Superior NBTI in High-k SiGe Transistors–Part I: Experimental

M. Waltl, G. Rzepa, A. Grill, W. Goes, J. Franco, B. Kaczer, L. Witters, J. Mitard, N. Horiguchi, and T. Grasser

***Abstract*— SiGe quantum-well pMOSFETs have recently been introduced for enhanced performance of transistors. Quite surprisingly, a significant reduction in negative bias temperature instability (NBTI) was also found in these devices. Furthermore, a stronger oxide field acceleration of the degradation in SiGe devices compared with Si devices was reported. These observations were speculated to be a consequence of the energetical realignment of the SiGe channel with respect to the dielectric stack. As these observations were made on large-area devices, only the average contribution of many defects to NBTI could be studied. In order to reveal the microscopic reasons responsible for the improved reliability, a detailed study of single defects is performed in nanoscale devices. To provide a detailed picture of single charge trapping, the step-height distributions for different device variants are measured and found to follow a unimodal and bimodal distribution. This finding suggests two conducting channels, one in the SiGe and one in the thin Si cap layer. We, furthermore, demonstrate that similar trap depth distributions are present among the device variants supported by a similar stress bias dependence of the capture times of the identified single defects. We conclude that NBTI is primarily determined by the dielectric stack and not by the device technology.**

***Index Terms*— Negative bias temperature instability (NBTI), pMOSFET, reliability, SiGe, time-dependent defect spectroscopy (TDDS).**

I. Introduction

MODERN transistor technologies employ gate stacks consisting of a silicon-dioxide SiO_2 insulating layer (IL) at the channel/oxide interface and a high-k layer (HK) of Hafniumoxide HfO_2 between the IL and the metal gate contact [1]–[4]. In such devices, the bias temperature instabilities (BTIs) are one of the most critical reliability issues. BTI is typically studied in large-area devices, which contain a large number of defects. There, BTI measurements record the *averaged* response of all defects [5]. In contrast, in a nanoscale transistor, there are only a few defects present, and their charging and discharging are clearly visible as discrete steps, for instance in the drain current. Thus the individual constituents to the degradation can be studied, for instance, by employing the time-dependent defect spectroscopy (TDDS) [6]–[8]. During TDDS measurements, the defects are repeatedly subjected to charging and discharging bias conditions and their stochastic response is recorded. As such, the TDDS can detect and monitor a single defect and extract their capture and emission times as a function of varying bias conditions and the device temperature.

So far, a large number of single defect studies have been successfully performed on nanoscale SiON transistors subjected to negative BTI (NBTI) [6], [9] and positive BTI stress [10] using the TDDS. Much less, however, is presently known about the behavior of single defects in high-k devices [11]–[13], particularly those with a SiGe channel.

In recent reports, it has been shown that the detrimental impact of NBTI on the performance of devices using HK gate stacks can be significantly reduced by using SiGe channels [14]–[17]. Although the SiGe channel devices have been initially introduced to exploit the higher mobility of the SiGe channel compared with conventional Si channel devices, the superior reliability with respect to NBTI is an incidental benefit of this particular device technology. Apparently counterintuitively, it was found that devices with a very thin Si cap layer on top of the SiGe channel resulted in the best reliability performance [18]. In order to provide a microscopic explanation for charge trapping in SiGe devices, we extend our initial study on this subject [19] by focusing on the experimental characterization and modeling of the response of nanoscale and large-area devices subjected to NBTI stress.

The focus of this paper is put on the experimental characterization of recently developed SiGe channel pMOS transistors using HK gate stacks (see Fig. 1). To achieve a comparable degradation caused by NBTI across the three different device types, the stress/relaxation measurement data are expressed in terms of an equivalent threshold voltage shift, defined as the difference in the two gate voltage levels necessary to achieve the drain–source current before and after the charge emission

Manuscript received August 25, 2016; revised December 25, 2016 and February 10, 2017; accepted March 17, 2017. Date of publication April 3, 2017; date of current version April 19, 2017. This work was supported in part by the Austrian Science Fund (FWF) under Project 26382-N30, in part by the European Community's FP7 Project (MoRV) under Grant 619234, and in part by the Intel Sponsored Research under Project 2013111914. The review of this paper was arranged by Editor Y. Momiyama.

M. Waltl, G. Rzepa, W. Goes, and T. Grasser are with the Institute for Microelectronics, TU Wien, 1040 Vienna, Austria (e-mail: waltl@iue.tuwien.ac.at).

