A Study of NBTI and Short-Term Threshold Hysteresis of Thin Nitrided and Thick Non-Nitrided Oxides

Hans Reisinger, Rolf-Peter Vollertsen, Paul-Jürgen Wagner, Thomas Huttner, Andreas Martin, *Senior Member, IEEE*, Stefano Aresu, Wolfgang Gustin, Tibor Grasser, *Senior Member, IEEE*, and Christian Schlünder

***Abstract*—Negative bias temperature instability (NBTI) degradation and recovery have been investigated for 7–50-nm non-nitrided oxides and compared to thin 1.8- and 2.2-nm nitrided oxides from a dual work function technology. A wide regime of stress fields from 2.5 to 10 MV/cm has been covered. Thermal activation has been studied for temperatures from 25 °C to 200 °C. The NBTI effect for the nitrided oxide is larger than for non-nitrided oxides. The percentage of threshold shift ΔV_{th} which is "lost" during a long measurement delay—which is the quantity leading to curved ΔV_{th} versus stress-time curves and to errors in extrapolated lifetimes—is about equal for nitrided or thick non-nitrided oxides. The fraction of recovered ΔV_{th} is strongly dependent on stress time but only weakly dependent on stress field. Recovery in thick oxides leads to exactly the same problems as for non-nitrided oxides, and clearly, a fast measurement method is needed. The effect of short-term threshold shifts has been studied for extremely short stress times down to 200 ns.**

***Index Terms*—MOSFET, negative bias temperature instability (NBTI), recovery, relaxation.**

I. Introduction

NEGATIVE bias temperature instability (NBTI) in pMOSFETs is one of the most critical device reliability issues. It is well known that the fast recovery phenomenon of NBTI leads to serious measurement and assessment problems for thin nitrided oxides [1]–[6]. Only a couple of studies deal with thickness dependence of NBTI [7], [8], and—to our knowledge—there is no publication investigating the importance of recovery for thick-oxide NBTI assessments. Thus, it is the purpose of this paper to compare the NBTI effect for thick oxides from 7 to 50 nm to thin plasma nitrided oxides (PNO) for a wide regime of electric fields from use condition around 3 MV/cm up to a high stress field of 10 MV/cm. This paper demonstrates, for the first time, that the impact of recovery in thick oxide devices is as critical as it is for thin PNO oxides.

Manuscript received December 22, 2008; revised March 13, 2009. First published April 21, 2009; current version published June 5, 2009.

H. Reisinger, R.-P. Vollertsen, T. Huttner, A. Martin, S. Aresu, W. Gustin, and C. Schlünder are with the Corporate Reliability Methodology Department, Infineon Technologies, 81739 München, Germany (e-mail: Hans.Reisinger@infineon.com; wolfgang.gustin@infineon.com; Christian.Schluender@infineon.com).

P.-J. Wagner and T. Grasser are with the Christian Doppler Laboratory for TCAD, Institute for Microelectronics, Technische Universität Wien, 1040 Wien, Austria (e-mail: pjwagner@iue.tuwien.ac.at; grasser@iue.tuwien.ac.at).

Digital Object Identifier 10.1109/TDMR.2009.2021389

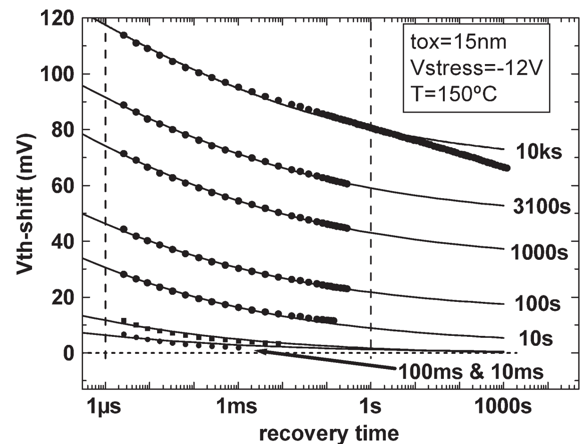


Fig. 1. Example of raw data as measured. Stress is interrupted at certain intervals (labels give total stress time). Each trace (dots) shows measured V_{th} as a function of the time after stress interruption (= recovery time). Lines are fitted using the universal relaxation and permanent components after (1), as described in [9].

II. Samples and Experiments

pMOSFETs with a width of 10 μm with three different non-nitrided oxides (7, 15, and 50 nm thick) with n^+-poly gates were used for this paper. Data are compared to 1.8- and 2.2-nm PNO oxide devices from a dual work function technology (boron-doped poly) as described in [5]. Threshold voltages V_{th}'s were extracted at a current criterion of 70 nA $* W/L$ in saturation at temperatures between 25 °C and 200 °C. Our fast measurement technique was used as described in [5]. The resolution and reproducibility of the measured ΔV_{th} at a 1-μs measuring delay has been pushed to a limit of ±0.03 mV. All results for the thick oxides were measured on the same wafer. The results for the thin nitrided 1.8- and 2.2-nm oxides were also gathered from a single wafer.

III. Experimental Results

An example of raw measured V_{th} data is shown in Fig. 1. The stress was interrupted 1–5 times per decade of stress time. For each interruption, recovery traces are measured (while the gate is at threshold voltage) for roughly a second. This way, it is possible to include the result from a conventional slow measurement with a 1-s measuring delay from the same measurement and the same sample. Any stress interruption was more than a

1530-4388/$25.00 © 2009 IEEE

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on November 16, 2009 at 04:03 from IEEE Xplore. Restrictions apply.

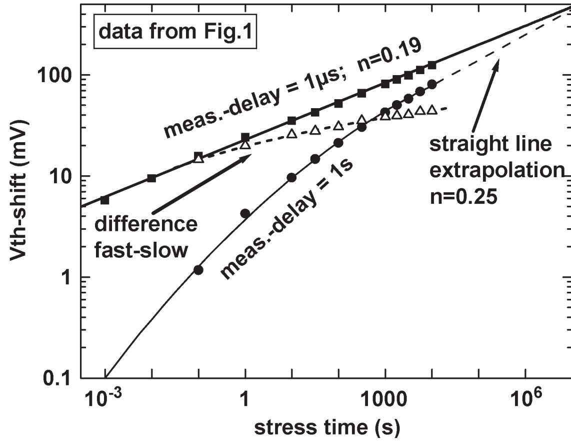


Fig. 2. ΔV_{th} as a function of stress time extracted from Fig. 1. A "fast" and a "slow" curve (for a measuring delay of 1 μs and of 1 s as for a conventional analyzer) is extracted along the vertical dashed lines in Fig. 1 (not all the measured traces are shown in Fig. 1). n denotes the power-law exponent (see Figs. 3 and 12).

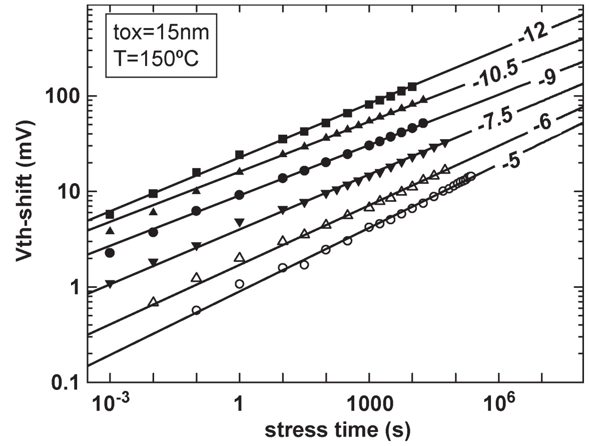


Fig. 3. ΔV_{th} versus stress time for 1-μs measuring delay. Curve labels denote the stress voltages. Measured points for stress time > 1 s are fitted to a power law (full straight lines). The curves follow a power law $\Delta V_{\mathrm{th}} = const * t^n$ for more than seven decades of time. The power-law exponents n are given in Fig. 12.

factor of 100 shorter than the following stress period. Therefore, the sample completely "forgets" the recovery caused by the interruption. The negligible effect of the interruption on ΔV_{th} versus stress time was experimentally verified.

