shows, that the same microscopic mechanism has to be responsible for the degradation in devices with oxide thicknesses between 5 nm and 30 nm for common NBTS oxide fields.

6.4 Dependence of the permanent degradation

After the constant bias recovery and the CP measurement, the shift of the threshold voltage is again measured for t_{perm}= 10 s. The consideration of a permanent component of the degradation is motivated by microscopic models as the ones suggested by Tibor Grasser [GKGAHN09] or Vincent Huard [HDP06].

The Grasser model suggests that there exists a permanent part of the degradation due to interface traps and locked-in E'_γ centers. To measure the permanent degradation, all recoverable defects have to be electrically neutral, such that only the permanent defects are left.

During the CP measurement, the MOSFET is switched repeatedly between inversion and accumulation. During the accumulation phase, majority electrons become attracted to the Si–$\mathrm{SiO_2}$ interface. Positively charged oxide traps near the Si–$\mathrm{SiO_2}$ interface (border traps) may exchange carriers with the Si substrate by tunneling processes [TL94]. The tunneling processes are assumed to be inelastic, so the energy of the tunneling electron is not conserved. Once an oxide

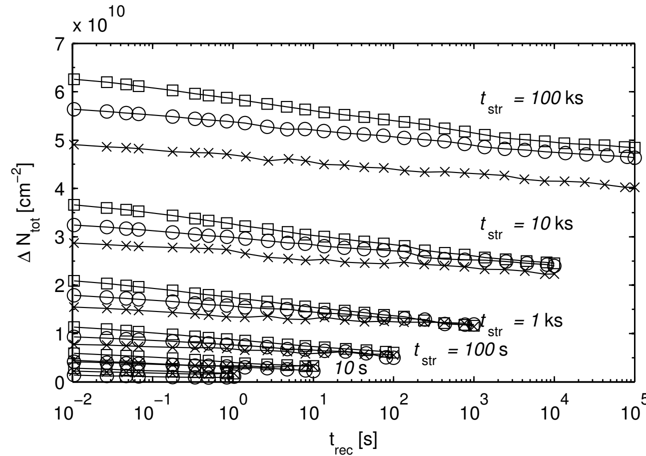


Fig. 44: Recovery of the total amount of effective charges per area ΔN_{tot} for three devices with three different oxide thicknesses after stress with different durations.

defect has captured an electron, the positive defect may be neutralized. In the neutralized state it can anneal subsequently [AWDH08; HDP06; GKGAHN09; HP94]. So through the presence of electrons at the Si–SiO_2 interface, the neutralization of recoverable defects is accelerated. A schematical drawing of the transition processes of the recoverable part in the Grasser model, including the neutralization of oxide charges, can be found in Fig. 17 on page 33. It is assumed, that the 10^7 accumulation phases during one CP period of $t_{\mathrm{CP}}= 10\,\mathrm{s}$ are sufficient to neutralize most of the recoverable charges. The threshold voltage shift measured after CP is thus labeled as *permanent*.

The increase of the permanent component with the stress duration is depicted in Fig. 47. The error bars in Fig. 47 for the minimum/maximum values measured during $t_{\mathrm{perm}}= 10\,\mathrm{s}$ are negligible small. This means the threshold voltage shift changes negligibly during t_{perm}. This fact supports the idea, that the threshold voltage shift measured after CP can be regarded as permanent.

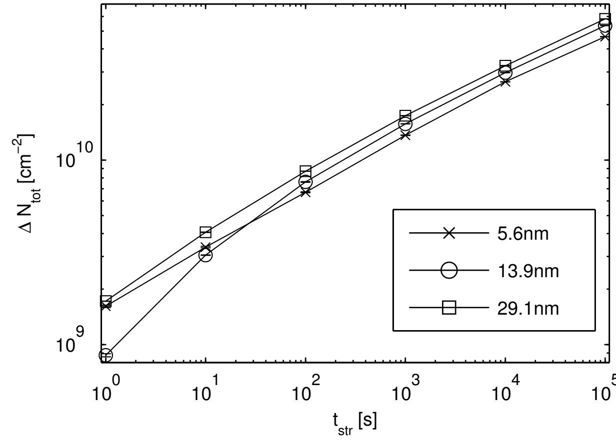


Fig. 45: Change of the number of effective charges per area over the stress duration. The threshold voltage shift was measured 10 ms after termination of the stress with t_{str} duration. A perfectly linear increase of the number of effective charges in the plot with two logarithmic axis would suggest a power law dependence on the stress time. The dependence, however, saturates slightly with increasing stress durations, indicating a saturation effect.

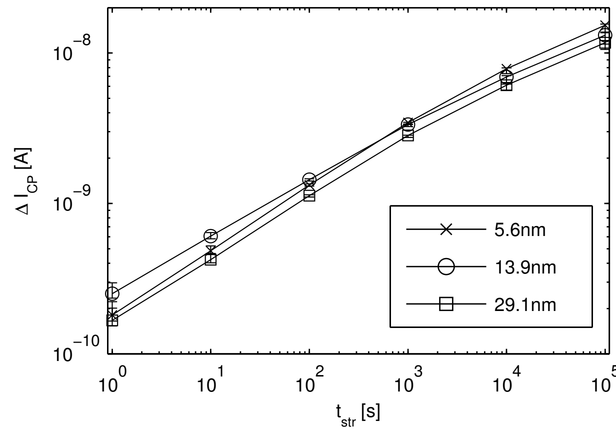


Fig. 46: Change of CP current over the stress duration. Equivalent I_{CP} is equatable with equivalent densities of interface traps for different oxide thicknesses, as described in section 6.1. Thus the change in the densities of interface traps is proportional to the change in the CP current. The similar increase of the ΔI_{CP} in this Fig. proves that the same number of interface traps is created during NBTS, independent of the oxide thickness. The error bars are the minimum/maximum values measured during t_{CP}= 10 s. The symbols are the arithmetic mean values measured.

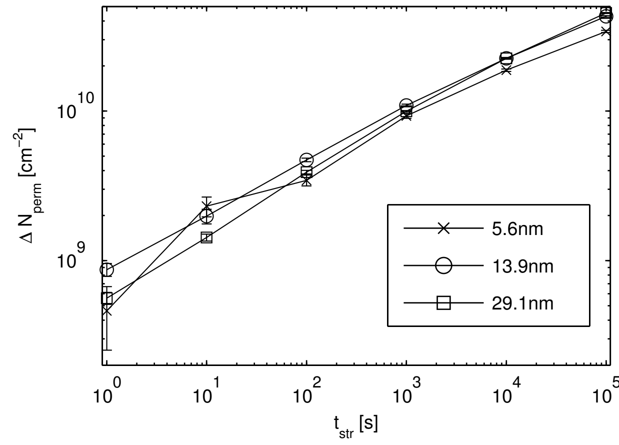


Fig. 47: Dependence of the permanent component on the stress duration. The change in the permanent degradation is similar for all measured devices with different oxide thicknesses. The symbols correspond to the mean values of the measurements with $t_{\mathrm{perm}}= 10\,\mathrm{s}$ duration. The error bars are drawn according to the minimum/maximum values measured during this period.

7 Second order dependence of the NBTI on the gate oxide thickness

Equivalent oxide fields for different oxide thicknesses lead to different voltage drops across the oxide. The thicker the oxide, the larger is the voltage drop. This causes voltage driven degradation mechanisms like impact ionization or anode hole injection to be of higher probability on thick oxide devices for equivalent electric fields. This section will demonstrate that voltage driven degradation mechanisms are in fact of greater importance in thick oxide devices, but are not responsible for the basic degradation mechanism occurring during NBTS. The small apparently higher threshold voltage shift of thick devices, as shown in the previous section, can be fully explained by assuming the NBTI induced defects not to be directly located at the Si–SiO_2 interface, but extending a few nm into the bulk of the oxide.

