Negative Bias Temperature Instability: Recoverable versus Permanent Degradation

Tibor Grasser
Christian Doppler Laboratory for TCAD
at the Institute for Microelectronics,
TU Wien, Gusshausstr. 27–29, A-1040 Wien, Austria

Ben Kaczer
IMEC
Kapeldreef 75, B-3001 Leuven, Belgium

***Abstract*— The analysis of negative bias temperature instability (NBTI) conventionally focuses on the stress phase where features like the power-law exponent, the temperature-dependence, and saturation of the observed threshold voltage shift have been extensively studied. As soon as the stress is removed, however, relaxation sets in, restoring at least some of the degradation. Although some studies on the relaxation phase have been presented, few authors have acknowledged the importance of the relaxation phase as a means of deepening our understanding of NBTI. We present a detailed analysis of NBTI relaxation, show that even at lower stressing voltages a permanent/slowly relaxing component is present, and demonstrate how this initially less dominant component might eventually determine the device lifetime.**

I. INTRODUCTION

The negative bias temperature instability (NBTI) is commonly regarded as one of the major reliability issues in highly scaled pMOSFETs [1]. It is normally expressed through the threshold voltage shift which increases with time under negative gate bias and at elevated temperatures. Characterization and modeling attempts have focused mainly on the stress phase, where the threshold voltage shift is frequently described using a power-law, $\Delta V_{\mathrm{th}}(t_{\mathrm{s}}) = A\, t_{\mathrm{s}}^{n}$. Many published models express the degradation behavior through a differential equation of the form

$$\frac{\partial V_{\mathrm{th}}(t_{\mathrm{s}})}{\partial t_{\mathrm{s}}} = \mathrm{ForwardRate} - \mathrm{BackwardRate}\ .$$

While the forward rate is assumed to strongly depend on the applied bias, the rate coefficient of the backward rate is taken as bias independent [2]. As a consequence, once the stress is removed, the backward rate dominates the relaxation of the threshold voltage shift, allowing an in-depth study.

In our previous analysis we found [3] that none of the published models seems to give satisfactory results during this relaxation phase, indicating a serious lack in our understanding of NBTI. In fact, since during the stress phase only the net effect of forward and backward rate is visible, any agreement with measurement based on an incorrectly modeled backward rate should be regarded with suspicion, making a clarification of this issue mandatory.

II. CHARACTERIZATION OF RELAXATION

It has frequently been argued that for the characterization of NBTI the on-the-fly method is the method of choice because it does not interrupt the stress during the measurement, thereby avoiding unintentional relaxation [2, 4–6]. However, the method suffers inherently from the difficulty of identifying the reference value for the threshold voltage shift, which is typically obtained after some unavoidable initial stress [7]. Furthermore, it is difficult to use during relaxation because the interface is in a different state compared to the stress phase, also because of the different position of the Fermi-level [3, 8].

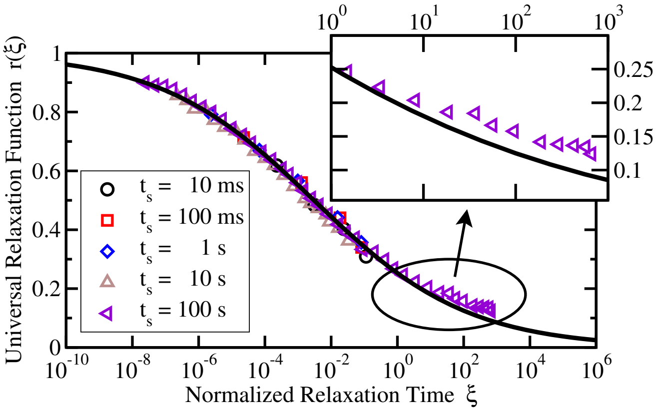

Fig. 1: Application of the universal relaxation scheme to the ultra-fast MSM data obtained by Reisinger *et al.* [8]. The individual data points can be mapped onto a universal curve, in this case (1). The inset highlights the slight deviation of the measurement data from the universal curve for $\xi > 10$, indicating a possible permanent component.

