The effect of recovery on NBTI characterization of thick non-nitrided oxides

H. Reisinger, R.P. Vollertsen, P.J. Wagner *, T. Huttner, A. Martin,
S. Aresu, W. Gustin, T. Grasser *, and C. Schlünder
Infineon Technologies, Corporate Reliability Methodology, D-81739 München, Germany;
* Christian Doppler Laboratory for TCAD at the Institute for Microelectronics, TU Wien, A-1040 Wien, Austria
e-mail: Hans.Reisinger@infineon.com

ABSTRACT

NBTI degradation and recovery have been investigated for 7 to 50nm oxides and compared to a thin 2.2nm nitrided oxide. A wide regime of stress fields 2.5MV/cm to 8MV/cm has been covered. NBTI effect for the nitrided oxide is larger than for non-nitrided oxides. The percentage of threshold shift *ΔVth* which is "lost" during a long measurement delay – which is the quantity leading to curved *ΔVth* vs stress-time curves and to errors in extrapolated lifetimes – is about equal for nitrided or thick non-nitrided oxides. The fraction of recovered *ΔVth* 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.

I. INTRODUCTION

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 work to compare the NBTI effect for thick oxides from 7nm to 50nm to thin plasma nitrided oxides (PNO) for a wide regime of electric fields from use condition around 3MV/cm up to high stress field 10MV/cm. This work demonstrates, for the first time, that the impact of recovery in thick oxide devices is as critical as it is for thin PNO oxides.

II. SAMPLES AND EXPERIMENTS

10µm wide pMOSFETs with 3 different non-nitrided oxides (7nm, 15nm and 50nm thick) with n^+-poly gates were used for this study.

Data are compared to 2.2nm PNO oxide devices from a dual work function technology (boron doped poly) as described in [5].

Threshold voltages *Vth* were extracted at a current criterion of 70nA*W/L in saturation at a temperature of 150°C, which is the desired maximum operation temperature. Our fast measurement technique with a 1µs measuring delay was used as described in [5]. All results for the thick oxides were gathered from the same wafer. Threshold shifts were reproducible within a ±1mV plus ±2% margin.

EXPERIMENTAL RESULTS

An example of raw measured *Vth* data is shown in Fig. 1. The stress was interrupted 1 to 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 1s measuring delay from the same measurement and the same sample. Any stress interruption was more than a 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 Δ*Vth* vs 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 7 traces in Fig. 1 are fitted by a single, global recovery law (containing the normalized recovery time *tr* over stress time). Each of the traces contains 2 more parameters, i.e. *ΔVth0* at *tr* = 0 and the permanent part *ΔVth∞* at *tr* = ∞. The fit provides a nice tool to perfectly fit and smooth the measured data – except the very long term recovery behavior as seen in the uppermost trace of Fig. 1 – and to extract *ΔVth(tr)* curves at any desired measuring delay, as practiced in Figs. 1 and 2. In general the recovery function (see full lines in Fig. 1) as well as measured recovery traces for thin oxides have the property that the derivative *dVth/d(log tr)* is monotonically decreasing with recovery time. For our thick oxides and long recovery times the slopes of the experimental traces (cmp. 10ks-trace for recovery times >100s in Fig. 1) get steeper with recovery time and do not obey the universal recovery law.

Fig. 2 shows a "fast" *ΔVth(t)* curve with a 1µs measuring delay and a "slow" one with a 1s 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 seen in Fig. 2 the difference between the fast and the slow curve is roughly constant with a weak stress time dependence only. The "slow" *ΔVth(t)* curve shows power-law exponents n>0.25 and the well known bending of the curve. This

