# Hydrogen-Related Volatile Defects as the Possible Cause for the Recoverable Component of NBTI

*T. Grasser*[◇], *K. Rott*[•], *H. Reisinger*[•], *M. Waltl*[◇], *P. Wagner*[◇], *F. Schanovsky*[◇], *W. Goes*[◇], *G. Pobegen*[†], *B. Kaczer*[○]

[◇]*TU Wien, Vienna, Austria* [•] *Infineon Munich, Germany* [†] *KAI, Austria* [○] *IMEC, Belgium*

## Abstract

The recently suggested time-dependent defect spectroscopy (TDDS) has allowed us to study the recoverable component of NBTI at the single-defect level. To go beyond our previous efforts, we have performed a long-term TDDS study covering also the kilo-second time window. We found that *even in this extended window NBTI recovery is due to a collection of first-order reactions*. In particular, there is *no trace of a diffusion-limited process as assumed in the reaction-diffusion model*. Most intriguingly, the responsible traps show various degrees of volatility, that is, they can disappear and reappear. Our observations lend strong support to the idea that the *recoverable component of NBTI is due to hydrogen-related defects* which are active when a hydrogen atom is at the defect site and inactive when not.

## Introduction

A considerable amount of information on the recoverable component of BTI [1–3] has recently been obtained on the single-defect level using the time-dependent defect spectroscopy [4–9] Because of the considerable experimental effort involved, stress and recovery times in our previous TDDS experiments were usually limited to stress times of about $t_s = 1\,\mathrm{s}$ and relaxation times $t_r = 1\,\mathrm{ks}$, respectively. In this window, NBTI recovery was found to be due to a collection of (roughly) independent first-order processes. We have postulated that this is also the case for larger stress times [10, 11] but never explicitly proven it using individual defect data. Our claim was recently challenged and it was argued that recovery after 1 s is due to a traditional reaction-diffusion mechanism [12, 13].

In order to demonstrate the correctness of our assertion, we extend the experimental TDDS window to $t_s = t_r = 1\,\mathrm{ks}$ to better cover what is normally considered the recoverable component $R$ [14]. In agreement with previous studies, we find that even in this much larger window, recovery is solely due to a collection of first-order processes and, in particular, that there is no diffusion-limited process. Furthermore, while behaving similarly to previously analyzed defects, *defects with larger time constants* are found to *show an even stronger volatility*, in the sense that they completely disappear and reappear over extended time intervals. Given the regularity with which this occurs, we conclude that this is not just an occasional second-order oddity but an essential feature of the defects constituting $R$. Since defects can rapidly disappear and reappear, we argue that previously suggested H related defects, such as hydrogen bridges [15, 16] and strained oxygen bridges [17–20] are very likely candidates for $R$.

## The Long-Term TDDS Setup

In a TDDS setup, a nanoscale device is repeatedly stressed and recovered using fixed stress/recovery voltages, $V_G^H$ and $V_G^L$ (Fig. 1). Since such devices contain only a handful of defects, the recovery of single defects is visible as discrete steps in the recovery trace (Fig. 2). While these discrete steps are of a characteristic height $\eta$, the charge emission events occur at exponentially distributed times $t_e$ (Fig. 4). The extracted $\{t_e, \eta\}$ pairs are collected in the spectral map (Fig. 5) where they produce clusters, clear fingerprints of individual defects. If the experimental window is wide enough and defect clusters do not overlap in the spectral map, then the average emission time can be estimated as $\tau_e^L = \tau_e(V_G^L) = \sum t_e / n_e$, with $n_e$ being the number of emission events. For larger stress times, however, clusters do overlap and the emission time may be larger than the maximum relaxation time $t_r^{max}$, resulting in truncated clusters. In such cases some care is required for the estimation of $\tau_e^L$. Furthermore, if $\tau_e \approx t_r^{max}$, the defect may not emit its charge and may thus be already charged at the beginning of the next stress phase. Such traces must be discarded for the estimation of $\tau_c^H = \tau_c(V_G^H)$. Finally, for long recovery traces, defects with $\tau_c^L = \tau_c(V_G^L) \lesssim t_r^{max}$ may spontaneously capture a charge as well as emit it again. In other words, the same defect can contribute to $R$ and produce random telegraph noise (RTN).