We therefore base our study on data obtained via conventional measurement-stress-measurement sequences (MSM) with the drain current at $V_{\mathrm{G}} \approx V_{\mathrm{th}}(t_{\mathrm{s}} = 0)$ monitored during relaxation [9]. Although the value of the threshold voltage shift at zero delay time is unknown, the method allows a straight forward determination of the most important reference value of the threshold voltage shift.

The most intriguing feature of NBTI relaxation is its universality [3, 10] which can be studied by (i) stressing the device using different stress times $t_{\mathrm{s},i}$, (ii) recording the relaxation $\Delta V_{\mathrm{th}}(t_{\mathrm{s},i}, t_{\mathrm{r}})$ during a certain interval of relaxation times t_{r}, (iii) normalizing $\Delta V_{\mathrm{th}}(t_{\mathrm{s},i}, t_{\mathrm{r}})$ to the last stress value $\Delta V_{\mathrm{th}}(t_{\mathrm{s},i}, 0)$, and (iv), normalizing the relaxation time to the stress time as $\xi = t_{\mathrm{r}}/t_{\mathrm{s},i}$. Then, all data obtained at different stress times fall on the same 'universal' curve, given by the universal relaxation function $r(\xi)$ [3]. This procedure can also be applied to data obtained from MSM measurements where the last (unknown) stress value $\Delta V_{\mathrm{th}}(t_{\mathrm{s},i}, 0)$ is obtained via extrapolation to $t_{\mathrm{r}} = 0$ using the universality (cf. Fig. 1). The exact functional form of r and its dependence on process

1-4244-1124-6/07/$25.00 ©2007 IEEE.

conditions, bias, and temperature remain to be clarified [3]. Nevertheless, close to perfect fits to experimental data have been obtained using the empirical relation [3, 9]

$$r(\xi) = 1/(1 + B\xi^{\beta}) \ . \tag{1}$$

The most important consequence of this universal relaxation behavior is that it allows one to reconstruct the 'true' degradation behavior based on standard delayed measurements [3].

In our previous study we have used existing NBTI relaxation data from [9] and data available in literature to convincingly confirm the universality [3]. It was found that only the very detailed data of Reisinger *et al.* [8] show a slight deviation from the universal behavior at very large relaxation times, see Fig. 1. To investigate this deviation we introduce a permanent component P [11, 12] in the overall observed degradation during stress, $S = \Delta V_{\mathrm{th}}$. We note that the permanent component could just as well be a very slowly relaxing component, however, as will be demonstrated, even more detailed data than presented here would be required to determine the exact nature of P.

The observed degradation during stress measured with a certain delay t_{M} will be written as

$$S_{\mathrm{M}}(t_{\mathrm{s}}, t_{\mathrm{M}}) = R_{\mathrm{M}}(t_{\mathrm{s}}, t_{\mathrm{M}}) + P(t_{\mathrm{s}}) \ , \tag{2}$$

where R_{M} gives the recoverable and P the permanent part of the degradation. The recovery depends on the relaxation time t_{r} and due to the universality can be written as $R_{\mathrm{M}}(t_{\mathrm{s}}, t_{\mathrm{M}}) = R(t_{\mathrm{s}})\, r(t_{\mathrm{M}}/t_{\mathrm{s}})$ [3]. In the following examples we assume P to follow a power-law, $P(t_{\mathrm{s}}) = A_{\mathrm{p}}\, t_{\mathrm{s}}^{n_{\mathrm{p}}}$, while the universal relaxation function will be assumed to be given by (1). Obviously, at zero delay, the true degradation would be given by

$$S(t_{\mathrm{s}}) = S_{\mathrm{M}}(t_{\mathrm{s}}, 0) = R(t_{\mathrm{s}}) + A_{\mathrm{p}}\, t_{\mathrm{s}}^{n_{\mathrm{p}}} \ . \tag{3}$$

Even if R follows a power law, S can only be approximated by a power-law when the permanent component P is of the same order of magnitude as the recoverable component R. Also, any measurement delay t_{M} will introduce an additional time-dependence into R_{M} and thus into S_{M}.