A. Grill is with the Institute for Microelectronics, TU Wien, 1040 Vienna, Austria, and also with the Christian Doppler Laboratory, Institute for Microelectronics, TU Wien, 1040 Vienna, Austria.

J. Franco, B. Kaczer, L. Witters, J. Mitard, and N. Horiguchi are with imec, Leuven 3001, Belgium.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TED.2017.2686086

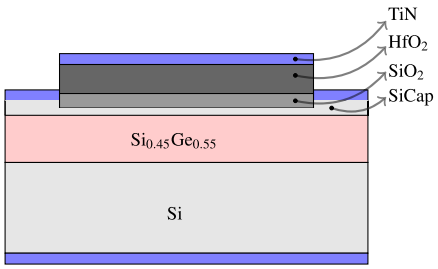

Fig. 1. Schematic view of the studied $Si_{0.45}Ge_{0.55}$ channel devices shows an HfO_2/SiO_2 high-k gate stack together with a thin Si cap spacer on top of the strained SiGe layer. In this paper, nanoscale and large-area devices with two different Si cap layer thicknesses of $d_{\mathrm{SiCap}} = 0.65$ nm and $d_{\mathrm{SiCap}} = 2$ nm and a reference Si transistor are studied in detail.

event occurred. First, we study charge trapping at the single defect level in two nanoscale SiGe devices with different Si cap layer thicknesses and compare them with a Si reference device. Afterward, we extend our investigations to large-area devices and confirm the increasing insensitivity of the devices to NBTI stress when a thin Si cap layer is introduced. Finally, we show that the Si cap layer increases the device lifetime by more than four decades in time. In the second part, we use our four-state nonradiative multiphonon (NMP) model [20] together with the Schrödinger Poisson solver VSP [21] and our device simulator Minimos-NT [22] to explain the recovery of large devices and the capture/emission time characteristics of the identified single traps to further confirm the validity of our conclusions.

II. Characterization of Nanoscale Devices

To analyze the detrimental impact of NBTI, we employ SiGe channel devices with Si cap layers of two different thicknesses ($d_{\mathrm{SiCap}} = 0.65$ nm and $d_{\mathrm{SiCap}} = 2$ nm) between the HfO_2/SiO_2 gate stack and the SiGe layer (see Fig. 1). Additionally all experiments are performed on an additional Si reference transistor.

The nanoscale devices under test (DUTs) have a gate length of $W = 90$ nm and a width of $L = 35$ nm. To study the properties of a single defect, i.e., its bias- and temperature-dependent average capture and emission times as well as its step height, the recently proposed TDDS is employed [6], [20]. Within TDDS, the DUTs are stressed at a constant gate voltage while the drain–source voltage is $V_{\mathrm{DS}}^{s} = 0$ V to prevent hot-carrier degradation effects. After the stress cycle, a recovery gate voltage is applied and the recovery behavior is measured. Finally, the recorded drain–source current is converted to a threshold voltage shift, and the discrete steps in the recovery traces are analyzed [6], [8], [20], [23].

The recovery traces shown in Fig. 2 show that a large number of steps are present in each device variant. Each discrete $\Delta V\mathrm{thstep}$ gives the impact of a single defect, which has been charged during the stress cycle and uncharges during the recovery cycle. Although all three DUTs nominally have the same gate stack, the number of steps in the recovery traces differs significantly in all three devices. In particular, for the same stress voltage, the devices with the thickest Si cap [Fig. 2 (middle)] have the largest number of active defects, followed by the Si reference device [Fig. 2 (right)]. The smallest number of active defects is observed in the SiGe device with the thinnest Si cap [Fig. 2 (left)].

The behavior of NBTI on the device threshold voltage shift is separated into a recoverable and permanent component. While the former is due to defects which become uncharged during the recovery phase, the latter is a consequence of defects which have been charged during the stress cycle too, but have not emitted their charge during the recovery cycle. The permanent contribution to $\Delta V\mathrm{th}$ can be seen as a remaining threshold voltage shift $\Delta V\mathrm{th}$ at the end of the measured recovery traces. An increase of the measurement window, i.e., the recovery time, would lead to a lower remaining $\Delta V\mathrm{th}$. It has to be noted that single traps can have emission times up to weeks, month, or even years [9]. Following from that, a negligible remaining $\Delta V\mathrm{th}$ would require ultralong recovery times.