## Extraction Methodology

In order to correctly estimate the defect time-constants from truncated clusters, it has to be determined whether a defect was charged prior to stress or not. This can be done by monitoring the last value of each trace, $L$ (Fig. 3). For short stress times, $L$ is reminiscent of an RTN signal, while with harsher stress conditions a more or less permanent contribution builds up. A wealth of information can be extracted from $L = \sum \eta_i$ by decomposing it into its constituents using step-heights of known defect seen at e.g. higher temperatures/biases and additional, yet unknown, defects with $\tau_e^L > t_r^{max}$ (Fig. 6). A simplified algorithm for the estimation of $\tau_c^H$ and $\tau_e^L$ based on these ingredients is sketched in Fig. 7. Without considering $L$, $\tau_c^H$ of defects with $\tau_e^L > t_r^{max}$ cannot be extracted.

## Results

While in previous TDDS studies it has already been observed that recovery is consistent with a nonradiative multiphonon charge exchange mechanism [21–23], it has not been demonstrated yet that this is also the case for larger stress times. As such, it has been recently speculated that NBTI recovery after longer stress is due to the back-diffusion of hydrogen, using the reaction-diffusion (RD) model [13]. While such a process also results in discrete steps in the recovery traces, the distribution of these emission times would be *logistic* rather than *exponential*, and would show a moving mean (Fig. 8), which is in stark contrast to our data.

Experimental long-term TDDS data are shown in Fig. 9, which appear, with the exception of an increased noise level due to a larger number of defects, very similar to short-term TDDS data. The most important difference, however, is that defects show much *stronger signs of volatility*, that is, they regularly disappear and reappear. This volatility can go largely unnoticed, with defects disappearing only occasionally (such as defects A1 and A6), or it can be so severe that no capture time can be extracted (while emission clusters always remain exponential). When active, the defect follows first-order kinetics, when inactive, the defect is neutral and cannot be charged. As an example, Fig. 10 shows

the volatility of defect 7 of device A. At 125 °C, $\tau_c^L = 240\,s$ and $\tau_e^L = 2\,ks$, meaning that the defect can produce RTN during the recovery trace. However, during stress $\tau_c^H(-1.5\,V) = 450\,\mu s$ and $\tau_e^H > 2\,ks$, meaning that the defect must always be charged after a 10 s stress. Still, for considerable amounts of time, the defect does not charge at all. While the times the defect is active or inactive are too large to be analyzed statistically as a function of bias in the available data set, defect A7 appears to be less likely to be active at higher temperatures. Similar observations are made for other defects with large $\tau_e^L$, see Fig. 10 and Fig. 11. Fig. 12 demonstrates that the emission events are exponentially distributed also for the defects with large $\tau_c^H$, while Fig. 13 shows that capture follows first-order kinetics. Fig. 14 shows $\tau_c^H$ and $\tau_e^L$ of the defects of device A and C with $\tau_e^L > 1\,s$. As with the previously studied defects from short-time TDDS, $\tau_c^H$ and $\tau_e^L$ appear uncorrelated with $\tau_c^H$ showing a strong exponential bias dependence.

## Discussion

The above results suggest that the defects responsible for NBTI recovery are of the *same origin in the window* $1\,\mu s \ldots 1\,ks$. Charging/discharging of these defects follows first-order kinetics and is consistent with nonradiative multiphonon theory [21–24]. Particularly the fact that the defects require a certain minimum $V_G$ to become chargeable (Fig. 13) suggests that we are dealing with oxide rather than interface defects.

The most *intriguing observation is the various degrees of volatility*. Since an active defect always has the same impact on $\Delta V_{th}$, the defect must be at a fixed geometrical position along the channel, ruling out mobile species as defect candidates. However, since the defects can transform into a neutral and nonchargeable form, the defect is unlikely to consist only of Si, O, and N atoms, which are rather immobile under typical NBTI conditions. Also, if an oxygen vacancy moved by hopping of a neighboring O atom, an oxygen vacancy would be created at the neighboring site, which should be observable as a defect with different properties. This, however, we have never observed. Rather, our results are consistent with the idea that *hydrogen can bind to suitable defect-host site, transforming it into an electrically active defect*. The defect-hydrogen complex could be charged and discharged until the hydrogen atom escapes, a mechanism suggested previously [17, 20, 25]. If the defect recaptures a H atom, the cycle starts afresh [17], see Fig. 17, which is essentially a 'hydrogenated' version [15] of the HDL model based on the E' center [26–29].