A. Experimental

PFET transistors with $W/L = 10\,\mu\mathrm{m}/0.5\,\mu\mathrm{m}$ were used. Nitrogen was incorporated with decoupled plasma nitridation (DPN) into in-situ steam generated (ISSG) 1.64 nm thick oxides. Post-nitridation anneal (PNA) in O_2 resulted in an SiON film with EOT of 1.4 nm [13]. Grouped Keithley 2602 instruments were used for electrical characterization. Linear I_{D} currents were measured at $V_{\mathrm{D}} = -50\,\mathrm{mV}$ and $V_{\mathrm{G}} = 300\,\mathrm{mV} \approx V_{\mathrm{th}}(0)$ and converted to threshold voltage shifts using the initial $I_{\mathrm{D}}V_{\mathrm{G}}$ curve [9].

B. Parameter Extraction Procedure

For N sets of relaxation data $S_{\mathrm{M}}(t_{\mathrm{s},i}, t_{\mathrm{r}})$ obtained at different stress times $t_{\mathrm{s},i}$ we determine the unknown parameters B, β, A_{p}, and n_{p} by minimizing the score function

$$\varepsilon_{\mathrm{t}} = \sum_{i=1}^{N} \int \left(\frac{r(t_{\mathrm{r}}/t_{\mathrm{s},i})}{r(t_{\mathrm{M}}/t_{\mathrm{s},i})} - \frac{S_{\mathrm{M}}(t_{\mathrm{s},i}, t_{\mathrm{r}}) - P(t_{\mathrm{s},i})}{S_{\mathrm{M}}(t_{\mathrm{s},i}, t_{\mathrm{M}}) - P(t_{\mathrm{s},i})} \right)^2 \mathrm{d}\log \xi_i \ .$$

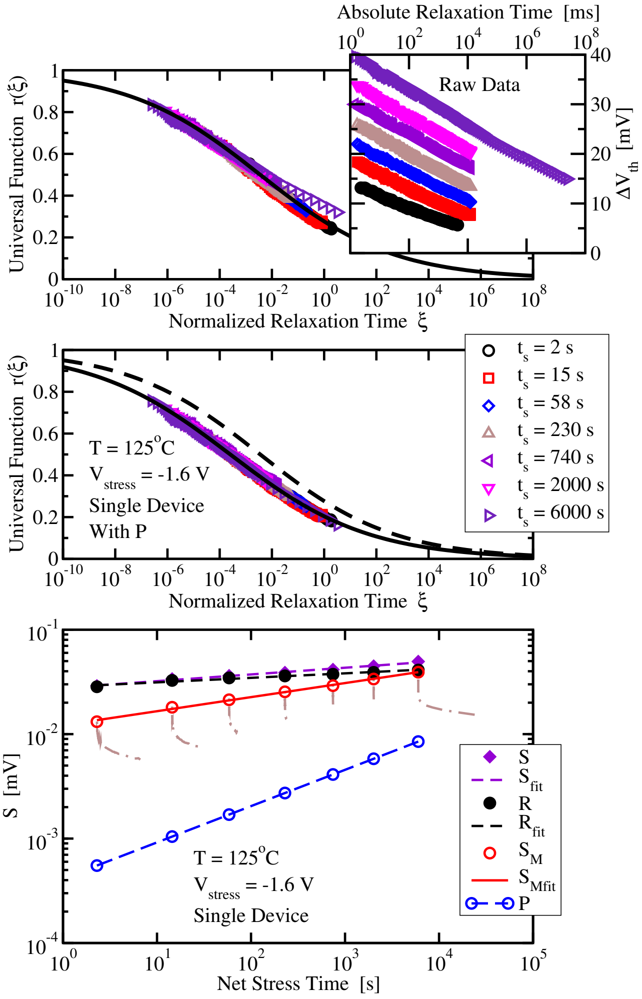

Fig. 2: *Top:* Universal relaxation plot for data recorded in a single MSM sequence at a stress voltage of $V_{\mathrm{G}} = -1.6\,\mathrm{V}$. The permanent component causes a visible deviation from the universal relaxation curve. In the inset the unscaled raw data is shown. *Middle:* By accounting for the permanent component, the universality is restored. The difference to the universal relaxation function extracted without P (dashed line) is already clearly visible. *Bottom:* The extracted degradation components as a function of the net stress time. At such a large stress voltage, the permanent component changes the slope of the corrected 'true' curve. Also schematically indicated on a relative time scale $t_{\mathrm{s},i} + t_{\mathrm{r}}$ is the relaxation data.