The complete set of recovery traces for one stress run was fitted using the method described in [9]. All of the seven traces in Fig. 1 are fitted by a single global recovery law

$$\Delta V_{\mathrm{th}_i}(\xi) = \frac{\Delta V_{\mathrm{th0}_i} - \Delta V_{\mathrm{th}\infty_i}}{1 + B\xi^{\beta}} + \Delta V_{\mathrm{th}\infty_i} \quad (1)$$

where $\xi = t_r/t_{\mathrm{stress}}$ is the normalized recovery time t_r. The first term in the sum of (1) is the recoverable part, and the second is the permanent part $\Delta V_{\mathrm{th}\infty}$ at $t_r = \infty$. The same "global" parameters B and ξ are valid for each of the seven traces (indexed i). In addition, each trace i contains two more parameters, i.e., ΔV_{th0} at $t_r = 0$ and the permanent part $\Delta V_{\mathrm{th}\infty}$ at $t_r = \infty$, making the total number of fit parameters in Fig. 1 $16 = 7 * 2 + 2$ or more general $2 \times i + 2$. The fit provides a nice tool to perfectly fit and smooth the measured data—except the very long term recovery behavior as shown in the uppermost trace of Fig. 1—and to extract $\Delta V_{\mathrm{th}}(t_r)$ curves at any desired measuring delay, as practiced in Figs. 1 and 2. In general, the recovery function (see full lines in Fig. 1) and the measured recovery traces for thin oxides have the property that the derivative $dV_{\mathrm{th}}/d(\log t_r)$ is monotonically decreasing with recovery time. For our thick oxides and long recovery times, the slopes of the experimental traces (cf. 10-ks trace for recovery times > 100 s in Fig. 1) get steeper with recovery time and do not obey the simple universal recovery law [see (1)], where the permanent part is assumed not to recover at all. Fig. 2 shows a "fast" $\Delta V_{\mathrm{th}}(t)$ curve with a 1-μs measuring delay and a "slow" one with a 1-s measuring delay as would result from a conventional measurement with a standard parameter analyzer. These fast and slow curves were extracted along the two dashed vertical lines in Fig. 1. As shown in Fig. 2, the difference between the fast and the slow curve is roughly constant with a weak stress-time dependence only. The "slow" $\Delta V_{\mathrm{th}}(t)$ curve shows power-law exponents $n > 0.25$ and the well-known bending of the curve. This bending has been often interpreted as a saturation in the past [10], [11]. As shown in Fig. 2, this "saturation" is just an artifact caused by recovery. The "fast" $\Delta V_{\mathrm{th}}(t)$ curves in Figs. 2 and 3 are perfectly linear (on the log–log scale), without any bending independent of the stress field and also independent of the oxide thickness for the non-nitrided oxides.

IV. Time Extrapolation of ΔV_{th} and Lifetime Prediction

For "slow" $\Delta V_{\mathrm{th}}(t)$ curves, like in Fig. 2, a very long stress time has to be spent before a meaningful extrapolation to longer times, up to the ΔV_{th} shift criterion, can be done [12]. In contrast, "fast" $\Delta V_{\mathrm{th}}(t)$ curves—as a consequence of the absence of any curvature—allow an extrapolation to longer times for a stress time which is 100 times shorter without any sacrifice, as can be seen in Fig. 2.

Fig. 4 shows an example for a lifetime extrapolation done using "fast" and "slow" $\Delta V_{\mathrm{th}}(t)$ measurements for measuring delays of 1 μs and 1 s, respectively. The "fast" and "slow" lifetimes were taken from extrapolations of the $\Delta V_{\mathrm{th}}(t)$ versus time curves as shown in the example (Fig. 2). It can be seen that the effect of recovery can result in an overestimation of lifetime as well as an underestimation. Overestimation occurs for high stress fields for which the drift criterion is reached during stress, but the measured ΔV_{th} is too low due to recovery. Underestimation occurs for tests at low stress fields for which only a fraction of the drift criterion is reached during stress and stress time is short. For example, if the actual stress time done in Fig. 2 were 100 s and the drift criterion were > 100 mV, then this extrapolation of the "slow" curve from 100 s would lead to a significant underestimation of lifetime. As shown in Fig. 4, an assessment with a slow measurement and tests done for high field (high acceleration and short stress times) will, in general, result in drastic underestimation of lifetime at use condition. A safe and favorable assessment technique is to do stress **at low fields** close to use condition (i.e., only slightly or not accelerated, for example, around -6 V in Fig. 4). The

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on November 16, 2009 at 04:03 from IEEE Xplore. Restrictions apply.

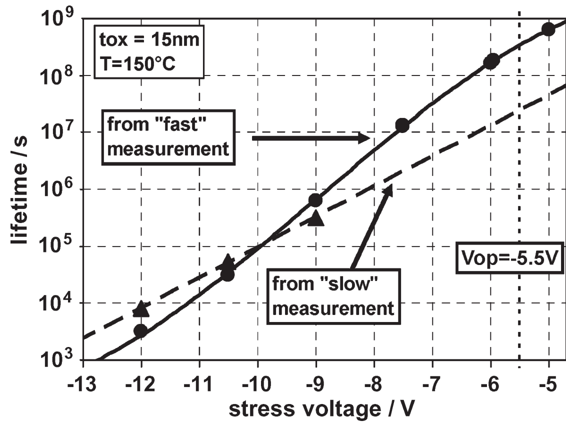


Fig. 4. Lifetime extrapolation from "fast" and "slow" measurements for a $\Delta V_{\mathrm{th}} = 100$ mV lifetime criterion. The extrapolation to operating voltage from the slow measurement from highly accelerated stress is significantly underestimating the lifetime. A most reliable assessment is done from low fields close to operation voltage with a fast measurement.

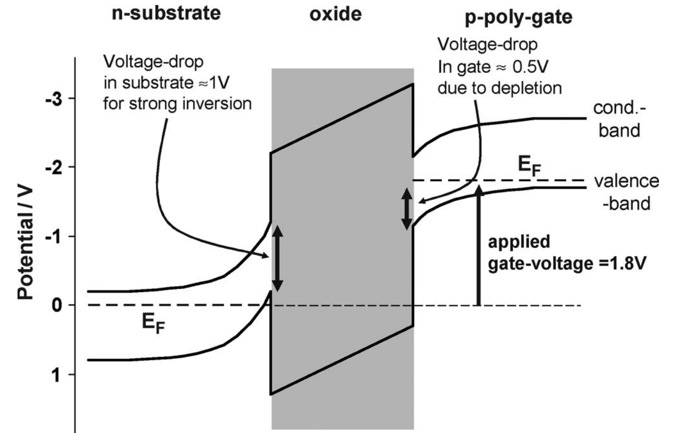


Fig. 5. Schematic band diagram of a pFET with a p-doped Si gate driven to p-inversion by applying a very high negative gate voltage corresponding to stress fields. To calculate the electric field in the oxide, the voltage drops in substrate and poly-gate have to be subtracted from the applied gate voltage.

straight line $\Delta V_{\mathrm{th}}(t)$ versus stress-time curve extracted by the fast measurement technique as shown in Fig. 3 will yield more reliable lifetimes than the ones from highly accelerated tests. Since the extrapolated straight line is independent of the stress time, this is true even for short stress times.