This explanation should be contrasted with the observation that recent data shows $R$ to only weakly depend on the H concentration [30, 31]. Still, NRA studies have demonstrated [32] that there is a massive amount of H stored in the near-interfacial regions ($> 10^{14}\,cm^{-2}$) suggesting that it is the availability of suitable defect-host sites rather than the H concentration that determines $R$. Fig. 15 shows that the overall loss of $R$ during NBTI stress/recovery cycling on large area devices depends significantly on the amount of H stored in the interfacial region.

Finally, it could be argued that while these first-order processes contribute to $R$, there is another component (like that of the RD model), which is actually more important. The proof that this is not so is given in Fig. 16, where the capture and emission times extracted via TDDS are *sufficient to completely explain* the measured average $\Delta V_{th}$ *recovery. This does not leave any room for any other hidden significant recovery mechanism* and, consequently, also implies that *degradation cannot be due to a reaction-diffusion process.*

## Conclusions

Using long-term TDDS experiments, we have investigated the traps constituting the recoverable component $R$ of NBTI. These traps behave like their faster counterparts, as *charging/discharging is reaction-limited* and *strongly temperature activated*. In particular, there is no sign of a diffusion-limited process, *ruling out the RD model* as a valid explanation. Interestingly, these traps show *various degrees of volatility*, that is, they can *disappear and reappear*. Our observations lend strong support to the idea that $R$ is due to *hydrogen-related defects* which are active when a H is at the defect site and inactive when not.

## Acknowledgments

The research leading to these results has received funding from the FWF project n°23390-M24 and the European Community's FP7 project n°261868 (MORDRED). Valuable discussions with P. Lenahan, J. Campbell, A. Shluger, and V. Afanas'ev are gratefully acknowledged.