Note that the parameter extraction is independent of the functional form of R and that the final form of R is directly related to the measurement data through the universal relaxation relation as $R(t_{\mathrm{s}}) = R_{\mathrm{M}}(t_{\mathrm{s}}, t_{\mathrm{M}})/r(t_{\mathrm{M}}/t_{\mathrm{s}})$ [3].

C. Single Device Relaxation Measurement

One possibility to obtain a sequence $S_{\mathrm{M}}(t_{\mathrm{s},i}, t_{\mathrm{r}})$ is to stress a single device in an MSM sequence, thereby interrupting the stress at the predefined stress times $t_{\mathrm{s},i}$. We interrupt the stress phase 6 times to start a relaxation measurement at $t_{\mathrm{r}} \approx 1\,\mathrm{ms}$ and record data up to $t_{\mathrm{r}} \approx 10\,\mathrm{s}$, resulting in four decades of relaxation data for each stress time $t_{\mathrm{s},i}$. The last relaxation curve is recorded up to $t_{\mathrm{r}} \approx 2 \times 10^4\,\mathrm{s}$, giving seven decades in time. Note that $t_{\mathrm{s},i}$ is the *net* stress time without the relaxation phases [3].

A typical example of such a measurement is given in Fig. 2, where a relatively large stress voltage of $V_{\mathrm{s}} = -1.6\,\mathrm{V}$ ($E_{\mathrm{ox}} \approx 9.3\,\mathrm{MV/cm}$) causes a considerable permanent component.

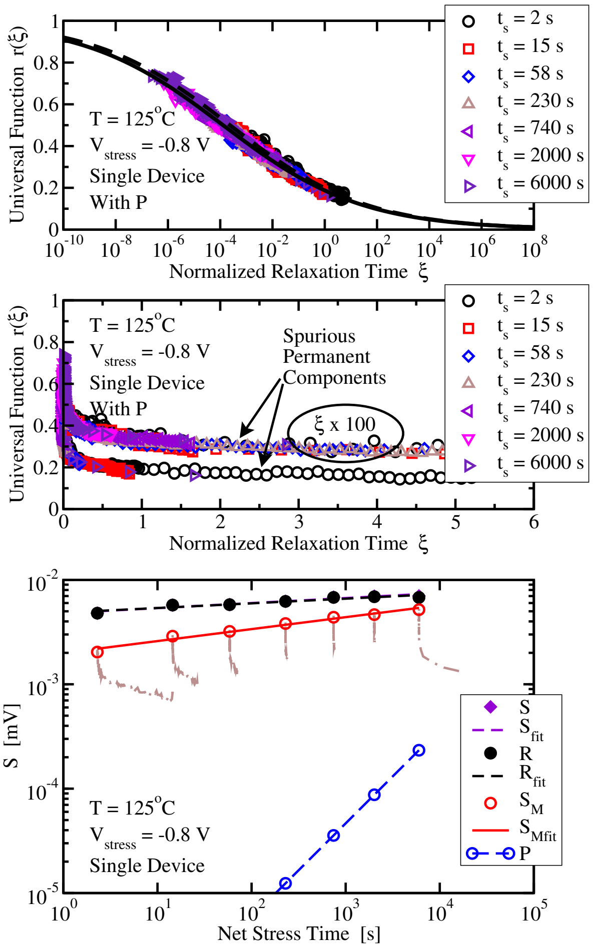

Fig. 3: *Top:* For a low stress voltage of $V_{\mathrm{G}} = -0.8\,\mathrm{V}$ the *apparent* influence of the permanent component is small and the universal relaxation function extracted without a permanent component (dashed line) is very similar to the full model. *Middle:* When the relaxation data is plotted on a lin-lin scale, the long-time tail of the relaxation could be mistaken for a permanent component. Note that the value of this 'spurious permanent component' depends on the time-interval chosen to plot the data and is thus completely arbitrary. *Bottom:* The extracted degradation components as a function of the net stress time. The permanent component does not influence the overall slope for $t_{\mathrm{s}} < 10^4\,\mathrm{s}$.