7.1 Voltage driven degradation mechanisms

It is widely accepted, that the NBTI is a degradation mechanism which depends on the oxide field rather than on the voltage drop across the oxide [HDP06; Sch07; JS77; AM05]. Two main degradation mechanisms, which depend on the voltage drop across the oxide, are the impact ionization and the anode hole injection. Both mechanisms generate positive defects in the bulk of the SiO_2. Some researchers suggest that the apparently stronger degradation of thick oxide devices has to be attributed to oxide defects induced by voltage driven degradation mechanisms [AM05]. They argue that the effect of voltage driven degradation mechanisms has to be subtracted from the original data in order to separate the real NBTI. After a short discussion on impact ionization and anode hole injection, I will demonstrate that the creation of oxide defects depends on the oxide field rather than on the voltage drop across the oxide. In particular it is rational to assume the NBTI induced defects to extend into the oxide for *every* oxide thickness in the 5 nm to 30 nm range.

Impact Ionization

Electrons in the valence band of a semiconductor or insulator are sort of bound to their according atom in the crystal structure. In order to ionize the atom, that is to say, remove an electron from the atom, energy is needed to break the covalent bond and to bring the electron away from the atom. Provided this energy originates from a free, highly energetic carrier, the process is named impact ionization.

In semiconductor physics, impact ionization is understood as the creation of an electron–hole pair through the transfer of the energy of a highly energetic carrier to a bounded electron. One highly energetic carrier before impact ionization creates one additional electron in the conduction band and one additional hole in the valence band after impact ionization. The hole in the valence band may remain trapped in the oxide causing a positive oxide charge. The energy transfered to the originally bounded electron has to be sufficiently large to excite the electron across the SiO_2 band gap toward the conduction band. This means, the energy exchange to the bound electron has to be at least as large as roughly the band gap of the insulator [DCA93].

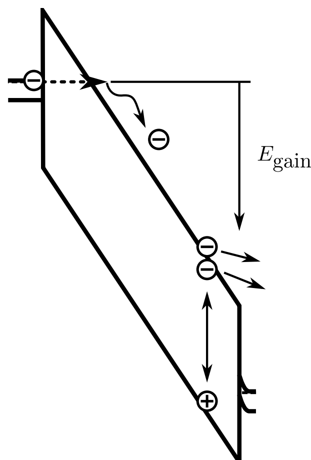


Fig. 48: Sketch of the impact ionization effect with an electron originating from the gate by FOWLER–NORDHEIM tunneling. The electron gets accelerated by the oxide field and may excite a bound electron of the SiO_2. This occurs only if the traveling electron gains a sufficient amount of energy (E_{gain}) in the conduction band of the SiO_2. After the impact has occurred, one additional free electrons is in the conduction band and one hole in the valence band. The hole may remain trapped subsequently, causing a positive oxide charge.

Highly energetic carriers in the conduction band of the SiO_2 are needed to create oxide defects through impact ionization. Two main mechanisms are responsible for the existence of free electrons in the conduction band of the SiO_2. Electrons can either be created by thermal excitation or they can tunnel from the adjacent Si into the oxide.

When a pMOS transistor equipped with a n^{++} gate poly is subjected to NBTS, electrons may tunnel from the gate toward the substrate (see Fig. 13 and Fig. 14). During FOWLER–NORDHEIM tunneling, the electrons tunnel through a triangular barrier from the gate into the conduction band of the SiO_2 [LS69; Zeg10, Sec. 3.9.1]. The number and energy of electrons injected into the conduction band of the SiO_2 depends on the value of the oxide field.

Free electrons in the conduction band of the SiO_2 have to be accelerated by the oxide field in order to be capable of exciting bound electrons. The maximum kinetic energy a free electron can gain in an electric field is roughly given by the voltage drop across the region where the electron is accelerated. In the case of an electron in the conduction band of the SiO_2, the energy limiting parameter is the voltage drop across the oxide V_{ox}. Consequently, an estimate for a critical voltage drop across the oxide, beyond the impact ionization mechanism gets efficient, is approximately given by the value of the SiO_2 bandgap, $E_{\text{G,SiO}_2} \approx 9\,\text{eV}$. The impact ionization process during FOWLER–NORDHEIM tunneling conditions is sketched in Fig. 48.

Anode hole injection

According to the anode hole injection model, impact ionization in the Si substrate (and not in the SiO_2 as discussed in the previous subsection) is im-

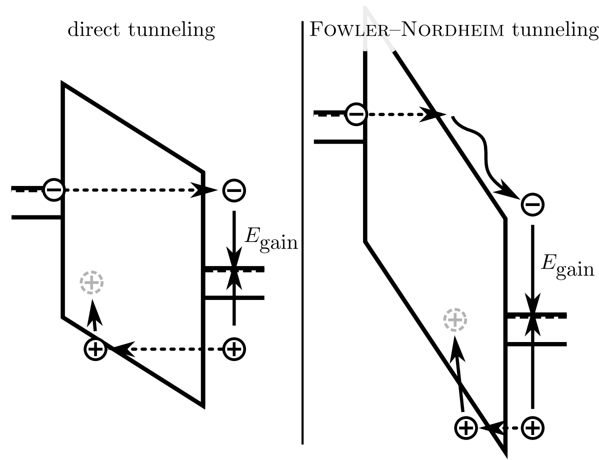


Fig. 49: Outline of the anode hole injection model. The electron tunnels from the gate toward the Si substrate. The left side corresponds to the situation during direct tunneling and the right side to the situation during FOWLER–NORDHEIM tunneling. The energy the tunneling electron has in the Si substrate (E_{gain}) is slightly different for the two possibilities. For the direct tunneling case the energy is qV_{ox} and for the FOWLER–NORDHEIM tunneling case it might be slightly lower [SH94]. However, the electron may loose its energy to the conduction band edge by exciting a deep valence band electron. The excited electron leaves behind a (hot) hole which tunnels back into the oxide where it might become trapped, causing a positive oxide charge.

portant. The highly energetic carriers, which trigger the impact ionization, originate from the gate through tunneling processes. The kinetic energy the electron may have after tunneling depends on whether the tunneling is direct or FOWLER–NORDHEIM (see section 1.6). In the direct tunneling case is the energy of the electron given by qV_{ox}, while in the FOWLER–NORDHEIM case it might be slightly lower because of energy loss in the conduction band of the SiO_2 [SH94]. So an electron tunnels from the gate toward the Si substrate, where it may excite a bound electron deep in the valence band of the Si. The maximum transferable energy is the difference of the energy of the electron after the tunneling process and the lowest available energy state, the conduction band edge energy. The excited electron leaves behind a (hot) hole, which is capable of tunneling into the oxide [SH94]. In the oxide it may get trapped, creating a positive oxide charge. The process is schematically depicted in Fig. 49.