During the initial stress/recovery experiments, the DUTs have been subjected to the same negative stress bias. The effective oxide stress field can be generally estimated by

$$E_{\mathrm{ox},s}^{\mathrm{eff}} \approx \frac{V_G^{\mathrm{ov}}}{d_{\mathrm{ox}}^{\mathrm{eff}}} = \frac{V_G^{s} - V\mathrm{th}}{d_{\mathrm{ox}}^{\mathrm{eff}}} \tag{1}$$

with the gate overdrive voltage V_G^{ov}, the stress voltage V_G^{s}, the device threshold voltage $V\mathrm{th}$, and the effective oxide thickness $d_{\mathrm{ox}}^{\mathrm{eff}}$. To determine the threshold voltage of the device variants, $I_D(V_G)$ characteristics are recorded (see Fig. 3). In order to suppress the impact of device-to-device variations, the average $I_D(V_G)$ characteristics are calculated from 40 devices for each device variant. Note that all three technologies have different threshold voltages. Using definition (1), apparently the device with the thinnest Si cap is subjected to the *largest overdrive, however, it still degraded less than all the others*.

It is important to note that all experimental methods using constant stress bias suffer from the dynamic change of $\Delta V\mathrm{th}$ due to charging and discharging of defects. As $\Delta V\mathrm{th}$ increases during the stress phase, the effective oxide stress field slightly decreases, hampering the creation of new defects. This effect is naturally considered in our self-consistent TCAD simulations. Note that this reduction of the stress field also occurs during regular device operation when the devices are subjected to constant supply voltages.

To analyze the number of active traps per device and to study the distribution of step heights, the cumulative distribution functions (CCDFs) of the step heights d of the hole traps present in more than 680 nanoscale transistors with $d_{\mathrm{SiCap}} = 0.65$ nm, $d_{\mathrm{SiCap}} = 2$ nm, and the reference pMOSFETs are created and shown in Fig. 4. At that point, it has to be noted that the recovery is very sensitive to the readout/recovery bias conditions. To ensure that the recovery traces are recorded at comparable bias conditions, a current criterion is used to determine the recovery voltage. As shown in Fig. 3, the recovery voltage is set to $V_G^{r} = V_G(I_{\mathrm{DS}} = -1\ \mu\mathrm{A})$. Thus a comparable oxide field is applied to all DUTs during recovery. Using the current criteria, the recovery traces used for computation of the CCDF are recorded.

Among our nanoscale device variants, we found unimodally and bimodally distributed step heights (see Fig. 4). A bimodal

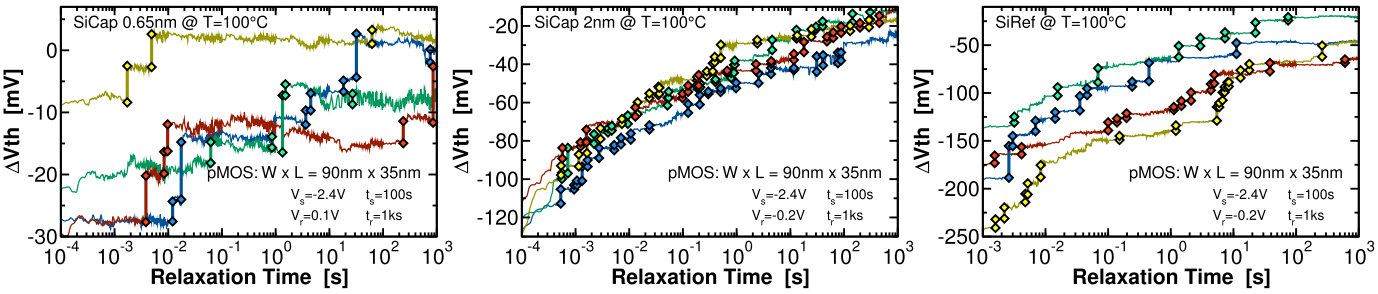

Fig. 2. Recovery traces of the studied p-channel nanoscale SiGe transistors recorded after NBTI stress clearly show a discrete recovery, which reveals the individual defects. Each discrete step corresponds to the emission of a single hole from a defect in the gate stack, while the step height shows its contribution to the threshold voltage shift ΔVth. As visible from the recovery traces, the number of active traps increase with larger Si cap layer thickness (**left**: $d_{\mathrm{SiCap}} = 0.65$ nm and **middle**: $d_{\mathrm{SiCap}} = 2$ nm). The traces from the reference pMOSFETs without a SiGe layer (**right**) show the largest threshold voltage shift among the studied devices. Note that relatively high stress voltages had to be used to cause a measurable degradation of the device with the thinnest Si cap layer.

