Understanding Temperature Acceleration for NBTI

Gregor Pobegen*§, Thomas Aichinger†, Michael Nelhiebel‡ and Tibor Grasser§
*KAI GmbH, A-9524 Villach, Austria; gregor.pobegen@k-ai.at †Pennsylvania State University, USA
‡Infineon Technologies, Austria §Institute for Microelectronics, TU Vienna, Austria

Abstract

Based on experimentally observed temperature-dependent charge exchange during NBTI, which is consistent with non-radiative multiphonon processes, we propose a new concept to understand degradation data at different temperatures. We show the impact of temperature-acceleration on both degradation and recovery, making it possible to perform long-term NBTI tests in a fraction of the time usually required.

Introduction

Recently, the negative bias temperature instability (NBTI) has been explained as a nonradiative multiphonon carrier exchange process between the silicon substrate and the gate oxide [1–3]. Within this framework capture cross sections σ for charge trapping are assumed to be thermally activated as [1, 4] $\sigma = \sigma_0 \exp(-E_\mathrm{B}/k_\mathrm{B}T)$, with a barrier energy E_B. Consequently, the defect time constants τ are thermally activated as

$$\tau \approx \tau_0 e^{\frac{-E_\mathrm{B}}{k_\mathrm{B}T}} \tag{1}$$

with a temperature-independent characteristic time constant τ_0 [5, 6]. Defect time constants τ_1 at temperature T_1 consequently change to

$$\tau_2 = \tau_0 \exp\Big(\frac{k_\mathrm{B}T_1 \ln \tau_1/\tau_0}{k_\mathrm{B}T_2}\Big) = \tau_0 \Big(\frac{\tau_1}{\tau_0}\Big)^{T_1/T_2} \tag{2}$$

at temperature T_2. This equation links time dependent data measured at T_1 to the equivalent values at T_2, thus introducing a new way to interpret NBTI data. Considering a distribution of time-constants, data recorded at T_ref will be equivalent to data recorded at any T, provided the data is plotted as a function of the *temperature-time*

$$\vartheta(T) = \tau_0 \Big(\frac{t}{\tau_0}\Big)^{T_\mathrm{ref}/T} \tag{3}$$

The benefit of (3) is that a fixed experimental time-window is transformed, allowing us to sample different fractions of the distribution at different temperatures and compile the results on a single time-scale.

Experimental details

The temperature-time concept becomes especially powerful when combined with the recently established poly-heater technique [5, 7] previously developed for fast wafer level reliability (fWLR) monitoring [8]. This technique allows fast on-chip device temperature switches through control of the power dissipated in two polysilicon wires situated next to the device (Fig. 1). To accurately determine the device temperature, one measures the drain current at a suitable chosen operating point (V_G, V_D) at different thermo chuck temperatures prior to stress. These measurements provide a second-order polynomial $I_\mathrm{D}(T)$ which is used to transform the drain current increase into an appropiate device temperature increase when supplying heating power with the poly-heater [7]. In order to access device temperatures beyond the chuck temperature range (-60°C to 200°C), an extrapolation of the change of the device temperature with the heater power was performed, which takes the change of the thermal resistances of the materials between the heater, the device and the thermo chuck into account. The extrapolation from a sweep range of 20°C is shown in Fig. 2 to cause a relative error of only a few percent even for a large temperature switch of 260°C.

Accelerated recovery

In Fig. 3 recovery traces at different temperatures following equal stress defined through temperature, time and oxide field, by using the poly-heater, are depicted [9]. When shifting the data along the time axis according to equation (3) all recovery traces can be mapped onto one single trace. Fig. 4 shows that this behavior is not technology dependent, since we made equivalent observations on devices having different gate oxides (plasma nitrided SiON or SiO_2) and different gate oxide thicknesses (2.2nm to 30nm). Having established this fact, we now switch the temperature to higher values *during* recovery, as depicted in Fig. 5 [5], which allows us to access the universal recovery trace of Fig. 3 in a single measurement.