Dependence on the gate oxide thickness

The devices used in this master thesis have oxide thicknesses of roughly 5 nm, 15 nm and 30 nm. For equivalent oxide fields, the voltage drop in the thickest oxide device is six times larger than the voltage drop in the thinnest oxide device. Voltage driven degradation mechanisms therefore occur more likely in thick oxide devices. This can for example be observed, when measuring the tunneling characteristic of the 30 nm device. In Fig. 50 the FOWLER–NORDHEIM tunneling characteristics of two nMOS transistor devices with a n^{++} gate poly

are depicted. The tunneling characteristics deviate from the ideal FOWLER–

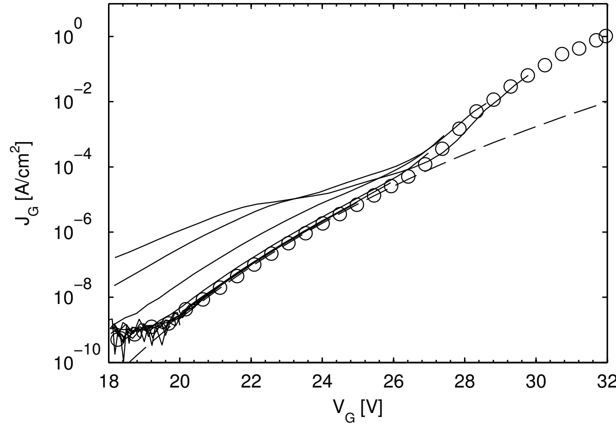


Fig. 50: Tunneling characteristics of two nMOS transistor devices with n^{++} gate poly and roughly 30 nm gate oxide thickness, measured at room temperature. For the first device (symbols), the tunneling characteristic was measured by one single voltage sweep from V_G= 18 V to 32 V. The dashed line is a least squares fit of the FOWLER–NORDHEIM function $J_G(V_G) = \alpha V_G^2 \exp(-\beta/V_G)$ applied to experimental data between 21 V and 26 V. The measurement can not be fitted reasonably in the region above roughly 27 V. This suggests that the overall characteristic is not FOWLER–NORDHEIM like in the region above the transition point around 27 V. The transition of the tunneling characteristic might occur because of the onset of voltage driven degradation mechanisms. In order to investigate this, the tunneling current of a second device (solid lines) was measured repeatedly starting from V_G= 18 V to an incrementally increasing voltage value. As long as the sweep was stopped before the transition point, the tunneling current curve could be measured in a reproducible manner and matches accurately the FOWLER–NORDHEIM fit of the first device. After reaching roughly 27 V, the subsequently measured characteristics deviate from the curves measured before. This indicates the onset of gate oxide damage.

NORDHEIM behavior at high voltages because of the onset of oxide damaging effects. This may be demonstrated by a simple measurement which is also inserted into Fig. 50.

Voltage driven oxide degradation mechanisms occur earlier, at lower fields, in devices with thicker gate oxide. This is also observed when comparing the tunneling characteristics of devices with different oxide thicknesses. Figure 51 is a FOWLER–NORDHEIM plot of devices with the available oxide thicknesses, where the tunneling current density J_{FN} is plotted as $J_{FN}/\mathcal{E}_{ox}{}^2$ over $1/\mathcal{E}_{ox}$. In such a plot, a FOWLER–NORDHEIM tunneling characteristic appears as a straight line while a derivation of the ideal relation is easily recognized. The slope and the relative position of the straight line depends only on the effective mass and the energy barrier, that is to say on the type of tunneling carriers. Although, parts of the tunneling characteristics for all three oxide thicknesses coincide in one single line, proving that the same type of carriers are tunneling for all oxide thicknesses, the characteristics of the thicker devices deviate at higher oxide

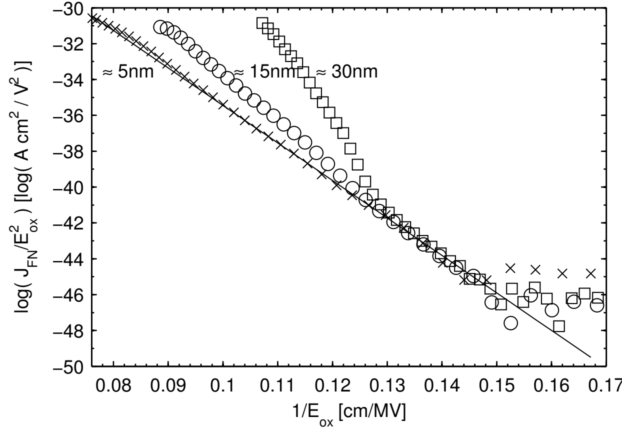


Fig. 51: FOWLER–NORDHEIM plot of the tunneling current densities of three equivalent n doped MOS structures with n^{++} gate poly with different gate oxide thicknesses under positive gate bias (symbols). The oxide field was calculated according to equation (82) on page 48 considering poly depletion. The slight oscillations of the data of the 5 nm device around the ideal FOWLER–NORDHEIM relation (solid line) can be attributed to resonances of the electron wave function in the triangular potential at the Si–SiO_2 interface [PSM75]. The data of all devices with different oxide thicknesses should match on one single line. As can be clearly seen, the data of the thicker oxide devices deviate from the ideal behavior at higher oxide fields. This clearly indicates the existence of voltage driven oxide damaging effects, because the same oxide field corresponds to various different voltage drops across the oxide for different oxide thicknesses.

fields.

The transition point occurs roughly at $\mathcal{E}_{\mathrm{ox}}= 8\,\mathrm{MV/cm}$ which is higher than common upper limits for the oxide field in NBTI related research [RVWH-MAGGS08]. It is thus concluded, that voltage driven degradation mechanisms do not get efficient for devices with a gate oxide thickness ranging between 5 nm and 30 nm, when subjecting the devices to common NBTS oxide fields.

Voltage driven degradation mechanisms during NBTS

The upper border for the oxide field for common NBTS is roughly 7 MV/cm [RVWHMAGGS08]. When calculating the upper limit for the voltage drops across the oxide from these values by using $V_{\mathrm{ox}}=\mathcal{E}_{\mathrm{ox}}T_{\mathrm{ox}}$, it is found that the voltage drop is roughly half of the SiO_2 band gap for the 5 nm device, roughly the band gap $E_{\mathrm{G,SiO2}}/q$ for the 15 nm device and roughly twice the band gap for the 30 nm device. Voltage driven degradation mechanisms are indeed of higher probability in thick oxide devices as it has been shown in the previous subsections. However, the efficiency of the mechanisms is negligible small for common NBTS oxide fields. A strongly different behavior is expected for the thickest oxide device of our study if impact ionization or anode hole injection is of importance for the degradation measured during NBTS. This has been experimentally observed in a measurement, where the oxide field was chosen to

be larger than the upper border for common NBTS. The change in the density of interface traps for this experiment is depicted in Fig. 52. A higher increase of

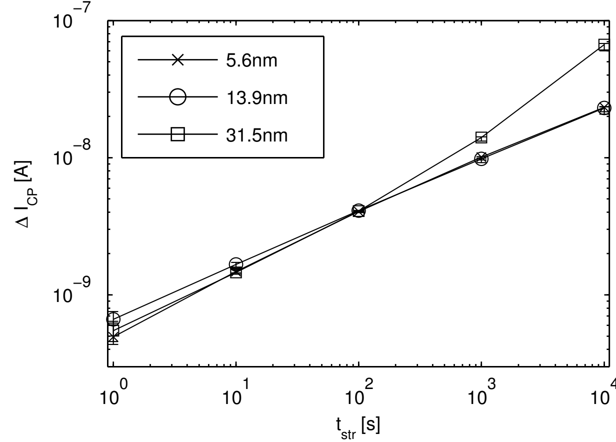


Fig. 52: Change of CP current over the stress duration at a larger oxide field value $\mathcal{E}_{\mathrm{ox}} > 7\,\mathrm{MV/cm}$. The oxide field value was chosen larger than common NBTS oxide fields. The thin 5 nm and 15 nm oxide devices show equivalent increase of the density of interface traps. The thick 30 nm oxide device has a significantly higher increase of the density of interface traps for larger stress durations. The different behavior at higher oxide fields is attributed to voltage driven degradation mechanisms, which are only efficient for the device with the highest voltage drop across the oxide.

the density of interface traps is observed for the 30 nm device. This is attributed to voltage driven degradation mechanisms which only occur for the device with the largest voltage drop across the oxide. As presented in section 6, the behavior of the devices with different oxide thicknesses was equivalent in the experiments with an oxide field smaller than the above mentioned border for common oxide fields during NBTS.

7.2 Spatial defect distribution

This section is addressed to the remaining discrepancy in the threshold voltage shift for the different oxide thicknesses as presented in section 6. It is proposed that the difference in the data measured by the change in the drain current is not due to a higher degradation of the thick oxide devices, as suggested by other research groups [AM05], but due to *NBTI induced defects* in the bulk of the SiO_2. The existence of NBTI induced charge in the bulk of the oxide after stress is supported by microscopic models as the ones suggested by Tibor Grasser [GKGAHN09] or Vincent Huard [HDP06].