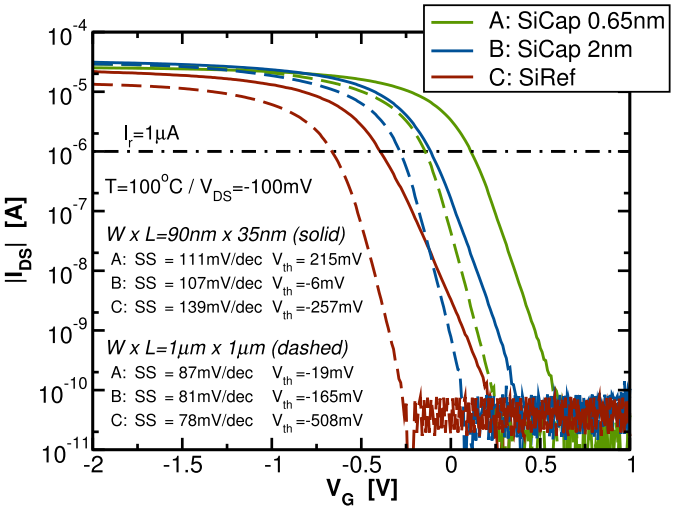

Fig. 3. IDVG characteristics of all studied large-area and nanoscale devices show that the threshold voltage and the subthreshold slope vary with the thickness of the Si cap layer. For comparable recovery conditions, the recovery voltage is defined as the gate voltage necessary to achieve $I_{\mathrm{DS}} = 1$ µA. The threshold voltage is extracted as the gate voltage at which $I_{\mathrm{DS}} = -70$ nA · W/L. As visible, the devices with the thin Si cap layer have the most negative Vth, whereas the reference devices show the most positive Vth.

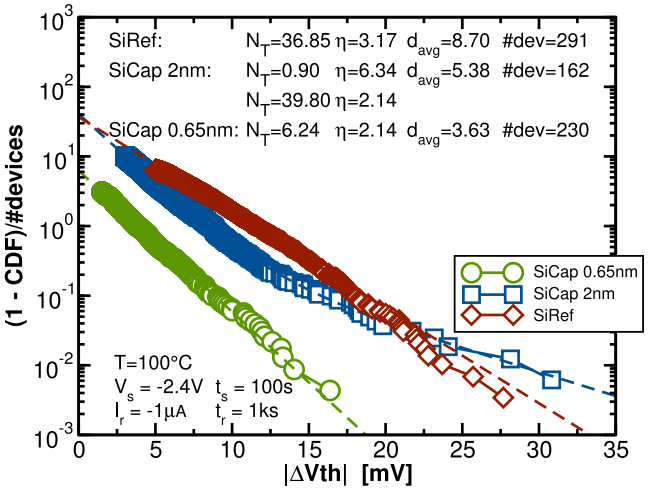

Fig. 4. Step heights d seen in Fig. 2 typically follow a unimodal distribution for the reference devices and the SiGe pMOSFETs with the thinnest Si cap layer of $d_{\mathrm{SiCap}} = 0.65$ nm. Conversely, the devices with $d_{\mathrm{SiCap}} = 2$ nm show a bimodal CCDF with a similar value of η for steps smaller than 12 mV as observed for the CCDF for the $d_{\mathrm{SiCap}} = 0.65$-nm devices. Quite interestingly, the number of accesible traps N_T and the averaged step heights d_{avg} increase with the Si cap layer thickness. The increasing d_{avg} is an indicator that the channel gets closer to the interface near the gate stack. As a consequence of the increased d_{avg}, the number of traps N_T, which can be monitored using TDDS, also increase. Thus, both the larger d_{avg} and the larger N_T indicate that the channel is located more closely to the IL/SiCap interface for the devices with the thick Si cap layer and the reference transistors compared with the devices with the thin Si cap layer.

CCDF can be described by the relation

$$1 - \mathrm{CDF} = A_1 e^{-\frac{\Delta V\mathrm{th}}{\eta_1}} + A_2 e^{-\frac{\Delta V\mathrm{th}}{\eta_2}} \tag{2}$$

with $A_{1,2} = N_{T\,1,2} \times n_T/n_d$ being the product of the trap number per device N_T and the ratio between the overall number of traps n_T and the number of devices n_d. $\eta_{1,2}$ denote $\langle d \rangle$ the mean contribution of a single charge capture/emission event to the threshold voltage shift. The indices 1 and 2 refer to branches of the bimodal distribution [13], [24], [25].