Accelerated recovery proves that short gate bias switches to positive values, as depicted in Fig. 6, discharge defects with small time constants [5, 10] and reveal the quasi-permanent part of the NBTI. This accelerated recovery experiment is consistent with earlier work concerning complete recovery through bake [11, 12], as being a consequence of recovery through bias switches and temperature acceleration.

Accelerated stress

A conventional way to track ΔV_TH over stress time is to briefly interrupt logarithmically increasing V_str phases by switching to V_TH to measure the degradation [13]. Comparing such measurements for different chuck temperatures using equation (3) is challenging, because at higher T not only stress but also recovery is accelerated. It is therefore insufficient to account only for temperature acceleration of the stress time without correcting for accelerated recovery (compare 'after same *t*' lines in Fig. 7). In principle, one could account for the different amount of recovery during stress interruption by using equal recovery times on the ϑ_rec scale. However, this

requires extrapolation of the recovery traces, because with a 10ms measurement delay at e.g. 200°C one would have to wait approximately 10^9s $\approx$ 30 years at -60°C. The errors caused by the extrapolation result in small deviations of the 'after same ϑ' lines in Fig. 7. As the above approach is clearly unsatisfactory, we again exploit the versatility of the poly-heater technique as depicted in Fig. 8, by measuring the ΔV_{TH} always at the same (substantially) lower chuck temperature. Thereby, no extrapolation is needed because the degradation is always measured with the same temperature-time delay and recovery is minimized due to degradation quenching [9]. Fig. 9 shows the result of an accelerated interrupted stress measurement using the poly-heater. The power law behavior which is observed for moderate temperatures/times clearly saturates outside the typical experimental window of approximately 10^6 to 10^7s [14]. The saturation of the power law can be perfectly explained by a log-normal distribution of capture time constants, see Fig. 10. We observe the same result also on a vastly different technology with 30nm SiO_2 gate oxides (c.f. Fig. 11) or when using another acceleration approach as depicted in Fig. 12. Of particular importance is that the degradation does not saturate abruptly, as for instance predicted by the reaction-diffusion model. The power law saturation is rather an essential feature of the degradation at all times, thereby confirming the wide distribution of time constants.

In analogy, one may also accelerate on-the-fly (OTF) NBTI tests through temperature switches to higher values during stress. In Fig. 13 we switched the device temperature from 100°C to 200°C for short periods of time while keeping the stress bias applied. When performing more intensive T accelerated OTF stress as depicted in Fig. 14 we observe again the same saturation of the degradation. Again, the data can be fitted by a log-normal cumulative distribution function. Comparing OTF data measured at different chuck temperatures requires a correction for the degradation of the first drain current measurement at stress level $I_{\mathrm{D},0}$, see Fig. 15.

In Fig. 16 we compared long-term NBTI degradation data from qualification tests up to 1500 hours (approximately 2 months) with temperature accelerated test using the poly-heater acquired within less than a day, giving more than satisfactory agreement. Fig. 17 shows how an accelerated test may be used to access the degradation of a pMOS transistor at use conditions.

Conclusions

By exploiting the versatility of the poly-heater technique, which allows for arbitrary temperature switches during measurement, together with the temperature-activated nature of NBTI, we have suggested a technique which allows for direct measurement of long-term degradation and recovery. Even though there remains some uncertainty due to the choice of τ_0, our method provides the most reliable way to directly measure device degradation outside any reasonable experimental window. Furthermore, if our method is employed for fWLR tests, where a very large number of transistors has to be measured in the shortest time possible, τ_0 may be calibrated to a reference measurement to further decrease the remaining uncertainty.

References

[1] M. J. Kirton and M. J. Uren, "Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/f) noise," *Advances in Physics*, vol. 38, no. 4, pp. 367–468, 11 1989.

[2] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel, "A two-stage model for negative bias temperature instability," in *IEEE International Reliability Physics Symposium*, 4 2009, pp. 33–44.

[3] T. Grasser, H. Reisinger, P.-J. Wagner, and B. Kaczer, "Time-dependent defect spectroscopy for characterization of border traps in metal-oxide-semiconductor transistors," *Physical Review B*, vol. 82, no. 24, p. 245318, 12 2010.

