# DISSERTATION

# ON THE ROLE OF HYDROGEN IN SILICON DEVICE DEGRADATION AND METALIZATION PROCESSING

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften

eingereicht an der Technischen Universität Wien Fakultät für Elektrotechnik und Informationstechnik von

MAG. THOMAS AICHINGER

Die Dissertation wurde betreut von Prof. Tibor Grasser (TU Wien) und Dr. Michael Nelhiebel (Infineon Technologies Austria) und ausgeführt am

KOMPETENZZENTRUM FÜR AUTOMOBIL- UND INDUSTRIE-ELEKTRONIK (KAI)

Villach, am 13. August 2010

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit selbständig sowie ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Villach, am 13. August 2010

# Acknowledgment

FIRST AND FOREMOST I want to thank Prof. Tibor Grasser and Dr. Michael Nelhiebel for supervising and supporting me during the entire three years of my PhD thesis. Thank you for all our fruitful discussions, for paper reviewing and for motivation.

I am also very grateful to my colleagues BSc. Schascha Einspieler, DI Gregor Pobegen and DI Stefan Puchner who supported me with measurements and discussions during their internships, master and PhD thesis at KAI.

I thank Prof. Herbert Grünbacher and Dr. Thomas Detzel for giving me the opportunity to join KAI, for providing me an excellent infrastructure and measurement equipment in the lab, and for the great industrial and academic network represented by KAI, Infineon and the TU Vienna.

Last but not least I want to thank all members of my family for financial and mental support. Many thanks, in particular, to my mother and my girlfriend who have always been taking care for me, provided me love and gave me a home.

# Kurzfassung

BAUELEMENT-Degradation und Gateoxid Zuverlässigkeit spielen in der modernen Halbleiterindustrie eine zunehmend wichtige Rolle, da die Anforderungen an die Bauteile mit zunehmender Miniaturisierung, höheren Belastungen in der Schaltung und innovativen Technologieentwicklungen stetig steigen. Neben der Zuverlässigkeit von mechanischen Lötstellen und Verdrahtungen im Produkt, ist es besonders wichtig, den Transistor schon in einem früheren Stadium der Prozessierung gegen Umweltkontaminationen und mechanische Verspannungen zu schützen. Um nebenbei noch die verlustarme Schaltung von hohen Strömen sowie Kurzschlussfestigkeit gewährleisten zu können, werden am Ende der Prozessierung eines Leistungshalbleiters (back-end of line) verschiedene Isolations- und Metallisierungsebenen (layer stacks) in einer gut durchdachten Zusammensetzung und Reihenfolge auf die elektrisch aktiven Bereiche des Transistors aufgebracht. Es ist bekannt, dass sich verschiedene Passivierungskonzepte und Schichtsequenzen auch auf die elektrische Charakteristik und Zuverlässigkeit des Siliziumbauteils auswirken, da im back-end Prozess Wasserstoff in grossen Mengen in das System eingebracht wird. Es wird angenommen, dass hohe Temperaturen, wie sie zum Beispiel bei der Abscheidung und dem thermischen Ausheilen auftreten, den eingebrachten Wasserstoff in Richtung Gateoxid diffundieren lassen, wo er Defekte im Oxid oder an der Grenzfläche absättigen kann. Unter elektrischem und thermischem Stress kann sich dieser Wasserstoff jedoch in weiterer Folge wieder ablösen, wodurch zuvor passivierte Defekte wieder freigelegt werden.

So wie viele Halbleiterhersteller weltweit, hat sich auch Infineon Villach in der Vergangenheit wiederholte Male mit dieser Problematik auseinandergesetzt und obwohl einzelne empirische Lösungsansätze für bestimmte Technologien bereits vorliegen, fehlt noch immer ein globales Verständnis der zugrundeliegenden Physik vom Wechselspiel zwischen Prozessierung und Bauelement-Degradation. Da zu erwarten ist, dass neue Probleme ähnlicher Natur bei zukünftigen Technologien auftreten werden, ist es besonders wichtig, ein physikalisch fundiertes Modell zu entwickeln, dass die Zusammenhänge von Defektgeneration und Passivierung mit einzelnen Prozessschritten in Verbindung bringt.

Neben den oben erwähnten Prozesseinflüssen ist auch die Physik hinter den Schädigungsmechanismen von Gateoxid und Grenzfläche noch lange nicht verstanden. Einer der wichtigsten Degradationsmechanismen ist die sogenannte ‘Negative Bias Temperature Instability’ (NBTI), die bei Belastung des Gateoxid mit hoher Temperatur und Spannung zeitabhängige Einsatzspannungdriften verursacht. NBTI ist in Hinblick auf den back-end Prozess von besonderem Interesse, da angenommen wird, dass die Degradation in unmittelbaren Zusammenhang zur Wasserstoffpassivierung der $Si–SiO_2$ Grenzfläche steht.

Ein Ziel dieser Dissertation ist es, die Rolle des Wasserstoffs bei der Bauelement-Degradation (NBTI) zu untersuchen und die Erkenntnisse anhand von Prozesssplits zu verifizieren. Ein spezielles Augenmerk liegt dabei auf Hochvolt (HV) MOSFET Bauelementen mit 30 nm $SiO_2$ Gateoxiden, wie sie bei Infineon Villach gefertigt werden. Bisweilen konzentrierte sich die Literatur vor allem auf ultradünne (nitridierte oder high–$\kappa$) Oxide, wohingegen Power Technologien bislang kaum untersucht wurden.

Zu Beginn werden grundlegende Signaturen der 'Bias Temperature Instability' (BTI) mit herkömmlichen Messmethoden auf 30 nm $SiO_2$ Bauelementen zusammengetragen und mit Resultaten neuerer Dünnoxid high–$\kappa$ Technologien verglichen. Diese einleitende Studie beinhaltet Überlegungen zur Elektrostatik und Dynamik von Degradation und Recovery und dient dazu, fundamentale elektrische Eigenschaften der auftretenden Defekte zu erfassen. Ein Literaturüberblick über das gegenwärtige mikroskopische Bild verschiedener Arten von Punktdefekten und deren Wechselwirkung mit Wasserstoff soll in weiterer Folge das physikalische Verständnis vertiefen. Die gesammelten Erkenntnisse werden dann benutzt, um neue Messvorschriften und experimentelle Prozeduren zu entwickeln, mit denen man in der Lage ist, Einsatzspannungsdriften auf verschiedene Arten zu charakterisieren und die jeweiligen Anteile bestimmten Defektklassen mit speziellen Eigenschaften zuzuordnen. Anhand der Daten und Schlussfolgerungen wird ein mikroskopisches Degradationsmodell vorgeschlagen, das (zumindest qualitativ) in der Lage ist, das Beobachtete anschaulich zu erklären. Um die Brücke zum Wasserstoff zu schlagen, werden die neuen Messmethoden auf verschiedene Prozesssplits angewandt, welche nachweislich über unterschiedliche Wasserstoffhaushalte im Gateoxid verfügen. Die Resultate werden mit den Vorhersagen des vorgeschlagenen mikroskopischen Degradationsmodells verglichen, Übereinstimmungen werden aufgezeigt und Schlussfolgerungen über den Prozesseinfluss werden gezogen.

Das aktuelle mikroskopische Model beschreibt die Bauelement-Degradation unter NBTI als einen Zweistufenprozess, dessen Ausgangsdefekt eine Sauerstofffehlstelle (oxygen vacancy) im amorphen $SiO_2$ Gateoxid ist. Unter thermischen und elektrischen Stress (NBTI) kann die verspannte Fehlstelle aufbrechen, wodurch ein positiv geladener Defekt im Oxid entsteht. Dieser Defekt ist unabhängig von der Wasserstoffpassivierung, kann umgeladen werden und wieder ausheilen, vorausgesetzt, er wurde zuvor elektrisch neutralisiert. Die zweite Defektklasse besteht aus Grenzflächenzuständen (dangling bonds), die durch feld- und temperaturunterstütztes Ablösen von Wasserstoff erzeugt werden. Untersuchungen deuten darauf hin, dass sich der abgelöste Wasserstoff mit einer aufgebrochenen Sauerstofffehlstelle verbinden kann, wodurch die Ausheilung des Oxiddefekts blockiert wird (locked-in oxide defect). Diese dadurch entstehenden quasi-permanenten Defekte spielen eine wichtige Rolle für die NBTI-Stabilität einer Technologie, da sie nur langsam und nur unter bestimmten Voraussetzungen ausheilen und wegen ihres Wasserstoffbezugs stark vom back-end Prozess abhängen.

# Abstract

DEVICE degradation and gate oxide reliability has become more and more important in modern semiconductor industry because the challenges for devices increase with miniaturization, larger stress within the circuit, and due to innovative technologies. Besides the reliability of mechanical soldering and wiring, it is of considerable importance to protect the transistor already in the early stages of processing from environmental impurities and mechanical stress. To furthermore guarantee the switching of high currents with minimum power dissipation as well as short-circuit stability, different isolation and metalization layers are deposited in a specific composition and sequence on the active areas of the transistor at the end of device fabrication (back-end of line). It is an accepted fact that different passivation concepts and layer sequences affect the electric characteristics and the reliability of silicon devices, because a lot of hydrogen is incorporated into the system during the back-end process. It is assumed that high temperatures going along with layer deposition and annealing support hydrogen diffusion toward the gate oxide where it may passivate defects within the oxide and at the interface. When subjecting the device afterwards to electrical and thermal stress, the hydrogen may be released again, thereby re-activating previously passivated defects.

Just like any semiconductor manufacturer worldwide, Infineon Villach has repeatedly encountered the problem in the past, and although particular solutions to some technologies have already been found, a global understanding of the underlying physics behind the interplay of processing and degradation is lacking. Because it is likely that new problems with the same roots may become relevant in future technology development, it is of fundamental importance to develop a physically based model which is able to link defect generation and passivation to single process steps.

Besides the above mentioned process impacts, also the physics behind the degradation mechanisms of gate oxide and interface are far from being understood. One of the most important degradation mechanisms is the so-called 'Negative Bias Temperature Instability' (NBTI) which leads to time dependent threshold voltage shifts when the gate oxide of the device is stressed at an elevated temperature and bias. NBTI is particularly interesting for the back-end process since it is assumed that the degradation mechanism is closely connected to hydrogen passivation of the Si–$SiO_2$ interface.

One aim of this PhD thesis is to investigate the role of hydrogen in NBTI and to verify the results by means of process splits. Specific emphasis is put on high voltage (HV) MOSFET devices having 30 nm $SiO_2$ gate oxides, as processed at Infineon Villach. Until now the scientific community has mainly been focusing on ultrathin (nitrided or high–$\kappa$) oxides whereas contributions on thicker oxides relevant for power technologies are hardly found in literature.

At first, basic signatures of the 'Bias Temperature Instability' (BTI) of 30 nm $SiO_2$ devices are collected using common measurement methods and the results are compared to state-of-the-art thinoxide high–$\kappa$ technologies. This preliminary study contains considerations on electrostatics and dynamics of degradation and recovery and helps to collect fundamental electric characteristics of emerging defects. A literature study on the current understanding of different species of point defects and their correlation to hydrogen is presented in order to enhance the physical understanding. The collected results are used to develop new measurements setups and experimental procedures which enable us to characterize threshold voltage shifts in various ways and assign the different contributions to certain defect classes with particular attributes. From the data and conclusions a microscopic degradation model is going to be suggested which is capable of explaining (at least qualitatively) the obtained results. In order to include hydrogen, the new measurement routines are performed on different split wafers which provide demonstrably different hydrogen budgets within the gate oxide. The results are cross-checked with the predictions of the suggested microscopic degradation model, agreements are highlighted and conclusions are drawn on the process impact.

The current microscopic model explains device degradation under NBTI as a two stage process, the precursor being an oxygen vacancy defect in the amorphous $SiO_2$ gate oxide. By subjecting the device to thermal and electrical stress (NBTI) the bond may be broken, thereby creating a positively charged defect. This defect is independent of hydrogen, rechargeable and can be annealed, provided it has been neutralized in advance. The second defect class consists of interface states (dangling bonds) which are created due to field- and temperature-assisted release of hydrogen. Studies suggest that the released hydrogen may become trapped in the broken oxygen vacancy, thereby blocking the relaxation of the oxide defect. These quasi-permanent defects play an important role for the NBTI-stability of a technology since they recover just slowly and only under certain conditions, and due to their correlation to hydrogen, their concentration depends strongly on the back-end process.

# Contents

# 1 The Negative Bias Temperature Instability

## 1.1 Industrial Background and Motivation

ONCE a new device technology is released from product line, it has to overcome several qualification procedures before it finally gets approved for delivery to the customer. Within those qualification tests it is checked for example whether the new product is working regularly and whether it meets the proposed parameter specifications. Besides initial parameter testing, also the reliability of random samples is monitored meticulously during different stages of the fabrication (i.e. wafer level, package level). This is done in order to collect statistics about extrinsic defect densities causing early device failure and parameter drifts causing gradual degradation of the device within years under operating conditions. Furthermore, fundamental conclusions are drawn on process changes which go along with device shrinking and upgraded innovation standards. Companies invest a lot of money and man power in this area to develop, produce, monitor and evaluate reliability test chips on split wafers that are specifically designed for gaining experience in process influences helping to optimize device performance while simultaneously minimizing production costs. Hence, physically based models explaining process influences on device performance are highly demanded and of great economical interest.

During reliability testing every important device parameter is assigned to a certain drift margin which has to be satisfied and finally determines the lifetime of the product. Considering that the aimed lifetimes of most semiconductor products lie in the range of several tens of years whereas a single qualification test has to be completed within hours, it would be highly inefficient and much too time consuming (if not impossible) to perform lifetime prediction tests under use conditions. Thus, in order to guarantee within a reasonable testing time that an airbag chip will still work at the end of

the cars' lifetime, it is necessary to subject devices to extreme test conditions (accelerated tests). For the analysis of the data, elaborated acceleration models are required allowing to estimate the actual parameter drift under use conditions from a parameter shift measured under accelerated conditions. Thus, in order to make reliable lifetime predictions, physically based acceleration models are urgently needed for each degradation mechanism. So far most existing models are empirical.

One of the most popular degradation mechanism affecting the gate oxide and the interface of metal oxide semiconductor field effect transistors (MOSFETs) is the so-called negative bias temperature instability (NBTI). As opposed to destructive failure modes like time dependent dielectric breakdown (TDDB), NBTI does not cause hard failures (short circuits across the oxide) which limit the possibilities of post stress characterization dramatically but rather leads to a creeping shift in the threshold voltage (typically several tens of mV) and to a gradual degradation of the channel mobility and the transconductance of a conventional metal oxide semiconductor (MOS) transistor. The resulting variations in the analog transfer characteristics of the device can cause severe problems especially when the afflicted transistor is incorporated in a digital network or when the demanded specification margin of the technology is very narrow. Since NBTI is very sensitive to variations in device processing, it is particularly hard to control the NBTI resistivity of a technology during device development. The main focus of this PhD thesis is to broaden the knowledge around NBTI by performing stress/recovery experiments on specifically designed test chips. Furthermore, correlations of the effect with process influences, in particular with hydrogen introduced during the back-end of line (BEOL) fabrication, are studied extensively by means of wafer splits produced at Infineon Technologies Austria.

## 1.2 Historical Background and Relevance

The bias temperature instability (BTI) in MOS devices has been known since the middle of the 1960s [1–4] and has been attributed to defects located at the Si–$SiO_2$ interface since the beginning of the 1970s [5]. Since that time BTI has been a vigorously discussed topic in the semiconductor community and numerous papers focusing on measurement techniques, modeling attempts and technology impact circulate in various scientific journals and conference proceeding. The increasing importance of the topic may be illustrated schematically by searching via 'Google Scholar' for publications containing variants of the keywords 'bias temperature stress', 'silicon' and 'oxide' itemized by the past decades of years. In Fig. 1.1 the result of such a simple keyword search is depicted as 'number of Google Scholar Hits' graded by five year intervals since the 1960s.

Naturally, the exact number of hits will depend considerably on the selected keywords making their absolute values somewhat arbitrary. Nevertheless, the overall roughly exponentially increasing trend of BTI related publications is likely to be reproduced independently of the subjective choice of the keywords.

The first attempt to describe the BTI phenomenon by a micro-physical degradation model is due to Jeppson and Svensson in 1977 [6] who assumed the surface trap growth as being caused by hydrogen release from Si–H bonds located at the interface. The dynamics of degradation were assumed to be

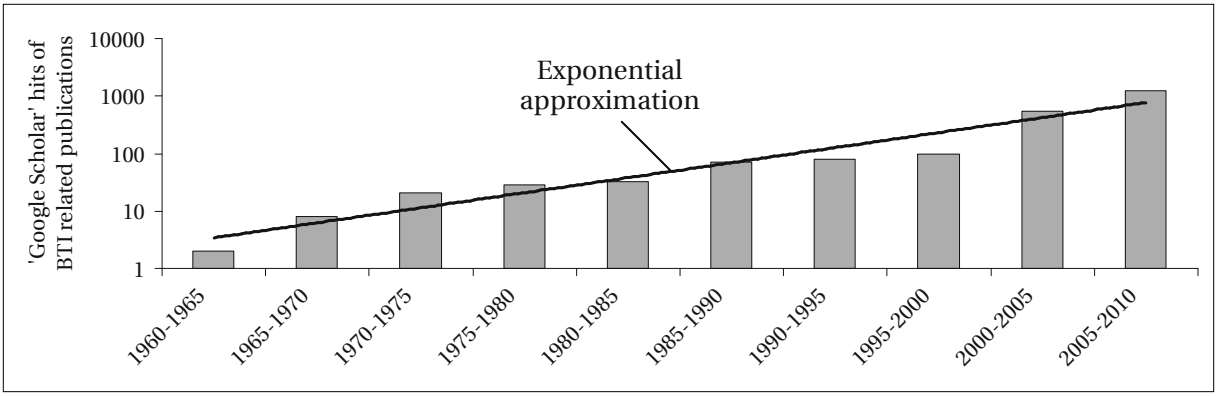


**Figure 1.1:** 'Google Scholar hits' of BTI related publications since the 1960s to the present date (2010). The number of hits (column bars) found by 'Google Scholar' increase roughly exponentially from the 1960s to 2010 (solid line). Within the past five years between 2005 and 2010 'Google Scholar' lists more than 1200 hits for the selected keywords.

diffusion controlled at low fields and tunneling limited at high fields. Since that time a lot of progress has been made to get the BTI under control. Although, numerous papers have been published on the topic, the issue is still far from being closed. There are a lot of open questions left which have to be resolved in order to understand discrepancies in experimental data and eliminate inconsistencies between suggested models. Since the BTI problem is going to become even more relevant in the future due to continuous processing and scaling changes, a universally valid model including process influences is urgently needed. Besides that a fundamental understanding of BTI may also lead to a better understanding of related degradation mechanisms like hot carrier injection (HCI) or TDDB.

## 1.3 Operating conditions causing bias temperature instability

When performing a bias temperature stress (BTS) on a MOS transistor, the device is usually heated to a defined stress temperature ($T_S$) and is then subjected to a relatively large electric field ($E_{OX}$) across the gate oxide (GOX) by applying a defined stress bias to the gate junction of the transistor. In general, the degradation characteristics following BTS depend strongly on the polarity of the applied gate bias during stress and on the doping type of the device under test (DUT). When applying for instance a *positive* gate bias during stress, one speaks of positive bias temperature instability (PBTI) [7], while degradation following a *negative* bias stress gives the classical NBTI [8–12]. The largest amount of damage is usually observed when subjecting a p-channel metal oxide semiconductor (PMOS) transistor to negative bias temperature stress (NBTS).

During NBTI stress, all transistor contacts except for the gate are grounded resulting in a relatively homogeneous stress field, hence generating uniformly distributed defects along the entire gate oxide area. When neglecting small variations of the doping profile along the transistor channel, in particular close to the source/drain implantations, the dimension of the problem is reduced by the symmetry of the stress profile from three to one, perpendicular to the gate oxide-substrate interface. The

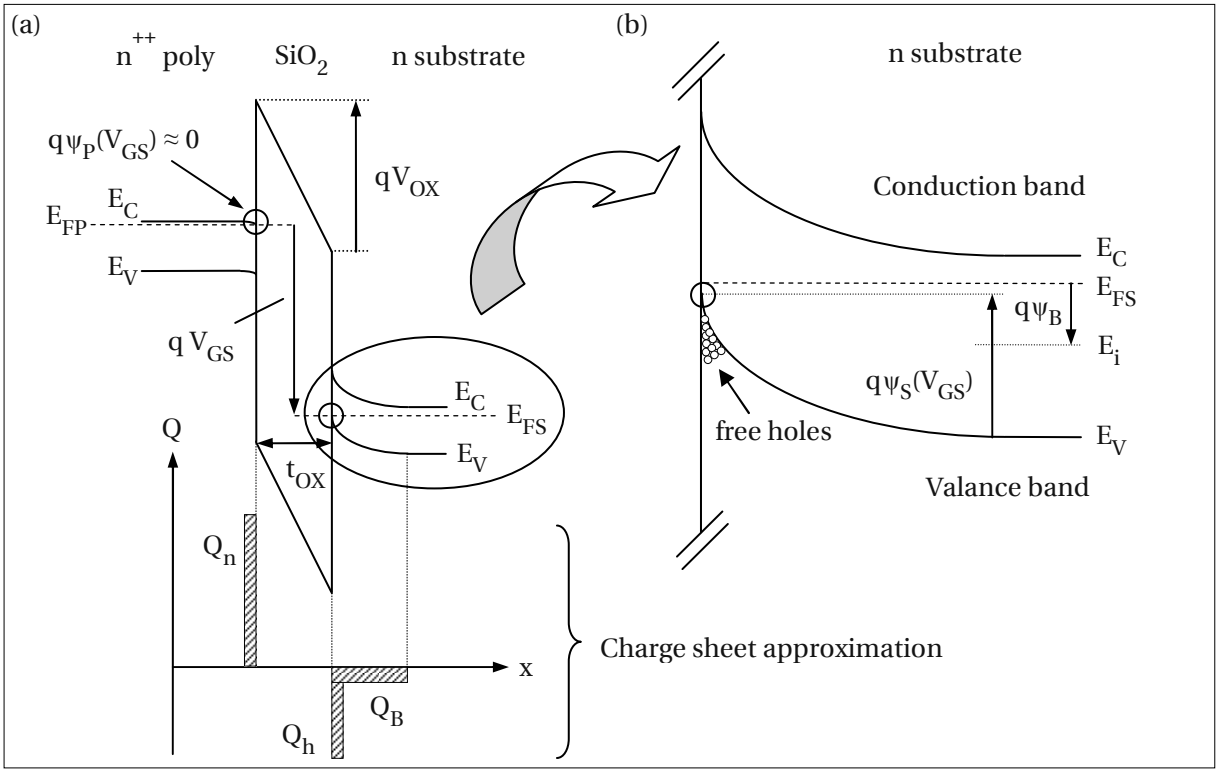


**Figure 1.2:** The band diagram and the charge balance of a PMOS transistor with an $n^{++}$ gate poly during NBTS (a). The interface and the depletion region within the Si substrate is enlarged in (b).

band banding situation during NBTS is schematically depicted in Fig. 1.2 (a) for a PMOS transistor equipped with an $n^{++}$ gate poly.

Due to the large negative gate voltage ($V_{GS} \ll V_{TH}$) applied during stress, the interface of the $n^{++}$ gate junction ($N_D \approx 10^{20}\,cm^{-3}$) becomes heavily accumulated with majority electrons ($Q_e$). Hence, the Fermi level ($E_{FP}$) remains pinned close to the conduction band edge at the polysilicon/silicon dioxide ($SiO_2$) interface, the band banding within the gate junction being approximately zero ($\psi_P \approx 0$). On the other side of the gate oxide the moderately n-doped silicon substrate ($N_D \approx 10^{16}\,cm^{-3}$) becomes heavily inverted attracting injected minority holes ($Q_h$) to the $SiO_2$/silicon (Si) interface. The depletion layer containing the bulk charge ($Q_B$) approaches its maximum extension increasing the surface potential ($\psi_S$), thereby bending the valence band edge at the $SiO_2$/Si interface toward the substrate Fermi level ($E_{FS}$). Following Fig. 1.2 (b), the band bending during stress (maximum surface potential $\psi_{S,max}$) may be estimated as

$$q\psi_{S,max} \approx E_i - E_V + q\psi_B, \tag{1.1}$$

where $E_i$ is the intrinsic energy, $E_V$ is the valence band edge, q is the elementary charge and $\psi_B$ is the bulk potential ($\psi_B$) which is given by

$$\psi_B = \frac{k_B T}{q} \ln\left(\frac{N_D}{n_i}\right). \tag{1.2}$$

In Eq. 1.2, $k_B$ is the Boltzmann constant, $N_D$ is the donor doping density of the substrate material and $n_i$ is the intrinsic carrier concentration. When performing a full voltage loop along the band edges in Fig. 1.2 (a) one obtains

$$V_{GS} = -V_{OX} - \psi_{S,max} - \psi_P - V_{FB}. \tag{1.3}$$

In Eq. 1.3, $V_{OX}$ is the voltage drop across the gate oxide and $V_{FB}$ is the flat band voltage which may be approximated as the difference in the work functions between the poly silicon gate junction and the silicon substrate assuming an ideal interface and no charges within the $SiO_2$ gate oxide:

$$V_{FB} \approx \Phi_P - \Phi_S \approx \frac{E_G}{q} - \psi_{S,max} - \psi_P. \tag{1.4}$$

By inserting Eq. 1.4 in Eq. 1.3, one finds

$$V_{GS} \approx -V_{OX} - \frac{E_G}{q}. \tag{1.5}$$

From Eq. 1.5 a simple approximation for the electric field $E_{OX}$ is obtained when stressing a PMOS transistor equipped with a $n^{++}$ gate poly under NBTS:

$$E_{OX}^{n++} = -\frac{V_{OX}}{t_{OX}} \approx \frac{V_{GS} + 1.1\,V}{t_{OX}}, \tag{1.6}$$

1.1 V being the silicon bandgap $E_G$ at a typical stress temperature of 125°C. A similar deviation for a PMOS transistor equipped with a $p^{++}$ gate poly yields the following approximation for the electric field during NBTS [13]:

$$E_{OX}^{p++} = -\frac{V_{OX}}{t_{OX}} \approx \frac{V_{GS} + |V_{Poly}|}{t_{OX}}, \tag{1.7}$$

where $V_{Poly}$ stands for the voltage drop within the gate poly junction (poly depletion) which is typically very small [14] due to the high poly doping density ($V_{Poly} \approx 0.1\,V$ for $N_A \approx 10^{20}\,cm^{-3}$).

When comparing Eq. 1.6 and Eq. 1.7, we come to the conclusion that a PMOS transistor equipped with a $n^{++}$ gate poly has to be stressed under a larger gate voltage of approximately 1.0 V compared to a PMOS transistor equipped with a $p^{++}$ gate poly in order to generate similar electric fields during NBTI stress. The same considerations hold for an n-channel metal oxide semiconductor (NMOS) transistor as well, the holes at the $SiO_2$/Si interface being then supplied by the p-doped silicon substrate.

Typical stress fields applied during NBTI are between 2.5 MV/cm and 8.0 MV/cm [15]. The magnitude of degradation caused by fields below 2.5 MV/cm is usually very small and hence often below the detection limit at typical stress temperatures (50°C – 200°C) within typical stress times (1 ms – 100,000 s). Fields above 8 MV/cm introduce different degradation mechanisms like impact ionization by 'hot' carriers which tunnel through the gate oxide causing intrinsic TDDB due to the gradual formation of a conductive percolation path across the oxide [16–18]. When stressing thick oxide devices, the effect emerges earlier since the voltage drop across the gate oxide ($V_{OX}$) is larger at the same $E_{OX}$.

# 2

# Electrostatics of NBTI degradation

THIS chapter addresses origins of NBTI induced drain current degradation in the linear and in the saturation operation regime of the transistor. The gate bias dependence of the threshold voltage shift is discussed theoretically and by means of a case study without taking time dependent degradation and recovery dynamics into account. An electrostatic treatment of the problem is basically justified when the time delay between stress and measurement is large enough so that defect relaxation during the measurement may be neglected.

Having identified the main causes of device degradation under DC bias conditions, the second part of this chapter introduces an experimental approach based on high frequency AC gate pulsing, allowing to separate charged defect classes which normally remain indistinguishable in a conventional DC experiment. This is because in DC experiments any kind of charged defect affects the threshold voltage shift in a similar way independent of its microscopic nature. Due to this superposition, it is not straight-forward to assign a certain contribution of the $V_{\mathrm{TH}}$ shift to a particular defect.

## 2.1 Effects of NBTI on analogue transfer characteristics

When stressing a MOS device under NBTS, we generally observe a variation in the drain current ($\Delta I_{\mathrm{D}}$) when measuring the same operating point (under the same bias conditions) again after stress. The linear and the saturation drain current of a long-channel PMOS transistor (such as mostly considered

here) above the threshold voltage of the device ($|V_{GS}| \geq |V_{TH}|$) can be approximated by the following equations [19]:

$$I_{D,lin}(V_{DS} < V_{DSS}) \approx -\frac{W}{L}\mu_{eff}C_{OX}\left((V_{GS} - V_{TH}) - \frac{1}{2}V_{DS}\right)V_{DS}, \tag{2.1}$$

$$I_{D,sat}(V_{DS} \geq V_{DSS}) \approx -\frac{1}{2}\frac{W}{L}\mu_{eff}C_{OX}\,(V_{GS} - V_{TH})\,V_{DSS}, \tag{2.2}$$

where the saturation drain voltage ($V_{DSS}$) is approximately given by

$$V_{DSS} \approx (V_{GS} - V_{TH}) \tag{2.3}$$

and the physically defined threshold voltage can be derived as

$$V_{TH} \approx V_{FB} - 2\psi_B - \frac{\sqrt{2\epsilon_{OX}qN_D(2\psi_B)}}{C_{OX}}. \tag{2.4}$$

In Eq. 2.1 and Eq. 2.2, the two unstable device parameters that may cause a degradation in the drain current during NBTS are the effective inversion carrier mobility ($\mu_{eff}$) and the threshold voltage ($V_{TH}$) of the device. The relative variation of the drain current as a function of the overdrive can be obtained by calculating the total derivative of Eq. 2.1 and Eq. 2.2:

$$\frac{\delta I_{D,lin}}{I_{D,lin}} \approx \frac{\delta\mu_{eff}}{\mu_{eff}} - \frac{\delta V_{TH}}{V_{GS} - V_{TH}}, \tag{2.5}$$

$$\frac{\delta I_{D,sat}}{I_{D,sat}} \approx \frac{\delta\mu_{eff}}{\mu_{eff}} - 2\frac{\delta V_{TH}}{V_{GS} - V_{TH}}. \tag{2.6}$$

Eq. 2.5 and Eq. 2.6 demonstrate that the total degradation of the drain current is directly proportional to the change of the effective carrier mobility and to the change of the threshold voltage shift divided by the gate bias overdrive ($V_{GS} - V_{TH}$). In the saturation region, the threshold voltage degradation has a larger influence on the drain current degradation than in the linear region making the signal vs. noise resolution as a function of the $V_{TH}$ shift about two times better when measuring in the saturation region of the device. From Eq. 2.5 and Eq. 2.6 we further conclude that the measurement resolution of the drain current degradation is expected to decrease linearly, when increasing the gate voltage overdrive ($|V_{GS}| \gg |V_{TH}|$).

Upon the build up of microscopic defects during NBTS, mobility and threshold voltage have been reported to degrade simultaneously which makes it difficult to differentiate between both effects. Once created, charged traps can act on the one hand as defect charges counterbalancing the applied gate potential ($\Delta V_{TH}^{q}$) and on the other hand as Coulomb scattering centers decreasing the effective carrier mobility ($\Delta\mu_{eff}$) and hence the on-resistance ($R_{ON}$) of the device. In general, charged defects are particularly effective scattering centers when they are located close to or directly at the $SiO_2$/Si interface. Following [20–22], the effective carrier mobility ($\mu_{eff}$) can be empirically approximated as

$$\mu_{eff} = \frac{\mu_0}{(1 + \alpha N_{it})(1 + \Theta(V_{GS} - V_{TH}))}. \tag{2.7}$$

In Eq. 2.7 $\mu_0$ is the temperature and doping dependent low field mobility and $N_{it}$ is the number of interface states per square centimeter. The first term of the denominator in Eq. 2.7 accounts for a reduction of this low field mobility as a consequence of enhanced Coulomb scattering at surface near charge centers ($N_{it}$) while the second term considers mobility degradation caused by enhanced surface scattering when applying a vertical electric field. The coefficients $\alpha$ and $\Theta$ are empirical fit parameters which have to be determined experimentally from the stress induced shift in the transconductance ($\alpha$) and from the curvature of the virgin transfer curve ($\Theta$). From a physical point of view, following [22] and [23], the parameter $\alpha$ is a scattering coefficient (capture cross section) that accounts for the interaction between charged interface states and inversion layer carriers. It has the unit [cm$^2$] and may take values between $10^{-12}$ cm$^2$ (depletion) and $10^{-13}$ cm$^2$ (inversion), the lower limit being smaller due to carrier screening in deep inversion [24–26]. The parameter $\Theta$ is the vertical field mobility coefficient [19] leading to a bending of the transfer curve in the triode region of the device. It has the unit [V$^{-1}$] and considers the reduction of the mobility as a consequence of the growing electric field perpendicular to the moving direction of the inversion charge carriers.

Differentiating Eq. 2.7 with respect to the interface state density ($N_{it}$) yields

$$\frac{\partial \mu_{eff}}{\partial N_{it}} = -\frac{\alpha}{1+\alpha N_{it}}\mu_{eff} + \frac{\Theta}{1+\Theta(V_{GS}-V_{TH})}\frac{\partial V_{TH}}{\partial N_{it}}\mu_{eff}, \tag{2.8}$$

$$\frac{\Delta \mu_{eff}}{\mu_{eff}} = -\frac{\alpha \Delta N_{it}}{1+\alpha N_{it}} + \frac{\Theta \Delta V_{TH}}{1+\Theta(V_{GS}-V_{TH})}. \tag{2.9}$$

Inserting Eq. 2.9 into Eq. 2.5 and Eq. 2.6, we may replace the variation in the effective mobility ($\mu_{eff}$) by empirical constants:

$$\frac{\Delta I_{D,lin}}{I_{D,lin}} \approx -\frac{\alpha \Delta N_{it}}{1+\alpha N_{it}} - \left(\frac{1}{(V_{GS}-V_{TH})(1+\Theta(V_{GS}-V_{TH}))}\right)\Delta V_{TH}^{q}, \tag{2.10}$$

$$\frac{\Delta I_{D,sat}}{I_{D,sat}} \approx -\frac{\alpha \Delta N_{it}}{1+\alpha N_{it}} - \left(\frac{2+\Theta(V_{GS}-V_{TH})}{(V_{GS}-V_{TH})(1+\Theta(V_{GS}-V_{TH}))}\right)\Delta V_{TH}^{q}. \tag{2.11}$$

Usually, when analyzing NBTI, the degradation of the drain current ($\Delta I_D$) is converted into a corresponding total threshold voltage shift ($\Delta V_{TH}$) by assigning the drain current measured post stress a corresponding gate voltage of the virgin transfer curve [27]. Following Fig. 2.1, the difference between the actually applied gate voltage during the drain current measurement and the hypothetical gate voltage corresponding to the same drain current recorded before stress on the virgin device gives the effective $V_{TH}$ shift ($\Delta V_{TH}$) describing a horizontal (voltage) shift of the transfer curve that is usually a function of the current gate bias due to the curvature of the transfer curve ($\Theta$) and due to mobility degradation caused by defect scattering ($\alpha$). Although the notation $\Delta V_{TH}$ is somehow related to the physically defined device threshold voltage (cf. Eq. 2.4), it is also common for shifts measured at arbitrary gate voltages.

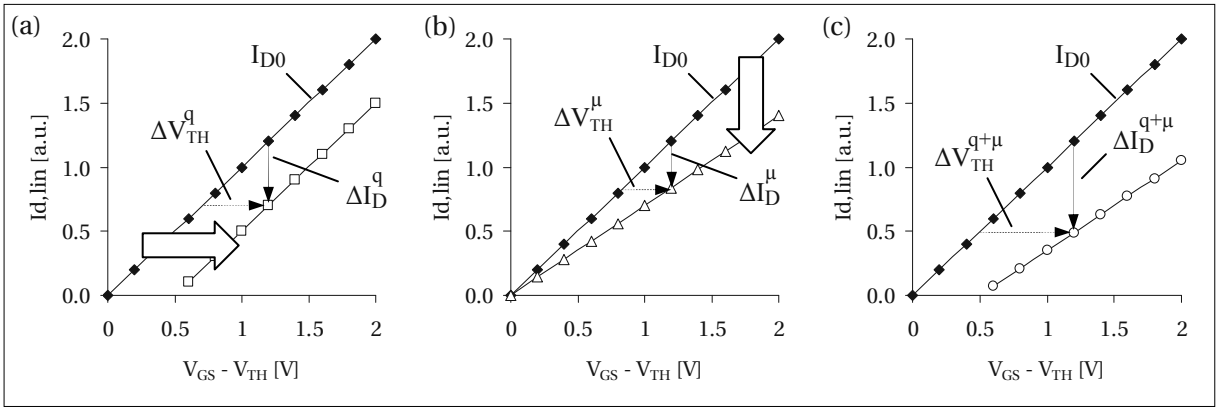


**Figure 2.1:** Schematic illustration of the individual components contributing to the threshold voltage shift depicted as a function of the gate bias overdrive. (a) Threshold voltage shift caused by a variation of the surface potential due to electrically active defect charges ($\Delta V_{TH}^{q}$); (b) Spurious $V_{TH}$ shift caused by a reduction in the channel transconductance due to mobility degradation ($\Delta V_{TH}^{\mu}$); (c) resulting total $V_{TH}$ shift caused by both components: mobility degradation and defect charges

### 2.1.1 Influence of mobility degradation and defect charges on the total $V_{TH}$ shift

As mobility and threshold voltage degrade upon the creation of various defects at the $SiO_2$/Si interface and within the gate oxide during stress, the total $V_{TH}$ shift calculated from $\Delta I_D$ is affected by both components simultaneously, however, their relative contribution may change as a function of the gate voltage overdrive.

Defect charges modify the surface potential directly, thereby shifting the transfer curve with respect to the gate voltage axis (x-axis) ($\Delta V_{TH}^{q}$). As a consequence, the drain current is degraded ($\Delta I_D^{q}$) when measuring the same operating point again after stress. On the other hand, mobility degradation reduces the drain current directly due to a change in the channel resistance thereby shifting the transfer curve with respect to the drain current axis (y-axis). When converting the parasitic current reduction ($\Delta I_D^{\mu}$) caused by the degradation of the channel resistance into an equivalent threshold voltage shift, it induces an extra $V_{TH}$ component ($\Delta V_{TH}^{\mu}$) superimposed to the 'real' $V_{TH}$ shift caused by defect charges ($\Delta V_{TH}^{q}$). The issue is discussed schematically in Fig. 2.1 where the influence of both components is illustrated separately and combined in the deep inversion regime of the transistor ($|V_{GS}| \geq |V_{TH}|$).

The drain current values were calculated directly form Eq. 2.1 (neglecting for simplicity the curvature of the transfer curve ($\Theta$) and the inversion carrier screening effect) and are plotted as a function of the gate bias overdrive. In Fig. 2.1 (a) – (c), the virgin drain current (labeled $I_{D0}$; full diamonds) is depicted as a reference. The individually degraded drain currents are illustrated by open symbols. Fig. 2.1 (a) shows the defect charge induced $V_{TH}$ shift component ($\Delta V_{TH}^{q}$) as a function of the gate voltage overdrive. According to the considerably simplified drain current model described by Eq. 2.1, the stress induced shift of the transfer curve (in the triode region of the device) caused by defect charges is parallel making the reduction of the drain current ($\Delta I_D^{q}$) independent of the gate voltage overdrive. Fig. 2.1 (b) illustrates the parasitic $V_{TH}$ shift of the transfer curve ($\Delta V_{TH}^{\mu}$) caused by a hypothetical 30 % reduction of the channel mobility. Since mobility degradation generally implies a change in the

channel transconductance (slope of the transfer curve), the reduction of the drain current and hence the spurious $V_{TH}$ shift ($\Delta V_{TH}^{\mu}$) caused by surface state scattering increases linearily with the overdrive. A combination of both effects is illustrated in Fig. 2.1 (c), demostrating that the total $V_{TH}$ shift ($\Delta V_{TH}$) gained from the conversion of the total drain current degradation ($\Delta I_D$) is generally a combination of a *parallel* shift due to the build up of defect charges and a *vertical* shift due to a change in the transconductance caused by enhanced scattering at charged defect centers:

$$\Delta V_{TH} = \Delta V_{TH}^{q} + \Delta V_{TH}^{\mu}. \tag{2.12}$$

The development of the total $V_{TH}$ shift with the gate voltage overdrive can be estimated from Eq. 2.10 and Eq. 2.11 by considering the curvature of the transfer curve ($\Theta$) and the scattering at charged defect centers located at the interface:

$$\Delta V_{TH,lin} = (V_{GS} - V_{TH})(1 + \Theta(V_{GS} - V_{TH}))\frac{\alpha \Delta N_{it}}{1 + \alpha N_{it}} + \Delta V_{TH}^{q}, \tag{2.13}$$

$$\Delta V_{TH,sat} = \frac{(V_{GS} - V_{TH})(1 + \Theta(V_{GS} - V_{TH}))}{2 + \Theta(V_{GS} - V_{TH})}\frac{\alpha \Delta N_{it}}{1 + \alpha N_{it}} + \Delta V_{TH}^{q}. \tag{2.14}$$

### 2.1.2 Spurious $V_{TH}$ shift due to mobility degradation ($\Delta V_{TH}^{\mu}$)

The first terms in Eq. 2.13 and Eq. 2.14 account for the spurious $V_{TH}$ shifts in the linear and the saturation region of the device caused by mobility degradation. In general, the interface scattering parameter $\alpha$ is not constant but depends on the gate bias overdrive due to the carrier screening effect [24–26]. As the density of inversion carriers increases when raising the gate voltage, they may crowd around charged point defects, thereby screening their Coulomb potential and reducing their scattering cross section. Experimental and theoretical investigations by [25, 26] revealed that the parameter $\alpha$ decreases proportionally to the square root of the density of inversion layer carriers:

$$\alpha \propto \frac{1}{\sqrt{N_{inv}}} \Rightarrow \alpha \approx \frac{\alpha_0}{\sqrt{|V_{OX}|}} \approx \frac{\alpha_0}{\sqrt{|V_{GS} - \psi_S - \psi_P - V_{FB}|}} \approx \frac{\alpha_0}{\sqrt{|V_{GS} - V_{TH}|}}. \tag{2.15}$$

Considering the inversion carrier screening effect, the spurious $V_{TH}$ shift in the linear and in the inversion regime of the device may be approximated as follows:

$$\Delta V_{TH,lin}^{\mu} \approx -|V_{GS} - V_{TH}|^{1/2}(1 + \Theta(V_{GS} - V_{TH}))\frac{\alpha_0 \Delta N_{it}}{1 + \alpha_0 N_{it}}, \tag{2.16}$$

$$\Delta V_{TH,sat}^{\mu} \approx -\frac{|V_{GS} - V_{TH}|^{1/2}(1 + \Theta(V_{GS} - V_{TH}))}{2 + \Theta(V_{GS} - V_{TH})}\frac{\alpha_0 \Delta N_{it}}{1 + \alpha_0 N_{it}}. \tag{2.17}$$

The result demonstrates that the mobility induced shift of the threshold voltage ($\Delta V_{TH}^{\mu}$) is proportional to the square root of the gate bias overdrive, the curvature of the transfer curve ($\Theta$) and the increase of the interface defect density ($\Delta N_{it}$). As already pointed out previously, the relative contribution $\Delta V_{TH}^{\mu}$ is found to be increasingly smaller (more than a factor 2) when measuring in the saturation region of the device.

In order to estimate $\Delta V_{TH}^{\mu}$ experimentally, we may express the interface scattering term in Eq. 2.16 and Eq. 2.17 by the shift of the transconductance recorded after electrical stress.

The transconductances in the linear and in the saturation regime representing the slopes of the transfer curve can be calculated by differentiating Eq. 2.1 and Eq. 2.2 by the applied gate potential. The derivations are performed in the appendix (9.1):

$$g_{m,lin} \approx \frac{I_{D,lin}}{V_{GS}-V_{TH}} \left( \frac{1}{1+\Theta(V_{GS}-V_{TH})} + \frac{\alpha N_{it}}{2(1+\alpha N_{it})} \right), \tag{2.18}$$

$$g_{m,sat} \approx \frac{I_{D,sat}}{V_{GS}-V_{TH}} \left( \frac{2+\Theta(V_{GS}-V_{TH})}{1+\Theta(V_{GS}-V_{TH})} + \frac{\alpha N_{it}}{2(1+\alpha N_{it})} \right). \tag{2.19}$$

The first terms in Eq. 2.18 and Eq. 2.19 represent the reduction of the transconductance due to field dependent scattering while the second terms consider inversion carrier screening. In particular, ons finds that the relative variations of the transconductances in the linear and in the saturation regime are directly related to the interface scattering factors in Eq. 2.16 and Eq. 2.17. The derivations are performed in the appendix (9.2). In the linear regime, we may express the interface scattering factor by the relative change in the maximum transconductance:

$$\frac{\Delta g_{m,lin}^{max}}{g_{m,lin}^{max}} \approx -\frac{\alpha_0 \Delta N_{it}}{1+\alpha_0 N_{it}}. \tag{2.20}$$

In the saturation regime, the deviation cannot be achieved in a straight-forward manner since the saturation transconductance has no maximum and therefore provides no characteristic operating point to refer on. In the vicinity of the threshold voltage, the variation of the transconductance may be approximated as follows:

$$\frac{\Delta g_{m,sat}^{V_{TH}}}{g_{m,sat}^{V_{TH}}} \approx -\frac{\alpha_0 \Delta N_{it}}{1+\alpha_0 N_{it}} - \frac{\Delta V_{TH}^{q}}{V_{GS}-V_{TH}}. \tag{2.21}$$

For gate voltages far away from the threshold voltage of the device, the equations proposed above are not valid anymore and may over- or underestimate the slope of the mobility induced threshold voltage shift. This is because the development of the transconductance is expected to change considerably due to field dependent scattering and inversion layer screening. By substituting the interface scattering factors in Eq. 2.16 and Eq. 2.17 by the transconductance related expressions derived in Eq. 2.20 and Eq. 2.21, we may express the spurious $V_{TH}$ shifts in the linear in in the saturation regime by experimentally measurable quantities.

In the linear regime, the mobility induced $\Delta V_{TH}^{\mu}$ shift yields

$$\Delta V_{TH,lin}^{\mu} \approx |V_{GS}-V_{TH}|^{1/2} \left(1+\Theta(V_{GS}-V_{TH})\right) \left( \frac{\Delta g_{m,lin}^{max}}{g_{m,lin}^{max}} \right). \tag{2.22}$$

In the saturation regime, the $\Delta V_{TH}^{\mu}$ shift may be approximated as

$$\Delta V_{TH,sat}^{\mu} \approx \frac{|V_{GS}-V_{TH}|^{1/2}\,(1+\Theta(V_{GS}-V_{TH}))}{2+\Theta(V_{GS}-V_{TH})}\left(\frac{\Delta g_{m,sat}^{V_{TH}}}{g_{m,sat}^{V_{TH}}}+\frac{\Delta V_{TH}^{q}}{V_{GS}-V_{TH}}\right), \tag{2.23}$$

demonstrating that $\Delta V_{TH}^{\mu}$ grows proportional to the relative change of the transconductance which is originated in enhanced Coulomb scattering at stress induced interface charges. As a consequence of inversion carrier screening and field dependent scattering, the mobility induced $V_{TH}$ shift does not actually increase linearly with the gate voltage overdrive as suggested in the simplest approximation illustrated in Fig. 2.1.

### 2.1.3 $V_{TH}$ shift due to defect charges ($\Delta V_{TH}^{q}$)

Once a defect is created during NBTS, it may exchange carriers with the silicon substrate thereby being charged either positively or negatively depending on the particular energy level of the defect and on the carrier situation at the $SiO_2$/Si interface. Statistical thermodynamics stipulates that at absolute zero the quantum energy of the highest occupied quantum state in a system of fermions is defined by the Fermi energy. At finite temperatures the concept of the Fermi energy is replaced by the Fermi level (electro-chemical potential) since the separation between occupied and unoccupied states is not infinitely sharp anymore but has a half-width of approximately $2k_BT$. In the operating temperature range of silicon devices this broadening is small (≈0.08 eV at 200°C) so that we may generally consider defects above the Fermi level as unoccupied while traps below the Fermi level are considered as occupied.

The net charge state of a defect level generally depends on the number of electrons trapped. In principle, every defect can be amphoteric which means that it has a donor- and an acceptor-level, the first describing the defect in the *positive* charge state, where it may become neutralized upon the capture of one electron, the latter describing the defect in the *neutral* charge state, where it may become negatively charged upon the capture of an additional electron. Capture of an electron is naturally accompanied by a level shift because a higher energetic state is required. The amount of level shift depends on the local bonding environment of the defect within the solid and decides whether a defect state actually appears only in one or in in both (donor and acceptor) configurations during a particular experiment. Due to localized inhomogeneities in the lattice stress involving slight variations in the individual bonding strengths, defects close to the interface usually do not have a single energy level but a distribution of states described by a density of states (DOS) profile. The boundary between the highest donor-like state and the lowest acceptor-like state in a network of defects of the same type is called the amphoteric transition level ($E_{amph}$). In thermal equilibrium, the total defect charge may be approximated as

$$Q_{defect} = q\int_{E_F}^{E_{amph}} D_{defect}(E)\,dE = \pm q\overline{D_{defect}}\Delta E_Q, \tag{2.24}$$

where $\Delta E_Q$ is the Fermi level dependent energy range of charged defects in thermal equilibrium:

$$\Delta E_Q = \left| E_{amph} - E_F \right|. \tag{2.25}$$

The integral in Eq. 2.24 is positive (resulting in a net positive defect charge) when the Fermi level is below $E_{amph}$, whereas the net defect charge is negative when the Fermi level lies above $E_{amph}$. It has to be mentioned that a certain time is required to restore thermal equilibrium between all kinds of defect species and the silicon substrate. The trap level only gives the equilibrium occupancy but does not say anything about the time constant. In particular, deep traps located in the bulk of the oxide having energy levels far away from the silicon bandgap, provide large carrier exchange time constants and small capture cross sections so that it may take a long time for them to align with the Fermi level. In particular, when taking inelastic carrier trapping/detrapping into account, some defects located energetically above the Fermi level may also have a finite chance to exchange carriers with the silicon substrate leading to a large variety of time constants. The dynamics of trap creation and annealing will be discussed in detail in Chapter 3.

As opposed to deep traps in the bulk of the gate oxide (slow states), defects being located energetically within the silicon bandgap and spatially close to the $SiO_2$/Si interface, exchange carriers very fast with the silicon substrate allowing to restore thermal equilibrium within a very short interval of time. Hence, the energy level of the highest occupied trap level may follow the development of the Fermi level (i.e. during a gate bias sweep) almost instantaneously. Interface states ($P_b$ centers) have such attributes, in particular, they provide a wide range of energy levels within the silicon bandgap having their amphoteric transition level around midgap ($E_{amph} \approx E_i$). Following Eq. 2.24, this implies that their net charge contribution is negative when the Fermi level is located in the upper half of the silicon bandgap, zero at midgap and positive when the Fermi level lies in the lower half of the silicon bandgap, cf. Fig. 2.2. In the subthreshold region of the device, there is an almost linear relationship between the Fermi level and the gate voltage allowing defects to trap and emit carriers very efficiently. Hence, during a gate bias ramp between accumulation and inversion interface states change their net charge contribution continuously as described by the standard Shockley Read Hall (SRH) model [28]. The gate bias dependent Fermi level depicted in Fig. 2.2 was simulated numerically for a PMOS device (SM6P/30/H1) at room temperature [29].

At 0.0 V, when the device is turned off, the Fermi level is pinned close to the conduction band edge $E_C$. Consequently, acceptor-like interface states located between midgap ($E_i$) and $E_F$ are occupied by two electrons resulting in a net negative defect charge (1). Approaching midgap (-0.5 V), previously negatively charged traps emit one of their two electrons resulting in an overall neutral interface at $E_F = E_i$ (2). As the gate bias drives the Fermi level deeper toward inversion, donor-like defect levels located between $E_i$ and $E_F$ emit an additional electron resulting in a net positive defect charge at the threshold voltage of the device (-1.0 V) (3). When driving the gate bias even deeper toward inversion (-2.0 V), the Fermi level position does not change significantly anymore. Hence, the net positive interface charge can be considered as virtually constant once the gate bias has exceeded the threshold voltage of the device.

Except for classical interface states (SRH), it has to be mentioned that there might exist also a specific type of oxide traps (E′ centers, cf. Section 4.3) located close to but not directly at the interface which may have similar carrier exchanging characteristics. However, as opposed to interface states, their charging-recharging time constants are probably larger since a thermodynamic barrier has to be overcome in order to communicate electrically with the silicon substrate.

While the energetic position of an individual trap type determines its equilibrium charge state, its spatial location within the gate oxide governs its relative impact on the observed threshold voltage shift ($\Delta V_{\mathrm{TH}}^{\mathrm{q}}$). On a PMOS device, one usually observes a negative $V_{\mathrm{TH}}$ shift after NBTS, indicating the creation of predominantly donor-like defects which become positively charged during stress and may keep their charge state for a certain time after termination of the stress. In the case of a PMOS transistor, termination of the stress means switching the gate bias from a negative stress level to a considerably lower but also negative threshold voltage. Consequently, when recording the degraded drain current after stress, most defects remain positively charged since the Fermi level remains pinned close to the valence band edge (cf. Fig. 2.2).

Assuming a spatial distribution of positive defects $\rho(x)$, with $x$ being the distance from the gate poly interface inside the $SiO_2$ gate oxide, the corresponding defect charge dependent $V_{\mathrm{TH}}$ shift ($\Delta V_{\mathrm{TH}}^{\mathrm{q}}$) is given by Gauss's law:

$$\Delta V_{\mathrm{TH}}^{\mathrm{q}} = -\frac{1}{C_{\mathrm{OX}}}\frac{1}{t_{\mathrm{OX}}}\int_0^{t_{\mathrm{OX}}} x\rho(x)\,\mathrm{d}x, \tag{2.26}$$

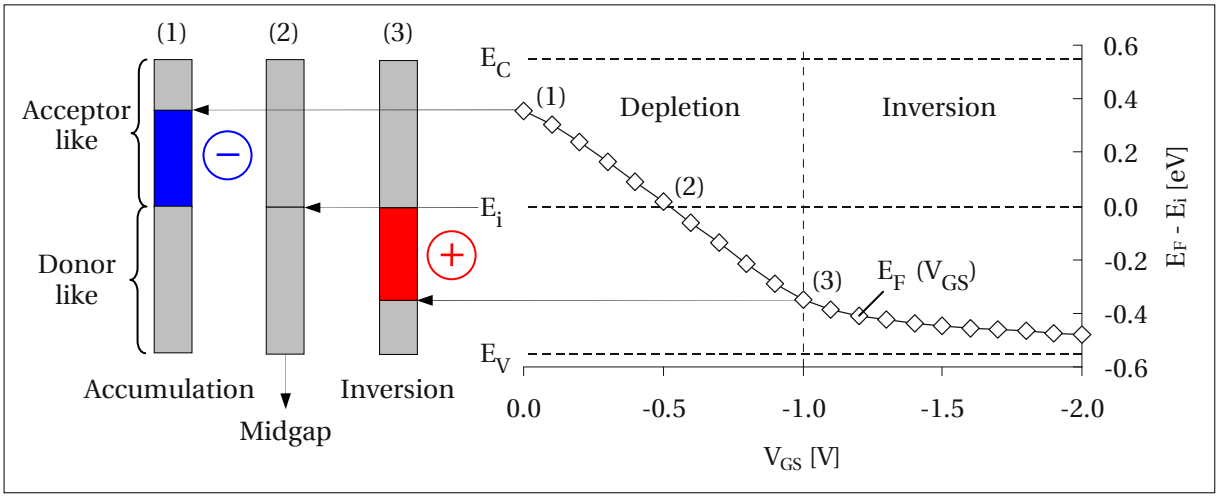


**Figure 2.2:** Dependence of the Fermi level on the gate bias in the subthreshold and inversion regime simulated numerically for a PMOS device (SM6P/30/H1) at room temperature. Within the depletion region of the device, their is a linear relationship between the gate bias and the Fermi level. In the inversion region, the Fermi level varies only slightly. On the left hand side of the figure the trap occupancy and the net interface trap charge is illustrated for three selected bias voltages. In accumulation (1) the net interface charge is negative, at midgap (2) zero and in inversion (3) positive.

where $C_{OX}$ is the area related oxide capacitance:

$$C_{OX} = \frac{\epsilon_0 \epsilon_{OX}}{t_{OX}} \approx \frac{3.45310^{-13}}{t_{OX}[\text{cm}]} [\text{F/cm}]. \tag{2.27}$$

Note that Eq. 2.27 is only accurate for thick oxide technologies ($t_{OX} > 10\,\text{nm}$), where the quantum-mechanical confinement of the inversion charge layer and the poly depletion may be neglected.

Assuming further all defects to be concentrated at the $SiO_2$/Si interface ($\rho(x) = q\Delta N_{it}\delta(x - t_{OX})$), the integral in Eq. 2.26 can be easily solved, giving the following simplified expression for the defect charge induced $V_{TH}$ shift:

$$\Delta V_{TH}^{q} = -\frac{q\Delta N_{it}}{C_{OX}} = -\frac{q\Delta N_{it}}{\epsilon_0 \epsilon_{OX}} t_{OX}. \tag{2.28}$$

The result demonstrates a linear correlation between the observed $\Delta V_{TH}$ shift and the oxide thickness. This is an important (albeit trivial) finding, in particular, when comparing $\Delta V_{TH}$ shifts of devices having different oxide thicknesses [15].

### 2.1.4 Relevance of defect charge induced $V_{TH}$ shift and mobility degradation

Having calculated in Eq. 2.28 the real threshold voltage shift caused by electrically active defect charges ($\Delta V_{TH}^{q}$) and in Eq. 2.16 and Eq. 2.17 the spurious threshold voltage shifts caused by field dependent mobility degradation ($\Delta V_{TH}^{\mu}$), we may estimate the relative contribution of $\Delta V_{TH}^{q}$ as a function of the gate voltage overdrive by assuming all charged defects to be located at the interface:

$$\frac{\Delta V_{TH}^{q}}{\Delta V_{TH,lin}} = \left(1 + |V_{GS} - V_{TH}|^{1/2} (1 + \Theta(V_{GS} - V_{TH})) \frac{\alpha_0}{1 + \alpha_0 N_{it}} \frac{C_{OX}}{q}\right)^{-1}, \tag{2.29}$$

$$\frac{\Delta V_{TH}^{q}}{\Delta V_{TH,sat}} = \left(1 + \frac{|V_{GS} - V_{TH}|^{1/2} (1 + \Theta(V_{GS} - V_{TH}))}{2 + \Theta(V_{GS} - V_{TH})} \frac{\alpha_0}{1 + \alpha_0 N_{it}} \frac{C_{OX}}{q}\right)^{-1}, \tag{2.30}$$

indicating that the $V_{TH}$ shift generated by defect charges is dominant when the capacitance is low (thick oxide devices), the overdrive is low and the overall interface state density is high. Considering a standard high voltage (HV) PMOS device with a 30 nm thick gate oxide, the curvature of the transfer curve ($\Theta$) being -0.13 $V^{-1}$, the interface density being $10^{10}\,cm^{-2}$, measured with an overdrive of -1.0 V ($\alpha_0 \approx 10^{-13}\,cm^2$ [24]), the relative contribution of $\Delta V_{TH}^{q}$ to the total $\Delta V_{TH}$ shift amount to approximately 92 % in the linear region and 96 % in the saturation region of the device which is quite large in comparison to the small remaining contribution of $\Delta V_{TH}^{\mu}$. In particular, in the presence of additional defect charge within the oxide which does not affect the mobility but increases the parallel $\Delta V_{TH}$ shift (i.e. charge centers located in the bulk of the oxide), the contribution of $\Delta V_{TH}^{\mu}$ may become almost negligible. However, as the oxide thickness shrinks or the overdrive increases (i.e. during on-the-fly (OTF) measurements, cf. Section 3.1) the mobility contribution can exceed up to 40 % of the total $\Delta V_{TH}$ shift [21] representing then an important component which has to be considered.

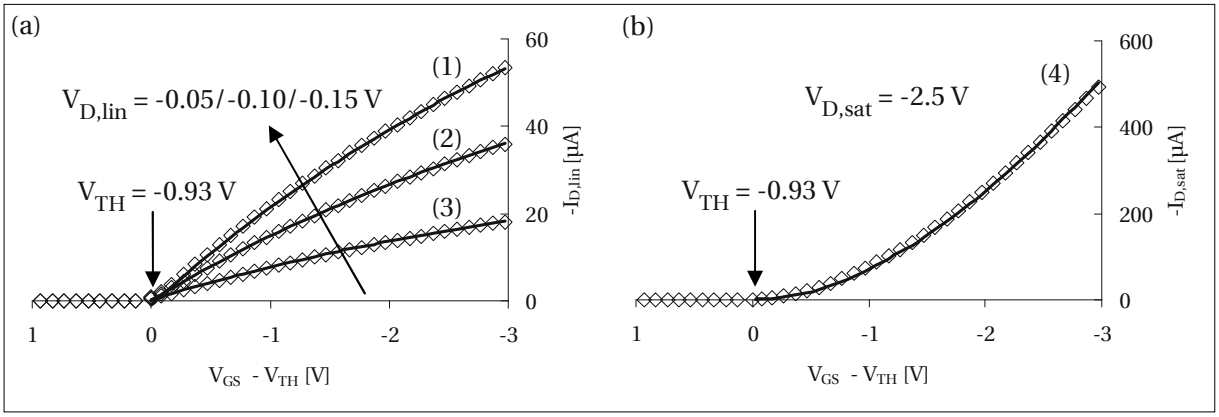


**Figure 2.3:** Virgin transfer curves measured at a temperature of 80°C in the linear (a) and in the saturation (b) drain current regime of a PMOS device (SM5P/30/H1) using drain biases of -0.05 V (1), -0.10 V (2), -0.15 V (3) and -2.5 V (4). The data points are depicted by open diamonds, the fits according to Eq. 2.1 and Eq. 2.2 are illustrated by thick solid lines. From parameter fitting a virgin threshold voltage of -0.93 V and a Θ of -0.13 $V^{-1}$ was extracted.

### 2.1.5 Case study

In order to demonstrate the impact of mobility degradation and defect charges on the drain current degradation and the $V_{TH}$ shift, a PMOS device (SM5P/30/H1) was stressed at a temperature of 80°C under an electric field of approximately -6.0 MV/cm for several thousand seconds. Before stress, virgin transfer curves were recorded between 0.0 V and -4.0 V (step size 0.01 V) using constant drain biases of -0.05 V (1), -0.10 V (2) and -0.15 V (3) in the linear regime ($I_{D,lin}$) and -2.5 V (4) in the saturation regime of the device ($I_{D,sat}$). The three transfer curves measured in the linear regime of the device are illustrated in Fig. 2.3 (a), the transfer curve recorded in the saturation regime of the device is depicted in Fig. 2.3 (b). From the measured data points (open symbols), the virgin threshold voltage $V_{TH}$ and the vertical field dependent scattering factor Θ is obtained by fitting the drain currents as a function of the gate bias overdrive according to Eq. 2.1 and Eq. 2.2. The fits are illustrated by thick solid lines yielding a threshold voltage of -0.93 V and a Θ of -0.13 $V^{-1}$.

Having determined the virgin threshold voltage and the curvature parameter Θ, we subject the device to NBTS. After stress, full transfer curves were recorded again serially in the linear and in the saturation regime of the device by ramping the gate voltage from accumulation toward inversion. It has to be mentioned that there were several seconds of delay following the termination of stress and the moment when the second set of transfer curves was finally recorded. Hence, some recovery has occurred during this floating period and during the measurement time itself. The recovery dynamics are going to be discussed in Chapter 3. In Fig. 2.4, the linear (a) and the saturation (b) drain currents recorded before (labeled '0') and after NBTI stress (labeled '1') at a constant drain bias of -0.1 V (linear regime) and -2.5 V (saturation regime) are illustrated as a function of the gate bias overdrive. The degraded transfer curves measured post stress are shifted slightly toward a more negative gate voltage indicating the build-up of positive charge.

Furthermore, the relative shift of the drain current is depicted as well in Fig. 2.4 (a) and (b) demonstrating the parameter degradation more clearly. As predicted by Eq. 2.10 and Eq. 2.11 ($|V_{GS}| > |V_{TH}|$), the relative variation of the drain current decreases gradually toward deeper inversion.

In Fig. 2.5, the transconductances in the linear (a) and in the saturation (b) regime recorded before (labeled '0') and after NBTI stress (labeled '1') are illustrated as a function of the gate bias overdrive. The thick solid line in (a) and (b) is a fit according to Eq. 2.18 and Eq. 2.19 which reflects the measured gradual decrease of the transconductance in the linear regime (a) and the convex curvature of the transconductance in the saturation regime (b) caused by vertical field dependent scattering ($\Theta$). Note that inversion layer screening has been neglected in the calculation of the fit. The relative degradation of the transconductance is also depicted in Fig. 2.5 (a) and (b), demonstrating mobility degradation and inversion carrier screening. Due to the superimposed parallel shift of the threshold voltage caused by defect charges, the relative variation of the transconductance is peaked for ($|V_{GS}| \mapsto |V_{TH}|$) and decreases in the deep inversion regime due to inversion layer screening. In the linear regime (a), the relative change of the transconductance at the gate voltage at which the transconductance reaches a maximum was found to be $6.2\times10^{-3}$. This is an important value since it can be used to determine the slope of the mobility induced $V_{TH}$ shift, cf. Eq. 2.22. In the saturation regime (b), a larger relative change of the transconductance is measured at the same gate voltage due to the parasitic influence of $\Delta V_{TH}^{q}$, cf. Eq. 2.23. When correcting this influence, a value of $8.0\times10^{-3}$ is obtained which is similar to the one extracted in the linear regime ($6.2\times10^{-3}$).

In the next step, the total $\Delta V_{TH}$ shift is calculated as a function of the gate bias overdrive from the drain current degradation. The results are illustrated in Fig. 2.6 (a) (linear regime) and (b) (saturation regime). In Fig. 2.6 (b) we have included one result measured in the linear regime ($V_{D,lin}$ = -0.1 V) as a reference. In both regimes a similar increase of the $\Delta V_{TH}$ shift is obtained below the threshold voltage of the device (I) which is assumed to be mainly caused by the gradual charging of the NBTI induced defects as the Fermi level moves from $E_C$ toward $E_V$ (cf. Eq. 2.24 and Fig. 2.2) [30–35].