The step heights of our reference transistors are found to be unimodally distributed, which is inline with CCDFs extracted from SiON transistors [24]. Additionally, the DUTs with $d_{\mathrm{SiCap}} = 0.65$ nm show an unimodally CCDF characteristics too (see Fig. 4). Furthermore, as reflected by the CCDF, the devices with $d_{\mathrm{SiCap}} = 0.65$ nm have a significantly lower mean number of traps, $N_T = 6.24$, compared with the reference transistor with $N_T = 36.85$. A particularly noteworthy observation is that the step heights for the devices with the thick Si cap layer follow a bimodal distribution. Quite remarkable, the unimodally distributed CCDF of the $d_{\mathrm{SiCap}} = 0.65$ nm devices and the first part of the bimodal CCDF of the transistors with $d_{\mathrm{SiCap}} = 2$ nm can both be fitted with the same average step height $\eta = 2.14$. The second part of the bimodal CCDF shows a significantly larger value for η, indicating that the traps are located closer to the channel than this is the case for smaller η.

The unimodal CCDFs appear to be a consequence of *only one* dominant conduction channel present at the SiCap/SiGe interface and the channel at the SiO_2/SiCap interface for the device with the thinnest Si cap layer and the reference Si

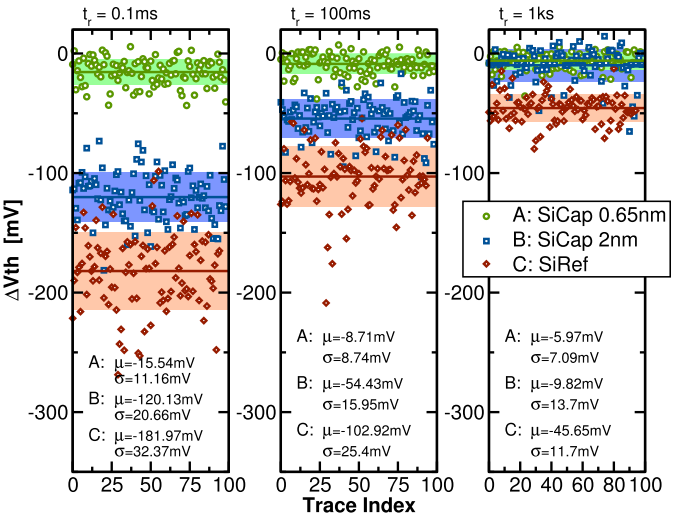

Fig. 5. Threshold voltage shift $_1V$th (symbols) measured after certain recovery times (left: $t_r = 0.1$ ms, middle: $t_r = 100$ ms, and right: $t_r = 1$ ks) is plotted for the different device variants versus the trace index. The experimental recovery traces have been used to calculate the average recovery of the nanoscale SiGe transistors [19]. Note that the standard deviation (marked by colored area) of the ΔVth distribution also decreases for larger recovery times. Also note that the standard deviation is related to the maximum observed step height of the device variants. The largest observed step height of a single trap for the devices with the thin Si cap layer is about half of the maximum step height of the two other technologies (see Fig. 4). This is also the case for the standard deviations from the ΔVth distributions analyzed at certain recovery times.