[4] C. H. Henry and D. V. Lang, "Nonradiative capture and recombination by multiphonon emission in GaAs and GaP," *Physical Review B*, vol. 15, no. 2, pp. 989–1016, 1 1977.

[5] T. Aichinger, M. Nelhiebel, and T. Grasser, "Unambiguous identification of the NBTI recovery mechanism using ultra–fast temperature changes," in *IEEE International Reliability Physics Symposium*, 4 2009, pp. 2–7.

[6] T. Grasser, H. Reisinger, P. Wagner, F. Schanovsky, W. Gös, and B. Kaczer, "The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability," in *IEEE International Reliability Physics Symposium*, 5 2010, pp. 16–25.

[7] 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 Transactions on Device and Materials Reliability*, vol. 10, pp. 3–8, 3 2010.

[8] C. Schlunder, 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 *European Solid State Device Research Conference*, 9 2007, pp. 131–134.

[9] T. Aichinger, M. Nelhiebel, and T. Grasser, "On the temperature dependence of NBTI recovery," *Microelectronics Reliability*, vol. 48, pp. 1178–1184, 9 2008.

[10] T. Grasser, T. Aichinger, G. Pobegen, H. Reisinger, P.-J. Wagner, J. Franco, M. Nelhiebel, C. Ortolland, and B. Kaczer, "The 'Permanent' Component of NBTI: Composition and Annealing," in *IEEE International Reliability Physics Symposium*, 6 2011, pp. 605–613.

[11] A. A. Katsetos, "Negative bias temperature instability (NBTI) recovery with bake," *Microelectronics Reliability*, vol. 48, pp. 1655–1659, 10 2008.

[12] C. Benard, G. Math, P. Fornara, J.-L. Ogier, and D. Goguenheim, "Influence of various process steps on the reliability of PMOSFETs submitted to negative bias temperature instabilities," *Microelectronics Reliability*, vol. 49, pp. 1008–1012, 11 2009.

[13] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, "Disorder–controlled–kinetics Model for Negative Bias Temperature Instability and Its Experimental Verification," in *IEEE International Reliability Physics Symposium*, 8 2005, pp. 381–387.

[14] H. Aono, E. Murakami, K. Okuyama, A. Nishida, M. Minami, Y. Ooji, and K. Kubota, "Modeling of NBTI degradation and its impact on electric field dependence of the lifetime," in *International Reliability Physics Symposium*, 4 2004, pp. 23–27.

[15] S. Mahapatra, A. E. Islam, S. Deora, V. D. Maheta, K. Joshi, A. Jain, and M. A. Alam, "A Critical Re-evaluation of the Usefulness of R-D Framework in Predicting NBTI Stress and Recovery," in *IEEE International Reliability Physics Symposium*, 6 2011, pp. 614–623.

[16] T. Grasser, in *IRPS*, ser. tutorial, 2011.

[17] T. Nagumo, K. Takeuchi, T. Hase, and Y. Hayashi, "Statistical Characterization of Trap Position, Energy, Amplitude and Time Constants by RTN Measurement of Multiple Individual Traps," in *IEEE International Electron Device Meeting*, 2010, pp. 28.3.1–28.3.4.

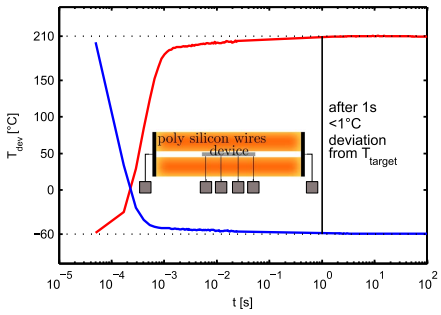

Fig. 1. The poly-heater consists of two polysilicon wires situated next to the device. The electrical power supplied to the poly wires increases the device temperature and thus the drain current of the transistor. The rise of the drain current with turn-on of the heating power (red line) or turn-off (blue line) is converted to the change of the device temperature with time to reveal that the temperature settles close to its target value within less than a second [7].