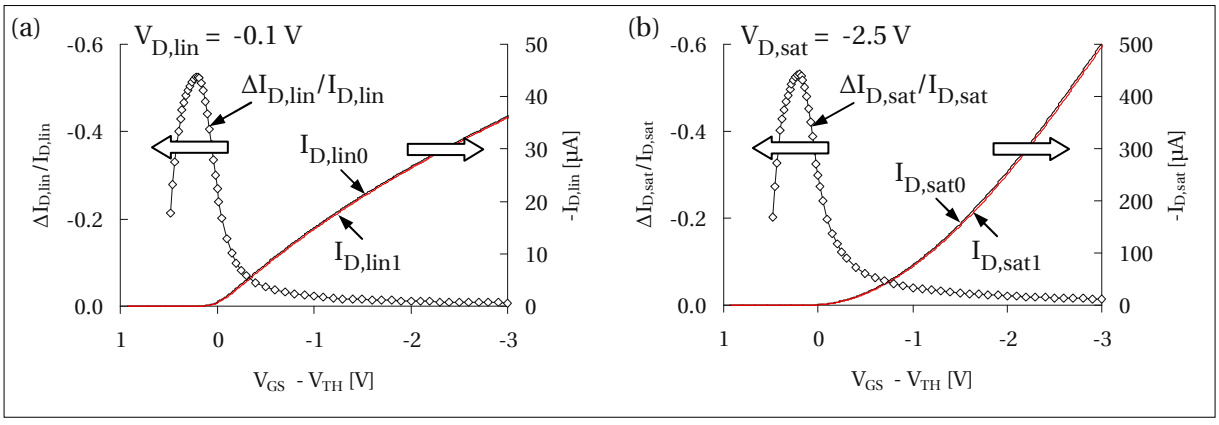


**Figure 2.4:** Drain currents (solid lines) and relative drain current shifts (open diamonds) recorded before and after stress in the linear ($V_{D,lin}$ = -0.1 V) (a) and in the saturation ($V_{D,sat}$ = -2.5 V) (b) regime of the device. The relative drain current variation is peaked close to the threshold voltage of the device and decreases then gradually toward deeper inversion.

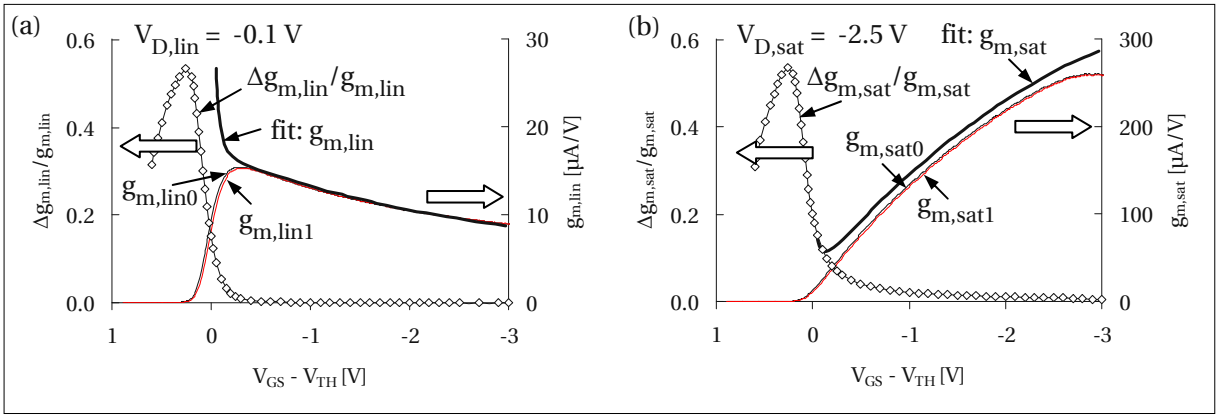


**Figure 2.5:** Transconductances (solid lines) and relative shifts of the transconductances (open diamonds) recorded before and after stress in the linear ($V_{D,lin}$ = -0.1 V) (a) and in the saturation ($V_{D,sat}$ = -2.5 V) (b) regime of the device. The thick solid lines represent fits of the transconductances in the linear (a) and saturation (b) regime according to Eq. 2.18 and Eq. 2.19. From the relative shifts of the transconductances we calculate the slope of spurious $V_{TH}$ shift caused by mobility degradation, cf. Eq. 2.20 and Eq. 2.21.

When exceeding the threshold voltage of the device, the defect charge contribution ($\Delta V_{TH}^{q}$) is assumed to reach a maximum (≈ -0.02 V). This assumption is likely not completely correct since the Fermi level still moves slightly toward the valence band when increasing the overdrive, thereby charging additional defects located close to the band edge, cf. Fig. 2.2. Neglecting this second-order effect, any additional $V_{TH}$ shift measured in the overdrive region (II) may be attributed to mobility degradation ($\Delta V_{TH}^{\mu}$). This is supported by the observation that $\Delta V_{TH,lin}$ and $\Delta V_{TH,sat}$ begin to drift apart not before entering the overdrive region (II) whereas their increase is similar in the subthreshold region (I) of the device.

As long as one only measures in the linear drain current region, the mobility dependent $V_{TH}$ shift is almost independent of the drain voltage (cf. Fig. 2.6 (a)), consistent with Eq. 2.13, however, in the saturation region of the device (cf. Fig. 2.6 (b)), only about half of the increase in $\Delta V_{TH}^{\mu}$ is observed, consistent with Eq. 2.13. The thick solid lines in Fig. 2.6 (a) and (b) are fit curves calculated analytically using Eq. 2.22 and Eq. 2.23, the offset being the defect charge dependent $V_{TH}$ shift at the threshold voltage ($\Delta V_{TH}^{q}$). By inserting the curvature parameter ($\Theta$) and the measured increase of the relative transconductance into Eq. 2.22 and Eq. 2.23, perfect agreement with the measurement data is obtained indicating that the spurious $\Delta V_{TH}$ component observed in the overdrive region of the device can actually be attributed to mobility degradation caused by enhanced Coulomb scattering ($\Delta V_{TH}^{\mu}$) at newly created charged defect centers.

From this case study we conclude that when recording the $\Delta V_{TH}$ shift at an overdrive of -1 V, the mobility component ($\Delta V_{TH}^{\mu}$) accounts for about 20 % in the linear region and for approximately 10 % in the saturation region of the device. However, when increasing the overdrive further toward -3 V, its contribution represents already about 35 % in the linear region and about 20 % in the saturation region of the device, which is then no longer negligible.

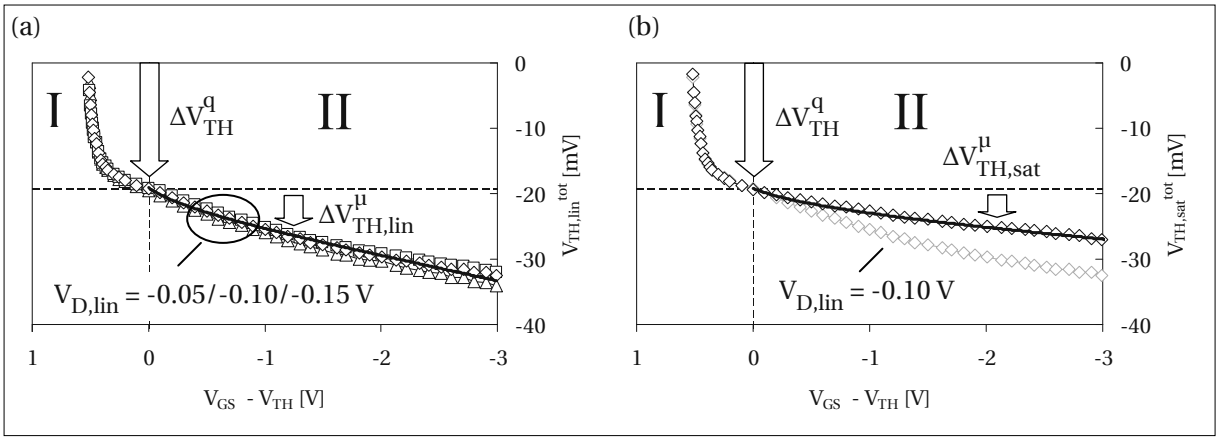


**Figure 2.6:** Total threshold voltage shifts measured after NBTI stress in the linear (a) and in the saturation (b) regime of the transistor. In the linear regime (a) similar results for three different drain biases (-0.05V triangles; -0.10V diamonds; -0.15V squares) are obtained. In the subthreshold region of the device (I) the increase in the $V_{TH}$ is attributed to the charging of NBTI induced defects. In the overdrive region (II) the obtained $V_{TH}$ shift in the linear region is about twice the shift measured in the saturation region. The thick solid lines in (a) and (b) are analytical fits according to Eq. 2.22 and Eq. 2.23 which are in perfect agreement with the measurement data.

## 2.2 Classification of defect charges by their electrical response time

When neglecting the spurious $V_{TH}$ shift caused by increased Coulomb scattering, the true reliability concern linked to negative bias temperature instability is originated in defects representing electrically active charges which counterbalance the applied gate potential causing a Fermi level dependent threshold voltage shift in the subthreshold region of the device. Besides their physical origin which will be discussed in detail in Chapter 4, defect charges may be classified empirically by their electrical response time, in particular, by their ability to communicate with the silicon substrate, thereby changing their charge state.

A very basic experimental classification of defect charges may differentiate between interface states and oxide traps, the first exchanging carriers with the silicon substrate very rapidly according to SRH theory, the latter being either permanently charged or provide at least carrier exchange time constants which are slow enough so that their capture and emission processes may be observed in low frequency or DC experiments as well. Naturally, such a classification is physically not well defined since the distinction is directly linked to a particular experimental setup. On the other hand, the differentiation of defects classes by their experimental response time seems obvious and reasonable not at least because their time and bias dependent appearance directly reflects their influence on device reliability under certain operating conditions.

To develop a basic understanding of the relative contributions of different defect types to the overall defect charge dependent threshold voltage shift, the basic experimental approach is usually to combine the results of a DC experiment being sensitive to both fast and slow states with the results of an AC experiment being sensitive only to fast interacting defects like for example interface states.

The DC component is basically represented by the threshold voltage shift calculated from the drain current degradation before and after NBTI stress (as discussed in Section 2.1). The $V_{TH}$ shift around the threshold voltage of the device combines the relative charge contributions of all kinds of defects located within the oxide and at the $SiO_2$/Si interface.

### 2.2.1 Charge pumping as a qualification tool

In order to investigate the contribution of fast interface states separately, additional methods have to be applied since a DC measurement is by far not fast enough to capture the fast recombination time constants associated with interface state charging. A well developed technique allowing to count the total number of interface states within a defined energy range of the silicon bandgap is the so-called charge pumping (CP) method [36, 37]. During CP, the gate junction is pulsed continuously between inversion and accumulation using a trapezoidal pulse shape and frequencies typically between 1 kHz and 1 MHz. The source and the drain junction of the transistor are short-circuited during the measurement acting as minority carrier sources during the inversion phase of the gate pulse. The source/substrate and drain/substrate diodes of the MOSFET structure are typically slightly reverse biased in order to suppress geometric leakage currents appearing at high gate pulsing frequencies [38, 39]. During the accumulation phase of the gate pulse the substrate junction (typically held at 0.0 V during the measurement) provides a source of majority carriers.

CP is a dynamic measurement method which drives the interface periodically into a non-steady state by switching the gate bias quickly between inversion and accumulation. In inversion, minority carriers coming from the highly doped source/drain junctions get trapped in interface states located energetically within the silicon bandgap, while during the subsequent accumulation phase, majority carriers coming from the silicon substrate recombine with previously trapped inversion charges giving rise to a DC substrate current ($I_{CP}^{max}$) which is proportional to the average density of interface traps ($\overline{D_{it}}$) within a defined energy range ($\Delta E_{CP}$). Hence, from an experimental point of view, we denote each trap fast enough to contribute to the charge pumping signal within half of a gate pulsing period as an interface state. This approximation is not too bad according to elastic tunneling simulations of Heh *et al.* [40] who calculated a spatial probing depth of approximately 5–6 Å in pure amorphous $SiO_2$ using high/low times of 1 $\mu$s which corresponds to a frequency of 500 kHz.

Following [37], the maximum CP current is given by

$$I_{CP}^{max} = q A_G^{eff} f \int_{E_{em}^h}^{E_{em}^e} D_{it}(E)\, dE = q A_G^{eff} f \overline{D_{it}} \Delta E_{CP}, \tag{2.31}$$

where q is the elementary charge, $A_G^{eff}$ is the effective gate area and $f$ is the gate pulsing frequency. The effective gate area defines the region below the gate oxide where trapping and detrapping may occur. It is generally smaller than the geometric gate area due to space charge regions which emerge when applying a reverse bias to the source/drain junctions. The pumped charge per area ($Q_{CP}^{max}$) and

pulse period is calculated by dividing the maximum CP current by the effective gate area and the pulse frequency:

$$Q_{CP} = qN_{CP} = \frac{I_{CP}^{max}}{A_G^{eff} f} = q\overline{D_{it}}\Delta E_{CP}, \tag{2.32}$$

where $N_{CP}^{max}$ is the number of pumped charges per area and pulse period.

The upper ($E_{em}^{e}$) and lower ($E_{em}^{h}$) boundary of the active energy interval $\Delta E_{CP}$ may be derived from the equations of the SRH theory, yielding

$$E_{em}^{h} = E_i - k_B T \ln\left(\frac{\Delta V_{G,pulse}}{\nu_{th,h}\sigma_h n_i t_r (V_{TH}^{CP} - V_{FB}^{CP})}\right), \tag{2.33}$$

$$E_{em}^{e} = E_i + k_B T \ln\left(\frac{\Delta V_{G,pulse}}{\nu_{th,e}\sigma_e n_i t_f (V_{TH}^{CP} - V_{FB}^{CP})}\right), \tag{2.34}$$

where $\Delta V_{G,pulse}$ is the gate pulse amplitude ($\Delta V_{G,pulse} = V_{GH} - V_{GB}$), $\nu_{th,h}$ and $\nu_{th,e}$ are the thermal drift velocities of holes and electrons, $\sigma_h$ and $\sigma_e$ are the capture cross sections for electron and hole capture, $n_i$ is the intrinsic carrier concentration, $t_f$ and $t_r$ are the fall and rise times of the trapezoidal gate pulse and $V_{TH}^{CP}$ and $V_{FB}^{CP}$ are the charge pumping threshold and flat band voltages.

From Eq. 2.33 and Eq. 2.34, the scanned energy interval $\Delta E_{CP}$ may be calculated as

$$\Delta E_{CP} = E_{em}^{e} - E_{em}^{h} = 2k_B T \ln\left(\frac{\Delta V_{G,pulse}}{\overline{\nu_{th}\sigma} n_i \sqrt{t_r t_f}(V_{TH}^{CP} - V_{FB}^{CP})}\right), \tag{2.35}$$

where the particular thermal drift velocities and capture cross sections for holes and electrons have been replaced by their average values $\overline{\nu_{th}}$ and $\overline{\sigma}$.

Following [41], the thermal drift velocity in Eq. 2.35 can be approximated as

$$\nu_{th} = \sqrt{\frac{3k_B T}{m_{eff}}}, \tag{2.36}$$

where $m_{eff}$ is the effective mass of the particular carrier. The experimental determination of the capture cross section is typically performed using three level charge pumping [42,43], deep level transient spectroscopy (DLTS) [44, 45], capacitance voltage (CV) sweeps [46] or from conductance measurements [47]. Although the capture cross sections reported in literature vary considerably dependent on the technique with which they have been determined, most authors report values between $10^{-14}\,cm^2$ and $10^{-16}\,cm^2$, the larger values corresponding typically to defect states located energetically close to midgap, the smaller values are often associated with trap levels located closer to the band edges. Some authors, i.e. [48], report that the capture cross sections of interface traps can may be considered as constant over a wide temperature and energy range.

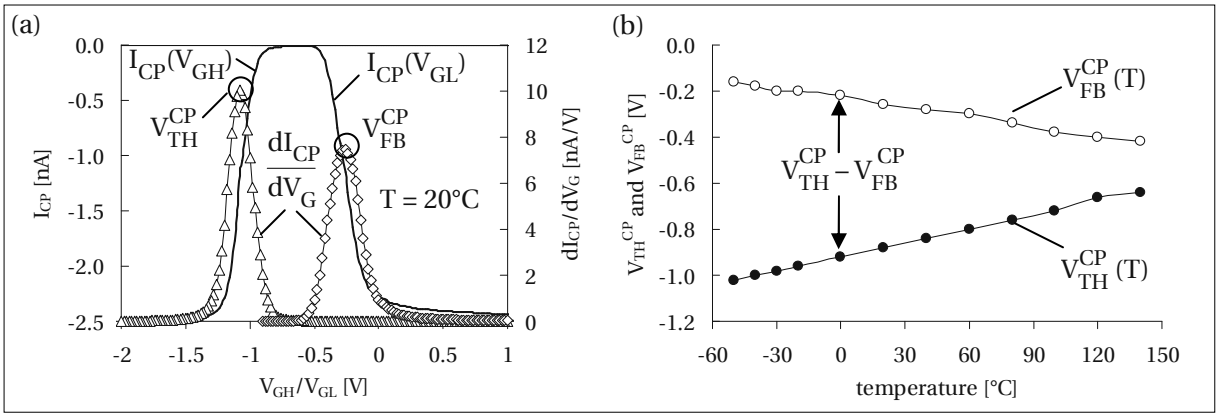


**Figure 2.7:** In (a) constant base level and constant high level CP curves (solid lines) and their derivatives (open symbols) are illustrated as a function of the gate bias. The measurement was performed on a PMOS device (SM5P/30/H2) measured with a source/drain to substrate reverse bias of -0.2V at a temperature of 20°C. The charge pumping threshold and flatband voltages correspond to the gate biases at which the derivatives are maximal. In (b) the measured CP threshold (full symbols) and flatband voltages (open symbols) are illustrated in a temperature range between -50°C and 140°C.

Following [41], the temperature dependent intrinsic carrier concentration is given by

$$n_i = \sqrt{N_C N_V} \exp\left(-\frac{E_G}{2k_B T}\right), \tag{2.37}$$

where $N_C$ is the effective density of states in the conduction band and $N_V$ is the effective density of states in the valence band.

In order to calculate the active energy interval $\Delta E_{CP}$ for a particular device and different experimental setups, the charge pumping related threshold $V_{TH}^{CP}$ and flatband $V_{FB}^{CP}$ voltages have to be determined experimentally from constant base level and constant high level charge pumping measurements. The method is described in detail in [49]. Fig. 2.7 (a) shows constant base level ($I_{CP}(V_{GH})$) and constant high level ($I_{CP}(V_{GL})$) charge pumping measurements performed on a PMOS device (SM5P/30/H2) measured with a source/drain to substrate reverse bias of -0.2V at a temperature of 20°C. The charge pumping related threshold voltage $V_{TH}^{CP}$ corresponds to the gate voltage at which the slope of the $I_{CP}(V_{GH})$ curve has its maximum (numerical derivative illustrated by open triangles), the charge pumping related flatband voltage $V_{FB}^{CP}$ corresponds to the gate voltage at which the slope of the $I_{CP}(V_{GL})$ curve has its maximum (numerical deviation illustrated by open diamonds). The same procedure has been performed for different temperatures ranging from -50°C to 140°C and the results for $V_{TH}^{CP}$ and $V_{FB}^{CP}$ are illustrated in Fig. 2.7 (b) showing that the difference between $V_{TH}^{CP}$ and $V_{FB}^{CP}$ shrinks with increasing temperature. This is due to the fact that the intrinsic carrier concentration (cf. Eq. 2.37) increases exponentially with the temperature, providing at higher temperatures a sufficiently large carrier concentration for trap filling already in weaker inversion, and accumulation, respectively.

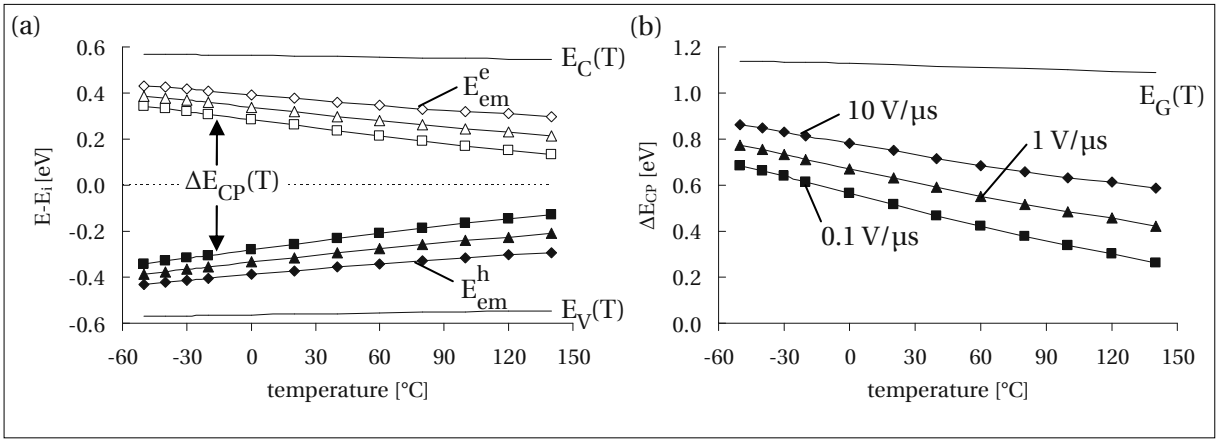


**Figure 2.8:** In (a) the emission boundaries $E^{h}_{em}$ (full symbols) and $E^{e}_{em}$ (open symbols) are illustrated as a function of temperature for three different rising/falling slopes of 10V/$\mu$s (diamonds), 1V/$\mu$s (triangles) and 0.1V/$\mu$s (squares). In (b) the resulting active energy intervals $\Delta E_{CP}$ are depicted as a function of temperature.

When inserting the thermal drift velocities (Eq. 2.36) and intrinsic carrier densities (Eq. 2.37) as well as the measured charge pumping threshold and flatband voltages into Eq. 2.33 and Eq. 2.34, we may calculate the charge pumping emission boundaries $E^{e}_{em}$ and $E^{h}_{em}$ for different experimental setups as a function of the temperature. The results for three different rising/falling slopes of the gate pulse ranging from 10V/$\mu$s to 0.1V/$\mu$s are illustrated in Fig. 2.8 (a) assuming a medium capture cross section of $\overline{\sigma} = 10^{-15}\,\mathrm{cm}^2$. The result demonstrates that the emission boundaries approach the band edges with lowering the measurement temperature (due to $n_i$) and with steeper rise and fall edges of the gate pulses. In particular, when measuring at -50°C using rising/falling slopes of 10V/$\mu$s, the detectable energy interval $\Delta E_{CP}$ covers nearly the entire silicon bandgap, the emission boundaries being then only about 100mV away from the silicon band edges. When measuring at higher temperatures using less steep rising/falling slopes, the active energy range becomes continuously narrower mapping then only a very small area around midgap. The development of $\Delta E_{CP}$ being symmetrically around midgap when using a symmetrical pulse shape, is illustrated as a function of temperature in Fig. 2.8 (b).

### 2.2.2 Energetic mismatch between the CP technique and the static DC approach

When applying charge pumping in order to estimate the interface state dependent contribution to the drain current degradation measured under DC bias conditions, it is necessary to convert the stress induced increase of the CP current ($\Delta I_{CP}$) as well as the stress induced increase of the pumped charge ($\Delta Q^{max}_{CP}$), correctly into a corresponding interface state dependent threshold voltage shift ($\Delta V^{it}_{TH}$). The conversion is generally not straight-forward if the energy interval of charged defects under DC bias conditions (cf. $\Delta E_Q$ in Eq. 2.25) does not coincide with the active energy interval profiled during charge pumping (cf. $\Delta E_{CP}$ in Eq. 2.35).

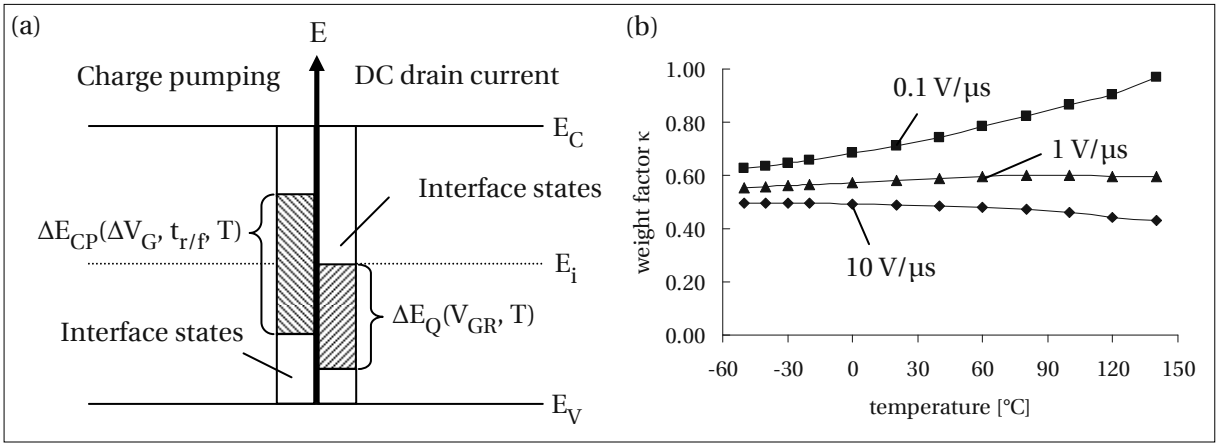


**Figure 2.9:** In (a) the energy mismatch between a charge pumping measurement ($\Delta E_{CP}$; left hand side) and a DC drain current measurement ($\Delta E_Q$; right hand side) is illustrated schematically for a PMOS device biased in inversion. In (b) the analytically derived weight factors $\kappa^{V_{TH}}$ (cf. Eq. 2.41) are plotted as a function of the temperature for three different rising/falling slopes: 10 V/$\mu$s (diamonds), 1 V/$\mu$s (triangles) and 0.1 V/$\mu$s (squares).

The situation is illustrated schematically in Fig. 2.9 (a) where $\Delta E_{CP}$ is depicted on the left hand side and $\Delta E_Q$ is illustrated on the right hand side. The energy interval $\Delta E_{CP}$ may be adjusted experimentally by varying the shape of the gate pulse and the temperature. In particular, $\Delta E_{CP}$ is independent of the charge state of the interface traps and generally contains information about parts of both the upper and lower half of the silicon bandgap when using a symmetrical pulse shape. By contrast, when measuring the drain current at a constant gate bias, $\Delta E_Q$ and the obtained threshold voltage shift are governed by the current Fermi level position during read-out and by the net charge state of electrically active interface states, cf. Eq. 2.25.

Fig. 2.9 (a) suggests that a reliable comparison of the charge pumping signal and the DC drain current measurement can only be achieved when taking the energetic mismatch either quantitatively into account (i.e. by a weight factor) or alternatively when tuning both measurement techniques in way so that their energy intervals overlap.

Assuming a homogeneous distribution of interface traps across the entire silicon bandgap, we may consider the energy mismatch illustrated in Fig. 2.9 (a) by means of a weight factor ($\kappa$):

$$\Delta V_{TH}^{it} = \frac{\Delta Q_{it}}{C_{OX}} = \frac{q\overline{\Delta D_{it}}\Delta E_Q}{C_{OX}} = \frac{\Delta E_Q}{\Delta E_{CP}} \frac{\Delta I_{CP}^{max}}{A_G^{eff} f C_{OX}} = \kappa \frac{\Delta I_{CP}^{max}}{A_G^{eff} f C_{OX}}. \tag{2.38}$$

Assuming the PMOS transistor is biased at its physical threshold voltage (cf. Eq. 2.4), the Fermi level position may be approximated as

$$E_F^{V_{TH}} = E_i - k_B T \ln\left(\frac{N_D}{n_i}\right), \tag{2.39}$$

yielding for $\Delta E_Q$ at the threshold voltage

$$\Delta E_Q^{V_{TH}} = k_B T \ln\left(\frac{N_D}{n_i}\right). \tag{2.40}$$

From Eq. 2.35 and Eq. 2.40 we may derive an analytic expression for the weight factor at the threshold voltage of the device:

$$\kappa^{V_{TH}} = \frac{\Delta E_Q^{V_{TH}}}{\Delta E_{CP}} = \frac{1}{2} \frac{\ln\left(\frac{N_D}{n_i}\right)}{\ln\left(\frac{\Delta V_{G,pulse}}{\overline{\nu_{th}\sigma} n_i \sqrt{t_r t_f}(V_{TH}^{CP} - V_{FB}^{CP})}\right)}. \tag{2.41}$$

In Fig. 2.9 (b) the weight factor $\kappa^{V_{TH}}$ is plotted as a function of the temperature for three different rising/falling slopes.

It has to be pointed out that the weight factor approach implies that the distribution of interface states within the silicon bandgap is uniform. As will be discussed in Chapter 4, this is a very crude approximation since real density of state profiles have rather a U-shape which is more or less symmetrically around midgap [50]. In particular, when $\Delta E_Q$ and $\Delta E_{CP}$ are significantly misaligned, the weight factor approach may introduce a large error due to the non-uniformity of the DOS profile.

The best way allowing a reliable conversion without making any assumptions on the density of state profile is to align $\Delta E_Q$ and $\Delta E_{CP}$ in advance. By combining the results of Eq. 2.34 and Eq. 2.39, we may derive an appropriate specification for the rising slope of the gate pulse in order to make the lower emission boundary ($E_{em}^h$) coinciding with the Fermi level position at the threshold voltage ($E_F^{VTH}$):

$$\frac{\Delta V_{G,pulse}}{t_r} = \nu_{th,h}\sigma_h N_D (V_{TH}^{CP} - V_{FB}^{CP}). \tag{2.42}$$

We may derive a second specification for the falling slope of the gate pulse in order to adjust the upper emission boundary ($E_{em}^e$) to the amphoteric transition level of the interface states which is assumed to be $E_i$:

$$\frac{\Delta V_{G,pulse}}{t_f} = \nu_{th,e}\sigma_e n_i (V_{TH}^{CP} - V_{FB}^{CP}). \tag{2.43}$$

The calculated rising and falling slopes leading to a 100 % overlap of $\Delta E_{CP}$ and $\Delta E_Q$ at the threshold voltage of the device are illustarted in Fig. 2.10 (a) as a function of the temperature. In Fig. 2.10 (b), the corresponding rise and fall times have been calculated assuming a pulse amplitude of 1 V. According to Fig. 2.10, a 100 % overlap would require a considerable assymetric pulse shape, the rising slope being much steeper than the falling slope. In principle, such a pulse shape can be generated by a conventional pulse generator, however, due to the large fall time, the pulsing frequency would become too small in order to achieve a reasonable measurement resolution of the maximum charge pumping current (cf. Eq. 2.31) making the measurement procedure unfeasible on MOS devices having typical device geometries.

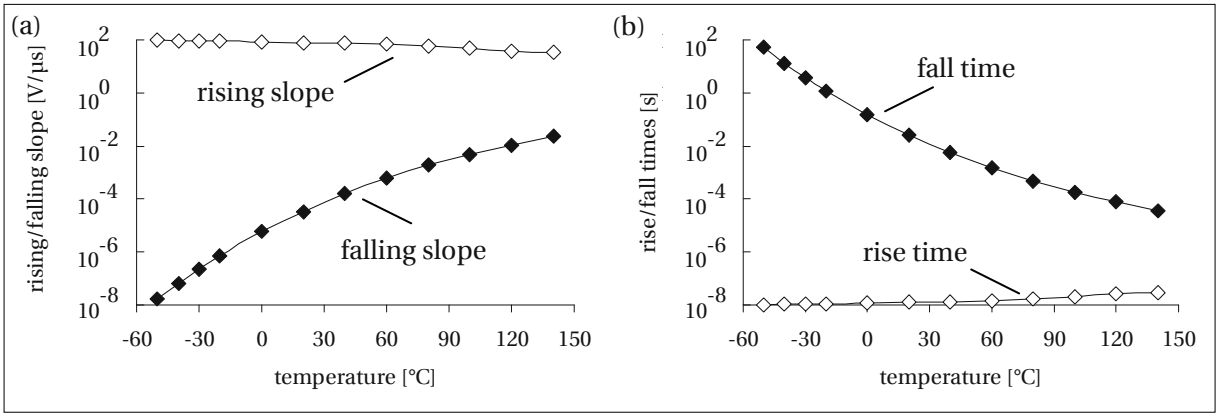


**Figure 2.10:** Fig. (a) illustrates the rising and falling slopes of a designed gate pulse making the energy intervals $\Delta E_{CP}$ and $\Delta E_Q$ coinciding at the threshold voltage of the device. For the calculation a donor doping density of $N_D = 10^{16}\,cm^{-3}$ and a unique capture cross section of $\sigma_h = \sigma_e = 10^{-15}\,cm^2$ were assumed. Fig. (b) shows the corresponding rise and fall times assuming a pulse amplitude of 1 V.

A reasonable compromise between the weight factor approach and the unfeasible 100 % overlap may be achieved by assuming a symmetrical density of state profile of interface traps around midgap. When doing so, we can use a symetrical pulse setup again, the rising and falling slopes being adjusted to make the lower emission boundary $E_{em}^{h}$ coinciding with the Fermi level position $E_F$ which guarantees at least that the whole range of $\Delta E_Q$ is covered by $\Delta E_{CP}$. Considering further that a comparable energy range in the upper half of silicon bandgap is profiled, the weight factor $\kappa$ becomes 1/2. Thus, except for the symmetry assumption, no additional approximation on the shape of the density of state profile has to be made in order to accomplish the conversion from $\Delta I_{CP}$ to $\Delta V_{TH}^{it}$ which is a considerable improvement in comparison to the uncorrelated weight factor approach.

# 3

# Dynamics of NBTI degradation and recovery

WHEN subjecting a PMOS device to NBTS, the amount of degradation is reported to increase gradually as a function of stress time, electric field and temperature, leading to a characteristic reduction of the MOSFET performance. However, once the stress bias is removed, at least a certain fraction of the total $V_{TH}$ shift is reported to recover very quickly, making the delay between the point of time of stress termination and the first measured point recorded at the threshold voltage of the device ($t_R$) an important parameter, which crucially determines the amount of degradation detected at the end of stress ($t_S$). The degradation and the recovery of the threshold voltage shift may be approximated empirically by power-laws:

$$\Delta V_{TH}(t_S, t_R) = A_S(E_{OX,S}, T_S)\, t_S^{n_S} - A_R(V_{GR}, T_R)\, t_R^{n_R} = A_S\, t_S^{n_S}\left(1 - \frac{A_R}{A_S}\frac{t_R^{n_R}}{t_S^{n_S}}\right), \tag{3.1}$$

where the pre-factors $A_S(E_{OX,S}, T_S)$ and $A_R(E_{OX,R}, T_R)$ are reported to depend on the oxide field during stress ($E_{OX,S}$) and recovery ($E_{OX,R}$) as well as on the stress temperature ($T_S$) and the recovery temperature ($T_R$). The power-law exponents $n_S$ and $n_R$ determine the degradation and recovery dynamics as a function of the stress ($t_S$) and the recovery time ($t_R$). $n_S$ and $n_R$ are extracted from the slope of the $V_{TH}$ shift in a double-logarithmic plot $\Delta V_{TH}^S$ vs. $t_S$, and $\Delta V_{TH}^R$ vs. $t_R$ respectively. Note that the second term in Eq. 3.1 accounts for power-law-like recovery after interrupting the stress phase. It may be neglected when the stress time ($t_S$) exceeds the measurement delay ($t_R$) by far (provided the overall dynamics of degradation and recovery are similar).

Long-term degradation ($t_S \geq 1 - 10\,s$) is reported to follow a typical power-law-like increase with $0.1 < n_S < 0.4$. On the other hand, short-term degradation ($t_S < 1\,s$) and recovery show much smaller power-law exponents ($n_S$ and $n_R \leq 0.1$), allowing to approximate Eq. 3.1 by logarithmic time depen-

dences. This can be easily demonstrated by expanding the power-law relations in Eq. 3.1 in Taylor series around small values of $t_{S0} > 0$, and $t_{R0} > 0$, respectively, yielding

$$\Delta V_{TH}(t_S, t_R) \approx A_S t_{S0}^{n_S} - A_R t_{R0}^{n_R} + n_S A_S \log\left(\frac{t_S}{t_{S0}}\right) - n_R A_R \log\left(\frac{t_R}{t_{R0}}\right), \tag{3.2}$$

$$\approx \log\left(\frac{t_{R0}^{B_R} t_S^{B_S}}{t_{S0}^{B_S} t_R^{B_R}}\right), \tag{3.3}$$

where $B_S = n_S A_S$ is the degradation rate and $B_R = n_R A_R$ is the recovery rate in mV/decade. Assuming NBTI to be caused by hole trapping within the gate oxide (as it is possibly the case for short stress times), $t_{S0}$ and $t_{R0}$ can be approximated by the minimum capture time constant $\tau_{c0}$ and the minimum emission time constant $\tau_{e0}$ at the particular stress/relax conditions after Huard *et al.* [51]. Universality of recovery traces ($\Delta V_{TH}$ vs. $t_S/t_R$) recorded at different stress times has been demonstrated by Grasser *et al.* in [52] and may be obtained when assuming $B_S = B_R$ and $t_{S0} = t_{R0} \gg t_S$ and $t_R$, yielding the following simplified expression for the threshold voltage shift:

$$\Delta V_{TH}(t_S, t_R) \approx B_R(E_{OX,S}, T_S, E_{OX,R}, T_R) \log\left(1 + \frac{t_S}{t_R}\right). \tag{3.4}$$

When neglecting recovery ($t_R \approx 0$), we may derive the degradation dynamics from Eq. 3.2 as follows:

$$\Delta V_{TH}(t_S, 0) = \Delta V_{TH}^{OTF}(t_S) \approx B_S(E_{OX,S}, T_S) \log\left(\frac{t_S}{t_{S0}}\right). \tag{3.5}$$

In [53] universal scalability of the short-term *and* long-term degradation and the recovery has been demonstrated for different stress fields, temperatures and oxide thicknesses suggesting either a single mechanism or at least tightly coupled phenomena to be responsible for NBTI induced degradation and recovery.

The dynamics of both branches, namely the evolution of the degradation as a function of the stress time, field and temperature and the evolution of the recovery as a function of the recovery time, field and temperature, are very important indicators used to evaluate the NBTI stability of a technology and to check theoretical predictions derived from particular physical models. Since degradation usually follows a power-law with $n_S > 0.1$ (except for short stress times) which cannot be perceived as log-like, it is typically described by the pre-factor $A_S$ and/or the power-law exponent $n_S$. On the other hand, (at least when being far away from recovery saturation) relaxation usually follows a power-law with $n_R \leq 0.1$ which is similar to log-like. Hence, the recovery rate per decade (pre-factor $B_R$) in mV/dec (when measuring $V_{TH}$) recovery or in A/dec (when measuring CP current recovery) can be used to fully describe relaxation dynamics.

In this chapter, different techniques are presented which allow to investigate the dynamics of defect generation and recovery under AC and DC bias conditions as a function of stress time, recovery time, temperature and electric field. By means of case studies performed optionally on thin high–$\kappa$ and/or thick $SiO_2$ devices, fundamental experimental signatures of NBTI are revealed helping to catch a glimpse of the multi-faceted features and complexity of the problem.

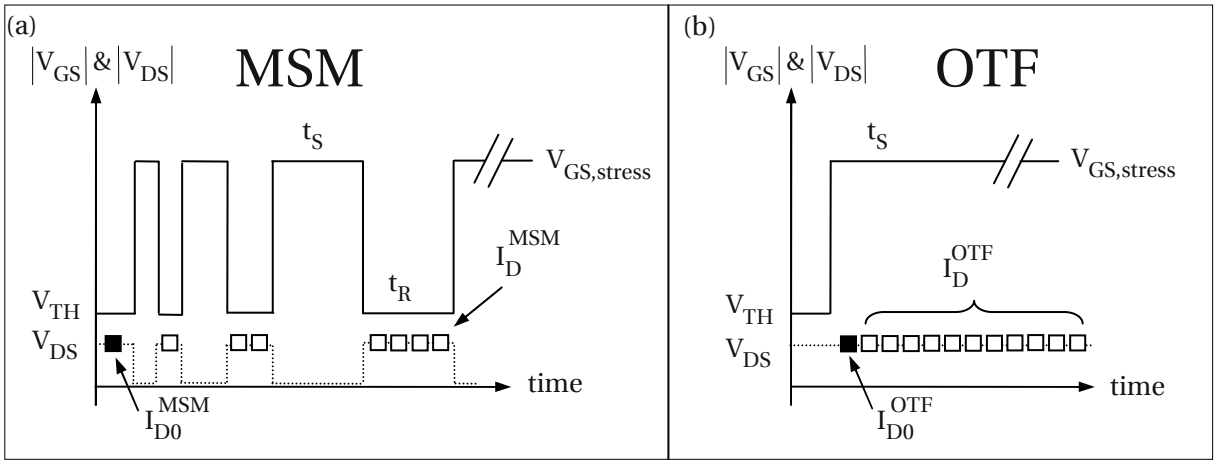


**Figure 3.1:** A schematic illustration of the gate and drain bias conditions applied during MSM (a) and OTF (b). When performing MSM (a), the drain current degradation and recovery is monitored after subsequent stress runs typically at the threshold voltage of the device (open squares) thereby interrupting the stress phase repeatedly by switching the gate bias from the stress level ($V_{GS,stress}$) to the $V_{TH}$. The stress-recovery intervals usually follow a geometric progression where the stress time equals the recovery time. For calculating the $V_{TH}$ shift the virgin drain current (full square) is taken as a reference. When performing OTF (b), the drain current degradation is measured directly under stress bias conditions making the measurement recovery-free ($t_R = 0\,s$). The first current value recorded under stress conditions ($I_{D0}^{OTF}$; full square) is taken as a reference for calculating the $V_{TH}$ shift.

## 3.1 Static MSM and OTF

So far, the two most frequently used measurement techniques for evaluating NBTI dynamics during stress and recovery are the classical measure/stress/measure (MSM) technique and the OTF measurement method [54]. Both techniques have certain advantages and drawbacks making the method of choice dependent on the particular problem [55, 56]. The biasing conditions applied during static MSM and OTF are schematically depicted in Fig. 3.1.

When applying the MSM technique, a single device is usually stressed repeatedly with increasing stress ($t_S$) and recovery times ($t_R$) following a geometric progression. Thereby it is usually assumed that the amount of defect relaxation occurring during the last recovery period is fully restored at the end of the subsequent longer-lasting stress run. After each stress run, the gate bias is switched from the stress level ($V_{GS,stress}$) to the threshold voltage of the device. In parallel the drain bias is switched from 0.0 V to the particular read-out bias ($V_{DS}$) initializing the drain current measurement $I_D^{MSM}$ as a function of the recovery time $t_R$, and the stress time $t_S$, respectively. In the data analysis, the drain current degradation (recorded at the threshold voltage) is converted into a corresponding $V_{TH}$ shift ($\Delta V_{TH}^{MSM}$) by assigning the degraded current values a corresponding gate voltage recorded at the virgin device, cf. Chapter 2.

Hence, the MSM technique provides the feature to analyze $V_{TH}$ shifts as a function of the stress and the recovery time, the gate bias being typically (but not necessarily) close to the threshold voltage of

the device during read-out. One important drawback of the technique is the fact that the extracted $V_{TH}$ shifts are always afflicted with an unknown amount of recovery occurring between the point of time after removal of the stress bias and the first measurement of the drain current at the threshold voltage of the device, cf. Eq. 3.1. For experimental reasons, the switching event and the following measurement cannot be performed arbitrarily fast ($\geq 1\,\mu$s), leaving much room for speculation what happens within the first couple of nanoseconds after terminating stress. Until now, only a few groups claim that they seriously succeed in finding a plateau in their $\Delta V_{TH}$ recovery curves at very early recovery times. By using the Ultra-Fast VT (UFV) measurement method [57], a minimum delay time in the range of microseconds was achieved [58, 59]. Although quite promising, a possible explanation for this observed plateau has been given in [60] as being due to the difficulty of synchronizing the recovery time scale with the real end of stress. In most reported cases, however, the extracted recovery curves are straight lines in a logarithmic time plot, indicating a very low starting time of the recovery event which remains apparently inaccessible even when advancing toward the microsecond regime [61].

When applying the OTF technique, the drain current degradation is monitored directly under stress bias conditions, making the measurement procedure itself quasi recovery-free. At the moment the gate bias is switched from the threshold voltage to the stress level ($V_{GS,stress}$), the drain current measurement is started instantaneously, with the read-out drain bias applied also during stress. In order to convert the degraded drain currents recorded under stress bias conditions ($I_D^{OTF}$) into corresponding $V_{TH}$ shifts ($\Delta V_{TH}^{OTF}$), the first current value recorded at $V_{GS,stress}$ is usually considered as virgin ($I_{D0}^{OTF}$), yielding the following expression for the threshold voltage shift measured during OTF (cf. Eq. 2.5):

$$\Delta V_{TH}^{OTF}(t_S) \approx (V_{GS} - V_{TH}) \frac{I_D^{OTF}(t_S) - I_{D0}^{OTF}}{I_{D0}^{OTF}} = (V_{GS} - V_{TH}) \frac{\Delta I_D^{OTF}(t_S)}{I_{D0}^{OTF}}. \tag{3.6}$$

As opposed to the $V_{TH}$ shift measured at the actual threshold voltage of the device, $\Delta V_{TH}^{OTF}$ contains a parasitic component caused by mobility degradation (cf. 2.1.2) making it difficult to compare $\Delta V_{TH}^{OTF}$ and $\Delta V_{TH}$ directly. In order to estimate the parasitic mobility impact of $\Delta V_{TH}^{OTF}$, the classical OTF technique has been extended to a so-called second level or three level OTF procedure were the gate voltage is varied slightly around the stress bias $V_{GS,stress}$ every time a drain current measurement is required allowing to estimate the variation of the transconductance around the stress voltage [54]. Besides the mobility influence, the OTF technique suffers from an additional handicap concerning the definition of the first measured drain current value $I_{D,lin0}$: Eq. 3.6 assumes $I_{D,lin0}$ to be the virgin drain current at the stress voltage, however, when considering degradation and recovery as similarly fast, $I_{D,lin0}$ is actually already degraded to a certain degree, again leaving much room for speculation about what happens in the first picoseconds after initiating stress [62]. From an application point of view, the OTF technique is only feasible when studying thin oxide technologies having their stress biases ($V_{GS,stress}$) not far away from their threshold voltages ($V_{TH}$). This is because the signal vs. noise resolution of the relative drain current degradation decreases linearly when increasing the gate voltage overdrive, cf. Eq. 2.5.

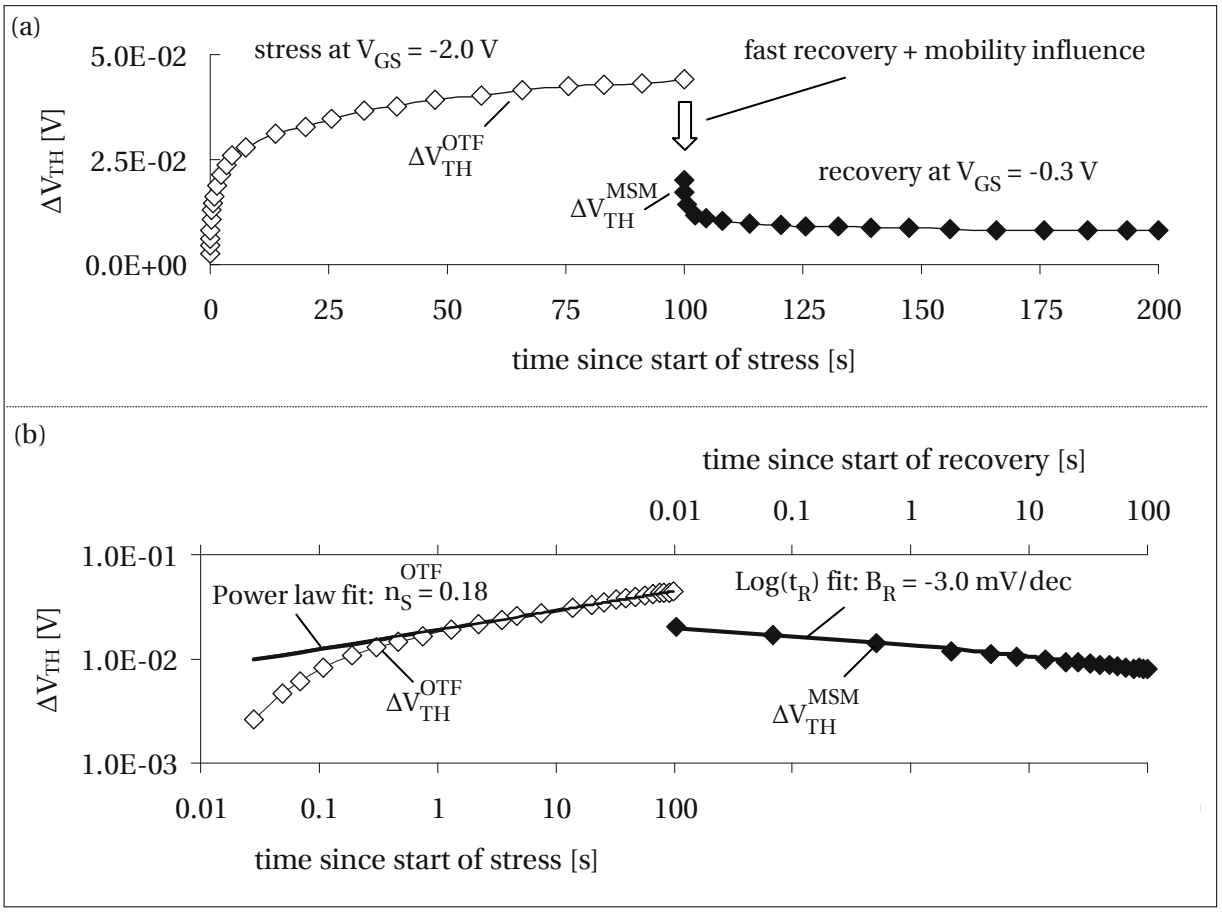


**Figure 3.2:** Degradation (open diamonds) and recovery (full diamonds) of the threshold voltage shift recorded for 100 s at 125°C on a 1.5 nm high–$\kappa$ PMOS device (HK2P/1.5/1). In (a), the $V_{TH}$ shifts are illustrated in a lin-lin plot, in (b), the same data is illustrated in a log-log plot. The threshold voltage shift during stress ($V_{GS,stress}$ = -2.0 V) was calculated using Eq. 3.6. $I_{D,lin0}$ was measured 10 ms after applying stress. The threshold voltage shift during recovery ($V_{GS,rec}$ = -0.3 V) was calculated by referencing the degraded drain current measured at $V_{TH}$ to the virgin transfer curve. Between the last measured point under stress conditions and the first measured point under recovery conditions there is a 24 mV gap which may be attributed to a mobility component in $\Delta V_{TH}^{OTF}$ and to fast recovery within the first 10 ms after removal of stress.

In order to demonstrate the dynamics of NBTI stress and recovery, a 1.5 nm high–$\kappa$ PMOS device equipped with a $p^{++}$ gate poly (HK2P/1.5/1) was stressed for 100 s at 125°C using a stress voltage of -2.0 V, cf. Fig. 3.2. Following Eq. 1.7, -2.0 V corresponds to an oxide field of approximately 6.6 MV/cm. During the stress period, the linear drain current ($V_{DS}$ = -0.1 V) was monitored and $\Delta V_{TH}^{OTF}$ was calculated using Eq. 3.6. $I_{D,lin0}$ was measured 10 ms after applying the stress bias. Subsequently to the stress phase, the gate bias was switched to the threshold voltage of the device (-0.3 V) recording the linear drain current then under recovery bias conditions. During recovery, $\Delta V_{TH}^{MSM}$ was calculated by referencing the degraded drain current to a gate voltage corresponding to the virgin transfer curve. As can be seen in Fig. 3.2 (a), during stress the threshold voltage shift ($\Delta V_{TH}^{OTF}$) increases rapidly achieving 44 mV after 100 s stress. At that moment, the stress bias is switched from -2.0 V to -0.3 V, the ob-

tained threshold voltage shift measured under recovery conditions ($\Delta V_{TH}^{MSM}$) is considerably reduced by 24 mV within the first 10 ms after removal of the stress bias. The large $\Delta V_{TH}$ gap between the last measured point of $\Delta V_{TH}^{OTF}$ and the first measured point of $\Delta V_{TH}^{R}$ indicates fast recovery, although probably not the entire 24 mV gap may be attributed to trap annealing since $\Delta V_{TH}^{OTF}$ is likely to contain an offset caused by mobility degradation.

Fig. 3.2 (b) illustrates that both $\Delta V_{TH}^{OTF}$ and $\Delta V_{TH}^{MSM}$ increase and decrease more or less linearly in a double logarithmic diagram indicating a power-law-like correlation between degradation and stress time ($t_S$) as well as between relaxation and recovery time ($t_R$) as suggested in Eq. 3.1. Note that due to the small power-law factor associated with recovery, we may describe the relaxation branch also as being log-like with a recovery rate of $B_R$ = -3.0 mV/dec. The power-law factor measured during OTF stress ($n_{tot}^{OTF}$) is found to be 0.18 which is pretty close to the factor 1/6 often reported in literature. The kink in the stress curve visible at early stress times may be a consequence of the fact that $I_{D,lin0}$, measured 10 ms after applying the stress bias, is actually already degraded to a certain degree, distorting the $\Delta V_{TH}^{OTF}$ shift evaluation in particular at the beginning of the stress phase where the overall $V_{TH}$ degradation is low.

## 3.2 CP MSM and OFIT

Besides the conventional static MSM and OTF techniques discussed in Section 3.1, recently similar methods including gate pulsing periods for charge pumping have been developed [63, 64]. During a CP MSM or on-the-fly interface trapping (OFIT) measurement, the gate junction is either pulsed *continuously* during stress and recovery or just *occasionally* for a short time when a CP measurement is performed. In the following, the first method, where the gate junction is pulsed continuously, will be indicated as pulsed CP MSM and pulsed OFIT while the latter method, where the gate junction is pulsed exclusively when measuring the CP current, will be denoted as interrupted CP MSM and interrupted OFIT. The gate biasing conditions applied during various kinds of CP MSM and OFIT are schematically depicted in Fig. 3.3. The stress-recovery intervals may follow a geometric progression in a similar way as discussed for the static MSM experiment.

Typically the base levels of the gate pulses ($V_{GB} < V_{FB}^{CP}$) are chosen identically during CP MSM and OFIT. The high levels of the gate pulses during OFIT ($V_{GH}^{S}$) equal the former static stress voltage ($V_{GS,stress}$) subjecting the device periodically to electrical stress while measuring in parallel the CP current under stress conditions ($I_{CP}^{OFIT}$). In [64] it was claimed that using the OFIT procedure the measurement is recovery free. The high levels of the gate pulse during CP MSM ($V_{GH}^{R}$) correspond to a gate bias which exceeds $V_{TH}^{CP}$ in order to measure the maximum CP current but does not subject the device to stress during the measurement of $I_{CP}^{MSM}$. During the CP MSM procedure the device is assumed to recover.

During OFIT, the stress phase is periodically interrupted since the gate junction has to be pulsed toward accumulation in order to enable CP. The number of stress interruption events (accumulation phases) is considerably increased when performing pulsed OFIT as when applying interrupted OFIT.

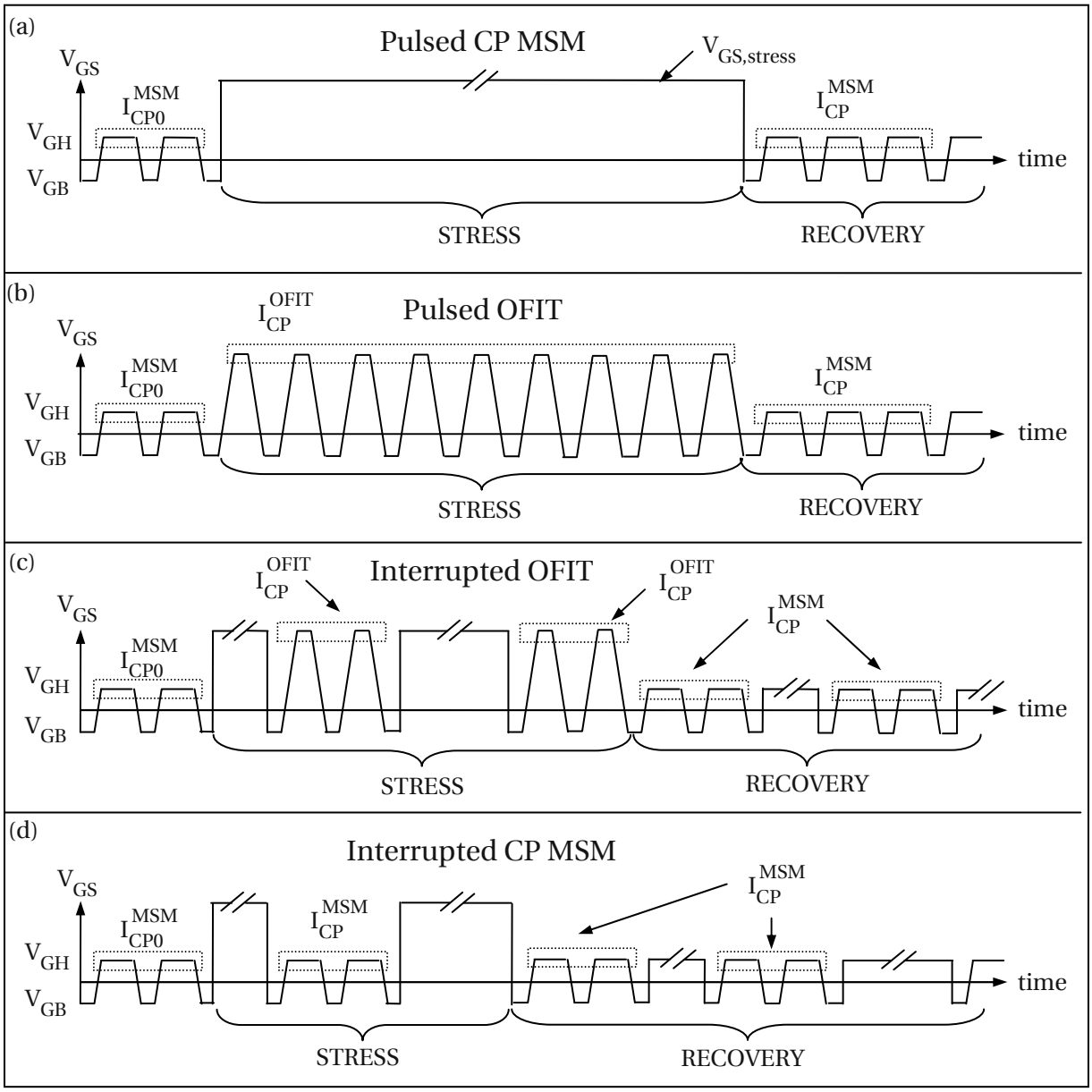


**Figure 3.3:** A schematic illustration of the gate bias conditions applied during 'pulsed CP MSM' (a), 'pulsed OFIT' (b), 'interrupted OFIT' (c) and 'interrupted CP MSM' (d). When performing 'pulsed CP MSM' (a) or 'interrupted CP MSM' (d), the charge pumping current ($I_{CP}^{MSM}$) is monitored under recovery conditions right after stress by pulsing the gate junction periodically between accumulation ($|V_{GB}| < |V_{FB}^{CP}|$) and inversion ($|V_{GH}^{R}| > |V_{TH}^{CP}|$). When performing pulsed OFIT (b) or interrupted OFIT (c), the CP current ($I_{CP}^{OFIT}$) is recoded directly under stress conditions by pulsing the gate junction between accumulation and inversion thereby extending the high level of the gate pulse to the stress voltage $V_{GS,stress}$. In order to scan a similar energy range during 'CP MSM' and 'OFIT', the rising/falling times of the gate pulses have to be adjusted appropriately in order to provide identical pulse slopes.

As a result, the total amount of degradation at the end of stress is decreased in comparison to an uninterrupted static stress experiment. To suppress recovery as efficiently as possible during OFIT,

the duty cycle during pulsed stress may be extended from 50 % to 99 %, thereby minimizing the time in accumulation where recovery may occur.

During CP MSM (Fig. 3.3 (a) and (d)) and during interrupted OFIT (Fig. 3.3 (c)) the effective stress time equals roughly the accumulated sum of the DC stress phases. During pulsed OFIT (Fig. 3.3 (b)), the effective stress time ($t_S^{eff}$) is defined by the accumulated time intervals where the stress bias ($V_{GS,stress}$) is applied:

$$t_S^{eff} = f(t_W - t_r)\, t_S \approx \frac{\text{duty cycle [\%]}}{100}\, t_S, \tag{3.7}$$

where $t_W$ is the pulse width, $t_r$ is the rise time and $f$ is the gate pulsing frequency. The effective stress time ($t_S^{eff}$) yields approximately half of the actual stress time ($t_S$) when using a symmetrical pulse shape (duty cycle 50 %).

The degradation of the CP current during CP MSM ($\Delta I_{CP}^{MSM}$) and OFIT ($\Delta I_{CP}^{OFIT}$) is typically calculated by subtracting the virgin CP current measured under CP MSM biasing conditions ($I_{CP0}^{MSM}$):

$$\Delta I_{CP}^{MSM} = I_{CP}^{MSM} - I_{CP0}^{MSM}, \tag{3.8}$$

$$\Delta I_{CP}^{OFIT} = I_{CP}^{OFIT} - I_{CP0}^{MSM}. \tag{3.9}$$

Note that the degradation of the CP current is calculated for both $\Delta I_{CP}^{MSM}$ and $\Delta I_{CP}^{OFIT}$ by referencing to the virgin CP current recorded under CP MSM biasing conditions ($I_{CP0}^{MSM}$). This approach is considerably different from the one applied during static OTF where the stress induced increase of the drain current ($\Delta I_D^{OTF}$) is referenced to the first drain current value measured under stress conditions ($I_{D0}^{OTF}$). A mandatory requirement for referencing to $I_{CP0}^{MSM}$ when calculating $\Delta I_{CP}^{OFIT}$ is, to arrange the pulse shapes in a way so that the profiled energy ranges during CP MSM and OFIT coincide, cf. Eq. 3.10. This yields a better agreement but is still not completely correct since Eq. 2.35 does not consider the field dependence of different contributing defects. Eq. 2.35 implies a specific rule for aligning the pulse shapes in order to fix the gate pulsing slopes and hence $\Delta E_{CP}$:

$$\frac{\Delta V_{G,pulse}^{S}}{\Delta V_{G,pulse}^{R}} = \frac{V_{GH}^{S} - V_{GB}}{V_{GH}^{R} - V_{GB}} = \frac{t_r^{S}}{t_r^{R}} = \frac{t_f^{S}}{t_f^{R}}. \tag{3.10}$$

In order to demonstrate the dynamics of interface state creation and recovery, the 1.5 nm high–$\kappa$ PMOS device (HK2P/1.5/1) was stressed for an effective stress time of 100 s at 125°C using a stress voltage of -2.0 V (oxide field approximately 6.6 MV/cm). Following the stress phase, a 100 s lasting recovery phase was recorded. Four different experiments according to the four stress-relax modes discussed in Fig. 3.3 were performed on different devices having the same geometry. The results are summarized in Fig. 3.4: left hand side – linear scale; right hand side – logarithmic scale. During OFIT, the CP current was measured by pulsing the gate junction between +0.75 V and -2.0 V, recording in parallel the CP current $I_{CP}^{OFIT}$. During CP MSM, the gate bias was pulsed between +0.75 V and -0.75 V, recording in parallel the CP current $I_{CP}^{MSM}$. During stress and recovery a symmetrical pulse shape was used with a duty cycle of 50 %, a gate pulsing frequency of 2 MHz and rising/falling slopes of 20 V/$\mu$s.

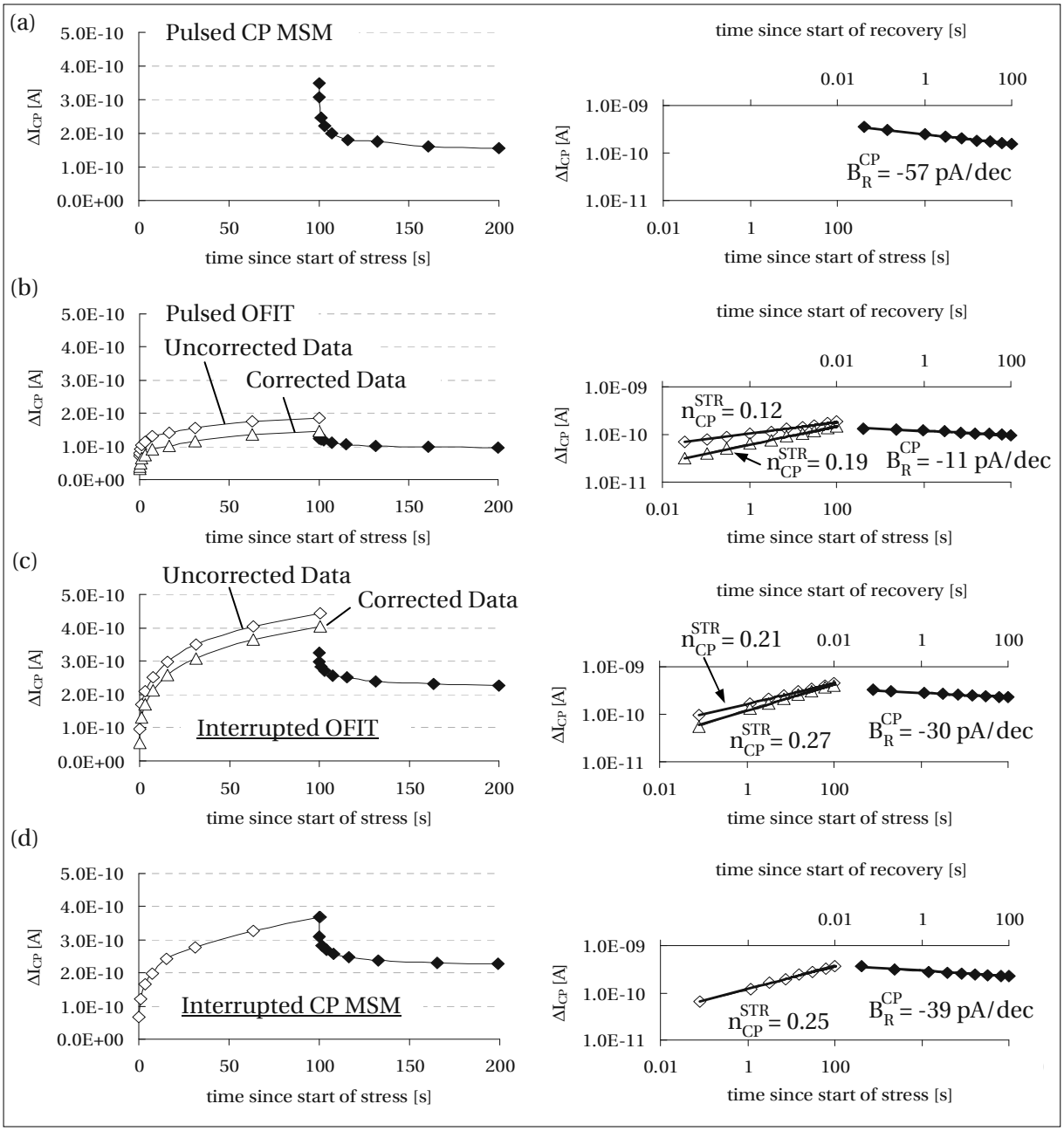


**Figure 3.4:** Degradation (open symbols) and recovery (full symbols) of the CP current measured in different CP modes for 100 s at 125°C on different 1.5 nm high–$\kappa$ PMOS devices (HK2P/1.5/1). On the left hand side, the data is plotted on a lin-lin scale (lines correspond to power-law fits), on the right hand side on a log-log scale (lines correspond to $\log(t_R)$ fits). Open diamonds reflect the uncorrected data, open triangles reflect the OFIT data corrected by the gate leakage currents and oxide charges. Full diamonds illustrates CP current recovery measured in the 'CP MSM' mode. Device (a) was stressed in the 'pulsed CP MSM' mode, device (b) in the 'pulsed OFIT' mode, device (c) in the 'interrupted OFIT' mode and device (d) in the 'interrupted CP MSM' mode.