V. Comparison of Different Oxide Thicknesses

A. Basic Preconditions for Comparing Different Oxide Thicknesses

It is clear that a meaningful comparison of V_{th} drifts involving different thicknesses, different processes for oxides and gates, etc., can be done only on a basis of equal electric stress fields. For oxides with 1-nm thickness, the simple $E = (V_g - V_{\mathrm{FB}})/t_{\mathrm{ox}}$ estimation ($V_{\mathrm{FB}} =$ flatband voltage) leads to a field value which is too high by roughly a factor of two. The voltage applied between gate and Si-substrate does not cause a voltage drop across the gate oxide only, but there are also voltage drops on the order of 1 V in the substrate and on the order of 500 mV in the poly-Si gate. Both effects are difficult to correct for and are hardly ever mentioned nor discussed in the literature. This is why we put some emphasis on this topic, do a detailed discussion, and show a simple way to measure the electric field in the gate oxide. The voltage drops in the substrate and the gate are schematically shown in Fig. 5: In the Si-substrate, there is a voltage drop in the depletion layer and in the inversion layer. The voltage drop in the depletion is a function of substrate doping level and temperature and could be calculated. The voltage drop in the inversion layer is determined by low quantum numbers of the subbands even at temperatures above room temperature and thus needs simultaneous solving of Schrödinger and Poisson equations. Commercial device simulators could do such calculations. However, the accuracy in the calculated potentials at high fields is unsatisfying and cannot be expected to be better than ± 100 mV due to the inevitably inappropriate treatment of the many-body effects in the space-charge layer [13]. In a boron-doped poly gate, there is a significant depletion layer width for negative gate bias at high field. The concentration of electrically activated boron in a poly gate is, in general, much less than that in an n-doped gate. This leads to voltage drops in the boron-doped poly on the order of 0.5 V at 10 MV/cm as shown in studies of MNOS memories [14]. To our knowledge, there is no way of direct determination of the electrically activated boron in a thin oxide pMOSFET. Thus, there is no way to calculate the voltage drop in the p-poly-gate. Fortunately, there is a simple way to calculate the electric field in the oxide from a CV curve which can be easily measured given a large square pFET with sufficient capacitance available. For a 2-D capacitor (i.e., thickness $\ll$ lateral dimension, as still given for any MOSFET), the relation between electric field in the dielectric and the charge density Q/A on the capacitor plates with area A is given by the Maxwell equation (Gauss' law)

$$E = \frac{Q}{A \times \varepsilon_0 \varepsilon_r} \tag{2}$$

where ε_r is the relative static permittivity of the dielectric. It is 3.9 for SiO_2, and it is only ≈5% higher if the oxide contains 10% nitrogen. The CV curve directly measures the quantity dQ/dV_g as a function of the gate voltage V_g. Therefore, the electric field E can just be calculated by measuring and integrating the CV curve as shown in Fig. 6. Still to be determined is the flatband voltage, which is the gate voltage where the electric field in the oxide and at both Si interfaces is zero. In the vicinity of flatband condition, there are no quantum mechanical effects in the Si space-charge layer. Thus, the classical approximation to determine the flatband condition of the CV curve as given in [15] is sufficient.

In conclusion, the calculation of the electric field from the CV curve implicitly contains all QM effects, gate depletion, and also work function differences. As can be seen from the data in Fig. 6, corrections have to be added to the 2.2-nm physical oxide thickness by QM and p^+-poly depletion of ≈0.5 nm and ≈0.7 nm, respectively (both equivalent oxide thickness). The correction is independent of the thickness of the dielectric, so the same correction of 1.2 nm would be valid for a 1.2-nm oxide, making the error of the simple $E = V_g/t_{\mathrm{ox}}$ approximation a factor of two.

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on November 16, 2009 at 04:03 from IEEE Xplore. Restrictions apply.

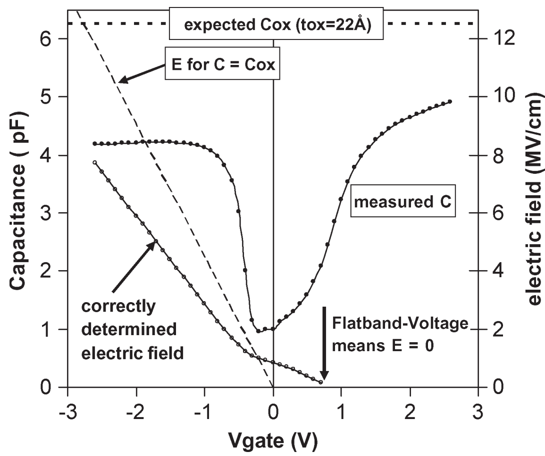


Fig. 6. Measured CV curve (left vert. axis), simple approximation $E = V_g/t_{\rm ox}$, and correctly determined electric field E for the 2.2-nm oxide (gate area = 400 μm^2). E is determined from the CV curve and Maxwell equation div$E = \rho$. For MOSFETs with nitrided oxides, the electric field is significantly reduced by depletion in the p$^+$-doped gate.

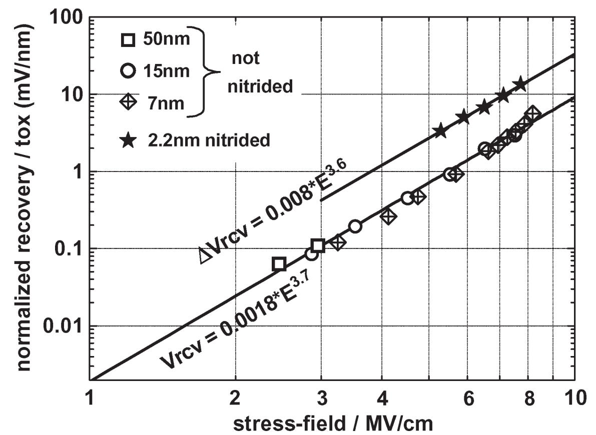


Fig. 8. Normalized **recovery** $\Delta V_{\rm rcv} = [\Delta V_{\rm th}(1\ \mu{\rm s}) - \Delta V_{\rm th}(1\ {\rm s})]/t_{\rm ox}$ as a function of stress field after a stress time of **10 ks**. Recovery in $\Delta V_{\rm th}$ was taken between 1 μs and 1 s. To save space, the equations of the straight-line fits given in Figs. 7 and 8 do not contain a prefactor with correct dimensions. When E is entered in megavolts per centimeter, then the result of $\Delta V_{\rm th}$ and $\Delta V_{\rm rcv}$ is in millivolts per nanometer. This result has to be multiplied by the oxide thickness in nanometers to get the real values for $\Delta V_{\rm th}$ and $\Delta V_{\rm rcv}$.

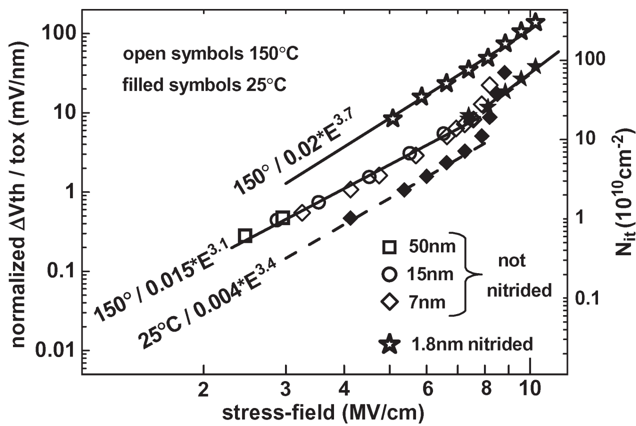


Fig. 7. Normalized $V_{\rm th}$ shift $\Delta V_{\rm th}/t_{\rm ox}$ as a function of stress field after a stress time of **10 ks** fitted to a power law in stress field. The right-hand scale shows the corresponding interface charge. Open and filled symbols are for temperatures of 150 °C and 25 °C, respectively. Labels give parameters of the power-law fits.