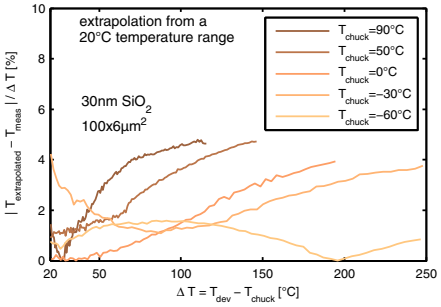

Fig. 2. Relative error when extrapolating the device temperature T_{dev} based on a sweep of the power dissipated in the poly-heater P_{PH} from a 20°C temperature range at chuck temperature. The extrapolation involves the changes of the thermal resistances of the materials surrounding the device and the heater. The relative error increases with the temperature difference ΔT between the the device and the chuck but stays well under a few percent for the accessible temperature range.

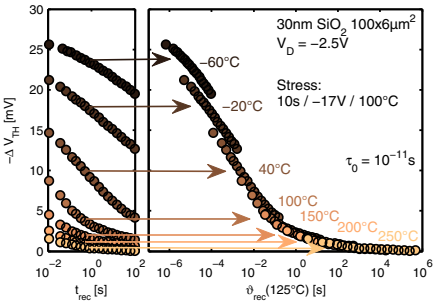

Fig. 3. On the left hand side recovery traces at different temperatures following exactly the same stress defined through temperature, duration and oxide field are depicted. On the right hand side the data was replotted as a function of the temperature-time. With a suitably chosen τ_0 all recovery traces can be matched.

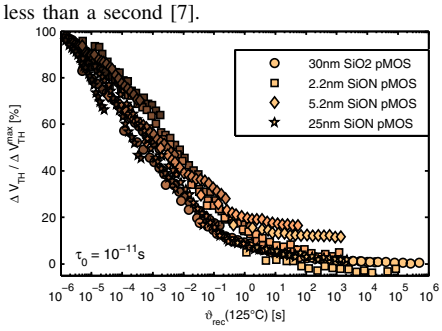

Fig. 4. The experiment of Fig. 3 was repeated on several different technologies with different gate dielectrics (SiO_2 and PNO SiON) and different gate oxide thicknesses. The temperature-time concept works for all observed technologies with one single value of τ_0. The graph was normalized to the maximum drift 10ms post stress at -60°C.

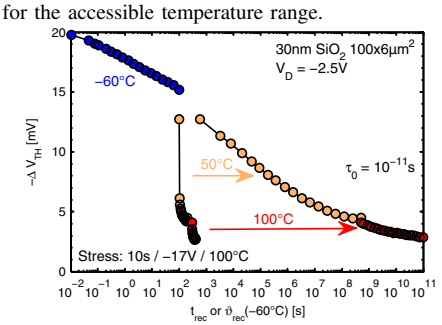

Fig. 5. When performing device temperature switches with the poly-heater during constant bias recovery at V_{TH}, substantial acceleration is observed [5]. Re-evaluation of the time data on the temperature-time scale results in a continuous recovery trace.

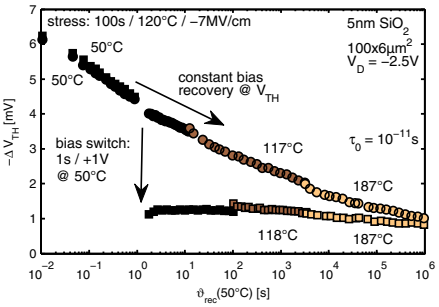

Fig. 6. Gate bias switches to positive values let defects with small time constants discharge quickly. The remaining degradation after the positive bias phase appears to be permanent but still undergoes slight recovery at long-enough times [10].

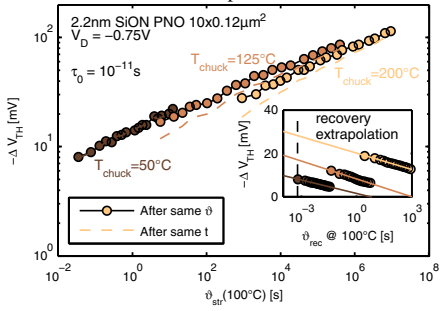

Fig. 7. When comparing interrupted stress measurements at different chuck temperatures without the use of the poly-heater, data interpretation in terms of the temperature-time becomes more sophisticated. When the stress temperature equals the recovery temperature both effects are accelerated in a similar manner. One has to compare the drift with an equal delay on the temperature-time scale. This demands extrapolation of the recovery data because even only 10ms delay for the first ΔV_{TH} measurement at 200°C corresponds to approximately 10^9s or 30 years at -60°C. This approach introduces some errors but already demonstrates the concept.