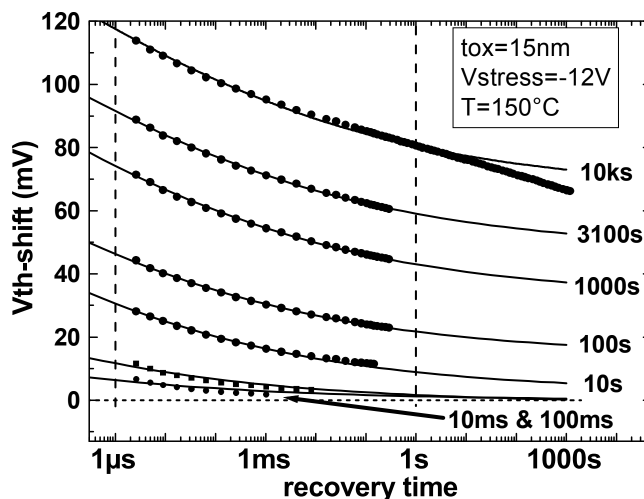

Fig.1: Example of raw data as measured. Stress is interrupted at certain intervals (labels give total stress time). Each trace (dots) shows measured Vth as a function of the time after stress interruption (= recovery time). Lines are fitted using universal relaxation law and permanent components as described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

978-1-4244-2195-4/08/$25.00 ©2008 IEEE

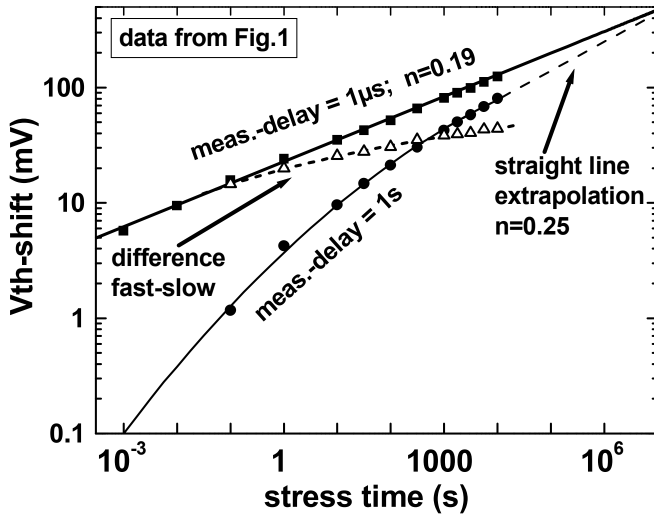

Fig. 2: ΔVth as a function of stress time extracted from Fig. 1. A "fast" and a "slow" curve (for a measuring delay of 1µs and of 1s as for a conventional analyzer) is extracted along the vertical dashed lines in Fig. 1 (not all the measured traces are shown in Fig. 1). n denotes the power-law exponent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

bending has been often interpreted as a saturation in past. As seen in Fig. 2 this "saturation" is just an artifact caused by recovery. The "fast" $\Delta Vth(t)$ 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.

III. Time extrapolation of ΔVth and lifetime prediction

For "slow" ΔVth(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 ΔVth shift criterion can be done [10]. In contrast "fast" ΔVth(t) curves - as a consequence of the absence of any curvature - allow an extrapolation to longer times for a stress time which is a 100 times shorter without any sacrifice as can be seen in Fig. 2.

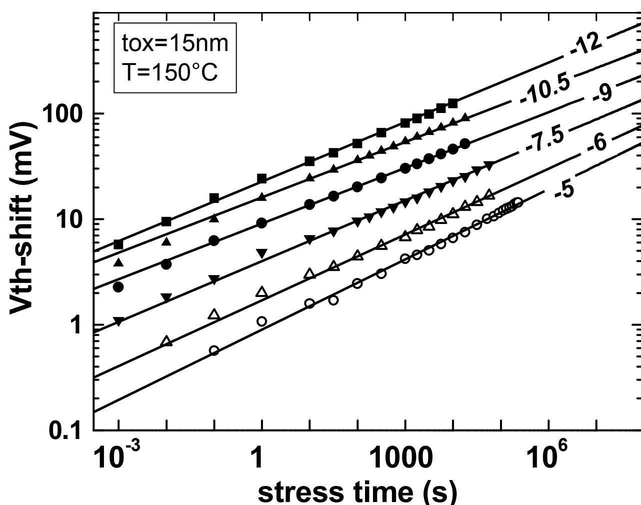

Fig. 3: ΔVth vs stress time for 1µs measuring delay. Curve labels denote the stress voltages. Measured points for stress time >1s are fitted to a power law (full straight lines). The curves follow a power law $\Delta Vth = const * t^n$ for more than 7 decades of time. The power law exponents n are given in Fig. 12.