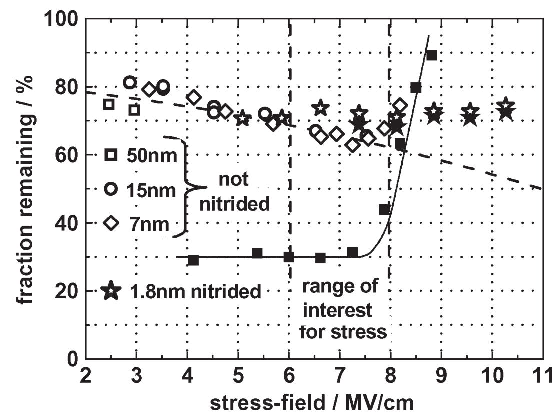


Fig. 9. Percentage of $\Delta V_{\rm th}$ which is measured with a conventional 1-s measuring delay taken at a stress time of **10 ks**. The $\Delta V_{\rm th}$ measured at a 1-μs delay was set as 100%. The 50-nm samples could be stressed at low field only due to a ±16-V output voltage limit of our experimental setup. Lines are just meant to guide the eye. Open symbols denote 150-°C data, and filled symbols (for 1.8-nm nitrided and 7-nm non-nitrided only) denote 25-°C data.

B. Results of Comparing Different Oxide Thicknesses

Fig. 7 shows the magnitude of the NBTI $V_{\rm th}$ shift for all investigated oxides. For a given amount of interface charge, the resulting $\Delta V_{\rm th}$ is proportional to the oxide thickness; thus, the vertical axes are normalized to allow a comparison of the amount of generated interface charge. Since the thick-oxide data cover a wide regime in electric field, we are able to clearly distinguish a power-law behavior from a behavior exponential in electric field. An exponential function does not fit the data. The field dependence of all thick-oxide NBTI shifts $\Delta V_{\rm th}(E)$ follows the same power law with a slightly lower exponent than for the thin PNO oxide. The PNO power-law exponent is 3.7, comparable to the value of 4.1 found in [8] (for much higher fields). Our thin PNO shows a $\Delta V_{\rm th}$ enhanced by a factor of ≈3 compared to the non-nitrided oxides. This may be attributed to a larger number and a different nature of interface states due to nitridation [16]. The **absolute** normalized amount of recovery is shown in Fig. 8. It follows a similar power law as the $\Delta V_{\rm th}(E)$ curves in Fig. 7. The **relative** amount of recovery and remaining $\Delta V_{\rm th}$ after a stress time of 10 ks is shown in Fig. 9 for temperatures of 150 °C and 25 °C. For 150 °C, the relative recovery is almost independent of the sample type. It is around 30% after a measuring delay of 1 s. $\Delta V_{\rm th}$ measured at a 1-μs measuring delay is taken as 100%. For 150 °C, the recovery is only slightly dependent on electric field and is **the same** for nitrided and non-nitrided oxides. For 25 °C, Fig. 9 shows that the recovery is significantly different for thick and thin oxides. More temperature-dependent recovery data will be given in Fig. 11. Similar data as shown in Fig. 9 for a 10-ks stress are shown in Fig. 10 for stress times of 0.1 s–100 ks with the same results that, for 150 °C, relative recovery is approximately independent of the technology (nitrided or not), independent of thickness, and only weakly dependent of electric field. Thus, for all oxides at **high** temperature and for the nitrided oxides at **high** and **low** temperatures, there is **only one** parameter that the remaining fraction depends on, i.e., the stress time. Not counted as trivial is the dependence on the measuring delay. In contrast for the

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on November 16, 2009 at 04:03 from IEEE Xplore. Restrictions apply.

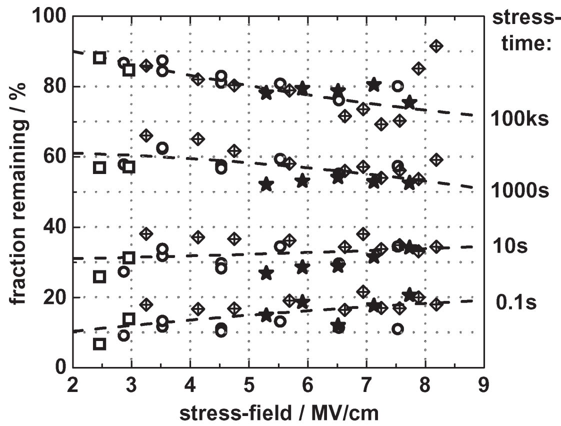


Fig. 10. Fraction remaining after 1-s recovery as a function of stress time. Data shown are for non-nitrided oxides at 150 °C and 2.2-nm nitrided oxide at 125 °C. Same symbols as in Fig. 9 used.

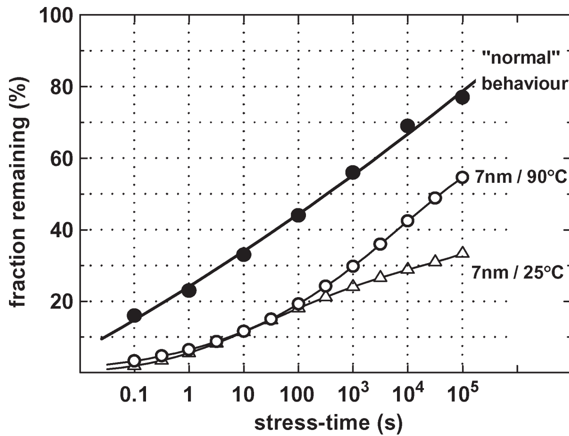


Fig. 11. Values of remaining fraction at 1-s measuring delay for a stress field of 6 MV/cm. Dots represent averages over all samples with "normal" recovery (1.8- and 2.2-nm nitrided from 150 °C to 25 °C and all non-nitrided at 150 °C). Open symbols are from the 7-nm oxide at 90 °C and 25 °C (15 and 50 nm not measured at 25 °C).

thick non-nitrided oxides, the recovery is strikingly higher at low temperature (25 °C) than at 150 °C. This behavior is also seen in Fig. 11. Fig. 11 extracts the remaining fractions of Fig. 10 for a stress field of 6 MV/cm which is a value most common for stress. Dots represent the remaining fraction given by the average over all data with "normal" behavior (i.e., dashed lines in Fig. 10) as a function of stress time. As expected, the remaining fraction increases with stress time but approaches 100% only after 10^7 s (four months). Also shown in Fig. 11 is the "anomalous" recovery behavior of a non-nitrided oxide at lower temperature. Low-temperature behavior will be discussed and investigated in more detail in Section VI. The most important conclusion to be drawn from the data in Figs. 9–11 is that—while NBTI in PNO is a factor of three stronger than in the non-nitrided oxides—at elevated temperature, there is no difference in the relative fraction of recovery between thin PNO oxides and thick non-nitrided oxides. Relative recovery is **the** quantity which is bending the ΔV_{th} versus stress-time curves, thus causing all the trouble with NBTI extrapolation and assessment. The impact of recovery on NBTI assessment is as severe as for "old technologies," for thick non-nitrided oxides as well as for modern thin nitrided oxides. There is also evidence for