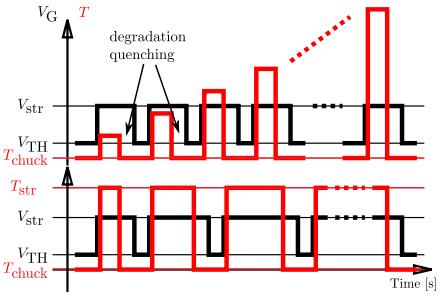

Fig. 8. For the measurements of Fig. 9-11 we interrupt the stress with logarithmically increasing intervals in order to measure the degradation level at the V_{TH} (lower plot). For the measurements of Fig. 12 we increased T with every interruption, keeping the stress time constant (upper plot). During stress we accelerate the degradation by supplying additional heat with the poly-heater. We turn off the heater prior to stress interruption in order to carefully avoid accelerated recovery (degradation quenching) [9]. The interruption of the test and the ΔV_{TH} measurement is always performed at a much lower chuck temperature T_{chuck} which minimizes undesirable recovery and makes the amount of recovery independent of the stress temperature T_{str}.

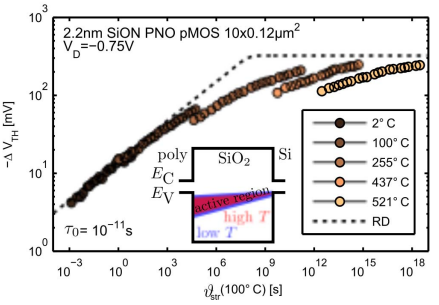

Fig. 9. Result of an interrupted stress measurement as introduced in Fig. 8 for a state-of-the-art small sized pMOS with 2.2nm thick SiON gate oxide. Stress times were transformed according to equation (3) to interpret the accelerated test. For conventional NBTI test temperatures we observe the well known $1/6$ power law coefficient as predicted by the reaction-diffusion theory [15]. However, the power law clearly saturates for large temperatures/times. This saturation occurs homogeneously at all times rather than abruptly as predicted by the reaction-diffusion model, and is as such an essential feature. The small offset in the drift is due to a decrease of the active energy region for the defects within the oxide due to the increasing stress temperatures [16, 17].

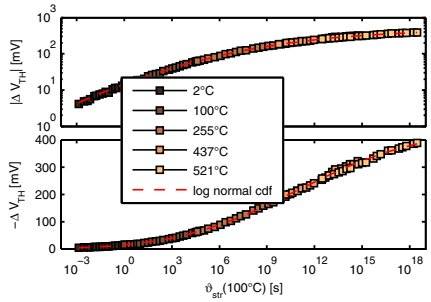

Fig. 10. When the data of Fig. 9 is corrected for the small offset due to the different active energy regions at different temperatures, it can perfectly be fitted by a log-normal cumulative distribution function (cdf) with large mean and large standard deviation.

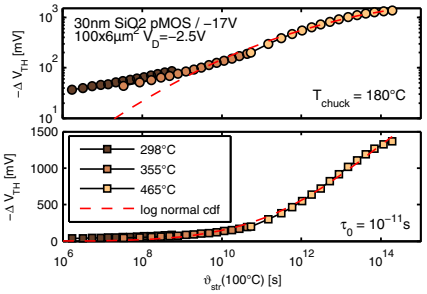

Fig. 11. Temperature accelerated interrupted stress measurements may also be performed on devices having 30nm SiO_2 gate oxides. Again, long-term high-temperature data can be perfectly fitted by a log-normal distribution.