Device (a) was stressed in the 'pulsed CP MSM' mode, recording the CP current recovery $\Delta I_{CP}^{MSM}$ during continuous gate pulsing following a 100 s DC stress phase. After terminating stress, the first measured CP current value was recorded with a time delay of 40 ms. Similar as observed for $\Delta V_{TH}^{MSM}$ during static MSM (cf. Fig. 3.2 (b)), the CP recovery follows a $\log(t_R)$ dependence, yielding a CP relaxation rate ($B_R^{CP}$) of -59 pA/dec.

Device (b) was stressed in the 'pulsed OFIT' mode, recording the CP current degradation $\Delta I_{CP}^{OFIT}$ during pulsed stress (effective stress time 100 s) and $\Delta I_{CP}^{MSM}$ during the subsequent 100 s pulsed recovery phase. Both, degradation and recovery, follow a power-law, however, with different exponents consistent with Fig. 3.2. The first measured CP current after initializing OFIT was recorded with an effective time delay of 40 ms. As can be seen in Fig. 3.2 (b) (left hand side), the uncorrected OFIT data (open diamonds) shows a significant offset ($\approx$ 71 pA) at the very beginning and at the end of the stress phase. In fact, this initial increase accounts already for 40 % of the total degradation (187 pA) obtained at the end of the 100 s lasting stress phase. Li *et al.* [64] attributed this initial offset to fast interface state creation at the very beginning of the degradation phase. Similarly, when switching from the pulsed OFIT mode to the pulsed CP MSM mode at the end of stress, a comparable decrease of 55 pA is obtained within the first 40 ms of recovery suggesting fast interface state repassivation as well when terminating stress. In order to check whether these initial offsets are actually caused by fast interface state degradation and recovery dynamics, as suggested by Li *et al.* in [64], we studied the OFIT and CP MSM technique in considerable detail [65]. Within this study we demostrated that the CP current is not stable once reaching the inversion regime, but keeps increasing slightly due to slow oxide traps increasingly contributing to the substrate current at elevated electric fields. Hence, data recorded during OFIT and CP MSM are not comparable in a straight-forward manner even when adjusting the rising/falling slopes in order to scan the same energy interval, cf. Eq. 3.10. By properly taking the contribution of oxide traps into account, we demonstrated [65] that no considerable fast degradation and recovery of interface states occurs at the very beginning and at the end of the stress phase revealing the major fraction of the obtained offsets between OFIT and CP MSM as a measurement artifact. After subtracting gate leakage currents and oxide trap contributions, we obtain a set of corrected OFIT data (open triangels) which displays considerably reduced offsets and different power-law dynamics: uncorrected data: $n_{CP}^{str} = 0.12$; corrected data: $n_{CP}^{str} = 0.19$. The relaxation rate associated with recovery during pulsed CP MSM ($B_R^{CP}$) yields -11 pA/dec which is considerably smaller than the one obtained in experiment (a) following DC stress. A further remarkable aspect of Fig. 3.4 may be found in the observation that the absolute amount of degradation obtained after NBTS is considerable larger when applying a constant stress bias (a) than when applying pulsed stress (OFIT), although both devices were stressed for the *same effective* stress time. In fact, the increase of the CP current measured 40 ms after terminating stress is 349 pA after constant bias stress but only 133 pA ($\approx$ 38 %) after pulsed stress, indicating that interface state creation proceeds much more efficiently under DC bias conditions.

Device (c) was stressed in the 'interrupted OFIT' mode, recording the CP current $\Delta I_{CP}^{OFIT}$ by sporadically interrupting the constant bias stress phase at -2.0 V by short (280 ms) lasting OFIT pulses. Within these 280 ms, three OFIT CP current values were recorded, the first one after 40 ms, the following ones after a sampling interval of 100 ms. During recovery, the CP currents $\Delta I_{CP}^{MSM}$ were monitored by interrupting the constant bias recovery phase at -0.75 V occasionally by short (280 ms lasting) CP

MSM measurements (pulsing the gate junction between -0.75 V and +0.75 V). Within these 280 ms, three CP MSM CP current values were recorded, the first one after 40 ms, the following ones after a sampling interval of 100 ms. The interruption cycles during stress are performed in the OFIT mode, hence, they are supposed to be recovery free after [64]. Note that the overall degradation is much larger as opposed to the 'pulsed OFIT' mode yielding a similar CP current at the end of interrupted stress as obtained after DC stress in (a). During stress, we obtain the following power-law coefficients: uncorrected data: $n_{CP}^{str}$ = 0.21; corrected data: $n_{CP}^{str}$ = 0.27 consistent with many studies dealing with degradation dynamics of the interface [8,10,12]. The recovery rate ($B_R^{CP}$) measured in the 'interrupted CP MSM' mode yields -30 pA/dec which is again considerably smaller than the one obtained during 'pulsed CP MSM' in experiment (a) although the degradation level is similar at the end of stress. This indicates that interface state recovery proceeds more readily when pulsing the gate junction continuously during recovery.

Device (d) was stressed in the 'interrupted CP MSM' mode, recording the CP current $\Delta I_{CP}^{MSM}$ as a function of the stress time by interrupting the constant bias stress phase at -2.0 V by short (280 ms) lasting CP MSM measurements. The procedure is similar as the one described in (c), however, this time the occasional measurements of the CP current during stress are performed in the CP MSM mode introducing possibly some additional recovery, since the gate junction is not pulsed toward the stress voltage during CP read-out. The subsequent recovery phase is performed identically as in (c) by interrupting the constant bias recovery phase at -0.75 V sporadically by short (280 ms lasting) CP MSM cycles (pulsing the gate junction between -0.75 V and +0.75 V). Although enhanced recovery might be involved as a consequence of an active stress interruption, the technique has considerable advantages over the OFIT method since it does not require any corrections on the CP current measured during the stress phase. The degradation and recovery dynamics are very similar as the ones obtained for the corrected data in experiment (c) ($n_{CP}^{str}$ = 0.25; $B_R^{CP}$ = -39 pA/dec) indicating that gate pulsing in the 'CP MSM' mode does not imply significantly more recovery than gate pulsing in the OFIT mode.

In order to check whether and how much recovery is involved in OFIT and/or CP MSM, we may analyze the evolution of the CP currents measured during the 280 ms stress interruption phases of the 'interrupted OFIT' and the 'interrupted CP MSM' experiment. The results are illustrated in Fig. 3.5.

Remarkably, both OFIT and CP MSM are affected with similar interface state recovery within the 280 ms lasting interruption phases of the constant bias stress challenging the proposal that the OFIT technique is actually recovery free. This statement holds at least for the used gate pulsing frequency of 2 MHz and a duty cycle of 50 %. Applying this setup, the maximum time in accumulation where recovery may proceed during OFIT is only 250 ns within a single pulse periode. From this result we conclude that interface state relaxation is accelerated by gate pulsing toward accumulation. Note that the recovery rates ($B_R^{CP}$) increase with the overall amount of degradation.

On the other hand, interface state relaxation is suppressed very efficiently when maintaining the gate bias constantly in inversion during recovery. To demonstrate this, we fade out the constant bias phases of the 'interrupted CP MSM' measurement and string together the CP currents measured during the 280 ms interruption phases. The results are illustrated in Fig. 3.6.

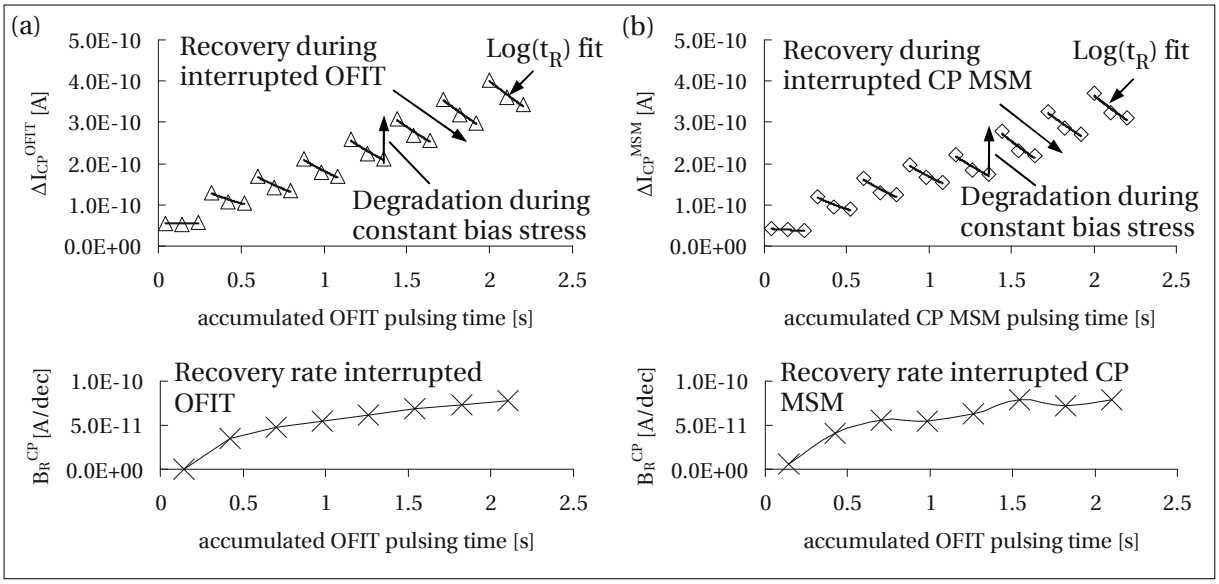


**Figure 3.5:** The evolution of the CP current as a function of the accumulated pulsing time during interrupted OFIT (a) and 'interrupted CP MSM' (b) stress (upper figures). The constant stress bias phases between the 280 ms CP intervals are omitted in the graphs. The lower illustrations show the recovery rates ($B_R^{CP}$) extracted from log($t_R$) fits (solid lines in the upper figures) of the original data.

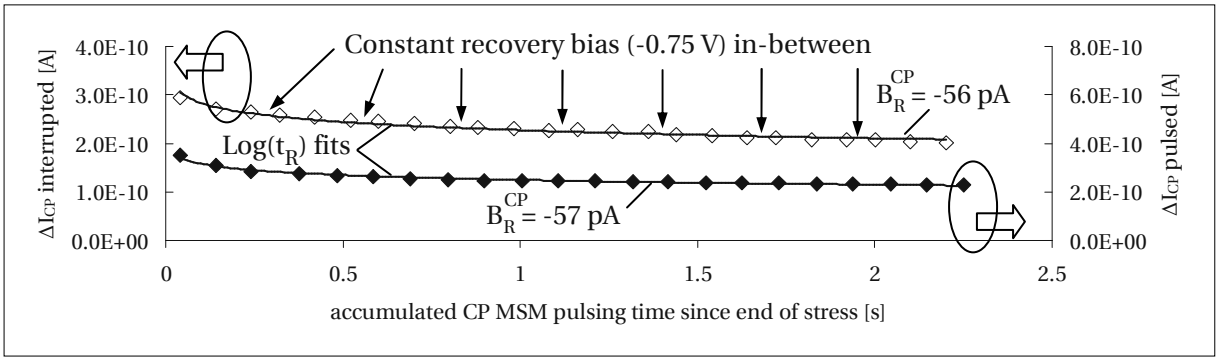


**Figure 3.6:** The evolution of the CP current as a function of the accumulated pulsing time during interrupted 'CP MSM' (open diamonds). For illustration purposes, the constant bias recovery phases at -0.75 V between the 280 ms CP intervals were omitted. As a reference, the CP recovery within the first 2.2 s recorded in the 'pulsed CP MSM' (full diamonds) mode has been included, cf. Fig. 3.4 (a).

Following Fig. 3.6 (open diamonds), we obtain that CP current recovery proceeds efficiently only during the gate pulsing periods. Within the intermediate constant bias phases at -0.75 V, interface state relaxation is almost negligible. In particular, when ignoring the constant bias recovery phases in-between completely, one obtains similar recovery characteristics ($B_R^{CP}$ = -56 pA/dec) as measured during pulsed CP MSM (cf. Fig. 3.4 (a); $B_R^{CP}$ = -57 pA/dec) where the gate is pulsed continuously (full diamonds) after stress. Hence, the overall recovery rate associated with interface state relaxation crucially depends on the way how the recovery cycle is performed. It is maximal when continuously pulsing the gate junction and becomes increasingly lower with increasing the ratio of the constant bias phases in-between. An exception to the rule is the 'pulsed OFIT' mode (cf. Fig. 3.4 (b)) where

$B_R^{CP}$ is small (-11 pA/dec) although the gate has been pulsed continuously during recovery. This may be explained by the lower overall degradation level at the end of 'pulsed stress' which may be affected already by a certain amount of recovery before switching to 'pulsed CP MSM' initiating the actual relaxation cycle.

In order to highlight the independence of the interface state dynamics on the oxide thickness and material, similar experiments have been performed on 30 nm $SiO_2$ PMOS devices equipped with $n^{++}$ gate poly (SM6P/30/H1). The HV device was stressed for an effective stress time of 100 s at 125°C using a stress voltage of -20.0 V (oxide field approximately 6.3 MV/cm). Following the stress phase, a 100 s recovery phase was recorded. The same four stress-relax modes discussed in Fig. 3.3 were performed on four HV devices (SM6P/30/H1), having, however, a different geometry (gate oxide area) as the previously discussed high–$\kappa$ devices (HK2P/1.5/1). The results for the HV devices are illustrated in Fig. 3.7: left hand side – linear scale; right hand side – logarithmic scale. During OFIT, the CP current is measured by pulsing the gate junction between +2.0 V and -20.0 V recording in parallel the CP current $I_{CP}^{OFIT}$. During CP MSM the gate bias was pulsed between +2.0 V and -2.0 V recording in parallel CP current $I_{CP}^{MSM}$. We have again used a symmetrical pulse shape with a duty cycle of 50 %, a gate pulsing frequency of 125 kHz and rising/falling slopes of 20 V/$\mu$s. Due to the much larger pulse amplitude during stress (22.0 V), a lower gate pulsing frequency had to be applied in order to adjust similar rising/falling slopes as applied to the high–$\kappa$ device (HK2P/1.5/1) discussed previously.

By comparing the results of Fig. 3.7 to the results of Fig. 3.4, we obtain very good qualitative agreement indicating that oxide thickness and material do not play a significant role in NBTI degradation and recovery dynamics. Although the obtained power-law exponents of the $SiO_2$ device (SM6P/30/H1) and the high–$\kappa$ device (HK2P/1.5/1) differ slightly, the overall good correlation indicates that the basic physical mechanisms behind NBTI are essentially the same when comparing thin high–$\kappa$ and thick $SiO_2$ devices. The small discrepancies (larger power-law exponent $n_{CP}^{STR}$ and lower recovery rate $B_R^{CP}$ for the thick oxide devices) may be attributed to the different pulsing frequencies used to characterize OFIT and CP MSM and naturally also to the fact that the tested wafers were fabricated in different factories having undergone a completely different process flow.

## 3.3 Duty cycle dependence of NBTI degradation

As indicated in the previous chapter, recovery leads to less severe device degradation under AC operation compared to DC operation. This considerably prolongs the lifetime of MOSFETs in digital circuits where the conventional static NBTI stress may lead to a significant underestimation of the NBTI lifetime [66–68]. In literature, the recovery effect occurring during the stress interruption phases is either attributed to interface state passivation [66] or to hole emission from previously charged oxide defects [8]. For instance, Chen *et al.* [66, 69] assumes that during stress, the interface acts as a hydrogen source injecting hydrogen from Si–H bonds into the oxide after bond distortion while during recovery the interface acts as a hydrogen sink recapturing hydrogen which diffuses immediately back to the interface after terminating stress. Following [70], the power-law exponent is reported to be lower for AC than for DC operation at least at sufficient large stress times.

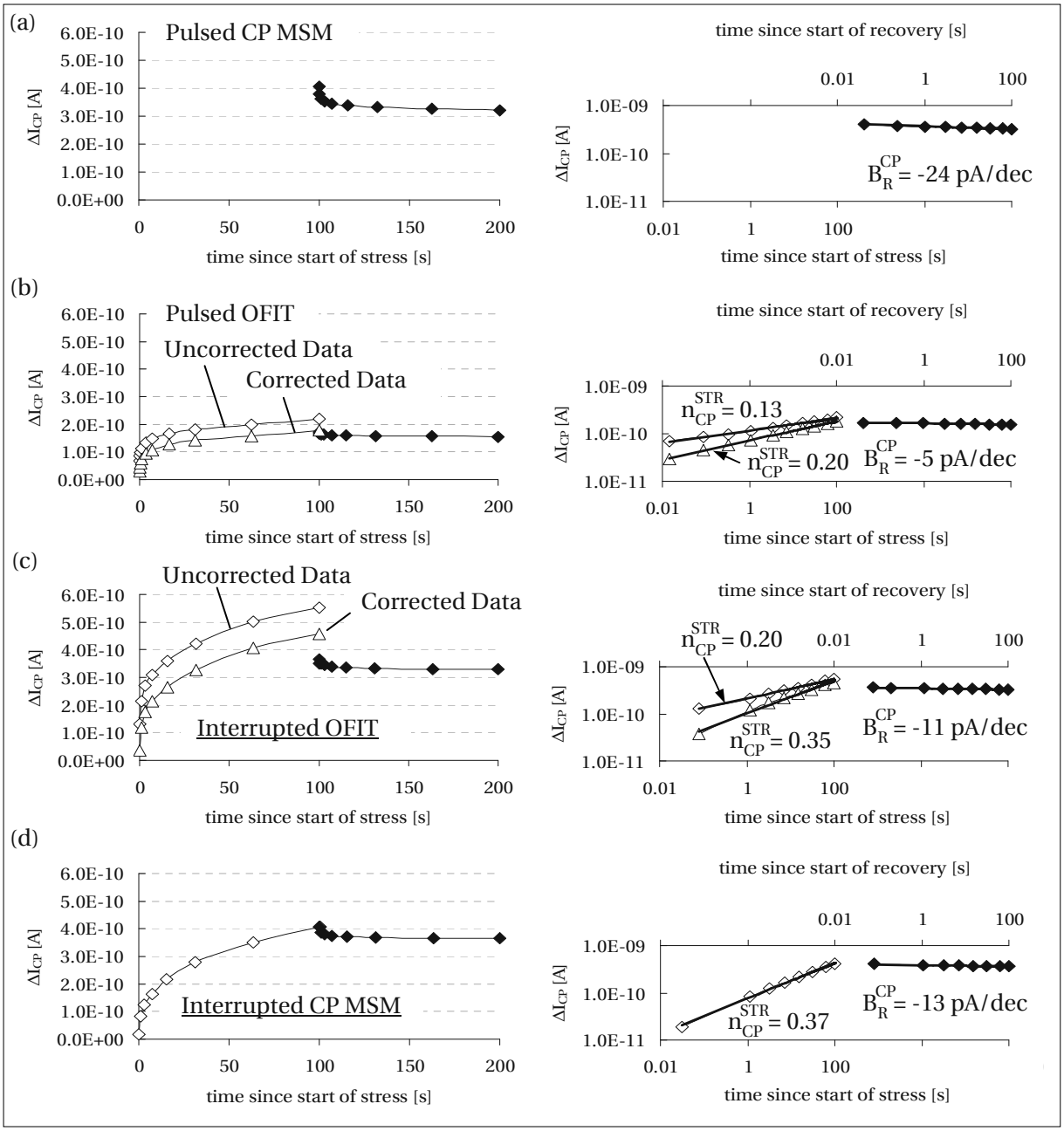


**Figure 3.7:** Degradation (open symbols) and recovery (full symbols) of the CP current measured in different CP modes for 100 s at 125°C on different 30 nm $SiO_2$ PMOS devices (SM6P/30/H1). On the left hand side the data is plotted in a lin-lin scale, on the right hand side in a log-log scale. Open diamonds correspond to uncorrected data, open triangles correspond to OFIT data corrected by gate leakage currents and oxide charges. Full diamonds illustrates CP current recovery measured in the 'CP MSM' mode. Device (a) was stressed in the 'pulsed CP MSM' mode, device (b) in the 'pulsed OFIT' mode, device (c) in the 'interrupted OFIT' mode and device (d) in the 'interrupted CP MSM' mode.

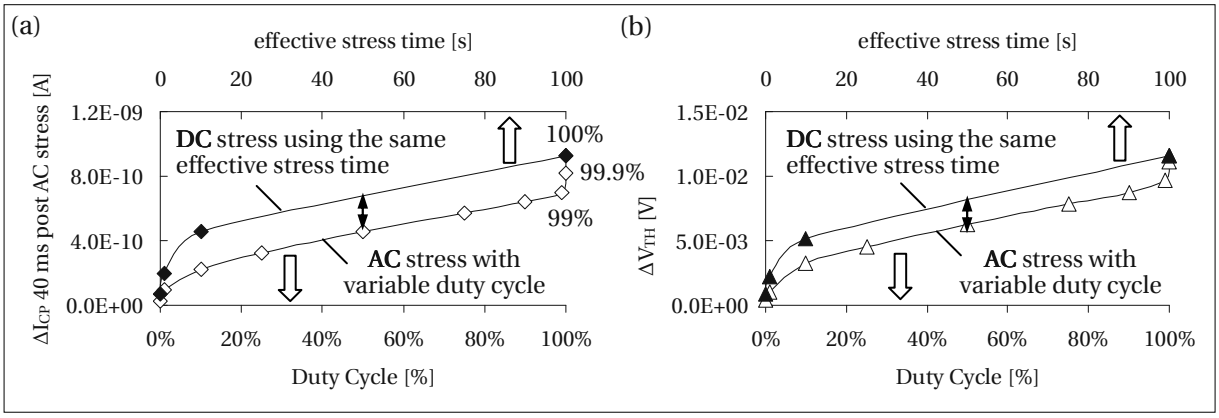


**Figure 3.8:** Duty cycle dependence of the CP current (a) and the $V_{TH}$ (b) degradation (open symbols). The full symbols correspond to a related DC stress experiment performed on separate devices using equivalent effective stress times. A hysteresis emerges between the AC stress and the DC stress procedure accounting for a reduced degradation as a consequence of stress interruption.

The dynamic NBTI effect was found to depend on temperature and gate oxide thickness but not on the stress frequency. The frequency independence was confirmed by several studies for the following band widths: 0.5 – 100 kHz by [69]; 2 – 20 MHz by [70]; 1 Hz – 2 GHz by [71]; 10 Hz – 2 GHz by [72].

As opposed to the pulsing frequency, the gate voltage duty cycle ('ON time' divided by 'total time') is reported to considerably influence the observed amount of degradation at the end of stress [70–72]. Most studies report a monotonic evolution of the degradation, with an apparent plateau around a duty cycle of 50 %.

In order to investigate the duty cycle dependence of NBTI, several PMOS devices (SM6P/30/H1) were stressed at a frequency of 1 kHz for 100 s (100,000 pulses) at 125°C with variable duty cycle ranging from 0.1 % to 100 %, cf. Fig. 3.8 (open symbols). During AC stress, the gate junction was pulsed between 0.0 V and -20.0 V corresponding to an effective oxide field of approximately 6.3 MV/cm during the high phase of the gate pulses. During read-out, the gate junction was pulsed at a frequency of 500 kHz (rising/falling slopes 20 V/$\mu$s; duty cycle 50 %) between +2.0 V and -2.0 V in order to record the maximum CP current. Immediatly after terminating stress, the maximum CP current was monitored with a minimum stress-measure delay of 40 ms, cf. Fig. 3.8 (a) (open diamonds). Subsequently to the CP measurement (total duration 1 s), the $\Delta V_{TH}$ degradation was recorded by monitoring the degraded drain current at -1.1 V corresponding approximately to the threshold voltage of the device, cf. Fig. 3.8 (b) (open triangles). Note that the effective stress time in seconds ($t_S^{eff}$) reflects the duty cycle in [%] when stressing the device for 100 s under AC bias conditions, cf. Eq. 3.7.

As can be seen in Fig. 3.8 (a) and (b), as a function of the duty cycle both the CP current and the $V_{TH}$ shift show a 'S'-like curve shape with a steep increase at low duty cycles (0.1 % – 10 %), a monotonic development at intermediate duty cycles (10 % – 90 %) and another steep increase at high duty cycles (90 % – 100 %). The results agree with the study of Fernández *et al.* [71] who applied AC stress pulses with a frequency of 10 kHz to 1.4 nm silicon oxynitride (SiON) gate dielectrics and with the results of

Huard *et al.* [72] who applied a stress pulsing frequency of 100 kHz to 1.7 nm plasma nitrided oxide (PNO) gate dielectrics. Both referenced authors performed their experiments at 125°C as well.

The first increase at low duty cycles in Fig. 3.8 (a) and (b) may be explained by the sharply rising effective stress times. The second increase at high duty cycles cannot be explained by this argument since the effective stress time changes only by 10 % when applying AC stress with duty cycles between 90 % and 100 %. On the other hand, when increasing the duty cycle, the effective relaxation periods between the stress intervals decrease, providing less time for recovery during the repated stress interruption phases. Hence, the appearance of the second increase suggests that it is important whether the entire stress phase is applied completely uninterrupted (DC) or interrupted (AC).

To elaborate this assumption in more detail, a second set of measurements was performed, where we have stressed separate PMOS devices having the same geometry for an 'equivalent' DC stress time. For example, stressing a device under AC bias conditions for 100 s using a duty cycle of 10 % correponds to an effective stress time of 10 s which is compared in the following to a 10 s lasting DC stress experiment. The results gained from these 'equivalent' DC experiments are illustrated as full symbols in Fig. 3.8 (a) and (b). When comparing the AC stress runs to the DC stress runs applied for the same effective stress times, a hysteresis emerges, representing a reduced degradation as a consequence of stress interruption and possibly involved recovery. Note that not before the recovery periods fall below 10 $\mu$s (duty cycle 99 %) the hysteresis disappears, giving then similar results as during DC stress. The experiment demostrates that interrupting the stress phase by only 10 $\mu$s or even less is already sufficient to influence the degradation dynamics considerably, which is consistent with the observation that OFIT is afflicted with recovery even when using a 2 MHz pulsing frequency, cf. Fig. 3.5.

Note that the CP results (cf. Fig. 3.8 (a)) agree very well with the $\Delta V_{TH}$ shifts (cf. Fig. 3.8 (b)) calculated from the drain current degradation (they are proportional) indicating a strong correlation between interface state creation and threshold voltage shift. Whether the obtained threshold voltage shifts can be explained completely by the existence of charged interface traps remains an open question which is going to be addressed in the following chapters.

In order to study the influence of the stress interruption time ($t_l$) on degradation and recovery dynamics, we have performed additional AC stress experiments on different devices at 125°C. In this particular set of experiments we keep the low level recovery times ($t_l$) of the gate pulses (0.0 V) constant, while the high level stress times ($t_h$) of the gate pulses (-20.0 V) are increased always by a factor 10 after an effective stress time of 100 s has elapsed. For example, beginning with a high level time of $t_h = 10^{-7}$ s and a constant low level time of $t_l = 10^{-6}$ s = 1 $\mu$s (duty cycle 10 %), the first stress run is performed for 1,000 s, corresponding to an effective stress time of 100 s. The second stress run is performed with $t_h = 10^{-6}$ s and $t_l = 10^{-6}$ s (duty cycle 50 %) for 200 s, corresponding again to an effective stress time of 100 s and to a total effective stress time of 200 s. The following stress runs (each lasting for an effective stress time of 100 s) are then performed serially on the same device with increasing $t_h$ until a total effective stress time of 1,000 s is reached (last high level time $t_h = 10^2$ s). At the end of each stress run, a CP measurement was performed for 1 s and subsequently a drain current measurement at -1.1 V in order to monitor the stress and recovery induced evolution of interface state creation and $V_{TH}$ shift.

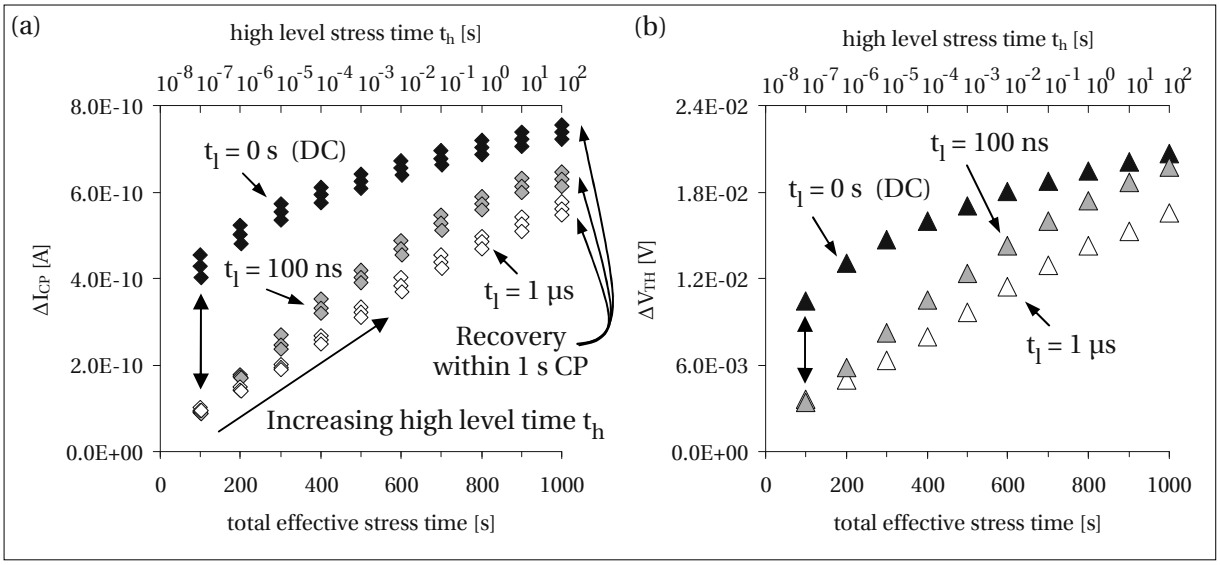


**Figure 3.9:** Evolution of the CP current (a) and the threshold voltage (b) degradation when performing AC stress with increasing stress intervals ($t_h$) and constant recovery intervals at 0.0 V ($t_l$). The full black symbols correspond to a DC stress without stress interruption phases ($t_l$ = 0 s). The gray shaded symbols correspond to an AC stress with $t_l$ = 100 ns and open symbols correspond to an AC stress with 1 $\mu$s. After an effective stress time of 100 s has elapsed, the CP current is monitored for 1 s giving three points at 40 ms, 100 ms and 1 s (a). After the CP measurement, the drain current degradation is monitored in order to determine $\Delta V_{TH}$ (b). 10 stress runs are performed on each device, thereby increasing the high level stress phase ($t_h$) always by a factor of 10 from one stress run to the other, cf. upper x-axis.

The experiment was performed on three different devices using different recovery intervals ($t_l$) of 0.0 s (DC), 100 ns and 1 $\mu$s. The degradation of the CP current and the increase of the threshold voltage shift are illustrated in Fig. 3.9 (a), and (b) respectively, as a function of the high level stress time ($t_h$) (upper x-axis) and as a function of the effective stress time (lower x-axis) for the three different recovery intervals ($t_l$).

From Fig. 3.9 (a) it becomes obvious that a stress interruption phase of only 100 ns is already sufficient to suppress degradation considerably. A huge gap (indicated by the double arrows) arises between the DC and the AC stress experiments which becomes narrower with increasing stress intervals in comparison to the constant recovery intervals (increasing duty cycle). A second remarkable detail of Fig. 3.9 (a) regards the CP recovery measured for 1 s after each complete stress run (visualized by three measurement points). Indeed, recovery is not observed before the stress interval exceeds the recovery interval by at least a factor 100 indicating a correlation between the degradation and the recovery rate. For instance, the device, subjected to AC stress with a low phase of 100 ns, begins to recovery not before the high level interval exceeds 10 $\mu$s. On the other hand, the device, subjected to AC stress with a low phase of 1 $\mu$s, begins to recover not before the high level interval exceeds 100 $\mu$s. Note that the total amount of recovery obtained for DC stress (full symbols) is always the same independently of the effective stress time and the degradation level. Again very good qualitative agreement is obtained between the CP results depicted in Fig. 3.9 (a) and the $V_{TH}$ shifts illustrated in Fig. 3.9 (b).

Based on the AC stress experiments discussed in this chapter we conclude that degradation is most efficient under DC bias conditions. The longer the stress bias is applied without interruption, the larger is the amount of degradation at the end of stress. Consequently, it is not the *effective stress time* that governs degradation but the *time interval* where the stress bias is continuously applied. Interrupting the stress phase by only 100 $\mu$s turned out to be already sufficient to influence the degradation dynamics considerably. Furthermore, a correlation between the time interval of uninterrupted stress and CP current recovery was found. When applying AC stress, CP current recovery does not proceed before the high level time exceeds the low level time by at least a factor of 100. Good qualitative agreement between the CP current increase and the threshold voltage shift was demonstrated for various AC stress conditions indicating that interface state creation plays a crucial role for the NBTI effect.

## 3.4 Temperature and oxide field – the driving forces of NBTI degradation

Having exemplary investigated in the previous sections the basic characteristics of time and duty cycle dependence of degradation and recovery for a singe stress temperature (125°C) and a single electric field (6.3 MV/cm), we are going to extend the study in the following to different stress conditions in order to draw fundamental conclusions on the temperature and field acceleration of NBTI. Following Eq. 3.1 and Eq. 3.2, the crucial parameters necessary to fully describe the degradation branch ($\Delta V_{TH}(t_S)$) are the pre-factor $A_S(E_{OX,S}, T_S)$ corresponding to the $V_{TH}$ shift at $t_S$ = 1 s and the power-law exponent $n_S$ representing the slope of the degradation traces in a double-logarithmic plot $\Delta V_{TH}(t_S)$ vs. $t_S$. On the other hand, the recovery branch may be fully described by the recovery rate $B_R(V_{GR}, T_R)$ alone, provided the relaxation traces follow a logarithmic time dependence ($\Delta V_{TH}(t_R) \propto B_R \log_{10}(t_R)$).

In this section we study only the dynamics of $V_{TH}$ degradation and recovery calculated from drain current variations recorded at a gate voltage close to the threshold voltage of the device. In particular, we abstain from analyzing the evolution of the CP current for different stress temperatures. Since the energy range, profiled during a CP measurement ($\Delta E_{CP}$), depends on the analyzing temperature (cf. Eq. 2.35), it could eventually lead to misleading conclusions, when comparing CP currents recorded at different temperatures. Consider that so far the simple MSM procedure is constrained to the limitation that the stress temperature ($T_S$) has to equal the recovery temperature ($T_R$). A more elaborate procedure allowing to overcome this constraint is presented in Chapter 6 which also includes a detailed discussion of the temperature dynamics of the CP current.

All experiments discussed in this section are performed on different PMOS devices (SM6P/30/STD1). Due to the fact that the static OTF procedure is unfeasible on 30 nm HV devices and because of other drawbacks (discussed in Section 3.1), we record the degradation dynamics by making use of the MSM technique, thereby accepting some 'undefined' recovery when switching from the stress level ($V_{GS}$) to the read-out bias (-1.1 V). The MSM procedure is performed by interrupting the DC stress phase once in a while for a short interval of time where the drain current is recorded with a stress-measurement delay is 10 ms at -1.1 V which is close to the threshold voltage of the device. After a total stress time of 6,000 s has elapsed, a continuous recovery trace is measured for 1,000 s in order to collect information on the recovery dynamics as well.

### 3.4.1 The role of the stress/recovery temperature

The first set of measurements addresses the temperature activation of the $V_{TH}$ shift and recovery. For that seven devices were stressed at the same oxide field of 5.5 MV/cm applying different stress/recovery temperatures ranging from -60°C to 200°C, which corresponds to the full temperature range of our thermo chuck.

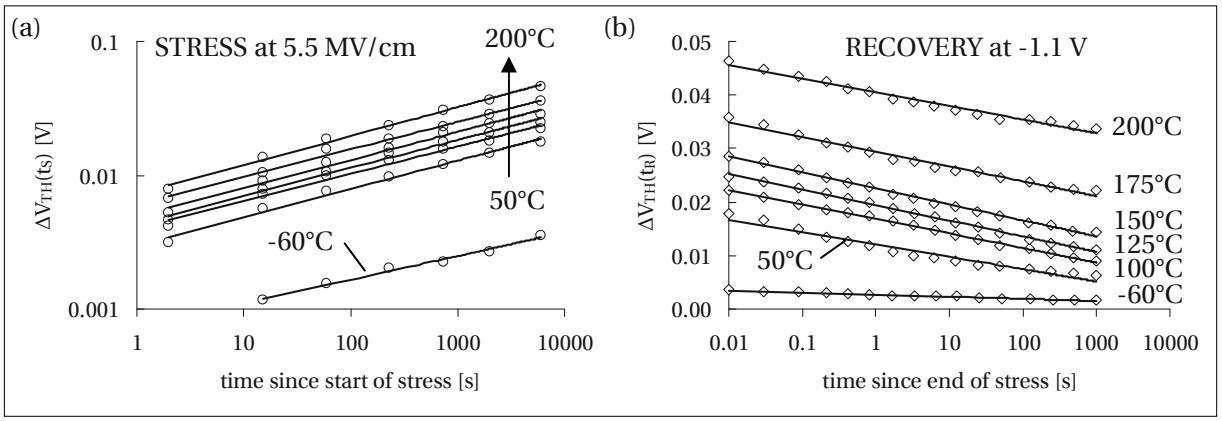


**Figure 3.10:** Threshold voltage degradation (a) and recovery (b) recorded for different stress and recovery temperatures ranging from -60°C to 200°C (-60/50/100/125/150/175/200°C), the stress field being 5.5 MV/cm. The data points in (a) and (b) are depicted as open symbols. The degradation branches in (a) are illustrated in a double-logarithmic plot because their evolution follows a power-law as indicated by the solid fit lines. The recovery branches in (b) are illustrated in a semi-logarithmic time plot since their evolution follows a log($t_R$) dependence as indicated by the solid fit lines.

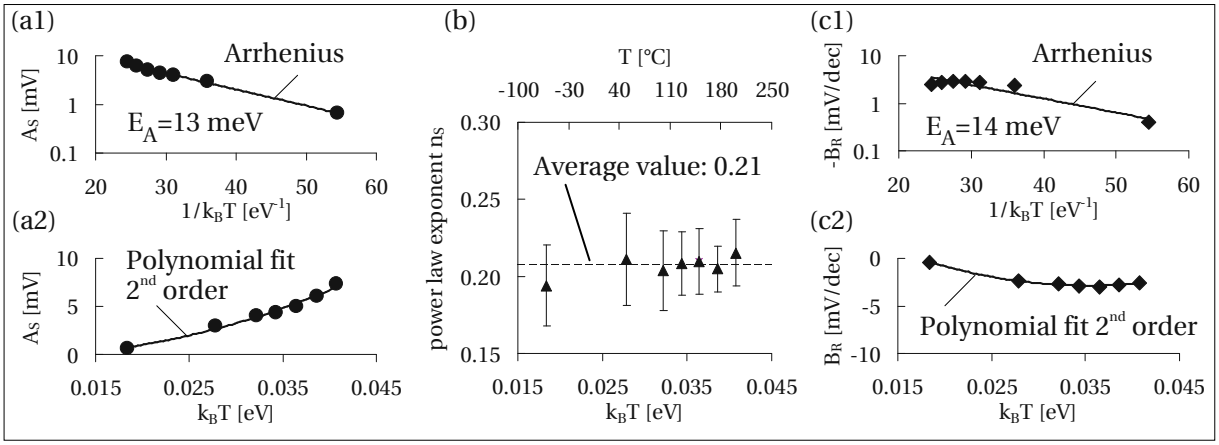


**Figure 3.11:** Temperature dependent variation of the representative parameters of Fig. 3.10. The values empirically describing the evolution of degradation are depicted in (a) (pre-factor $A_S$) and (b) (power-law exponent $n_S$). (a1) is an Arrhenius plot and (a2) is a linear plot of $A_S$. In (b) $n_S$ is found to be more or less independent of the temperature, the error bars indicating the scattering of $n_S$ caused by slight deviations of the data points from the perfect power-law-like evolution. The recovery rate $B_R$, which describes the relaxation branch, is depicted in (c). (c1) is an Arrhenius plot and (c2) is a linear plot of $B_R$.

The data illustrating $V_{TH}$ degradation and recovery are depicted in Fig. 3.10 (a), and (b), respectively. Fig. 3.10 (a) shows a power-law-like evolution of the degradation branch and Fig. 3.10 (b) a log-like evolution of the recovery traces. It is obvious that elevated temperatures cause larger degradation justifying the letter 'T' in NBTI.

For closer inspection, the data has been analyzed with respect to the temperature dependence of the coefficients describing the degradation ($A_S$ and $n_S$) and the recovery branch ($B_R$). The results are given in Fig. 3.11. Actually, the dependence of $A_S$ on T may be approximated either by an Arrhenius law $\propto \exp(-E_A/k_B T)$ with $E_A$ = 13 meV as suggested by Schroder *et al.* [10] and by Kaczer *et al.* [27, 73, 74] or by a polynomial fit of second-order. Following Kaczer *et al.*, an exponential relationship between the pre-factor $A_S$ and the stress temperature may be explained by considerations regarding disorder controlled hydrogen diffusion kinetics. On the other hand, following Bindu *et al.* [75], a polynomial T-dependence may be explained by inelastic oxide hole trapping, where the linear contribution is assumed to account for the temperature dependent enhancement of the surface hole concentration during stress and the quadratic contribution accounts for a multiphonon-emission process which is assumed to govern threshold voltage degradation predominantly at low stress times.

As opposed to the pre-factor $A_S$, the power-law exponent $n_S$ is found to be temperature independent which is clearly visible in Fig. 3.11 (b). The error bars consider the standard deviation of $n_S$ from the proposed perfect power-law-like evolution of the degradation branch. The small deviations are likely due to parasitic recovery occurring within the short time delay between removal of stress and the actual measurement (10 ms). It has to be emphasized that the observation of a temperature independent power-law exponent is in contradiction to some proposed NBTI models in literature, where it has been suggested that at least the dynamics of interface state creation are supposed to produce temperature dependent power-law exponents due to either dispersive hydrogen diffusion within the oxide [27, 73, 74] or (according to an alternative model) due to dispersive reaction kinetics [8] controlling hydrogen release from Si–H bonds at the interface. We remark that those conclusions were drawn either from MSM experiments suffering from a considerably larger stress-measure delay than ours' [12, 73] or from CP current measurements [8] which profile different ranges within the silicon bandgap when performed at different temperatures, cf. Fig. 2.8. In agreement with our findings, Alam *et al.* have demonstrated in [76] that the temperature dependence of the power-law exponent disappears (at least for thin oxide technologies) when applying the 'recovery free' OTF technique.

Fig. 3.11 (c1) and (c2) show the recovery rates $B_R$ as a function of T in an Arrhenius plot, and in a linear plot, respectively. Within the time scale of our experiment $B_R$ follows apparently a $\log(t_R)$ dependence. Except for the data recorded at -60°C, the recovery rates are found to be only weakly temperature dependent, yielding a universal recovery rate of approximately -3.0 mV/dec, similar to the one obtained for the high–$\kappa$ device (HK2P/1.5/1) in Fig. 3.2. The evolution of the pre-factor $B_R$ as a function of temperature may be described best by a second-order polynomial fit (cf. Fig. 3.2 (c2)). The Arrhenius-like exponential approximation gives unsatisfactory results. Note that the universal recovery rate of -3.0 mV/dec does not necessarily imply that recovery is generally temperature independent considering that the prior stress phase was performed at variable temperatures as well, leading to different points of origin of the individual recovery traces.

In order to study the phenomenon in a more sophisticated way, it would be highly expedient to have a technique available which allows to bring devices to the same degradation level (by stressing them under the same oxide field and temperature) but monitor their recovery at arbitrary relaxation temperatures. A tool being able to switch the device temperature with maximum precision within a minimum of time would have the power to overrule the so far strict constraint that the stress temperature has to equal the recovery temperature. Such a tool was found in the so-called 'in-situ polyheater' technique. The implementation, calibration and application of particularly designed polyheater was one of the main achievements of this PhD thesis, allowing to investigate the influence of temperature on NBTI degradation and recovery in an unprecedented manner. The technique is introduced and utilized in Chapter 6.

### 3.4.2 The role of the stress field

The second set of measurements addresses the field activation of $V_{TH}$ degradation and its implication on recovery. Again, seven different PMOS devices (SM6P/30/STD1) were stressed, however, this time a unique temperature of 100°C was applied, but varying the stress field from 4.5 MV/cm to 7.2 MV/cm. The data illustrating $V_{TH}$ degradation and recovery are depicted in Fig. 3.12 (a), and (b), respectively.

As already obtained previously for different stress temperatures, Fig. 3.12 (a) reflects a power-law-like evolution of the degradation branch for different stress fields as well and Fig. 3.12 (b) the $\log(t_R)$ like evolution of the recovery traces. An exception to the rule in found for the 7.2 MV/cm data which begins to deviate considerably from the power-law-like characteristic after 1,000 s of stress, indicating the onset of a different degradation mechanism. At elevated stress biases (fields) a larger amount of degradation is observed, justifying the letter 'B' in NBTI. An important difference to the temperature characteristics in Fig. 3.10 regards the shape of the recovery traces recorded at -1.1 V after electrical stress at different fields. As opposed to the recovery characteristics obtained for different temperatures, the recovery traces recorded at different stress fields become considerably steeper with increasing stress field. For example, the '175°C at 5.5 MV/cm' data in Fig. 3.10 (b) and the '100°C at 6.5 MV/cm' data in Fig. 3.12 (b) display similar degradation levels ($\approx$ 40 mV) 10 ms post stress, but recover with completely different relaxation rates (175°C at 5.5 MV/cm: $-2.8$ mV/dec;100°C at 6.5 MV/cm: $-4.4$ mV/dec). The different relaxation behavior indicates that elevated stress temperatures may generate defects providing much larger relaxation time constants than defects generated by elevated electric fields. Those defects created at high stress temperatures (175°C) and moderate electric fields (5.5 MV/cm) obviously do not recover significantly within our experimental time scale, although the recovery temperature itself is as large as well. This is a fundamental finding raising the question whether possibly more than one single defect type is involved in NBTI.

For closer inspection, the data has been analyzed with respect to the field dependence of the coefficients describing the degradation ($A_S$ and $n_S$) and the recovery branch ($B_R$). Fig. 3.13 (a) reveals a considerable field dependence of the initial degradation rate ($A_S$). The full symbols reflect the analysis of the 100°C data whereas the open symbols show the 200°C data (not given in Fig. 3.12).

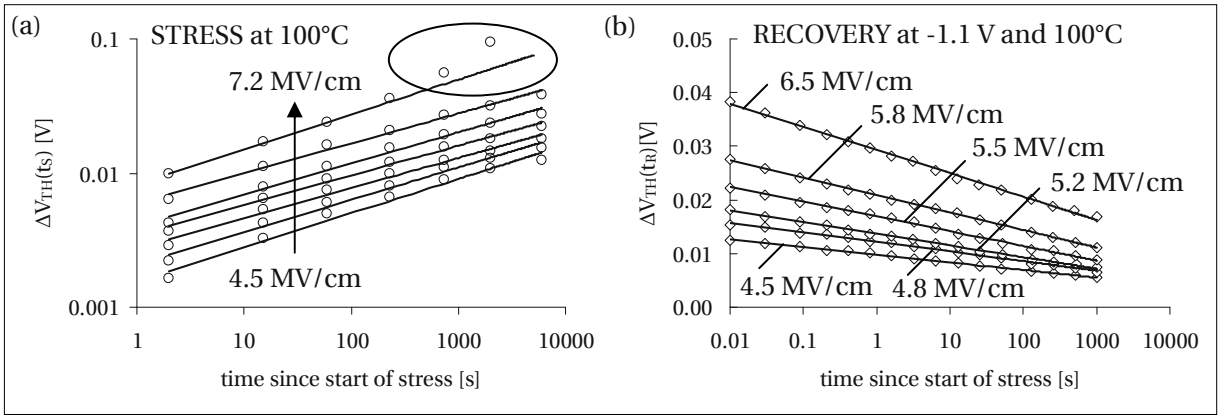


**Figure 3.12:** Threshold voltage degradation (a) and recovery (b) recorded at a temperature of 100°C for different stress fields ranging from 4.5 MV/cm to 7.2 MV/cm (4.5/4.8/5.2/5.5/5.8/6.5/7.2 MV/cm). The data points in (a) and (b) are depicted as open symbols. The degradation branches in (a) are illustrated in a double-logarithmic plot because their evolution follows a power-law as indicated by the solid fit lines. An exception to the rule is the 7.2 MV/cm data which begins to deviate considerably from the power-law-like characteristic after 1,000 s of stress. The recovery branches in (b) are illustrated in a semi-logarithmic time plot since their evolution follows a $\log(t_R)$ dependence as indicated by the solid fit lines.

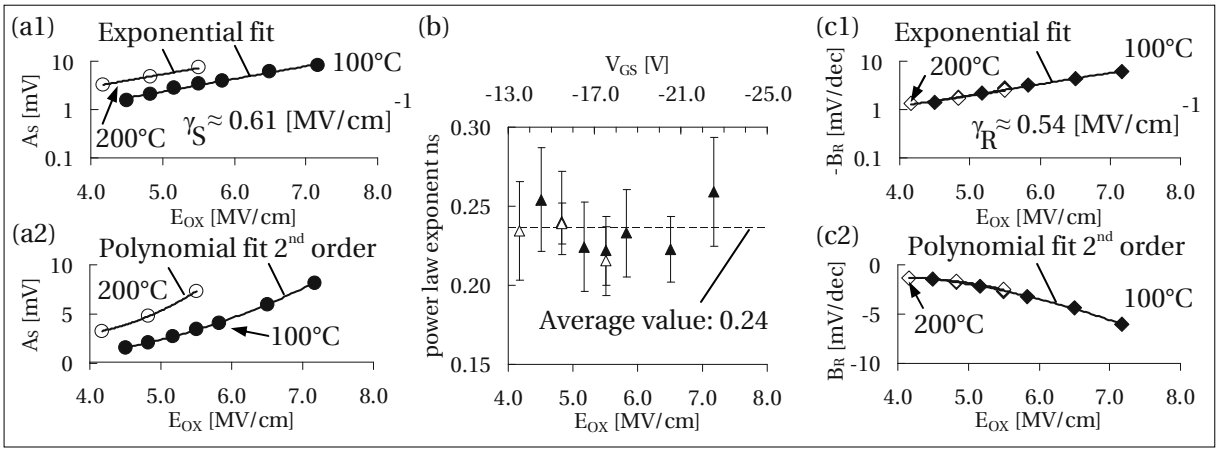


**Figure 3.13:** Field dependent variation of the representative parameters of Fig. 3.12 (100°C; full symbols). Additional data for 200°C (not given in Fig. 3.12) was added (open symbols). The values empirically describing the evolution of the power-law-like degradation are depicted in (a) (pre-factor $A_S$) and (b) (power-law exponent $n_S$). (a1) is an exponential plot and (a2) is a linear plot of $A_S$. In (b) $n_S$ is found to be more or less independent of the stress field, the error bars indicating the scattering of $n_S$ caused by slight deviations of the data points from the perfect power-law-like evolution. The recovery rate $B_R$, which describes the relaxation branch, is depicted in (c). (c1) is an exponential plot and (c2) is a linear plot of $B_R$.

Again, the development of $A_S$ with $E_{OX}$ may be approximated either in an exponential manner (Arrhenius) $\propto \exp(\gamma E_{OX})$ with $\gamma_S = 0.61\,[\mathrm{MV/cm}]^{-1}$ as suggested by Schroder *et al.* [10] or by a polynomial fit of second-order as suggested by Grasser *et al.* and Bindu *et al.* [53, 75]. Note that the 'electric

field factor' $\gamma_S$ in Fig. 3.13 (a1) is identically for 100°C and 200°C. The polynomial field dependence in Fig. 3.13 (a2) may be explained when assuming inelastic hole trapping during the early stages of stress: The quadratic $E_{OX}$ dependence is then reflected by the $E_{OX}^2$ field dependence of the multiphonon-emission process while the linear contribution accounts for the linear field dependence of the surface hole concentration.

Similar to the observations at different stress temperatures, the power-law exponent $n_S$ is found to be widely field-independent which is clearly visible in Fig. 3.12 (b) (full symbols 100°C; open symbols 200°C) and consistent with the bulk of literature [8, 76–78]. The result is, however, in contradiction to the work of Reisinger *et al.* [15] who reported a decrease of $n_S$ for similar thick oxide devices (7 – 15 nm) in the stress field range of 3.0 MV/cm to 6.0 MV/cm.

Fig. 3.13 (c1) and (c2) show the recovery rates $B_R$ as a function of $E_{OX}$ in an Arrhenius plot, and in a linear plot, respectively. As opposed to the variable temperature experiment, the recovery rates recorded after NBTS at different stress fields are found to be exponentially stress bias dependent giving a similar temperature independent 'electric field factor' $\gamma_R$ of 0.54 $[\mathrm{MV/cm}]^{-1}$ as obtained during stress ($\gamma_S = 0.61\,[\mathrm{MV/cm}]^{-1}$), cf. Fig. 3.12 (c1). The recovery rates may be approximated by a polynomial fit of second-order as well, cf. Fig. 3.12 (c2). The larger the stress bias, the steeper the recovery slope, indicating that a larger amount of defects having time constants within the time scale of our experiment are activated during stress, and anneal during recovery, respectively. The evolution of $B_R$ shows apparent similarities to the evolution of $A_S$, suggesting a symmetry of stress and relaxation when investigating NBTI at different stress fields but at a single stress/relaxation temperature. This symmetry is violated when introducing a different stress temperature (i.e. 200°C), cf. open symbols). While the recovery rate $B_R$ remains independent of temperature, the degradation rate $A_S$ shows an offset at elevated stress temperatures. This indicates that there might be a quasi-permanent contribution of degradation which becomes activated during stress but does not recover within the time scale of our experiment.

### 3.4.3 Recovery saturation

In the previous subsections the recovery branch was found to follow a $\log(t_R)$ dependence with $B_R$ being the recovery rate per decade. This behavior was, however, so far only demonstrated for the cases where the stress time considerably exceeds the recovery time ($t_R \ll t_S$). To elaborate whether the recovery changes its characteristic time dependence for $t_R \gg t_S$ and whether a permanent offset remains after long relaxation times is subject of this subsection.

In order to address this question, we have stressed a PMOS device at 200°C and 5.5 MV/cm for a short time of only 1 s and subsequently monitored its recovery for 50,000 s at -1.1 V. The result of this experiment is illustrated in Fig. 3.14. Following Fig. 3.14, we clearly observe two different recovery rates, indicating the presence of at least two different kinds of defects with considerably different relaxation time constants. The first is dominant within a 10 s lasting period after terminating stress ($B_R^A$) and has a similar magnitude as obtained in previous experiments (-1.8 mV/dec). Between 10 s and 1,000 s the recovery slope levels off considerably toward a value $B_R^B$ of only -0.08 mV/dec.

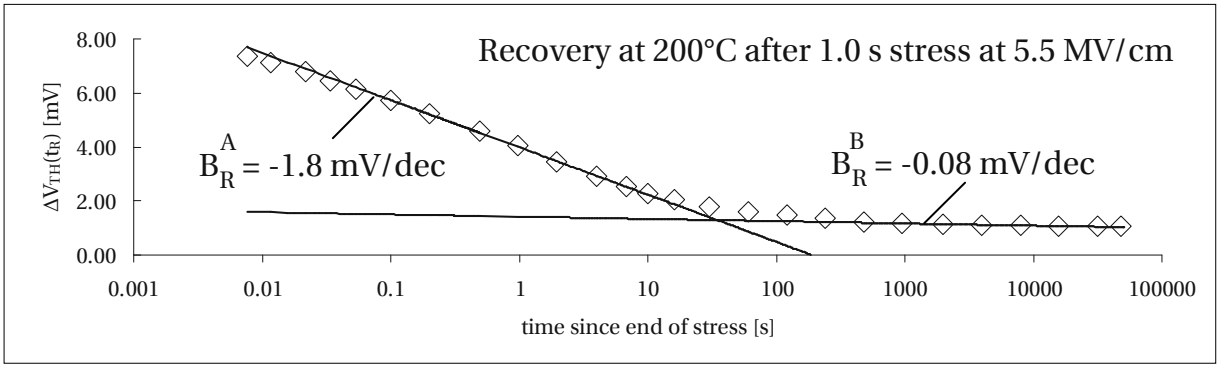


**Figure 3.14:** Long-term recovery after stressing the device for 1 s at 200°C and 5.5 MV/cm. Within the first 10 s after terminating stress, the $V_{TH}$ recovery follows a regular $\log(t_R)$ dependence with a recovery rate of $B_R^A$ = -1.8 mV/dec. Between 10 s and 1,000 s the recovery rate levels off considerably toward $B_R^B$ = -0.08 mV/dec.

Complete stabilization of the threshold voltage shift could not be obtained even after 50,000 s of recovery. A final, 'quasi-permanent' degradation plateau of approximately 1.0 mV remains. We call the remaining degradation plateau 'quasi-permanent' since it is not completely constant but shows considerable slower relaxation dynamics compared to the initial steep recovery branch.

# 4

# Point defects and their correlation to hydrogen

## 4.1 Terminology and defect classes

DEFECT research in crystalline and amorphous $SiO_2$ trace back to 1956 when Robert A. Weeks of Oak Ridge National Laboratory [79] first used electron paramagnetic resonance (EPR) to study radiation-induced defects in quartz. In 1978, a committee was established by the Electronics Division of the Electrochemical Society and the IEEE-sponsored Semiconductor Interface Specialists Conference in order to clarify the terminology of defects associated with reliability issues in thermally oxidized MOS devices. Following Bruce E. Deal [80, 81], the Chairman of this committee, four different types of oxide charges could be identified, cf. Fig. 4.1.

The defects were entitled in the following way: *Fixed Oxide Charge* $Q_f$, *Oxide Trapped Charge* $Q_{ot}$, *Interface Trapped Charge* $Q_{it}$ and *Mobile Ionic Charge* $Q_m$. At that time the different types of charges could not be ascribed to a microscopic defect, but could at least be classified by their process sensitivity, their ability to exchange carriers with the silicon substrate, and by their response to thermal treatment. (i) *Fixed Oxide Charges*: oxidation process dependent fixed positive charges which do not exchange carriers with the silicon substrate. (ii) *Oxide Trapped Charges*: stress induced (avalanche injection or ionizing radiation) positive or negative charges, depending on whether holes or electrons are injected; cannot be neutralized electrically but can be annealed out by low temperature treatment (<500°C). (iii) *Interface Trapped Charges*: structural oxidation-induced defects located at the $Si/SiO_2$ interface, which communicate (exchange carriers) rapidly with the underlying silicon substrate and thus can be either positive or negative, depending on the surface potential; most of the interface trapped charge can be neutralized by low-temperature (450°C) hydrogen annealing. (iv) *Mobile Ionic*

*Charge*: Ionic impurities such as $Li^+$, $Na^+$, $K^+$ and possibly $H^+$, which move across the oxide when a field is applied.

Concerning the degradation of the oxide as a consequence of NBTI, all of those charge types may come into consideration, except for the mobile ionic charges which are rather process contaminations than real defects. In the last 30 years a lot of progress has been made to understand the microscopic nature and degradation dynamics behind the above listed charge species. Against Deal's request, not to modify the terminology of charges, scientists began to invent new names for the defect precursors and defect states causing the observed oxide and interface charge. Furthermore, due to process innovations like PNOs and high–$\kappa$ dielectrics, additional defects and charge types had to be incorporated to the above framework.

## 4.2 The $P_b$ center

The defect corresponding to the interface trapped charge observed after NBTI stress is the so-called $P_b$ center [82–84]. The $P_b$ center is an interface trivalent Si atom with an unsaturated unpaired valence electron at the $SiO_2$/Si interface [85] ($Si_3{\equiv}Si\cdot$), cf. Fig. 4.2 (a). It is originated in the lattice mismatch between the amorphous gate oxide and the crystalline silicon substrate resulting in silicon atom sites with a reduced coordination number (dangling bonds). The energy levels of the $P_b$ centers are located within the silicon bandgap, the DOS profile being peaked twice in the vicinity of the silicon band edges ($E_V$ + 0.26 eV and $E_V$ + 0.84 eV) [50, 86], cf. Fig. 4.2 (b). The characteristic dispersion of the peaks in the DOS profile is suggested to be caused by localized inhomogeneities in the lattice stress involving slight variations in the individual bonding strengths [87, 88]. Superimposed to the signal of the $P_b$ centers, there exists an U-shaped continuum background (band tails) of traps which is caused by deviations in the silicon bond length and angle arising from structural disorder at the interface.

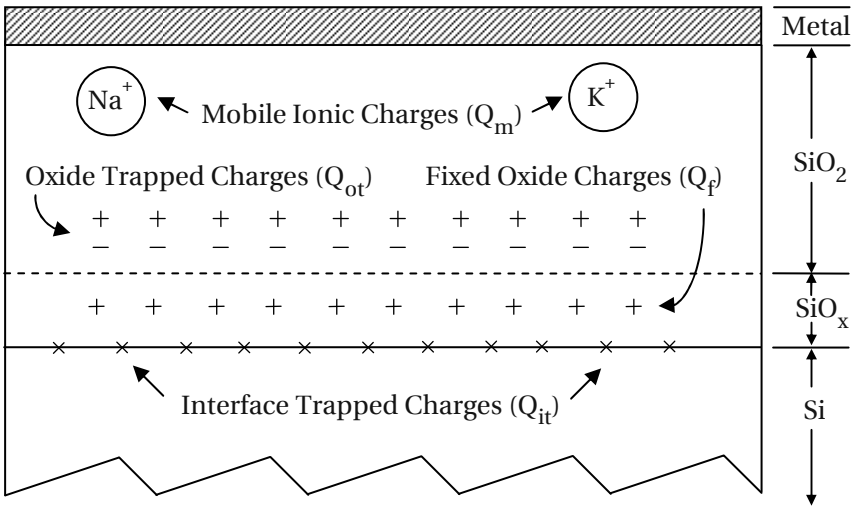


**Figure 4.1:** Terminology and location of charges in thermally oxidized silicon, after B.E. Deal [81]

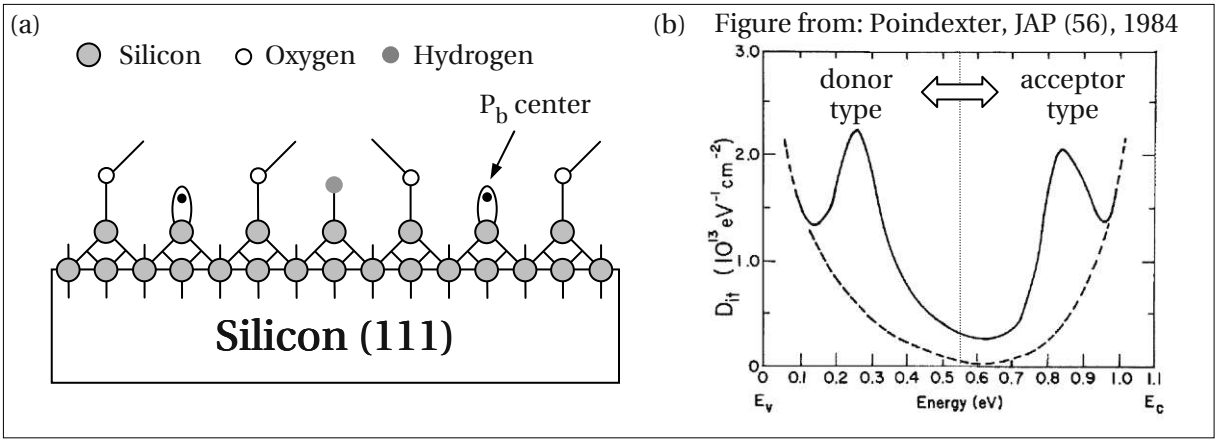


**Figure 4.2:** (a) A schematic illustration of the $Si/SiO_2$ interface. Due to the lattice mismatch, the interface is relatively rough containing silicon atoms with a reduced coordination number and dangling bonds which may be passivated by hydrogen atoms forming Si–H bonds. (b) Bandgap distribution of interface trap density ($D_{it}$) in as-oxidized p-type MOS sample measured by low frequency C-V technique (figure from [50]). The solid line is the total $D_{it}$, the dashed line is assumed to be a continuum component of $D_{it}$ without $P_b$'s. The amphoteric transition level between donor and acceptor type defects is at midgap.

From an electrical point of view, the interface trap is assumed to be amphoteric [89, 90] which means that defects located in the lower half of the silicon bandgap are considered donor-like (positively charged when empty; neutral when occupied), while defects located in the upper half of the silicon bandgap are considered acceptor-like (neutral when empty; negatively charged when occupied). Hence, the net interface charge can be either positive or negative, depending on whether the Fermi level at the interface (surface potential) lies in the lower ($E_V < E_F < E_i$) or in the upper ($E_i < E_F < E_C$) half of the silicon bandgap.

As a consequence of various hydrogen incorporation steps during fabrication, a large number of interface dangling bonds are assumed to be passivated by hydrogen atoms forming Si–H precursors which reduce the electrically active defect density in a fresh device considerably. Due to disorder-induced variations in the bond properties, the Si–H binding energies are dispersed symmetrically around a medium dissociation energy ($E_{dm}$). Following [91, 92], the energetic dispersion can be described by a broadened Fermi derivative distribution with a spread ($\sigma$) of about 0.1 eV. Note that $\sigma$ in Eq. 4.1 is not the variance of the Fermi derivative distribution:

$$g(E_d) = \frac{1}{\sigma} \frac{\exp\left(\frac{E_{dm} - E_d}{\sigma}\right)}{\left(1 + \exp\left(\frac{E_{dm} - E_d}{\sigma}\right)\right)^2}. \tag{4.1}$$

The reported medium dissociation energies of the Si–H complexes ($E_{dm}$) vary between 5.7 eV and 1.5 eV, depending on the bonding environment, the nature of the released hydrogen species (H, $H^+$, $H_2$) and whether a hot hole was captured before dissociation (weakening the Si–H bond) or not [93–96].

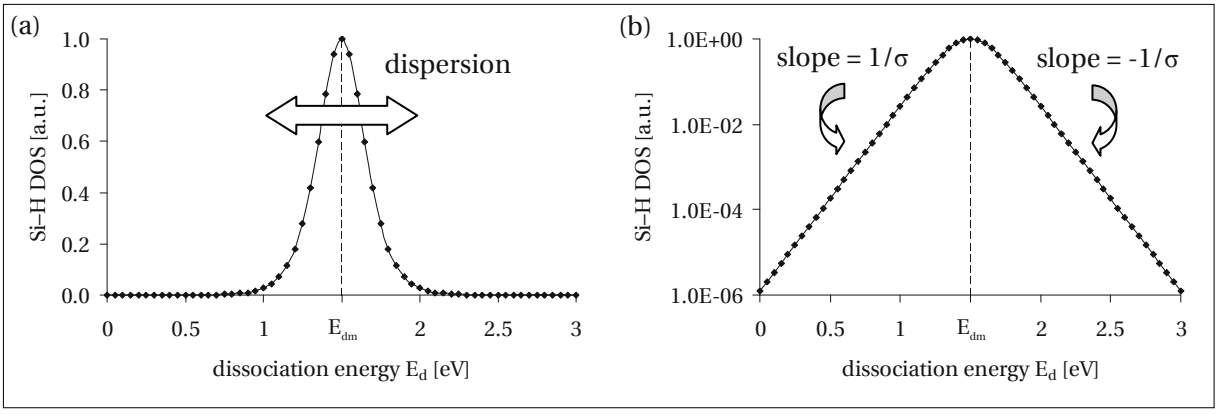


**Figure 4.3:** (a) Hypothetical density of state profile of Si–H bonds following a Fermi derivative distribution ($E_{dm}$ = 1.5 eV, $\sigma$ = 0.1 eV). (b) Same as (a) but with logarithmic y-axis.