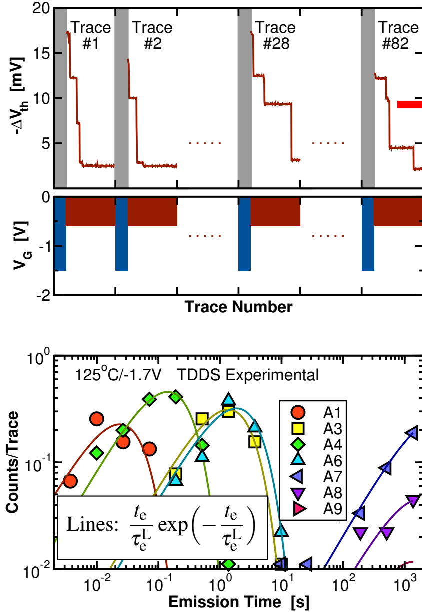
## References

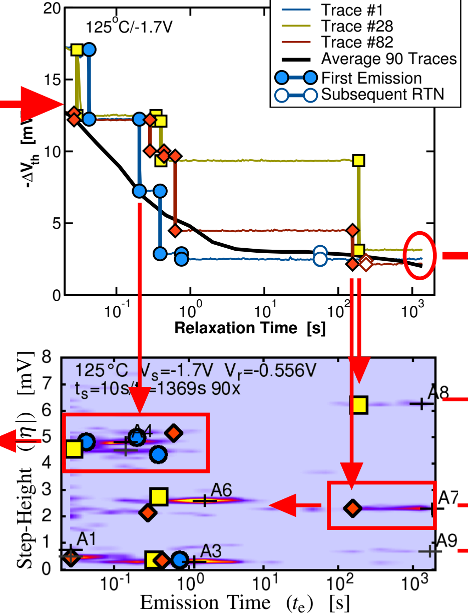
[1] S. Rangan, N. Mielke, and E. Yeh, in *IEDM* (2003), pp. 341–344.
[2] V. Huard, M. Denais, and C. Parthasarathy, MR **46**, 1 (2006).
[3] T. Aichinger, M. Nelhiebel, and T. Grasser, in *IRPS* (2009), pp. 2–7.
[4] C.-T. C. T. Wang AND, C.-J. Tang, C.-W. Tsai, H. Wang, M.-H. Chi, and D. Tang, T-ED **53**, 1073 (2006).
[5] T. Grasser, H. Reisinger, P.-J. Wagner, W. Goes, F. Schanovsky, and B. Kaczer, in *IRPS* (2010), pp. 16–25.
[6] M. Toledano-Luque, B. Kaczer, E. Simoen, P. J. Roussel, A. Veloso, T. Grasser, and G. Groeseneken, ME **88**, 1243 (2011).
[7] M. Toledano-Luque, B. Kaczer, P. Roussel, T. Grasser, G. Wirth, J. Franco, C. Vrancken, N. Horiguchi, and G. Groeseneken, in *IRPS* (2011), pp. 364–371.
[8] M. Toledano-Luque, B. Kaczer, J. Franco, P. Roussel, T. Grasser, T. Hoffmann, and G. Groeseneken, in *VLSI Symp.* (2011).
[9] T. Grasser, K. Rott, H. Reisinger, P.-J. Wagner, W. Goes, F. Schanovsky, M. Waltl, M. Toledano-Luque, and B. Kaczer, in *IRPS* (2013), pp. 2D.2.1–2D.2.7.
[10] B. Kaczer, T. Grasser, P. Roussel, J. Franco, R. Degraeve, L. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger, in *IRPS* (2010), pp. 26–32.
[11] T. Grasser, P.-J. Wagner, H. Reisinger, T. Aichinger, G. Pobegen, M. Nelhiebel, and B. Kaczer, in *IEDM* (2011), pp. 27.4.1–27.4.4.
[12] S. Mahapatra, A. Islam, S. Deora, V. Maheta, K. Joshi1, A. Jain, and M. Alam, in *IRPS* (2011), pp. 614–623.
[13] S. Mahapatra, N. Goel, S. Desai, S. Gupta, B. Jose, S. Mukhopadhyay, K. Joshi, A. Jain, A. Islam, and M. Alam, T-ED **60**, 901 (2013).
[14] T. Aichinger, M. Nelhiebel, and T. Grasser, MR **48**, 1178 (2008).
[15] J. Conley Jr. and P. Lenahan, T-NS **40**, 1335 (1993).
[16] P. Blöchl, PRB **62**, 6158 (2000).
[17] J. de Nijs, K. Druijf, V. Afanas'ev, E. van der Drift, and P. Balk, APL **65**, 2428 (1994).
[18] V. Afanas'ev and A. Stesmans, PRL **80**, 5176 (1998).
[19] J. M. Soon, K. P. Loh, S. S. Tan, T. P. Chen, W. Y. Teo, and L. Chan, APL **83**, 3063 (2003).
[20] A. Edwards, PRB **69**, 125318 (2004).
[21] K. Huang and A. Rhys, Proc.R.Soc.A **204**, 406 (1950).
[22] C. Henry and D. Lang, PRB **15**, 989 (1977).
[23] A. Stoneham, Rep.Prog.Phys. **44**, 1251 (1981).
[24] F. Schanovsky, W. Goes, and T. Grasser, JVST B **29**, 01A2011 (2011).
[25] M. Houssa, V. Afanas'ev, A. Stesmans, M. Aoulaiche, G. Groeseneken, and M. Heyns, APL **90**, 043505 (2007).
[26] A. Lelis and T. Oldham, T-NS **41**, 1835 (1994).
[27] J. Conley Jr., P. Lenahan, A. Lelis, and T. Oldham, T-NS **42**, 1744 (1995).
[28] J. Campbell, P. Lenahan, C. Cochrane, A. Krishnan, and S. Krishnan, T-DMR **7**, 540 (2007).
[29] J. Ryan, P. Lenahan, T. Grasser, and H. Enichlmair, in *IRPS* (2010), pp. 43–49.
[30] V. Huard, in *IRPS* (2010), pp. 33–42.
[31] G. Pobegen, M. Nelhiebel, and T. Grasser, in *IRPS* (2013), pp. XT.10.1–XT.10.6.
[32] Z. Liu, S. Fujieda, H. Ishigaki, M. Wilde, and K. Fukutani, ECS T. **35**, 55 (2011).
[33] Z. Teo, D. Ang, and K. See, in *IEDM* (2009), pp. 737–740.
[34] D. Ang, Z. Teo, T. Ho, and C. Ng, T-DMR **11**, 19 (2011).
[35] M. Duan, J. Zhang, Z. Ji, W. Zhang, B. Kaczer, S. De Gendt, and G. Groeseneken, EDL **33**, 480 (2012).
[36] T. Aichinger, M. Nelhiebel, S. Einspieler, and T. Grasser, T-DMR **10**, 3 (2010).
[37] F. Schanovsky, W. Goes, and T. Grasser, in *SISPAD* (2013), pp. 1–4.