Fig. 4 shows an example for a lifetime extrapolation done using "fast" and "slow" ΔVth(t) measurements for measuring delays of 1µs and 1s, respectively. The "fast" and "slow" lifetimes were taken from extrapolations of the ΔVth(t) vs. 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. Underestimation occurs for high stress fields for which the drift criterion is reached during stress but the measured ΔVth is too low due to recovery. Underestimation occurs for low stress fields were 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 100s and the drift criterion were >100mV then this extrapolation of the "slow" curve from 100s 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, 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 -6V in Fig. 4). The straight line ΔVth(t) vs. stress time curve extracted by the fast measurement technique as shown in Fig. 3 will yield correct lifetimes. Since the extrapolated straight line is independent of the stress time this is true even for short stress times .

IV. Comparison of different oxide thicknesses

A. Basics for comparison of different oxide thicknesses

It is a necessary precondition for a meaningful comparison of Vth 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 1nm thickness the simple $E=(Vg-VFB)/tox$ estimation (VFB = flatband voltage) leads to a field value which is too high by roughly a factor of 2. 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 in the order of 1V in the substrate and of order 500mV 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

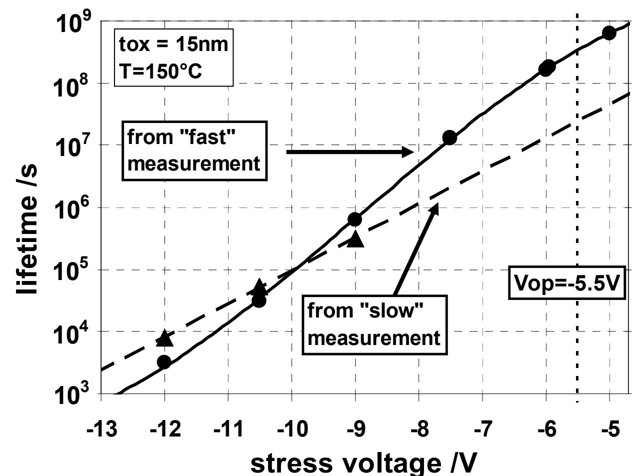

Fig. 4: Lifetime extrapolation from "fast" and "slow" measurements for a $\Delta Vth = 100$mV lifetime criterion. The extrapolation to operating voltage from the slow measurement from highly accelerated stress is significantly **under**estimating the lifetime. A most reliable assessment is done from low fields close to operation voltage with a fast measurement.

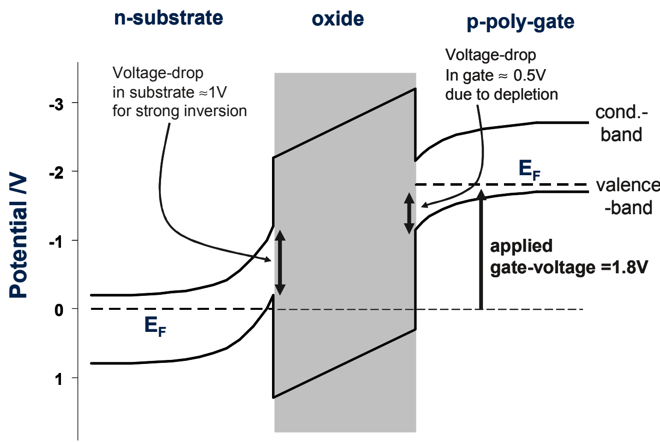

Fig. 5: Band-diagram of a pFET with a p-doped Si gate driven to p-inversion by applying a very high negative gate voltage corresponding to stress fields. To calculate the electric field in the oxide the voltage drops in substrate and poly-gate have to be subtracted from the applied gate voltage.

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-equation. 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 ±100mV due to the inevitably inappropriate treatment of the many-body effects in the space charge layer [11].