Impact of charges in the oxide

In section 1.4, the influence of oxide charges on the potential situation in a MOS structure or a MOSFET has been studied. The voltage drop across the oxide

originating from a charge distribution ρ is given by (equation (27) on page 20)

$$V = -\frac{1}{C_{\mathrm{ox}}}\left(\frac{1}{T_{\mathrm{ox}}}\int_0^{T_{\mathrm{ox}}} x\rho(x)\mathrm{d}x\right). \qquad (115)$$

The integration variable x is chosen such, that $x = 0$ corresponds to the poly–SiO_2 interface and $x = T_{\mathrm{ox}}$ to the Si–SiO_2 interface. The integral is similar to the expression of a linearly weighted mean, so charges located near the Si–SiO_2 interface have a stronger impact on the voltage drop than charges near the poly–SiO_2 interface. During the subsequent argumentations is assumed that the oxide charge density is negligible in a virgin device. Thus, ρ_0 is regarded as the change of the charge density, and the voltage drop as calculated from equation (115) is the shift in the threshold voltage.

The threshold voltage shift as a function of the gate oxide thickness

Figure 53 shows the dependence of the threshold voltage shift on the gate oxide thickness. The dependence is clearly not linearly, as it would be the case when

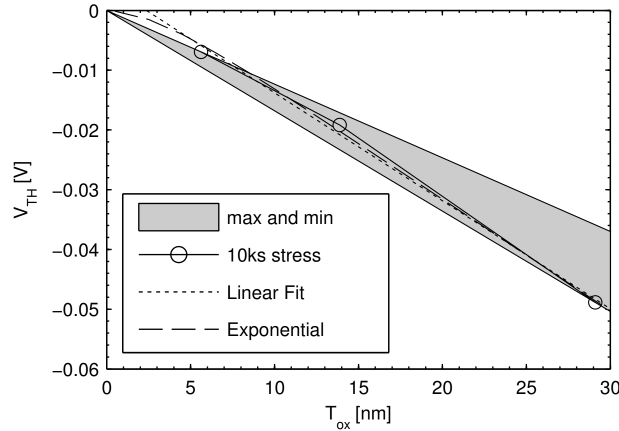


Fig. 53: Comparison of proposed spatial defect density distributions with the total threshold voltage shift measured 10 ms after the termination of NBTS with t_{str}= 10 ks duration. The gray area corresponds to maximum and minimum values of a capacitance scaling approach (see equation (113) on page 74). The dotted line is a linear fit through the data points. The dashed line is a fit of the exponential charge distribution of equation (116) on page 88.

only charges located at the interface would be created during NBTS. The upper border of the gray area in Fig. 53 corresponds to the approach, that devices with thicker gate oxides drift apparently more than those with thinner oxides. In order to understand the measured relation, spatial charge densities have been proposed and the resulting $V_{\mathrm{TH}}T_{\mathrm{ox}}$ curve was calculated by solving the integral in equation (115). Some possible charge distributions are depicted in Fig. 54. Their according shifts are listed in Table 5. Among several proposed spatial charge distributions, only those with a defect layer near the Si–SiO_2 interface propose a $V_{\mathrm{TH}}T_{\mathrm{ox}}$ relation which does not intersect in the origin. As can be

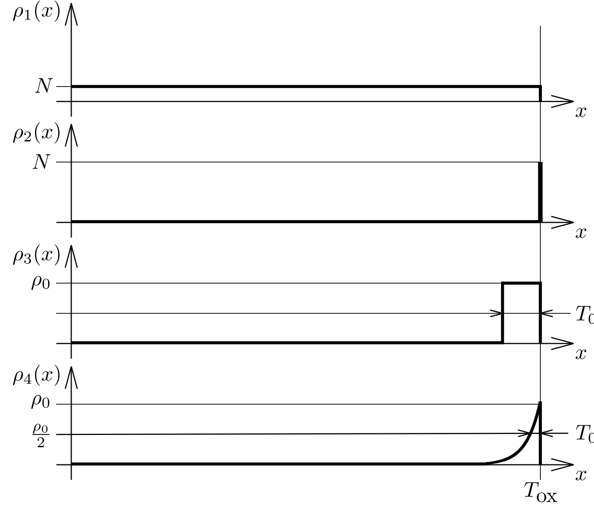


Fig. 54: Possible defect distributions as listed in Table 5.

seen in Fig. 53, the apparently best fit of the experimentally gathered data is obtained assuming an exponential decay of the charge density from the Si–SiO_2 interface toward the bulk of the oxide,

$$\rho(x) = \rho_0 \exp\left(\frac{x - T_{\text{ox}}}{T_0/\ln 2}\right). \tag{116}$$

The constant $\ln 2$ causes the characteristic length T_0 to be the depth into the bulk of the oxide, where the charge density decreases to the half of its initial value. The according voltage shift caused by this charge distribution is calculated as,

$$\begin{aligned}\Delta V &= -\frac{1}{C_{\text{ox}}}\left(\frac{1}{T_{\text{ox}}}\int_0^{T_{\text{ox}}} x\rho_0 \exp\left(\frac{x - T_{\text{ox}}}{T_0/\ln(2)}\right)\mathrm{d}x\right) \\ &= \frac{\rho_0}{\varepsilon_{\text{SiO2}}}\left(\frac{T_0}{\ln 2}\right)^2\left(1 - \mathrm{e}^{-\frac{T_{\text{ox}}\ln 2}{T_0}}\right) - \frac{\rho_0 T_0}{\varepsilon_{\text{SiO2}}\ln 2}T_{\text{ox}}. \end{aligned} \tag{117}$$

In Fig. 55 the exponential spatial charge distribution is depicted for three different oxide thicknesses. The *same* spatial defect distribution has to be assumed for *all* oxide thicknesses in order to explain the measurement data.

The $V_{\text{TH}}T_{\text{ox}}$ relation for the exponential charge distribution can be fitted to total threshold voltage shifts measured 10 ms after termination of NBTS, in order to determine the half-value charge layer depth and the density of charges per cubic centimeter. The result of the fit of the experimental data after different stress durations is depicted in Fig. 56. The according results for the half-value charge layer thicknesses and the charge densities are given in Table 6. The half-value of the charge layer is roughly 2 nm and decreasing slightly with increasing stress duration.

When performing the same experiment using the poly heater technique described in section 5.2, the fits become even more accurate, as depicted in Fig. 57. The half-value charge layer thickness is slightly smaller compared to the result

Table 5: Summary of a few possible spatial charge distributions and their voltage shifts according to equation (115). For the uniform spatial charge distribution it is proposed that equivalent NBTS causes an equivalent overall number of oxide charges in the oxide. T_0 is a characteristic length for the thickness of a charge layer near the Si–SiO_2 interface. The deviation of the voltage shift for the exponential charge distribution is found in equation (117). The numbers in the first column (Nr) refer to the outlines of the distribution functions in Fig. 54.