**Fig. 1:** TDDS example: Device $A$ is stressed at $V_G^H = -1.7\,V/125\,°C$ for $t_s = 10\,s$ and recovered at $V_G^L = -0.58\,V$ for $t_r = 1389\,s$. This cycle is repeated 90 times to extract the capture time under stress, $\tau_c^H$, and the emission time under recovery bias, $\tau_e^L$.

**Fig. 2:** Each recovery trace is analyzed for discrete steps of height $\eta$ occurring at emission time $t_i$. Each step corresponds to the capture/emission of a single hole. The first emission is recorded in the spectral map below, all subsequent ones are treated as RTN.

**Fig. 3:** The last value $L$ of each trace for various stress times. For short stress times ($\lesssim 1\,s$), $L$ fluctuates but does not increase. For larger stress times, defects with $\tau_e^L > t_r^{max}$ become charged, cannot recover, and accumulate.



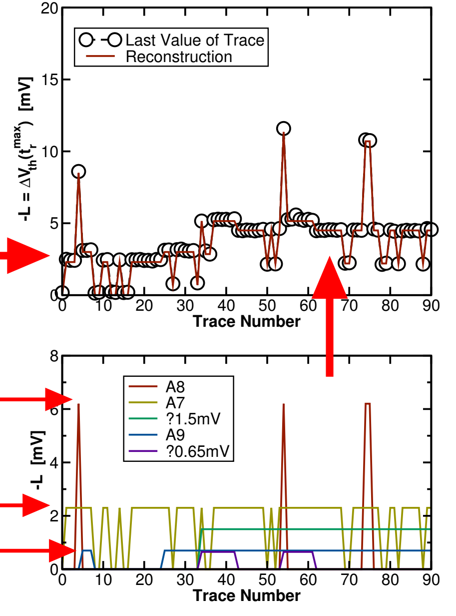




**Fig. 4: All** experimentally observed emission distributions are **exponential**, shown above for the defects of device A visible in Fig. 5. The mean of these distributions, $\bar{\tau}_e$, is independent of $t_s$, consistent with a **first-order reaction-limited** process.

**Fig. 5:** Discrete clusters appear in the **spectral map** shown above, each corresponding to a single defect. In order to separate defects, emission events are only considered inside a filtered region (red boxes), given by $(t_1, \eta_1, t_2, \eta_2)$. Note that some clusters with large $\tau_e^L$ are truncated (A7-A9), cf. Fig. 7.

**Fig. 6:** Using the best-fitting combination of step-heights of known defects (A7 - A9) and two unknown defects with $\tau_e^L > t_r^{max}$ and $\eta_x = 1.5\,mV$ and $0.65\,mV$, $L = \sum \eta_i$ can be reconstructed (see top) and its contributors identified (bottom).

| $m_c/n_c$ | $m_e/n_e$ | $m_l/n_l$ | |
|---|---|---|---|
| 0/+1 | 0/0 | 0/0 | Could have captured |
| +1/+1 | +1/+1 | 0/0 | Captured and emitted |
| +1/+1 | 0/0 | +1/+1 | Captured but did not emit |
| 0/0 | +1/+1 | 0/0 | Was already captured, emitted now |
| 0/0 | 0/0 | +1/+1 | Was already captured, did not emit |