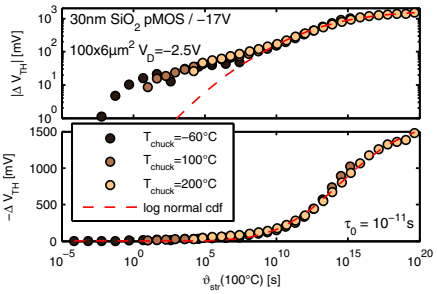

Fig. 12. In Fig. 11 we accelerated the test using one temperature and increasing stress times. Now, we always used 10s of stress but with increasing temperatures from T_{chuck} to the maximum achievable temperature with the poly-heater (upper plot of Fig. 8).

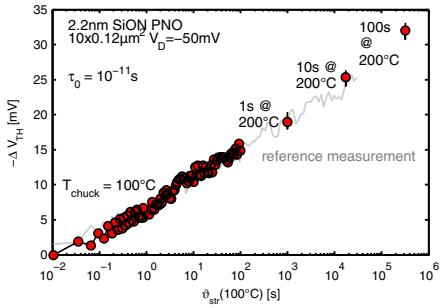

Fig. 13. Switching the device temperature to a higher value during stress accelerates the degradation. For small temperature switches, the correct interpretation of the acceleration can be demonstrated by a comparison with long-term OTF data (reference measurement).

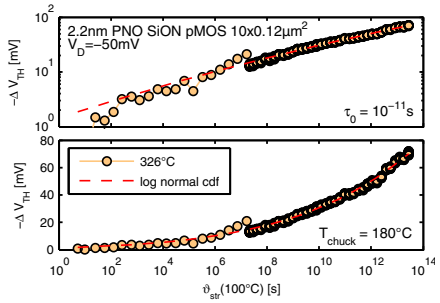

Fig. 14. Temperature accelerated OTF degradation data shows the same behavior as the data from interrupted stress measurements. Again, the saturation of the power law can be perfectly matched by a log-normal distribution of capture time constants.

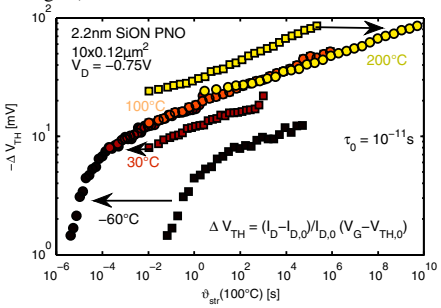

Fig. 15. Even without a poly-heater, the temperature-time concept is useful. The above shows OTF data measured at different chuck temperatures aligned onto a single degradation curve as a function of ϑ. The only difference is that a correction for the degradation in $I_{D,0}$ is needed. If $I_{D,0}$ is performed with 10ms delay at 30°C one may add the drift after approximately 60s at -60°C to the 30°C data. With this approach all degradation traces can be aligned.

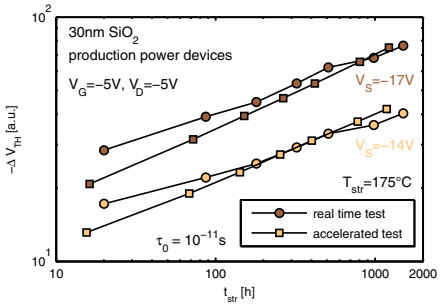

Fig. 16. Long-term NBTI degradation data of state-of-the-art power pMOS devices (gathered over several months) are reproduced within less than a day by using temperature accelerated testing with the poly-heater. Considering the vastly different measurement method used for the reference data, which included large (manual) delay times, which could only be roughly emulated by the poly-heater technique, the agreement is satisfactory.

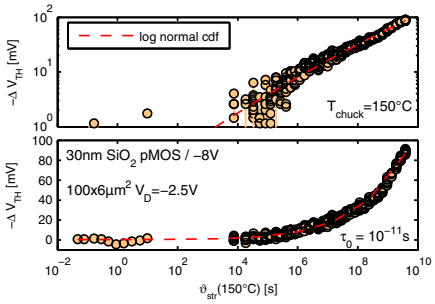

Fig. 17. The temperature-time concept makes it possible to access the degradation of power devices at use conditions. Depicted is an example of such a measurement where the degradation at -8V gate bias was achieved by accelerating with 300°C. We remark that based on the results in this paper, temperature acceleration is a far more promising approach to access lifetime data compared to conventional downscaling from large stress oxide fields.