Nr	Description	$\rho(x)$	ΔV
1	Uniform	N/T_{ox}	$-NT_{ox}/2\varepsilon_{SiO2}$
2	Interface	$N\,\delta\,(x - T_{ox})$	$-NT_{ox}/\varepsilon_{SiO2}$
3	Defect layer	$\rho_0\Theta\,(T_{ox} - x)\,\Theta\,(T_{ox} - T_0 - x)$	$\rho_0 T_0/\varepsilon_{SiO2}\,(T_0/2 - T_{ox})$
4	Exponential	$\rho_0 \exp\left(\ln 2\,(x - T_{ox})\,/\,(T_0)\right)$	$\rho_0 T_0^2/\left(\varepsilon_{SiO2}\ln^2 2\right)$
		$(1 - \exp(-T_{ox}\ln 2/T_0))$	
		$-\rho_0 T_0 T_{ox}/\varepsilon_{SiO2}$	

Table 6: Fit parameters resulting from the assumption of an exponential charge distribution with the data of Fig. 56. T_0 is the half-value charge layer thickness and ρ_0 the charge density at the Si–SiO_2 interface. The defect layer thickness after t_{str}= 1 s may be erroneous because of the overall small measurement signal after the short stress period.

t_{str}	T_0	ρ_0
1 s	32.5 nm	$1.6 \times 10^{15}\,cm^{-3}$
10 s	4.8 nm	$8.4 \times 10^{15}\,cm^{-3}$
100 s	2.8 nm	$2.7 \times 10^{16}\,cm^{-3}$
1 ks	2.4 nm	$6.3 \times 10^{16}\,cm^{-3}$
10 ks	2.1 nm	$1.3 \times 10^{17}\,cm^{-3}$
100 ks	2.2 nm	$2.3 \times 10^{17}\,cm^{-3}$

Table 7: Fit parameters resulting from the assumption of an exponential charge distribution with the data of Fig. 57. The experiment was performed using the poly heater. The degradation was thus measured at −60°C.

t_{str}	T_0	ρ_0
1 s	3.5 nm	$3.1 \times 10^{16}\,cm^{-3}$
10 s	3.2 nm	$4.5 \times 10^{16}\,cm^{-3}$
100 s	2.5 nm	$7.2 \times 10^{16}\,cm^{-3}$
1 ks	1.7 nm	$1.4 \times 10^{17}\,cm^{-3}$
10 ks	1.2 nm	$2.6 \times 10^{17}\,cm^{-3}$

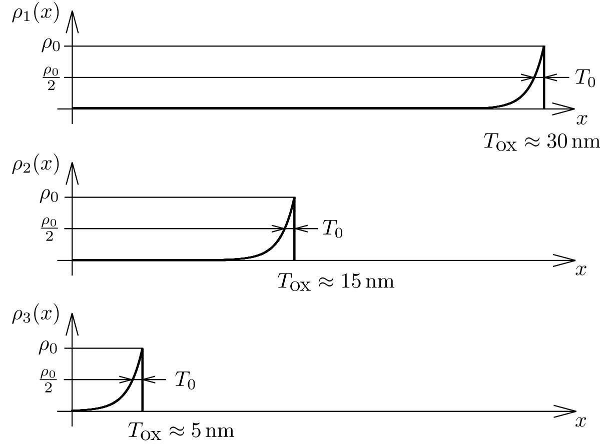


Fig. 55: Spatial defect density distributions for different oxide thicknesses T_{ox}^1, T_{ox}^2, T_{ox}^3 under the assumption that equivalent NBTS causes an equivalent defect density distribution.

obtained at stress temperature, as listed in Table 7. The charge densities are comparable higher than the charge densities obtained at stress temperature. This is because the recovery dynamics of NBTS induced defects are decelerated at low temperatures. The remaining charge, after the short recovery period before the first measurement is performed, is significantly decreased at high temperatures.

The evolution of the charge layer thickness over the stress duration is depicted in Fig. 58. Similar values are obtained for the thickness, nearly independent of the temperature during recovery. The thickness decreases with increasing stress duration, which may also be explained by an increasing dominance of the charge stored in interface traps.

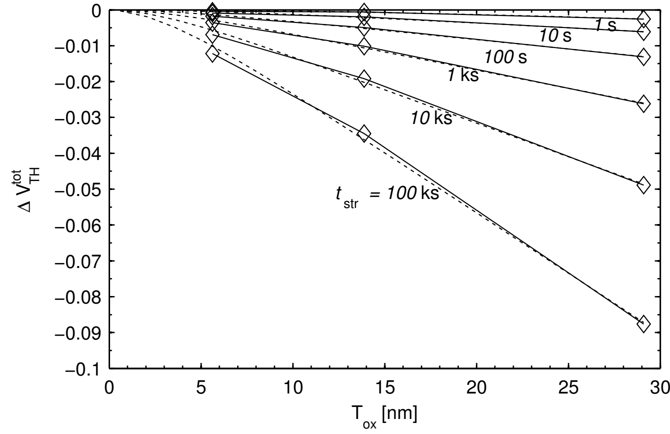


Fig. 56: Total threshold voltage shifts, measured 10 ms after termination of stress with t_{str} duration, over the gate oxide thickness. The measurement was performed at stress temperature of 125°C. The results of the fits of the exponential charge distribution of equation (116) on page 88 (dashed lines) are given in Table 6.

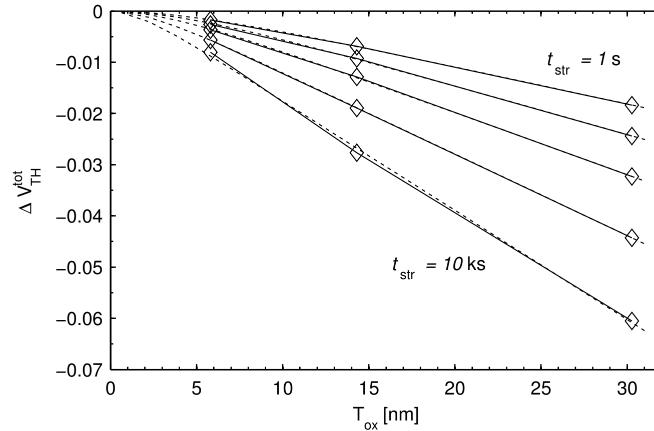


Fig. 57: $V_{\mathrm{TH}}T_{\mathrm{ox}}$ plot of an experiment using the poly heater technique. The degradation was therefore not measured at the stress temperature of 125°C, but at −60°C. The dashed lines are fits of the exponential charge distribution of equation 116. The resulting values for the half-value charge layer thickness are listed in Table 7.

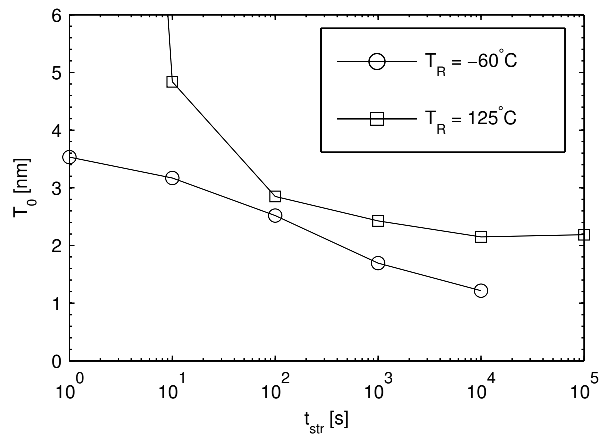


Fig. 58: Thickness of the defect layer of the exponential distribution function over stress duration. The circles indicate the measurement with the use of the poly heater. The squares indicate the experiment where the temperature during recovery was the same as during stress. The first measurement point for the 125°C curve is outside the plot because it is regarded to be highly affected by measurement errors.

Bibliography

[Aic07] Thomas Aichinger. "Implementation of the charge pumping method for MOS characterization into existing soft- and hardware laboratory environment." MA thesis. Karl–Franzens University Graz, Apr. 2007.

[AM05] M. A. Alam and S. Mahapatra. "A comprehensive model of PMOS NBTI degradation." In: *Microelectronics Reliability* 45.1 (2005), pp. 71–81.

[AN08] Thomas Aichinger and Michael Nelhiebel. "Advanced Energetic and Lateral Sensitive Charge Pumping Profiling Methods for MOSFET Device Characterization – Analytical Discussion and Case Studies." In: *IEEE Transactions on Device and Materials Reliability* 8 (2008), pp. 509–518.