$$\bar{p}_e = \frac{n_e}{n_e + n_l}$$

$$\bar{t}_e = \frac{1}{n_e}\sum t_{e,i}$$

$$\bar{p}_e \approx \int_{t_1}^{t_2} p(t_e)\,dt_e \quad (1)$$

$$\bar{t}_e \approx \int_{t_1}^{t_2} t_e\, p(t_e)\,dt_e \quad (2)$$

$$p(t_e) = \frac{1}{\tau_e^L}\exp\left(-\frac{t_e}{\tau_e^L}\right)$$

**Fig. 7:** Simplified version of the algorithm used to extract the time constants of defects which have a significant number of emission events outside the filter window $t_1 < t_e < t_2$ (red boxes in Fig. 5). The most important cases for $t_1 \ll \tau_e^L$ and $t_2 = t_r^{max}$ are shown to the left. The counters $m_x$ (number of events) and $n_x$ (number of cases where event could have occurred) are incremented as shown by the table in the middle based on the extracted steps and the information provided by the decomposition of $L$ (cf. Fig. 6). The average emission time $\tau_e^L$ is estimated by solving the simple nonlinear equations (1) and (2). Of course, for $t_1 \ll \tau_e^L \ll t_2$, $\bar{t}_e \to \tau_e^L$ as assumed previously. The complete algorithm also considers the number of RTN events, see Fig. 10. In the simplest case the capture time $\tau_c$ is estimated by inverting $1 - \exp(-t_s/\tau_c^H) = m_c/n_c$. This is done for many stress times $t_s$ to minimize the error in $\tau_c$, see also Fig. 13.

**Fig. 8:** *In complete contradiction* to the experiment, the **RD model** predicts very wide clusters where $\log(t_e/t_s)$ is logistically distributed with parameter $s \approx 2$. These distributions are $\approx 3\times$ too wide and their mean shifts with $t_s$, while the experimental clusters remain fixed. Furthermore, the RD clusters depend on $t_r^{max}$, because the H atoms cannot diffuse back completely. Apart from the trivial effect of truncation, experimental clusters are independent of $t_r^{max}$.

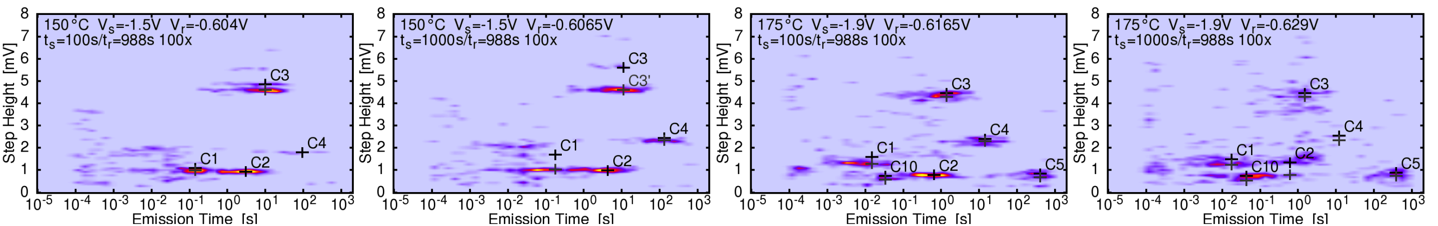



**Fig. 9:** Even at longer stress times (100 s and 1 ks) and higher temperatures, 150 °C (left) and 175 °C (right), **all** clusters are exponential. Due to the increasing number of defects contributing to the emission events and $L$, the data becomes noisier with increasing stress bias, temperature, and time. With increasing stress, **all** defects show **signs of volatility** implying that the defects could be related to a species which can easily form and re-arrange bonds, such as H. For instance, at 175 °C the intensity of the C4 cluster is much weaker at $t_s = 1000\,s$ compared to 100 s. This volatility often occurs at timescales much larger than typical experiments (couple of days) and appears completely erratic. What has been previously reported as an occasional curiosity [5] is in fact an *essential feature of the defects.*