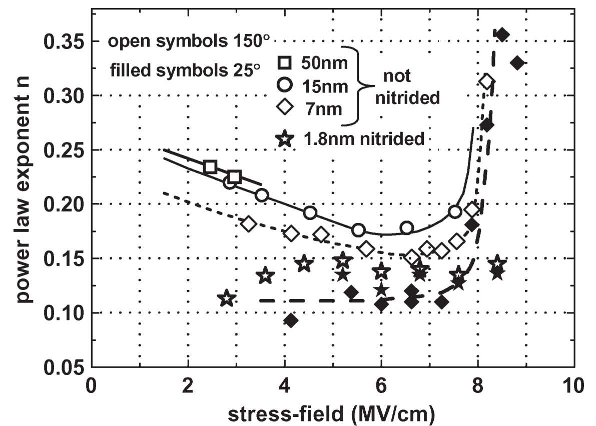


Fig. 12. Power-law exponents of $\Delta V_{\mathrm{th}}(t)$ curves (cf. Figs. 2 and 3) as a function of stress field for all investigated samples. Lines are meant to guide the eye. n of the non-nitrided 7-nm oxide is significantly lower at 25 °C than at 150 °C.

another conclusion: Both 1.8- and 2.2-nm oxides and thicker oxides obey a similar power law for degradation (cf. Fig. 7), and recovery is very similar (cf. Figs. 8–10). This would support the hypothesis that any diffusion of hydrogen into the oxide beyond 2 nm from the interface does not play a role or does not occur. Otherwise, the 1.8-nm data would look different from the thick-oxide data. Fig. 12 shows the power-law exponents n of the $\Delta V_{\mathrm{th}} \propto t^n$ fits from Fig. 3. For $T = 150$ °C, there is a slight but systematic increase of $\boldsymbol{n}$ with thickness. There is also a significantly lower n of the PNO oxide compared to the non-nitrided oxide. This difference might be due to a larger number of fast bulk traps in the PNO which are known to decrease n [2]. For the nitrided oxides at $T = 25$ °C, n is significantly lower at 25 °C than at 150 °C (see further discussion in Section VI).

For high fields above 7 MV/cm, there is a sharp increase in slope n for the thick oxides. This high-field "deviation" is also seen in the 7-nm data in Figs. 7–10, 12, and 15. This anomalous effect is **not due to NBTI**—which is an effect occurring without injected current—but due to the damage done by electrons tunneling from the gate to the anode and—for thick oxides with stress voltages > 5 V—having kinetic energies > 5 eV at the anode. This energy is dissipated in the oxide close to the anode (i.e., the Si interface) and leads to oxide wear-out and finally to dielectric breakdown. The oxide wear-out apparently creates permanent interface and bulk traps. An increase in n correlated with an increase in stress-induced gate leakage has also been seen in [17]. It should be noted that the increase in n is seen only at high fields > 7 MV/cm and not from the start of stress but first after > 1000 s when oxide wear-out gets effective (cf. Fig. 10). Another hint suggesting that oxide wear-out creates another species of traps is that these traps correlated with the increase in n in Fig. 12 also show less recovery effect than "real" NBTI as shown in Figs. 9 and 10. This effect is pronounced at low temperature. To further support the conclusion that oxide wear-out dominates the anomalous high-field degradation, the correlation of NBTI degradation to the measured Fowler–Nordheim current is shown in Fig. 13. The anomalous degradation (i.e., the deviation from power law) starts at fields > 7.5 MV/cm. This corresponds to a charge fluence of roughly 1 C/cm^2 as shown in Fig. 13. Dielectric breakdown typically

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on November 16, 2009 at 04:03 from IEEE Xplore. Restrictions apply.

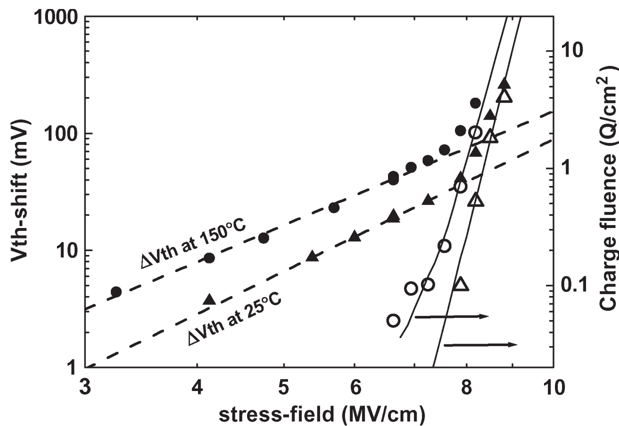


Fig. 13. High field "anomaly" of the 7-nm non-nitrided oxide at $T = 25\ ^\circ\mathrm{C}$ and $T = 150\ ^\circ\mathrm{C}$. Triangles denote 25 °C, and circles denote 150 °C. The figure shows a set of **4 curves** #1 to #4 for each temperature: $\#\mathbf{1} = \textbf{filled symbols}$ is the measured ΔV_{th} after a stress time of 10 ks. $\#\mathbf{2} = \textbf{dashed lines}$ is a power-law fit to the "normal" low field ΔV_{th} up to 7 MV/cm. $\#\mathbf{3} = \textbf{open symbols}$ is the calculated **deviation** in ΔV_{th} between power law and high field measured ΔV_{th} and thus is the deviation between full dots and dashed lines, i.e., between normal and anomalous, which is around zero at the low fields. $\#\mathbf{4} = \textbf{full lines}$ (is the only set which corresponds to right scale) is the measured charge fluence due to Fowler–Nordheim current integrated over the 10-ks stress time. The right-hand scale covers three decades exactly like the left-hand scale and is adjusted to illustrate the overlay of curves #3 and #4. The **exact** congruence of the ΔV_{th} deviation and the charge fluence for 25 °C and 150 °C leads to the conclusion that the charge fluence is the root cause of the deviation, i.e., the ΔV_{th} anomaly.

happens at 10 Q/cm^2. A perfect correlation of the anomalous part of degradation (open symbols in Fig. 13) with the Fowler-Nordheim fluence in field and temperature is shown in Fig. 13. As a conclusion, any NBTI assessment for thick oxides with charge fluences > 1 Q/cm^2 or fields > 7.5 MV/cm will cause serious extrapolation errors and must definitely be avoided. In contrast to the thick oxides ($\geq$ 7 nm), the stress voltages in the thin < 2.2-nm oxide are below 3 V, so below the $\mathrm{Si{-}SiO_2}$ band offset and thus in the direct tunneling regime with no energy dissipated in the oxide.

VI. Voltage Acceleration and Thermal Activation for Short Times

It is frequently speculated that short-time effects in NBTI—although small and not dominating the NBTI effect—have a significant influence on ΔV_{th} versus time curves and extrapolations. In the NBTI community, there is a faction favoring the idea that NBTI consists of a fast bulk trapping contribution and a reaction/diffusion contribution and another faction claiming that hole trapping and interface state generation are equally important. Arguments for the first idea are given in [18]. It is assumed that the "real" NBTI part is given by the reaction/diffusion theory and that the corresponding degradation ΔV_{th} versus time follows a $t^{1/6}$ power law. Subtracting this $t^{1/6}$ part from the experimental data, there remains a "fast" bulk trapping part with a low activation energy. Since the reaction/diffusion model completely fails to explain the long time constants involved in recovery, the basic and crucial assumption of the $t^{1/6}$ power law seems to be risky. Arguments supporting the hole trapping and interface state theory are given in [4]. An interesting observation in that context is that the hole trapping and interface state components could be coupled [19]. There, it is shown that all experimental ΔV_{th} versus time curves for a given sample type can be made to overlap by just using the right scaling factor for ΔV_{th}. This can be done **for all stress fields** and **temperatures**, which means that the short-term ΔV_{th}'s can also be described by a coupled NBTI mechanism. A drawback of the data presented in [19] is that they were taken with a slow "on-the-fly" technique. This technique defines an already degraded drain current taken at 1 ms after the beginning of stress as the unstressed $\Delta V_{\mathrm{th}} = 0$ level (i.e., at stress time = 0) and thus does not provide information about contributions faster than 1 ms.