This also influences the slope of the corrected overall power-law slope n_{s}. For a low stress voltage of $V_{\mathrm{s}} = -0.8\,\mathrm{V}$ ($E_{\mathrm{ox}} \approx 3.6\,\mathrm{MV/cm}$), see Fig. 3, the permanent component is barely visible on the universal relaxation plot but can still be extracted from the data set. This is consistent with earlier reports which recommended an oxide field below $6\,\mathrm{MV/cm}$ for the investigation of NBTI [14].

D. Multiple Devices Relaxation Measurement

To increase the accuracy of the parameter extraction we have used several individual devices stressed with varying stress times $t_{\mathrm{s},i}$ and recorded the relaxation starting from $t_{\mathrm{r}} = 1\,\mathrm{ms}$ up to $t_{\mathrm{r}} \approx 10^5\,\mathrm{s}$, giving eight decades in time for each data set. Although these data allow for a more detailed study of the relaxation phase, small device variations result in a somewhat less smooth universal scaling as shown in Fig. 4. However, the increased relaxation time interval reveals that also for very small stress times the data deviates from the universal curve if no permanent component is considered.

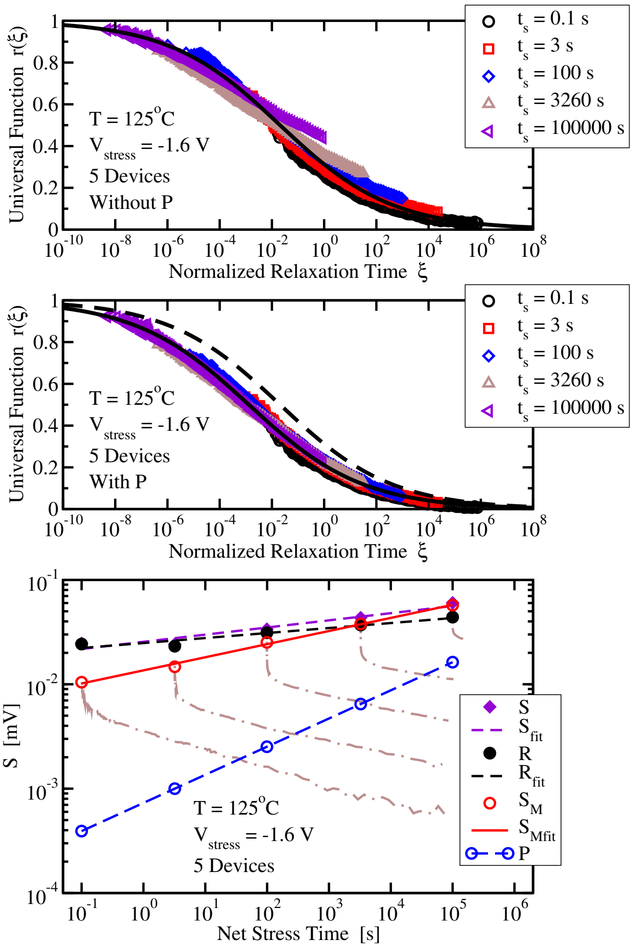

Fig. 4: Same as Fig. 2 but based on data from five individually stressed devices which considerably improves the sensitivity of the parameter extraction. Without consideration of a permanent component P, the deviation from the universal plot is now obvious even for very small stress times (top). After accounting for the permanent component, the universality is observed (middle).