In a boron-doped poly-gate there is a depletion for negative gate bias at high field. The concentration of electrically activated boron in a poly gate is in general much less than in a n-doped gate. This leads to voltage drops in the boron doped poly in the order of 0.5V at 10MV/cm as seen in studies of MNOS memories [12]. 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.

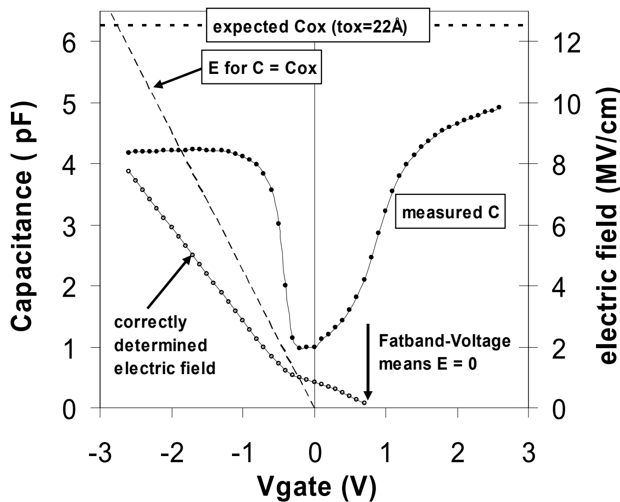

Fig. 6: Measured CV-curve (left vert. axis), simple approximation $E=Vg/tox$ and correctly determined electric field E for the thin 2.2nm oxide (gate area=400μm^2). E is determined from the CV-curve and Maxwell-eq. $divE=\rho$. For pMOSFETs the electric field is significantly reduced by depletion in the p+-doped 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 is available.

For a two-dimensional capacitor (i.e. thickness<<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):

$$E = Q / A / \kappa \quad (1)$$

κ is the static dielectric constant of the dielectric. It is 3.9 for SiO2 and only ≈5% percent higher if the oxide contains 10% Nitrogen. The CV-curve directly measures the quantity dQ/dVg as a function of the gate voltage Vg. So 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 [13] is sufficient.

In conclusion a calculation of the electric field from the CV curve implicitly contains all QM-effects, gate depletion and also work function differences. As can be from data in Fig. 6 corrections have to be added to the 22Å physical oxide thickness by QM and p+-poly depletion of ≈5Å and ≈7Å, respectively (both equivalent oxide thickness). The correction is independent of the thickness of the dielectric, so the same correction of 12Å would be valid for a 12Å oxide making the error of the simple $E=Vg/tox$ approximation a factor 2.

B. Results of comparing different oxide thicknesses

Fig. 7 shows the magnitude of the NBTI Vth shift for all investigated oxides. For a given amount of interface charge the resulting ΔVth is proportional to the oxide thickness (to be correct: to 1/capacitance), 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 $\Delta Vth(E)$ follows the

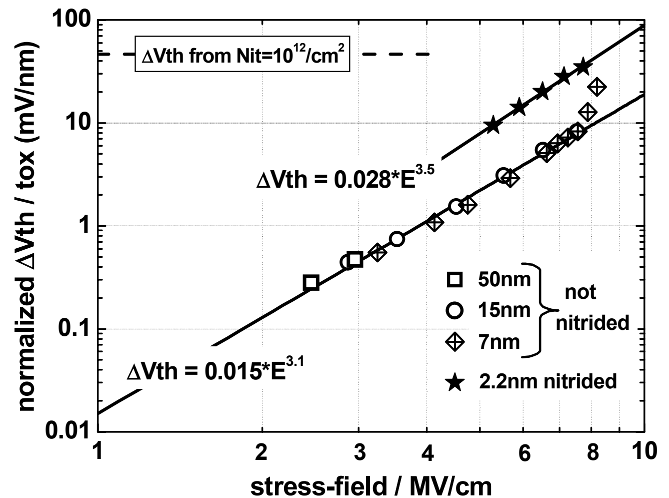

Fig. 7: Normalized ΔVth as a function of stress field after a stress time of **10ks** fitted to a power law in stress-field. The horizontal line at ΔVth=46mV shows ΔVth caused by an interface charge of 10^{12}/cm^2.