[ANDG10] Thomas Aichinger, Michael Nelhiebel, Stefan Decker, and Tibor Grasser. "Energetic Distribution of Oxide Traps Created under Negative Bias Temperature Stress and their Relation to Hydrogen." In: *Applied Physics Letters* 96.13 (Mar. 2010), pp. 133511–133511–3.

[ANEG10] Thomas Aichinger, Michael Nelhiebel, Sascha Einspieler, and Tibor Grasser. "In Situ Polyheater – A reliable tool for performing fast and defined temperature switches on chip." In: *IEEE Transactions on Device and Materials Reliability* 10 (2010), pp. 3–8.

[ANG08] Thomas Aichinger, Michael Nelhiebel, and Tibor Grasser. "On the temperature dependence of NBTI recovery." In: *Microelectronics Reliability* 48 (2008), pp. 1178–1184.

[ANG09a] Thomas Aichinger, Michael Nelhiebel, and Tibor Grasser. "A Combined Study of p- and n-Channel MOS Devices to Investigate the Energetic Distribution of Oxide Traps After NBTI." In: *IEEE Transactions on Electron Devices* 56.12 (Dec. 2009), pp. 3018–3026.

[ANG09b] Thomas Aichinger, Michael Nelhiebel, and Tibor Grasser. "Unambiguous identification of the NBTI recovery mechanism using ultra–fast temperature changes." In: Apr. 2009, pp. 2 –7.

[AWDH08] D.S. Ang, S. Wang, G.A. Du, and Y.Z. Hu. “A Consistent Deep-Level Hole Trapping Model for Negative Bias Temperature Instability.” In: *IEEE Transactions on Device and Materials Reliability* 8.1 (2008), pp. 22–34.

[Ber66] C.N. Berglund. “Surface states at steam-grown silicon-silicon dioxide interfaces.” In: *IEEE Transactions on Electron Devices* 13.10 (Oct. 1966), pp. 701–705.

[BJ69] J.S. Brugler and P.G.A. Jespers. “Charge pumping in MOS devices.” In: *IEEE Transactions on Electron Devices* 16.3 (1969), pp. 297–302.

[Bre78] J.R. Brews. “A charge-sheet model of the MOSFET.” In: *Solid-State Electronics* 21.2 (1978), pp. 345–355.

[CKRMK04] S. Chakravarthi, A. Krishnan, V. Reddy, C.F. Machala, and S. Krishnan. “A comprehensive framework for predictive modeling of negative bias temperature instability.” In: *IEEE International Reliability Physics Symposium.* Apr. 2004, pp. 273 –282.

[CLKK06] J.P. Campbell, P.M. Lenahan, A.T. Krishnan, and S. Krishnan. “Observations of NBTI-induced atomic-scale defects.” In: *IEEE Transactions on Device and Materials Reliability* 6.2 (June 2006), pp. 117–122.

[DCA93] D. J. DiMaria, E. Cartier, and D. Arnold. “Impact ionization, trap creation, degradation, and breakdown in silicon dioxide films on silicon.” In: *Journal of Applied Physics* 73.7 (1993), pp. 3367–3384.

[Dea80] B.E. Deal. “Standardized terminology for oxide charges associated with thermally oxidized silicon.” In: *IEEE Transactions on Electron Devices* 27.3 (Mar. 1980), pp. 606–608.

[DVMMH95] M. Depas, B. Vermeire, P. W. Mertens, R. L. Van Meirhaeghe, and M. M. Heyns. “Determination of tunnelling parameters in ultra-thin oxide layer poly-Si/SiO2/Si structures.” In: *Solid-State Electronics* 38.8 (1995), pp. 1465–1471.

[Edw91] Arthur H. Edwards. “Interaction of H and H_2 with the silicon dangling orbital at the (111) Si/SiO_2 interface.” In: *Physical Review B* 44.4 (July 1991), pp. 1832–1838.

[Fle92] D.M. Fleetwood. “‘Border traps’ in MOS devices.” In: *IEEE Transactions on Nuclear Science* 39.2 (Apr. 1992), pp. 269–271.

[GFSLWCH97] Ashawant Gupta, Peng Fang, Miryeong Song, Ming-Ren Lin, Don Wollesen, Kai Chen, and Chenming Hu. “Accurate Determination of Ultrathin Gate Oxide Thickness and Effective Polysilicon Doping of CMOS Devices.” In: *IEEE Electron Device Letters* 18.12 (Dec. 1997), pp. 580–582.

[GKGAHN09] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel. “A two-stage model for negative bias temperature instability.” In: *IEEE International Reliability Physics Symposium*. 2009, pp. 33–44.

[Gmb10] Advanced Temperature Test Systems GmbH. *Specification ATT Low temperature chuck system M200*. Advanced Temperature Test Systems GmbH. Apr. 2010. URL: `http://www.att-systems.com/m200.htm`.

[GMBDK84] G. Groeseneken, H.E. Maes, N. Beltran, and R.F. De Keersmaecker. “A reliable approach to charge-pumping measurements in MOS transistors.” In: *IEEE Transactions on Electron Devices* 31.1 (1984), pp. 42–53.

[GN66] A. Goetzberger and H.E. Nigh. “Surface charge after annealing of Al-SiO2-Si structures under bias.” In: *Proceedings of the IEEE* 54.10 (Oct. 1966), pp. 1454–1454.

[HAGGHS05] M. Houssa, M. Aoulaiche, S. De Gendt, G. Groeseneken, M. M. Heyns, and A. Stesmans. “Reaction-dispersive proton transport model for negative bias temperature instabilities.” In: *Applied Physics Letters* 86.9, 093506 (2005), p. 093506.

[HDP06] V. Huard, M. Denais, and C. Parthasarathy. “NBTI degradation: From physical mechanisms to modelling.” In: *Microelectronics and Reliability* 46.1 (2006), pp. 1–23.

[HP94] C R Helms and E H Poindexter. “The silicon-silicon dioxide system: Its microstructure and imperfections.” In: *Reports on Progress in Physics* 57.8 (1994), pp. 791–852.

[JS77] Kjell O. Jeppson and Christer M. Svensson. “Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices.” In: *Journal of Applied Physics* 48.5 (1977), pp. 2004–2014.

[KADCGG05] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin. “Disorder–controlled–kinetics Model for Negative Bias Temperature Instability and Its Experimental Verification.” In: *IEEE International Reliability Physics Symposium*. 2005, pp. 381–387.

[Len07] P.M. Lenahan. “Deep level defects involved in MOS device instabilities.” In: *Microelectronics Reliability* 47.6 (2007). Modelling the Negative Bias Temperature Instability, pp. 890–898.

[Lil30] Julius Edgar Lilienfeld. “Method and Apparatus for controlling electric currents.” English. Pat. 1745175. Jan. 1930.

[LMJFW01] P.M. Lenahan, T.D. Mishima, J. Jumper, T.N. Fogarty, and R.T. Wilkins. “Direct experimental evidence for atomic scale structural changes involved in the interface-trap transformation process.” In: *IEEE Transactions on Nuclear Science* 48.6 (Dec. 2001), pp. 2135 –2135.

[LO94] A.J. Lelis and T.R. Oldham. "Time dependence of switching oxide traps." In: *IEEE Transactions on Nuclear Science* 41.6 (Dec. 1994), pp. 1835–1843.

[LS69] M. Lenzlinger and E. H. Snow. "Fowler-Nordheim Tunneling into Thermally Grown SiO_2." In: *Journal of Applied Physics* 40.1 (1969), pp. 278–283.

[MAVIGMSA07] S. Mahapatra, K. Ahmed, D. Varghese, A.E. Islam, G. Gupta, L. Madhav, D. Saha, and M.A. Alam. "On the Physical Mechanism of NBTI in Silicon Oxynitride p-MOSFETs: Can Differences in Insulator Processing Conditions Resolve the Interface Trap Generation versus Hole Trapping Controversy?" In: *IEEE International Reliability Physics Symposium*. Apr. 2007, pp. 1–9.