III. DISCUSSION

The extracted parameters B, β, A_{p}, and n_{p} are shown in Fig. 5 as a function of the stress voltage. For the sake of comparison, also S, S_{M}, and R are approximated by a power-law as $A_{\mathrm{s}}\,t_{\mathrm{s}}^{n_{\mathrm{s}}}$, $A_{\mathrm{sM}}\,t_{\mathrm{s}}^{n_{\mathrm{sM}}}$, and $A_{\mathrm{r}}\,t_{\mathrm{s}}^{n_{\mathrm{r}}}$, respectively. The parameters of these power-law approximations are obtained by a linear least-squares fit on a log-log plot for the total interval of available stress times (cf. Figs. 2–4 bottom).

The voltage dependence of all power-law prefactors can be nicely fitted using expressions of the form $C|V_{\mathrm{s}}|^a$ while the exponents are found to be roughly independent of the stress bias. Most notably, the corrected power-law exponent of the recoverable component $n_{\mathrm{r}} \approx 0.05$ is considerably smaller than the frequently cited 0.15 obtained by OTF measurements [5] but in good agreement with some recent results [7].

With these parameters the influence of the measurement delay on the apparent power-law exponent n_{s} can be studied in unprecedented detail. Fig. 6 shows the dependence of n_{s} on the measurement delay t_{M}. Interestingly, this analysis predicts that even with ultra-fast measurements using $t_{\mathrm{M}} = 1\,\mu\mathrm{s}$, a 40% error in the slope is obtained and only a measurement

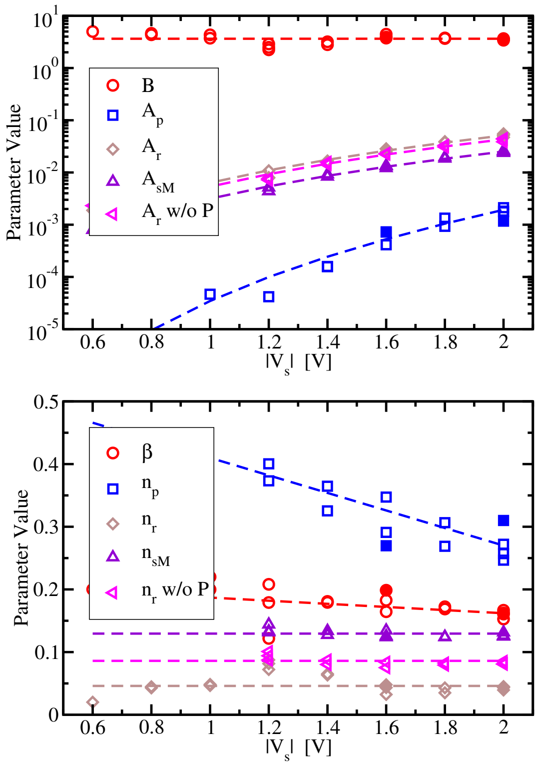

Fig. 5: *Top:* The extracted parameters of the model as a function of the stress voltage with the open symbols extracted from a single device measurement and the closed symbols from the multiple-device measurements. The line for the parameter B is the average value of 3.6 while the other lines are obtained by a power-law fit which delivers better results compared to an exponential fit. *Bottom:* The parameter β and the extracted power-law exponents as a function of the stress voltage. Here, the lines are obtained via a linear fit. Data with higher scatter at low V_{G} are excluded from the plot.

resolution in the pico-second range would be able to reveal the correct slope.

Furthermore, the model (3) can be used to more accurately estimate the device lifetime. Interestingly, even though the permanent component does not affect the power-law slope n_{s} during the measurement interval at the low stress bias of $-0.8\,\mathrm{V}$, it determines the lifetime, which is commonly considerably larger than the measurement interval. This can be clearly seen in Fig. 7 which shows that the lifetime at low voltages is determined by the permanent component, while the recoverable component influences the lifetime in the high voltage range. This apparently contradictory behavior is a consequence of the prefactor A_{r} being considerably larger than A_{p} and that at higher stress voltages the recoverable part alone causes the degradation to reach the failure criterion at very short times.