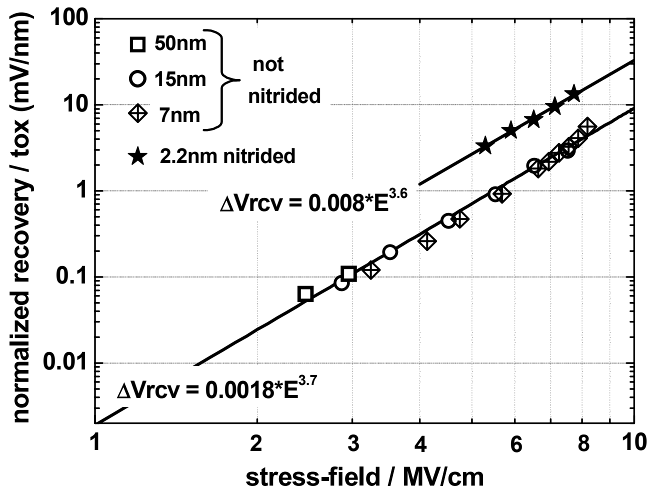

Fig. 8 Normalized **recovery** $\Delta Vrcv$ as a function of stress field after a stress time of **10ks**. Recovery was taken between 1µs and 1s. To save space equations in Figs.7, 8 do not contain correct dimensions. E has to be entered in MV/cm and the result to be multiplied by tox/nm to get ΔVt and $\Delta Vrcv$ in mV.

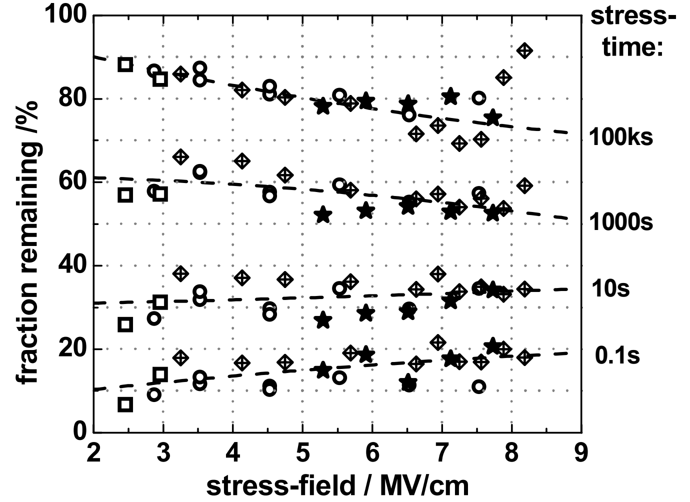

Fig. 10: Same as Fig. 9 for other stress times, same symbols used.

same power-law with a slightly lower exponent than for the thin PNO-oxide. The PNO power-law exponent is 3.5, comparable to the value of 4.1 found in [8] (for much higher fields). Our thin PNO shows a ΔVth enhanced by a factor of ≈3 compared to the non-nitrided oxides. This can be attributed to a larger number and a different nature of interface states due to nitridation [14]. The **absolute** normalized amount of recovery is shown in Fig. 8. It follows a similar power-law as the $\Delta Vth(E)$ curves in Fig. 7. The **relative** amount of recovery and remaining ΔVth after a stress time of 10ks is plotted in Fig. 9. The relative recovery after a measuring delay of 1s is around 30%. ΔVth measured at a 1µs measuring delay is taken as 100%. Recovery is only slightly dependent on electric field and is **exactly the same** for nitrided and non-nitrided oxides. The same data shown in Fig. 9 for a 10ks stress are shown in Fig. 10 for stress times 0.1s to 100ks with the same results that relative recovery is approximately independent of the technology (nitrided or not), independent of thickness and only weakly dependent of electric field. Thus the only parameter the remaining fraction depends on is the stress time and of course the measuring delay.