In Fig. 4.3, the scaled shape of the Fermi derivative distribution of Si–H bond energies (cf. Eq. 4.1) is illustrated for $E_{dm}$ = 1.5 eV and $\sigma$ = 0.1 eV.

It is assumed that during NBTS, some Si–H bonds may be cracked as a consequence of the applied electric field (lowering the medium dissociation barrier) and due to thermal excitation (increasing the vibrational mode of the bond). As a result of hydrogen release, previously passivated $P_b$ centers become electrically reactivated as they become positively charged when the interface is flooded by holes during NBTS, cf. Fig. 1.2.

## 4.3 The E′ center

The defect most likely corresponding to the oxide trapped charge and maybe to the fixed oxide charge as well is the so-called E′ center [97–99]. Due to its multifarious appearance and its ability to optionally exchange charge carriers with the silicon substrate, the E′ defect family has been given various names in the past like anomalous positive charge (APC) [100], switching oxide traps [101–103], border traps [104] or hole traps [105]. To date, EPR and spin dependent recombination (SDR) are the most widely utilized techniques to identify the physical and chemical nature of oxide defects in MOS technologies [106]. The EPR resonances are labeled as $E'_n$, where n can be either a serial number in crystalline $\alpha$-quartz or a Greek letter in amorphous $SiO_2$ for distinguishing individual signals. The prime and double prime indicate the number of electrons responsible for the EPR signal.

Efforts of [105, 107–109] identified oxygen vacancies within the gate insulator as the precursors for the E′ centers. Oxygen vacancies have long been known to be the dominant intrinsic defects in crystalline and amorphous $SiO_2$. They consist of two adjacent silicon atoms, each back-bonded to three oxygen atoms ($O_3 \equiv Si–Si \equiv O_3$). Following [108–110], the oxygen vacancy has the ability to exchange carriers with the silicon substrate thereby changing its electrical charge state and/or even its molecular configuration. The suggested electrical and molecular transitions which possibly occur during NBTI stress and recovery as well are illustrated in Fig. 4.4.

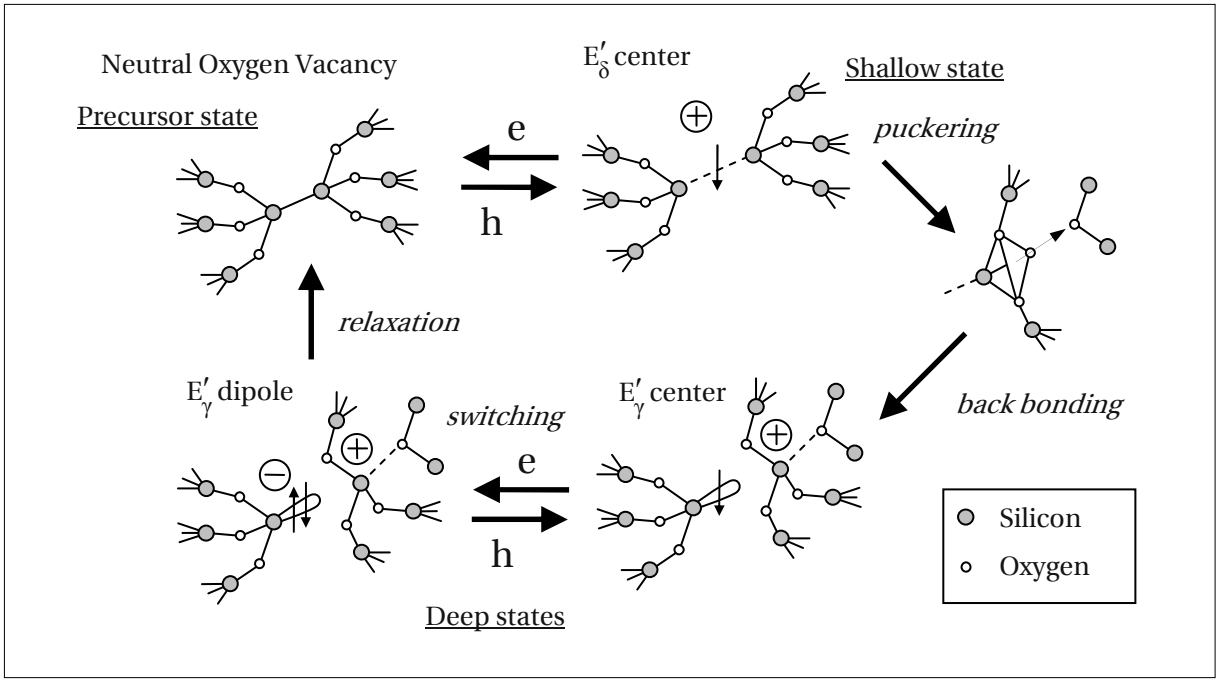


**Figure 4.4:** A schematic state diagram of the oxygen vacancy and the proposed electronic/atomic transitions following electrical or ionizing radiation stress, after [108–110]

During stress, the stretched bond between the two silicon atoms of the oxygen vacancy may be weakened by temperature and field assisted hole capture from the silicon substrate. The resulting positive defect was identified theoretically by Chavez *et al.* [111] as the $E'_\delta$ center. Retaining the dimer configuration, the two adjacent Si atoms relax symmetrically increasing the Si–Si bond distance, the residual single electron being shared with both silicon atoms, cf. Fig. 4.4. The energy levels of the positive $E'_\delta$ centers have been suggested to be shallow states. They are energetically wide-spread around the silicon valence band edge [112] allowing trap neutralization via electron capture from the silicon valence band. Once neutralized, the initial oxygen vacancy configuration is restored, shifting the defect level considerably downwards in energy. Thus, the defect can only be charged positively again when the electrical stress remains applied or during a subsequent stress run. The wide energetic distribution and the metastability of the $E'_\delta$ hole trap makes it also a candidate for the defect type responsible for the temporal dispersion of hole transport through amorphous $SiO_2$ after exposure to ionizing radiation or high field stress (TDDB).

Because of local variations in the lattice stress in the amorphous $SiO_2$, there will be a distribution of separation distances between the two Si atoms, allowing alternative defect configurations after bond distortion. The key point is that the strain between the two Si atoms is likely not the same for all $E'_\delta$ centers, favoring asymmetric relaxation for selected defect sites after hole capture [113]. In that case one of the two Si atoms retains the tetrahedral bonding to its three backside oxygens while the other Si atom relaxes toward the plane defined by its three backside oxygen atoms, thereby puckering into a four-fold coordinated structure with an additional oxygen atom in the vicinity of the former vacancy [97]. A five-fold coordinated structure is reported as well where the puckered silicon atom

bonds to an additional oxygen *and* an additional silicon atom in its molecular neighborhood. The four- and five-fold family of puckered defects are called $E'_\gamma$ centers. After hole capture and asymmetric relaxation, the cluster around the regularly bond silicon atom remains neutral, providing an $sp^3$ hybrid orbital (dangling bond) with an unpaired electron, while the puckered silicon cluster is positively charged resulting in a paramagnetic center 3.5 eV above the $SiO_2$ valence band [108]. The energy levels of the $E'_\gamma$ centers are known as deep states and are located within or close to the silicon bandgap [108, 112]. If an electron is added to this configuration, it pairs with the electron in the dangling bond, making the three-fold silicon cluster negatively charged, while the puckered silicon atom remains positively charged. The resulting neutral state is a metastable dipole configuration. In this so-called Harry Diamond Laboratory (HDL) model proposed by Lelis *et al.* [110, 114], the switching behavior of the trap is explained by reversible carrier exchange between the dangling bond orbital and the silicon substrate without significantly changing the electrical structure of the positively charged puckered silicon complex. Permanent relaxation (reflecting true annealing of the defect instead of charge compensation) can only be achieved in the neutral dipole configuration as the Si–Si bond reforms, thereby restoring the original oxygen vacancy precursor. It is important to mention that the time constants associated with true bond reformation are assumed to be much larger than the switching time constants. Hence, many switching events can occur before the defect finally anneals back to its initial configuration. Once annealed, only a re-stress can activate the same defect again. The previously discussed five-fold puckered configuration (not illustrated in Fig. 4.4) represents an exception to the rule since it may anneal immediately upon electron capture, hence, not showing any switching behavior.

The switching time constants strongly depend on the spatial location and energy level of the individual defect. Hence, the differentiation between a switching oxide trap and a fixed oxide trap is not only a matter of the defect type but also depends on the individual defect energy and how the measurement is done. In a single defect model $E'$ centers with wide-spread energy levels can account for both non-switching oxide hole traps (fixed oxide traps) and switching oxide traps [101].

Although the classical $E'$ center originates from a hydrogen-free oxygen vacancy, there are also related defect classes cited in literature that arise from hydrogenated oxygen precursors. When exposing films of irradiated amorphous $SiO_2$ to molecular hydrogen, even at room temperature some $E'$ centers may be converted into hydrogen coupled complexes [115, 116]. The most prominent candidates observed in electron spin resonance (ESR) spectra are the so-called 74 G doublet [117, 118], the 10.4 G doublet [116] and the $E'_\beta$ centers [119], cf. Fig. 4.5. The precursor for the 74 G doublet center is an oxygen vacancy where one oxygen atom of the silicon back-bonded tetrahedron is substituted by a hydrogen atom. In the case of the 10.4 G doublet center, the precursor has not been identified yet. The only thing one knows about this defect is its microscopic defect structure involving a paramagnetic silicon atom bonded to three oxygens. One of the oxygens is bonded to a proton instead of a silicon atom. The precursor for the hydrogen bridge is assumed to be an oxygen vacancy [93, 120] and forms when replacing one oxygen atom of the regular $SiO_2$ structure by one hydrogen atom. Depending on whether a hydrogen atom is incorporated in the oxygen precursor or not, the energy levels of the defects shift moderately, thereby creating new defect levels within the $SiO_2$ bandgap which may contribute to leakage currents and defect charge centers after electrical or ionizing-radiation stress.

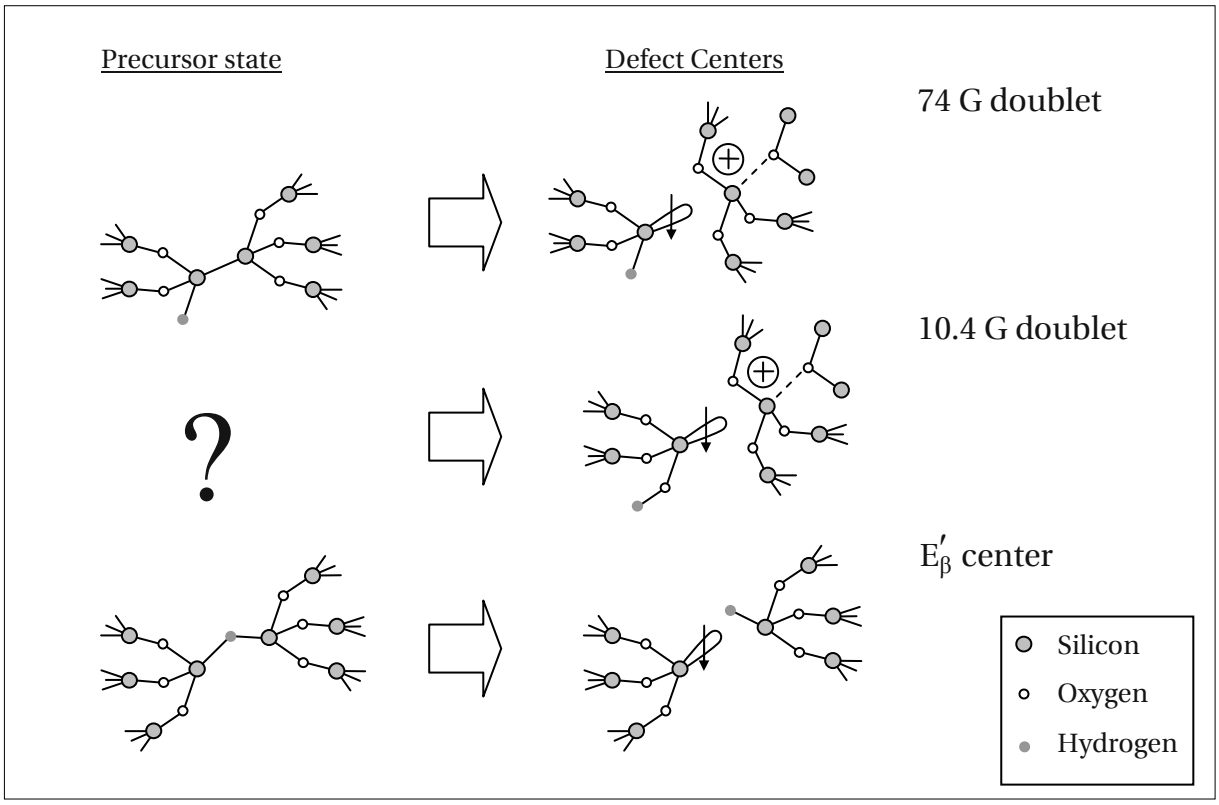


**Figure 4.5:** A schematic diagram of hydrogenated oxygen vacancies and their hydrogen related defect centers.

It has also been suggested that $H_2$ molecules can be cracked at E′ centers, thereby creating various hydrogenated species of oxygen vacancies [93, 121, 122] and free hydrogen atoms that may interact with interface defects. Furthermore, Lenahan *et al.* [83] suggested that there may be hydrogen exchange between passivated $P_b$ centers and stress induced hydrogen-free E′ centers. He argued that the presence of a large number of hydrogen passivated silicon dangling bond sites at the $Si/SiO_2$ interface and a large number of unpassivated silicon dangling bond sites in the adjacent oxide (E′ centers) corresponds to a well-ordered configuration which cannot be favorable in thermal equilibrium since its Gibbs free energy would be very high. Thus, basic statistical thermodynamics arguments based on entropy considerations indicate that the exchange of a large number of hydrogen atoms from interface dangling bonds toward the E′ centers is thermodynamically favorable, thereby creating a direct link between hydrogen release from the interface and the creation of oxide defects.

## 4.4 The K′ center

Due to process innovations going along with continuous device scaling, an alternative type of gate oxide insulator increasingly gained importance, namely silicon nitride ($Si_3N_4$). Silicon nitrides and ni-

trided oxides (NOs) have a number of advantages over pure silicon dioxide ($SiO_2$) insulators, for example, they have a higher dielectric constant (high–$\kappa$), a lower gate leakage current due to the larger oxide thickness, a denser structure and a better resistance to hydrofluoric (HF) acid than $SiO_2$ [123]. The higher dielectric constant provides larger capacitances at lower oxide thicknesses, while the denser structure makes the oxide a better barrier against the diffusion of various impurities like doping atoms, mobile ions and moisture [124].

The nitridation process is realized in different ways depending on the oxide thickness. While thin oxide devices (<10 nm) undergo a so-called thermal nitridation where a pure silicon surface is exposed to $N_2$ or $NH_3$ at relatively high temperatures (950°C – 1300°C), thick oxide devices (>10 nm) are usually processed by plasma enhanced nitridation where the oxide is grown at lower temperatures by adding a plasma (photons, electrons or ions). A third method is realized by ion implantation where nitrogen ions are implanted directly beneath the surface of the silicon wafer followed by a high temperature anneal. The thickness of the nitride film is controlled by the implantation energy, the ion mass and the ion dose, thereby leading to a Gaussian-like implantation profile with poor interfacial properties.

By introducing oxygen either subsequently to the actual nitridation process (re-oxidized nitroxide (RONO) or rapid thermal oxynitridation (RTNO)), or directly during low pressure chemical vapor deposition (LPCVD) or plasma enhanced chemical vapor deposition (PECVD), some nitrogen atoms are replaced by oxygen, resulting in an SiON film ($SiO_xN_y$) which has a similar structure but much less tensile stress as opposed to the pure $Si_3N_4$ layer. The bandgap energy of the $SiO_xN_y$ layer can be adjusted between 5 eV and 9 eV depending on the [O]/[N] ratio [124].

As a consequence of nitrogen incorporation, the micro structure of the oxide is modified considerably which introduces some new defect types [125, 126] in plasma nitrided oxides (PNOs) and thermally nitrided oxides (TNOs). One of the new defect classes is labeled K′ centers in pure $Si_3N_4$ and $K'_N$ centers in silicon oxynitrides [127, 128]. Following Fig. 4.6, both trapping centers are likely dangling bond defects in which the silicon is back-bonded to three nitrogen atoms, the dangling orbital electron being located at the central silicon atom [129–132]. Indeed, there are only small discrepancies in the magnetic-resonance parameters between K′ centers and $K'_N$ centers which might arise from slightly different bonding environments in silicon nitride and SiON [133].

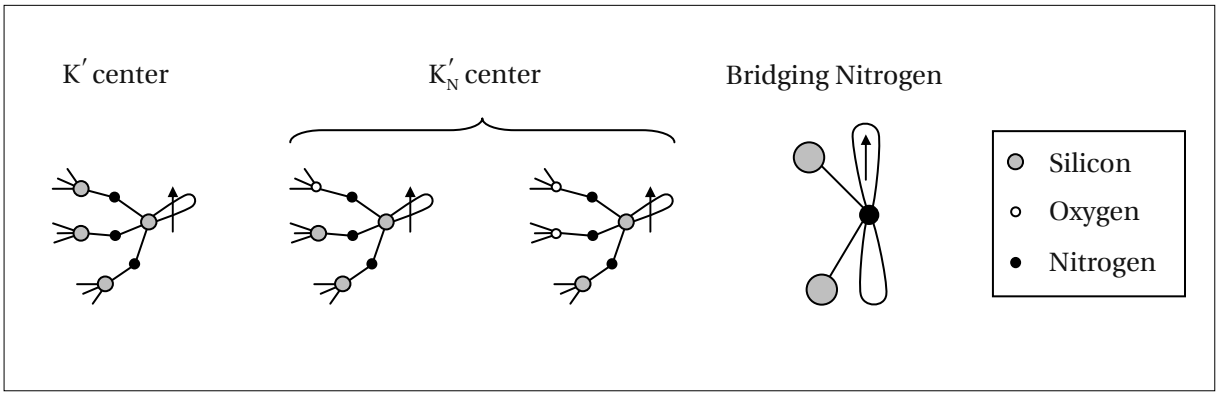


**Figure 4.6:** A schematic diagram of nitrogen related defects in silicon nitride and oxynitrides

In SiON one or two of the second nearest neighbors of the central silicon atom may be oxygen atoms instead of silicon atoms. The K′ center defects are assumed to be located in the near interfacial region [134] of the dielectric and have a narrow-peaked effective density of state profile near the middle of the silicon bandgap, hence, constituting very effective trapping centers in MOSFET devices. Besides K′ centers, other defects like bridging nitrogen configurations have been identified in nitrided oxides as well, cf. Fig. 4.6. ESR spectroscopy revealed that one (paramagnetic) or two (diamagnetic) electrons can be captured in the N $2p_\pi$ non-bonding orbital of the brigding nitrogen defect which is oriented normally to the $Si_2N$ plane [135, 136].

The PNO process is known to enhance NBTI considerably because a lot of hydrogen is incorporated into the gate oxide during PECVD forming Si–H and N–H bonds within the gate oxide and at the interface [137]. Assuming that K′ centers in nitrided oxides can be passivated by hydrogen in a similar way as $P_b$ centers in $SiO_2$ oxides [138], oxynitrides provide a large number of hydrogen precursors which may be distorted during NBTI stress [139], explaining the similar degradation dynamics as observed in $SiO_2$ devices [140]. Besides that K′ centers and other nitrogen related defects have a much lower hole trapping barrier than E′ centers, hence, positive oxide charge is assumed to be prevalent in nitrided oxides. It is also reported that K′ centers represent effective hydrogen traps [141] and can therefore trigger hydrogen desorption from Si–H bonds at the interface and hence $P_b$ center creation [8]. This would explain the link between hole trapping and interface trap creation observed in nitrided oxides.

# 5

# Bias dependence of degradation and recovery

THE aim of this chapter is to study the effect of read-out gate bias and gate pulsing on the measured $V_{TH}$ shift and the CP current response. In general, every variation in the gate bias corresponds to a shift of the Fermi level across the silicon bandgap, cf. Fig. 2.2. Depending on the Fermi level position, different charge densities at the interface may trigger carrier exchange mechanisms between stress-induced defects and the silicon substrate. Concerning interface states located energetically within the silicon bandgap, the carrier exchange process is assumed to happen quasi instantaneously (SRH), allowing to adjust the equilibrium charge state very quickly after a bias switch (relative to the time scales of our measurements). Following Subsection 2.1.3, this means that traps above the Fermi level are unoccupied while traps below $E_F$ are occupied. The net charge state changes as the Fermi level crosses the silicon bandgap during a standard gate bias ramp. As a result, we observe a variation in the subthreshold slope of the transfer curve which is usually attributed to charging of stress-induced interface states [30–33, 33–35, 142]. Recent research suggests, however, that oxide traps can change their occupancy with respect to $E_F$ as well, which may result in an additional contribution to the variation in the subthreshold slope [35, 143, 144]. However, due to a certain spatial depth of the trapping centers within the gate oxide and due to a possible inelastic carrier tunneling process governing carrier exchange between the silicon substrate and stress-induced oxide defects, oxide trap neutralization/charging by electron/hole capture is likely to be afflicted with larger time constants as compared to SRH recombination.

Here, we extend the idea of rechargeable oxide traps with larger carrier exchange time constants than interface states and will prove that such traps have to be considered in order to explain all features of the $V_{TH}$ shifts recorded at different gate biases. The position of the Fermi level at which we evaluate the threshold voltage shift will turn out to be of fundamental importance since it has a considerable

impact on the extracted degradation and recovery characteristics. In a combined study, we compare threshold voltage shifts and recovery of NMOS (SM5N/30/H1) and PMOS (SM5P/30/H1) devices which allows us to gain access to the full silicon bandgap by appropriate gate biasing [145].

## 5.1 PMOS & NMOS Combination Technique

As we drive a device from accumulation to inversion by a gate bias sweep, we cannot reasonably measure an appropriate exponential growth in the drain current before the density of minority carriers does not exceed the density of majority carriers within the channel. In the subthreshold region of the device the Fermi level changes almost linearly with the gate voltage (cf. Fig. 2.2) and the drain current grows proportional to the majority carrier density:

$$I_{\mathrm{D}}^{\mathrm{NMOS}} \propto n = n_{\mathrm{i}} \exp\left(\frac{E_{\mathrm{F}} - E_{\mathrm{i}}}{k_{\mathrm{B}} T}\right), \tag{5.1}$$

$$I_{\mathrm{D}}^{\mathrm{PMOS}} \propto p = n_{\mathrm{i}} \exp\left(\frac{E_{\mathrm{i}} - E_{\mathrm{F}}}{k_{\mathrm{B}} T}\right). \tag{5.2}$$

Consequently, when evaluating $V_{\mathrm{TH}}$ shifts from transfer curves, the energy range between the valence band edge $E_{\mathrm{V}}$ and the intrinsic energy $E_{\mathrm{i}}$ ($n < n_{\mathrm{i}}$) is not accessible for the NMOS device. Similarly, the energy range between $E_{\mathrm{i}}$ and the conduction band edge $E_{\mathrm{C}}$ ($p < n_{\mathrm{i}}$) remains hidden for the PMOS device. Only by combining the study of both devices, we can maximize the accessible range of the $V_{\mathrm{TH}}$ shift evaluation versus gate voltage: on a PMOS device we have the possibility to scan roughly the lower half of the silicon bandgap by varying the Fermi level between $E_{\mathrm{i}}$ and $E_{\mathrm{V}}$ (hole current). By using otherwise an identically processed but oppositely doped NMOS device, which has a completely different Fermi level position for gate biases around its threshold voltage, we can scan roughly the upper half of the silicon bandgap by varying the Fermi level between $E_{\mathrm{i}}$ and $E_{\mathrm{C}}$ (electron current). A schematic drawing of the individual Fermi level positions is given in Fig. 5.1.

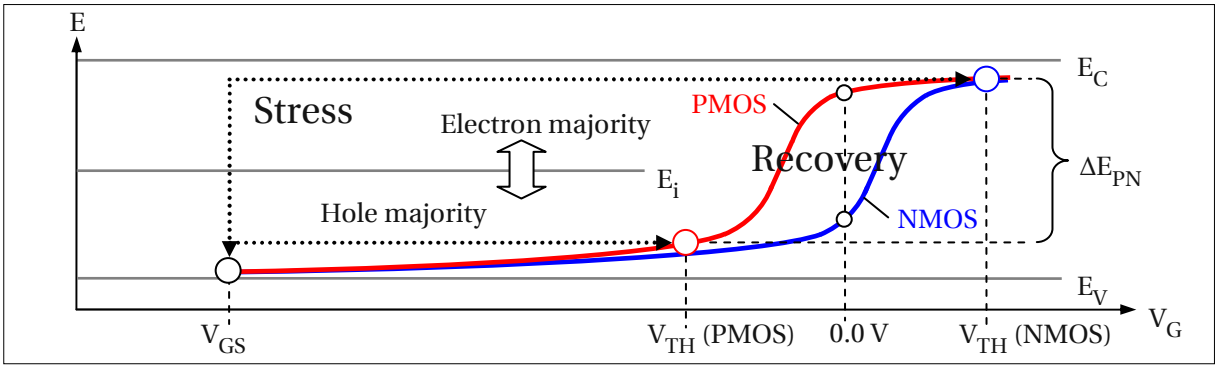


**Figure 5.1:** A schematic illustration of the Fermi level position at the interface during stress and recovery for the PMOS and the NMOS device. The values at the stress level and at the particular $V_{\mathrm{TH}}$ are highlighted by circles. During stress, the Fermi level position and the oxide field is identical for both types of devices. The energy interval $\Delta E_{\mathrm{PN}}$ represents the difference in Fermi level position between the NMOS and the PMOS device during recovery at their particular $V_{\mathrm{TH}}$.

In order to generate identical stress levels by identical stress biases, we have fabricated our devices with a single gate poly process $\longrightarrow$ $n^{++}$ gate material for both NMOS (SM5N/30/H1) and PMOS (SM5P/30/H1) devices. The threshold voltage of the PMOS device is approximately -1.1 V, while the threshold voltage of the NMOS device is roughly +1.1 V. We remark that despite of the different substrate and junction doping, the same stress biases applied during NBTS result in the same oxide fields and identical carrier concentrations under the gate oxide. This statement holds at least for thick oxide technologies stressed at gate biases far beyond threshold voltage (cf. Section 1.3). By studying differently doped hardware it is possible to monitor the impact of comparable degradation levels in arbitrary regions of the entire silicon bandgap. In particular, the method allows to study degradation and recovery with a majority of holes at the interface (PMOS) as well as with a majority of electrons at the interface (NMOS).

All devices discussed in the following are stressed identically for one thousand seconds at a temperature of 125°C and a constant oxide field of 5.5 MV/cm. During recovery, also at 125°C, we monitor the response of the charge pumping current and the threshold voltage shift by applying positive (NMOS device) or negative (PMOS device) gate biases around the $V_{TH}$ of the individual devices. For these investigations, it is necessary to compare different sets of identically processed devices (except for the doping) with the same geometry on the same wafer (SM5P/30/H1 and SM5N/30/H1) by subjecting them to NBTI stress. In order to achieve a good correlation of the degradation levels after stress, the DUTs were selected carefully by their virgin CP characteristics. This was done because preliminary measurements have indicated a good correlation between the degradation levels observed in equally fabricated devices provided they had identical virgin CP characteristics. The zero-hour CP current gives information about the initial interface state density of the device.

## 5.2 Bias dependence of the CP current recovery

In Fig. 5.2 (a) the recovery of $\Delta I_{CP}^{max}$ is illustrated for the NMOS and the PMOS device stressed under the same bias and temperature conditions. During CP, the gate junction of the NMOS device was continuously pulsed between -1.0 V and +2.0 V, the gate junction of the PMOS device between -2.0 V and +1.0 V, respectively. Except for the 1 V variation in the biasing conditions (because of the different threshold voltages) both devices were measured identically using the same gate pulsing setup (rising/falling slopes 10 V/ţs; frequency 500 kHz). After stress, the NMOS device provides a CP signal similar to the PMOS device, indicating a similar level of interface damage. Note that within the first second of uninterrupted gate pulsing there is no significant decrease in $\Delta I_{CP}^{max}$ neither for the PMOS nor for the NMOS device. This finding is in contradiction to results obtained by on-the-fly charge pumping (OFIT) measurements reported in [64] which showed fast interface state recovery immediately after terminating stress. On the other hand, within the first 100 s of uninterrupted gate pulsing, the CP currents of both devices recover in a similar way by about 15 %, indicating time dependent re-passivation of dangling bonds at the interface [12, 140, 146, 147]. In the experiment illustrated in Fig. 5.2 (b) continuous gate pulsing was replaced by sequences of constant gate biasing (NMOS +1.1 V; PMOS -1.1 V) followed by gate pulsing. Both cycles were repeated four times with increasing cycle durations (10/100/1,000/10,000 s).

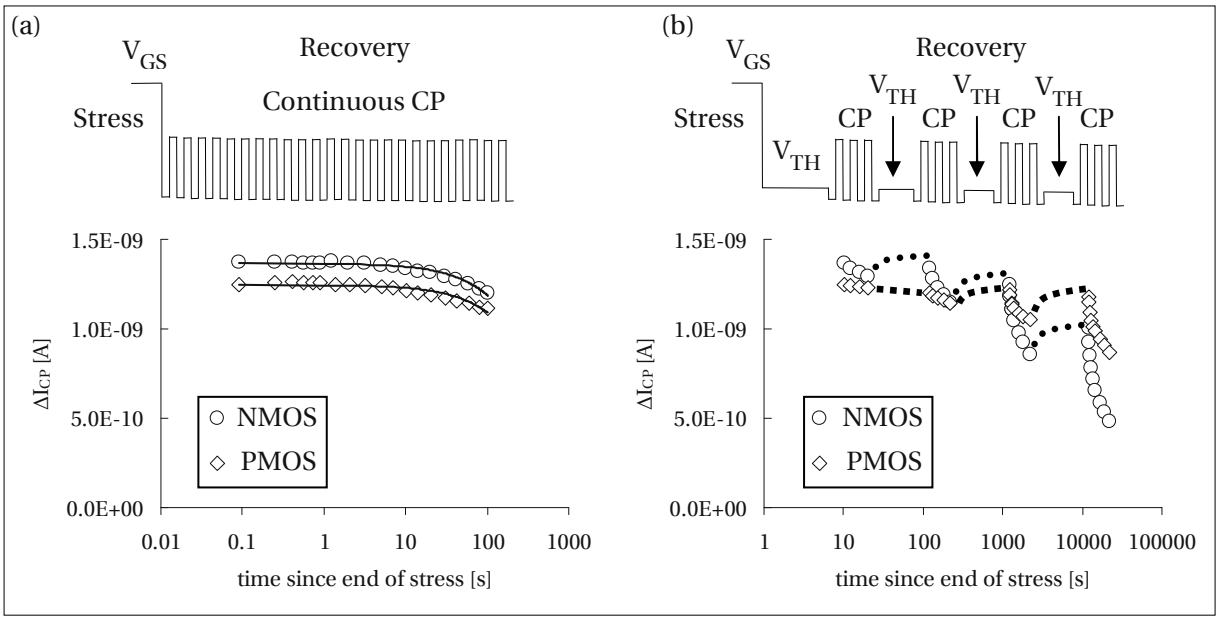


**Figure 5.2:** (a) The recovery of $\Delta I_{CP}^{max}$ for a PMOS and a NMOS device after stressing both devices under the same bias and temperature conditions. Both devices show similar degradation of $I_{CP}^{max}$, indicating a comparable interface state density at the end of stress. Within the first second of recovery the reduction of $\Delta I_{CP}^{max}$ is insignificant and independent of the bias conditions. (b) The recovery of $\Delta I_{CP}^{max}$ for a PMOS and a NMOS device with intermediate constant bias phases at the particular threshold voltages of the devices (NMOS +1.1 V; PMOS -1.1 V). Only during the gate pulsing periods interface state recovery is observed. During the constant bias periods some of the previously recovered interface states can be restored again.

During the gate pulsing periods, we observe enhanced interface state recovery similar to that reported in [140, 148]. We also note that the recovery of the NMOS device is faster than the recovery of the PMOS device which may be due to the different biasing conditions during gate pulsing [149, 150]. Conversely, between two pulsing periods, where the gate bias is constant, there is rather an increase than a decrease in $\Delta I_{CP}^{max}$ which indicates that previously recovered interface states may be restored again, provided the gate junction is not pulsed. The apparent 'degradation' of the CP current during the constant bias phases is estimated by dots in Fig. 5.2 (b). Remarkably, the increase is similar for the PMOS and NMOS devices despite of the different biasing conditions (± 1.1 V) indicating that the gate voltage and the carrier concentration at the interface are not crucially determining this phenomenon.

In contrast to Yang *et al.* [149], who concluded that the passivation of interface states is accelerated by an applied positive gate bias, and Ang [150], who suggested that the CP current recovery is suppressed by an applied positive gate bias, our experiments show unambiguously that it is the high frequency pulsing, i.e. the CP measurement itself which accelerates interface state recovery and makes it look like the 'classic' recovery curves illustrated in Fig. 5.2 (a). The pulsing levels and the sign of the constant gate voltage between the CP samples seems to be just of second-order importance as long as the gate voltage does not subject the device to electrical stress like in conventional OTF or OFIT measurements. The results obtained by [149] and [150] might be also misleading due to the long lasting

CP measurement periods (e.g. 120 s in [150]) and the large measurement delay used to determine the $N_{it}$ contributions. These observations challenge $V_{TH}$ recovery theories based on interface state re-passivation [12, 140, 146, 147] and indicate that long continuous gate pulsing is not appropriate to investigate $D_{it}$ recovery during constant gate biasing.

In the following section it will be demonstrated that the interface state contribution stays quasi-permanent, consistent with [78], if we keep our CP samples shorter than one second and abstain from long continuous gate pulsing periods. This finding indicates that the observed increase of the CP current during constant gate biasing (cf. Fig. 5.2 (b)) is rather a restoration of previously recovered interface traps than a true degradation.

## 5.3 Bias dependence of the $\Delta V_{TH}$ recovery

In this section, the influence of the gate bias on the total $V_{TH}$ shift is investigated. In particular, we expect that CP has a significant impact on the carrier exchange properties and therefore on the $V_{TH}$ and oxide trap recovery since gate pulsing during CP corresponds to periodic Fermi level switches across the entire silicon bandgap. Standard $V_{TH}$ recovery curves observed after NBTI stress of p- and

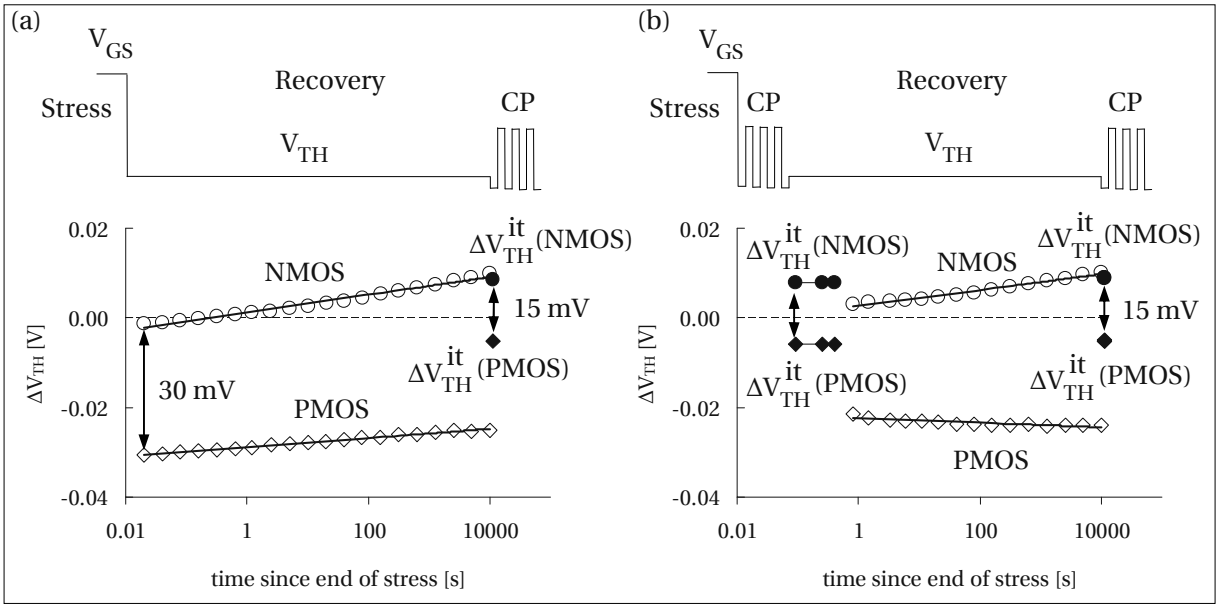


**Figure 5.3:** (a) The constant bias recovery recorded at the particular $V_{TH}$ for the PMOS (-1.1 V) and for the NMOS (+1.1 V) device. After 10,000 s recovery a short CP measurement was performed and the $\Delta I_{CP}^{max}$ currents were converted into appropriate interface state dependent $V_{TH}$ shifts ($\Delta V_{TH}^{it} \rightarrow$ full symbols). (b) The same measurement procedure as in (a) including an additional short CP cycle right after stress. The corresponding $\Delta V_{TH}^{it}$ shifts are shown by full symbols. The CP measurement has a dramatic impact on the recovery of the $V_{TH}$ shift. No CP current recovery is observed during the intermediate constant gate bias phases.

n-channel MOSFETs are displayed in Fig. 5.3 (a). During recovery the gate-source bias was kept at a value close to the threshold voltage of the PMOS (-1.1 V), and NMOS (+1.1 V) device , respectively.

According to Fig. 5.3 (a), the first important finding is that both p- and n-channel devices recover in a comparable way (recovery rate NMOS = +2.0 mV/dec; recovery rate PMOS = +1.0 mV/dec), yet starting from completely different $\Delta V_{TH}$ values visible immediately after the end of stress. Note that the $V_{TH}$ shift of the NMOS device is negative within the first 200 ms post stress. Only after 200 ms the shift becomes positive indicating that the balance between negative and positive contributions to the $\Delta V_{TH}$ shift has turned in favor of the negative charge.

In order to understand these results, a short CP measurement was appended right after the constant bias recovery period and changes in $I_{CP}^{max}$ where converted into appropriate $D_{it}$ dependent $\Delta V_{TH}$ shifts ($\Delta V_{TH}^{it}$). The conversion was performed according to Eq. 2.38 assuming a flat DOS and an amphoteric nature of interface traps. Recalling Subsection 2.1.3, the active interface state charge can be either positive or negative depending on the current Fermi level position during read-out. This implies that although the net charge build up after NBTI is positive in most reported cases (resulting in a negative threshold voltage shift after NBTI), individual contributions to the $V_{TH}$ shift can interfere either constructively or destructively with respect to their present charge state. Hence, in a PMOS device biased at $V_{TH}$ (-1.1 V) both interface traps and oxide traps are positively charged accumulating in a high negative threshold voltage shift. Conversely, in a NMOS device biased at $V_{TH}$ (+1.1 V) oxide traps are positively charged while interface states are negatively charged, resulting in either a net positive or net negative threshold voltage shift depending on the dominating contribution.

In Fig. 5.3 (a), the $D_{it}$-equivalent $\Delta V_{TH}$ shifts for both devices are illustrated by full symbols. Clearly, after a long recovery period of 10,000 s, the remaining threshold voltage degradation is essentially due to interface states for the NMOS ($\approx$ +8.0 mV)), whereas for the PMOS the major part ($\approx$ -2.5 mV) of the visible shift must be attributed to positive oxide charge, only -5 mV being due to interface states.

In order to elaborate the role of interface states and gate pulsing on the $V_{TH}$ recovery, we perform another key experiment using a second set of p- and n-channel devices, cf. Fig. 5.3 (b). On this second set we start immediately with CP after the end of stress, i.e. before recording the $V_{TH}$ recovery curve at a constant gate bias. The CP periods were kept as short as possible (< 1 s) in order to prevent the previously discussed $D_{it}$ recovery induced by gate pulsing, cf. Fig. 5.2 (a). A comparison of the $V_{TH}$ recovery characteristics illustrated in Fig. 5.3 (a) and Fig. 5.3 (b) shows that CP right after stress influences the subsequent $V_{TH}$ shift considerably, consistent with [149, 151]. The $\Delta V_{TH}$ levels of both NMOS and PMOS devices shift in the positive direction indicating a reduction in net positive charge. The recovery curve of the PMOS device even changes its direction after the CP measurement. When comparing the $\Delta V_{TH}^{it}$ contributions of Fig. 5.3 (a) and Fig. 5.3 (b), one finds that they are nearly identical right after stress and after 10,000 s recovery at constant gate bias. This again indicates that (at least for this particular wafer) interface state recovery is negligible provided the CP measurement cycles are kept short.

## 5.4 Bias dependence of the $\Delta V_{TH}$ extraction

The occupancy of interface states can follow changes in $E_F$ quasi instantaneously (SRH theory), with the fastest response times being in the pico- and nano-second range and the slowest response times for traps located at midgap in the millisecond regime [152], both outside our measurement window. Oxide traps, on the other hand, are expected to require a longer time interval to restore equilibrium with the silicon substrate since a large quantum mechanical barrier has to be overcome by elastic or inelastic tunneling. Also, charging and discharging is likely to be accompanied by structural relaxation [102, 110, 153]. Generally this means, that if the Fermi level is changed abruptly (for example when we switch the gate bias from stress level to $V_{TH}$) a new equilibrium condition is generated. Assuming an elastic tunneling process, stress-induced oxide traps within the energy range $E_{F,new} - E_{F,old}$ would tend to become or stay positively charged while traps below $E_{F,new}$ would rather be neutralized. Such a 'switching' behavior of oxide traps was already reported for E′ centers in MOS structures after irradiation [101, 103].

In Fig. 5.4, we generate such new equilibrium conditions by switching the gate bias during recovery of the NMOS and PMOS devices in a defined way between weak and deep inversion: after 10 s of recovery at ±1.1 V (1), we change the gate bias for another 100 s towards weaker inversion (±0.85 V) (2). This moves the Fermi level immediately from the band edges closer toward midgap, cf. Fig. 5.4. A second gate bias switch from ±0.85 V (2) to ±0.6 V (3) for another 1,000 s moves the Fermi level even closer to midgap. The last recovery cycle was performed again at ±1.1 V (1) and lasted for 10,000 s.

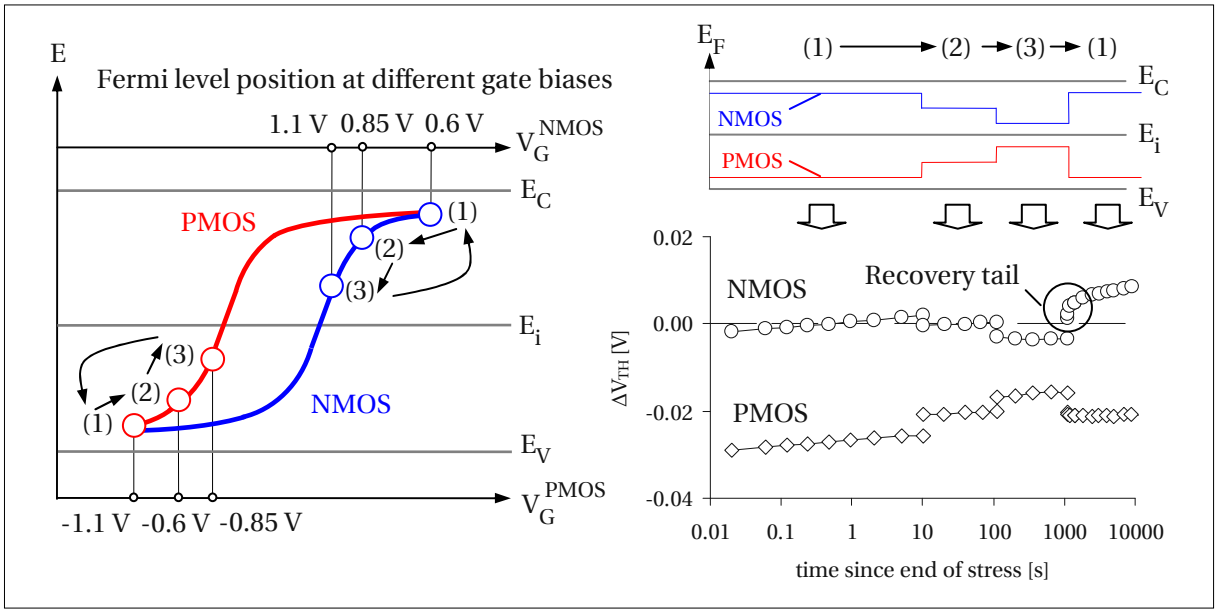


**Figure 5.4:** $V_{TH}$ shift during recovery at different gate biases for the PMOS and the NMOS device. On the left hand side, the Fermi level position is illustrated for the individual gate biases. As the gate voltage drives the Fermi levels closer toward midgap (1) → (2) → (3), the difference in the $V_{TH}$ shift between the PMOS and the NMOS device decreases. After switching the gate bias from depletion back toward inversion (3) → (1), transients emerge as tails in the recovery curves.

According to the data shown in Fig. 5.4, a Fermi level variation toward accumulation lowers the total $V_{TH}$ shift for both the NMOS and PMOS transistor. Consequently, the difference in net positive charge between NMOS and PMOS devices is reduced. Recalling the previous discussion, this is because the energy interval of electrically active charge is narrower as we bias both devices closer toward midgap. After the last bias switch from ±0.6 V (3) back to ±1.1 V (3), we observe a huge step in the $V_{TH}$ shift indicating that previously neutralized defects can be recharged again. This behavior is typically attributed to interface states and not fully appreciated for oxide defects. Furthermore, these charges do not react instantaneously to the bias switch. Instead, moderately fast transients are observed (tails in the recovery curve) after the bias switch indicating a relatively slow (i.e. time constants in the range of several thousand seconds) dynamic carrier exchange process which may be due to inelastic tunneling between oxide defects and the silicon substrate. SRH recombination of interface states or elastic tunneling would be definitely too fast to produce recovery tails within our experimental time resolution.

## 5.5 Conclusions – Bias dependence of the $\Delta V_{TH}$ shifts

Based on the upper key experiments performed on identically stressed NMOS and PMOS devices, one may draw the following conclusions on bias dependence of the measured $\Delta V_{TH}$ shifts and its recovery:

1. One must use separate sets of devices and short CP intervals for reliable $\Delta V_{TH}$ and $\Delta V_{TH}^{it}$ investigations during constant bias recovery. These conclusions are drawn from the findings that $\Delta I_{CP}$ and $\Delta V_{TH}$ recovery are influenced considerably by gate pulsing.

2. Interface states are quasi-permanent as long as the gate bias is constant between two short CP measurements, cf. Fig. 5.3 (b). Consequently, they cannot account for the threshold voltage recovery observed at constant gate bias.

3. During $V_{TH}$ recovery performed at positive or negative gate bias, both NMOS and PMOS devices relax in the positive direction, cf. Fig. 5.3 (a). This indicates that in both cases net positive charge is lost. The recovery rate, however, is found to be larger in the NMOS device (biased at a positive gate voltage). This is consistent with previous observations and with the idea of positive oxide trap neutralization via electron capture, the latter being much more efficient at positive gate biases (majority of electrons at the interface).

4. The net charge at the end of a long recovery phase is positive for the PMOS and negative for the NMOS device. Indeed, the $V_{TH}$ shift of the NMOS device after 10,000 s recovery at +1.1 V equals approximately its interface state contribution (measured via CP). Conversely, 10,000 s recovery at -1.1 V (PMOS) leaves a considerable amount of remaining positive oxide charge, consistent with [27, 78]. The explanation agreeing with the bulk of literature on oxide and interface traps is as follows: Oxide traps (E′ centers) are donor-like while interface states are amphoteric [50, 89]. Consequently, the positive $V_{TH}$ shift of the NMOS device after recovery must originate from a dominance of negative interface state charge.

5. The observed difference of ≈ 30 mV between the total $V_{TH}$ shift of the NMOS and PMOS device visible in Fig. 5.3 (a) cannot be explained by interface charge alone, cf. Fig. 5.3 (b). The $D_{it}$ related fraction ($\Delta V_{TH}^{it}$), which is no more than ≈ 15 mV, is only about half of the total $V_{TH}$ shift difference.

6. After NBTI significantly more (equivalent another ≈ 15 mV) positive oxide charge remains visible on PMOS device than on the NMOS device, although the stress field and temperature have been identical for both devices.

7. The difference in the $V_{TH}$ shift can be reduced by biasing both transistors closer toward accumulation, the Fermi levels approaching midgap, cf. Fig. 5.4. This suggests a certain amount of defects to be rechargeable, which is accepted for interface states but not fully appreciated for oxide defects. As opposed to interface states (which align their charge state very fast according to SRH theory) oxide charges do not react instantaneously to the bias switch, making them observable as slow tails in the recovery curves within our experimental time resolution.

# 6

# The role of temperature in NBTI characterization

HAVING evaluated and summarized fundamental features of NBTI degradation and recovery in the previous chapters, this chapter recalls the drawbacks of conventional characterization techniques and introduces a new hardware tool, enabling us to overcome previously strict experimental limitations regarding temperature. The new experimental approach is based on in-situ polyheaters which enable us to switch the temperature of the device in a very fast and defined way. Having broadened one's mind to the possibility of switching the temperature at arbitrary points of time, the polyheater technique opens up unprecedented experimental capabilities for NBTI characterization.

This chapter addresses the calibration procedure of the polyheater tool, investigates its capabilities and demonstrates dedicated experimental setups which become feasible once having a well calibrated and stable temperature switching tool at hand. Having embedded the polyheater in the measurement software environment, the technique allows to run stress/recovery experiments with variable temperature. From these experiments new fundamental features of NBTI arise, helping to clarify conflicting issues in literature and allowing to draw new conclusions which finally lead to a more consistent microscopic picture of the degradation and the recovery mechanism.

## 6.1 Fast heating and cooling using in-situ polyheaters

In this section the in-situ polyheater measurement technique is introduced and its power for device characterization and reliability issues is discussed. Furthermore, the calibration procedure and performance of the polyheater tool is elaborated. A particular emphasis is put on the accessible temperature range, the heating and cooling dynamics as well as on the impact of the temperature gradient

between heater, active device and thermo chuck. Once calibrated, the polyheater technique provides a reliable solution for fast and arbitrary temperature switches and offers the possibility to reach device temperatures far beyond the operating range of conventional thermo chuck systems [154].

### 6.1.1 Hardware assembly and polyheater design

Poly resistors surrounding a silicon device can be used to perform fast in-situ heating on a single device on wafer level which is commonly applied in time-critical fast wafer level reliability (fWLR) monitoring [155–159]. The main advantage of the polyheater is that it provides an elevated stress temperature without the use and the limitations associated with a conventional heating system like a thermo chuck. Although, this alone is already a valuable feature, it does not even begin to explore the multitude of possibilities one finds in a more scientific use of the tool. By correct calibration and automation of the polyheater-device system, the temperature becomes a quasi arbitrary experimental parameter which can be switched easily by more than ±200 K within a couple of seconds. For such an application conventional thermo chuck systems are unsuitable because they are slow and their heating/cooling durations depend strongly on the difference to the target temperature. Furthermore, when attempting to keep the junction biases applied during the temperature switch, it is a mandatory requirement not to lose probe needle contact. Since heating and cooling of a wafer on a thermo chuck involves considerable thermal expansion of wafer, needles and pads, this request cannot be fulfilled over a wide temperature range without continuous manual readjustments. As opposed to heating the whole wafer on the thermo chuck, in-situ heating by poly wires is very local. Hence, there is no thermal expansion of needles and pads, which saves us from losing mechanical contact during the temperature switch.

Once calibrated and implemented in the software, the polyheater feature opens up unprecedented possibilities for device characterization and reliability testing. Its application potential is thereby way beyond the scope of NBTI characterization. The concept of in-situ heating by implanted poly wires has been taken up also for instance for thermo cycling [160] or for performing on chip high temperature annealing of irradiated PMOS dosimeters [161].

A common challenge of in-situ heating can be found in the fact that the device we want to heat is in most cases a distance away from the actual heating source. Consequently, the poly wires are always hotter than the tested structure itself, which results in a temperature gradient between radiator, device, and ambient. From a technology point of view the distance between the polyheater wires and the active areas of the device must be large enough to prevent a direct current flow between the individual components. However, the larger the distance between the heater and the device, the higher is the thermal resistance of the system and the more power has to be applied to the wires in order to generate a certain elevated device temperature. Furthermore, the time delay for restoring thermal equilibrium must also be taken into account as we switch the heater on or off abruptly. The farther the heater is away from the device, the longer it takes to restore thermal equilibrium. Thus, in order to find the optimal distance between the heating source and the active device for a particular application, the above specified points have to be considered carefully.

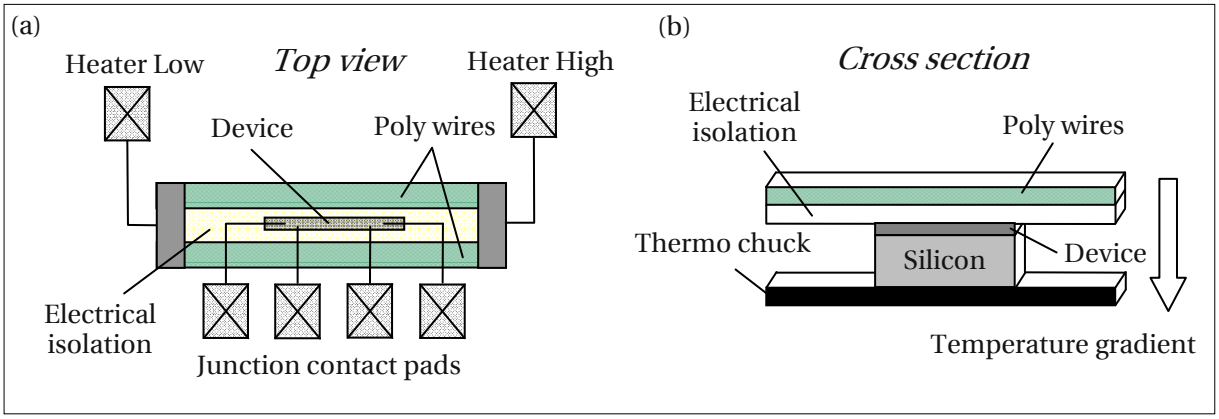


**Figure 6.1:** (a) A schematic illustration of a polyheater-device system placed on a thermo chuck. Electrically isolated poly wires surround the silicon device. When a voltage is applied to the wires, a current flows and the dissociated heat elevates the temperature within the subjacent electrically active device regions. A schematic cross section of the polyheater-device system is given in (b). Due to the lower chuck temperature, a temperature gradient arises between heater and wafer bottom.

The layout of an active transistor surrounded by a polyheater system and placed on a thermo chuck is illustrated schematically in Fig. 6.1. Depending on the ground temperature of the thermo chuck, a certain power has to be applied to the heater in order to reach a specified device temperature. From a layout point of view the heater should overlap the device considerably in order to guarantee a homogeneous temperature distribution over the whole active device area. The theoretical determination of the polyheater and the device temperature as a function of the power is very difficult [162] since considerable simplifications are required due to the fact that various back-end layers and materials usually have non-linear thermal conductivities [163]. In the following subsections an experimental procedure is presented, allowing to determine the power supply which is necessary to bring an active transistor to a certain device temperature. Thereby, we evaluate both the heater and device temperature which enables us to estimate the temperature gradient between poly wires and active device. Furthermore, we extract the transient heating and cooling characteristics for different target temperatures and discuss effects that may delay the restoration of thermal equilibrium and destabilize an adjusted temperature.

### 6.1.2 Calibration of heater and device temperature

In order to extract the polyheater and device temperature as a function of the applied power, we call on an appropriate temperature dependent parameter of the material. In the case of the highly doped poly silicon wire, the temperature dependent electric resistance ($R_{PH}$) would be such a parameter. In the case of the MOS transistor the forward current of the source/bulk diode or the source/drain current (around the threshold voltage of the device) can be used as an appropriate thermometer. For reliability issues, the source/drain current ($I_D$) is the preferred reference since it directly reflects the temperature of the interface between the silicon substrate and the gate oxide. This interface is of major interest because most studies suggest this region to be the location of concern for NBTI.

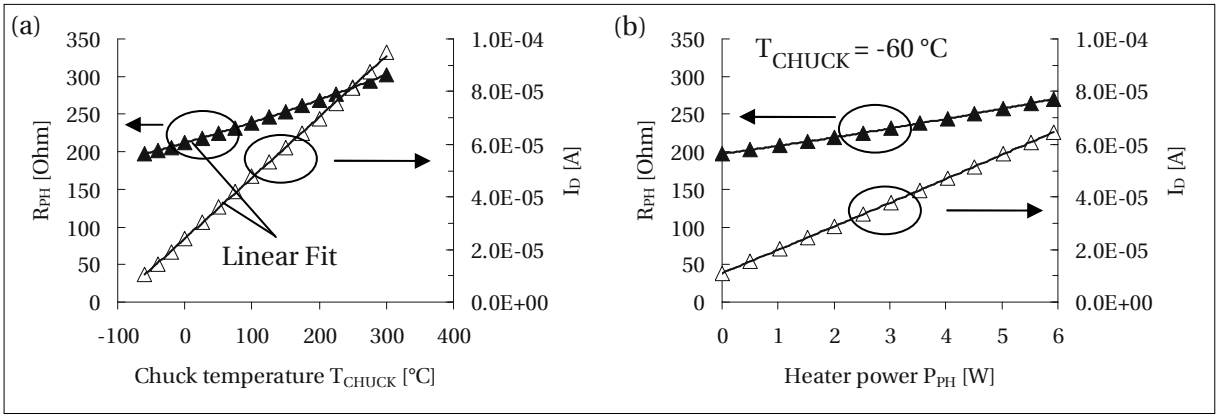


**Figure 6.2:** (a) The poly resistance $R_{PH}$ (full triangles) and the drain current $I_D$ (open triangles) as a function of the chuck temperature $T_{CHUCK}$. $R_{PH}$ is characteristic for the poly temperature while $I_D$ reflects the interface temperature of the active device. (b) The poly resistance $R_{PH}$ (full triangles) and the drain current $I_D$ (open triangles) as a function of the heater power $P_{PH}$ ($T_{CHUCK}$ = -60°C). The increase in the heater power causes a linear increase of $R_{PH}$ and $I_D$.

For the initial calibration we heat the wafer on the thermo chuck from -60°C to 300°C and record the poly resistance ($R_{PH}$) and the drain current ($I_D$) of our PMOS device at selected temperatures, cf. Fig. 6.2 (a). The sense currents and voltages must be chosen carefully in order to prevent self-heating during the measurement. Within the scanned temperature range the increase of the poly resistance and the drain current can be fitted very well by a polynomial of first (linear) or of second-order. From a physical point of view, the increase in the poly resistance can be explained by a reduction of the carrier mobility due to enhanced lattice scattering at higher temperatures while the increase in the drain current is originated in an enhancement of the concentration of thermally activated inversion carriers ($n_i$). Having determined the coefficients of the polynomial fit, we can calculate the poly and device temperature from $R_{PH}$ and $I_D$, respectively, when a power is applied.

In Fig. 6.2 (b) the poly resistance and the drain current is illustrated as a function of the power supply ($P_{PH}$). In this example the chuck temperature was -60°C. When applying zero watts ($P_{PH}$ = 0 W), the poly resistance and the drain current correspond exactly to their values extracted for -60°C in Fig. 6.2 (a). As we increase the power supply from 0 W toward 6 W, the poly resistance and the drain current grow simultaneously. In order to provide enough time for restoring thermal equilibrium, the power supply steps must be moderate. In the analysis $R_{PH}(P_{PH})$ can be converted into the temperature of the polyheater ($T_{PH}$) and $I_D(P_{PH})$ into the device temperature ($T_{DV}$).

The temperatures as a function of the power supply are illustrated in Fig. 6.3 (a). As can be seen, when applying a heater power of 6 W at a chuck temperature of -60°C, one may reach a device temperature of approximately 175°C. This corresponds to a maximum temperature range of $\Delta T_{DV}$ = 235K. Naturally, the correlation between the heater power and the accessible device temperature depends on the particular poly design. The higher the supplied power, the further the poly and the device temperature drift apart. For example, at a poly temperature of 150°C the device temperature is just 125°C. This is caused by the vertical temperature gradient between the heater and the device, cf. Fig. 6.1.

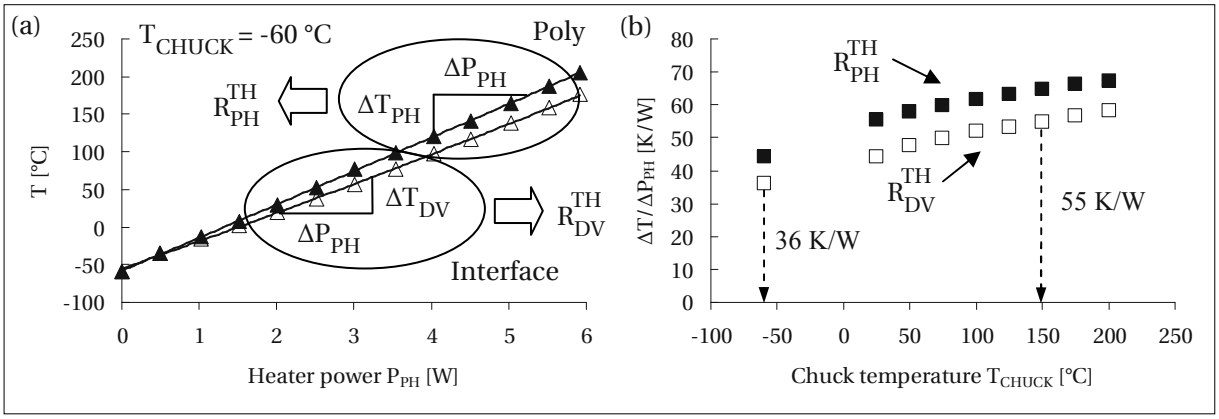


**Figure 6.3:** (a) The poly ($T_{PH}$; full triangles) and the interface temperature ($T_{DV}$; open triangles) when increasing the heater power ($P_{PH}$) at a chuck temperature of -60°C. The individual temperatures drift apart with $P_{PH}$ due to the vertical temperature gradient. The thermal resistances $R_{PH}^{TH}$ and $R_{DV}^{TH}$ are given by the slopes $\Delta T_{PH}/\Delta P_{PH}$, and $\Delta T_{DV}/\Delta P_{PH}$, respectively. (b) The thermal resistances of the polyheater ($R_{PH}^{TH}$; full squares) and the device ($R_{DV}^{TH}$; open squares) extracted for different chuck temperatures. The higher the chuck temperature, the steeper the temperature increase with heating power. $R_{DV}^{TH}$ is lower than $R_{PH}^{TH}$ because of the vertical temperature gradient.

From the slope $\Delta T/\Delta P_{PH}$ one can determine the individual thermal resistances. In Fig. 6.3 (b) the thermal resistances $R_{PH}^{TH}$ and $R_{DV}^{TH}$ of our particular heater design are depicted for different chuck temperatures. Obviously, $R_{PH}^{TH}$ and $R_{DV}^{TH}$ increase as the ambient temperature increases. This means that less power has to be applied at higher ambient temperatures in order to bridge the same temperature range $\Delta T$. This is due to the non-linear thermal conductivity of silicon. In fact, the thermal conductivity decreases as the temperature increases [162–164]. Consequently, the dissipated heat by the poly wires is distributed to a smaller area the higher the temperature. Due to the more concentrated power dissipation, poly and device heating become more efficient at higher chuck temperatures.

In many cases one is not interested in the thermal resistance and in the heating characteristics, making the calibration procedure for a single target temperature much less laborious. In principle, it is sufficient to determine one single target drain current at a certain target temperature. Therefore, we heat the thermo chuck to the desired temperature and record the corresponding target current at an arbitrary operating point (i.e. the threshold voltage of the device). When later (at a lower chuck temperature) determining the power supply necessary to reach this elevated target temperature, we simply increase the power supply incrementally while measuring the drain current in parallel at the same operating point as before. Once the measured drain current corresponds exactly to its calibrated target value, the required power supply is found.

### 6.1.3 Maximum accessible temperature range

A remarkable application of the polyheater technique can be found in the ability of reaching device temperatures far beyond the scope of conventional thermo chuck systems. For instance, an ultra

high target temperature can be reached when providing additional heating power at the maximum temperature range accessible by the thermo chuck. This feature enables us for example to determine activation energies of NBTI induced defects in a much wider temperature range than provided by our thermo chuck.

However, the calibration procedure of the polyheater in the high temperature regime (beyond the scope of the thermo chuck) requires more effort. This is because we cannot determine a target drain current in a temperature regime which is not accessible by the thermo chuck. Furthermore, previously established thermometers like the drain current change their temperature dependent characteristics as the semiconductor becomes intrinsic.

The high temperature calibration is illustrated in Fig. 6.4. In Fig. 6.4 (a) the same experiment as in Fig. 6.2 (b) was performed, however, this time at a chuck temperature of 150°C. As we increase the power supply at an ambient temperature of 150°C, the heater resistance ($R_{PH}$) and drain current ($I_D$) initially show a similar behavior as observed in Fig. 6.2 (b). Both parameters increase regularly until we reach a power supply of approximately 3 W. According to the thermal resistance of 55 K/W (cf. Fig. 6.3 (b)), 3 W applied at a ground temperature of 150°C should elevate the device temperature to approximately 315°C. Beyond this temperature range the doped silicon substrate becomes intrinsic which changes the current/voltage characteristics and temperature development of the drain current significantly. As can be seen in Fig. 6.4 (a), between 3 W and 5 W the drain current begins to deviate from its linear temperature dependent relationship (cf. Fig. 6.2 (a)) and can therefore no longer serve as a reliable thermometer. On the other hand, the poly resistance abides by its regular development which validates $R_{PH}$ still for poly temperature extrapolation.

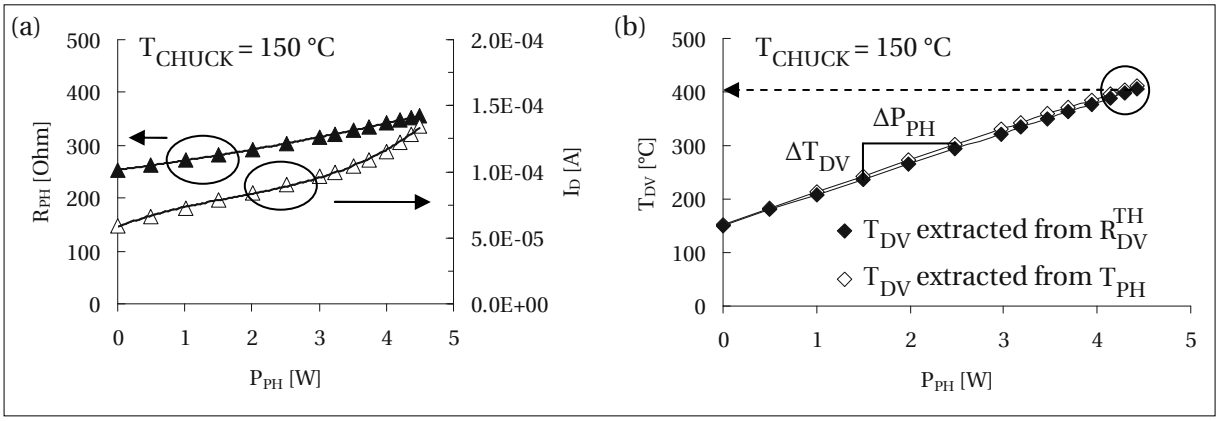


**Figure 6.4:** (a) The poly resistance $R_{PH}$ (full triangles) and the drain current $I_D$ (open triangles) as we increase the heater power at a fixed chuck temperature of 150°C. While the poly resistance still develops regularly in the high temperature regime, the drain current begins to deviate from its linear development as the silicon substrate becomes intrinsic ($T_{DV} \geq 300$°C). (b) The interface temperature ($T_{DV}$) versus the power supply in the high temperature regime; (i) extracted from the poly resistance, the poly temperature, respectively (open diamonds); (ii) extracted from the thermal resistance $R_{DV}^{TH}$ (full diamonds). Although both techniques are based on independent calculations, they give similar results for $T_{DV}$. At a chuck temperature of 150°C, interface temperatures up to 400°C (at 4.2 W) can be reached by providing additional heating power.