[MM66] Yoshio Miura and Yasuo Matukura. "Investigation of Silicon-Silicon Dioxide Interface Using MOS Structure." In: *Japanese Journal of Applied Physics* 5.2 (1966), pp. 180–180.

[NB02] E. H. Nicollian and J. R. Brews. *MOS (Metal Oxide Semiconductor) Physics and Technology*. first. John Wiley & Sons, 2002, p. 928.

[PSM75] Göran P. Petersson, Christer M. Svensson, and J. Maserjian. "Resonance effects observed at the onset of Fowler-Nordheim tunneling in thin MOS structures." In: *Solid-State Electronics* 18 (1975), pp. 449–451.

[RBHGS07] H. Reisinger, O. Blank, W. Heinrigs, W. Gustin, and C. Schlunder. "A Comparison of Very Fast to Very Slow Components in Degradation and Recovery Due to NBTI and Bulk Hole Trapping to Existing Physical Models." In: *IEEE Transactions on Device and Materials Reliability* 7.1 (Mar. 2007), pp. 119 –129.

[RVWHMAGGS08] H. Reisinger, R.P. Vollertsen, P.J. Wagner, T. Huttner, A. Martin, S. Aresu, W. Gustin, T. Grasser, and C. Schlunder. "The effect of recovery on NBTI characterization of thick non-nitrided oxides." In: *IEEE International Integrated Reliability Workshop Final Report*. Oct. 2008, pp. 1–6.

[Sch06] Dieter K. Schroder. *Semiconductor Material and device characterization*. 3rd. John Wiley & Sons, Inc., 2006, p. 779.

[Sch07] Dieter K. Schroder. "Negative Bias Temperature Instability: What do we understand?" In: *Microelectronics Reliability* 47 (2007), pp. 841–852.

[See] P. Seegebrecht. *Elektronische Bauelemente und Schaltungen*. Technische Fakultät der Christian-Albrechts-Universität zu Kiel, Lehrstuhl für Halbleiterphysik,

[SH94] K.F. Schuegraf and Chenming Hu. “Hole injection SiO2 breakdown model for very low voltage lifetime extrapolation.” In: *Electron Devices, IEEE Transactions on* 41.5 (1994), pp. 761–767.

[SMPD98] Ying Shi, T. P. Ma, Sharad Prasad, and Sumit Dhanda. “Polarity dependent gate tunneling currents in dual-gate CMOSFETs.” In: *IEEE Transactions on Electron Devices* 45.11 (Nov. 1998), pp. 2355–2360.

[SN06] Simon M. Sze and Kwok K. Ng. *Physics of Semiconductor Devices*. third. John Wiley & Sons, 2006.

[SZ06] J.H. Stathis and S. Zafar. “The negative bias temperature instability in MOS devices: A review.” In: *Microelectronics and Reliability* 46.2-4 (2006), pp. 270–286.

[TL94] T. L. Tewksbury and H.-S. Lee. “Characterization, modeling, and minimization of transient threshold voltage shifts in MOSFETs.” In: *IEEE Journal of Solid-State Circuits* 29.3 (1994), pp. 239–252.

[Zeg10] Bart Van Zeghbroeck. *Principles of Semiconductor Devices*. Apr. 2010. URL: `http://ecee.colorado.edu/~bart/book/`.

Acronyms

$I_\mathrm{D}V_\mathrm{G}$ drain current–gate voltage.

AC alternating current.

BTI bias temperature instability.

BTS bias temperature stress.

CMOS complementary metal oxide semiconductor.

CP charge pumping.

CV capacitance voltage.

Fig. figure.

H neutral hydrogen.

H_2 diatomic hydrogen.

HDL Harry–Diamond–Laboratories.

HF CV high frequency capacitance voltage.

LF CV low frequency capacitance voltage.

MOS metal oxide semiconductor.

MOSFET MOS field effect transistor.

MSM measurement–stress–measurement.

NBTI negative bias temperature instability.

NBTS negative bias temperature stress.

nMOS transistor n channel MOS field effect transistor.

O oxygen.

pMOS transistor p channel MOS field effect transistor.

poly polycrystalline silicon.

RD model reaction–diffusion model.

Si silicon.

Si–H hydrogen passivated silicon dangling bond.

SiO_2 silicon dioxide.

SRH Shockley–Read–Hall.

Symbols

Symbol	Description	Unit
A	Effective area of a MOS structure or MOSFET	cm^2
C	Capacitance	F/cm^2
C_{FB}	Capacitance value at the flat–band voltage of MOSFET or a MOS structure	F/cm^2
C_S	Capacitance of the semiconductor	F/cm^2
C_{ox}	Oxide capacitance	F/cm^2
D_X	Diffusion constant of the species X	cm/s
E_B	Barrier height	eV
E_C	Conduction band edge energy	eV
E_F	FERMI level energy	eV
E_G	Band gap energy	eV
E_T	Energy level of an interface trap	eV
E_V	Valence band edge energy	eV
E_i	Energy level of the intrinsic semiconductor	eV
I_{CP}	Charge pumping current	A
I_D	Current through the drain contact of a MOSFET	A
J_{FN}	FOWLER–NORDHEIM tunneling current density	A/cm^2
L	MOSFET channel length	cm
N	Net effective number of charges	cm^{-2}
N_A	Acceptor impurity concentration	cm^{-3}
N_A^-	Ionized acceptor impurity concentration	cm^{-3}
N_C	Effective density of states in the conduction band	cm^{-3}
N_D	Donor impurity concentration	cm^{-3}
N_D^+	Ionized donor impurity concentration	cm^{-3}
N_{IT}	Number of interface traps	cm^{-2}
N_{SiH}	Number of hydrogen passivated Si dangling bonds at the interface	cm^{-2}
N_S	Number of semiconductor charges	cm^{-2}
N_X	Number of diffusing species in the RD model	cm^{-2}
N_p	Effective gate poly impurity concentration	cm^{-3}
Q	Net effective charge	C/cm^2
Q_G	Net effective charge in the gate	C/cm^2
Q_{IT}	Total charge in interface traps	C/cm^2
Q_S	Net effective charge of the semiconductor in the MOS structure	C/cm^2
Q_i	Inversion layer charge	C/cm^2
Q_m	Mobile ionic oxide charge	C/cm^2
Q_{ot}	Oxide trapped charge	C/cm^2

Symbol	**Description**	**Unit**
Q_{ox}	Fixed oxide charge	C/cm^2
S	Closed surface in a surface integral	cm^2
T	Temperature	K
T_0	Half-value charge layer thickness of an exponential charge layer	cm
T_{ox}	Effective oxide thickness	cm
V	Voltage value	V
V_{B}	Voltage applied to the bulk of the semiconductor of a MOSFET with respect to the source	V
$V_{\text{D,sat}}$	Drain saturation voltage	V
V_{D}	Voltage applied to the drain of a MOSFET with respect to the source	V
V_{FB}	Voltage at which the energy band edges in the semiconductor are not bended	V
V_{G}	Potential difference between the gate contact and the bulk of the semiconductor in MOS structures and between the gate and the source in MOSFETs	V
V_{H}	High level of a trapezoidal pulse	V
V_{L}	Low level of a trapezoidal pulse	V
V_{R}	Value of the bias applied to the gate during recovery after BTS	V
V_{S}	Value of the stress bias applied to the gate during BTS	V
V_{TH}	Threshold voltage at which the semiconductor at the Si–SiO_2 interface is fully inverted	V
V_{ox}	Potential drop across the oxide	V
V_{p}	Voltage drop across the gate poly because of poly depletion	V
W	MOSFET channel width	cm
Z	Electrical impedance	Ω
ΔE_{CP}	Electrical active energy region within the band gap during CP	eV
ΔI_{CP}	Change of the charge pumping current	A
ΔI_{D}	Change of the drain current of a MOSFET	A
ΔN_{tot}	Change of the total number of effective charges per area	cm^{-2}
ΔV_{TH}	Change of the threshold voltage (threshold voltage shift)	V
α	Power law exponent of NBTI induced interface trap creation as a function of stress time	1
χ	Electron affinity	eV
γ	Slope of a linear function	
$\hbar$	Reduced PLANCK constant $\hbar = \frac{h}{2\pi} = 6.5821 \times 10^{-16}\,\text{eV s} = 1.05458 \times 10^{-34}\,\text{J s}$ [SN06]	J s
$\mathcal{E}$	Electric field	V/cm
$\mathcal{E}_{\text{S}}$	Electric field in the semiconductor	V/cm
$\mathcal{E}_{\text{ox}}$	Electric field in the oxide	V/cm
μ	Electrical mobility	$\text{cm}^2/(\text{V s})$
μ_{X}	Electrical mobility of the species X	$\text{cm}^2/(\text{V s})$
ϕ_{Bn}	Potential in the bulk of an n doped semiconductor	V
ϕ_{Bp}	Potential in the bulk of an p doped semiconductor	V
ϕ_{B}	Potential in the bulk of a semiconductor	V
ϕ_{S}	Surface potential at the Si–SiO_2 interface	V
ϕ_{ms}	Work function difference between the gate and the substrate	V
ψ_{S}	Total band bending in the semiconductor	V