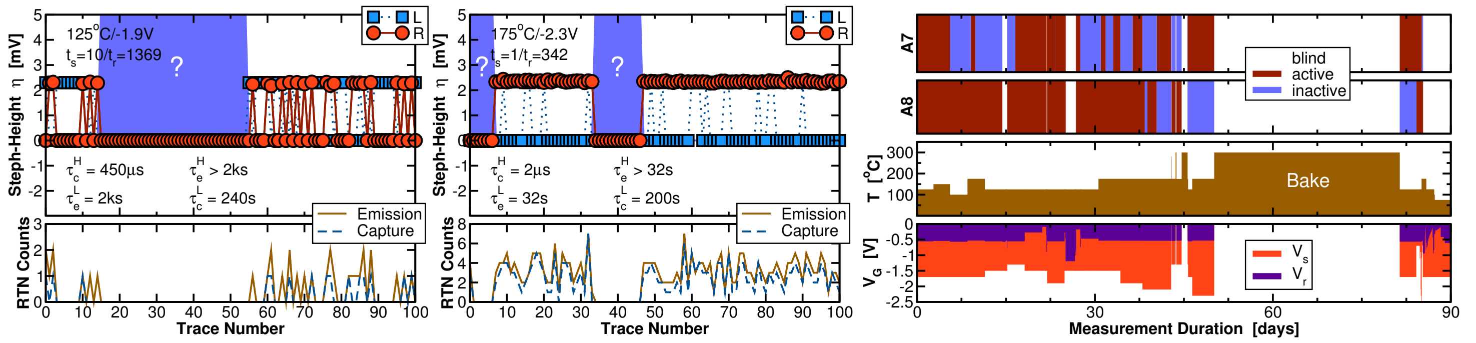



**Fig. 10:** Defects show varying degrees of volatility. As an example, the activity of defect A7 is shown above at 125 °C (**left**) and 175 °C (**middle**). Since the capture time at stress bias ($\tau_c^H$) is in the microsecond range and the emission time $\tau_e^H$ sufficiently large, after a long stress the capture probability $\tau_e^H/(\tau_e^H+\tau_c^H)$ is nearly 1. Still, for certain periods of time the defect does not capture a charge and is also not already charged ($L$), it has become *inactive*. A defect is considered active when it emits a charge in a trace or when it contributes to $L$. In addition, when $\tau_e^L \approx \tau_c^L$, the defect must produce RTN at $V_G^L$ in sufficiently long relaxation traces. **Right**: The activity of defects A7 and A8 monitored over a long period. After a 1 month-long bake step at 300 °C, the previously inactive defects are reactivated, cf. Fig. 15.

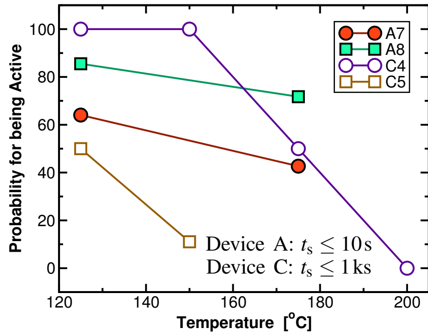



**Fig. 11:** The activity of defects is here defined as the ratio of the active time over the total observed time. At higher temperatures, the defects are more likely to be inactive. Also, the two defects revealed in the long-term TDDS measurements on device C are also more prone to disappear.

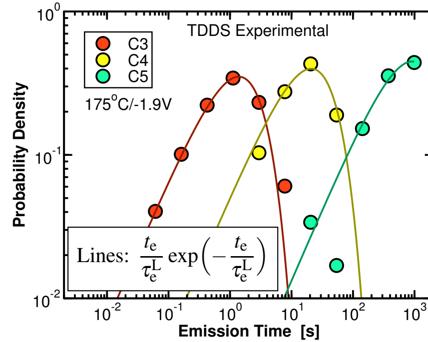



**Fig. 12:** As in Fig. 4, when the defect is active, **all** experimentally observed emission distributions with large emission times are **exponential**, shown above for three defects of device C. The mean emission times $\bar{\tau}_e$ are independent of $t_s$. This is consistent with a **first-order reaction-limited** process but *inconsistent with the RD model.*

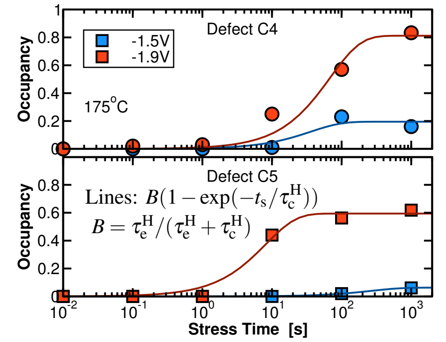



**Fig. 13:** In addition, **all** experimentally observed occupancies are **exponential** in $t_s$, shown above for the defects with the largest capture times, C4 and C5, again consistent with a **first-order reaction-limited** process. The strong bias dependence of the maximum occupancy indicates that these **defects are not directly at the interface but in the oxide**.

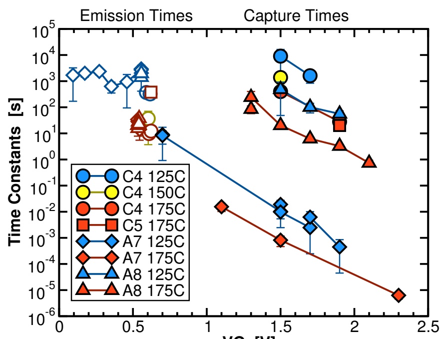



**Fig. 14:** The extracted capture and emission times for defects with large $\tau_e^L$. Due to the limited activity of these defects, the extraction was not always possible. However, when the defects are active, they behave in a very similar manner as those studied previously which only disappear occasionally, demonstrating that these oxide defects are the dominant contributor to the recoverable component of NBTI.