In Fig. 6.4 (b) we have determined the interface temperature of the device for an increasing power supply in two different ways. The first way is simply linked to the thermal resistance evaluated in Fig. 6.3 (b) for 150°C. According to $R_{DV}^{TH}$ at 150°C, the device temperature should increase by 55 K/W. We can use this result to calculate the device temperature in a straight-forward manner. The second way is a little bit trickier. From the poly resistance, which still follows its regular development in the high temperature regime, we can extrapolate at first the poly temperature as a function of the power supply. In a second step the device temperature is calculated from the poly temperature. The procedure is similar to the one demonstrated in Fig. 6.3 (a), however, this time for a chuck temperature of 150°C. As can be seen in Fig. 6.4 (b), the results of both methods agree very well and give similar device temperatures with a maximum discrepancy of about ±5 K. We remark that at a chuck temperature of 150°C one is able to exceed device temperatures up to 400°C which is far beyond the scope of most industrial thermo chuck systems. At the moment we turn off the heater, the temperature drops back to 150°C almost immediately. The detailed heating and cooling dynamics of the heater-device system are investigated in the next subsection.

### 6.1.4 Heating and cooling dynamics

In this subsection we elaborate on the time dependent heating and cooling dynamics of the device as a heating voltage/power is applied or removed, respectively. The chuck (ambient) temperature was -60°C during these experiments. By applying a certain heating voltage to the poly wires, the device temperature quickly elevates and stabilizes after a couple of seconds. On the other hand, when removing the heating voltage, the device cools down immediately. In order to determine the exact heating voltage necessary to reach a certain device temperature, the heater was calibrated in advance. The output of this initial calibration were 8 different heating voltages appropriate to heat the device from -60°C to -40/-20/0/25/50/75/100/125°C. In the experiments illustrated in Fig. 6.5, a heating voltage is abruptly applied while recording in parallel the heater current for 100 s. After this 100 s heating period the supply voltage was removed abruptly recording again the heater current in parallel for another 100 s. From the heating voltage and current characteristics the time dependent power dissipation of the heater is calculated.

As can be seen in Fig. 6.5 (a), when turning the heater on, it takes up to 1 ms until the maximum power dissipation is reached. This delay time is mainly limited by the finite speed of the voltage source. Using our particular heater design, heating voltages up to 34 V are required in order to overcome a temperature range of 185K (-60°C → 125°C). After the voltage source has stabilized (> 1 ms), the heater power tends to decrease slightly for a couple of seconds. This is because some time is needed to restore thermal equilibrium between heater, wafer, and thermo chuck. Consequently, the initial power decrease is more significant the larger the temperature difference between heater and chuck. When turning the heater off, the heater power vanishes within approximately 1 ms. Again, this 1 ms is originated in the finite speed of the voltage source.

In Fig. 6.5 (b) the same experimental sequence as in Fig. 6.5 (a) was performed but this time the drain current of the device was recorded as a representative for the interface temperature. By using the results of Fig. 6.2 (a) one can calculate the evolution of the device temperature $T_{DV}$ from $I_D$.

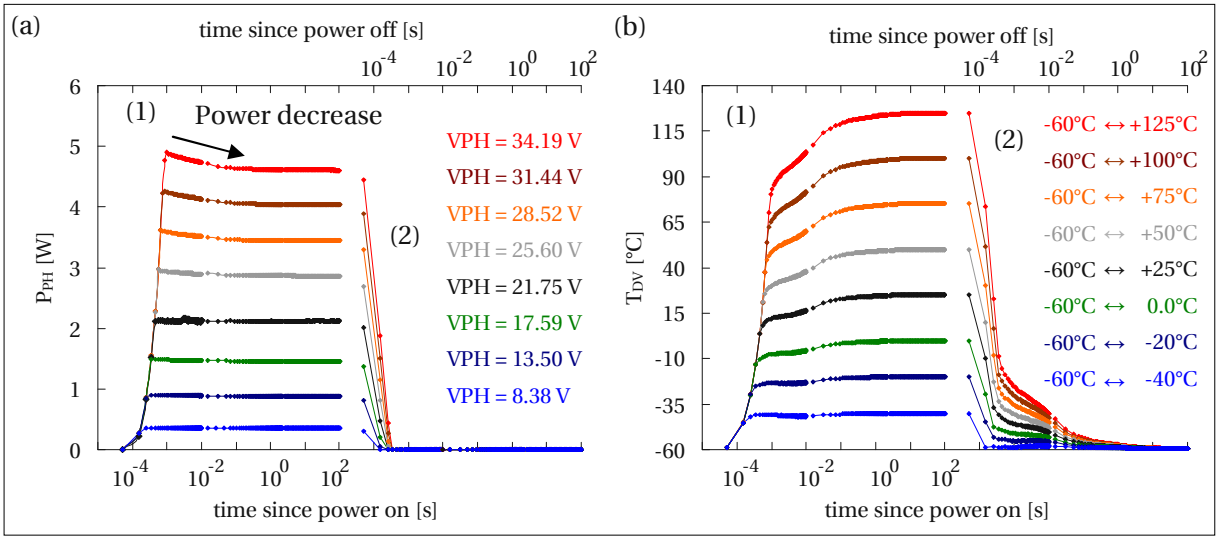


**Figure 6.5:** (a) The heating power when turning the heater voltage abruptly on (1) and off (2). The chuck temperature was -60°C. During the heater calibration specific voltages were determined to reach certain device temperatures. At the moment the heating voltage is turned on (1) or off (2), we record the heating current in parallel and calculate $P_{\mathrm{PH}}$. (b) The heating and cooling characteristics of the device when turning the heater power abruptly on (1) or off (2). The heater/device/chuck system needs a couple of seconds to restore thermal equilibrium causing a shoulder in the device temperature at the very beginning of the heating and cooling procedure.

As can be seen in Fig. 6.5 (b), when turning the heater on abruptly, it takes up to 10 s until the device has stabilized at its calibrated target temperature. The larger the temperature difference, the longer it takes to reach the target temperature. The larger time delay is due to the finite time interval necessary to restore thermal equilibrium a distance away from the actual heating source. The shoulder visible in the evolution of $T_{\mathrm{DV}}$ during heating is due to the power decrease illustrated in Fig. 6.5 (a). We remark that although heater power and heater temperature reach a maximum 1 ms after turning the heating voltage on, the device temperature does never exceed its target value due to the delayed thermal coupling between the polyheater and the device. This is an important aspect since we do not want to subject the device to an elevated pre-stress at the moment the heater is turned on. When turning the heater off, the situation is similar as during heating. It takes a couple of seconds until the excess heat generated by the polyheater can be removed by the thermo chuck. Following Fig. 6.5 (b), we come to the conclusion that at -60°C ambient temperature, any temperature switch up to ±200K can be executed safely within a time interval of maximal 10 s. In fact it takes approximately 0.1 s to reach the target temperature by 3 %, 1 s to obtain a 1 % accuracy and after 10 s the target temperature is adjusted by 0.1 % which corresponds to the measurement resolution as well.

### 6.1.5 Summary of the polyheater features

In the previous subsections the features and performance of the in-situ polyheater measurement technique was demonstrated. The temperature calibration procedure for determining the poly tem-

perature and device temperature was elaborated in detail for different ambient temperatures and power supplies. The thermal resistances of the polyheater and the device were found to depend on ambient temperature which is consistent with the non linear thermal conductivity reported for silicon. It was shown that a temperature range of more than 200K can be bridged by additional power supply provided by the polyheater. In particular, the ability of reaching device temperatures far beyond conventional thermo chuck ranges was pointed out. A thorough study on the heating and cooling dynamics of the device has revealed that a maximum time of 10 s is needed to switch the temperature within an interval of ±200K with a maximum precision of 0.1K. Thereby, the heating and cooling procedure was found to be nearly independent of the difference between ambient temperature (chuck temperature) and target temperature. Equipped with this features the polyheater tool exhibits a remarkable and unique tool for device reliability testing and characterization purposes which will be applied in the following sections for NBTI investigation.

## 6.2 On the temperature dependence of NBTI recovery

As demonstrated in the previous section, poly resistors around the device can be used to perform fast and reliable in-situ heating on a single device on wafer level. The following sections address the question how such a feature can be used to perform NBTI stress at a certain stress temperature, which generates a certain degradation level, while the recovery itself can be studied at arbitrary recovery temperatures. By turning the heater on during stress and switch it off during recovery, the tool enables us on the one hand (i) to bring identically processed devices to the same degradation level and on the other hand (ii) to fix a different temperature or vary the temperature in a defined way during recovery. By using this technique, our understanding of the recovery physics can be probed in a novel manner.

The motivation for temperature switches is, for example, to clarify whether the observed recovery dynamics, demonstrated in Chapter 3 are thermally activated or not. From a physical point of view, thermodynamic models are often linked to mobile hydrogen which can diffuse toward the gate oxide during recovery thereby passivating or creating dangling bonds at the $SiO_2$/Si interface and inside the bulk of $SiO_2$ [94, 96, 147, 165, 166]. However, by contrast previous experiments dealing with bias switches during recovery, cf. Chapter 5 and [78, 145, 167], demonstrated that hydrogen reaction and diffusion cannot be made responsible alone for degradation and recovery. As a solution carrier trapping via elastic tunneling was suggested [167]. However, an elastic tunneling process would be definitely too fast to explain long term recovery over many decades in time for ultra thin gate oxide devices. From a theoretical point of view such fundamentally different physical phenomena should show completely different dependencies on experimental parameters like the gate bias, the temperature and the recovery time. Also, the often observed log-like time dependence of $V_{TH}$ recovery over many decades of time or the bias and temperature dependent acceleration of degradation (cf. Section 3.4) requires a consistent physical explanation. To put in a nutshell, until now the microscopic degradation and recovery mechanisms as well as the relative amount and correlation of different contributions to the total threshold voltage shift and recovery are still controversial points in literature which require an unambiguous clarification in order to probe and formulate reliable degradation and recovery models [168, 169].

### 6.2.1 The principle of degradation quenching

A trivial problem encountered in the observation of temperature effects in NBTI recovery is the need to dispose of a set of (i) comparable devices that are brought to (ii) the same degradation level by identical stress, but which then (iii) recover at different temperatures.

The first condition (comparable devices) may be solved by careful sample selection, involving thorough characterization before stress. Statistics indicate that a good correlation between the degradation levels observed in equally fabricated devices on the same wafer is to classify them by their virgin CP current. The zero hour CP current gives information about the initial interface state density of the device and about the concentration of hydrogen stored in the oxide. The second condition (identical degradation level before recovery), however, cannot be solved by classically available stress and characterization methods, if the third condition (various recovery temperatures) has to be maintained. Classical methods are either based on performing stress and recovery at the same temperature, or strictly separate stress and recovery by a long, scarcely observable and basically undefined transition period. While the first approach can provide the observation of recovery at very good time resolution [170], it obviously always violates condition (ii) if condition (iii) is fulfilled and vice versa.

Therefore, in order to conserve the degradation level during cooling, the temperature switch has to be fast, well controlled and independent of the difference between stress temperature and recovery temperature. This demand cannot be fulfilled by a conventional thermo chuck system since the cooling duration of such systems is very long (>30 min) and strongly depends on the target temperature. As a consequence, one has to deal either with *additional degradation* when maintaining the stress bias applied during cooling or with *uncontrolled recovery* when leaving the device floating during cooling. Furthermore, applying a bias during the temperature switch is difficult (if not impossible) due to contact difficulties related to thermal expansion of probe needles and metal pads. In short, using the thermo chuck for temperature switches suffers from a lot of unacceptable systematic errors and drawbacks.

Our approach to harmonize all conditions and to get rid of the above described technical difficulties is to use the polyheater technique, cf. Fig. 6.6 (a). During stress a certain stress bias ($V_{GS}$) is applied to the gate and the (previously calibrated) heater generates an elevated device temperature ($T_S$) for a defined stress time $t_S$. Before initiating the recovery cycle, the heater is switched off, the device reaching ambient (chuck) temperature within a couple of seconds ($t_D$). In order to prevent any relaxation during the temperature switch, the stress bias remains applied within the delay time $t_D$. During $t_D$ stress continues in an undefined way, but this additional degradation is assumed to be very small compared to the degradation occurring within the main stress period typically performed at a much higher stress temperature ($T_S \gg T_R$). This was verified in Fig. 6.6 (b) where we have stressed different PMOS devices (SM6P/30/H3) for 1,000 s at $T_S$ = 125°C and $E_{OX}$ = 6.0 MV/cm and afterwards let them recover at $T_R$ = -60°C and $V_{TH}$. The cooling delay time $t_D$ between turning off the heater and switching the gate bias from $V_{GS}$ to $V_{TH}$ was varied between 0 s and 1,000 s. When using a delay time between 0 s and 1 s, the device has not reached the target temperature $T_R$ at the moment the stress bias is removed, cf. Fig. 6.5. Consequently, due to the larger temperature the measured $V_{TH}$ shift is afflicted with an error affecting predominantly the first couple of seconds after removal of the stress

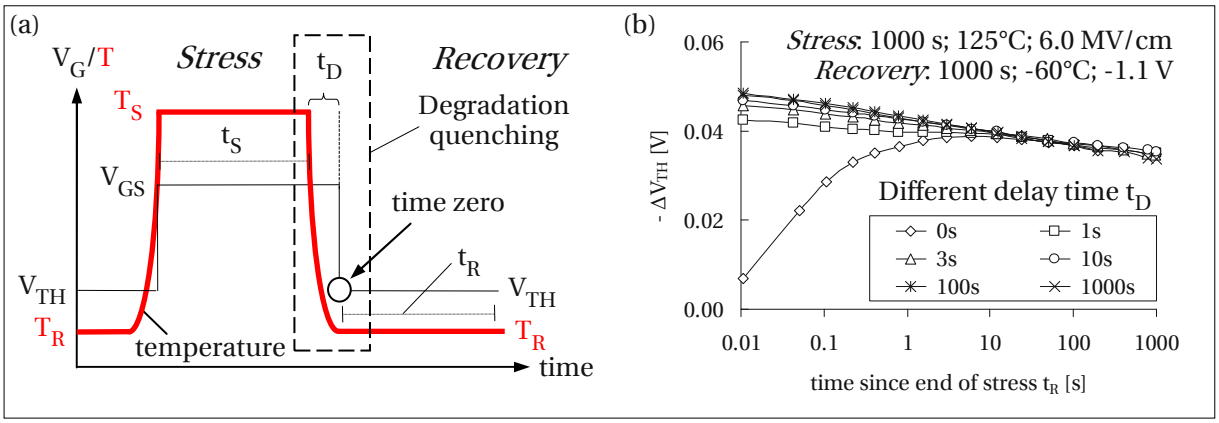


**Figure 6.6:** (a) The principle of degradation quenching. Subsequently to the initial characterization phase at the analyzing temperature $T_R$ and the gate bias $V_{TH}$, the heater is turned on, elevating the device temperature quickly toward the stress temperature $T_S$. Once at $T_S$, the stress phase is initiated by switching the gate bias from $V_{TH}$ to $V_{GS}$. After the stress time $t_S$ has elapsed, the heater is turned off. The stress bias remains applied for a delay time $t_D$ until the device is at $T_R$ (degradation quenching). The recovery cycle ($t_R$) is then initiated by switching the gate bias from $V_{GS}$ to $V_{TH}$. (b) Threshold voltage shift as a function of the delay time $t_D$. Different PMOS devices (SM6P/30/H3) were stressed for 1,000 s at $T_S$ = 125°C and $E_{OX}$ = 6.0 MV/cm. After degradation quenching, the $V_{TH}$ shift was monitored at $T_R$ = -60°C and $V_{TH}$ using different delay times $t_D$.

bias. However, when using a delay time ≥ 3 s, we obtain similar recovery characteristics for arbitrary delay times suggesting (i) that 3 s is sufficiently enough to reach the target temperature and (ii) that we can safely neglect additional degradation or recovery during $t_D$ provided the stress temperature exceeds the recovery temperature by far. Since the cooling time of the polyheater-device system is nearly independent of the temperature difference ($T_S$-$T_R$), statement (i) and (ii) hold independent of the recovery temperature provided $T_R$ is lower than $T_S$.

Having demonstrated the ability of the polyheater technique of conserving the degradation level during the temperature switch, the actual initialization of the recovery cycle ('time zero') may be defined in the following as the point in time when switching the gate bias from the stress level to the threshold voltage of the device. The entire procedure is called 'degradation quenching'. By 'quenching' the degradation level to an arbitrary recovery temperature in the way described above, we are going to thoroughly investigate the role of temperature in NBTI recovery in the following sections.

### 6.2.2 Recovery of identically stressed devices recorded at different temperatures

In order to investigate the temperature dependence of NBTI recovery, the following experimental procedure was performed on different PMOS devices (SM6P/30/STD1) by making use of the previously described degradation quenching method. During stress, the heater generates a defined interface temperature and a certain stress bias is applied to the gate, subjecting the devices to an oxide field of approximately $E_{OX}$ = 5.5 MV/cm. Source and drain are at zero volts. The stress phase was performed identically for all samples, creating a unique degradation level of each device at the end of $t_S$.

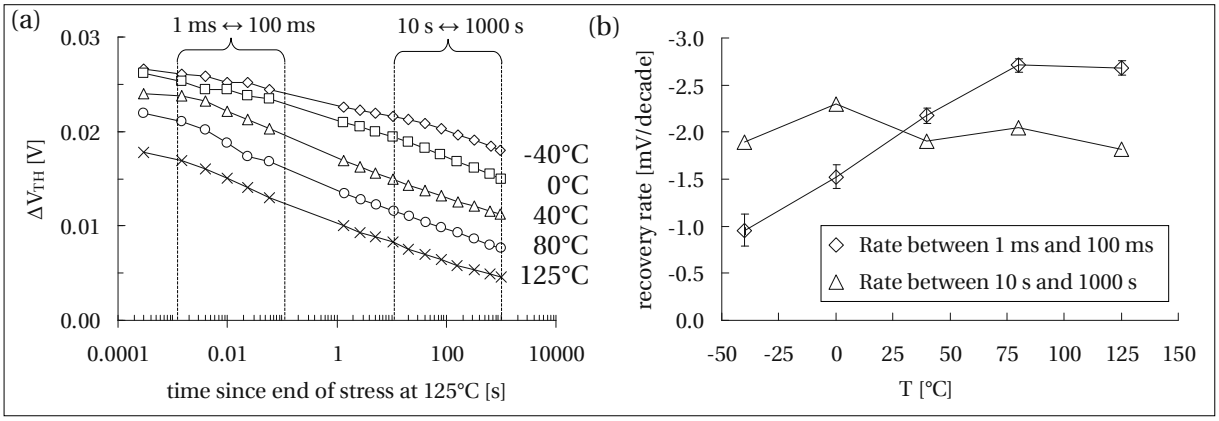


**Figure 6.7:** (a) Recovery of the threshold voltage shift recorded at different temperatures after stressing all samples at $E_{OX}$ = 5.5 MV/cm and $T_S$ = 125°C. Recovery conditions: $V_{GR}$ = -1.1 V; $V_{DR}$ = -2.5 V; $T_R$ = -40/0/40/80/125°C. (b) Temperature dependent recovery rate between 1 ms and 100 ms (diamonds) and between 10 s and 1,000 s (triangles).

After 1,000 s of NBTS, the heating current is taken away and the device cools down rapidly toward the individual ambient temperature which is defined by the temperature of the underlying thermo chuck. For example, when planning to study recovery at -40°C, the chuck has to be at that temperature already before stress. Note that the required power supply for the polyheater necessary to reach the unique stress temperature of 125°C, is the greater the lower the base temperature of the thermo chuck. One second after the heater has been turned off, the gate bias is switched to the threshold voltage ($V_{GR}$ = -1.1 V) of the device which terminates the stress abruptly and initiates the recovery cycle ($t_R$). In parallel, the drain bias is set to its read-out value ($V_{DR}$ = -2.5 V) in order to measure the recovery of the saturation drain current. The transition from stress to read-out bias conditions requires approximately 200 $\mu$s and is limited by the speed of the voltage source unit. The first measured current at the individual recovery temperature and bias conditions can be recorded about 300 $\mu$s after removal of the stress voltage and about one second after the heater was turned off. The time dependent evolution of the saturation drain current is converted into a stress/recovery induced threshold voltage shift [27], cf. Chapter 2.

The result of such a temperature quenched recovery measurement is illustrated in Fig. 6.7 (a). The unique stress temperature, supplied by the polyheater, was 125°C. The individual recovery temperatures were -40/0/40/80/125°C. Stress and recovery durations were 1,000 s respectively. As can be seen in Fig. 6.7 (a), the recovery curves look quite similar except for a temperature dependent offset which is already present at the first measurement point recorded 300 $\mu$s after removal of the stress field. Although the recovery temperature was varied by more than 160K with respect to the individual analyzing temperatures, there is no significant long-term temperature dependence visible in the recovery slopes. The recovery traces are apparently parallel.

In Fig. 6.7 (b) the recovery rate per decade is evaluated more precisely for the first one hundred milliseconds and for the last two decades of the recovery traces. On closer inspection, there is a temperature dependence visible within the first 100 ms right after stress. We observe an increasing slope of

the recovery curves with increasing temperatures. While at -40°C the amount of recovery is only 1 mV per decade, it is about three times larger for temperatures above 80°C. A reason for the initial temperature dependence might be the fact that the device is probably not exactly at the target temperature after the short cooling delay time ($t_D$) of only 1 s, cf. Fig. 6.6 (b). A few seconds after the termination of stress all traces become nearly parallel independent of $T_R$. After 10 s the decrease of the threshold voltage shift has leveled off to about 2 mV/dec for all samples. This holds at least for two or three decades in time.

The offset can be explained qualitatively by assuming an inelastic tunneling process and a homogeneous distribution of trap energy levels responsible for the observed log-like recovery traces. When lowering the analyzing temperature, all recovery time constants increase simultaneously, thereby shifting the entire recovery characteristics to larger recovery times. After demonstrating further examples, the issue is going to be discussed in detail by means of a first order model in Subsection 6.3.4. It has to be mentioned that a possibly remaining small offset can be explained by the temperature dependent position of the Fermi level at the threshold voltage. While at low temperatures the Fermi level is closer to the valence band edge, it moves further toward midgap the higher the analyzing temperature. Considering creation of interface states within the silicon bandgap as a result of NBTS, their charge state (occupation probability) is governed by the position of the Fermi level, cf. Fig. 2.2. Consequently, at lower temperatures more of them tend to be positively charged which is reflected by a larger threshold voltage shift. The temperature dependent variation of the Fermi level apparently causes a systematic error which is, however, believed to be too little to explain the full offset.

In summary, the obtained recovery characteristics are quite surprising, in particular, when considering that degradation during stress has usually a significant temperature dependence, cf. Section 3.4. A crucial point here is that the temperature independence of the recovery rate challenges models based on hydrogen dominated relaxation. For instance, in dispersive diffusion models hydrogen is believed to be stored in traps within the oxide, at the interface or somewhere else close to the interface. In order to passivate stress induced damage, the H atoms or $H_2$ molecules have to be released from there and overcome a thermodynamic barrier. At lower temperatures the probability of release as well as the diffusion rate of hydrogen is much lower. Considering such a mechanism to be the controlling process, one would expect freezing of recovery at -40°C. However, time dependent recovery is still observed even at such relatively low temperatures. The same argument also holds for the switching of hydrogen from a bonding to an antibonding Si–H configuration, as proposed by [171]. Since the transfer from a bonding to an antibonding configuration is a thermodynamical process, it should be highly temperature activated.

In order to investigate the role of interface states in the recovery process, CP measurements were performed at the end of the 1,000 s lasting constant bias recovery phases which are performed at different temperatures. Considering that the obtained $\Delta V_{TH}$ shifts are considerably different (at the end of the recovery phases), one would expect a similar difference in the remaining CP current as well provided interface state re-passivation is the dominating recovery mechanism. Such a difference is actually obtained, the -40°C data showing a considerably larger $\Delta I_{CP}$ than for example the 125°C data, cf. uncorrected data in Fig. 6.8 (b).

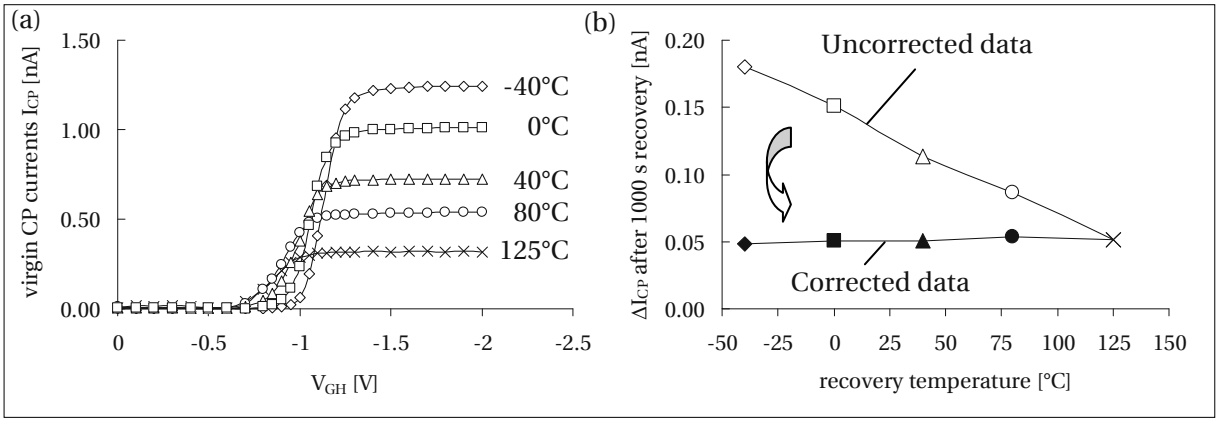


**Figure 6.8:** (a) Virgin CP current characteristics recorded at the individual recovery temperatures. The lower the temperature, the higher the CP signal due to a larger profiled active energy range ($\Delta E_{CP}$). (b) The remaining CP current degradation at the end of the constant bias recovery phases performed at different temperatures. CP setup: $V_{GB}$ = 1.0 V; $V_{GH}$ = -2.0 V; $f$ = 500 kHz; $t_r$ = $t_f$ = 375 ns. The remaining CP signal is the larger the lower the temperature, however, when accounting for the temperature dependent active energy interval ($\Delta E_{CP}$), the obtained differences in the uncorrected $\Delta I_{CP}$ data (open symbols) is removed. The corrected data (full symbols) reveals a similar remaining degradation level of the interface after 1,000 s constant bias recovery at different temperatures.

However, considering the temperature dependence of the scanned energy interval (cf. Fig. 2.8 (a)), the obtained difference turns out to be afflicted with a systematic error. The error may be corrected by referencing to the initial offset of the virgin CP currents recorded at the individual analyzing temperatures, cf. Fig. 6.8 (a). We remark that all samples show a similar virgin CP current, when recorded at the same temperature. After accounting for the temperature dependent offset, the remaining shift in the CP currents becomes independent of temperature, cf. corrected data in Fig. 6.8 (b). The result indicates that either no interface state recovery takes place at all, or interface state recovery is independent of temperature. We further note that the remaining offsets in the $\Delta V_{TH}$ shifts at the end of the recovery phases ($t_R$) cannot be explained by different interface state densities.

In order to investigate the temperature and time dependence of CP current recovery, a similar experiment as the one presented in Fig. 6.7 (a) was performed on a different PMOS devices (SM6P/30/STD2). After stressing all DUTs under the same stress conditions (125°C; $E_{OX}$ = 5.5 MV/cm, $t_S$ = 1,000 s), we have quenched degradation and monitored this time the CP current recovery immediately after the termination of stress. The recovery phase was performed under continuous gate pulsing conditions for 1,000 s at three different temperatures ranging from -60°C to 80°C.

As can be seen in Fig. 6.9 (a), due to the temperature dependence of the active energy interval, a systematic offset in the CP signal and different recovery rates appear. However, when scaling all curves to the first measured point (0.1 s post stress), they coincide (cf. Fig. 6.9 (b)) showing a unique recovery rate of 6.5–7.2 %/dec. In particular, there is a decrease in all CP signals of about 30 % within four decades in time (0.1 s – 1,000 s). The relative decrease is the same for all recovery temperatures indicating temperature independent interface state recovery during continuous gate pulsing.

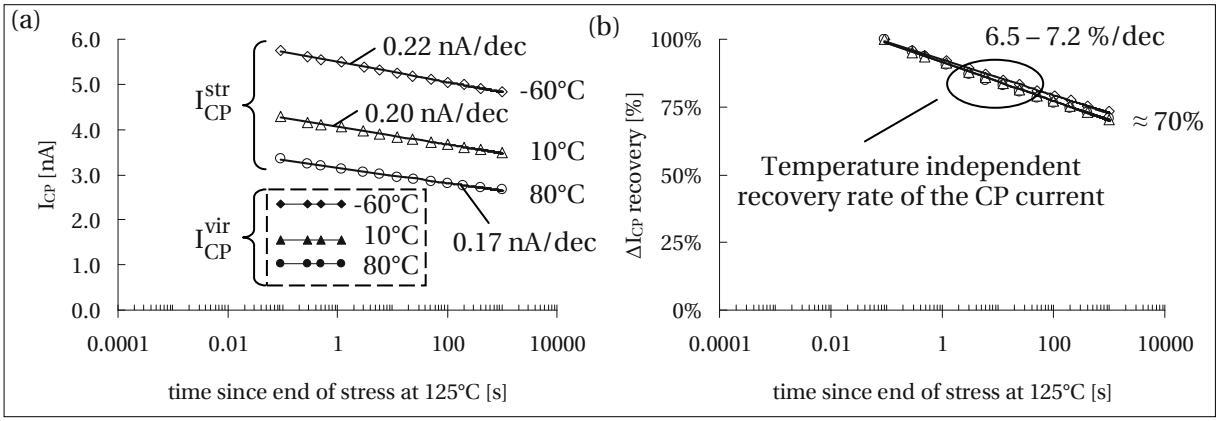


**Figure 6.9:** (a) Recovery of the maximum CP current (continuous CP) recorded at different temperatures after stressing all samples at $E_{OX}$ = 5.5 MV/cm and $T_S$ = 125°C. Bias and temperature conditions during recovery: $V_{GB}$ = 2.0 V; $V_{GH}$ = -2.0 V; $f$ = 1 MHz; $t_r = t_f$ = 125 ns; $T_R$ = -60/10/80°C. The CP currents measured at the individual analyzing temperatures are illustrated before stress ($I_{CP}^{vir}$ - full symbols) and after stress ($I_{CP}^{str}$ - open symbols). (b) The scaled recovery traces of $\Delta I_{CP}$ after stress. Independently of the recovery temperature, all samples show a similar recovery rate leaving behind approximately 70 % of the original degradation (measured 0.1 s post stress) after 1,000 s of pulsed recovery.

### 6.2.3 Identically stressed devices subjected to abrupt temperature switches

In Fig. 6.10 (a) the $V_{TH}$ recovery is illustrated for four different PMOS devices (SM6P/30/STD1). All devices were stressed at an oxide field of 5.5 MV/cm and at a temperature of 125°C for 1,000 s. Three reference devices recovered at a constant temperature of -40°C, 40°C, and 125°C, respectively. The forth device was subjected to two abrupt temperature switches by making use of the polyheater technique. Right after stress it recovered for 1 s at -40°C, then for 100 s at 40°C and finally for another 10,000 s at 125°C providing two decades of inspection at each particular temperature. As can be seen in Fig. 6.10 (a), when elevating the device temperature abruptly, $\Delta V_{TH}$ relaxation becomes accelerated approaching gradually the 40°C reference curve, respectively, the 125°C reference curve after a couple of seconds. Temperature accelerated recovery at constant gate bias conditions can definitely not be ascribed to elastic tunneling.

In Fig. 6.10 (b) we have performed the complementary experiment to Fig. 6.10 (a): at first, the device recovered at 80°C for 10 s. Afterwards, the heater power was lowered so that the device cooled down to 40°C. However, as opposed to accelerated recovery as a consequence of heating, cooling leads to frozen recovery for a certain interval of time. Indeed, recovery does not proceed before the cooled measurement curve reaches the 40°C reference curve. Again, frozen recovery at constant gate bias conditions cannot be ascribed by an elastic hole trapping model.

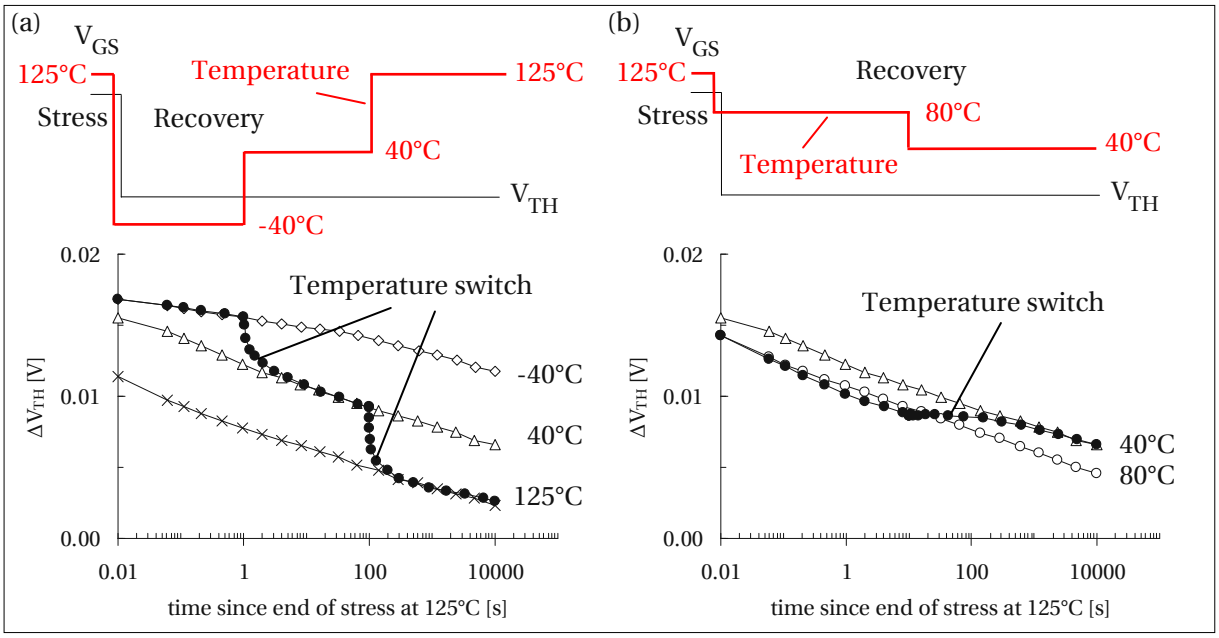


**Figure 6.10:** (a) Two step heating performed during constant bias recovery. Reference measurements at -40/40/125°C are illustrated by open diamonds/triangles/crosses. Recovery can be accelerated twice as we heat the device abruptly from -40°C to 40°C (after 3 s) and form 40°C to 125°C (after 100 s), cf. full symbols. (b) Cooling performed during constant bias recovery. Reference measurements at 40/80°C are illustrated by open triangles/circles. Lowering the recovery temperature (80°C → 40°C) leads to frozen recovery until the cooled (full symbols) reaches the reference curve at 40°C.

### 6.2.4 Identically stressed devices subjected to temperature ramps

In order to investigate the temperature dependence of the CP current and the $V_{TH}$ recovery in a more sophisticated way, the 'simple' temperature switch experiments was extended to full temperature ramps. This means, the device temperature was not changed abruptly by one or two large steps but was carefully increased in many small steps beginning at a relatively low temperature (i.e.: -60°C) and ending up finally at a much higher temperature (i.e.: the stress temperature: 125°C). Note that a careful heater and device calibration has to be performed in advance in order to accomplish the heater power → device temperature conversion and in order to investigate the individual threshold voltage shifts recorded at different device temperatures. All PMOS devices (SM6P/30/H1) discussed in the following were stressed at an oxide field of 5.5 MV/cm and at a temperature of 125°C for 1,000 s.

After degradation quenching toward -60°C, the temperature ramp experiment can be divided in four separate sections (A/B/C/D), cf. Fig. 6.11. The gate bias remains constant at $V_{TH}$ during the entire recovery phase. Section (A) was performed immediately after the termination of stress and lasts for 100 s at a temperature of -60°C. Section (B) is the actual temperature ramp where the device temperature was increased incrementally within 125 s from -60°C toward 125°C in 48 equidistant (3.85°C) steps. Having reached 125°C, the temperature remains fixed for another time interval of 1,000 s in section (C). Before section (D) the polyheater was switched off thereby returning abruptly back to the initial characterization temperature of -60°C for another 1,000 s.

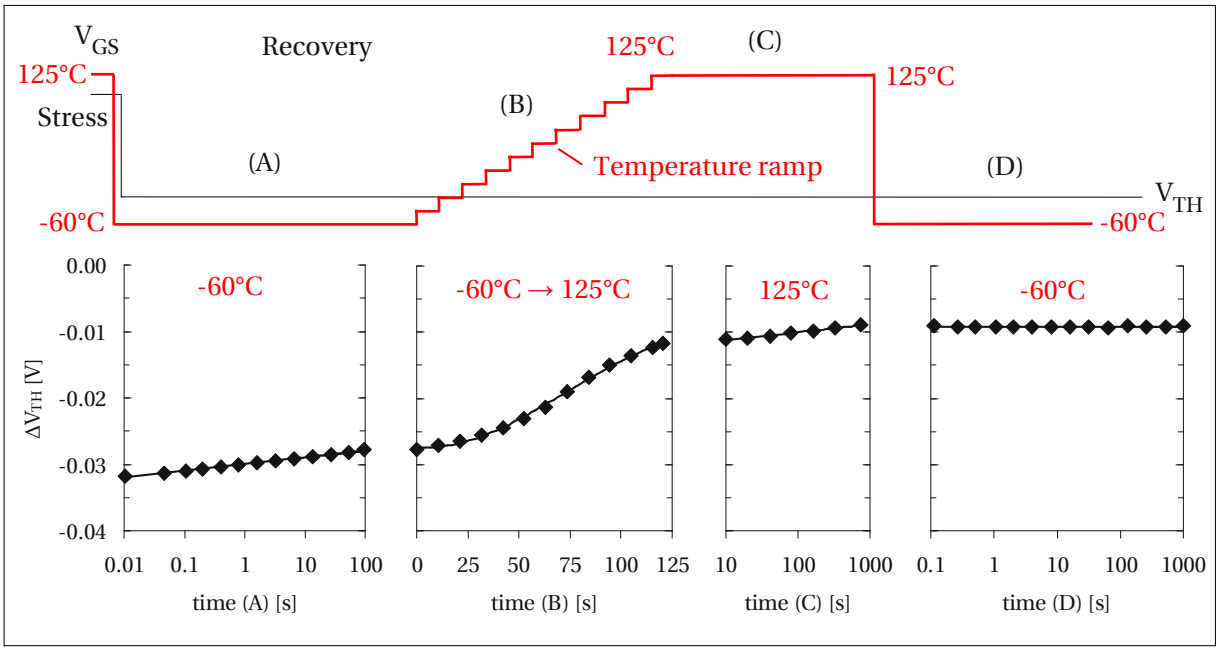


**Figure 6.11:** The recovery of the threshold voltage shift within the different stages of the temperature ramp experiment. The recovery sections (A/B/C/D) are depicted separately as a function of time. Note that during the temperature ramp in section (B) the time axis is linear. In section (A) at -60°C a standard log-like $V_{TH}$ recovery rate of 1.06 mV/dec is observed. During the temperature ramp in section (B) the $V_{TH}$ shift is reduced considerably by 16 mV. Within section (C) at 125°C recovery proceeds again log-like with are recovery rate of 1.17 mV/dec. As we finally cool down abruptly from 125°C to -60°C in section (D), the $V_{TH}$ degradation level remains frozen for the next 1,000 s.

In section (A) a log-like recovery characteristic is observed at -60°C. The recovery rate per decade within this section is approximately 1.0 mV/dec. After 100 s, we start a linear temperature ramp in section (B) beginning at -60°C and ending up at 125°C. As a consequence of heating, the observed $V_{TH}$ shift is reduced considerably by 16 mV indicating temperature accelerated defect annealing. In the following section (C) performed at 125°C, we obtain again a similar log-like recovery characteristic of the $V_{TH}$ shift as in section (A). The recovery rate in section (C) is approximately 1.2 mV/dec. In the last section (D), where we finally switch the temperature abruptly from 125°C back to -60°C, the degradation level remains frozen for the next 1,000 s. Note that the last measured $V_{TH}$ shift value of section (C) and the first measured $V_{TH}$ shift value of section (D) agree perfectly, although measured at vastly different characterization temperatures. This, on the one hand, confirms the reliability of our $V_{TH}$ shift extraction method for different temperatures and, on the other hand, proves that the observed temperature dependent recovery acceleration in section (B) is actually a real chemical relaxation effect and not a measurement artifact associated with different characterization temperatures.

To investigate CP current recovery under comparable experimental conditions, a similar experimental procedure was performed on a different PMOS device, cf. Fig. 6.12. To record the maximum CP current, the gate junction was pulsed between accumulation (2.0 V) and inversion (-2.0 V) during recovery, using a pulse frequency of 500 kHz and rising/falling times of 100 ns.

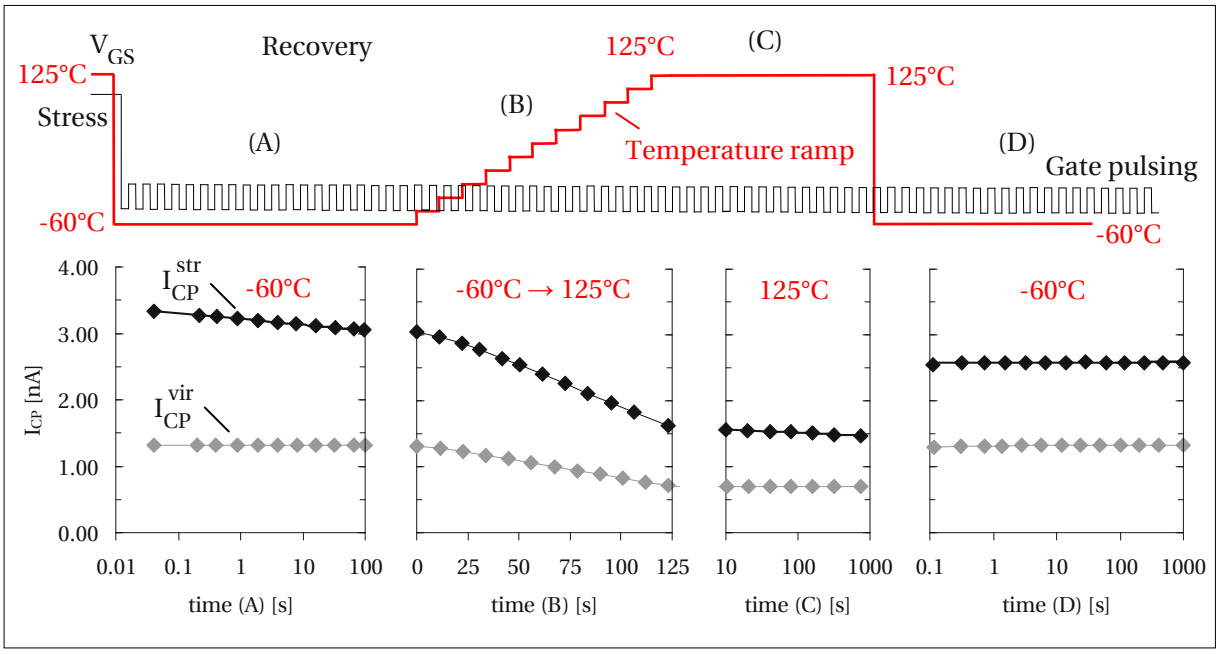


**Figure 6.12:** The virgin (gray diamonds) and the stressed (black diamonds) CP currents measured during the temperature ramp experiment. The sections (A/B/C/D) are depicted separately as a function of time. Note that during the temperature ramp in section (B) the time axis is linear. In section (A) we observe a log-like CP current recovery at -60°C. During the temperature ramp in section (B) both $I_{CP}^{vir}$ and $I_{CP}^{str}$ decrease considerably due to the shrinking of the active energy interval at elevated temperatures. Within section (C) at 125°C, only very small additional CP current recovery is obtained. When finally cooling down from 125°C to -60°C in section (D), the CP signal increases again abruptly due to the expanding of the active interval. The CP signal is frozen for the following 1,000 s at -60°C.

In order to estimate the systematic error in the CP signal caused by the temperature dependence of the active energy interval, an initial temperature ramp experiment was performed without subjecting the device to NBTI stress. Note that the evolution of the resulting $I_{CP}^{vir}$ curve reflects solely the temperature dependence of the CP current without being distorted by any recovery. As discussed in Subsection 2.2.2, by increasing the temperature, the scanned energy interval within the silicon bandgap becomes narrower because of the higher emission probability of previously captured inversion/accumulation carriers during the falling/rising edges of the gate pulses. As opposed to the unstressed device, the stressed CP signal $I_{CP}^{str}$ may be influenced by time and temperature dependent recovery mechanisms as well. In general, we expect that during the temperature ramp $I_{CP}^{str}$ will decrease (i) because of a contraction of the active energy interval ($\Delta E_{CP}$) and (ii) because of chemical interface state repassivation of stress induced $P_b$ centers.

In Fig. 6.12, the virgin (gray diamonds) and stressed (black diamonds) CP currents are illustrated within the different sections of the experiment as a function of time. $I_{CP}^{vir}$ was recorded before stress, $I_{CP}^{str}$ was recorded after stress. Following Fig. 6.12, we obtain a decrease in the virgin CP current during the temperature ramp in section (B) which is due to the narrower active energy interval at higher temperatures. Within the sections (A), (C) and (D), the virgin CP current $I_{CP}^{vir}$ is constant. After the last temperature switch from 125°C to -60°C (section (D)) the same $I_{CP}^{vir}$ is obtained as in section (A)

indicating that neither recovery nor stress has occurred as long as the device was not subjected to electrical stress. The reduction of $I_{CP}^{vir}$ within section (B) can be attributed completely to a reduction in the active energy interval.

As opposed to the virgin CP current $I_{CP}^{vir}$, the stressed CP current $I_{CP}^{str}$ does not exclusively decease in section (B) but also in section (A) and (C) where the temperature is constant. This suggests that moderate interface state recovery probably also occurs during the 125 s lasting temperature ramp in section (B). At the beginning of section (C) (-60°C), the CP current increases again considerably because of the temperature switch and nearly reaches its previous value obtained at the end of section (A). This indicates that the amount of real chemical interface state relaxation during section (B) and (C) is actually small (10 %). Consequently, the main reason for the observed decrease of the $I_{CP}^{str}$ signal in section (B) must be the active energy interval which is getting narrower with increasing temperature. We remark that during section (B) $I_{CP}^{vir}$ and $I_{CP}^{str}$ agree perfectly when multiplying $I_{CP}^{vir}$ by a factor of 2.3. This again suggests that chemical relaxation during heating is almost negligible which is consistent with the observation that CP current recovery is apparently independent of temperature, cf. Subsection 6.2.2.

### 6.2.5 Conclusions – T-dependence of $\Delta V_{TH}$ and CP current recovery

Based on the upper key experiments performed on identically stressed PMOS devices, one may draw the following conclusions on the temperature dependence of $\Delta V_{TH}$ and CP current recovery:

1. Threshold voltage recovery is accelerated considerably by elevating the temperature. On the other hand, when decreasing the temperature during recovery, the degradation level remains frozen for a certain interval of time.
2. The mechanism causing $\Delta V_{TH}$ recovery at elevated temperature is a true chemical relaxation process which is not reversible by subsequent device cooling.
3. CP current recovery is only marginally influenced by either heating or cooling, suggesting interface state repassivation to play only a minor role in recovery as long as the gate bias is maintained constant around the $V_{TH}$.
4. The recovery rates of the $V_{TH}$ shift and the CP current (recorded under continuous gate pulsing conditions) are widely independent of temperature. This holds at least for long-term recovery measurements recorded between 1 s and 1,000 s after the termination of stress.

Since our measurements show both bias (cf. Chapter 5) and temperature dependence (cf. Chapter 6), but support neither elastic tunneling nor interface state repassivation, a different mechanism has to be responsible for the observed recovery characteristics. Because bias dependence is totally incompatible with a diffusion process of neutral hydrogen species, we take the observed read-out voltage sensitivity as an indication for a trapping/detrapping phenomenon and attempt to expand the idea of elastic carrier exchange to a temperature sensitive model.

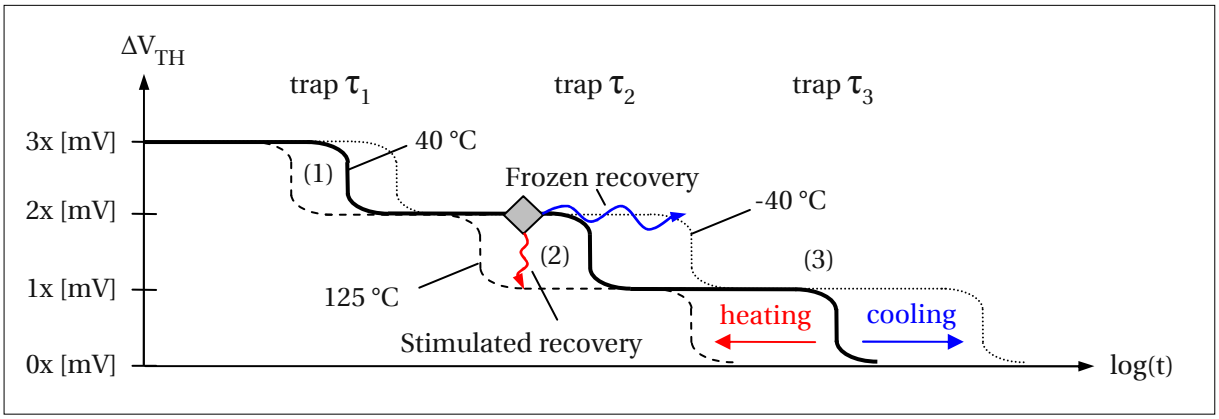


**Figure 6.13:** First order model of temperature dependent recovery effects using a simple picture of three different traps having three different time constants, barrier heights, respectively. Each trap is assigned to an arbitrary threshold voltage shift of 1x mV. Heating or cooling shifts the recovery curve to the left or the right (shorter or longer time constants) resulting in stimulated or frozen recovery. The diamond indicates a hypothetic temperature switching event.

As opposed to elastic tunneling, inelastic phonon assisted tunneling is temperature dependent [172]. Oxide defects and valence band electrons having different energetic positions cannot exchange carriers elastically. However, if they gain energy from lattice vibration (phonons) they may pass the thermodynamical tunneling barrier $\Delta E_B$ with a certain temperature dependent probability [173].

The lifetime of a single trap can be expressed by an Arrhenius law:

$$\tau(\Delta E_B, T_R) = \tau_0 \exp\left(\frac{\Delta E_B}{k_B T_R}\right), \tag{6.1}$$

where $\tau(\Delta E_B, T_R)$ is the inelastic tunneling lifetime of a single trap, $\tau_0$ is the pseudo-elastic tunneling exchange time between a trap and a substrate carrier at a barrier height zero, $\Delta E_B$ is the thermodynamical tunneling barrier, $k_B$ is the Boltzmann constant and $T_R$ is the analyzing temperature. At a constant temperature the time constants of different traps are solely determined by their individual barrier heights $\Delta E_B$.

When assuming NBTI recovery to be mainly determined by the neutralization of positive oxide defects via electron capture from the silicon substrate (respectively hole emission into the silicon substrate), the observed threshold voltage recovery can be interpreted as a continuous decay of traps with different barrier heights $\Delta E_B$.

Based on this idea, we can schematically illustrate the $\Delta V_{TH}$ recovery curve for three different traps having different thermodynamical barrier heights, cf. Fig. 6.13: According to their individual barrier heights, each trap has a certain characteristic time constant at which it recovers with maximum probability. A variation of temperature ($T_{R1} \longrightarrow T_{R2}$) impacts all time constants in parallel thereby shifting the plateaus along the time axis in log scale.

The respective shift in time for a trap with barrier height $\Delta E_B$ can be calculated as

$$\log(\tau(\Delta E_B, T_{R1})) - \log(\tau(\Delta E_B, T_{R2})) = \frac{\Delta E_B}{k_B T_{R1}} - \frac{\Delta E_B}{k_B T_{R2}}. \tag{6.2}$$

For $T_{R2} > T_{R1}$, the plateaus will shift to the left since the time constants of all traps will decrease, for $T_{R2} < T_{R1}$ the time constants increase leading to a shift to the right. Note that the widths of the plateaus are proportional to $\Delta E_B$. The trap level which recovers first ($\tau_1$), has the lowest barrier ($\Delta E_{B1}$) and is therefore least temperature dependent. On the other hand, trap levels with higher barriers ($\Delta E_{B2}$ and $\Delta E_{B3}$) depend stronger on temperature which results in a more significant temperature impact on the plateau broadness.

In this first order model, heating or cooling the device during recovery leads to stimulated recovery (at the diamond, stepping from the solid to the dashed line in Fig. 6.13, cf. Fig. 6.10 (a)) or frozen recovery (at the diamond, stepping from the solid to the dotted line in Fig. 6.13, cf. Fig. 6.10 (b)) compatible with our measurement results. When further assuming that the barrier $\Delta E_B$ itself can be lowered by a bias change, the model covers also bias change experiments and includes 'mathematically' elastic tunneling in the limit $\Delta E_B = 0$. Furthermore, homogeneously distributed thermodynamical barriers would lead to a large variety of time constants resulting in a large number of small steps like the ones described in Fig. 6.13. In a realistic experiment (large device), one would therefore expect large amounts of small steps to be smeared out as a straight line in a $\log(t)$ diagram consistent with our recovery experiments.

## 6.3 Classification of defects by their recovery characteristics

Having identified the gate bias and the temperature as the main impact factors governing defect neutralization and annealing during NBTI recovery, the following investigations address a combined study of gate bias ramps and temperature switches aiming to unambiguously clarify the basic characteristics of recoverable defects introduced during NBTS. In this context, the polyheater technique allows to control temperature and gate bias independently, thereby enabling us to perform low temperature measurements right after stressing the device at a much higher stress temperature (degradation quenching). A low characterization temperature is not only beneficial since it expands the accessible energy range during CP, thereby improving the measurement resolution and the DOS approximation but also because it decelerates thermodynamical trapping/detrapping mechanisms making a larger portion of the initial degradation level visible after the termination of stress [174].

The fundamental approach of this study is to classify defects by their individual recovery characteristics. The experimental setup is particularly adapted to probe an established damage/recovery hypothesis based on E′ centers [83, 175] that are believed to be generated as a result of NBTS and then can act as switching traps during recovery [108, 152, 176]. This so-called Harry Diamond Laboratories (HDL) model was introduced by Lelis following irradiation studies [110] later confirmed by ESR studies [102] and finally revisited and extended by Grasser *et al.* in the context of NBTI [144].

The microscopic model transitions involved in the classical and extended HDL model are rigorously discussed in Section 4.3.

The study clearly reveals three different types of defects. All defect types have in common that their charge state depends on the position of the Fermi level and that they introduce a broad density of state profile in the vicinity of or within the silicon bandgap. Defect (I) is fully recoverable, defect (II) is similar to defect (I) in terms of DOS but does not recover that easily, while defect (III) can be attributed to the conventional interface state ($P_b$ center). From this study we come to the conclusion that the carrier trapping and detrapping characteristics of stress induced defects can be controlled by temperature and Fermi level in a similar way. However, there is also irrevocable structural relaxation which is mainly influenced by the temperature. Based on these ideas, a measurement method has been developed which can be used to energetically profile the relaxation of stress induced oxide defects.

### 6.3.1 Basic experimental approach and proposed underlying model transitions

To investigate bias dependent trap annealing, we study hysteresis in low temperature (-60°C) gate bias ramps ('down sweep' and 'up sweep') after stressing PMOS devices (SM6P/30/STD2) for 1,000 s under NBTI (-5.5 MV/cm; 125°C). Without loss of generality, it is assumed that during stress interface and oxide traps can be created and exchange carriers with the silicon substrate. Since during stress the device is in strong inversion, the Fermi level is pinned close to the valence band edge and most interface states are considered as positively charged. Since there are almost no electrons present at the interface as long as the gate bias is at the stress level, hole capture dominates the carrier exchange process between NBTS induced traps and the silicon substrate, resulting in predominantly positively charged defects at the end of stress [21, 140, 177].

After stress, degradation is quenched [168] from 125°C to -60°C. The basic experimental setup following electrical stress and the proposed microscopic model transitions linked to the gradual Fermi level sweeps are discussed in Fig. 6.14. The start of trap characterization during recovery is given by the switch of the gate voltage from the stress level (-17.0 V) to a much lower recovery bias (-1.7 V). As long as the new gate bias value still corresponds to strong inversion, the Fermi level remains pinned close to the valence band edge. From this point on a 'down sweep' toward accumulation is started turning the device gradually off. As soon as the gate bias approaches and then passes the threshold voltage (≈-1.0 V) during the 'down sweep', the Fermi level quickly moves from the valence band edge toward the conduction band edge which influences the free carrier concentrations at the interface considerably. The Fermi level positions at the particular gate voltages were simulated numerically for the tested device [29]. When approaching accumulation, new equilibrium conditions for carrier exchange between the silicon substrate and the positively charged oxide defects develop, increasingly favoring electrical neutralization of the switching traps the closer the Fermi level approaches the conduction band edge (transition (2) → (3)). Basically, this (2) → (3) transition (which we propose to call electrical neutralization, since electrically visible damage disappears upon neutralization of positively charged defects) is reversible, that is, a subsequent 'up sweep' leads to recapture of a hole from the substrate and restores the electrically visible damage in state (2) (transition (3) → (2)).

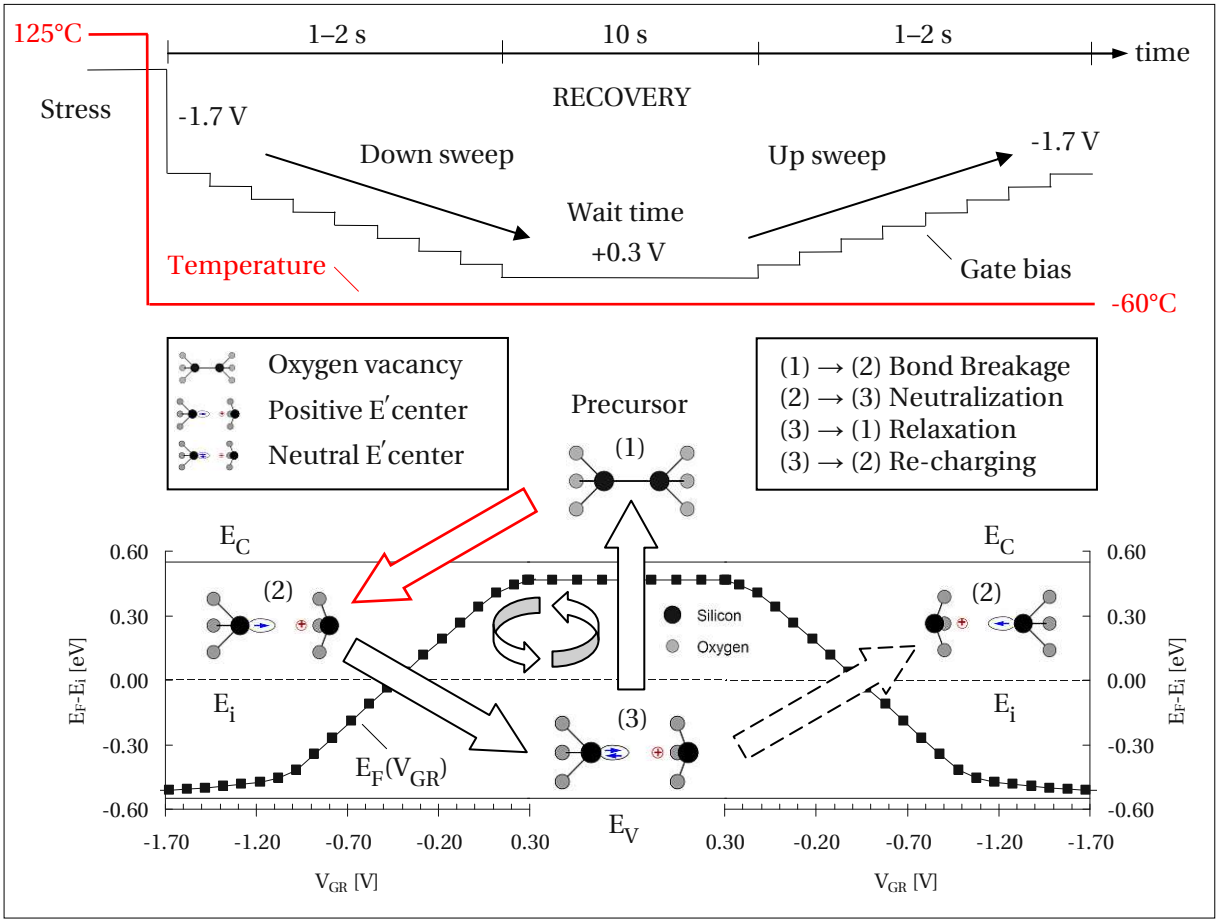


**Figure 6.14:** A schematic description of the trap characterization experiment. During stress at 125°C and -5.5 MV/cm oxygen vacancy precursors (1) are believed to break up, forming positively charged E′ centers (2). After the termination of stress, a 'down sweep' toward accumulation is performed at -60°C driving the Fermi level gradually from the valence band edge toward the conduction band edge, thereby neutralizing previously positively charged E′ centers (3). Once neutralized, the E′ center may undergo structural relaxation during the intermediate wait phase thereby restoring the original precursor state. Neutralized E′ centers which do not undergo structural relaxation may become positively charged again during the subsequent 'up sweep'.

However, provided the defect remains electrically neutral for a sufficiently long time, the trap can undergo a structural relaxation ((3) → (1)) which lowers the defect level in energy and brings it back to the initial precursor state (1). This recovery is irrevocable (as opposed to the reversible electrical recovery mentioned before) in the sense that a re-stress is required to trigger the transition (1) → (2). Since relaxation can only occur when the trap state is neutral (that is, in state (3)), it will happen predominantly during the 10 s 'wait phase' of our experiment which lasts much longer than the actual ramping durations (1–2 s). Optionally, the target bias of the 'down sweep' can be varied or the temperature during the 'wait phase' can be increased which will enhance the relaxation process (3) → (1), provided it is temperature activated. Note that increasing the temperature at a constant gate bias does not modify the Fermi level position considerably.

Our analysis is based on the observed difference in the $V_{TH}$ shift between the two gate bias ramps ('down sweep' and 'up sweep')) which provides insight into the above mentioned transitions: Pure electrical neutralization ((2) → (3)) is reversible, meaning that the positive charge can be restored in the 'up sweep'. On the other hand, a 'real chemical' relaxation ((3) → (1)) is irreversible under the applied bias conditions of the sweep which results in observable differences in the 'up sweep' compared to the 'down sweep'.

Having explained the concept of our basic experimental approach, tailored to the physical mechanisms which we expect to observe, we will briefly summarize the detailed experimental conditions and present the results in the following subsection.

### 6.3.2 Individual experimental setups and results

After the stress phase, every device discussed in the following is at -60°C and 'per definition' at the same degradation level. From that point on, the experiment was carried out in four different ways. The results and the individual measurement sequences are illustrated in Fig. 6.15.

**Discussion on the experimental sequences**

(a) Right after stress, the gate voltage is ramped down from deep inversion (-1.7 V) toward depletion (-0.7 V). The 'down sweep' takes about one second and consists of 50 equidistant voltage steps of 20 mV. After remaining for 10 s at -0.7 V, the gate junction is ramped back from -0.7 V to -1.7 V ('up sweep'). The sequence is completed by a final charge pumping measurement.

(b) Same sequence as in (a), however, during the wait phase in depletion, the polyheater was turned on in order to heat the device to 125°C for 10 s.

(c) Same sequence as in (a), however, 'sweep down' and 'sweep up' were recorded from deep inversion (-1.7 V) till accumulation (+0.3 V) the intermediate 10 s wait period being at +0.3 V.

(d) Mixture of sequence (b) and (c). During the 10 s wait period in accumulation (+0.3 V) a heating pulse of 125°C was applied using the polyheater.

In the analysis of the experiments (a) to (d) a gate voltage dependent hysteresis effect is evaluated that emerges, when the 'down seep' and the 'up sweep' are compared to the virgin transfer curve. These particular stress/recovery induced shifts are depicted as gate voltage dependent threshold voltage variations in Fig. 6.15. Note that as the gate bias exceeds depletion (-0.7 V) and approaches accumulation (+0.3 V) in experiment (c) and (d), the $V_{TH}$ shift cannot be monitored anymore. This is a general experimental limitation originating from the lack of an inversion channel as the Fermi level approaches and finally exceeds the intrinsic level $E_i$ (more electrons than holes at the interface), cf. Section 5.1.