In this paper, we decided not to make any assumption and just to do our best to separate "real" NBTI from fast trapping effects—if existent. Differences between NBTI and fast trapping would be seen in the following.

1) In the characteristic time constants of the effects: The longest tunneling time constants for a 2-nm oxide would be on the order of 1 s and decreasing by a factor of 10 for each 0.2-nm reduction in thickness (estimation from WKB approximation and the oxide barrier heights).
2) In the thermal activation of the processes, tunneling is expected to be nearly independent of temperature.
3) In the field dependence of the effects: For tunneling, the field acceleration depends on the spatial depths of a trap, but at least, the field dependence would most probably be different from "real" NBTI.
4) In the time dependence of the effect: "Real" NBTI would follow t^n and trapping $\log(t)$. It is difficult to distinguish between the two; the $\log(t)$ behavior must be seen for more than four decades in time.

As a consequence, in order to separate the effects, we did a series of measurements, varying the temperature between 25 °C and 200 °C and stress times from typically 10 ks down to extremely short stress times of 200 ns. For the accurate extraction of acceleration factors for field and temperature for the short stress times, we had to improve our resolution in ΔV_{th} to a value better than 0.1 mV. In order to compare our data to data in recent literature [18], [19], we chose to measure 1.8-nm nitrided oxide devices.

Fig. 14 shows an example for measured V_{th} shifts in response to very short stress pulses. All V_{th} shifts in Figs. 15 and 16 are taken from fast measurements like in Fig. 13 or Fig. 1 with a measurement delay of 1.0 μs after stress. Fig. 15 shows the result of stress field "sweeps" done at 25 °C and 150 °C for thin nitrided and thick non-nitrided oxides. Fig. 16 shows the result of a temperature "sweep" for high stress fields of 8 MV/cm for the nitrided oxide and 6.6 MV/cm for the non-nitrided oxide, yielding an activation energy E_A for each stress time. E_A values from Fig. 16 are shown in Fig. 17.

The results from Figs. 14–16 can be summarized as follows.

1) For the **1.8-nm nitrided oxide**, the power-law exponent of the voltage dependence is always close to four. There is no effect of temperature on the voltage dependence of NBTI. We would expect such an effect if fast bulk trapping would have a higher contribution to NBTI at 25 °C than at 150 °C. In addition, there is no effect of

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on November 16, 2009 at 04:03 from IEEE Xplore. Restrictions apply.

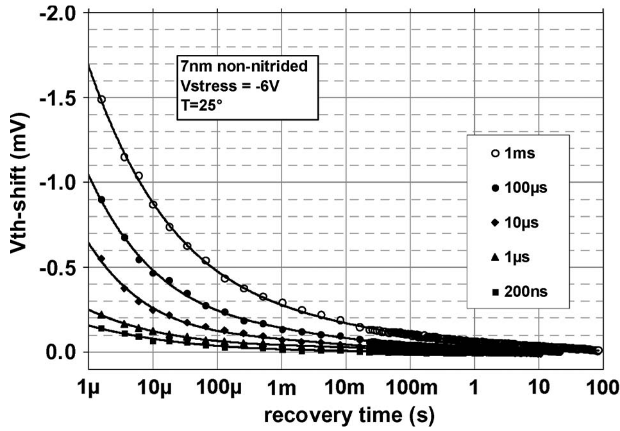


Fig. 14. Example for the measured fast high-resolution threshold hysteresis as a response to short stress pulses (pulsewidth given in legend). Each trace consists of typically 16 single measurements which have been averaged for noise reduction. After each stress pulse and measurement, a wait time is inserted until V_{th} is completely recovered. A reproducibility and resolution of ±0.03 mV can be achieved (cf. Fig. 15, lower traces).

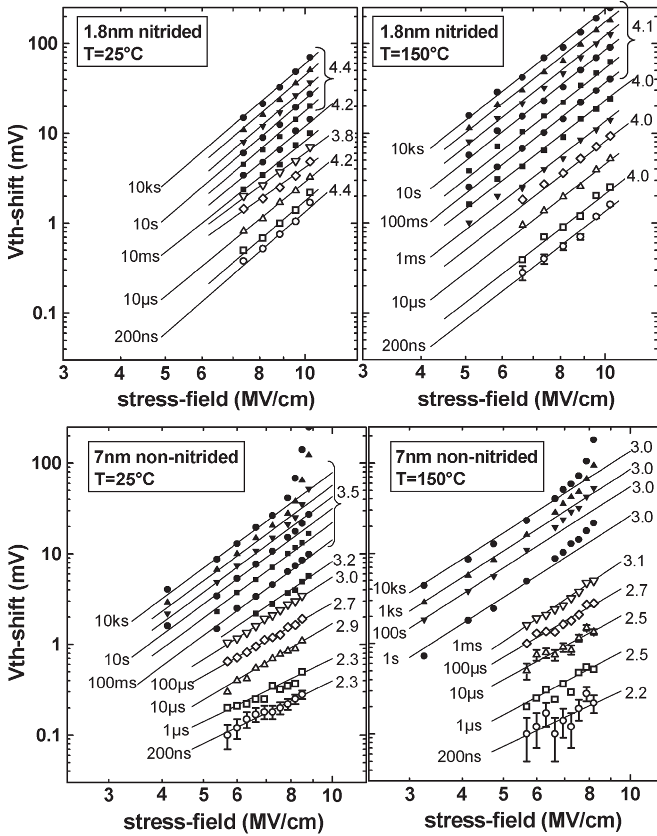
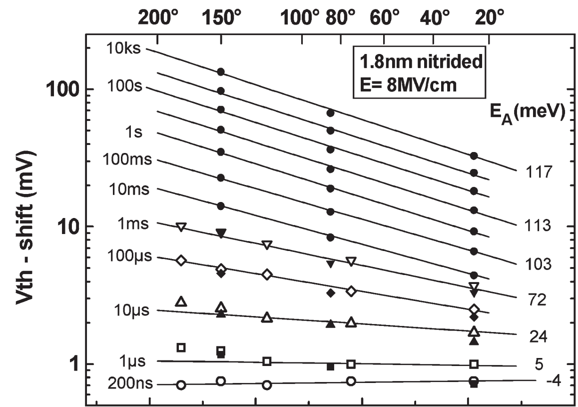


Fig. 15. Result of stress voltage "sweeps" at high and low temperatures. Parameter is the stress time. The experimental ΔV_{th}'s (dots) are plotted together with power-law fits to $\Delta V_{\mathrm{th}} = \mathrm{const}^* E^{\alpha}$. The power-law exponent is given as right-hand label for each curve. The stress times are given as left-hand labels. Open symbols for stress times ≤ 1 ms correspond to measurements like the one in Fig. 13. All these measurements were done using only a single FET. Filled symbols (> 1 ms) are from standard measurements using a new FET for each stress voltage.



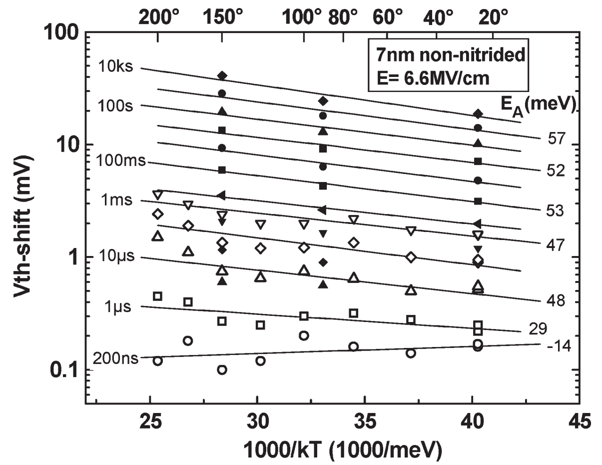


Fig. 16. Result of temperature "sweeps" at high stress field with stress time as parameter. Activation energies E_A are given as right-hand label for each curve. The stress times are given as left-hand labels. Open symbols for stress times ≤ 1 ms correspond to measurements like the one in Fig. 13. All these measurements were done using only a single FET. Filled symbols (> 1 ms) are from standard measurements using a new FET for each stress voltage.