IV. CONCLUSION

We have studied the relaxation phase following NBT stress in detail, thereby identifying a universally relaxing component on top of a permanent or slowly relaxing component. It has been shown that at least seven decades of relaxation data are required for a reliable determination of the two components. If the available new data is limited to two decades, the permanent component becomes unidentifiable, leading to errors in the extracted parameters. Accurate parameters for the permanent

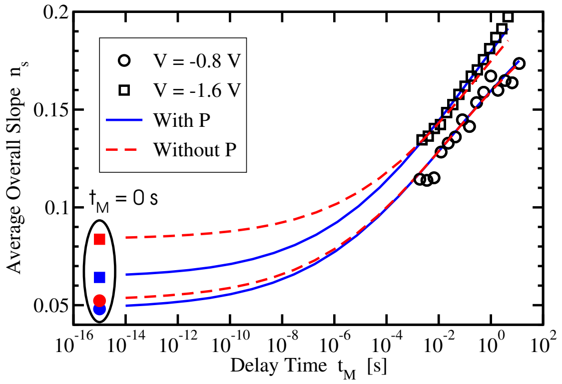

Fig. 6: Observed slope in a delayed measurement as a function of the measurement delay. The symbols are the measurement data while the lines give the extrapolation to the 'true' slope using our algorithm. Consideration of the permanent component P gives a smaller slope. For larger stress voltages the influence of P is considerably larger. Note that only for delay times in the pico-second range the true slope could be measured.

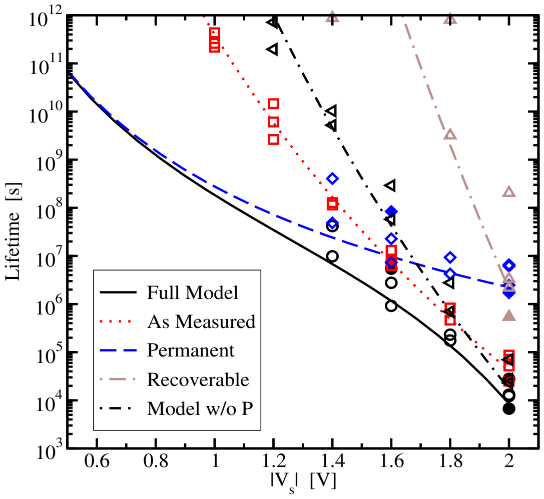

Fig. 7: Extracted lifetimes for a failure criterion of $S(t_{\mathrm{s}}) = S_{\mathrm{F}} = 100\,\mathrm{mV}$. The permanent component dominates the lifetime even at low stress voltages. Extraction of the lifetime without considering the permanent component or an extraction based on the as measured (delayed) data may result in a considerable overestimation of the lifetime. The lines are calculated using the fits in Fig. 5 while the symbols are derived from the symbols of Fig. 5.

component could be of fundamental importance for the prediction of the lifetime.

Furthermore, a detailed understanding of the involved components is mandatory for the development of refined NBTI models that not only capture the relaxation phase but lead to an improved description during the stress phase as well.

REFERENCES

[1] D. Schroder, Microelectr.Reliab. (2006), (online version).
[2] M. Alam *et al.*, Microelectr.Reliab. (2006), (online version).
[3] T. Grasser *et al.*, in *Proc. IRPS* (2007), pp. 1–13.
[4] M. Denais *et al.*, in *Proc. IEDM* (2004), pp. 109–112.
[5] D. Varghese *et al.*, IEEE Electron Device Lett. **26**, 572 (2005).
[6] A. Krishnan *et al.*, in *Proc. IEDM* (2005), pp. 688–691.
[7] C. Shen *et al.*, in *Proc. IEDM* (2006), pp. 333–336.
[8] H. Reisinger *et al.*, in *Proc. IRPS* (2006), pp. 448–453.
[9] B. Kaczer *et al.*, in *Proc. IRPS* (2005), pp. 381–387.
[10] M. Denais *et al.*, in *Proc. IRPS* (2006), pp. 735–736.
[11] V. Huard *et al.*, Microelectr.Reliab. **46**, 1 (2006).
[12] J. Zhang *et al.*, IEEE Trans.Electr.Dev. **51**, 1267 (2004).
[13] B. Kaczer *et al.*, in *Proc. INFOS* (2005), pp. 1–4.
[14] S. Mahapatra *et al.*, Microelectr.Eng. **80**, 114 (2005).