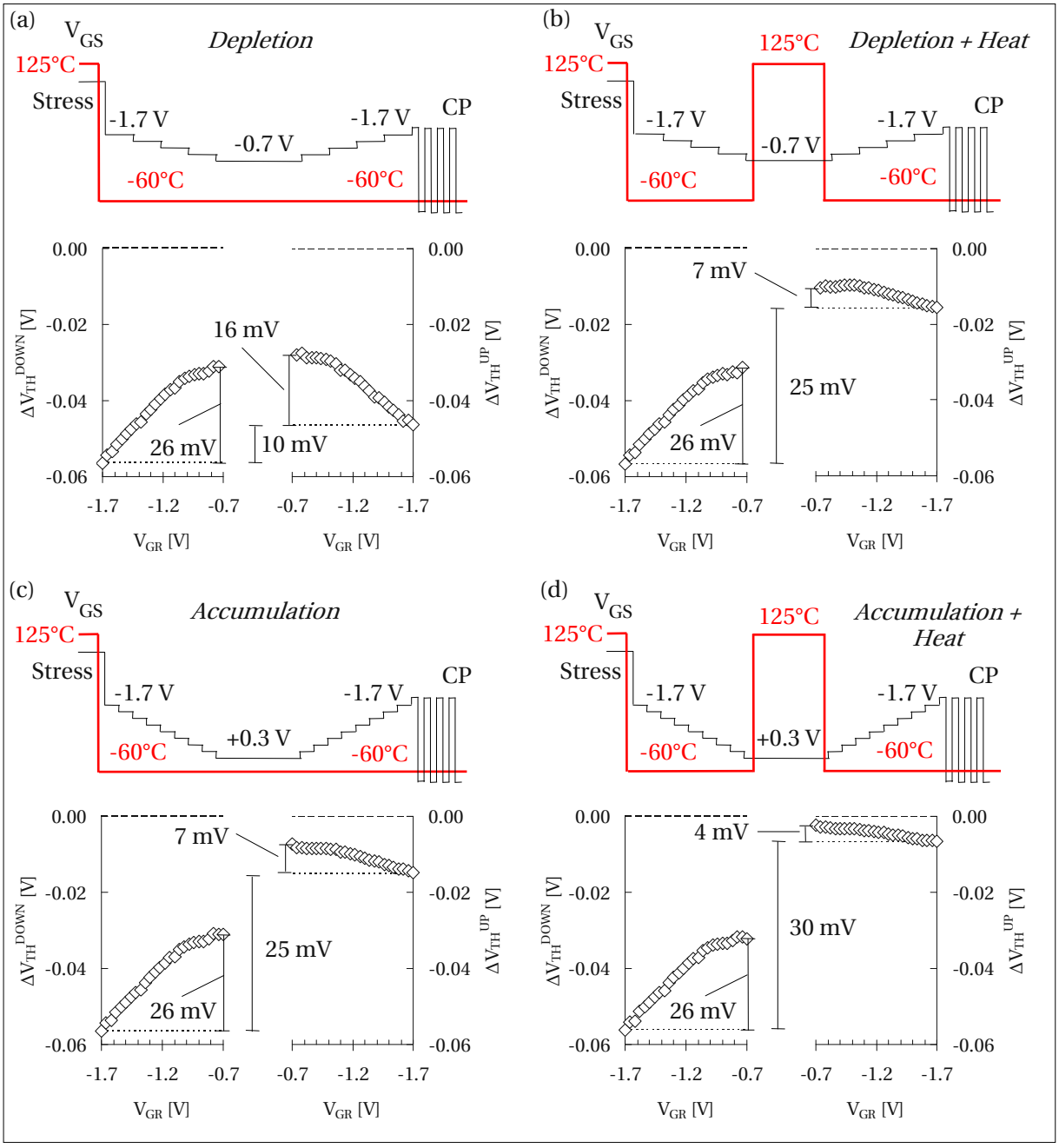


**Figure 6.15:** The $V_{TH}$ shift during 'sweep down' (left hand side) and 'sweep up' (right hand side) as a function of the read-out gate bias ($V_{GR}$). Experiments were performed according to the measurement sequences (a), (b), (c) and (d).

The $V_{TH}$ shift in the 'down sweep' is identical for all experiments discussed in Fig. 6.15. A considerable reduction of $\Delta V_{TH}$ by 26 mV is obtained in all setups when driving the gate voltage from deep inversion (-1.7 V) toward depletion (-0.7 V) right after stress. The fact that this reduction is the same for all devices serves as a consistency check considering that the measurement sequences are identical for (a), (b), (c) and (d) until the gate bias exceeds the depletion voltage of -0.7 V. The good correla-

tion is evidence for a comparable degradation level of all devices at the end of stress. Keeping the model of energetically widely distributed oxide defects in mind, the 26 mV reduction in $\Delta V_{TH}$ during 'sweep down' is mainly caused by electrical oxide trap neutralization as the Fermi level travels from the valence band edge toward midgap (transition (2) → (3)), cf. Fig. 6.14. Possible alternative interpretations and objections are discussed in detail by means of separate experiments in the Appendix of [174].

While 'sweep down' is identical for all sequences, the results of 'sweep up' depend significantly on the gate bias and on the temperature applied during the 10 s wait phase. According to the oxide trapping/detrapping model, the difference in the $\Delta V_{TH}$ shift between 'sweep down' and 'sweep up' is mainly caused by irreversible relaxation of neutralized switching oxide traps ((3) → (1)). Repassivation of interface states and time delay effects are assumed to be negligible. Justification for these assumptions is given the Appendix of [174].

In the following, the results of the particular experimental sequences are discussed separately and agreement with the predicted structural transitions is highlighted.

**Discussion of the experimental results**

(a) After a 10 s wait time at -60°C and a gate voltage of -0.7 V, the gate bias is ramped up again toward inversion (-1.7 V). During 'sweep up' an increase in the $V_{TH}$ shift of about 16 mV is obtained, which is, however, not as large as the amount of $\Delta V_{TH}$ lost during 'sweep down' (26 mV). According to the hole trapping model of Grasser *et al.* [144] the final difference in the $V_{TH}$ shift of 10 mV between 'sweep down' and 'sweep up' can be attributed to permanently recovered oxide defects (transition (3) → (1)). On the other hand, the remaining increase of 16 mV in 'sweep up' indicates that electrical neutralization of oxide traps does not necessarily result in a structural relaxation which would erase them irrecoverably. Some of the defects can obviously be recharged positively again as the gate bias re-approaches inversion during 'sweep up' (transition (3) → (2)).

(b) This sequence studies the influence of temperature on the permanent recovery of oxide defects. As opposed to experiment (a), the device is heated up to 125°C during the 10 s wait phase in depletion between 'sweep down' and 'sweep up'. Remarkably, as a consequence of the 10 s lasting temperature pulse, a 25 mV reduction in the $V_{TH}$ shift emerges, the remaining increase in 'sweep up' being only 7 mV. This indicates that device heating supports structural relaxation suggesting transition (3) → (1) to be temperature activated.

(c) This sequence studies the influence of a Fermi level shift toward the conduction band edge on the permanent recovery of oxide defects. As opposed to sequence (a), 'sweep down' in sequence (c) drives the device from inversion (-1.7 V) till accumulation (+0.3 V), while keeping the gate bias there during the 10 s wait phase at -60°C. During the subsequent 'sweep up' again a hysteresis of about 25 mV emerges in the $\Delta V_{TH}$ shift, indicating that during accumulation irrevocable oxide trap recovery (transition (3) → (1)) is much more effective than during depletion, cf. sequence (a). This is likely due to the fact that in accumulation almost all state (2) traps

become electrically neutralized in state (3). Consequently, the more traps are in state (3), the higher is the possibility of a (3) → (1) transition. The remaining increase in the $\Delta V_{TH}$ shift during 'sweep up' is considerably reduced to about 7 mV just like in sequence (b) which is consistent with the suggestion that most oxide traps visible in 'sweep down' have already been erased irrevocably during the wait phase in accumulation. The origin of the obtained equivalence of sequence (b) and (c) can be explained as follows: During the low temperature accumulation phase in experiment (c) most oxide traps become immediately electrically neutralized (transition (2) → (3)). Consequently, at the beginning of the 10 s accumulation phase nearly all traps are in state (3) and therefore available for the (3) → (1) transition. Although the probability of structural relaxation is low at -60°C, some of the state (3) traps will still recover permanently within the 10 s wait phase. On the other hand, when biasing the device in depletion during the wait phase, a smaller fraction of the previously generated state (2) traps are in state (3) at the beginning of the 10 s wait phase. As a consequence, less (3) → (1) transitions are expected, cf. sequence (a). However, when heating the device in depletion (cf. sequence (b)), the temperature activated (3) → (1) transition becomes accelerated considerably. Furthermore, additional (2) → (3) → (1) transitions are likely to occur as a consequence of inelastic tunneling ((2) → (3)) and subsequent structural relaxation ((3) → (1)).

(d) A combination of sequence (b) and (c), accumulation phase plus heating pulse, leads to the highest degree of permanent relaxation. The hysteresis in the $V_{TH}$ shift between 'sweep down' and 'sweep up' exceeds 30 mV which correspond to almost full recovery except for a small apparently permanent offset. It seems that nearly the entire oxide charge contribution, visible as 26 mV in 'sweep down', has been neutralized (transition (2) → (3)) and subsequently recovered permanently (transition (3) → (1)) by heating the device in accumulation. This is consistent with the Fermi level dependence of the (2) → (3) neutralization and the temperature acceleration of (3) → (1) relaxation. The small remaining $\Delta V_{TH}$ increase of about 4 mV in 'sweep up' and the constant offset of 3 mV visible at -0.7 V can be attributed to interface state charging and possibly positive locked-in oxide defects emerging as a result of hydrogen exchange between passivated interface traps (Si–H bonds) and state (2) oxide traps (E′ centers), cf. Section 4.3.

**Discussion on the role of interface states**

It is an important finding that interface states ($P_b$ centers) cannot be made responsible for the observed hysteresis between 'sweep down' and 'sweep up'. This is demonstrated clearly by comparing the remaining $V_{TH}$ shifts recorded at the end of 'sweep up' at -1.7 V to the remaining CP current signal recorded immediately after 'sweep up'. The comparison is illustrated in Fig. 6.16. As a reference the $V_{TH}$ shift and the CP current measured directly after the termination of stress is included and labeled 'post stress' at the left hand side of the figure.

Following Fig. 6.16, the measured CP signal is widely independent of the individual measurement sequence and also quite the same right after the termination of stress. This is because the CP signal is mainly sensitive to interface states which are considered to be quasi-permanent. On the other hand, the remaining $V_{TH}$ shift measured at -1.7 V at the end of 'sweep up' depend strongly on the

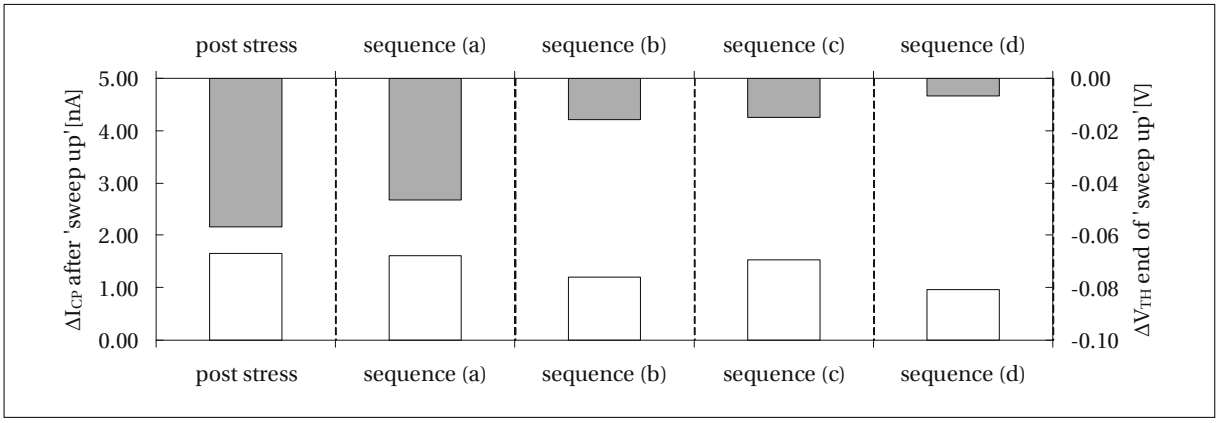


**Figure 6.16:** Comparison of the remaining $V_{TH}$ shifts (gray bars) to the remaining CP current signals (white bars) for the individual sequences at the end of 'sweep up'. As a reference, the $\Delta V_{TH}$ and the $\Delta I_{CP}$ signals recorded immediately after the termination of stress are illustrated to the left and labeled as 'post stress'.

measurement sequence. When measuring the $V_{TH}$ shift right after the termination of stress (first point of 'down sweep' labeled 'post stress' in Fig. 6.16), it is found to be largest. This is because the total $V_{TH}$ shift is assumed to consist of two components, the first being attributed to recoverable oxide defects (E′ centers), the latter being attributed to quasi-permanent locked-in oxide charges and interface states ($P_b$ centers). While the recoverable contribution may be annealed with time, temperature and bias switches, the quasi-permanent contributions remain particularly the same within the scope of our measurement sequences.

Indication for additional locked-in positive oxide charges is given by the small remaining degradation level observed at -0.7V in 'sweep up'. Due to the fact that $P_b$ centers are amphoteric [89, 90] and therefore commonly assumed to be neutral, when the Fermi level is close to midgap, the remaining $V_{TH}$ shift at -0.7V may be due to a different type of oxide defect equipped with larger time constants than conventional E′ centers, cf. Section 4.3.

### 6.3.3 Energetic profiling of recoverable oxide defects

The results and conclusions drawn by the experiments described above identified the Fermi level position and the temperature as the main parameters influencing the microscopic model transitions discussed in Fig. 6.14. In order to correlate the Fermi level position to the amount of $V_{TH}$ recovery, an additional extended experimental setup was developed: This 'incremental sweep' procedure linked to the Fermi level positions at individual gate voltages is schematically depicted in Fig. 6.17.

Rather than starting with a full 'down sweep' immediately after stress as was done in Fig. 6.15, a 100 s lasting constant gate bias phase at -1.7V was included prior to the actual sweeping procedure. This was done in order to differentiate between 'time dependent' and 'bias dependent' recovery effects. Within this initial 100 s, the $\Delta V_{TH}$ shift was determined as a function of time ($\Delta V_{TH}^{pstr}$). After

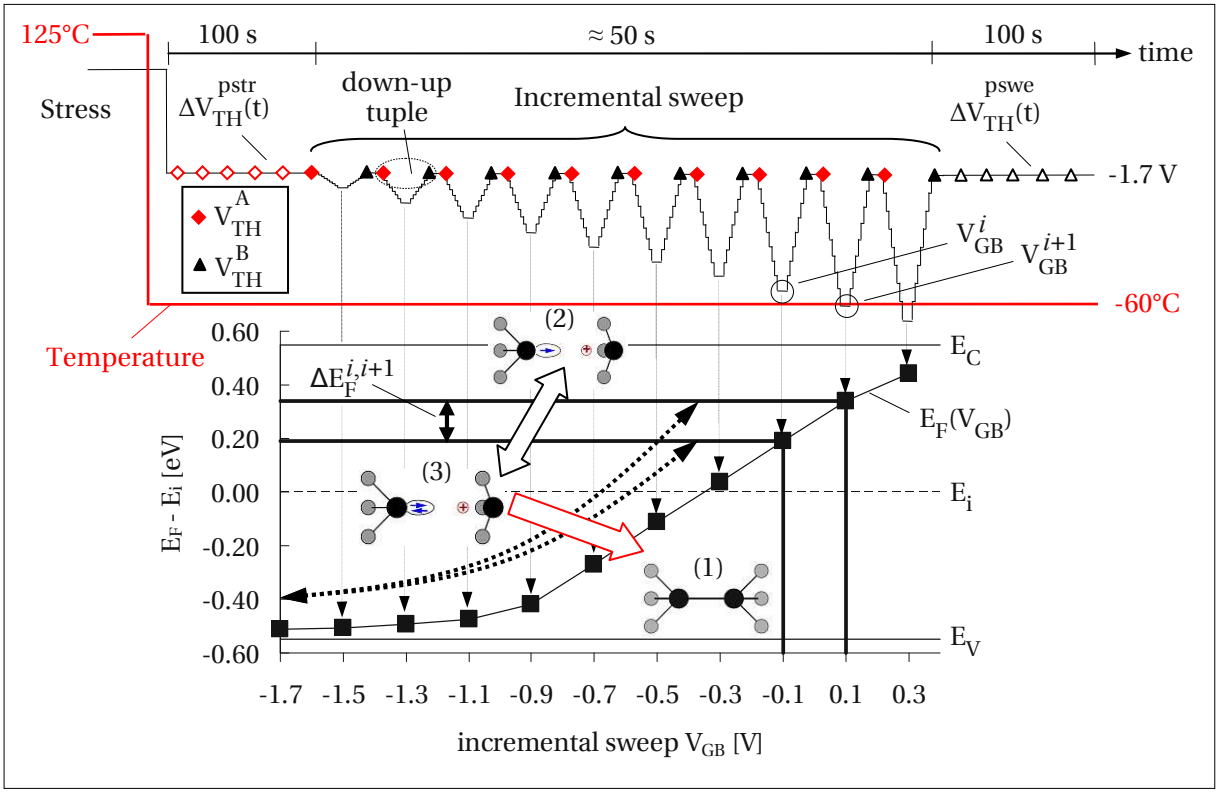


**Figure 6.17:** A schematic illustration of the 'incremental sweep' technique. The $V_{TH}$ shift is measured for 100 s at a constant gate bias of -1.7 V directly post stress ($\Delta V_{TH}^{pstr}$) and directly after the incremental sweep procedure ($\Delta V_{TH}^{pswe}$). In between, $V_{TH}$ shift tuples are measured right before (diamonds) and right after (triangles) each double-staircase ramp ($V_{GH} \longrightarrow V_{GB}^{i} \longrightarrow V_{GH}$). The Fermi level positions ($E_F^i$) corresponding to the base potentials ($V_{GB}^i$) of the double-ramps are illustrated below the experimental setup. As an example, the predicted model transitions and the additionally accessible energy range ($\Delta E_F^{i,i+1}$) is depicted when going from an arbitrary voltage level $V_{GB}^i$ to the next level $V_{GB}^{i+1}$.

the constant bias phase, the incremental down sweep was initiated beginning at $V_{GH}$ = -1.7 V till $V_{GB}^1$ = -1.5 V followed by a constant bias wait phase of 1 s at $V_{GB}^1$. Subsequently, the gate bias was ramped back toward $V_{GH}$ = -1.7 V which was maintained for another second. In the following, the $V_{GH} \longrightarrow V_{GB}^{i} \longrightarrow V_{GH}$ double-staircase sweeps are repeated keeping $V_{GH}$ constantly at -1.7 V while increasing $V_{GB}^i$ incrementally by +0.2 V till a final value of +0.3 V. Right before and immediately after such a double-staircase sweep, tuples of $\Delta V_{TH}$ values are recorded and referenced to the Fermi level positions corresponding to $V_{GB}^i$, cf. Fig. 6.17. After the last down-up cycle, the $\Delta V_{TH}$ recovery is monitored again for 100 s at a constant gate bias of -1.7 V ($\Delta V_{TH}^{pswe}$).

As the $V_{GB}^i$ level incrementally moves from inversion (-1.7 V) toward accumulation (+0.3 V), the Fermi level $E_F\left(V_{GB}^i\right)$ increasingly approaches the conduction band edge, cf. Fig. 6.17. In fact, when going from an arbitrary voltage level $V_{GB}^i$ to the next level $V_{GB}^{i+1}$, the accessible energy range for the (2) → (3) transition within the silicon bandgap extends by

$$\Delta E_F^{i,i+1} = E_F\left(V_{GB}^{i+1}\right) - E_F\left(V_{GB}^{i}\right). \tag{6.3}$$

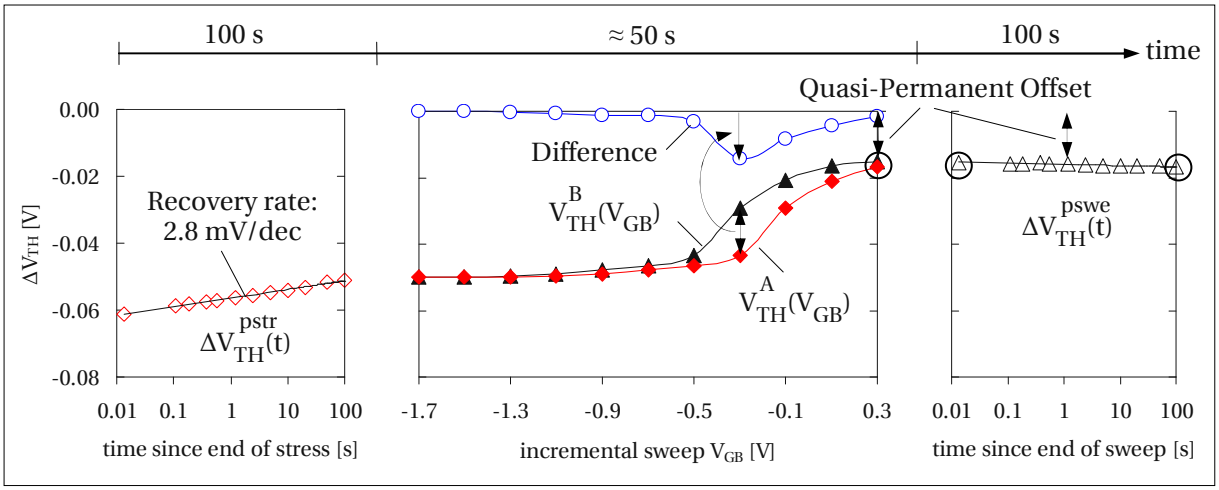


**Figure 6.18:** The $V_{TH}$ shifts measured during the incremental sweep experiment. The time dependent recovery of the threshold voltage shift, recorded for 100 s at $V_{GH}$ = -1.7 V directly post stress is illustrated to the left ($\Delta V_{TH}^{pstr}$). In the middle the tuples $\Delta V_{TH}^{A}$ and $\Delta V_{TH}^{B}$ corresponding to the particular base levels $V_{GB}^{i}$ are illustrated. The difference $\Delta V_{TH}^{A}(V_{GB}^{i}) - \Delta V_{TH}^{B}(V_{GB}^{i})$ is representative for the number of (3) → (1) transitions happening during the $V_{GB}^{i}$ phase. From that difference, a density of state profile may be evaluated by referencing to energy gap $\Delta E_{F}^{i,i+1}$ and to the position of the Fermi level within the silicon bandgap ($E_{F}^{i}$).

Traps within this energy range may be neutralized during the following down sweep ($V_{GH} \longrightarrow V_{GB}^{i+1}$). Once neutralized, some of the additional state (3) traps within $\Delta E_{F}^{i,i+1}$ may relax permanently during the 1 s wait phase at $V_{GB}^{i+1}$ (transition (3) → (1)). The ones who made the (3) → (1) transition are finally missing in the subsequent up sweep ($V_{GB}^{i+1} \longrightarrow V_{GH}$) and therefore account for a difference between $\Delta V_{TH}^{A}(E_{F}^{i+1})$ and $\Delta V_{TH}^{B}(E_{F}^{i+1})$.

By referencing the differences $\Delta V_{TH}^{A}(E_{F}^{i+1})$ - $\Delta V_{TH}^{B}(E_{F}^{i+1})$ to the current Fermi level position $E_{F}^{i+1}$ at $V_{GB}^{i+1}$ and to the corresponding energy gap $\Delta E_{F}^{i,i+1}$, an average density of state (DOS) profile can be determined. The DOS profile may be calculated under the assumption that all traps are located close to the $SiO_2$/Si interface:

$$D_{OX}^{rec}\left(\frac{E_{F}^{i}+E_{F}^{i+1}}{2}\right) = \frac{C_{OX}}{q}\frac{\Delta V_{TH}^{A}(E_{F}^{i}) - \Delta V_{TH}^{B}(E_{F}^{i})}{\Delta E_{F}^{i,i+1}}. \tag{6.4}$$

In Eq. 6.4, $C_{OX}$ is the area related gate oxide capacitance, q is the elementary charge and $E_{F}^{i}$ is the Fermi level position at the base level $V_{GB}^{i}$.

The $\Delta V_{TH}$ shifts measured during the 'incremental sweep' procedure are illustrated in Fig. 6.18. The results of the 100 s time dependent threshold voltage recovery recorded right after stress ($\Delta V_{TH}^{pstr}$) are depicted at the left hand side of Fig. 6.18. Within this time interval, a logarithmic recovery characteristic is obtained showing a slope of +2.8 mV/dec. The incremental sweep routine, which takes approximately 50 s, follows right after the initial 100 s performed at a constant gate bias of -1.7 V. The results for $\Delta V_{TH}^{A}$ and $\Delta V_{TH}^{B}$ are illustrated in the middle of Fig. 6.18 as a function of the double-sweep base

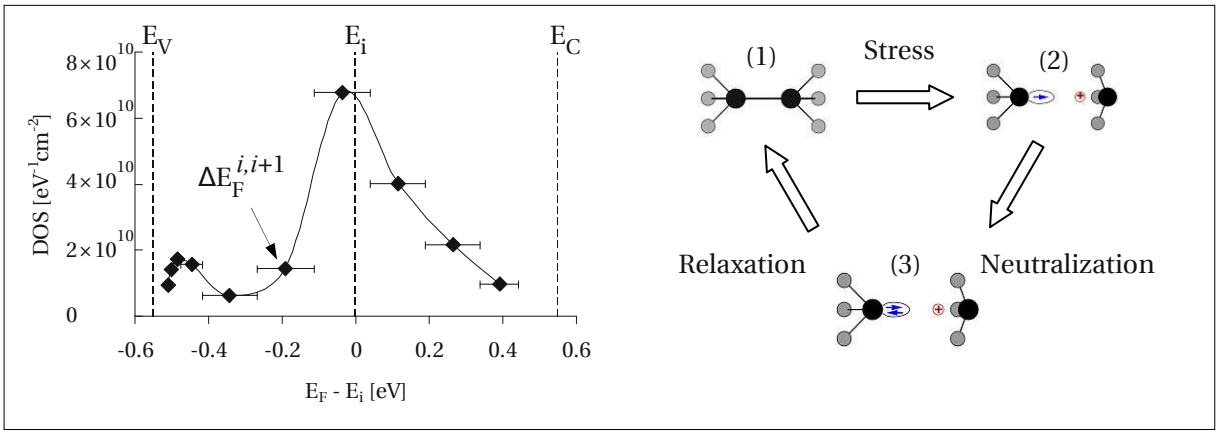


**Figure 6.19:** Extracted density of state profile of permanently recovered E′ centers which have undergone the transitions (2) → (3) → (1) during the incremental sweep procedure. The energy is referenced to the position of the Fermi level within the silicon bandgap ($E_F^i$).

potential $V_{GB}$. Any kind of time dependent recovery associated with the large voltage drop from stress bias (-17.0 V) to $V_{GH}$ (-1.7 V) is assumed to be negligible during the sweep phase due to the 100 s time delay following the actual stress phase. What remains is mostly Fermi level and temperature controlled carrier exchange. This is consistent with the observation that the degradation level is frozen at the end of the sweep routine, cf. $\Delta V_{TH}^{pswe}$ in Fig. 6.18 right hand side. Note that the final degradation level of -16 mV after $V_{GB}$ = +0.3 V in Fig. 6.18 is quasi-permanent at least within the time scale of our experiment and correlates perfectly with the 'up sweep' in sequence (b) and (c) in Fig. 6.15.

By making use of Eq. 6.4 and Fig. 6.18, a DOS profile may be calculated for the particular stress conditions. It is illustrated in Fig. 6.19 on the left hand side.

From a physical point of view, this DOS corresponds to E′ centers which are generated during stress ((1) → (2)), primarily neutralized during recovery ((2) → (3)) and finally recovered via structural relaxation ((3) → (1)). It has to be remarked that the stated energy levels $E_F^i$ in Eq. 6.4 and Fig. 6.19 need not necessarily correspond exactly to the actual energy levels of the defects located within the oxide, especially when assuming that the carrier exchange mechanism between the silicon substrate and the state (2) trap is inelastic. The DOS profile shows a significant peak around midgap. The occurrence of this peak can in general have two different reasons:

(i) There is a large DOS of type (2) traps located around midgap. After being neutralized by Fermi level variations ((2) → (3)), many (3) → (1) transitions may occur during the 1 s wait phase at $V_{GB}$ causing a significant reduction of positive charge at the end of the subsequent up sweep.

(ii) In general, oxide trap neutralization ((2) → (3)) and subsequent relaxation ((3) → (1)) of state (2) traps might become efficient not before the free electron concentration exceeds the free hole concentration at the oxide-substrate interface.

Based on our previous experiments, explanation (i) has to be favored: It has been shown in Fig. 6.15 that an increase in temperature during depletion has a similar impact on permanent oxide trap recovery as biasing the device in accumulation for the same amount of time, cf. sequence (b) and (c) in Fig. 6.15. Since elevating the temperature does not change the ratio of free interface carriers significantly but considerably enhances the number of (2) → (3) transitions by inelastic tunneling and simultaneously accelerates the thermally activated (3) → (1) transition, a dominance of electrons does not seem to be a mandatory requirement for relaxation.

### 6.3.4 Conclusions on the microscopic origin of NBTI induced defects

In this subsection switching oxide traps created during NBTS have been identified and examined. Recoverable oxide defects (most likely E′ centers) were found to be positively charged during stress. These defects can change their charge state during recovery. According to the HDL model, one must differentiate strictly between electrical neutralization and chemical bond reconstruction. The neutralization process is influenced considerably by the temperature and by the position of the Fermi level during read-out. In the neutral charge state (3) the switching oxide trap can either undergo a structural relaxation which anneals the trap permanently ((3) → (1)) or can re-emit a previously captured carrier and become positively charged again ((3) → (2)). Although electrical neutralization was identified to be a basic requirement for chemical relaxation, the intermediate neutral trap state requires thermal activation and time in order to lower in energy and relax irrevocably. By incrementally moving the Fermi level position into the silicon bandgap, oxide trap neutralization and subsequent permanent relaxation could be stimulated in a controlled manner. By referencing the observed threshold voltage shift reduction to the position of the Fermi level, an oxide trap DOS profile could be obtained which has a significant peak around midgap. The remaining degradation, after having annealed the majority of switching traps, seems to consist of a mixture of interface states and another type of chargeable oxide trap (locked-in oxide defects).

## 6.4 NBTI at low temperatures

This section represents an extension of Section 3.4 where we have analyzed NBTI dynamics of arbitrarily stressed devices at different temperatures. By using the in-situ polyheater technique, the so far strict constraint that the stress temperature has to equal the recovery temperature could be removed, revealing new insights in the temperature and the bias dependence of the power-law exponent, the recovery rates and on the ratio between interface and oxide trapped charge. In particular, the power-law exponent for interface state generation is found to be independent of the stress temperature and the stress bias, challenging previous investigations performed either at different characterization temperatures or being afflicted with large time delays for cooling from stress temperature to a unique characterization temperature [8, 78]. After erasing the recoverable contribution by gate pulsing toward accumulation, a universal factor of 2.5 between the interface state dependent $V_{TH}$ shift component and the quasi-permanent part of the total $V_{TH}$ shift is obtained for all analyzed stress

temperatures, fields and times. The appearance of this universality helps to extend the model of recoverable E′ centers to quasi-permanent components represented by interface states and locked-in oxide charges, cf. Section 4.3 [178].

### 6.4.1 Experimental constraints associated with conventional MSM techniques

From an experimental point of view, it is quite challenging to compare degradation and recovery dynamics of devices stressed at arbitrary fields and temperatures. When applying for example the standard MSM technique (cf. Section 3.1), the stress phase has to be interrupted repeatedly in order to record CP currents by gate pulsing or $V_{TH}$ shifts around the threshold voltage of the device. In particular, to monitor the total $V_{TH}$ shift, it is necessary to switch the gate bias from its stress level to a much lower recovery level around the threshold voltage of the device and record in parallel the linear or the saturation drain current. Alternatively, in order to measure the degraded maximum CP current right after stress, one must switch from a constant stress bias to gate pulsing between accumulation and inversion, thereby measuring the maximum charge pumping current at the substrate junction of the device. Since the switch from constant gate bias stress to alternating gate pulsing is experimentally harder to perform in a minimum of time, the CP data are typically afflicted with larger time delays compared to $\Delta V_{TH}$ shift data. During the unavoidable time delay involved with the change of the biasing conditions and the initialization of the subsequent recovery measurement, an unknown amount of recovery may already have occured when measuring the first point after the termination of stress. Consequently, the evaluation of the actual amount of degradation at the end of stress becomes distorted.

It has been shown in Section 6.2 that the temperature plays a significant role in defect recovery and relaxation. In a first-order approximation a temperature activated recovery process may be expressed by an Arrhenius law, cf. Eq. 6.1.

However, when aspiring to determine the true degradation level at the very beginning of the recovery phase, the exact value of $\tau_0$ has to be investigated. This is very difficult since $\tau_0$ is supposed to be very small and may be linked to the inverse phonon frequency which is around $10^{13}$ Hz. To approach $\tau_0$, sophisticated MSM methods have been developed in the past that accomplish $\Delta V_{TH}$ measurements only a couple of micro seconds after removal of the stress bias [58, 59, 170] thereby attaching considerable importance on the transition event between the stress and the relaxation phase [27, 52, 54]. However, following Eq. 6.1 and Section 6.2, a comparison of threshold voltage shifts and CP currents recorded at different temperatures may be misleading due to the temperature dependence of NBTI recovery. To put it in a nutshell, the limitation that the stress temperature has to equal the recovery temperature distorts a reliable case study on NBTI dynamics for different stress temperatures. Furthermore, particular theoretical difficulties with the temperature arise when attempting to correlate the $V_{TH}$ shift to the generation of interface states. This is mainly due to the fact that the charge pumping technique, traditionally used to characterize interface states [37], is quite inefficient at stress temperature since it covers just a very narrow part of the silicon bandgap when performed at high device temperatures, cf. Subsection 2.2.1. Consequently, when aspiring to estimate the threshold voltage shift caused by charged interface states ($\Delta V_{TH}^{it}$) it is necessary to assume a flat DOS in order

to account for the energetic mismatch, cf. Subsection 2.2.2. However, the flat density approach is definitely a very crude approximation [49, 86, 179] which is likely to introduce a considerable error when comparing data measured at different stress/recovery temperatures. Similar difficulties arise when comparing $V_{TH}$ shifts measured at different characterization temperatures. Due to the fact that the intrinsic carrier concentration within a semiconductor is exponentially temperature dependent, the Fermi level position at a particular read-out bias becomes a function of the temperature as well leading to different carrier concentrations at the interface for different characterization temperatures. Thus, in order to overcome those theoretical and experimental discrepancies, it is highly expedient to compare devices stressed at different stress temperatures directly post stress at always the same recovery temperature. Ideally, the recovery temperature should be much lower than the stress temperature in order to achieve a reasonable CP resolution and in order to decelerate thermally activated recovery processes. According to Eq. 6.1 the recovery time constants decrease exponentially with temperature which can be interpreted as a stretching on the time axis. The polyheater technique allows to generate different stress temperatures by applying different heater powers. On the other hand, the recovery temperature can be chosen independently and at a much lower level, i.e. -60°C. When applying degradation quenching, one might see a larger degradation level 10 ms post stress at -60°C than for example 1 $\mu$s post stress at the stress temperature. This feature is a particular strength of the polyheater measurement technique since it allows us to see more of the actual degradation level at the end of stress even though it might take a couple of milliseconds to measure the first current after removal of the stress bias.

### 6.4.2 Experimental setup for low temperature characterization

Fig. 6.20 (a) illustrates the basic MSM procedure applied to PMOS devices (SM6P/30/STD2). During stress, various electric fields (5.6/5.0/4.3 MV/cm) and temperatures (125/100/75°C) are applied to nine different devices. During recovery, the device temperature is *always* -60°C and the drain current is recorded at the threshold voltage (-1.1 V). During stress, the polyheater provides the individual stress temperature ($T_S$) and a certain stress bias ($V_{GS}$) is applied to the gate junction. The recovery cycle at -60°C is split in three sections. During section (I), the threshold voltage recovery ($\Delta V_{TH}^{tot}$) is monitored immediately after terminating the stress, the recovery time ($t_R$) equaling the previous stress time ($t_S$). During section (II), a 0.5 s lasting CP measurement is appended to the previous recovery cycle performed at $V_{TH}$ ($t_{CP}$). Since CP implies gate bias switches between inversion and accumulation, thereby bringing the stress induced oxide defects (E′ centers) repeatedly to the neutral charge state which is prone to structural relaxation (cf. Subsection 6.3.3), the $V_{TH}$ shifts recorded after the CP cycle for 0.5 s during $t_{R'}$ in section (III) are expected to be influenced considerably. The same MSM procedure is performed on each device with increasing stress ($t_S$) and recovery ($t_R$) durations (1/10/100/1,000 s). The stress temperature ($T_S$) and the stress field ($E_{OX}$) is varied.

Fig. 6.20 (b) illustrates a representative example of $\Delta V_{TH}$ shifts recorded during the sections (I) and (III) of the the MSM experiment. Right after the stress runs 'regular' log-like recovery traces are obtained for $\Delta V_{TH}^{tot}$ during $t_R$. As a consequence of CP, the measured degradation level is reduced considerably in section (III), the $\Delta V_{TH}$ shift being constant during $t_{R'}$ ($\Delta V_{TH}^{perm}$).

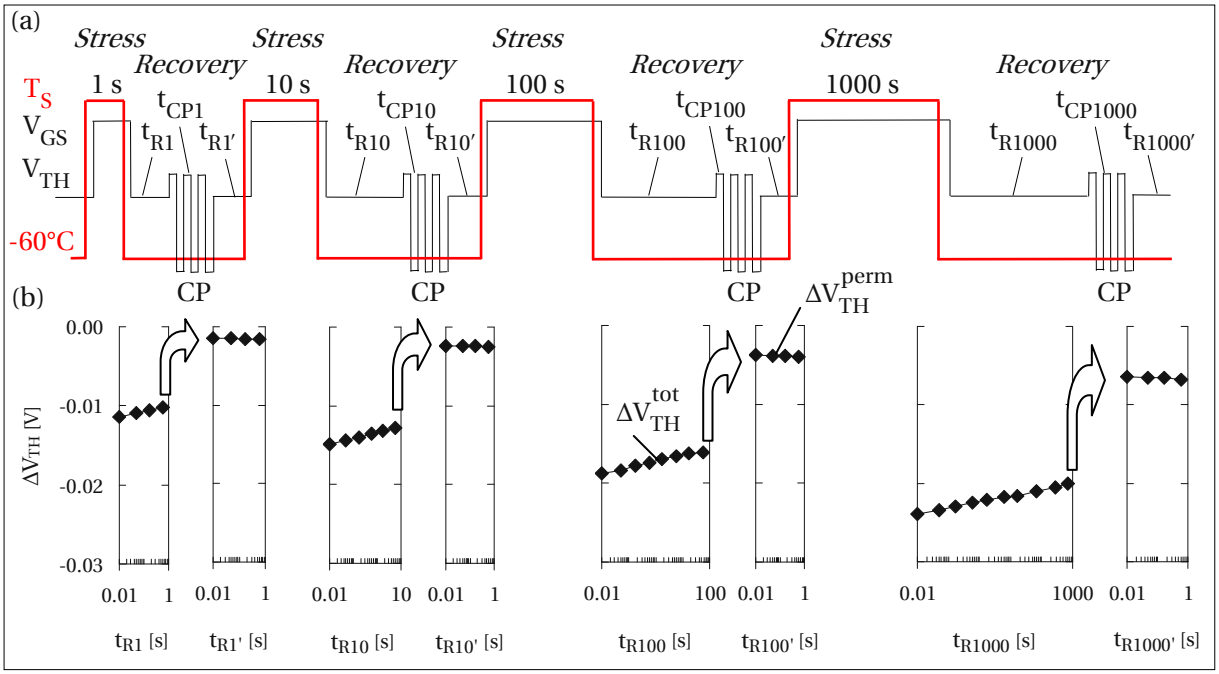


**Figure 6.20:** (a) The MSM procedure used to monitor the degradation and recovery dynamics of the CP current and the $V_{TH}$ shift after stressing the device at various gate bias ($V_{GS}$) and temperature ($T_S$) conditions. During stress, the polyheater provides the stress temperature $T_S$. During recovery, the heater is switched off and the device approaches the constant recovery temperature of -60°C. Different devices are used for each stress temperature and stress bias. The MSM procedure is performed with increasing stress ($t_S$) and recovery ($t_R$) durations (1/10/100/1,000 s). The charge pumping cycles $t_{CP}$ and the recovery phase $t_{R'}$ last for 0.5 s independently of $t_S$ and $t_R$. (b) illustrates a representative example of the $\Delta V_{TH}$ shifts recorded after stress ($\Delta V_{TH}^{tot}$ during $t_R$) and after CP ($\Delta V_{TH}^{perm}$ during $t_{R'}$). As a consequence of gate pulsing between inversion and accumulation the $\Delta V_{TH}$ shift measured after CP is considerably reduced and quasi-permanent.

### 6.4.3 Discussion on the stress dependent recovery rate

In Section 3.4 it was found that after 6,000 s seconds of stress the recovery rate is nearly independent of the stress temperature but depends considerably on the stress field. This former study was, however, limited to the constraint that the stress temperature has to equal the recovery temperature. By means of the new experimental setup and the availability of the polyheater technique, the study is now extended to different stress times where the recovery rate and the degradation level at the end of stress is monitored always at the same analyzing temperature of -60°C.

The results of the $V_{TH}$ shifts recorded during $t_R$ after different stress times are illustrated in Fig. 6.21. For illustration purposes, different stress biases and stress temperatures were grouped in nine separate graphs on the left hand side of Fig. 6.21. There are four separate measurement curves in every graph corresponding to four subsequent stress runs on every device with 1/10/100/1,000 s stress and recovery durations, respectively. Obviously, the total $V_{TH}$ shift increases with stress time, stress bias and stress temperature. Note that as opposed to Section 3.4 the results illustrated in Fig. 6.21 have

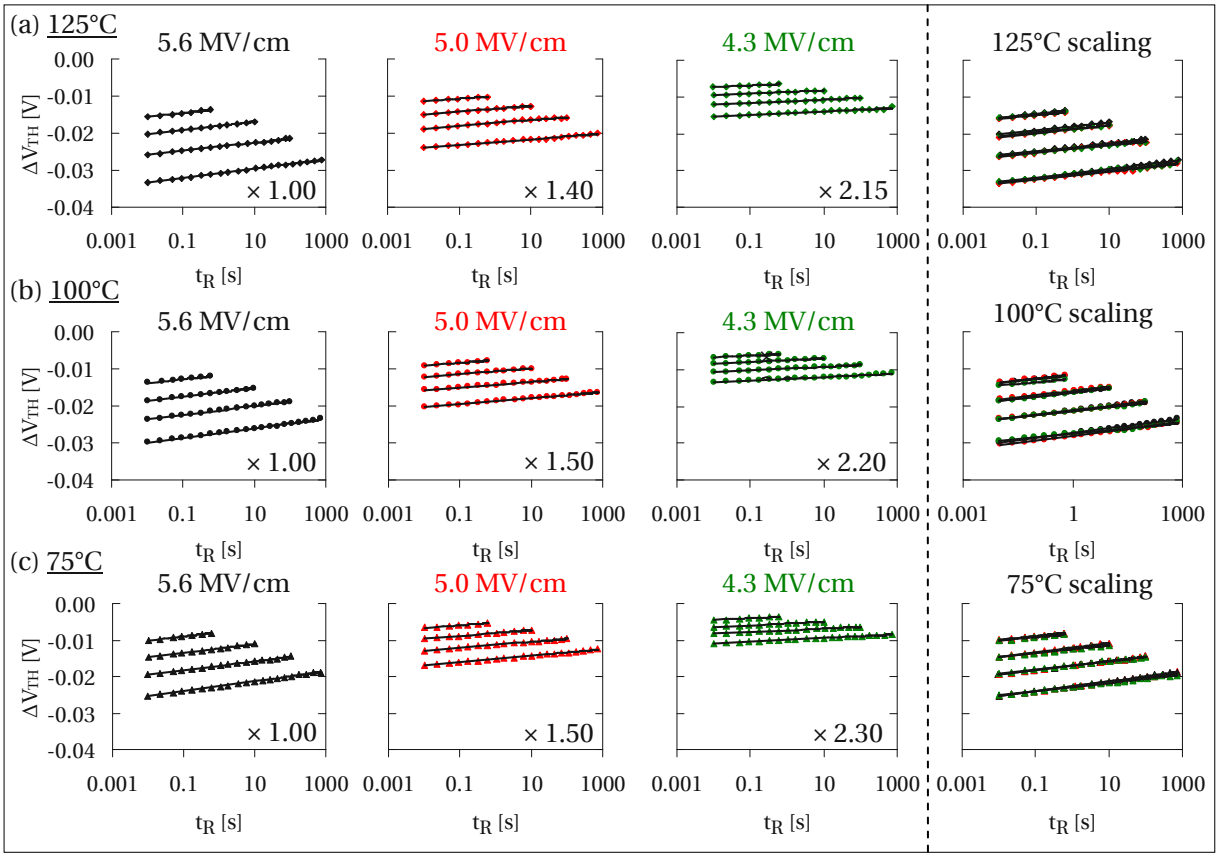


**Figure 6.21:** The $V_{TH}$ recovery curves measured during $t_R$ after different stress times. Nine separate devices were stressed for 1/10/100/1,000 s at three different stress temperatures (125/100/75°C) and three different stress fields (5.6/5.0/4.3 MV/cm). The various stress fields are depicted from left to right, while different stress temperatures are illustrated in (a) (125°C), (b) (100°C) and (c) (75°C). The unscaled $V_{TH}$ recovery data are depicted on the left hand side. On the right hand side different stress fields recorded at the same stress temperature were scaled by multiplying the data by appropriate scaling factors. Perfect scalability can be obtained for every stress temperature. Note that the field scaling factors of different stress temperatures are very similar. The $V_{TH}$ shifts recorded at the same stress field but at different stress times or temperatures differ rather by an additive than by a multiplicative factor.

been recorded at exactly the same recovery conditions directly post stress ($T_R$ = -60°C; $V_{GR}$ = -1.1 V) although the devices have been stressed at arbitrary temperatures and electric fields.

All recovery traces show a perfect linear decrease on the semi-logarithmic time plot. Basically, two features define the shape of the recovery plot: (i) the slope ($B_R$ in mV/decade) and (ii) the offset at an arbitrary time ($A_S$ in mV). Considering that every recoverable trap has a particular time constant that equals a point in time at which the probability of relaxation is largest, it is reasonable to suggest that a steeper recovery slope indicates a larger number of recoverable traps having similar time constants, whereas a larger offset with respect to the y-axis ($\Delta V_{TH}$ axis) indicates an enhanced creation of quasi-permanent defects, or defects having at least larger recovery time constants than observed in this

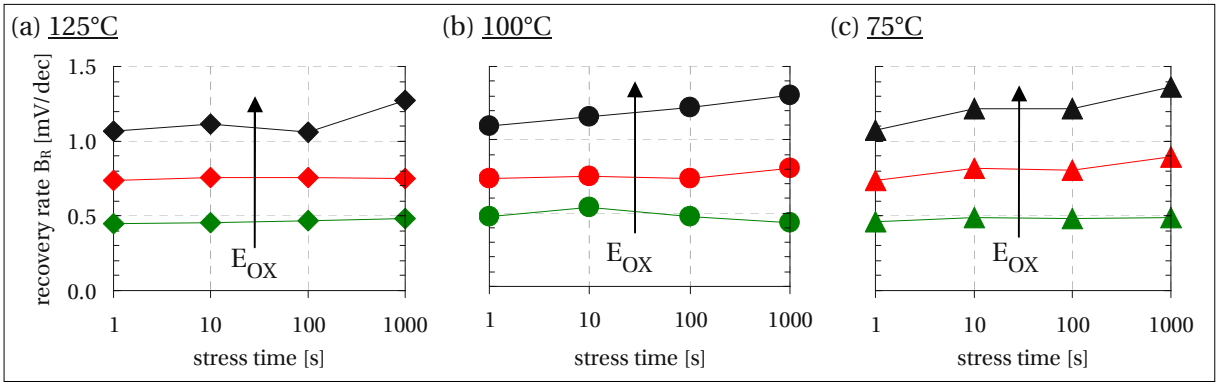


**Figure 6.22:** The extracted recovery rates $B_R$ (slopes of the recovery curves) per decade for different stress biases and stress temperatures. Since the recovery characteristics of all curves illustrated in Fig. 6.21 are approximately linear in a logarithmic time scale, an average recovery rate per decade has been evaluated for every tested device (stress condition). Different stress temperatures are illustrated in (a) (125°C), (b) (100°C) and (c) (75°C). The stress field has the most considerable influence on the recovery rate and therefore probably also on the total amount of recoverable traps created under NBTI. It seems that increasing stress time or stress temperature rather creates quasi-permanent defects or at least defects with much larger recovery time constants.

particular recovery experiment. In the following, all defect types which remain apparently constant within the time scale of our experiment will be denoted as 'quasi-permanent' defects.

Note that only the stress field influences the recovery slope considerably whereas stress time and temperature rather shift the recovery traces by an additive factor, indicating the activation of defects having larger recovery time constants. To highlight this observation in more detail, the calculated recovery slopes for all stress biases and stress temperatures have been depicted as a function of the stress time in Fig. 6.22.

According to the discussion above, one may conclude that increased stress fields activate a larger number of both recoverable (E′ centers) and quasi-permanent defects (interface traps and fixed positive oxide charge). This is reflected by the steeper recovery slope at higher stress fields, indicating a larger amount of defects having similar emission time constants, and the increased offsets, corresponding to more permanent or quasi-permanent damage. On the other hand, stress time and temperature seem to enhance predominantly the creation of defects with larger time constants while keeping the total number of recoverable defects which have time constants smaller than $t_R$ almost unaffected. In particular, one can find a set of empirical scaling factors that make the recovery traces overlap for different stress fields and the same stress temperature, cf. Fig. 6.21. Remarkably, the scaling factors are almost identical for all stress times and stress temperatures indicating a strong coupling of recoverable and quasi-permanent damage. It has to be remarked that one may find similar bias scaling factors when recording stress and recovery classically at the same temperature, as it was done in Section 3.4 and by others [77,144], however, the particular factors for different stress temperatures are only identical when comparing recovery traces recorded at one single recovery temperature. The scaled recovery curves for different stress fields are illustrated on the right hand side of Fig. 6.21.

Interface state re-passivation does not play a significant role in the observed $V_{TH}$ recovery. This was checked in [178] by comparing the maximum CP currents right after stress to the maximum CP currents recorded during $t_{CP}$ after constant bias recovery. Upon those measurements and due to all results discussed in the previous chapters, the interface state contribution is in the following argumentation considered as quasi-permanent as long as the gate bias is constant and as long as one abstains from long continuous gate pulsing periods, cf. Section 5.2.

### 6.4.4 Correlation between interface states and $V_{TH}$ shift

Having measured time, bias and temperature dependent $V_{TH}$ shifts and CP current degradation at identical recovery conditions, one main aim of this study is to check whether there is a correlation between the interface state generation and the total $V_{TH}$ shift. To accomplish this, it is necessary to convert the increase of the maximum CP current ($\Delta I_{CP}$) into an interface state dependent threshold voltage shift ($\Delta V_{TH}^{it}$). Similar attempts have been already made by others, however, previous studies were always bound to the constraint that the analyzing temperature equals the stress temperature, the implications of which being quite significant, as shown in the following.

The conversion is performed according to Subsection 2.2.2. The position of the Fermi level at a gate bias of $V_{GR}$ = -1.1 V and at a temperature of -60°C was simulated numerically [29] to be about 120 meV above the silicon valence band edge. Using our particular pulse setup ($f$ = 500 kHz; $t_r = t_f$ = 300 ns; $V_{GH}$ = -2.0 V; $V_{GB}$ = +2.0 V) and under the assumption of an energetically homogeneous capture cross section of $\sigma_p = 10^{-15}$ cm$^2$ [42, 46], the lower emission boundary of $\Delta E_{CP}$ at -60°C was calculated to be approximately 150 meV above the silicon valence band edge. We consider this to be almost equivalent to the 120 meV calculated before for $\Delta E_Q$ at -1.1 V. Note that at -60°C the energy interval $\Delta E_{CP}$ covers nearly the entire silicon bandgap due to the low characterization temperature. This would not be the case when recording the CP current at the much higher stress temperature of i.e. 125°C where $\Delta E_{CP}$ would be centered narrowly around midgap [50], cf. Fig. 2.8 in Subsection 2.2.1. Since a symmetrical pulse setup was used, it is necessary to consider that the CP signal covers approximately the same energy range in the upper and lower half of the silicon bandgap. Thus, in order to make the two energy intervals $\Delta E_{CP}$ and $\Delta E_Q$ coincide, one must divide the CP signal by a factor 2 and assume that the shape of the density of state (DOS) profile is symmetrical around midgap yielding a weight factor $\kappa$ of approximately 1/2, cf. Eq. 2.41. Except for this established assumption (symmetric DOS), no additional approximation on the shape of the density of state profile is necessary in order to accomplish the conversion from $\Delta I_{CP}^{max}$ to $\Delta V_{TH}^{it}$. The ability to accomplish the conversion without any additional assumption on the DOS is a particular benefit of our low temperature polyheater measurement technique.

Following Eq. 2.38 in Subsection 2.2.2, the increase of the maximum CP current ($\Delta I_{CP}^{max}$) may be converted into an interface state dependent threshold voltage shift ($\Delta V_{TH}^{it}$):

$$\Delta V_{TH}^{it} = \frac{\Delta I_{CP}^{max}}{2 A_G^{eff} f C_{OX}}, \tag{6.5}$$

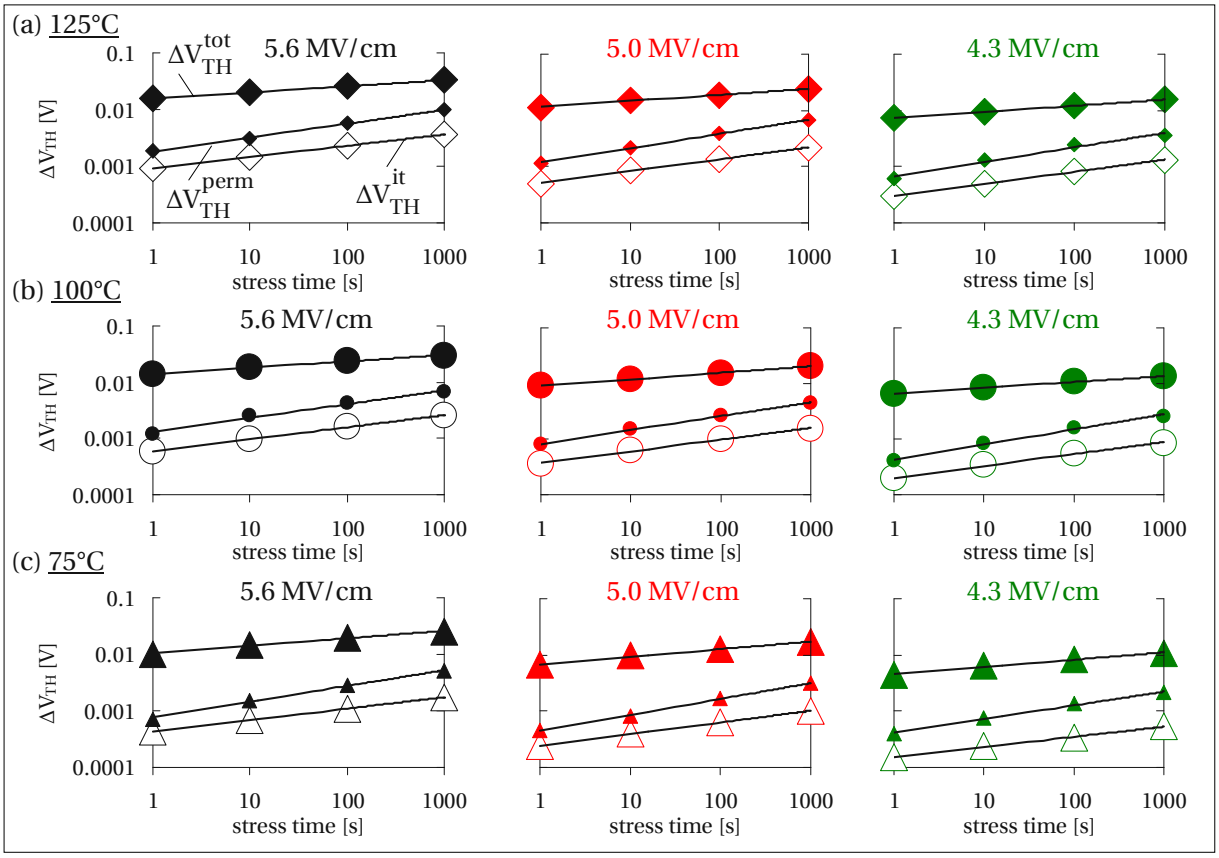


**Figure 6.23:** Time dependent degradation dynamics of $\Delta V_{TH}^{tot}$ (large full symbols), $\Delta V_{TH}^{perm}$ (small full symbols) and $\Delta V_{TH}^{it}$ (open symbols) recorded for three different stress temperatures ((a) 125°C; (b) 100°C; (c) 75°C) and three different stress fields (5.6/5.0/4.3 MV/cm). All data was measured 10 ms after the termination of stress at -60°C.

where the factor 2 considers that only donor-like interface traps (in the lower half of the silicon bandgap) contribute to a negative threshold voltage shift at $V_{GR}$ = -1.1 V. Once being able to isolate the interface state dependent $V_{TH}$ shift component from the total $\Delta V_{TH}$, one may attempt to find correlations between the threshold voltage shift measured from the drain current degradation and the number of generated $P_b$ centers measured from the increase in the CP current. Such correlations are investigated in this subsection.

Fig. 6.23 illustrates the individual evolution of $\Delta V_{TH}^{tot}$, $\Delta V_{TH}^{perm}$ and $\Delta V_{TH}^{it}$ for different stress temperatures and electric fields. As can be seen, all types of shifts show a power-law-like increase with the stress time (cf. Eq. 3.1), however, the steepness (power-law exponent $n$) and the offset (pre-factor $A_S$) of the individual shifts differ considerably. Note that $\Delta V_{TH}^{perm}$ is much lower than $\Delta V_{TH}^{tot}$, indicating recoverable oxide trap neutralization and annealing as a consequence of accumulation phases during CP consistent with the results obtained in Section 6.3.

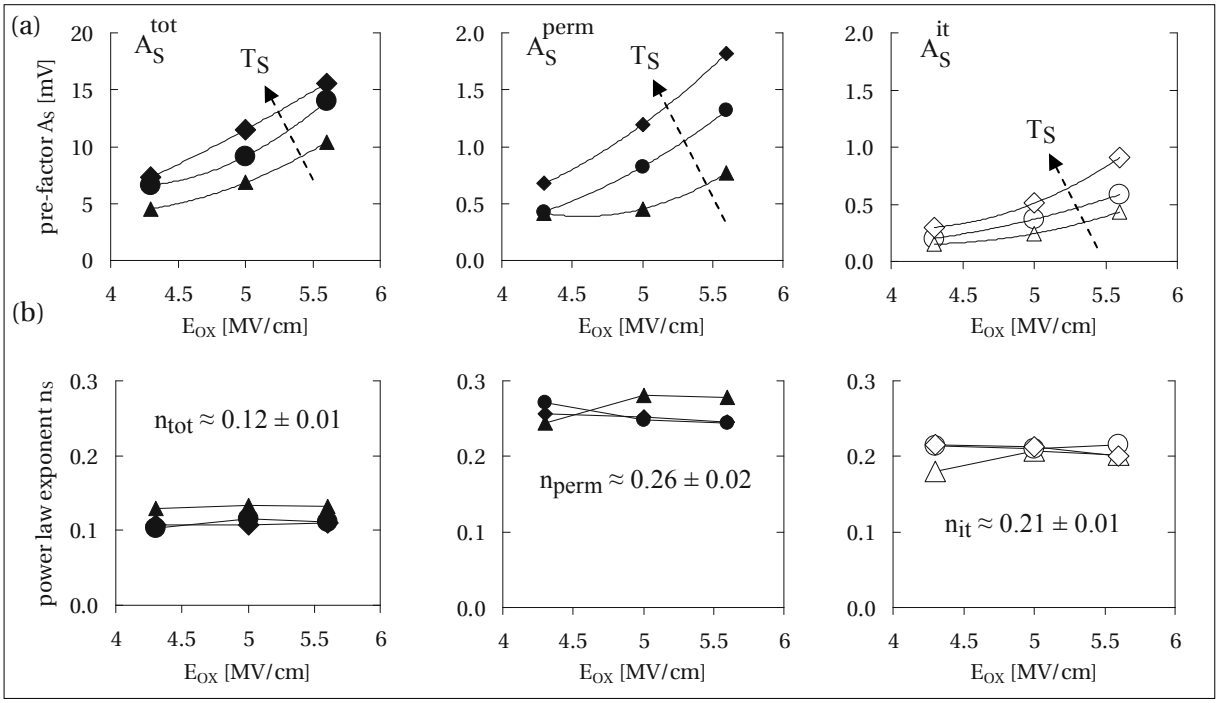


**Figure 6.24:** (a) Pre-factors $A_S$ and power-law exponents $n$ (b) as a function of the oxide field ($E_{OX}$) and the stress temperature ($T_S$). The data was extracted from Fig. 6.23. Different stress temperatures are illustrated as diamonds (125°C), circles (100°C) and (triangles 75°C).

Based on the model discussed in Section 6.3, it is reasonable to suggest that the total $V_{TH}$ shift ($\Delta V_{TH}^{tot}$) consists of quasi-permanent ($\Delta V_{TH}^{perm}$) and recoverable ($\Delta V_{TH}^{rec}$) components:

$$\Delta V_{TH}^{tot} = \Delta V_{TH}^{rec} + \Delta V_{TH}^{perm}. \tag{6.6}$$

After having annealed most recoverable traps during CP, the remaining quasi-permanent damage is assumed to be combination of interface state charge ($P_b$ centers) and locked-in positive oxide defects [144]:

$$\Delta V_{TH}^{perm} = \Delta V_{TH}^{it} + \Delta V_{TH}^{ox}. \tag{6.7}$$

Provided the generation of interface states and locked-in oxide defects is coupled somehow, one would expect a universal correlation between $\Delta V_{TH}^{it}$ and $\Delta V_{TH}^{perm}$ for all analyzed times, temperatures and stress fields. In particular, when assuming that $P_b$ centers and locked-in oxide traps are created simultaneously, i.e. as a consequence of hydrogen exchange between passivated $P_b$ centers (Si–H bonds) and E′ centers, as discussed in Section 4.3 and in [83, 144], $\Delta V_{TH}^{it}$ and $\Delta V_{TH}^{ox}$ should appear in a 50:50 relation. Consequently, $\Delta V_{TH}^{it}$ would represent about half of the $\Delta V_{TH}^{perm}$ component which then consists of both components in equal parts.

In Fig. 6.24, the individual pre-factors $A_S$ and the power-law exponents $n$ are illustrated as a function of stress field and stress temperature. In Fig. 6.24 (a), $A_S$ is found to increase with stress temperature and shows quadratic dependence on the oxide field, consistent with the results presented in Chap-

ter 3. $A_S^{tot}$ is about 16 times larger than $A_S^{it}$ and 8 times larger than $A_S^{perm}$, indicating a huge amount of recoverable damage (E′ centers) which becomes annealed during the intermediate CP cycle. The individual power-law exponents displayed in Fig. 6.24 (b) are found to be independent of the oxide field and the stress temperature. $n_S^{tot}$ is smaller than $n_S^{perm}$ and $n_S^{it}$ indicating that oxide trap creation has different degradation dynamics than quasi-permanent damage including interface states. Note that the pre-factors $A_S^{perm}$ and $A_S^{it}$ as well as the power-law exponents $n_S^{perm}$ and $n_S^{it}$ are very similar suggesting a tight coupling between the two components.

The tight coupling between $\Delta V_{TH}^{it}$ and $\Delta V_{TH}^{perm}$ is most explicitly demonstrated in Fig. 6.25. Fig. 6.25 is identical to Fig. 6.23 but the interface state component ($\Delta V_{TH}^{it}$) was multiplied by a universal factor of 2.5. As a consequence, perfect agreement between $\Delta V_{TH}^{perm}$ and $\Delta V_{TH}^{it}$ is obtained for all stress fields, temperatures and times. A similar factor making $\Delta V_{TH}^{it}$ overlapping with $\Delta V_{TH}^{tot}$ cannot be obtained due to the fact that $\Delta V_{TH}^{tot}$ consists of two independent components which provide different field and temperature acceleration. Note that the extracted factor of 2.5 is close to the physically predicted 50:50 relation between quasi-permanent $V_{TH}$ shift and interface state creation which would account for a factor of 2 according to Eq. 6.7. It has to be emphasized that a physical 50:50 ratio need not necessarily correspond to the relation of electrically active traps since the relation of created and charged defects also depends on the individual density of state profiles.

Alternatively to the coupling argument, it may be argued that $\Delta V_{TH}^{perm}$ is solely due to $\Delta V_{TH}^{it}$, for example when assuming donor-like defects which cover the whole silicon bandgap [140]. However, this assumption is in contradiction to the Fermi Level dependence of the $V_{TH}$ shift. In particular, it has been shown in Section 5.1 that interface states charge negatively in the NMOS device, where the Fermi level is pinned close to the conduction band edge during read-out, causing a net smaller or even positive threshold voltage shift after NBTS [10, 145].

### 6.4.5 Discussion on the power-law exponent

The power-law exponent is a crucial parameter for reliability life-time prediction since it is used to extrapolate the degradation dynamics (measured within a limited interval of time) to a long period of time corresponding to the product operation time. In this study is was found that the power-law exponent of the total $V_{TH}$ shift ($n_S^{tot}$) as well as the power-law exponents for interface state creation $n_S^{it}$ and quasi-permanent $\Delta V_{TH}$ degradation is independent of the stress field and the stress temperature (cf. Fig. 6.24 (b)) provided the DUTs are characterized always at the same analyzing temperature. In Section 3.4 the power-law exponents $n_S^{tot}$ were evaluated for the case where the stress temperature ($T_S$) equals the recovery temperature $T_R$. Even then $n_S^{tot}$ was found to be independent of the stress temperature within a range between -60°C and 200°C. In this previous study the temperature dependency of $n_S^{it}$ was not investigated due to the temperature sensitivity of the CP signal which is expected to introduce an error.

While the independence of the power-law exponent on the electric field has been observed already by others, the independence of the stress temperature is in contradiction to previous studies [8, 76] who suggested a linear T-dependence of $n_S^{it}$. This linear T-dependence is either derived from a dispersive

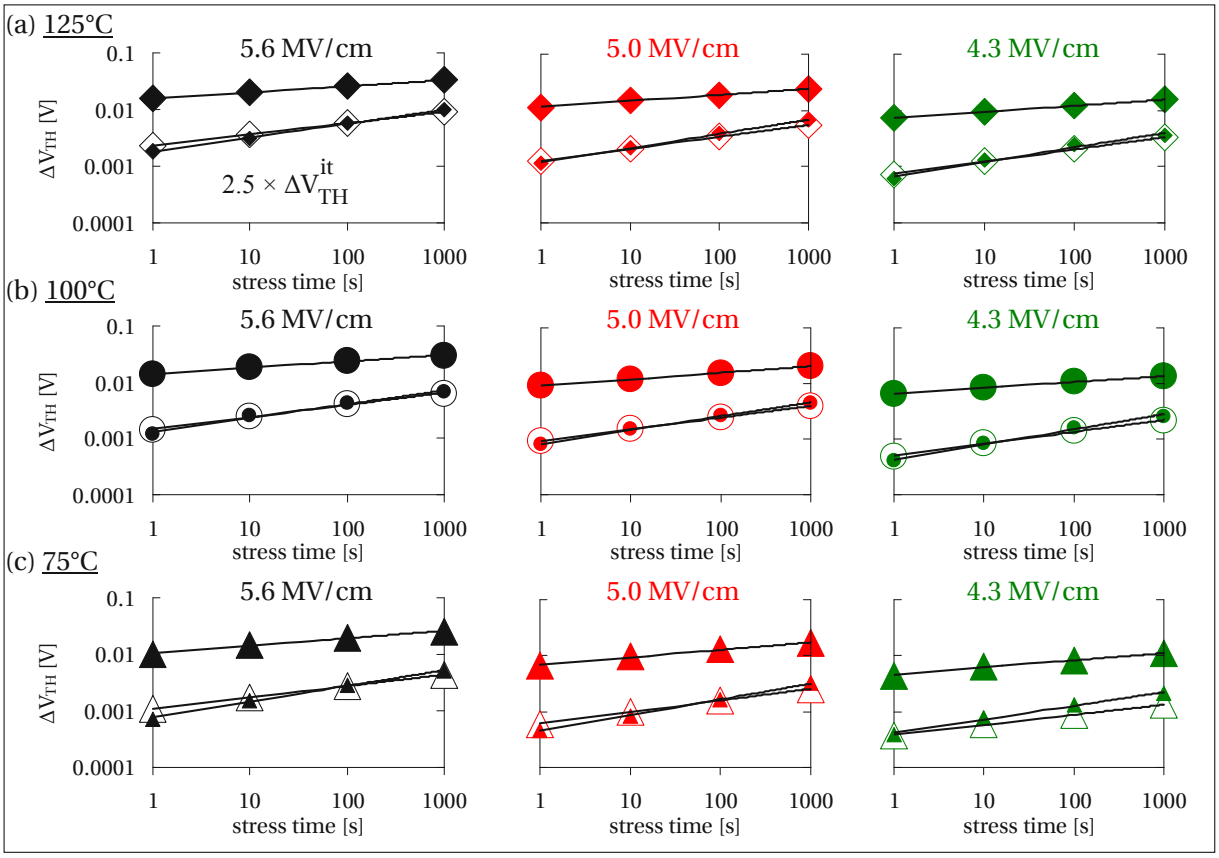


**Figure 6.25:** Similar illustration as Fig. 6.23, however, the interface state dependent $V_{TH}$ shift component ($\Delta V_{TH}^{it}$) was multiplied by a universal factor of 2.5 leading to a nearly perfect agreement of $\Delta V_{TH}^{it}$ (open symbols) and $\Delta V_{TH}^{perm}$ (small full symbols).

reaction controlled hydrogen release model [8, 78] (cf. Section 4.2) or from a dispersive hydrogen diffusion model [27, 76]. Both models suggest the power-law exponent $n_S^{it}$ to increase linearly with the stress temperature. A comparison between our results and the results of Huard *et al.* [8, 78] is given in Fig. 6.26.