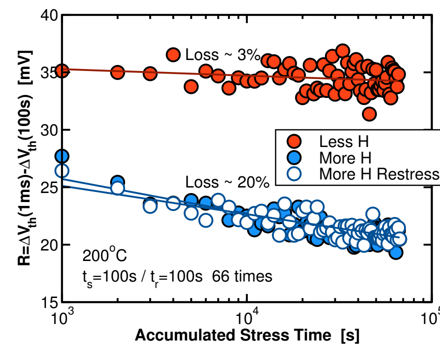
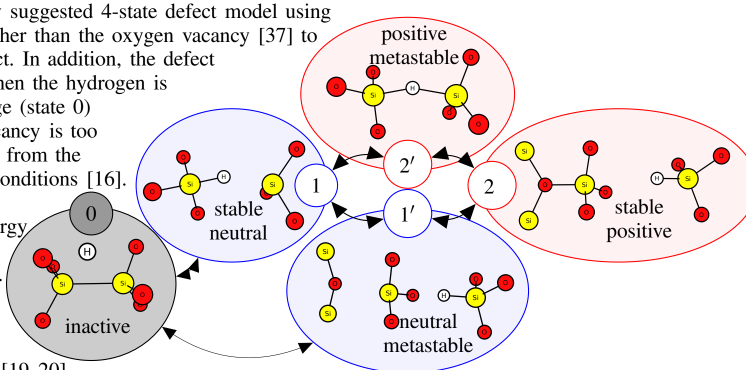
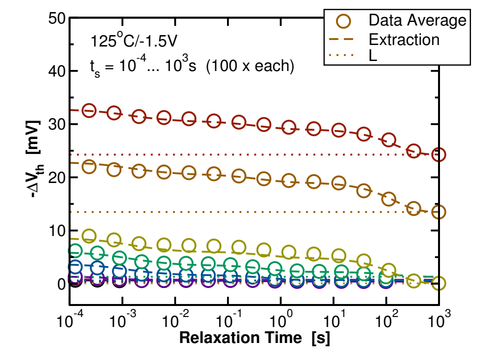



**Fig. 15:** On large-area devices, a loss in the recoverable component has been observed by various groups in numerous technologies [5, 11, 33–35]. Interestingly, in wafer splits with different H concentrations in the gate stack, it is observed that $R$ disappears faster in wafers with higher H content. A 1min bake at 400 °C with an in-situ polyheater [36] demonstrates that the procedure is repeatable.

**Fig. 17:** Our previously suggested 4-state defect model using the hydrogen bridge rather than the oxygen vacancy [37] to describe an active defect. In addition, the defect can become inactive when the hydrogen is removed from the bridge (state 0) because the oxygen vacancy is too stable to capture a hole from the substrate under NBTI conditions [16]. However, by capturing an H atom [15], its energy level moves closer to the Si bandgap [16, 37]. A similar state diagram can be constructed with the strained oxygen bridge [19, 20].



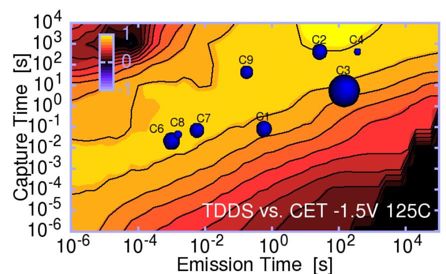




**Fig. 16: Top**: Using the time constants extracted from the long-term TDDS data, it is possible to **fully** reconstruct the average recovery trace (expectation value) as $\Delta V_{th} = \sum B_i(1-\exp(-t_s/\tau_{c,i}^H))\exp(-t_r/\tau_{e,i}^L)$ for all $t_s$ and $t_r$ with $B_i = \eta_i\tau_{e,i}^H/(\tau_{e,i}^H+\tau_{c,i}^H)$. **Bottom**: The extracted values of $\tau_c^H$ and $\tau_e^L$ shown against a CET map extracted from the same technology [11], which is fully consistent with this particular configuration observed by TDDS in device C. The size of the dots is given by $B_i$, the maximum impact on $\Delta V_{th}$.