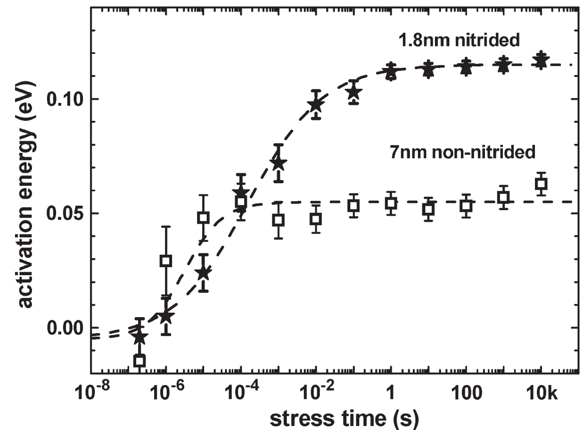


Fig. 17. Activation energies E_A as a function of stress time, extracted from the data in Fig. 16. A decrease of E_A is seen for short stress times which could be attributed to bulk trapping at short times. However, the corresponding V_{th} shifts are only 2 mV, hardly above the detection limit, and thus cannot have a significant influence on the ΔV_{th} versus stress-time curves.

the stress time on the voltage dependence down to stress times of 200 ns. Such an effect would be expected if bulk trapping would dominate at short times. The activation energy for long stress times is around 100 meV. As shown in Figs. 16 and 17, there is indeed a decrease in E_A for short stress times ≤ 30 μs as would be expected if tunneling into bulk traps would be dominating. As shown in Fig. 16, however, the effect is very small at levels of ΔV_{th} around 2 mV, which are hardly detectable.

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on November 16, 2009 at 04:03 from IEEE Xplore. Restrictions apply.

2) For the **7-nm non-nitrided oxide**, there actually is an effect of temperature and stress time on the power-law exponent of the voltage dependence. The magnitude of the effect is not really striking. We cannot associate the effect to a physical reason. The activation energy for long stress times is lower than for the nitrided oxide, around 60 meV for long stress times. Similar to the nitrided oxide (cf. Figs. 16 and 17), there is a decrease in E_A for short stress times which could be attributed to bulk trapping. The corresponding level of ΔV_{th}, which could be caused by bulk traps again, is **extremely** small below 1 mV (cf. Fig. 16).

The remaining open questions are as follows.

a) Why is the activation energy smaller for the non-nitrided oxide?
b) Why is there significantly more recovery (cf. Figs. 9 and 11) and a much lower power-law exponent for the time dependence for the non-nitrided thick oxide at low temperature (cf. Fig. 12)?

It should be mentioned that the thick oxides ≥ 7 nm are—due to the thick tunneling barriers—able to "provide" trapping and detrapping time constants $> 10^7$ s as occurring in NBTI recovery. A dominant contribution of bulk trapping for the thick oxides could explain the lower activation energies of the thick oxides, the differences in recovery, and the power-law exponent of the ΔV_{th} versus stress-time curves. Because the fraction in ΔV_{th}, which could be attributed to fast bulk trapping (with low E_A), is very small (cf. Figs. 15–17), we do not believe, however, that our measured data in sum do support such a hypothesis. To clarify the aforementioned questions, measurements at temperatures much lower than 25 °C should be done. Also an experiment using samples with a dense mesh of oxide thicknesses 1.6 nm to 5 nm in 0.2 nm steps, each thickness both with nitrided and with non-nitrided oxide, would certainly help. Unfortunately, such an experiment would currently exceed our capabilities.

VII. Summary

NBTI drift and recovery have been measured for non-nitrided oxides of 7–50 nm for stress field E from use condition 2.5 to 8 MV/cm. A comparison to thin nitrided oxides of 1.8 and 2.2 nm has been done. As expected, the NBTI shift in thick non-nitrided oxides is less than in thin nitrided oxides (cf. Fig. 7). However, the effect of recovery on measurements, extrapolations, and assessments is (even for low fields) equally significant in non-nitrided oxides (see Figs. 9 and 10)—which is in contrast to widespread belief. For ultrashort measuring delays, ΔV_{th} versus stress-time curves are straight lines (in log–log, power law) for all fields and all thicknesses (in contrast to the bent curves from slow measurements, cf. Fig. 2). Thus, for the thick oxides, the fastest and most reliable assessment technique is to do stress measurements and extrapolations at stress fields close to use condition with a fast measuring technique. The electric field dependence of the V_{th} shift is power law—clearly distinguishable from exponential behavior—with very similar exponents around 4 for PNO and slightly above 3 for non-nitrided oxides (cf. Fig. 7). For thick oxides, the stress field has to be kept below 7.5 MV/cm to prevent degradation due to Fowler–Nordheim current. Like for thin PNO oxides, the short and long time constants involved in the recovery **cannot** be explained by the reaction/diffusion model. The same degradation/recovery behavior for 1.8-nm oxide and thicker oxides at elevated temperature suggests that all NBTI processes take place in a thin interface layer < 1.8 nm. Major differences between thin nitride and thick non-nitrided oxides have been shown to exist at temperatures < 100 °C. Our experiments were not able to reveal the origin of these differences. Other physical characterization methods of oxide and interface (like ESR [16]) might be helpful for clarification.

References

[1] 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, Feb.–Apr. 2006.

[2] C. Shen, M.-F. Li, C. E. Foo, T. Yang, D. M. Huang, A. Yap, G. S. Samudra, and Y.-C. Yeo, "Characterization and physical origin of fast Vth transients in NBTI of pMOSFETs with SiON dielectric," in *IEDM Tech. Dig.*, 2006, pp. 333–336.

[3] S. Rangan, N. Mielke, and E. C. C. Yeh, "Universal recovery behavior of negative bias temperature instability," in *IEDM Tech. Dig.*, 2003, pp. 341–344.

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

[5] 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. Device Mater. Rel.*, vol. 7, no. 1, pp. 119–129, Mar. 2007.

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

[7] A. T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John, and S. Krishnan, "NBTI impact on transistor & circuit: Models, mechanisms & scaling effects," in *IEDM Tech. Dig.*, 2003, pp. 349–352.

[8] S. Tsujikawa, T. Mine, K. Watanabe, Y. Shimamoto, R. Tsuchiya, K. Ohnishi, T. Onai, J. Yugami, and S. Kimura, "Negative bias temperature instability of pMOSFETs with ultra-thin SiON gate dielectrics," in *Proc. IRPS*, 2003, pp. 183–188.

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

[10] S. Zafar, "A model for negative bias temperature instability in oxide and high κ pFETs," in *ICICDT*, Austin, TX, May 2007, pp. 1–5.

[11] 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.

[12] C. Schlünder, W. Heinrigs, W. Gustin, K. Goser, and H. Reisinger, "On the impact of NBTI recovery phenomenon on lifetime prediction of modern p-MOSFETs," in *IRW*, 2006, pp. 1–4.

[13] T. Ando, A. B. Fowler, and F. Stern, "Electronic properties of two-dimensional systems," *Rev. Mod. Phys.*, vol. 54, no. 2, pp. 437–672, Apr. 1982.