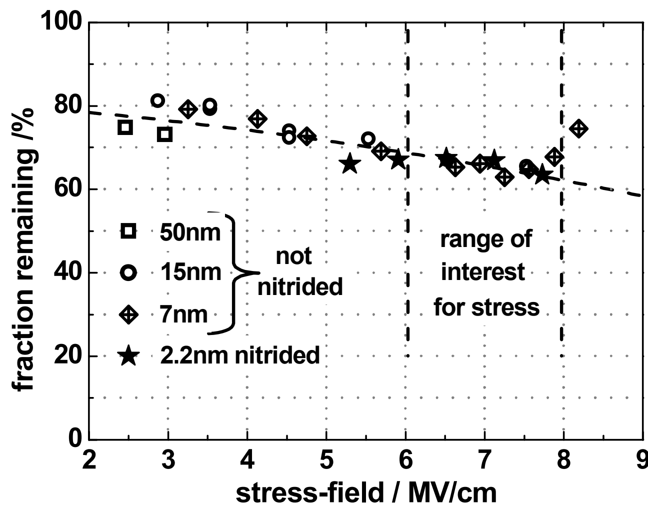

Fig. 9: Percentage of ΔVth which is measured with a conventional 1s measuring delay taken at a stress time of **10ks**. The ΔVth measured at a 1µs delay was set as 100%. The 50nm samples could be stressed at low field only due to a ±16V output voltage limit of our exp. setup. The dashed line is a 2nd order polynomial through all data (except 7nm data at >7.7MV/cm, see text)

Fig. 11 extracts the remaining fractions of Fig. 10 for a stress field of 6MV/cm which is a value most common for stress. Plotted is the remaining fraction given by the averaged, 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 10^7s (4 months).

The most important conclusion to be drawn from the data in Figs. 9 to 11 is that - while NBTI in PNO is a factor 3 stronger than in the non-nitrided oxides – 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 ΔVth vs stress time curves, thus causing all the trouble with NBTI extrapolation and assessment. The impact of recovery on NBTI assessment is as severe for "old technologies", for thick, non-nitrided oxides as for modern, thin nitrided oxides. There is also evidence for another conclusion: both 2.2nm and thicker oxide obey a similar power law for degradation (cmp. Fig. 7) and recovery is very similar (cmp. Figs. 8, 9, 10). This would support the hypothesis that any diffusion of Hydrogen into the oxide beyond 2nm from the interface does not play a role or does nor occur. Otherwise the 2.2nm data would look different from the thick-oxide data.

Fig. 12 shows the power law exponents n of the $\Delta Vth \propto t^n$ fits from Fig. 3. There is a slight but systematic increase of $\boldsymbol{n}$ with thickness.

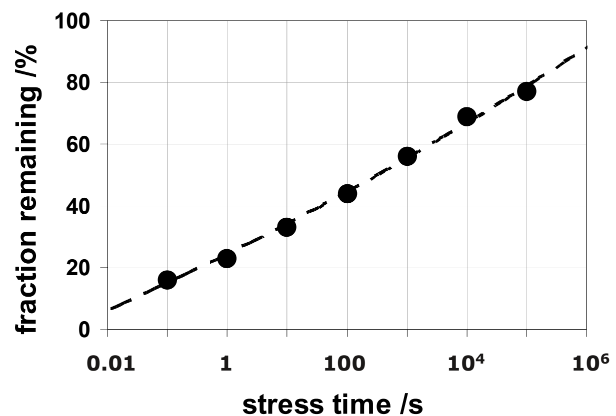

Fig. 11: Values of remaining fraction at 1s measuring delay for a stress field of 6MV/cm. Taken are averages over all samples = dashed lines at 6MV/cm in Fig. 10.

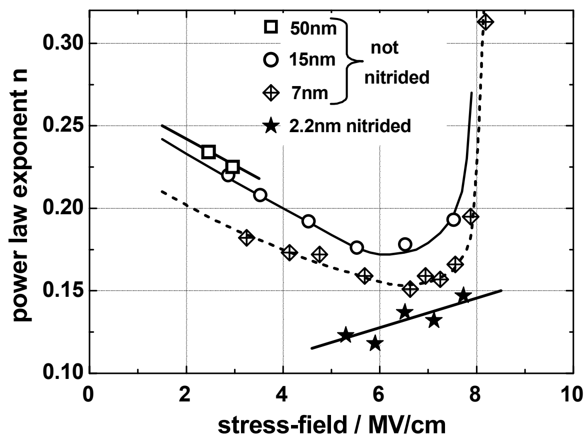

Fig. 12: Power law exponents of $\Delta Vth(t)$ curves (cmp. Figs. 2, 3) as a function of stress field for all investigated samples. Lines are splines through the data points.