For stress temperatures between 75°C and 125°C the values of $n_S^{it}$ in literature are reported to be in the range of 0.25 to 0.35 (cf. Fig. 6.26 (b)) which is much larger than the values 0.18 to 0.22 measured by our technique (cf. open diamonds in Fig. 6.26 (a)). We suggest that the discrepancies between our investigations and previous attempts arise from the fact that we analyze the CP current increase (recorded 10 ms post stress) at always the same recovery temperature (-60°C) while others compare either CP currents measured at different temperatures, thereby profiling different ranges of the silicon bandgap, or accept a long time delay at undefined biasing conditions between stress and measurement for cooling. In order to check this hypothesis, a similar experiment was performed using a stress field of 5.6 MV/cm, but this time the stress temperature equaling the recovery temperature (cf. full squares in Fig. 6.26 (a)). Remarkably, by the conventional approach, completely different power-law exponents for $n_S^{it}$ are obtained, which increase slightly with temperature in a similar but not that distinct way as reported in literature [8, 76]. The result suggests that the larger values of the power-law

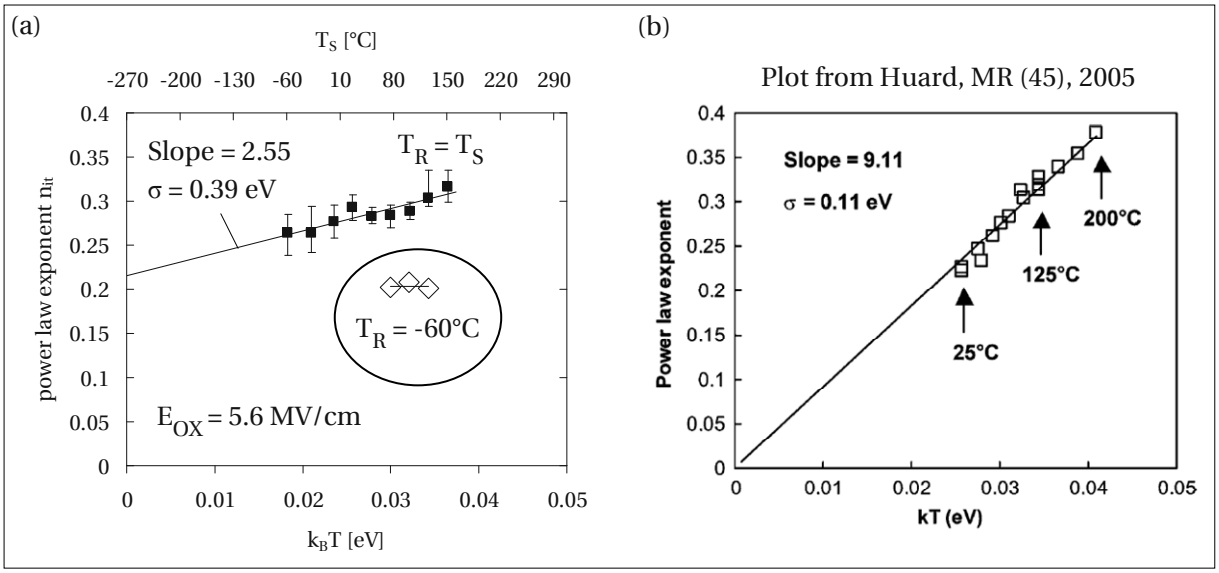


**Figure 6.26:** (a) Extracted power-law exponents for interface states creation ($n_S^{it}$) measured with stress temperature equals recovery temperature (full squares) and with the same stress temperatures but at a constant recovery temperature of -60°C (open diamonds) as a function of the stress temperature. When measuring $n_S^{it}$ conventionally (stress temperature equals recovery temperature) the extracted power-law exponents for interface states creation are much larger and increase slightly with the stress temperature (slope 2.55). On the other hand, when measuring $n_S^{it}$ at a constant recovery temperature of -60°C, the extracted power-law exponents are smaller and stable with temperature. (b) Temperature dependence of $n_S^{it}$ measured on p-channel MOSFETs with a 2 nm thick nitrided gate oxide (data from [8]). The power-law exponent increases significantly with stress temperature (slope 9.11).

exponents $n_S^{it}$ and their temperature development, as measured by conventional techniques, might be originated in the different energy ranges ($\Delta E_{CP}$) profiled and probably also in enhanced interface state recovery when recording the CP currents at an undefined time post stress at the particular stress temperatures.

### 6.4.6 Extension of the microscopic model explaining NBTI induced defects

In this section the $V_{TH}$ degradation and recovery of different PMOS devices (SM6P/30/STD2) stressed at different electric fields and temperatures were investigated. By making use of the polyheater technique, it became feasible to subject devices to different stress temperatures while characterizing them at a much lower characterization temperature of -60°C. This procedure allows to evaluate degradation and recovery phenomena of differently stressed devices under identical recovery conditions. Such investigations reveal that the power-law dynamics of interface state generation ($n_S^{it}$) and $V_{TH}$ shifts ($n_S^{tot}$) are considerably different when characterizing arbitrarily stressed devices at identical low temperature recovery conditions (i.e. $n_S^{it}$ turns out to be independent of the stress temperature in contradiction to previous studies where the stress temperature equaled the recovery temperature).

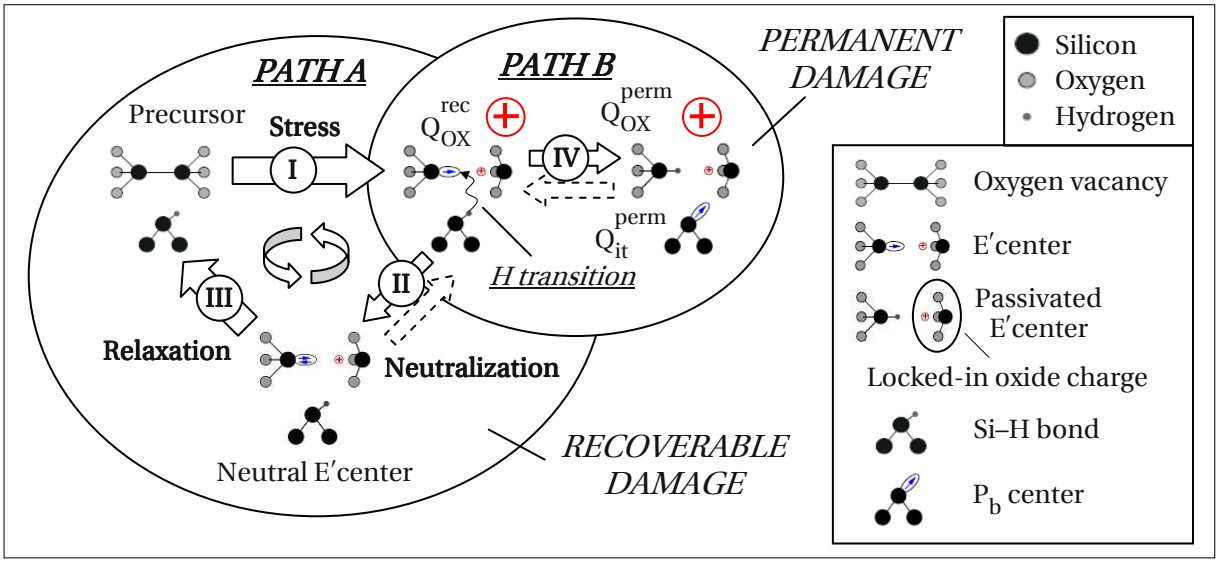


**Figure 6.27:** A schematic illustration of the microscopic transitions predicted by the Grasser model [144]. During stress, recoverable oxide traps (E′ centers) are created by temperature and field induced bond breakage of intrinsic oxygen vacancies. In the positive charge state the E′ center may become neutralized and anneals permanently (Path A). Alternatively it may get locked-in by hydrogen capture form the interface (Path B) thereby creating quasi-permanent damage.

A particularly adapted MSM setup made it possible to convert the increase of the maximum CP current ($\Delta I_{CP}^{max}$) recorded under identical recovery conditions directly (10 ms) after stress into an interface state dependent $V_{TH}$ shift ($\Delta V_{TH}^{it}$). By comparing the separate $V_{TH}$ shifts extracted from $\Delta I_D$ and $\Delta I_{CP}^{max}$ measurements, we conclude that a direct correlation between $\Delta V_{TH}^{tot}$ and $\Delta V_{TH}^{it}$ cannot be obtained as long as the total $V_{TH}$ shift contains recoverable components, i.e. recoverable oxide defects like E′ centers. Gate pulsing between inversion and accumulation (i.e. in the form of CP) turned out to anneal this recoverable component very efficiently. Finally, the quasi-permanent part of the $V_{TH}$ shift, determined from $\Delta I_D$ after CP, was found to be directly proportional to the interface state dependent threshold voltage shift. Universal correlation (factor 2.5) between $\Delta V_{TH}^{perm}$ and $\Delta V_{TH}^{it}$ has been demonstrated for various stress biases, temperatures and times. The results are strong evidence for a very tight coupling between interface state generation and the quasi-permanent part of $V_{TH}$ degradation measured under DC bias conditions. Based on these findings the microscopic model discussed in Subsection 6.3.1 may be extended to quasi-permanent defects, hydrogen and interface states. This was done by Grasser *et al.* in [144], where we have proposed the following complete model including recoverable and quasi-permanent damage, cf. Fig. 6.27.

Following Fig. 6.27: *Path A*: During NBTS, oxygen vacancies located close to the interface are assumed to break up and become positively charged (transition I) due to the presence of the high electric field and a majority of holes at the gate oxide substrate interface (E′ centers). During recovery, where the field and the carrier situation at the interface is quite different, some of these E′ centers ($Q_{OX}^{rec}$) may become neutralized by hole emission (transition II). Once in the neutral charge state, the E′ center can anneal permanently via structural relaxation, thereby restoring the initial precursor state again

(transition III). Structural relaxation is assumed to be highly temperature activated (relaxation barrier), while neutralization is supposed to be very fast and mainly Fermi level driven. In particular, when the temperature is very low (i.e. -60°C) some neutralized traps (which did not manage transition III) may become positively charged again during a subsequent sweep toward inversion (hashed arrow), cf. Fig. 6.15 (a). *Path B*: Once created during stress, the dangling bond of the E′ center can optionally attract a hydrogen atom from the interface which converts the recoverable oxide defect ($Q_{\mathrm{OX}}^{\mathrm{rec}}$) and the passivated interface state (Si–H bonds) into a locked-in positive oxide defect ($Q_{\mathrm{OX}}^{\mathrm{perm}}$) and an electrically active $P_b$ center ($Q_{\mathrm{it}}^{\mathrm{perm}}$) (transition IV). In principle, the reverse reaction of Path B, where the H atom is released from the dangling bond of the E′ center and travels back to the un-passivated interface state, is feasible as well (hashed arrow). However, in a first order approximation, this back transition is neglected assuming that the Si–H bond is stable within the E′ center. Consequently, once created, locked-in oxide defects and interface states are considered as quasi-permanent charge centers which cannot relax but may exchange carriers with the silicon substrate.

In principle, the model implies the following fundamental statements:

1. The total $V_{\mathrm{TH}}$ degradation consists of the following components: (1) Quasi-permanent interface states and locked-in oxide charges which emerge in a 50:50 relation due to entropy driven hydrogen exchange between Si–H bonds at the interface and positively charged E′ centers; (2) Recoverable positively charged E′ centers which may be annealed by time, temperature and Fermi level level switches toward accumulation.

2. The appearance of quasi-permanent damage depends highly on the number of Si–H precursors at the interface (passivation degree).

3. Recoverable damage is largely independent of hydrogen (at least, when neglecting the loss of recoverable E′ centers by transition IV).

4. Hydrogen incorporation into the gate oxide affects the creation of quasi-permanent damage considerably while the number of recoverable defects should remain more or less unaffected.

In this subsection the basic features of NBTI degradation and recovery dynamics have been investigated with a particular focus on the interplay of temperature and gate bias. New experimental procedures and setups have been developed which allow a deeper insight into the detailed attributes of defects, their response to environmental conditions and their dynamic transitions among each other. Based on these results a microscopic model has been developed which explains NBTI degradation and recovery dynamics as a two-stage process distinguishing clearly between recoverable and quasi-permanent damage. Our considerations suggest an intrinsic precursor within the $SiO_2$ gate oxide, namely an oxygen vacancy, to be the major candidate causing threshold voltage degradation in PMOS devices. Once created during stress, it acts on the one hand as a positive defect charge (E′ center) counterbalancing the applied gate potential and on the other hand a catalyst triggering hydrogen release from Si–H bonds at the interface thereby generating quasi-permanent damage in the form of interface states ($P_b$ centers) and locked-in oxide charges. Hydrogen comes into play concentrating on interface state generation during stress. In fact, the higher the passivation degree of the interface

(number of Si–H bonds), the larger the expected amount of quasi-permanent damage associated with electrical stress. As opposed to quasi-permanent degradation, the fundamental oxygen vacancy precursor is supposed to be independent of hydrogen and should hence show only little response to BEOL process steps which incorporate hydrogen into the gate oxide. In the next chapter, the developed experimental features are applied to differently processed wafers where the hydrogen budget of the gate oxide is tuned during BEOL fabrication. This was done in order to check the fundamental statements of our model with respect to the impact of hydrogen.

# 7

# The role of hydrogen in NBTI degradation

HAVING a precise degradation/recovery model in mind, this chapter aims to identify the link between hydrogen incorporation into the gate oxide and the formation of quasi-permanent and recoverable damage. Hydrogen incorporation is controlled by different modifications in the back-end process which improve the performance and defect densities of virgin silicon devices [180]. This can be done either directly by exposing the wafers to pure hydrogen or forming gas anneals [181, 182] or indirectly as a consequence of plasma-enhanced chemical vapor deposition (PECVD) of silicon nitride (SNIT) layers [183]. Such layers are considered to be efficient hydrogen sources since they contain a large concentration of hydrogen [184, 185] which may be released during or after deposition to diffuse toward the gate oxide. Provided there is no diffusion barrier below the SNIT, some hydrogen may reach the gate oxide where it can passivate dangling bonds at the interface, thereby improving the 'zero hour' performance of the MOS device. However, once passivated, previously captured hydrogen may be released from the interface during NBTS, leaving behind donor-like $P_b$ centers that are reported to cause a negative shift in the threshold voltage of a PMOS transistor. Thus, one may expect that the initial passivation degree of the interface, namely the total number of Si–H bonds present at the interface before stress, crucially determines the NBTI sensitivity of the technology. This is a generally accepted fact often reported in literature. However, concerning the underlying micro-structural physics behind degradation and recovery, different models come to different conclusions and hence make different predictions. In order to check fundamental statements of our model with respect to the impact of hydrogen, wafer splits were fabricated, where the hydrogen budget within the gate oxide (and hence the number of Si–H bonds at the interface) was modified. Hydrogen incorporation is measured analytically by time of flight secondary ion mass spectrometry (TOFSIMS) analysis (counting secondary H ions) and electrically by CP measurements (counting dangling bonds at the interface).

## 7.1 Basic signatures linked to BEOL hydrogen incorporation

The first wafer split which is going to be investigated consists of three selected wafers which differ in the back-end processing with respect to the layer and annealing sequences and the distance between the SNIT layer and the gate oxide. A schematic drawing of the individual layer stacks is given in Fig. 7.1. During the BEOL fabrication of wafer #1 (SM5P/30/H1) the SNIT layer was deposited directly on metal 1. Wafer #2 (SM5P/30/H2) was fabricated similarly as wafer #1 except for the fact that the sequence of power-metalization and SNIT layer was transposed (SNIT above power metal) and an inter-level dielectric was deposited between metal 1 and the power-metalization. Also, the annealing sequences of Wafer #1 and Wafer #2 were different. The fabrication of wafer #3 (SM5P/30/H3) was stopped after metal 1 deposition.

Considering that SNIT deposition introduces a lot of hydrogen into the device stack, Wafer #1 is supposed to have the best passivated interface because the distance between the hydrogen source (SNIT) and the gate oxide is smallest. During annealing, hydrogen is assumed to diffuse from the SNIT layer toward the gate oxide. Wafer #2 has its SNIT layer above the power-metalization and above an additional inter-level dielectric. Hence, the distance toward the gate oxide is much larger, and consequently, the interface of wafer #2 is supposed to be less well passivated than the interface of wafer #1. The processing of wafer #3 was aborted after metal 1 and no SNIT layer was deposited. Hence, wafer #3 is supposed to have the worst passivated interface after fabrication.

The speculation concerning the hydrogen incorporation into the gate oxide and the passivation degree of the interface was confirmed electrically by means of CP measurements (cf. Fig. 7.2 (a)) and analytically by means of TOFSIMS measurements (cf. Fig. 7.2 (b)).

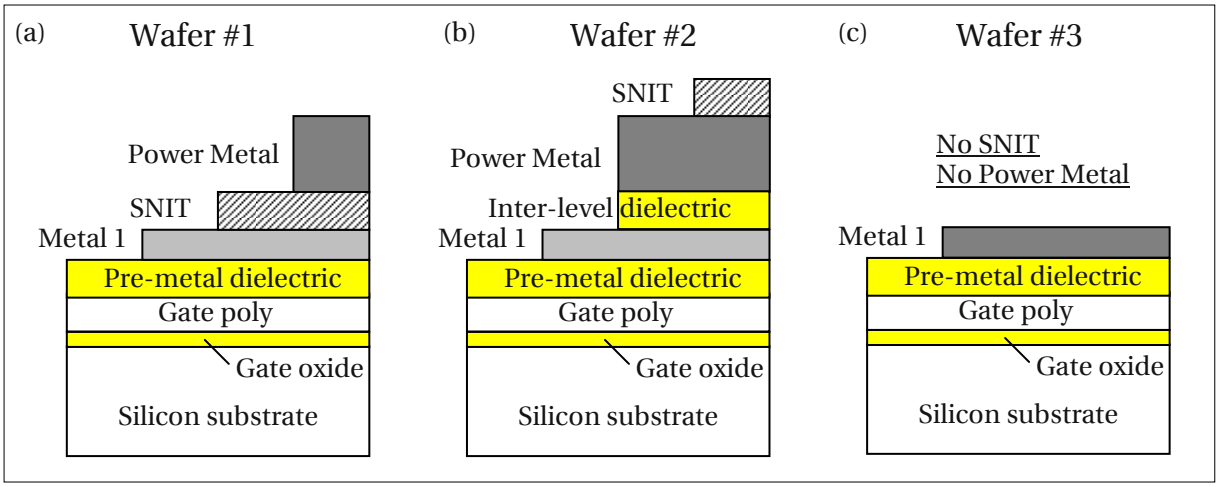


**Figure 7.1:** Schematic drawing of the BEOL wafer split. (a) Wafer #1 (SM5P/30/H1) has its hydrogen rich SNIT layer between metal 1 and the power-metalization. (b) Wafer #2 (SM5P/30/H2) has its hydrogen rich SNIT layer above the power-metalization and an additional inter-level dielectric. (c) The processing of wafer #3 (SM5P/30/H3) was aborted after deposition of metal 1.

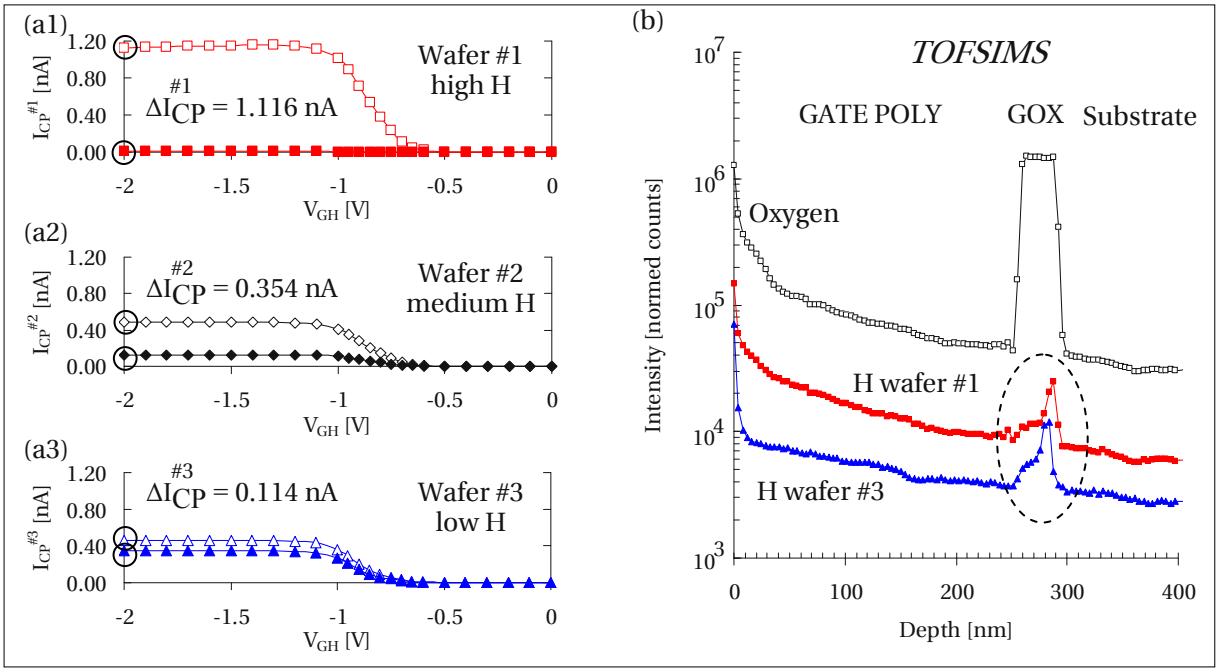


**Figure 7.2:** (a) CP currents of the selected split wafers measured before (full symbols) and after (open symbols) 1,000 s NBTI stress (125°C; 5.2 MV/cm). CP setup: $V_{GB}$ = 1.0 V; $f$ = 500 kHz; rising/falling slopes 2.7 V/$\mu$s. (a1) Wafer #1 has the lowest initial CP signal and shows maximum drift. (a2) Wafer #2 has a medium initial CP signal and a medium drift. (a3) Wafer #3 has the largest initial CP signal but minimum drift. (b) TOFSIMS image of the sub-metal BEOL stack. The oxygen signal indicates the gate oxide (GOX) and is illustrated by open symbols. The hydrogen signals of wafers #1 and #3 are illustrated by full squares, and triangles, respectively.

In agreement with the TOFSIMS results, wafer #1 provides the lowest initial CP signal, wafer #2 has an intermediate CP current and wafer #3 has the highest CP signal before stress. Note that the hydrogen signal in Fig. 7.2 (b) is peaked at the interface. According to the generally accepted assumption that $P_b$ centers at the interface become passivated by hydrogen, a well saturated gate oxide is supposed to degrade more rapidly under NBTI stress since it provides initially more Si–H precursors at the interface [180]. Following Fig. 7.2 (a), this hypothesis is confirmed by the variable increase of the CP current measured after subjecting PMOS devices of the particular split wafers for 1,000 s to NBTI stress (125°C; 5.2 MV/cm). Within the stated stress time wafer #1 degrades most heavily, wafer #2 shows a medium shift and wafer #3 degrades least. Another remarkable aspect of Fig. 7.2 (a) is the fact that wafer #1 provides the largest CP signal at the end of stress thereby exceeding wafer #2 and wafer #3. Actually, the CP current order is inverted after 1,000 s NBTI stress suggesting that the hydrogen incorporated during the back-end process does not only passivate dangling bonds at the interface but optionally also creates new Si–H precursors making the total amount of Si–H precursors plus $P_b$ centers larger for a well passivated gate oxide, i.e. wafer #1.

In order to further check statements of the Grasser model with respect to the impact of hydrogen, a certain recovery experiment was performed on all three process split wafers, cf. Fig. 7.3 (a).

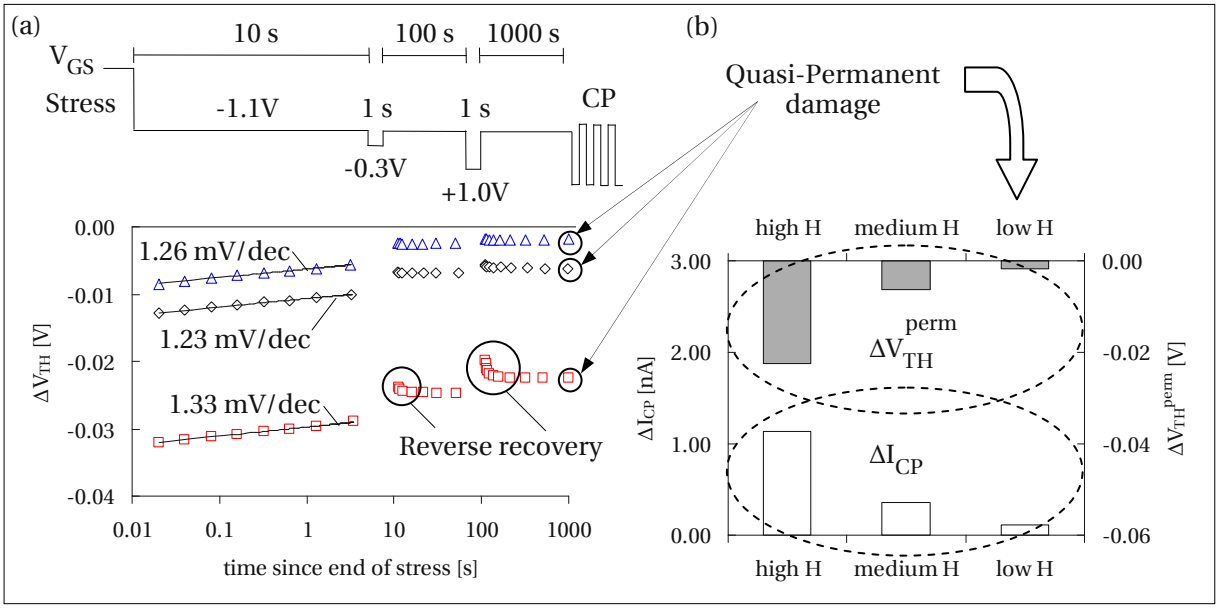


**Figure 7.3:** (a) Recovery experiment performed on each wafer: wafer #1 (squares); wafer #2 (diamonds); wafer #3 (triangles). After 1,000 s NBTI stress at 125°C and 5.2 MV/cm, the recovery was monitored for 10 s at -1.1 V, followed by a 1 s lasting bias switch toward -0.3 V. After the first bias switch, the recovery was monitored again for another 100 s at -1.1 V, followed by a second bias switch toward +1.1 V (1 s) and a third recovery cycle at -1.1 V (1,000 s). (b) A comparison of the remaining quasi-permanent threshold voltage shift at the end of recovery ($\Delta V_{TH}^{perm}$) and the CP current ($\Delta I_{CP}$) measured at the same time.

Right after stress, the $V_{TH}$ recovery was monitored for 10 s at -1.1 V in order to determine the recovery slope ($B_R$). Note that all wafers show a similar recovery rate of about 1.2–1.4 mV/dec, although the overall degradation level is significantly different. This is consistent with our model (cf. Fig. 6.27), which predicts the recoverable component (E′ centers) to be independent of hydrogen while the quasi-permanent contribution depends considerably on hydrogen. In order to extract the quasi-permanent component from the total $V_{TH}$ shift, two bias switches were performed, one toward depletion (-0.3 V) and the other toward accumulation (+1.0 V). At the end of the last recovery cycle the remaining $V_{TH}$ shift is assumed to be permanent within the time scale of our experiment. Obviously, the remaining quasi-permanent damage is largest for the well passivated wafer #1 and least for the metal 1 wafer #3. A subsequent CP measurement shows perfect agreement between the quasi-permanent $V_{TH}$ shift ($\Delta V_{TH}^{perm}$) and the increase in the CP signal ($\Delta I_{CP}$) consistent with the model and with previous observations. A remarkable detail of Fig. 7.3 (a) is the 'reverse recovery' observed in the $V_{TH}$ shifts after the gate bias switches. A certain amount of 'reverse recovery' is obtained for each device, however, it is most significant for wafer #1. This is consistent with our model when assuming 'reverse recovery' caused by charging and discharging of locked-in oxide defects. Considering that such locked-in oxide traps are created as a result of hydrogen exchange between passivated $P_b$ centers and electrically active E′ centers, wafer #1 is expected to show enhanced 'reverse recovery' since it is assumed to provide more locked-in oxide traps (with larger time constants) after NBTI stress.

## 7.2 Interaction of hydrogen with degradation and recovery dynamics

In the previous section it has been demonstrated that the recovery rate is independent of the hydrogen budget within the gate oxide. This is a strong indication that the oxygen vacancy is the responsible precursor for $Q_{OX}^{rec}$. The following subsections study the time evolution and the energetic distribution of recoverable and quasi-permanent damage. Experiments are performed on selected hydrogen split wafers.

### 7.2.1 Energetic distribution and dynamic evolution of recoverable oxide traps

In this subsection, the energetic distribution and dynamics of the recoverable component is investigated in detail for three selected split wafers with different power-metalization and hydrogen budget within the gate oxide [186]. The applied technique is the 'incremental sweep technique' demonstrated in Subsection 6.3.3 [174].

During BEOL processing, titanium layers of different thicknesses were incorporated below the metalization in order to control the hydrogen diffusion from the upper hydrogen rich SNIT layer toward the gate oxide during fabrication. Titanium is known to be an effective barrier against hydrogen diffusion [187]. Hence, it is assumed that wafers with thick titanium barriers have less hydrogen within the gate oxide than wafers with thin titanium barriers. In this study three selected split wafers which provide vastly (about an order of magnitude) different hydrogen concentrations within the gate oxide are analyzed. This is demonstrated in Fig. 7.4 by the considerably different virgin CP characteristics.

Wafer #1 (SM6P/30/H1) and wafer #2 (SM6P/30/H2) are supposed to have a well passivated interface (thin Ti liner ⟶ much hydrogen ⟶ low CP signal) while wafer #3 (SM6P/30/H3) has a weakly passivated interface (thick Ti liner ⟶ few hydrogen ⟶ high CP signal). In the following, different PMOS devices of the selected split wafers were stressed for 10/100/1,000/10,000 s at 125°C.

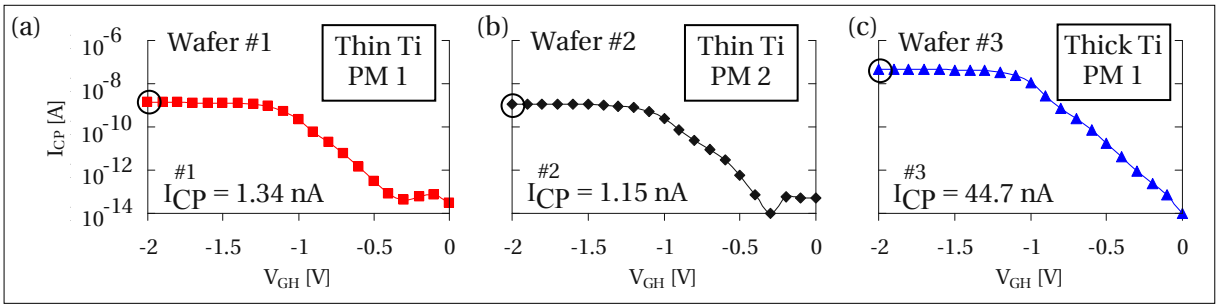


**Figure 7.4:** CP currents of the selected split wafers measured before stress at -60°C. CP setup: $V_{GB} = 2.0$ V; $f = 1$ MHz; rising/falling slopes 6.4 V/$\mu$s. (a) Wafer #1 (SM6P/30/H1) has a thin titanium barrier and a standard power-metalization (PM 1). (b) Wafer #2 (SM6P/30/H2) has a similar thin titanium barrier as wafer #1 but a different power-metalization (PM 2). (c) Wafer #3 (SM6P/30/H3) has a thick titanium barrier and a standard power-metalization (PM 1). Wafer #1 and Wafer #2 have initially a much lower CP signal than Wafer #3, suggesting a better passivated interface.

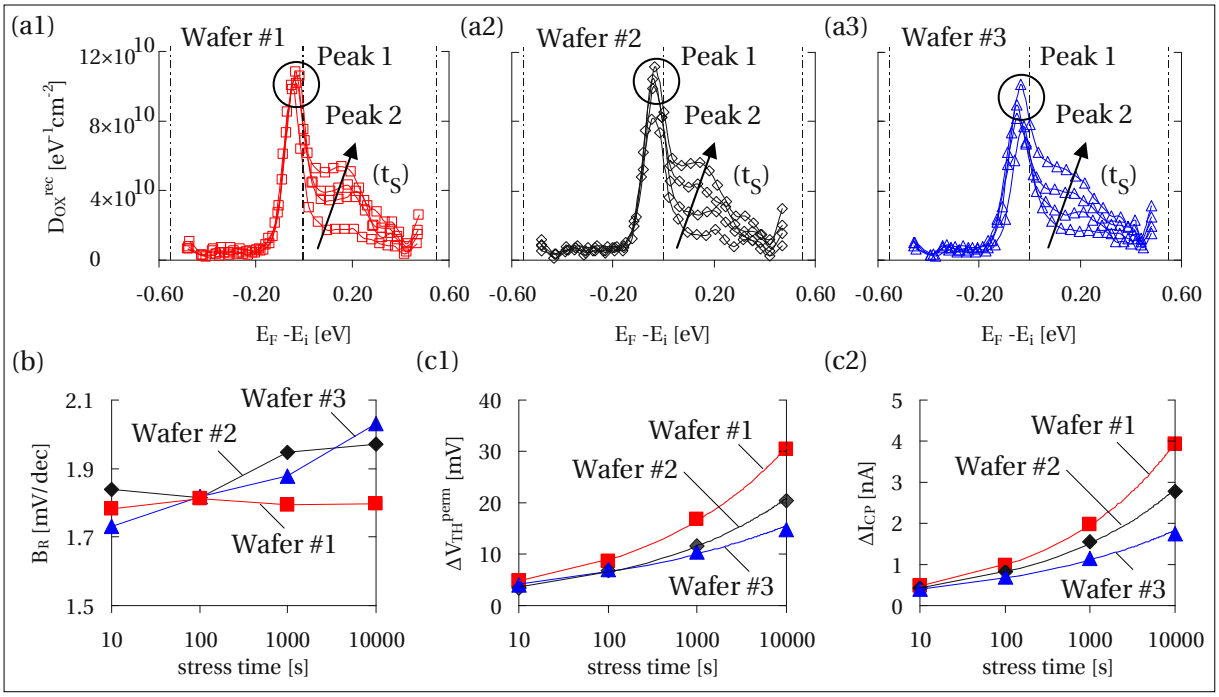


**Figure 7.5:** (a) Recoverable oxide trap profiles ($D_{OX}^{rec}$) of the selected split wafers recorded after different stress times. All samples show two peaks of similar amplitude and shape; the first is located close to midgap and almost fully developed after ten seconds of stress; the second is located in the upper half of the silicon bandgap and develops gradually as a function of stress time. (b) The recovery rates ($B_R$) of the selected split wafers recorded within the 100 s constant gate bias phase (-1.5 V) directly post stress. All samples show similar time dependent recovery rates of approximately 1.9±0.2 mV/dec. (c1) The quasi-permanent $V_{TH}$ shift ($\Delta V_{TH}^{perm}$) and (c2) the increase in CP current ($\Delta I_{CP}$) of all samples as a function of stress time. Wafer #1 shows a considerable higher quasi-permanent $V_{TH}$ shift than #2 and #3. The CP currents in (c2) correlate with the quasi-permanent $V_{TH}$ shifts in (c1) by a universal multiplicative factor. $\Delta V_{TH}^{perm}$ and $\Delta I_{CP}$ show a power-law-like increase with stress time (fit lines), wafer #1 having a larger power-law exponent than wafer #2 and wafer #3 ($n_{perm}^{\#1} = 0.27$; $n_{perm}^{\#2} = 0.25$; $n_{perm}^{\#3} = 0.19$).

After each stress run, degradation was quenched to -60°C and the devices were analyzed using the 'incremental sweep' technique, cf. Subsection 6.3.3. The outcome includes (i) the time dependent recovery rates within the first 100 s ($B_R$) after the termination of stress, (ii) the recoverable oxide trap profiles extracted from the incremental sweep ($D_{OX}^{rec}$), (iii) the quasi-permanent $V_{TH}$ shifts ($\Delta V_{TH}^{perm}$) and (iv) the CP currents measured at the end of each characterization run ($\Delta I_{CP}$). The results are illustrated for increasing stress times in Fig. 7.5.

The effective density of state profiles of all split wafers are illustrated in Fig. 7.5 (a). All samples show very similar recoverable DOS profiles indicating that the trap precursor is actually independent of hydrogen, consistent with the results gained from the alternative wafer split discussed in Section 7.1. The measured recovery rates during the first 100 s constant gate bias phase are illustrated in Fig. 7.5 (b). In agreement with the DOS profiles, all samples show similar recovery rates of 1.9 ± 0.2 mV/dec which slightly increase with stress time. On closer inspection of the energetic

profile ($D_{OX}^{rec}$) two characteristic peaks emerge; the first one is located near midgap and is almost fully developed after ten seconds of stress; the second one is located in the upper half of the silicon bandgap and develops gradually with further increasing stress times. The quasi-permanent $V_{TH}$ shifts ($\Delta V_{TH}^{perm}$; measured for 100 s directly after the sweep procedure at -1.5 V) and the increase in the CP current ($\Delta I_{CP}$; measured directly after $\Delta V_{TH}^{perm}$ by gate pulsing) are illustrated in Fig. 7.5 (c1) and (c2) as a function of the stress time. $V_{TH}^{perm}$ and $\Delta I_{CP}$ correlate by a multiplicative factor consistent with previous results. The quasi-permanent damage is considerably larger for wafer #1 (well passivated interface) than for wafer #3 (weakly passivated interface) indicating that both components are directly linked to each other and to hydrogen. An exception is found in wafer #2 (equipped with a different power-metalization PM 2) which has a similar 'zero hour' CP signal as wafer #1 (cf. Fig. 7.4) and hence a comparably passivated interface but shows considerably less quasi-permanent damage. Except for the metalization, wafer #1 was identically fabricated as wafer #2. The quasi-permanent damage follows a power-law-like increase for all tested structures, the value of the exponent, however, increases proportionally to the hydrogen budget within the gate oxide.

In this subsection the density of state profiles of recoverable oxide traps was analyzed by making use of the 'incremental sweep' technique. In particular, two peaks located energetically in the middle and in the upper half of the silicon bandgap were investigated. By comparing three PMOS devices taken from selected split wafers, recovery was found to be independent of the hydrogen budget and the metalization process while the increase in the CP current and the quasi-permanent $V_{TH}$ shift is strongly connected to the BEOL fabrication.

### 7.2.2 Interaction of hydrogen with recoverable and quasi-permanent damage

Two selected hydrogen split wafers (wafer #1 (SM6P/30/H1) and wafer #3 (SM6P/30/H3)) were subjected to a particularly designed experimental procedure in order to separate time and bias dependent recoverable damage from apparently quasi-permanent degradation, which is assumed to consist of interface states and locked-in oxide charges, cf. Fig. 6.27. Wafer #1 has a thin Ti barrier and hence a lot of hydrogen within the gate oxide while wafer #3 has a thick Ti barrier and hence a worse passivated interface. This is demonstrated by the considerably different initial CP currents and the different hydrogen signals detected with TOFSIMS, cf. Fig. 7.6 (a), and (b), respectively. Both wafers were fabricated with a standard power-metalization (PM 1).

The TOFSIMS image in Fig. 7.6 (b) shows the sub-metal BEOL layer stack of the two selected split wafers. Displayed are the oxygen and the titanium signals for the orientation within the BEOL stack. Underneath the Ti liners considerably different hydrogen concentrations are measured for wafer #1 and #3 in the post metal dielectric (PMD), the gate-poly and the gate oxide (GOX). In perfect agreement with the TOFSIMS results we obtain in Fig. 7.6 (a) that the initial CP signal of wafer #1 (0.3 nA) is about 30 times lower than the 'zero hour' CP signal of wafer #3 (9.0 nA). This is consistent with the assumption that the interface of wafer #1 ($D_{it} = 2.3\times10^{9}$ eV$^{-1}$ cm$^{-2}$) is more efficiently passivated with hydrogen than the interface of wafer #3 ($D_{it} = 6.9\times10^{9}$ eV$^{-1}$ cm$^{-2}$). The CP currents were recorded at a temperature of 50°C using a pulsing frequency of 500 kHz and rising/falling slopes of 10 V/$\mu$s, scanning roughly 500 meV of the silicon bandgap around midgap.

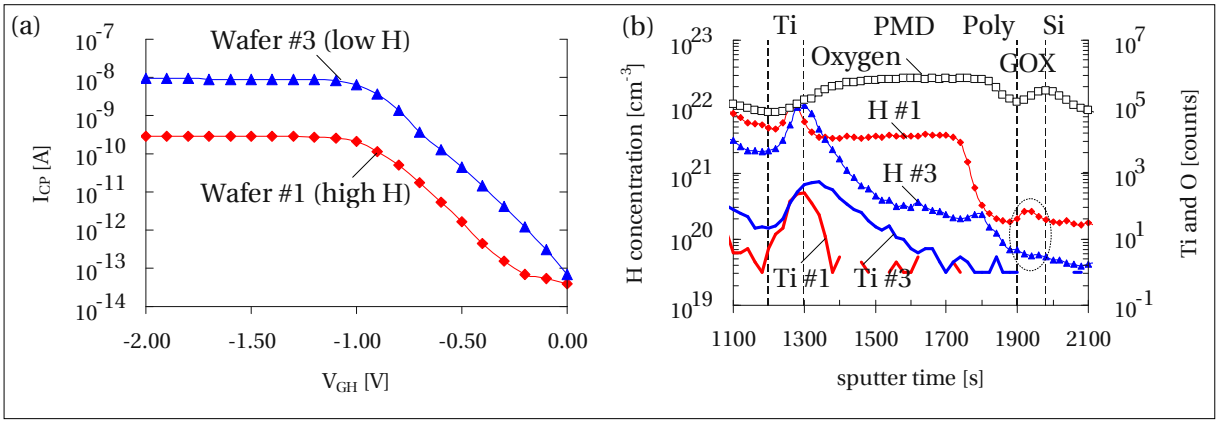


**Figure 7.6:** (a) CP currents of the BEOL process split wafers #1 (diamonds) and #3 (triangles). Due to the higher hydrogen concentration within the gate oxide, wafer #1 has a more efficiently passivated interface and a lower CP signal than wafer #3. (b) TOFSIMS images of wafers #1 and #3. By modifying the titanium layer thickness, the hydrogen budget within the post metal dielectric (PMD), the gate-poly and within the gate oxide (GOX) can be controlled. Wafer #1 has a thinner Ti layer than wafer #3 and hence a higher hydrogen concentration within the layers below, in particular, within the GOX.

The characterization procedure following NBTS (200°C; 7.0 MV/cm) for a defined stress time ($t_S$) is illustrated in Fig. 7.7 (a). By making use of the in-situ polyheater technique, the recovery phase can be performed at a much lower characterization temperature of 50°C which decelerates thermodynamical recovery mechanisms and improves the charge pumping measurement resolution (degradation quenching).

After stress, the characterization procedure is initiated by a 1,000 s lasting recovery phase at -2.0 V ($t_{R1}$). The relative amount of $V_{TH}$ recovery during $t_{R1}$ between the first measured point after removal of the stress bias (40 ms post stress) and the last measured point (1,000 s post stress) is denoted as the time dependent recovery contribution ($\Delta V_{TH}^{time}$). Subsequently to $t_{R1}$, the gate bias is ramped down in 20 mV steps from strong inversion (-2.0 V) toward depletion (0.0 V) ($S_{D1}$). In parallel, the $V_{TH}$ shift is monitored as a function of the gate bias. One full gate bias ramp takes approximately 10 s. Approaching depletion, the Fermi level moves from the valance band edge toward the conduction band edge, thereby gradually changing the ratio of free holes and electrons at the interface. After staying for 10 s at 0.0 V ($t_{wait1}$), the gate bias is ramped back to -2.0 V ($S_{U1}$). The difference in the $V_{TH}$ shift recorded at -2.0 V at the beginning of $S_{D1}$ and at the end of $S_{U1}$ is denoted as the bias-dependent recovery contribution ($\Delta V_{TH}^{bias}$). After the first ramp down-up cycle, the maximum CP current is recorded for 10 s by pulsing the gate junction between strong inversion (-2.0 V) and accumulation (+1.0 V) at a frequency of 500 kHz ($t_{CP}$). In the analysis, the maximum CP signal is converted into an interface state dependent threshold voltage shift ($\Delta V_{TH}^{it}$) by assuming an amphoteric nature of interface traps [89] and a flat density of state profile [37], cf. Subsection 2.2.2. After the CP cycle, a short 10 s lasting constant gate bias phase at -2.0 V ($t_{R2}$) is performed followed by a second down-up ramp ($S_{D2}$; $t_{wait2}$; $S_{U2}$). This basic MSM procedure is repeated six times on both devices of the wafer split with increasing stress times $t_S$ (1/10/100/1,000/10,000/100,000 s).

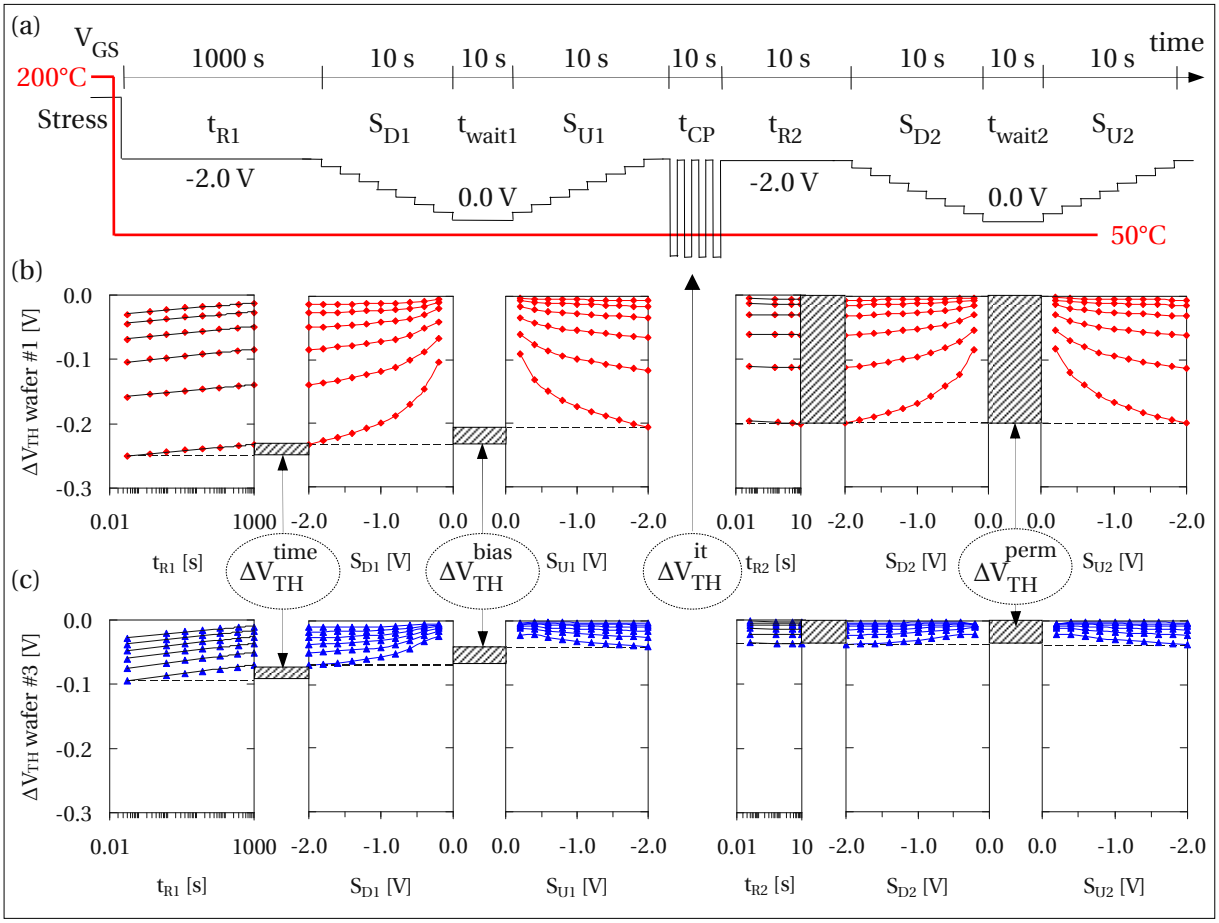


**Figure 7.7:** (a) Basic MSM procedure used for degradation and recovery analysis. During stress, the polyheater tool is used to generate an elevated stress temperature of 200°C. During recovery, gate bias sweeps and CP measurements are performed in order to monitor time and bias dependent $V_{TH}$ recovery and interface state creation. (b) The individual $V_{TH}$ shifts recorded at 50°C after six subsequent stress runs (1/10/100/1,000/10,000/100,000 s): (b) wafer #1; (c) wafer #3. The time dependent $V_{TH}$ recovery ($\Delta V_{TH}^{time}$) is recored at a constant gate bias of -2.0 V directly post stress for 1,000 s ($t_{R1}$). The bias dependent $V_{TH}$ recovery component ($\Delta V_{TH}^{bias}$) is the difference in the $V_{TH}$ shift between $S_{D1}$ and $S_{U1}$ recorded at -2.0 V. After $S_{U1}$, the maximum CP current is recorded for 10 s ($t_{CP}$) and changes in $I_{CP}^{max}$ are converted into corresponding interface state dependent $V_{TH}$ shifts ($\Delta V_{TH}^{it}$). After gate pulsing, the remaining $V_{TH}$ shift is considered to be quasi-permanent ($\Delta V_{TH}^{perm}$) since it is constant and independent of time and bias within the scope of the experiment, cf. $t_{R2}$, $S_{D2}$ and $S_{U2}$ in (b) and (c).

The $V_{TH}$ shifts measured during the different stages of the experiment are illustrated for wafer #1 (thin Ti/high H) in Fig. 7.7 (b) and for wafer #3 (thick Ti/low H) in Fig. 7.7 (c). Shown are six curves corresponding to the six subsequent stress runs. In Fig. 7.8, the individual $V_{TH}$ shifts are depicted separately for both wafers as a function of the stress time.

The following characteristics are obtained: (i) within the initial 1,000 s constant bias phase in strong inversion (-2.0 V), a similar amount of time dependent recovery ($\Delta V_{TH}^{time}$) is obtained for both H-levels, cf. Fig. 7.8 (a1); (ii) the total $V_{TH}$ shift decreases during $S_{D1}$ and increases during $S_{U1}$, cf. Fig. 7.7 (b)

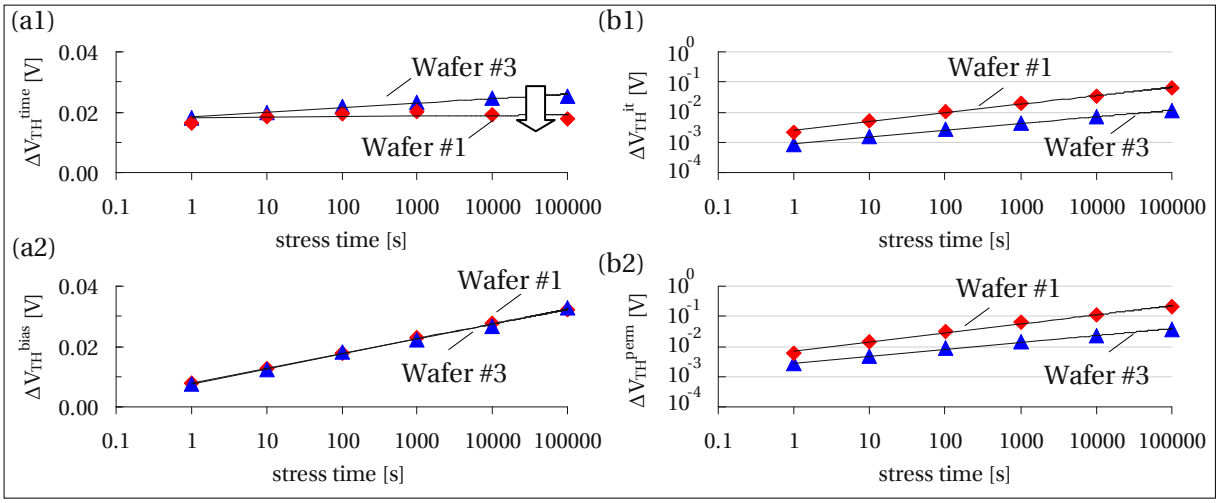


**Figure 7.8:** The individual $V_{TH}$ shifts depicted for both wafers as a function of stress time: (a1) $\Delta V_{TH}^{time}$; (a2) $\Delta V_{TH}^{bias}$; (b1) $\Delta V_{TH}^{it}$ and (b2) $\Delta V_{TH}^{perm}$. $\Delta V_{TH}^{time}$ and $\Delta V_{TH}^{bias}$ show a log-like evolution with stress time and their contribution is similar for both split wafers. $\Delta V_{TH}^{it}$ and $\Delta V_{TH}^{perm}$ follow a power-law-like evolution with stress time and scale by a factor 3. The weakly passivated wafer #3 provides considerably lower interface and quasi-permanent damage than the well passivated wafer #1. Furthermore, the well passivated wafer #1 provides larger power-law exponents ($n_{perm}^{\#1} = 0.30$; $n_{it}^{\#1} = 0.29$) than the weakly passivated wafer #3 ($n_{perm}^{\#3} = 0.23$; $n_{it}^{\#3} = 0.22$).

and (c); (iii) a significant bias dependent reduction in the $V_{TH}$ shift is observed after $S_{U1}$ which is again similar for both H-levels, cf. Fig. 7.8 (a2); (iv) after the intermediate CP cycle, the remaining degradation level is quasi-permanent and cannot be reduced further by an additional gate bias ramp toward 0.0 V, cf. Fig. 7.7 (b) and (c); (v) The remaining quasi-permanent $V_{TH}$ shift ($\Delta V_{TH}^{perm}$) and the interface state dependent $V_{TH}$ shift ($\Delta V_{TH}^{it}$) are much larger for wafer #1 than for wafer #3, cf. Fig. 7.8 (b1) and (b2); (vi) the interface state dependent $V_{TH}$ shift is smaller than the quasi-permanent $V_{TH}$ shift but scales with $\Delta V_{TH}^{perm}$ when multiplying $\Delta V_{TH}^{it}$ by a factor 3; (vii) The power-law exponents ($n_{perm}$ and $n_{it}$) are larger for wafer #1 than for wafer #3 causing faster degradation of the well passivated sample.

These seven findings on the bias and time dependence of the recovery, on interface state creation and on quasi-permanent damage, as a function of the H budget within the gate oxide, represent a comprehensive collection of NBTI characteristics. In the following, the single statements are cross-checked against our microscopic model presented in Fig. 6.27.

(i) Recovery over time is independent of the H-level: This agrees with our model which predicts the recoverable path (path A) to be nearly independent of the hydrogen budget within the gate oxide except for the small fraction of recoverable damage being converted to locked-in oxide charge by hydrogen capture from the interface (path B), cf. Fig. 7.8 (a1).

(ii) The $V_{TH}$ shift depends on the read-out gate bias: This agrees with our model which suggests that oxide traps and/or interface states may be neutralized and become positively charged again as the Fermi level crosses the silicon bandgap, cf. Subsection 6.3.1.

(iii) Ramping the gate bias toward 0.0 V accelerates $V_{TH}$ recovery, the effect being independent of the H-level: Agrees with our model which suggests bias dependent neutralization and relaxation of positively charged hydrogen-independent oxide defects.

(iv) The degradation level is quasi-permanent after the first down-up double ramp and the intermediate CP cycle: Our model explains this by the bias-accelerated relaxation of recoverable oxide defects. A large portion of oxide defects is assumed to be removed after the first double ramp, making the remaining $V_{TH}$ shift appear as quasi-permanent within the scope of our experiment.

(v) The remaining quasi-permanent $V_{TH}$ shift is much larger for the H-rich wafer: Agrees with our model considering that the larger quasi-permanent damage of the H-rich wafer #1 is a logical consequence of its initially higher Si–H precursor concentration at the interface.

(vi) $\Delta V_{TH}^{it}$ is smaller but proportional to $\Delta V_{TH}^{perm}$: Provided that the conversion of the CP signal into an interface state dependent $V_{TH}$ shift is correct, this statement agrees with our model which suggests $\Delta V_{TH}^{perm}$ to be the sum of locked-in oxide defects and a comparable amount of charged interface states ($Q^{perm} = Q_{OX}^{perm} + Q_{it}^{perm}$). Due to the simultaneous creation of $Q_{it}^{perm}$ and $Q_{OX}^{perm}$ via hydrogen exchange between passivated interface states and E′ centers, the model suggests a physical 50:50 relation and a strong correlation between $\Delta V_{TH}^{it}$ and $\Delta V_{TH}^{perm}$. The deviation in the measured factor 3 from the proposed factor 2 (50:50) may be due to different energy distributions of interface and oxide charges, leading to a different electrical responses of both trap types.

(vii) The power-law exponent of quasi-permanent damage depends on the passivation degree of the interface: Consistent with our model considering that a well passivated interface (wafer #1) provides a larger amount of weak Si–H bonds which get broken more readily upon electrical stress.

## 7.3 Conclusions on the role of hydrogen in the NBTI

By varying the BEOL processing (layer sequence – Section 7.1; Ti barrier thickness – Section 7.2), different split wafers were produced which provide vastly different hydrogen contents within the gate oxide. The initial variance in the hydrogen budget has been evaluated electrically by means of CP measurements and physically by means of TOFSIMS analysis. After subjecting the individual samples to NBTS, the degradation and recovery signatures have been investigated by making use of the newly developed characterization techniques. In agreement with our proposed microscopic model, it has been demonstrated that the recoverable part of NBTI degradation is largely independent of hydrogen while the quasi-permanent $V_{TH}$ shift component is strongly linked to the total hydrogen budget within the gate oxide. Also, a strong correlation between the increase in the CP signal and the quasi-permanent $V_{TH}$ shift component was found which is also fully consistent with our model.

# 8

# Summary and Conclusions

Silicon device degradation and its correlation to hydrogen incorporation was investigated with a particular focus on the negative bias temperature instability (NBTI) which is typically observed when subjecting PMOS devices at elevated temperatures to a relatively large electric field across the gate oxide. As a consequence, a negative shift in the threshold voltage is usually observed causing a degradation of the drain current when measuring the same operating point again after stress.

Although the basic facts of the problem seem almost trivial at a first glance, an elaborated inspection of the electrostatics and dynamics of the effect quickly reveals that considerable precautions have to be taken for accurate extraction and correct interpretation of the phenomenon. For example, in the triode region of the device ($|V_{GS}| \geq |V_{TH}|$), the change in the drain current may be attributed to conventional defect charges counterbalancing the applied gate potential as well as to mobility degradation affecting the channel resistance as a consequence of Coulomb scattering at charged point defects. Different types of defects provide a broad distribution of energy levels located energetically close or within the silicon band edges. The trap states were found to be rechargeable which makes the gate bias (Fermi level position) during read-out an important parameter which crucially determines the detected amount of degradation. In the positive charge state the defect causes a negative contribution to the threshold voltage shift while in the neutral charge state the defect is electrically invisible and may undergo structural relaxation. The time constants for carrier exchange with the silicon substrate and for structural relaxation were found to be both bias and temperature dependent, suggesting a phonon assisted inelastic tunneling mechanism governing defect neutralization and a transition over a thermodynamic barrier for the subsequent relaxation event.

Due to the fact that the created damage starts to recover almost immediately after removal of stress, the delay time between the termination of stress and the actual measurement becomes an important

parameter. In most reported cases $V_{TH}$ recovery follows a log-like evolution in a semi-logarithmic time plot, suggesting a broad and homogeneous distribution of recovery time constants, and energy levels, respectively. Since the nature of the carrier exchange mechanism turned out to be inelastic, recovery is not only linked to the 'time since end of stress' but also to the characterization temperature. Indeed, defects recover more rapidly the larger the temperature. As a consequence, different analyzing temperatures complicate a direct comparison of measurement signals. Besides such dynamic inconsistencies, different analyzing temperatures affect the electrostatic magnitude of CP signals and transfer characteristics as well. This is because the scanned energy interval during CP and the Fermi level position as a function of the gate voltage depend on temperature. Thus, when characterizing degradation after stressing devices at different temperatures, a reasonable comparison of measurement signals is very challenging, if not impossible, as long as one is tied to the constraint of conventional measurement techniques that the stress temperature has to equal the recovery temperature.

This constraint was resolved in this study by the development and application of the 'degradation quenching' technique. By using in-situ polyheaters which surround the active silicon device and heat it up to a defined temperature when a power is applied, we are able to stress devices at different temperatures while characterizing them at a unique, much lower analyzing temperature. The degradation level remains conserved during the temperature switch by maintaining the stress bias until the lower characterization temperature has stabilized. This takes only a couple of seconds.

The application of the polyheater technique combined with bias ramp experiments like for example the 'incremental sweep' procedure has revealed several new characteristics of NBTI induced defects. It was found that there exists a specific defect type (probably an E′ center) which provides a large distribution of time constants. In particular, this defect is *recoverable* which means that it can be neutralized and annealed by moderate temperature enhancement or gate bias switches toward accumulation. Its energetic distribution with respect to the Fermi level position within the silicon substrate was investigated, showing two significant peaks, the first being located around midgap, the second being located in the upper half of the silicon bandgap. A second defect type was identified which is *quasi-permanent*. 'Quasi-permanent' means that the defect does not recover significantly due to single gate bias switches or moderate temperature enhancement and it appears also relatively stable in time within the scope of our experiments. However, even this quasi-permanent defect is not a 100 % 'permanent' degradation. Long continuous gate pulsing at high frequencies or high temperature treatment (400°C) let the defect recover almost completely. Nevertheless, in comparison to the previously described recoverable defect, the quasi-permanent defect appears often constant and shows completely different degradation and recovery dynamics. As opposed to the recoverable defect, the quasi-permanent defect correlates universally with the increase of the CP signal. This correlation was demonstrated for different temperatures and electric fields. The increase in the CP current is typically attributed to the creation of interface states. However, by converting the CP signal into an interface state dependent threshold voltage shift, it was found that only 30–50 % of the quasi-permanent $V_{TH}$ shift can be explained by fast SRH-like defects. According to our present microscopic model, the missing contribution is assigned to locked-in oxide defects which are assumed to be created when a hydrogen atom gets trapped into a recoverable E′ center, thereby blocking the relaxation of the formally recoverable oxide defect. The hydrogen atom is likely to come from broken

Si–H bonds at the interface. Such a coupling of interface states and locked-in oxide defects would explain the observed universial correlation between the quasi-permanent $V_{TH}$ shift and the CP current.

Besides the electrostatics and dynamics of individual defects, their correlation to hydrogen was investigated by means of BEOL process splits. In particular, it was found that the density of recoverable defects is practically independent of hydrogen, suggesting the precursor of $Q_{OX}^{rec}$ to be an hydrogen-free oxygen vacancy. If this is the case, the E′ center is a reasonable candidate representing the recoverable defect. On the other hand, the quasi-permanent defects were found to strongly depend on the hydrogen budget within the gate oxide. The universal correlation between the quasi-permanent $V_{TH}$ shift and the CP current has proven true also for different process splits. The strong link to hydrogen suggests a passivated dangling bond at the interface to be the precursor for quasi-permanent damage. The dynamic creation of quasi-permanent defects may be explained by the field- and temperature-assisted release of hydrogen from passivated Si–H bonds. Due to disorder induced variations in the bond properties, the Si–H binding energies are assumed to be dispersed symmetrically around a medium dissociation energy. As a consequence of this dispersion, different time constants for the release of hydrogen emerge. Based on the sum of defects properties collected within this PhD thesis, a microscopic model has been suggested which is capable to explain (at least qualitatively) the observed characteristics.

# 9 Appendix

## 9.1 Transconductance as a function of the gate bias overdrive

The transconductance in the linear regime is calculated by differentiating Eq. 2.1 by $V_{GS}$:

$$\begin{aligned} g_{m,lin} = \frac{\partial I_{D,lin}}{\partial V_{GS}} &\approx I_{D,lin}\left(\frac{1}{V_{GS}-V_{TH}} + \frac{1}{\mu_{eff}}\frac{\partial \mu_{eff}}{\partial V_{GS}}\right), \\ &\approx I_{D,lin}\left(\frac{1}{V_{GS}-V_{TH}} - \frac{\Theta}{1+\Theta(V_{GS}-V_{TH})} - \frac{N_{it}}{1+\alpha N_{it}}\frac{\partial \alpha}{\partial V_{GS}}\right), \\ &\approx I_{D,lin}\left(\frac{1}{(1+\Theta(V_{GS}-V_{TH}))(V_{GS}-V_{TH})} + \frac{\alpha N_{it}}{2(1+\alpha N_{it})(V_{GS}-V_{TH})}\right), \\ &\approx -\frac{W}{L}\frac{\mu_0 C_{OX} V_{DS}}{(1+\alpha N_{it})}\left(\frac{1}{(1+\Theta(V_{GS}-V_{TH}))^2} + \frac{\alpha N_{it}}{2(1+\alpha N_{it})(1+\Theta(V_{GS}-V_{TH}))}\right). \end{aligned}$$

The transconductance in the saturation regime is calculated by differentiating Eq. 2.2 by $V_{GS}$:

$$\begin{aligned} g_{m,sat} = \frac{\partial I_{D,sat}}{\partial V_{GS}} &\approx I_{D,sat}\left(\frac{2}{V_{GS}-V_{TH}} + \frac{1}{\mu_{eff}}\frac{\partial \mu_{eff}}{\partial V_{GS}}\right), \\ &\approx I_{D,sat}\left(\frac{2}{V_{GS}-V_{TH}} - \frac{\Theta}{1+\Theta(V_{GS}-V_{TH})} - \frac{N_{it}}{1+\alpha N_{it}}\frac{\partial \alpha}{\partial V_{GS}}\right), \\ &\approx I_{D,sat}\left(\frac{2+\Theta(V_{GS}-V_{TH})}{(1+\Theta(V_{GS}-V_{TH}))(V_{GS}-V_{TH})} + \frac{\alpha N_{it}}{2(1+\alpha N_{it})(V_{GS}-V_{TH})}\right), \\ &\approx -\frac{W}{L}\frac{\mu_0 C_{OX}(V_{GS}-V_{TH})}{(1+\alpha N_{it})}\left(\frac{2+\Theta(V_{GS}-V_{TH})}{(1+\Theta(V_{GS}-V_{TH}))^2} + \frac{\alpha N_{it}}{2(1+\alpha N_{it})(1+\Theta(V_{GS}-V_{TH}))}\right). \end{aligned}$$

The linear transconductance ($g_{m,lin}$) decreases gradually in the triode region of the device ($|V_{GS}| \geq |V_{TH}|$) due to enhanced field dependent scattering ($\Theta$). The saturation transconductance ($g_{m,sat}$) increases more or less linearly in the deep inversion region of the device. Due to field enhanced scattering ($\Theta$), a positive curvature is superimposed.