Symbol	Description	Unit
ψ_n	Band bending in an n doped semiconductor with respect to the bulk	V
ψ_p	Band bending in an p doped semiconductor with respect to the bulk	V
ρ	Charge density	$\mathrm{C/cm^3}$
ρ_0	Charge density at the Si–SiO_2 interface of a spatial charge distribution	$\mathrm{C/cm^3}$
σ	Capture cross section	$\mathrm{cm^2}$
τ	Emission/capture time constant	s
ε	Permittivity	$\mathrm{A\,s/\,(V\,cm)}$
$\varepsilon_{\mathrm{SiO_2}}$	Permittivity of SiO_2 $\varepsilon_{\mathrm{SiO_2}} = 3.9 \times 8.854188 \times 10^{-14}\,\mathrm{A\,s/V\,cm}$[SN06]	$\mathrm{A\,s/\,(V\,cm)}$
$\varepsilon_{\mathrm{Si}}$	Permittivity of Si $\varepsilon_{\mathrm{Si}} = 11.9 \times 8.854188 \times 10^{-14}\,\mathrm{A\,s/V\,cm}$[SN06]	$\mathrm{A\,s/\,(V\,cm)}$
$\vec{u}$	Unit vector perpendicular to the surface S pointing outward	1
f	AC signal frequency	Hz
k	BOLTZMANN constant $k = 8.6174 \times 10^{-5}\,\mathrm{eV/K} = 1.38066 \times 10^{-23}\,\mathrm{J/K}$ [SN06]	J/K
k_F	RD model forward dissociation rate constant	$\mathrm{s^{-1}}$
k_R	RD model reverse dissociation rate constant	$\mathrm{s^{-1}}$
$m_{\mathrm{SiO_2}}$	Effective mass of an electron in the conduction band of SiO_2	kg
m_e	Rest mass of an electron $m_\mathrm{e} = 9.1095 \times 10^{-31}\,\mathrm{kg}$	kg
n	Free electron density	$\mathrm{cm^{-3}}$
p	Free hole density	$\mathrm{cm^{-3}}$
q	Elementary charge $1.60218 \times 10^{-19}\,\mathrm{C}$[SN06]	C
t	Time	s
t_CP	Duration of a CP measurement after constant bias recovery and BTS	s
t_H	High time of a trapezoidal pulse	s
t_L	Low time of a trapezoidal pulse	s
t_f	Fall time of a trapezoidal pulse	s
t_perm	Duration of the measurement of the permanent component of BTS induced degradation	s
t_p	Pulse period of a trapezoidal pulse	s
t_rec	Duration of constant bias recovery after BTS	s
t_r	Rise time of a trapezoidal pulse	s
t_str	Duration of stress in a bias temperature instability (BTI) experiment	s
v	Average carrier velocity	cm/s
v_th	Thermal velocity	cm/s
x	Dimension perpendicular to the Si–SiO_2 interface	cm
x_D	Depletion width	cm

List of Figures

1 Charges in the MOS structure 9
2 Band diagram of a MOS structure without electric contact 10
3 FERMI level for Si as a function of temperature 11
4 Band diagram of a MOS structure with electric contact 11
5 Value of the FERMI–DIRAC integral 13
6 Semiconductor charge over surface potential 15
7 Band diagram of a n^{++}/n MOS structure in accumulation 16
8 Band diagram of a n^{++}/n MOS structure in depletion 17
9 Band diagram of a n^{++}/n MOS structure in inversion 18
10 Pb0 and Pb1 defects at the Si–SiO_2 interface 20
11 E'_γ center in SiO_2 21
12 Theoretical low frequency CV curve 23
13 Schematic band situation during FOWLER–NORDHEIM tunneling 24
14 Schematic band situation during direct tunneling 25
15 Sketch of a pMOSFET 26
16 Evolution of the power law exponent according to the RD model 31
17 Schematic picture of the recoverable part of the GRASSER model 33
18 Schematic picture of the permanent part of the GRASSER model 34
19 Schematic picture of the GRASSER model 36

20 Theoretical dependence of the gate oxide field on the gate voltage 39
21 Band diagram of a n^{++}/n MOS structure in strong inversion . . 41
22 Tunneling current plots and extrapolation 43
23 Integration of GAUSS's law in a MOS structure 44
24 Electric oxide field from LF CV curve 56
25 Tunneling current plots for three oxide thicknesses 57
26 Surface potentials for three oxide thicknesses 57
27 Comparison between experimental and theoretical LF CV curve . 58
28 Determination of N_D from MOSFET threshold voltage 58
29 Doping concentration extracted with different methods 59
30 Trapezoidal pulse output of a pulse generator unit 60
31 Drift of the $I_D V_G$ curve after stress 61
32 Calculation of ΔV_{TH} from ΔI_D 62
33 Minimal measurement–stress–measurement experiment 63
34 Device embedded in poly heater structure 64
35 Outline of a device with a poly heater 64
36 Minimal MSM experiment with poly heater 65

37 The author in the Julian Alps 69
38 Comparison drain current criterion and BERGLUNDs method . . 71
39 MSM experiment for the comparison of oxide thicknesses 72
40 $I_\mathrm{D}V_\mathrm{G}$ characteristics before and after stress 74
41 Recovery of the drain current 75
42 Recovery of the threshold voltage 76
43 Recovery of the effective charges per area 77
44 Recovery of the effective charges per area after all stress durations 78
45 Total threshold voltage shift over stress duration 79
46 Change of CP current over stress duration 79
47 Change of the permanent component over stress duration 80
48 Impact ionization after FOWLER–NORDHEIM tunneling 82
49 Anode hole injection 83
50 Incremental sweep for 30 nm device 84
51 FOWLER–NORDHEIM plot for three oxide thicknesses 85
52 Change of CP current for stress at higher oxide fields 86
53 Ideal defect distributions fitted to experiment 87
54 Some possible defect distributions 88
55 Equivalent defect distribution for all oxide thicknesses 90
56 Threshold voltage shift over gate oxide thickness 91
57 Threshold voltage shift over gate oxide thickness using the poly heater 91
58 Defect layer thickness over stress duration 92

List of Tables

1 Approximating functions for F 19

2 Approximations for ψ_{S} in strong inversion/accumulation 40
3 V_{FB} for all combinations of gate poly and substrate toping types 41
4 Approximations for $\mathcal{E}_{\mathrm{ox}}$ in strong inversion or accumulation . . . 42

5 Shifts of ideal charge distributions 89
6 Charge layer thicknesses for the exponential charge distribution . 89
7 Charge layer thicknesses for the exponential charge distribution using the poly heater 89