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 high fields above 7MV/cm there is a sharp increase in slope for the thick oxides. This high-field deviation is also seen in the 7nm-data in Fig. 7, 8, 9, 10. We think that this 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 having kinetic energies >5eV 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. Such an increase in n correlated with an increase in stress induced gate leakage also has been seen in ref. [15]. It should be noted that the increase in n is seen only at high fields >7MV/cm and not from the start of stress but first after >1000s when oxide wear out gets effective (cmp. 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 seen in Figs. 9,10. To further support the assumption that oxide wear out plays a role at high fields time dependent dielectric breakdown data (TDDB) are shown in Fig. 13. TDDB for the 7nm-oxide was done at constant gate voltage stress. At a stress field of 7.5MV/cm the oxide is close to the dielectric breakdown after a stress time of 10ks and so is already heavily damaged.

In contrast to the thick oxides ($\geq$7nm) the stress voltages in the thin 2.2nm oxide are below 3V, so below the Si-SiO2 band-offset and thus in the direct tunneling regime with no energy dissipated in the oxide.

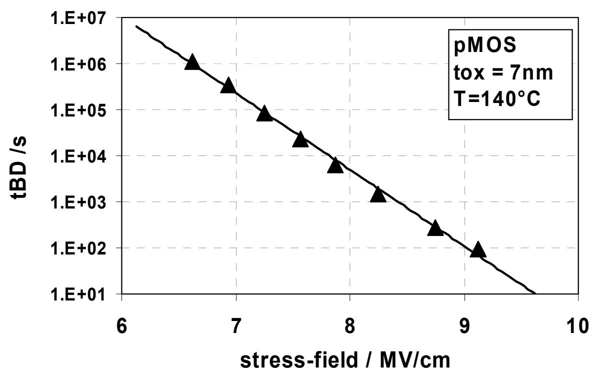

Fig. 13: Time to breakdown tBD after constant voltage stress for the 7nm non-nitrided oxide.

Currently there are not enough experimental data to clearly separate the observed high-field deviation from real NBTI and to attribute it to dielectric wear out. Since dielectric wear out is not thermally activated in contrast to NBTI the temperature dependence would be a means to distinguish the two effects.

The high field effects will have to be investigated further.

SUMMARY

NBTI drift and recovery have been measured for non-nitrided oxides 7nm to 50nm for stress-field E from use condition 2.5MV/cm to 8MV/cm. A comparison to a thin 2.2nm nitrided oxide has been done. As expected the NBTI shift in thick non-nitrided oxides is less than in thin nitrided oxides (cmp. Fig. 7). However, the effect of recovery on measurements, extrapolations and assessments is (even for low fields) equally significant in non-nitrided oxides (see Fig. 9, 10) - which is in contrast to wide spread belief. For ultra short measuring delays ΔVth vs 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, cmp. 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 Vth shift is power-law, with very similar exponents 3.5 for PNO and 3.1 for non-nitrided oxides (cmp. Fig. 7). 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 2.2nm oxide and thicker oxides suggests that all NBTI processes take place in a thin interface layer <2nm. Outlook: This study has been done at a single temperature of 150° and with a single nitrided oxide thickness oxide (2.2nm). A more systematic study employing nitrided and non-nitrided oxides below 2nm and between 2nm and 7nm and other temperatures would be desirable.