## 9.2 Variation of the transconductance with the interface state density

Close to the threshold voltage the linear transconductance reaches a maximum when its increase in the subthreshold region ($|V_{GS}| < |V_{TH}|$) is compensated by its decrease in the triode region ($|V_{GS}| \geq |V_{TH}|$) of the device. This maximum value of the transconductance may be approximated in the linear regime as follows:

$$g_{m,lin}^{max} \approx -\frac{W}{L}\frac{\mu_0 C_{OX} V_{DS}}{1+\alpha_0 N_{it}}.$$

The variation of the maximum transconductance with the increase in interface state density may be calculated by differentiating the maximum transconductance:

$$\frac{\partial g_{m,lin}^{max}}{\partial N_{it}} \approx \frac{W}{L}\mu_0 C_{OX} V_{DS}\frac{\alpha_0}{(1+\alpha_0 N_{it})^2},$$
$$\frac{\Delta g_{m,lin}^{max}}{g_{m,lin}^{max}} \approx -\frac{\alpha_0 \Delta N_{it}}{1+\alpha_0 N_{it}}.$$

In the saturation regime the transconductance does not reach a maximum which makes it considerably difficult to evaluate an appropriate operating point to refer to. The absence of this reference point is the reason why mobility degradation is usually analyzed in the linear drain current regime and not in the saturation regime. Close to the threshold voltage the saturation transconductance may be approximated as follows:

$$g_{m,sat}^{V_{TH}} \approx -2\frac{W}{L}\frac{\mu_0 C_{OX}(V_{GS}-V_{TH})}{1+\alpha_0 N_{it}}.$$

The variation of this saturation transconductance with the increase in interface state density may be calculated by differentiating the upper equation:

$$\frac{\partial g_{m,sat}^{V_{TH}}}{\partial N_{it}} \approx 2\frac{W}{L}\frac{\mu_0 C_{OX}(V_{GS}-V_{TH})}{1+\alpha_0 N_{it}}\left(\frac{\alpha_0}{1+\alpha_0 N_{it}}+\frac{1}{V_{GS}-V_{TH}}\frac{\partial V_{TH}}{\partial N_{it}}\right),$$
$$\frac{\Delta g_{m,sat}^{V_{TH}}}{g_{m,sat}^{V_{TH}}} \approx -\frac{\alpha_0 \Delta N_{it}}{1+\alpha_0 N_{it}}-\frac{\Delta V_{TH}^{q}}{V_{GS}-V_{TH}}.$$

Note that both equations are very crude approximations which are only valid close to the threshold voltage of the device.

# Hardware

The following table contains information about the different hardware used in this PhD thesis.

| Wafer ID | $t_{OX}$ | Oxide | Poly | Type | Process | Comment |
|---|---|---|---|---|---|---|
| SM5P/30/H1 | 30 nm | $SiO_2$ | $n^{++}$ | PMOS | SNIT/PM | high H |
| SM5N/30/H1 | 30 nm | $SiO_2$ | $n^{++}$ | NMOS | SNIT/PM | high H |
| SM5P/30/H2 | 30 nm | $SiO_2$ | $n^{++}$ | PMOS | Standard | med H |
| SM5P/30/H3 | 30 nm | $SiO_2$ | $n^{++}$ | PMOS | Metal 1 | low H |
| SM6P/30/STD1 | 30 nm | $SiO_2$ | $n^{++}$ | PMOS | Standard | med H |
| SM6P/30/STD2 | 30 nm | $SiO_2$ | $n^{++}$ | PMOS | Standard | med H |
| SM6P/30/H1 | 30 nm | $SiO_2$ | $n^{++}$ | PMOS | thin Ti/PM1 | high H |
| SM6P/30/H2 | 30 nm | $SiO_2$ | $n^{++}$ | PMOS | thin Ti/PM2 | high H |
| SM6P/30/H3 | 30 nm | $SiO_2$ | $n^{++}$ | PMOS | thick Ti/PM1 | low H |
| HK2P/1.5/1 | 1.5 nm | high–$\kappa$ | $p^{++}$ | PMOS | IMEC wafer | |

# Bibliography


[1] A. Goetzberger and H.E. Nigh, "Surface charge after annealing of Al–($SiO_2$)–Si structures under bias," *Proc.IEEE*, vol. 54, no. 10, pp. 1454–1454, 1966.

[2] Y. Miura and Y. Matukura, "Investigation of silicon-silicon dioxide interface using MOS structure," *Jpn.J.Appl.Phys.*, vol. 5, no. 2, pp. 180–180, 1966.

[3] B.E. Deal, M. Sklar, A.S. Grove, and E.H. Snow, "Characteristics of the surface-state charge ($Q_{SS}$) of thermally oxidized silicon," *J.Electrochem.Soc.*, vol. 114, no. 3, pp. 266–274, 1967.

[4] D. Frohman-Bentchkowsky, "A fully decoded 2048-bit electrically programmable FAMOS read-only memory," *IEEE J.Solid-State Circuits*, vol. 6, no. 5, pp. 301–306, 1971.

[5] A. Goetzberger and H.E. Nigh, "On the formation of surface states during stress aging of thermal Si–$SiO_2$) interfaces," *J.Electrochem.Soc.*, vol. 120, no. 1, pp. 90–96, 1973.

[6] K.O. Jeppson and M. Svensson, "Negative bias stress of MOS devices at high electric fields and degradation of NMOS devices," *J.Appl.Phys.*, vol. 48, no. 5, pp. 2004–2014, 1977.

[7] J.F. Zhang and W. Eccleston, "Positive bias temperature instability in MOSFET's," *IEEE Trans.Elect.Dev.*, vol. 45, no. 1, pp. 116–124, 1998.

[8] V. Huard, M. Denais, F. Perrier, N. Revil, C. Parthasarathy, A. Bravaix, and E. Vincent, "A thorough investigation of MOSFETs NBTI degradation," *Microelectron.Reliab.*, vol. 45, no. 1, pp. 83–98, 2005.

[9] D.K. Schroder and J.A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *J.Appl.Phys.*, vol. 94, no. 1, pp. 1–18, 2003.

[10] D.K. Schroder, "Negative bias temperature instability: What do we understand?," *Microelectron.Reliab.*, vol. 94, no. 1, pp. 841–852, 2007.

[11] J.H. Stathis and S. Zafar, "The negative bias temperature instability in MOS devices: A review," *Microelectron.Reliab.*, vol. 46, no. 2-4, pp. 270–286, 2006.

[12] M.A. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," *Microelectron.Reliab.*, vol. 45, no. 1, pp. 71–81, 2005.

[13] Y. Shi, T.P. Ma, S. Prasad, and S. Dhanda, "Polarity dependent gate tunneling currents in dual-gate CMOSFET's," *IEEE Trans.Elect.Dev.*, vol. 45, no. 11, pp. 2355–2360, 1998.

[14] K.F. Schuegraf, C.C. King, and C. Hu, "Impact of polysilicon depletion in thin oxide MOS technology," in *Proc.VLSI*, 1993, pp. 86–90.

[15] H. Reisinger, R.P. Vollertsen, P.J. Wagner, T. Huttner, A. Martin, S. Aresu, W. Gustin, T. Grasser, and C. Schlünder, "The effect of recovery on NBTI characterization of thick non-nitrided Oxides," in *Proc.IIRW*, 2008, pp. 1–6.

[16] J.H. Stathis, "Percolation models for gate oxide breakdown," *J.Appl.Phys.*, vol. 86, no. 10, pp. 5757–5766, 1999.

[17] J.W. McPherson and R.B. Khamankar, "Molecular model for intrinsic time-dependent dielectric breakdown in $SiO_2$ dielectrics and the reliability implications for hyper-thin gate oxide," *Semicond.Sci.Technol.*, vol. 45, no. 5, pp. 462–470, 2000.

[18] A. Teramoto, H. Umeda, K. Azamawari, K. Kobayashi, K. Shiga, J. Komori, Y. Ohno, and A. Shigetomi, "Time-dependent dielectric breakdown of $SiO_2$ films in a wide electric field range," *Microelectron.Reliab.*, vol. 41, no. 1, pp. 47–52, 2001.

[19] B.J. Sheu, D.L. Scharfetter, P.-K. Ko, and M.-C. Jeng, "BSIM: Berkeley short-channel IGFET model for MOS transistors," *IEEE J.Solid-State Circuits*, vol. 22, no. 4, pp. 558–566, 1987.

[20] H. Haddara and D.S. Cristoloveanu, "Two-dimensional modeling of locally damaged short-channel MOSFET's operating in the linear region," *IEEE Trans.Elect.Dev.*, vol. 34, no. 2, pp. 378–385, 1987.

[21] A.T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John, and S. Krishnan, "NBTI impact on transistor & circuit: Models, mechanisms & scaling effects," in *Proc.IEDM*, 2003, pp. 14.5.1–14.5.4.

[22] S.C. Sun and J.D. Plummer, "Electron mobility in inversion and accumulation layers on thermally oxidized silicon surfaces," *IEEE Trans.Elect.Dev.*, vol. 27, no. 8, pp. 1497–1508, 1980.

[23] F.-C. Hsu and S. Tam, "Relationship between MOSFET degradation and hot-electron-induced interface-state generation," *IEEE Electron Device Lett.*, vol. 5, no. 2, pp. 50–52, 1984.

[24] J.E Chung, P.-K. Ko, and C. Hu, "A model for hot-electron-induced MOSFET linear-current degradation based on mobility reduction due to interface-state generation," *IEEE Trans.Elect.Dev.*, vol. 38, no. 6, pp. 1362–1370, 1991.

[25] A. Koga, S. Takagi, and A. Toriumi, "A comprehensive study of MOSFET electron mobility in both weak and strong inversion regimes," in *Proc.IEDM*, 1994, pp. 18.6.1–18.6.4.

[26] A.K.M. Ahsan and D.K. Schroder, "Impact of channel carrier displacement and barrier height lowering on the low-frequency noise characteristics of surface-channel n-MOSFETs," *Solid State Electron.*, vol. 49, no. 4, pp. 654–662, 2005.

[27] 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 *Proc.IRPS*, 2005, pp. 381–387.

[28] W. Shockley and W.T. Read, "Statistics of the recombinations of electrons and holes," *Phys.Rev.*, vol. 87, no. 5, pp. 835–842, 1952.

[29] S. Decker, *Numerical simulations performed by Stefan Decker*, Infineon Munich (Germany), process simulator TSUPREM4, device simulator MEDICI.

[30] J.-S. Lyu, K.-S. Nam, and Choochon Lee, "Determination of the interface traps density in metal oxide semiconductor field-effect transistor through subthreshold slope measurement," *Jpn.J.Appl.Phys.*, vol. 32, no. 10, pp. 4393–4397, 1993.

[31] N.S. Saks and M.G. Ancona, "Generation of interface states by ionizing radiation at 80K measured by charge pumping and subthreshold slope techniques," *IEEE Trans.Nucl.Sci.*, vol. 34, no. 6, pp. 1347–1354, 1987.

[32] S.C. Witczak, J.S. Suehle, and M. Gaitan, "An experimental comparison of measurement techniques to extract Si–$SiO_2$ interface trap density," *Solid State Electron.*, vol. 35, no. 3, pp. 345–355, 1992.

[33] C. Schlünder, M. Hoffmann, R.-P. Vollertsen, G. Schindler, W. Heinrigs, W. Gustin, and H. Reisinger, "A novel multi-point NBTI characterization methodology using smart intermediate stress (SIS)," in *Proc.IRPS*, 2008, pp. 79–86.

[34] D. Brisbin and P. Chaparala, "The effect of the subthreshold slope degradation on NBTI device characterization," in *Proc.IIRW*, 2008, pp. 96–99.

[35] J.F. Zhang, Z. Ji, M.H. Chang, B. Kaczer, and G. Groeseneken, "Real Vth instability of pMOSFETs under practical operation conditions," in *Proc.IEDM*, 2007, pp. 817–820.

[36] J.S. Brugler and P.G.A. Jespers, "Charge pumping in MOS devices," *IEEE Trans.Elect.Dev.*, vol. 16, no. 3, pp. 297–284, 1969.

[37] G. Groeseneken, H.E. Maes, N. Beltran, and R.F. De Keersmaecker, "A reliable approach to charge-pumping measurements in MOS transistors," *IEEE Trans.Elect.Dev.*, vol. 31, no. 1, pp. 42–53, 1984.

[38] G. Van den Bosch and G. Goeseneken, "On the geometric component of charge-pumping current in MOSFET's," *IEEE Electron Device Lett.*, vol. 14, no. 3, pp. 107–109, 1993.

[39] T.N. Duyet, H. Ishikuro, M. Takamiya, T. Saraya, and T. Hiramoto, "Suppression of geometric component of charge pumping current in thin film silicon on insulator metal-oxide-semiconductor-field-effect-transistors," *Jpn.J.Appl.Phys.*, vol. 37, pp. L855–L858, 1998.

[40] D. Heh, G. A. Brown, A. Diebold, and J. B. Bernstein, "Spatial distributions of trapping centers in $HfO_2/SiO_2$ gate stack," *IEEE Trans.Elect.Dev.*, vol. 54, no. 6, pp. 1338–1345, 2007.

[41] S.M. Sze and K.K. Ng, *Physics of semiconductor devices*, A John Wiley & Sons, 3rd edition, 2006.

[42] N.S. Saks, "Determination of interface trap capture cross sections using three-level charge pumping," *IEEE Electron Device Lett.*, vol. 11, no. 8, pp. 339–341, 1990.

[43] N.S. Saks, "Measurement of single interface trap capture cross sections with charge pumping," *Appl.Phys.Lett.*, vol. 70, no. 25, pp. 3380–3382, 1997.

[44] D. Vuillaume, R. Bouchakour, M. Jourdain, and J.C. Bourgoin, "Capture cross section of Si–$SiO_2$ interface states generated during electron injection," *Appl.Phys.Lett.*, vol. 55, no. 2, pp. 153–155, 1989.

[45] D. Goguenheim, D. Vuillaume, G. Vincent, and N.M. Johnson, "Accurate measurements of capture cross sections of semiconductor insulator interface states by a trap-filling experiment: The charge-potential feedback effect," *J.Appl.Phys.*, vol. 68, no. 3, pp. 1104–1113, 1990.

[46] J. Albohn, W. Füssel, N.D. Sinh, K. Kliefoth, and W. Fuhs, "Capture cross sections of defect states at the Si/$SiO_2$ interface," *J.Appl.Phys.*, vol. 88, no. 2, pp. 842–849, 2000.

[47] E.H. Nicollian and A. Goetzberger, "MOS conductance technique for measuring surface state parameters," *Appl.Phys.Lett.*, vol. 7, no. 8, pp. 216–219, 1965.

[48] G. Van den Bosch, G.V. Groeseneken, P. Heremans, and H.E. Maes, "Spectroscopic charge pumping: A new procedure for measuring interface trap distributions on MOS transistors," *IEEE Trans.Elect.Dev.*, vol. 38, no. 8, pp. 1820–1831, 1991.

[49] T. Aichinger and M. Nelhiebel, "Advanced energetic and lateral sensitive charge pumping profiling methods for MOSFET device characterization, analytical discussion and case studies," *IEEE Trans.Dev.Mater.Rel.*, vol. 8, no. 8-9, pp. 509–518, 2008.

[50] E.H. Poindexter, G.J. Gerardi, M.-E. Rueckel, N.M. Johnson P.J. Caplan, and D.K. Biegelsen, "Electronic traps and $P_b$ centers at the $SiO_2$/Si interface: Band-gap energy distribution," *J.Appl.Phys.*, vol. 56, no. 10, pp. 2844–2849, 1984.

[51] V. Huard, C.R. Parthasarath, C. Guerin, and M. Denais, "Physical modeling of negative bias temperature instabilities for predictive extrapolation," in *Proc.IRPS*, 2006, pp. 733–734.

[52] T. Grasser, W. Gös, V. Sverdlov, and B. Kaczer, "The universality of NBTI relaxation and its implications for modeling and characterization," in *Proc.IRPS*, 2007, pp. 268–280.

[53] T. Grasser, B. Kaczer, T. Aichinger, W. Gös, and M. Nelhiebel, "Defect creation stimulated by thermally activated hole trapping as the driving force behind negative bias temperature instability in $SiO_2$, SiON, and high–$\kappa$ gate stacks," in *Proc.IIRW*, 2008, pp. 91–95.

[54] M. Denais, Bravaix, V. Huard, C. Parthaarathy, G Ribes, F. Perrie, Y. Rey-Tauriac, and N. Revil, "On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's," in *Proc.IEDM*, 2004, pp. 5.2.1–5.2.4.

[55] T. Grasser, P.-J. Wagner, P. Hehenberger, W. Gös, and B. Kaczer, "A rigorous study of measurement techniques for negative bias temperature instability," in *Proc.IEDM*, 2007, pp. 6–11.

[56] S. Aresu, R. Pufall, M. Goroll, and W. Gustin, "NBTI on smart power technologies: A detailed analysis of two concurrent effects using a re-examined on-the-fly technique," *Microelectron.Reliab.*, vol. 48, no. 8-9, pp. 1310–1312, 2008.

[57] V.D. Maheta, V.D, E.N. Kumar, S. Purawat, C. Olsen, K. Ahmed, and S. Mahapatra, "Development of an ultrafast on-the-fly $I_{DLIN}$ technique to study NBTI in plasma and thermal oxynitride p-MOSFETs," *IEEE Trans.Elect.Dev.*, vol. 55, no. 10, pp. 2614–2622, 2008.

[58] Z. Ji, J.F. Zhang, M.H. Chang, B. Kaczer, and G. Groeseneken, "An analysis of the NBTI-induced threshold voltage shift evaluated by different techniques," *IEEE Trans.Elect.Dev.*, vol. 56, no. 5, pp. 1086–1093, 2009.

[59] A.E. Islam, E.N. Kumar, H. Das, S. Purawat, V. Maheta, H. Aono, E. Murakami, S. Mahapatra, and M.A. Alam, "Theory and practice of on-the-fly and ultra-fast VT measurements for NBTI degradation: Challenges and opportunities," in *Proc.IEDM*, 2007, pp. 805–808.

[60] P. Hehenberger, P.-J. Wagner, H. Reisinger, and T. Grasser, "Comparison of fast measurement methods for short-term negative bias temperature stress and relaxation," in *Proc.ESSDERC*, 2009, pp. 311–314.

[61] H. Reisinger, O. Blank, W. Heinrigs, W. Gustin, and C. Schlünder, "A Comparison of very fast to very slow components in degradation and recovery due to NBTI and bulk hole trapping to existing physical models," *IEEE Trans.Dev.Mater.Rel.*, vol. 7, no. 1, pp. 119–129, 2007.

[62] P. Hehenberger, P.-J. Wagner, H. Reisinger, and T. Grasser, "On the temperature and voltage dependence of short-term negative bias temperature stress," *Microelectron.Reliab.*, vol. 49, no. 9-11, pp. 1013–1017, 2009.

[63] W.J. Liu, Z.Y. Liu, D. Huang, C.C. Liao, L.F. Zhang, Z.H. Gan, W. Wong, C.Shen, and M.-F. Li, "On-the-fly interface trap measurement and its impact on the understanding of NBTI mechanism for p-MOSFETs with SiON gate dielectric," in *Proc.IEDM*, 2007, pp. 813–816.

[64] M.-F. Li, D. Huang, C. Shen, T. Yang, W.J. Liu, and Z. Liu, "Understand NBTI mechanism by developing novel measurement techniques," *IEEE Trans.Dev.Mater.Rel.*, vol. 8, no. 1, pp. 62–71, 2008.

[65] P. Hehenberger, T. Aichinger, T. Grasser, W. Gös, O. Triebl, B. Kaczer, and M. Nelhiebel, "Do NBTI-induced interface states show fast recovery? A study using a corrected on-the-fly charge-pumping measurement technique," in *Proc.IRPS*, 2009, pp. 1033–1038.

[66] G. Chen, M.-F. Li, C.H. Ang, J.Z. Zheng, and D.L. Kwong, "Dynamic NBTI of p-MOS transistors and its impact on MOSFET scaling," *IEEE Electron Device Lett.*, vol. 23, no. 12, pp. 734–736, 2002.

[67] M. Ershov, S. Saxena, H. Karbasi, S. Winters, S. Minehane, J. Babcock, R. Lindley, P. Clifton, M. Redford, and A. Shibkov, "Dynamic recovery of negative bias temperature instability in p-type metal-oxide-semiconductor field-effect transistors," *Appl.Phys.Lett.*, vol. 83, no. 8, pp. 1647–1649, 2003.

[68] B. Zhu, J.S. Suehle, and J.B. Bemstein, "Mechanism for reduced NBTI effect under pulsed bias stress conditions," in *Proc.IRPS*, 2004, pp. 689–690.

[69] G. Chen, K. Y. Chuah, M.-F. Li, SH Chan, C. H. Ang, J.Z. Zhcng, Y. Jin, and D. L. Kwon, "Dynamic NBTI of p-MOS transistors and its impact on device lifetime," in *Proc.IRPS*, 2003, pp. 196–202.

[70] W. Abadeer and W. Ellis, "Behavior of NBTI under AC dynamic circuit conditions," in *Proc.IRPS*, 2003, pp. 17–22.

[71] R. Fernández, B. Kaczer, A. Nackaerts, S. Demuynck, R. Rodríguez, M. Nafría, and G. Groeseneken, "AC NBTI studied in the 1 Hz – 2 GHz range on dedicated on-chip CMOS circuits," in *Proc.IEDM*, 2006, pp. 1–4.

[72] V. Huard, C. Parthasarathy, N. Rallet, C. Guerin, M. Mammasel, D. Barge, and C. Ouvrard, "New characterization and modeling approach for NBTI degradation from transistor to product level," in *Proc.IEDM*, 2007, pp. 797–800.

[73] B. Kaczer, V. Arkhipov, M. Jurczak, and G. Groeseneken, "Negative bias temperature instability (NBTI) in $SiO_2$ and SiON gate dielectrics understood through disorder-controlled kinetics," *Microelectron.Eng.*, vol. 80, no. 1, pp. 122–125, 2005.

[74] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwinc, "Temperature dependence of the negative bias temperature instability in the framework of dispersive transport," *Appl.Phys.Lett.*, vol. 86, no. 14, pp. (143506–1)–(143506–3), 2005.

[75] B. Bindu, W. Gös, B. Kaczer, and T. Grasser, "Analytical solution of the switching trap model for negative bias temperature stress," in *Proc.IIRW*, 2009, pp. 93–96.

[76] M.A. Alam, H. Kufluoglu, D. Varghese, and S. Mahapatra, "A comprehensive model for PMOS NBTI degradation: Recent progress," *Microelectron.Reliab.*, vol. 47, no. 6, pp. 853–862, 2007.

[77] T. Grasser and B. Kaczer, "Evidence that two tightly coupled mechanisms are responsible for negative bias temperature instability in oxynitride MOSFETs," *IEEE Trans.Elect.Dev.*, vol. 56, no. 5, pp. 1056–1062, 2009.

[78] V. Huard, C.R. Parthasarathy, C. Guerin, and M. Denais, "Physical modeling of negative bias temperature instabilities for predictive extrapolation," in *Proc.IRPS*, 2006, pp. 733–734.

[79] R.A Weeks, "Paramagnetic resonance of lattice defects in irradiated quartz," *J.Appl.Phys.*, vol. 27, no. 11, pp. 1376–1381, 1956.

[80] B.E. Deal, "The current understanding of charges in the thermally oxidized silicon structure," *J.Electrochem.Soc.*, vol. 121, no. 6, pp. 198–205, 1974.

[81] B.E. Deal, "Standardized terminology for oxide charges associated with thermally oxidized silicon," *IEEE Trans.Elect.Dev.*, vol. 27, no. 3, pp. 606–608, 1980.

[82] A.H. Edwards, "Theory of the $P_b$ center at the $\langle 111 \rangle$ $Si/SiO_2$," *Phys.Rev.B*, vol. 36, no. 18, pp. 9638–9648, 1987.

[83] P.M. Lenahan, "Atomic scale defects involved in MOS reliability problems," *Microelectron.Eng.*, vol. 69, no. 2-4, pp. 173–181, 2003.

[84] A. Stesman, "Dissociation kinetics of hydrogen-passivated $P_b$ defects at the ⟨111⟩ Si/$SiO_2$ interface," *Phys.Rev.B*, vol. 61, no. 12, pp. 8393–8403, 2000.

[85] P.M. Lenahan, K.L. Brower, and P.V. Dressendorfer, "Radiation-induced trivalent silicon defect buildup at the Si–$SiO_2$ interface in MOS structures," *IEEE Trans.Nucl.Sci.*, vol. 28, no. 6, pp. 4105–4106, 1981.

[86] L.A Ragnarssona and P. Lundgren, "Electrical characterization of $P_b$ centers in (100) Si–$SiO_2$ structures: The influence of surface potential on passivation during post metallization anneal," *J.Appl.Phys.*, vol. 88, no. 2, pp. 938–942, 2000.

[87] H. Sakaki, K. Hoh, and T. Sugano, "Determination of interface-state density and mobility ratio in silicon surface inversion layers," *IEEE Trans.Elect.Dev.*, vol. 17, no. 10, pp. 892–896, 1970.

[88] T. Sakurai and T. Sugano, "Theory of continuously distributed trap states at Si–$SiO_2$ interfaces," *J.Appl.Phys.*, vol. 52, no. 4, pp. 2889–2896, 1981.

[89] P.V. Gray and D.M. Brown, "Density of $SiO_2$–Si interface states," *Appl.Phys.Lett.*, vol. 8, no. 2, pp. 31–33, 1966.

[90] P.J. McWhorter, P.S. Wimkur, and R.A. Pastorek, "Donor/acceptor nature of radiation-induced interface traps," *Appl.Phys.Lett.*, vol. 35, no. 6, pp. 1154–1159, 1988.

[91] A. Haggag, W. McMahon, K. Hess, K. Cheng, J. Lee, and J. Lyding, "High-performance chip reliability from short-time-tests," in *Proc.IRPS*, 2001, pp. 271–279.

[92] R.A.B. Devine, J.-L. Autran, W.L. Warren, K.L. Vanheusdan, and J.-C. Rostaing, "Interfacial hardness enhancement in deuterium annealed 0.25 mm channel metal oxide semiconductor transistors," *Appl.Phys.Lett.*, vol. 70, no. 22, pp. 2999–3001, 1997.

[93] P.E. Bunson, M. Di Ventra, S.T. Pantelides, D.M. Fleetwood, and R.D. Schrimpf, "Hydrogen-related defects in irradiated $SiO_2$," *IEEE Trans.Nucl.Sci.*, vol. 47, no. 6, pp. 2289–2296, 2000.

[94] R. Biswas, Y.-P. Li, and B.C. Pan, "Enhanced stability of deuterium in silicon," *Appl.Phys.Lett.*, vol. 72, no. 26, pp. 3500–3502, 1998.

[95] S.T. Pantelides, S.N. Rashkeev, R. Buczko, D.M. Fleetwood, and R.D. Schrimpf, "Reactions of hydrogen with Si–$SiO_2$ interfaces," *IEEE Trans.Nucl.Sci.*, vol. 47, no. 6, pp. 2262–2268, 2000.

[96] K.L. Browler, "Dissociation kinetics of hydrogen passivated (111) Si–$SiO_2$ interface defects," *IEEE Trans.Nucl.Sci.*, vol. 42, no. 6, pp. 3444–3454, 1990.

[97] J.K. Rudra and W.B. Fowler, "Oxygen vacancy and the $E_1'$ center in crystalline $SiO_2$," *Phys.Rev.B*, vol. 35, no. 15, pp. 8223–8230, 1987.

[98] R.K. Freitag, D.B. Brown, and C.M. Dozier, "Experimental evidence of two species of radiation induced trapped positive charge," *IEEE Trans.Nucl.Sci.*, vol. 40, no. 6, pp. 1316–1322, 1993.

[99] J.P. Campbell, P.M. Lenahan, A.T. Krishnan, and S. Krishnan, "Direct observation of the structure of defect centers involved in the negative bias temperature instability," *Appl.Phys.Lett.*, vol. 87, no. 20, pp. (204106–1)–(204106–3), 2005.

[100] L.P. Trombetta, F.J. Feigl, and R.J. Zeto, "Positive charge generation in metal-oxide-semiconductor capacitors," *J.Appl.Phys.*, vol. 69, no. 4, pp. 2512–2521, 1991.

[101] J.F. Conley and P.M. Lenahan, "Electron spin resonance evidence that $E'_\gamma$ centers can behave as switching oxide traps," *IEEE Trans.Nucl.Sci.*, vol. 42, no. 6, pp. 1744–1749, 1995.

[102] J.F. Conley, P.M. Lenahan, A.J. Lelis, and T.R. Oldham, "Electron spin resonance evidence for the structure of a switching oxide trap: Long term structural change at silicon dangling bond sites in $SiO_2$," *Appl.Phys.Lett.*, vol. 67, no. 15, pp. 2179–2181, 1995.

[103] V.S. Pershenkov, S.V. Cherepko, A.V. Sogoyan, V.V. Belyakov, V.N. Ulimov, V.V. Abramov, A.V. Shalnov, and V.I. Rusanovsky, "Proposed two-level acceptor-donor (AD) center and the nature of switching traps in irradiated MOS structures," *IEEE Trans.Nucl.Sci.*, vol. 43, no. 6, pp. 2579–2586, 1996.

[104] D.M. Fleetwood, M.R. Shaneyfelt, W.L. Warren, J.R. Schwabk, T.L. Meisenheimer, and P.S. Winokur, "Border traps: Issues for MOS radiation response and long-term reliability," *Microelectron.Reliab.*, vol. 35, no. 6, pp. 403–428, 1995.

[105] P.M. Lenahan and P.V. Dressendorfer, "Hole traps and trivalent silicon centers in metal/oxide/silicon devices," *J.Appl.Phys.*, vol. 55, no. 10, pp. 3495–3499, 1984.

[106] P.M. Lenahan and J.F. Conley, "What can electron paramagnetic resonance tell us about the $Si/SiO_2$ system?," *J.Vac.Sci.&Techn.B*, vol. 16, no. 4, pp. 2134–2153, 1998.

[107] F.J. Feigl, W.B. Fowler, and K.L. Yip, "Oxygen vacancy model for the $E'_1$ center in $SiO_2$," *Solid State Commun.*, vol. 14, no. 3, pp. 225–229, 1974.

[108] C.J. Nicklaw, Z.-Y. Lu, D.M. Fleetwood, R.D. Schrimpf, and S. T. Pantelides, "The structure, properties, and dynamics of oxygen vacancies in amorphous $SiO_2$," *IEEE Trans.Nucl.Sci.*, vol. 49, no. 6, pp. 2667–2673, 2002.

[109] S.T. Pantelides, Z.-Y. Lu, C. Nicklaw, T. Bakos, S.N. Rashkeev, D.M. Fleetwood, and R.D. Schrimpf, "The $E'$ center and oxygen vacancies in $SiO_2$," *J.Non-Cryst.Solids*, vol. 254, no. 2-9, pp. 217–223, 2008.

[110] A.J. Lelis and T.R. Oldham, "Time dependence of switching oxide traps," *IEEE Trans.Nucl.Sci.*, vol. 41, no. 6, pp. 1835–1843, 1994.

[111] J.R. Chavez, S.P. Karna, K. Vanheusden, C.P. Brothers, R.D. Pugh, B.K. Singaraju, W.L. Warren, and R.A.B. Devine, "Microscopic structure of the $E'_\delta$ center in amorphous $SiO_2$: A first principles quantum mechanical investigation," *IEEE Trans.Nucl.Sci.*, vol. 44, no. 6, pp. 1799–1803, 1997.

[112] W. Gös and T. Grasser, "Charging and discharging of oxide defects in reliability issues," in *Proc.IIRW*, 2007, pp. 27–32.

[113] K.L. Yip and W.B. Fowler, "Electronic structure of $E_1'$ centers in $SiO_2$," *Phys.Rev.B*, vol. 11, no. 6, pp. 2327–2338, 1975.

[114] A.J. Lelis and T.R. Oldham, "The nature of the trapped hole annealing process," *IEEE Trans.Nucl.Sci.*, vol. 36, no. 6, pp. 1808–1815, 1989.

[115] J.-W. Lee, M. Tomozawa, and R.K. MacCrone, "Annihilation of E′ center defects in silica glass by hydrogen treatment," *J.Non-Cryst.Solids*, vol. 67, no. 12, pp. 3510–3512, 2008.

[116] J.F. Conley and P.M. Lenahan, "Room temperature reactions involving silicon dangling bond centers and molecular hydrogen in amorphous $SiO_2$ thin films on silicon," *IEEE Trans.Nucl.Sci.*, vol. 39, no. 6, pp. 2186–2191, 1992.

[117] T. Takahashi, B.B. Triplett, K. Yokogawa, and T. Sugano, "Electron spin resonance observation of the creation, annihilation, and charge state of the 74-Gauss doublet in device oxides damaged by soft x rays," *Appl.Phys.Lett.*, vol. 51, no. 17, pp. 1334–1336, 1987.

[118] P.M. Lenahan, W.L Warren, D.T. Krick, P.V. Dressendorfer, and B.B. Triplett, "Interaction of molecular hydrogen with trapped hole E′ centers in irradiated and high field stressed metal/oxide/silicon oxides," *J.Appl.Phys.*, vol. 67, no. 12, pp. 7612–7614, 1990.

[119] J.K. Rudra, W.B. Fowler, and F.J. Feigl, "Model for the $E_2'$ center in alpha quartz," *Phys.Rev.Lett.*, vol. 55, no. 23, pp. 2614–2617, 1985.

[120] P.E. Blöchl and J.H. Stathis, "Hydrogen electrochemistry and stress-induced leakage current in silica," *Phys.Rev.Lett.*, vol. 83, no. 3, pp. 372–375, 1999.

[121] M. Vitiello, N. Lopez, F. Illas, and G. Pacchioni, "$H_2$ cracking at $SiO_2$ defect centers," *J.Phys.Chem.A*, vol. 104, no. 20, pp. 4674–4684, 2000.

[122] R.E. Stahlbush, B.J. Mrstik, and R.K. Lawrence, "Post-irradiation behavior of the interface state density and the trapped positive charge," *IEEE Trans.Nucl.Sci.*, vol. 37, no. 6, pp. 1641–1649, 1990.

[123] B. Tavel, M. Bidaud, N. Emonet, D. Barge, N. Planes, H. Brut, D. Roy, J.C. Vildeuil, R. Difrenza, K. Rochereau, M. Denais, V. Huard, P. Llinares, S. Bruyere, C. Parthasarthy, N. Revi, R. Pantel, F. Guyader, L. Vishnubotla, K. Barla, F. Amaud, P. Stolk, and M. Woo, "Thin oxynitride solution for digital and mixed-signal 65 nm CMOS platform," in *Proc.IEDM*, 2003, pp. 27.6.1–27.6.4.

[124] G. Barbottin and A. Vapaille, *Instabilities in silicon devices*, Elsevier Science B.V., 3st edition, 1999.

[125] A.E. Islam, G. Gupta, S. Mahapatra, A.T. Krishnan, K. Ahmed, F. Nouri, A. Oates, and M.A. Alam, "Gate leakage vs. NBTI in plasma nitrided oxides: characterization, physical principles, and optimization," in *Proc.IEDM*, 2006, pp. 1–4.

[126] J.P. Campbell, P.M. Lenahan, C.J. Cochrane, A.T. Krishnan, and S. Krishnan, "Atomic-scale defects involved in the negative-bias temperature instability," *IEEE Trans.Dev.Mater.Rel.*, vol. 7, no. 4, pp. 540–557, 2007.

[127] D.T. Krick, P.M. Lenahan, and J. Kanicki, "Electrically point defects in amorphous silicon nitride: An illumination and charge injection study," *J.Appl.Phys.*, vol. 64, no. 7, pp. 3558–3563, 1988.

[128] J.P. Campbell, P.M. Lenahan, A.T. Krishnan, and S. Krishnan, "Identification of atomic-scale defect structure involved in the negative bias temperature instability in plasma-nitrided devices," *Appl.Phys.Lett.*, vol. 91, no. 3, pp. (133507–1)–(133507–3), 2007.

[129] P.M. Lenahan and S.E. Curry, "First observation of the $^{29}$Si hyperfine spectra of silicon dangling bond centers in silicon nitride," *Appl.Phys.Lett.*, vol. 56, no. 2, pp. 600–603, 1990.

[130] W.L. Warren, P M. Lenahan, and S.E. Curry, "First observation of paramagnetic nitrogen dangling-bond centers in silicon nitride," *Phys.Rev.Lett.*, vol. 65, no. 2, pp. 207–210, 1990.

[131] W.L. Warren and P.M. Lenahan, "Electron-nuclear double-resonance and electron-spin-resonance study of silicon dangling-bond centers in silicon nitride," *Phys.Rev.B*, vol. 42, no. 3, pp. 1773–1780, 1990.

[132] W.L. Warren, J. Kanicki, and E.H. Poindexter, "Paramagnetic point defects in silicon nitride and silicon oxynitride thin films on silicon," *Colloids Surf., A*, vol. 115, pp. 311–317, 1996.

[133] D. Jousse, J. Kanicki, and J.H. Stathis, "Observation of multiple silicon dangling bond configurations in silicon nitride," *Appl.Phys.Lett.*, vol. 54, no. 11, pp. 1043–1045, 1989.

[134] J.P. Campbell, P.M. Lenahan, A.T. Krishnan, and S. Krishnan, "Location, structure, and density of states of NBTI-induced defects in plasma nitrided pMOSFETs," in *Proc.IRPS*, 2007, pp. 503–510.

[135] H. Wong and V.A. Gritsenko, "Defects in silicon oxynitride gate dielectric films," *Microelectron.Reliab.*, vol. 42, no. 4-5, pp. 597–605, 2002.

[136] J.T. Yount, P.M. Lenahan, and J.T. Krick, "Comparison of defect structure in $N_2O$- and $NH_3$-nitrided oxide dielectrics," *J.Appl.Phys.*, vol. 76, no. 3, pp. 1754–1758, 1994.

[137] F.H.P.M. Habraken, "Characterization of LPCVD and PECVD silicon oxynitride films," *Appl.Surf.Sci.*, vol. 30, pp. 186–196, 1987.

[138] V.J. Kapoor, R.S. Baifey, and H.J. Stein, "Hydrogen-related memory traps in thin silicon nitride films," *J.Vac.Sci.&Techn.A*, vol. 1, no. 2, pp. 600–603, 1983.

[139] J.P. Campbell, P.M. Lenahan, A.T. Krishnan, and S. Krishnan, "Atomic-scale defects involved in NBTI in plasma-nitrided pMOSFETs," in *Proc.IIRW*, 2006, pp. 12–17.

[140] 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 *Proc.IRPS*, 2007, pp. 1–9.

[141] S.S. Tan, T.P. Chen, J.M. Soon, K.P. Loh, C.H. Ang, and L. Chan, "Nitrogen-enhanced negative bias temperature instability: An insight by experiment and first-principle calculations," *Appl.Phys.Lett.*, vol. 82, no. 12, pp. 1881–1883, 2003.

[142] J.-S. Lyu, K.-S. Nam, and C. Lee, "Determination of the interface traps density in metal oxide semiconductor field-effect transistor through subthreshold slope measurement," *Jpn.J.Appl.Phys.*, vol. 32, no. 10, pp. 4393–4397, 1993.

[143] D.S. Ang, S. Wang, G.A. Du, and Y.Z. Hu, "A consistent deep-level hole trapping model for negative bias temperature instability," *IEEE Trans.Dev.Mater.Rel.*, vol. 8, no. 1, pp. 22–34, 2008.

[144] T. Grasser, B. Kaczer, W. Gös, T. Aichinger, P. Hehenberger, and M. Nelhiebel, "A two-stage model for negative bias temperature instability," in *Proc.IRPS*, 2009, pp. 33–44.

[145] T. Aichinger, M. Nelhiebel, and T. Grasser, "A combined study of p- and n-channel MOS devices to investigate the energetic distribution of oxide traps after NBTI," *IEEE Trans.Elect.Dev.*, vol. 56, no. 12, pp. 3018–3026, 2009.

[146] M.A. Alam, "A critical examination of the mechanics of dynamic NBTI for PMOSFETs," in *Proc.IEDM*, 2003, pp. 14.4.1–14.4.4.

[147] Y. Wang, "On the recovery of interface state in pMOSFETs subjected to NBTI and SHI stress," *Solid State Electron.*, vol. 52, no. 2, pp. 264–268, 2008.

[148] S. Mahapatra, D. Saha, D. Varghese, and P.B. Kumar, "On the generation and recovery of interface traps in MOSFETs subjected to NBTI, FN, and HCI stress," *IEEE Trans.Elect.Dev.*, vol. 53, no. 7, pp. 1583–1592, 2006.

[149] T. Yang, C. Shen, M.F. Li, C.H. Ang, C.X. Zhu, Y.-C. Yeo, G. Samudra, and D.-L. Kwong, "Interface trap passivation effect in NBTI measurement for p-MOSFET with SiON gate dielectric," *IEEE Electron Device Lett.*, vol. 26, no. 10, pp. 758–760, 2005.

[150] D.S. Ang, "Observation of suppressed interface state relaxation under positive gate biasing of ultrathin oxynitride gate p-MOSFET subjected to negative bias temperature stressing," *IEEE Electron Device Lett.*, vol. 27, no. 5, pp. 412–415, 2006.

[151] B. Kaczer, T. Grasser, R. Fernández, and G. Goeseneken, "Toward understanding of the wide distribution of time scales in negative bias temperature instability," *ECS Trans.*, vol. 6, no. 3, pp. 265–281, 2007.

[152] T.L. Tewksbury, *Relaxation effects in MOS devices due to tunnel exchange with near-interface oxide traps*, PhD. Thesis MIT, 1992.

[153] P.V. Sushko, S. Mukhopadhyay, A.S. Mysovsky, V.B. Sulimov, A. Taga, and A.L. Shluger, "Structure and properties of defects in amorphous silica: New insights from embedded cluster calculations," *J.Phys.-Condens.Matter*, vol. 17, pp. 2115–2140, 2005.

[154] T. Aichinger, M. Nelhiebel, S. Einspieler, and T. Grasser, "In-situ polyheater, a reliable tool for performing fast and defined temperature switches on chip," *IEEE Trans.Dev.Mater.Rel.*, vol. 76, no. 7, pp. 1–7, 2009.

[155] W. Muth and W. Walter, "Bias temperature instability assessment of n- and p-channel MOS transistors using a polysilicon resistive heated scribe lane test structure," in *Proc.ESSDERC*, 2007, pp. 1251–1262.

[156] C. Schluender, R.-P. Vollertsen, W. Gustin, and H. Reisinger, "A reliable and accurate approach to assess NBTI behavior of state-of-the-art pMOSFETs with fast-WLR," in *Proc.ESSDERC*, 2007, pp. 131–134.

[157] T.-K Kang, C.-S Wang, and K.-C. Su, "Self-heating p-channel metal-oxide-semiconductor field-effect-transistors for reliability monitoring of negative-bias temperature instability," *Jpn.J.Appl.Phys.*, vol. 46, no. 12, pp. 7639–7342, 2007.

[158] C.-S. Wang, W.-C. Chang, W.-S. Ke, C.-T. Chiang, C.-F. Lee, and K.-C. Su, "Characterization of embedded poly-heater pMOSFETs and its application on in-line wafer level NBTI monitor," in *Proc.SSDM*, 2005, pp. 580–581.

[159] C.-S. Wang, W.-C. Chang, W.-S. Ke, and K.-C. Su, "Ultra-fast negative bias temperature instability monitoring and end-of-life projection," in *Proc.IIRW*, 2006, pp. 136–138.

[160] H. Köck, V. Košel, C. Djelassi, M. Glavanovics, and D. Pogany, "IR thermography and FEM simulation analysis of on-chip temperature during thermal-cycling power-metal reliability testing using in-situ heated structures," *Microelectron.Reliab.*, vol. 49, no. 9-11, pp. 1132–1136, 2009.

[161] A. Kelleha and W. Lane, "Investigation of on-chip high temperature annealing of PMOS dosimeters," *IEEE Trans.Nucl.Sci.*, vol. 43, no. 3, pp. 997–1001, 1996.

[162] V. Košel, R. Sleik, and M. Glavanovics, "Transient non-linear thermal FEM simulation of smart power switches and verification by measurements," in *Proc.THERMINIC*, 2007, pp. 110–114.

[163] A. Ramalingam, F. Liu, S.R. Nassif, and D.Z. Pan, "Accurate thermal analysis considering nonlinear thermal conductivity," in *Proc.ISQED*, 2006, pp. 643–649.

[164] C.J. Glassbrenner and G.A. Slack, "Thermal conductivity of silicon and germanium from 3K to the melting point," *Phys.Rev.*, vol. 134, no. 4, pp. 1058–1069, 1964.

[165] V. Huard, F. Ivfonsieur, G. Ribes, and S. Brnyere, "Evidence for hydrogen-related defects during NBTI stress in p-MOSFETs," in *Proc.IRPS*, 2003, pp. 178–182.

[166] Y. Mitani, T. Yamaguchi, H. Satake, and A. Toriumi, "Reconsideration of hydrogen-related degradation mechanism in gate oxide," in *Proc.IRPS*, 2007, pp. 226–231.

[167] T.L. Tewksbury and H.-S. Lee, "Characterization, modeling, and minimization of transient threshold voltage shifts in MOSFET's," *IEEE J.Solid-State Circuits*, vol. 29, no. 3, pp. 239–252, 1994.

[168] T. Aichinger, M. Nelhiebel, and T. Grasser, "On the temperature dependence of NBTI recovery," in *Proc.ESREF*, 2008, pp. 1178–1184.

[169] T. Aichinger, M. Nelhiebel, and T. Grasser, "Unambiguous identification of the NBTI recovery mechanism using ultra fast temperature changes," in *Proc.IRPS*, 2009, pp. 2–7.

[170] H. Reisinger, O. Blank, W. Heinrigs, A. Mühlhoff, W. Gustin, and C. Schlünder, "Analysis of NBTI degradation- and recovery-behavior based on ultra fast Vth-measurements," in *Proc.IRPS*, 2006, pp. 448–453.

[171] B. Tuttle, "Structures, energies, and vibrational properties of Si–H bond dissociation in silicon," *Phys.Rev.B*, vol. 59, no. 20, pp. 12884–12889, 1999.

[172] S.D. Ganichev, E. Ziemann, W. Prettl, I.N. Yassievich, A.A. Istratov, and E.R. Weber, "Distinction between the Poole-Frenkel and tunneling models of electric-field-stimulated carrier emission from deep levels in semiconductors," *Phys.Rev.B*, vol. 61, no. 15, pp. 61–65, 2000.

[173] W. Gös, M. Karner, S. Tyaginov, P. Hehenberger, and T. Grasser, "Level shifts and gate interfaces as vital ingredients in modeling of charge trapping," in *Proc.SISPAD*, 2008, pp. 69–72.

[174] T. Aichinger, M. Nelhiebel, S. Einspieler, and T. Grasser, "Observing two stage recovery of gate oxide damage created under negative bias temperature stress," *J.Appl.Phys.*, vol. 107, pp. (024508–1)–(024508–8), 2010.

[175] E.H. Poindexter and W.L. Warren, "Paramagnetic point defects in amorphous thin films of $SiO_2$ and $Si_3N_4$," *J.Electrochem.Soc.*, vol. 142, no. 7, pp. 2508–2516, 1995.

[176] A.V. Kimmel, P.V. Sushko, A.L. Shluger, and G. Bersuker, "Positive and negative oxygen vacancies in amorphous silica," *ECS Trans.*, vol. 19, no. 2, pp. 3–17, 2009.

[177] V. Reddy, A.T. Krishnan, A. Marshall, J. Rodriguez, S. Natarajan, T. Rost, and S. Krishnan, "Impact of negative bias temperature instability on digital circuit reliability," in *Proc.IRPS*, 2002, pp. 248–254.

[178] T. Aichinger, M. Nelhiebel, and T. Grasser, "Investigation of the NBTI mechanism by low temperature characterization of arbitrarily stressed PMOS devices," *IEEE Trans.Elect.Dev.*, 2010, submitted for publication.

[179] D. Bauza, "Rigorous analysis of two-level charge pumping: Application to the extraction of interface trap concentration versus energy profiles in metal-oxide-semiconductor transistors," *J.Appl.Phys.*, vol. 94, no. 5, pp. 3239–3248, 2003.

[180] M. Nelhiebel, J. Wissenwasser, T. Detzel, A. Timmerer, and E. Bertagnolli, "Hydrogen-related influence of the metalization stack on characteristics and reliability of a trench gate oxide," *Microelectron.Reliab.*, vol. 45, no. 9-11, pp. 1355–1359, 2005.

[181] Y. Nissan-Cohen, "The effect of hydrogen on hot carrier and radiation immunity of MOS devices," *Appl.Surf.Sci.*, vol. 39, no. 1-4, pp. 511–522, 1989.

[182] L.J. Jin, H.P. Kuan, D. Sim, and M. Mukhopadhyay, "Influence of hydrogen annealing on NBTI performance," in *Proc.IPFA*, 2008, pp. 1–4.

[183] J.Z. Xie and S.P. Murarka, "Stability of hydrogen in silicon nitride films deposited by low-pressure and plasma enhanced chemical vapor deposition techniques," *J.Vac.Sci.&Techn.B*, vol. 7, no. 2, pp. 150–152, 1989.

[184] C.Y. Chang and S.M. Sze, *ULSI Technology*, McGraw-Hill, 1st edition, 1996.

[185] I. Jonak-Auer, R. Meisels, and F. Kuchar, "Determination of the hydrogen concentration of silicon nitride layers by Fourier transform infrared spectroscopy," *Infrared Phys. Technol.*, vol. 38, no. 4, pp. 223–226, 1997.

[186] T. Aichinger, M. Nelhiebel, and T. Grasser, "Energetic distribution of oxide traps created under negative bias temperature stress and their relation to hydrogen," *Appl.Phys.Lett.*, vol. 96, pp. (133511–1)–(133511–3), 2010.

[187] D.E. Woon, D.S. Marynick, and S.K. Estreicher, "Titanium and copper in Si: Barriers for diffusion and interaction with hydrogen," *Phys.Rev.B*, vol. 45, no. 23, pp. 13383–13389, 1992.

# Own Publications


[1] T. Aichinger, M. Nelhiebel, S. Einspieler, and T. Grasser, "Observing two stage recovery of gate oxide damage created under negative bias temperature stress," *J.Appl.Phys.*, vol. 107, pp. (024508–1)–(024508–8), 2010.

[2] T. Aichinger, M. Nelhiebel, and T. Grasser, "Energetic distribution of oxide traps created under negative bias temperature stress and their relation to hydrogen," *Appl.Phys.Lett.*, vol. 96, pp. (133511–1)–(133511–3), 2010.

[3] T. Aichinger, S. Puchner, M. Nelhiebel, T. Grasser, and H. Hutter, "Impact of hydrogen on recoverable and permanent damage following negative bias temperature stress," in *Proc.IRPS*, 2010, pp. 1063–1068.

[4] T. Aichinger, M. Nelhiebel, and T. Grasser, "Investigation of the NBTI mechanism by low temperature characterization of arbitrarily stressed PMOS devices," *IEEE Trans.Elect.Dev.*, 2010, submitted for publication.

[5] T. Grasser, T. Aichinger, H. Reisinger, J. Franco, P.-J. Wagner, M. Nelhiebel, C. Ortolland, and B. Kaczer, "The 'permanent' component of NBTI: Creation, composition, and annealing," in *Proc.IIRW*, 2010, submitted for publication.

[6] T. Grasser, B. Kaczer, W. Gös, H. Reisinger, T. Aichinger, P. Hehenberger, P.-J. Wagner, F. Schanovsky, J. Franco, P. Roussel, and M. Nelhiebel, "Recent advances in understanding the bias temperature instability," in *Proc.IEDM*, 2010, submitted for publication.

[7] G. Pobegen, T. Aichinger, M. Nelhiebel, and T. Grasser, "Dependence of the negative bias temperature instability on the gate oxide thickness," in *Proc.IRPS*, 2010, pp. 1073–1077.

[8] T. Aichinger, M. Nelhiebel, and T. Grasser, "Unambiguous identification of the NBTI recovery mechanism using ultra fast temperature changes," in *Proc.IRPS*, 2009, pp. 2–7.

[9] T. Aichinger, M. Nelhiebel, and T. Grasser, "A combined study of p- and n-channel MOS devices to investigate the energetic distribution of oxide traps after NBTI," *IEEE Trans.Elect.Dev.*, vol. 56, no. 12, pp. 3018–3026, 2009.

[10] T. Aichinger, M. Nelhiebel, S. Einspieler, and T. Grasser, "In-situ polyheater, a reliable tool for performing fast and defined temperature switches on chip," *IEEE Trans.Dev.Mater.Rel.*, vol. 76, no. 7, pp. 1–7, 2009.

[11] T. Grasser, B. Kaczer, W. Gös, T. Aichinger, P. Hehenberger, and M. Nelhiebel, "A two-stage model for negative bias temperature instability," in *Proc.IRPS*, 2009, pp. 33–44.

[12] T. Grasser, B. Kaczer, W. Gös, T. Aichinger, P. Hehenberger, and M. Nelhiebel, "Understanding negative bias temperature instability in the context of hole trapping," *Microelectron.Eng.*, vol. 86, no. 7-9, pp. 1876–1882, 2009.

[13] P. Hehenberger, T. Aichinger, T. Grasser, W. Gös, O. Triebl, B. Kaczer, and M. Nelhiebel, "Do NBTI-induced interface states show fast recovery? A study using a corrected on-the-fly charge-pumping measurement technique," in *Proc.IRPS*, 2009, pp. 1033–1038.

[14] T. Grasser, H. Reisinger, W. Gös, T. Aichinger, P. Hehenberger, P.-J. Wagner, M. Nelhiebel, J. Franco, and B. Kaczer, "Switching oxide traps as the missing link between negative bias temperature instability and random telegraph signal," in *Proc.IEDM*, 2009, pp. 1–4.

[15] P.-J. Wagner, T. Aichinger, T. Grasser, M. Nelhiebel, and K.J. Vandamme, "Possible correlation between flicker noise and bias temperature stress," in *Proc.ICNF*, 2009, pp. 621–624.

[16] T. Aichinger, M. Nelhiebel, and T. Grasser, "On the temperature dependence of NBTI recovery," in *Proc.ESREF*, 2008, pp. 1178–1184.

[17] T. Aichinger, M. Nelhiebel, and T. Grasser, "On the temperature dependence of NBTI recovery," *Microelectron.Reliab.*, vol. 48, no. 3, pp. 1178–1184, 2008.

[18] T. Aichinger and M. Nelhiebel, "Advanced energetic and lateral sensitive charge pumping profiling methods for MOSFET device characterization, analytical discussion and case studies," *IEEE Trans.Dev.Mater.Rel.*, vol. 8, no. 8-9, pp. 509–518, 2008.

[19] T. Grasser, B. Kaczer, T. Aichinger, W. Gös, and M. Nelhiebel, "Defect creation stimulated by thermally activated hole trapping as the driving force behind negative bias temperature instability in $SiO_2$, SiON, and high–$\kappa$ gate stacks," in *Proc.IIRW*, 2008, pp. 91–95.

[20] T. Aichinger and M. Nelhiebel, "Charge pumping revisited, the benefits of an optimized constant base level charge pumping technique for MOSFET analysis," in *Proc.IIRW*, 2007, pp. 63–69.

# Acronyms

| | |
|---|---|
| APC | anomalous positive charge |
| BEOL | back-end of line |
| BTI | bias temperature instability |
| BTS | bias temperature stress |
| CP | charge pumping |
| CV | capacitance voltage |
| DLTS | deep level transient spectroscopy |
| DOS | density of states |
| DUT | device under test |
| EPR | electron paramagnetic resonance |
| ESR | electron spin resonance |
| fWLR | fast wafer level reliability |

| | |
|---|---|
| GOX | gate oxide |
| HCI | hot carrier injection |
| HDL | Harry Diamond Laboratory |
| HF | hydrofluoric |
| HV | high voltage |
| LPCVD | low pressure chemical vapor deposition |
| MOS | metal oxide semiconductor |
| MOSFET | metal oxide semiconductor field effect transistor |
| MSM | measure/stress/measure |
| NBTI | negative bias temperature instability |
| NBTS | negative bias temperature stress |
| NMOS | n-channel metal oxide semiconductor |
| NO | nitrided oxide |
| OFIT | on-the-fly interface trapping |
| OTF | on-the-fly |
| PBTI | positive bias temperature instability |
| PECVD | plasma enhanced chemical vapor deposition |
| PMD | post metal dielectric |
| PMOS | p-channel metal oxide semiconductor |
| PNO | plasma nitrided oxide |

| | |
|---|---|
| RONO | re-oxidized nitroxide |
| RTNO | rapid thermal oxynitridation |
| SDR | spin dependent recombination |
| Si | silicon |
| $Si_3N_4$ | silicon nitride |
| $SiO_2$ | silicon dioxide |
| SiON | silicon oxynitride |
| SNIT | silicon nitride |
| SRH | Shockley Read Hall |
| TDDB | time dependent dielectric breakdown |
| TNO | thermally nitrided oxide |
| TOFSIMS | time of flight secondary ion mass spectrometry |

# Symbols

| | |
|---|---|
| $A_G^{eff}$ | effective gate area during charge pumping in $cm^{-2}$ |
| $C_{OX}$ | oxide capacitance in $F/cm^2$ |
| $D_{it}$ | density of interface traps in $eV^{-1}$ $cm^{-2}$ |
| $E_C$ | conduction band edge in eV |
| $E_{FP}$ | Fermi level within the polysilicon gate in eV |
| $E_{FS}$ | Fermi level within the silicon substrate in eV |
| $E_F$ | Fermi level in eV |
| $E_F^{VTH}$ | Fermi level position at the threshold voltage in eV |
| $E_G$ | silicon bandgap in eV |
| $E_{OX,R}$ | oxide field during recovery in V/cm |
| $E_{OX,S}$ | oxide field during stress in V/cm |
| $E_{OX}$ | oxide field in V/cm |
| $E_V$ | valence band edge in eV |
| $E_{amph}$ | amphotheric transition level in eV |
| $E_{dm}$ | medium dissociation energy in eV |
| $E_{em}^e$ | electron emission boundary in eV |
| $E_{em}^h$ | hole emission boundary in eV |
| $E_i$ | intrinsic Fermi level in eV |

| | |
|---|---|
| $I_{CP}^{MSM}$ | charge pumping current measured during CP MSM in A |
| $I_{CP}^{OFIT}$ | charge pumping current measured during OFIT in A |
| $I_{CP}^{max}$ | maximum charge pumping current in A |
| $I_{D,lin}$ | linear drain current in A |
| $I_{D,sat}$ | saturation drain current in A |
| $I_D$ | drain current in A |
| $I_D^{MSM}$ | drain current measured during MSM in A |
| $I_D^{OTF}$ | drain current measured during OTF in A |
| $N_A$ | acceptor doping density in $cm^{-3}$ |
| $N_{CP}^{max}$ | total number of pumped charges per area in $cm^{-2}$ |
| $N_C$ | effective density of states in the conduction band in $cm^{-3}$ |
| $N_D$ | donor doping density in $cm^{-3}$ |
| $N_V$ | effective density of states in the valence band in $cm^{-3}$ |
| $N_{it}$ | number of interface traps in $cm^{-2}$ |
| $P_{PH}$ | polyheater power in W |
| $Q_B$ | bulk charge density $C/cm^3$ |
| $Q_{CP}^{max}$ | pumped charge per area in $C/cm^{-2}$ |
| $Q_{OX}^{perm}$ | locked-in oxide charge $C/cm^2$ |
| $Q_{OX}^{rec}$ | recoverable oxide charge $C/cm^2$ |
| $Q_e$ | electron charge density $C/cm^3$ |
| $Q_h$ | hole charge density $C/cm^3$ |
| $Q_{it}^{perm}$ | permanent interface charge $C/cm^2$ |
| $R_{DV}^{TH}$ | thermal resistivity of the device in K/W |
| $R_{PH}^{TH}$ | thermal resistivity of the polyheater in K/W |
| $R_{ON}$ | on resistance of the device in Ω |
| $R_{PH}$ | resistance of the polyheater in Ω |

| | |
|---|---|
| $T_{DV}$ | device temperature in K |
| $T_{PH}$ | polyheater temperature in K |
| $T_R$ | recovery temperature in K |
| $T_S$ | stress temperature in K |
| $V_{DSS}$ | saturation drain voltage in V |
| $V_{DS}$ | drain voltage in V |
| $V_{FB}$ | flat band voltage in V |
| $V_{FB}^{CP}$ | charge pumping threshold voltage in V |
| $V_{GB}$ | base level of the gate pulse in V |
| $V_{GH}$ | high level of the gate pulse in V |
| $V_{GH}^{R}$ | high level of the gate pulse during CP MSM in V |
| $V_{GH}^{S}$ | high level of the gate pulse during OFIT in V |
| $V_{GR}$ | gate voltage during read-out in V |
| $V_{GS}$ | gate voltage during stress in V |
| $V_{OX}$ | voltage drop across the gate oxide in V |
| $V_{Poly}$ | voltage drop within the poly gate junction in V |
| $V_{TH}$ | threshold voltage in V |
| $V_{TH}^{CP}$ | charge pumping threshold voltage in V |
| $\Delta E_B$ | general thermodynamical barrier in eV |
| $\Delta E_{CP}$ | active energy range during charge pumping in eV |
| $\Delta E_Q$ | charged energy range in thermal equilibrium in eV |
| $\Delta I_{CP}$ | increase of the charge pumping current in A |
| $\Delta I_{CP}^{MSM}$ | increase of the charge pumping current measured during CP MSM in A |
| $\Delta I_{CP}^{OFIT}$ | increase of the charge pumping current measured during OFIT in A |
| $\Delta I_{CP}^{max}$ | increase of the maximum charge pumping current in A |
| $\Delta I_D$ | degradation of the drain current in A |

| | |
|---|---|
| $\Delta I_{\mathrm{D}}^{\mathrm{OTF}}$ | stress induced increase of the drain current measured during OTF in A |
| $\Delta I_{\mathrm{D}}^{\mu}$ | degradation of the drain current caused by mobility degradation in A |
| $\Delta I_{\mathrm{D}}^{\mathrm{q}}$ | degradation of the drain current caused by defect charges in A |
| $\Delta N_{\mathrm{it}}$ | increase of the number of interface traps $\mathrm{cm}^{-2}$ |
| $\Delta Q_{\mathrm{CP}}^{\mathrm{max}}$ | increase of the pumped charge per area in $\mathrm{C/cm}^{-2}$ |
| $\Delta V_{\mathrm{G,pulse}}$ | gate pulse amplitude in V |
| $\Delta V_{\mathrm{TH}}$ | threshold voltage shift in V |
| $\Delta V_{\mathrm{TH}}^{\mathrm{MSM}}$ | threshold voltage shift measured during MSM in V |
| $\Delta V_{\mathrm{TH}}^{\mathrm{OTF}}$ | threshold voltage shift measured during on-the-fly in V |
| $\Delta V_{\mathrm{TH}}^{\mathrm{R}}$ | threshold voltage shift measured during recovery in V |
| $\Delta V_{\mathrm{TH}}^{\mu}$ | threshold voltage shift caused by mobility degradation in V |
| $\Delta V_{\mathrm{TH}}^{\mathrm{it}}$ | interface state dependent threshold voltage shift in V |
| $\Delta V_{\mathrm{TH}}^{\mathrm{q}}$ | threshold voltage shift caused by defect charge in V |
| $\Delta \mu_{\mathrm{eff}}$ | change in the effective carrier mobility in $\mathrm{cm}^2/\mathrm{Vs}$ |
| $\gamma_{\mathrm{R}}$ | electric field factor during recovery in $[\mathrm{MV/cm}]^{-1}$ |
| $\gamma_{\mathrm{S}}$ | electric field factor during stress in $[\mathrm{MV/cm}]^{-1}$ |
| $\kappa$ | charge pumping weight factor |
| $\mu_{\mathrm{eff}}$ | effective carrier mobility in $\mathrm{cm}^2/\mathrm{Vs}$ |
| $\nu_{\mathrm{th,e}}$ | thermal drift velocity of electrons in cm/s |
| $\nu_{\mathrm{th,h}}$ | thermal drift velocity of holes in cm/s |
| $\overline{D_{\mathrm{it}}}$ | average density of interface traps in $\mathrm{eV}^{-1}\mathrm{cm}^{-2}$ |
| $\overline{\nu_{\mathrm{th}}}$ | average thermal drift velocity in cm/s |
| $\overline{\sigma}$ | average capture cross section in $\mathrm{cm}^2$ |
| $\psi_{\mathrm{B}}$ | bulk potential in V |
| $\psi_{\mathrm{P}}$ | poly surface potential/band bending in V |
| $\psi_{\mathrm{S}}$ | substrate surface potential/band bending in V |

| | |
|---|---|
| $\sigma_e$ | capture cross section for electrons in $cm^2$ |
| $\sigma_h$ | capture cross section for holes in $cm^2$ |
| $f$ | frequency of the gate pulse in Hz |
| $k_B$ | Boltzmann constant in eV/K |
| $m_{eff}$ | effective mass in kg |
| $n_{CP}^{STR}$ | charge pumping power-law factor measured during stress |
| $n_i$ | intrinsic carrier density in $cm^{-3}$ |
| $n_{tot}^{OTF}$ | power-law factor measured during OTF |
| $t_D$ | cooling delay time for degradation quenching in s |
| $t_R$ | recovery time in s |
| $t_S$ | stress time in s |
| $t_S^{eff}$ | effective stress time in s |
| $t_W$ | pulse width of the gate pulse in s |
| $t_f$ | fall time of the gate pulse in s |
| $t_h$ | high time of the gate pulse in s |
| $t_l$ | low time of the gate pulse in s |
| $t_r$ | rise time of the gate pulse in s |

# Curriculum vitæ

## MAG. THOMAS AICHINGER

### General information

| | |
|---|---|
| DATE OF BIRTH | May 28th 1983 |
| MARRIAGE STATUS | Unmarried |
| NATIONALITY | Austrian |
| ADDRESS | Völkendorferstrasse 33<br>9500 Villach<br>Austria |
| E-MAIL | thomas.aichinger@k-ai.at, aichinger.thomas@gmx.at |

### Education

| | | |
|---|---|---|
| 2007–2010 | PhD thesis | TU Vienna |
| 05/2007 | University graduation (MSc. Physics) | KF University Graz |
| 2002–2007 | Studies of experimental physics | KF University Graz |
| 07/2001 | High school graduation (A Levels) | Perau Gymnasium Villach |
| 1998–2001 | High school | Perau Gymnasium Villach |

## Jobs

| | | |
|---|---|---|
| 2007–2010 | Industry related PhD thesis | KAI Villach |
| 2006–2007 | Industry related master thesis | Infineon Austria |
| 2002–2006 | Professional summer internships | Infineon Austria |
| 2001–2002 | Military service | Rohrkaserne Villach |
| 1997–2001 | Pupil summer internships | miscellaneous |

## Skills

| | |
|---|---|
| LANGUAGE | German (native), English |
| TECHNICAL | Handling & programming of bench tester and parameter analyzer |
| OFFICE | Microsoft office, Latex |

## Awards

| | |
|---|---|
| BEST PAPER AWARD | ESREF 2008 |