[14] H. Bachhofer, H. Reisinger, E. Bertagnolli, and H. Philipsborn, "Transient conduction in multidielectric silicon-oxide-nitride-oxide semiconductor structures," *J. Appl. Phys.*, vol. 89, no. 5, pp. 2791–2800, 2001.

[15] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. New York: Wiley, 2007, ch. 4.

[16] J. P. Campbell, P. M. Lenahan, A. T. Krishnan, and S. Krishnan, "NBTI: An atomic-scale defect perspective," in *Proc. IRPS*, 2006, pp. 442–447.

[17] S. Mahapatra and M. A. Alam, "A predictive reliability model for PMOS bias temperature degradation," in *IEDM Tech. Dig.*, 2002, pp. 505–508.

[18] S. Mahapatra, V. D. Maheta, A. E. Islam, and M. A. Alam, "Isolation of NBTI stress generated interface trap and hole-trapping components in PNO p-MOSFETs," *IEEE Trans. Electron Devices*, vol. 56, no. 2, pp. 236–242, Feb. 2009.

[19] T. Grasser, B. Kaczer, and W. Goes, "An energy-level perspective of bias temperature instability," in *Proc IRPS*, 2008, pp. 28–38.

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on November 16, 2009 at 04:03 from IEEE Xplore. Restrictions apply.

Hans Reisinger received the diploma and the Ph.D. degree in physics from the Technical University of Munich, Munich, Germany, in 1979 and 1983, respectively.

He joined the Siemens Semiconductor Department (now Infineon Technologies), Munich, in 1986, focusing mostly on the study of thin dielectrics and interfaces in DRAMs and NVMs. He is currently with the Corporate Reliability Methodology Department, Infineon Technologies, working on MOSFET degradation.

Rolf-Peter Vollertsen received the Ph.D. degree in material science engineering from the University of Erlangen, Erlangen, Germany, in 1987.

After joining Siemens AG, Munich, Germany, he focused on thin single and multilayer dielectric reliability. He worked for ten years in the DRAM Development Alliance at the IBM Microelectronics Division, Essex Junction, VT, where he collected experience in device and product reliability. He is currently with the Corporate Reliability Methodology Department, Infineon Technologies, Munich, working on fast wafer level reliability test and methodology development. His focus is on the reliability assessment of dielectrics and devices.

Paul-Jürgen Wagner was born in Vienna, Austria, in 1979. He received the degree of Diplomingenieur in electrical engineering in 2007 from the Technische Universität Wien, Vienna, where he is currently working toward the Ph.D. degree in the Institute for Microelectronics.

Thomas Huttner was born in Marktoberdorf, Germany, in 1967. He received the B.S. and Ph.D. degrees in physics from the Technical University Munich, Munich, Germany, in 1994 and 2000, respectively. The subject of his Ph.D. thesis was the physics of hot-carrier phenomena in silicon-on-insulator MOSFETs.

He is currently with the Infineon Technologies, Munich. He is also currently working on gate oxide reliability.

Andreas Martin (S'94–A'96–M'98–SM'05) received the M.Eng.Sc. degree in electrical and electronic engineering from the Technical University of Darmstadt, Darmstadt, Germany, in 1992.

In 1992, he joined the Silicon Technology Characterisation Group of the Tyndall Institute (formerly National Microelectronics Research Centre), Cork, Ireland, and worked on dielectric reliability assessment and reliability simulation. Since 1998, he has been with the Corporate Reliability Methodology Department, Infineon Technologies AG, Munich, Germany, in the field of fWLR Monitoring. He is involved in advanced and novel test structure design, development of new stress and measurement methodologies, and data analysis techniques on the topics dielectrics, plasma-induced damage, metallization, and devices. He manages PID process reliability qualifications and fWLR Monitoring projects covering processes from 0.25 μm down to 32 nm for all Infineon processes worldwide. He has published and coauthored numerous papers.

Mr. Martin has served for many years in the management and technical committees of the IEEE International Integrated Reliability Workshop, the IEEE International Reliability Physics Symposium, and the Workshop on Dielectrics in Microelectronics. He is an Alternate of the JEDEC subcommittee 14.2 and a member of the German Integrated Technology Group 8.5.6 on WLR and reliability simulations.

Stefano Aresu received the degree in electronic engineering from the University of Cagliari, Cagliari, Italy, in 2000 and the Ph.D. degree in physics in 2005.

He was with the Institute of Material Research in Microelectronics (division IMEC), where he worked on hot-carrier effect on power devices. He is currently with the Corporate Reliability Methodology Department, Infineon Technologies, Munich, Germany, where he is responsible for technology qualification of smart power devices for automotive and industrial application. His main research work is in the area of NBTI and hot-carrier effect.

Wolfgang Gustin received the diploma in physics from the University of Stuttgart, Stuttgart, Germany, in 1990, and the Ph.D. degree in physics from the Max-Planck-Institut, Stuttgart, in 1994.

From 1994 to 1998, he was with Philips and IBM, working on integration and unit process issues for logic and DRAM Technologies. In 1998, he joined the DRAM Development Group, Infineon Technologies, Munich, Germany, where he is currently the Manager of the device group of the Corporate Reliability Methodology Department.

Tibor Grasser (M'05–SM'05) was born in Vienna, Austria, in 1970. He received the Dipl.Ing. degree in communications engineering, the Ph.D. degree in technical sciences, and the Venia Docendi degree in microelectronics from the Technische Universität Wien (TU Wien), Wien, Austria, in 1995, 1999, and 2002, respectively.

He is currently an Associate Professor with the Institute for Microelectronics, TU Wien. Since 1997, he has headed the Minimos-NT development group, working on the successor of the highly successful MiniMOS program. He was a Visiting Research Engineer with Hitachi Ltd., Tokyo, Japan, and with the Alpha Development Group, Compaq Computer Corporation, Shrewsbury, MA. Since 2003, he has been the Head of the Christian Doppler Laboratory for TCAD in Microelectronics, an industry-funded research group embedded in the Institute for Microelectronics. He is the coauthor or author of over 200 articles in scientific books, journals, and conference proceedings, and is the Editor of a book on advanced device simulation. His current scientific interests include circuit and device simulation, device modeling, and reliability issues.

Dr. Grasser has been involved in the program committees of conferences such as SISPAD, IWCE, ESSDERC, IRPS, IIRW, and ISDRS. He was also a Chairman of SISPAD 2007.

Christian Schlünder received the Dipl.Ing. degree in electrical engineering, focusing on microelectronics, and the Dr.Ing. degree in engineering science from the University of Dortmund, Dortmund, Germany, in 1999 and 2006, respectively.

From 1998 to 1999, he was with a cooperative program between Siemens Corporate Research Laboratories, Munich, Germany, and the University of Dortmund in the field of characterization, modeling, and reliability of analog CMOS circuits. After the founding of Infineon Technologies, Munich, he acted for a short time as one of its Technical Consultants until he finally joined as a member of the Corporate Research Department, where he was active in the research on hot-carrier stress in mixed-signal applications. Since 2000, he has been with the Corporate Reliability Methodology Department, Infineon Technologies, where he manages process qualification projects for various state-of-the-art CMOS technologies and evaluates the reliability of innovative technologies like SOI, strained silicon, multigate/FIN-FETs, metal gates, etc. He has published several papers in various conference proceedings and microelectronic journals. His main research work is in the area of NBTI. Furthermore, he is involved in JEDEC standard developments.

Dr. Schlünder is a member of the Technical Program Committee of the IEEE conferences, International Reliability Physics Symposium (IRPS), and Integrated Reliability Workshop. He is a referee of several IEEE journals. In addition, he has presented invited talks and tutorials at many conferences such as the IRPS.

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on November 16, 2009 at 04:03 from IEEE Xplore. Restrictions apply.