REFERENCES

[1] J.H. Stathis a. S. Zafar, "The Negative Bias Temperature Instability in MOS Devices: A Review", Microelectronics Reliability, Vol.46, no.2-4, 2006, p. 270

[2] C. Shen, M.-F. Li, C.E. Foo, T. Yang, D.M. Huang, A. Yap, G.S. Samudra, Y.-C. Yeo, "Characterization and Physical Origin of Fast Vth Transients in NBTI of pMOSFETs with SiON Dielectric", IEDM 2006 Technical Digest, p. 333

[3] S. Rangan, N. Mielke, E.C.C. Yeh, "Universal Recovery Behavior of Negative Bias Temperature Instability", IEDM 2003 Technical Digest, p. 341

[4] V. Huard, C. Parthasarathy, N. Rallet, C. Guerin, M. Mammase, D. Barge, C. Ouvrard, "New characterization and modeling approach for NBTI degradation from transistor to product level", IEDM 2007 Technical Digest, p. 797

[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 TDMR, Vol. 7, No. 1, p. 119 (2007)

[6] M.A. Alam, S. Mahapatra, "A comprehensive model of PMOS. NBTI degradation " , MR Vol.45, p. 71 (2005)

[7] A.T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John, S. Krishnan, "NBTI Impact on Transistor & Circuit: Models, Mechanisms & Scaling Effects", IEDM 2003 Technical Digest, p. 345

[8] S. Tsujikawa, T. Mine, K. Watanabe, Y. Shimamoto, R. Tsuchiya, K. Ohnishi, T. Onai, J. Yugami, a. S. Kimura,

"Negative Bias Temperature Instability of pMOSFETs with Ultra-thin SiON Gate Dielectrics", Proc. IRPS 2003, p. 183

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

[10] 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", IRW 2006, pp. 1-4

[11] T. Ando, A.B. Fowler, F. Stern, "Electronic properties of two-dimensional systems", Reviews of Modern Physics, Vol. 54, p. 437 (1982)

[12] H. Bachhofer; H. Reisinger; E. Bertagnolli; H. Philipsborn, "Transient conduction in multidielectric silicon-oxide-nitride-oxide semiconductor structures", JAP, Vol.89, no.5, p.2791 (2001)

[13] S.M. Sze a. K.K. Ng in "Physics of semiconductor devices", Wiley & Sons, third edition 2007, chapter 4

[14] J.P Campbell a. P.M. Lenahan, "NBTI: An atomic-scale defect perspective", Proc. IRPS 2006, p. 442

[15] S. Mahapatra and M.A. Alam, "A Predictive Reliability Model for PMOS Bias Temperature Degradation," IEDM 2002 Technical Digest, p. 505

Questions and Answers

Q: What do you call recovery, is it ΔVth? (Slide 12 and Fig.8)

A: Yes, we define here "recovery" as the measured ΔVth between measuring delay 1µs and 1s. This recovery is then normalized by dividing it by the oxide thickness.

Q: Has your team also done any independent experiments where you fix electric field and vary flux of current? Does this also have any effect? Does gate leakage change NBTI?

A: Normal NBTI stress fields for thick oxides are 5-7MV/cm. Up to 6MV/cm the Fowler-Nordheim current is below the detection limit, thus the flux is practically zero. Up to 7MV/cm a finite stress time around 10^5s still means a gate oxide wearout corresponding to below 1% of charge-to-breakdown. This means that for a given gate voltage the current is constant and there is no difference between constant field and constant current. For any thick, non-nitrided oxide any current above the detection limit ($\approx 10^{-10}$A/cm^2) "gate leakage" means a Fowler-Nordheim current. So a variation in "gate leakage" does not exist.

Q: All processes happen in <2nm interface layer (Slide 17). Does this mean the reaction diffusion model is wrong?

A: 1. It is well known that - for example after a 100s stress - there is recovery continuously ranging from "very fast" with time constant of 1µs and below to "very slow" with time constants up to 10^5s and above. Recovery according to the reaction diffusion model contains only a single time constant of 100s and thus cannot explain recovery.

A: 2. To our knowledge there are no studies being able to extract diffusion constants D from experimental data and thus to present any number for a diffusion length $\sqrt{Dt}$. Such a number would be the precondition to judge if the reaction diffusion model is right or wrong.

A: 3. On the other hand 2nm correspond to a number of ≈ 6 atomic layers only which clearly is not a regime covered by classical diffusion theory. So we think yes, the reaction diffusion theory is wrong.