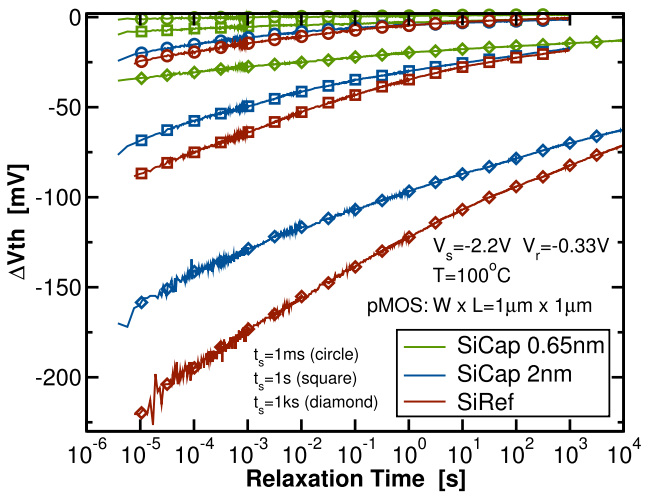

Fig. 6. On large-area devices, the recovery traces show the same trend as their nanoscale counterparts. Whereas only a small threshold voltage shift is measured for the devices with $d_{\mathrm{SiCap}} = 0.65$ nm, a significantly larger sensitivity to NBTI is observed for $d_{\mathrm{SiCap}} = 2$ nm and the reference transistor. The symbols are the mean values of the recovery data used for further modeling of the recovery in the second part of this paper [26]. It is visible that the total threshold voltage shift shows a strong dependence on the stress time. Although a large relaxation time of $t_r = 10$ ks is used, a large permanent component remains at the end of each recovery trace.

devices, respectively. For a bimodal CCDF, the traps interact with *two* conduction path inside the device. Furthermore, in nanoscale devices, the position of random discrete dopands within the channel influences the percolation path of the drain–source current. Considering unfavorably located randomly distributed dopands inside the Si cap, it is conceivable that such dopands can pinchoff one of the two conduction paths, and as a consequence, only a single current path either in the SiGe layer or at the SiO_2/SiCap interface remains. Moreover, it is possible that traps located near the SiO_2/SiCap interface can influence the potential of both channels, leading to very large step heights. From that it follows that there is no unique relation between the step height of a single trap and the corresponding channel the single trap is interacting with. However, the CCDF is created from a large number of defects (more than 4000 for the devices with the thick Si cap layer) and devices. Based on the η values observed from our experimental data for the devices with the thick Si cap layer, we link the first part of the bimodal CCDF to charge trapping interaction with a channel present in the SiGe layer and the tail of the distribution to a superposition of charge trapping interactions between the gate stack and the SiGe and/or SiO_2/SiCap channel. In part two of our study [26], we perform detailed quantum-mechanical simulations confirming that there are a significant amount of hole carriers in both layers for the devices with $d_{\mathrm{SiCap}} = 2$ nm.

The evaluation of the threshold voltage shift of 100 devices for each of the three device variants at certain recovery times is shown in Fig. 5. By considering Fig. 5 together with the CCDF, a remarkable correlation between the standard deviation σ and the maximum observed step height for a device variant is visible as the devices with larger step height have a larger σ is.

III. Characterization of Large Devices

In the following, the recovery behavior after application of NBTI stress to the large-area DUTs, $W = 1$ μm and $L = 1$ μm, is studied. As mentioned previously, in the nanoscale devices, the small number of defects present allows for their individual characterization. In contrast, for large-area devices, the average recovery of a large number of defects is recorded. It has to be noted that for an increasing device area $A = W \cdot L$, a smaller average contribution of a single charge $\eta_{\mathrm{um}}/\eta_{\mathrm{nm}} = A_{\mathrm{um}}/A_{\mathrm{nm}}$ is obtained. The trap density, however, increases with increasing device area $N_{T,\mathrm{um}} = A \cdot N_{T,\mathrm{nm}}$ (the indices *nm* and *um* denote the nanoscale and large-area devices, respectively). To monitor the recovery of the large-area transistors, a conventional extended measure-stress-measure (eMSM) scheme is used [27]. Each eMSM cycle is recorded at the same stress and recovery bias conditions and at the same device temperature. For a measurement cycle, the stress and recovery times are varied as given in

$$\begin{pmatrix} t_s \\ t_r \end{pmatrix} = \begin{pmatrix} 1\ \mathrm{ms} & 10\ \mathrm{ms} & \ldots & 10\ \mathrm{s} & 100\ \mathrm{s} & 1\ \mathrm{ks} \\ 100\ \mathrm{s} & 100\ \mathrm{s} & \ldots & 100\ \mathrm{s} & 1\ \mathrm{ks} & 10\ \mathrm{ks} \end{pmatrix}.$$

Furthermore, the gate bias during stress is varied in the range of $V_G^s \in [-1.6\ \mathrm{V}, -1.7\ \mathrm{V}, \ldots, -2.4\ \mathrm{V}]$, the gate voltage during recovery is set to $V_G^r = 0.33$ V, and the device temperature is not changed during a complete eMSM sequence. For each measurement cycle, a fresh device has been used.

As can be seen from the recovery traces shown in Fig. 7, the devices with $d_{\mathrm{SiCap}} = 0.65$ nm show the smallest threshold voltage shift, whereas the devices with $d_{\mathrm{SiCap}} = 2$ nm as well

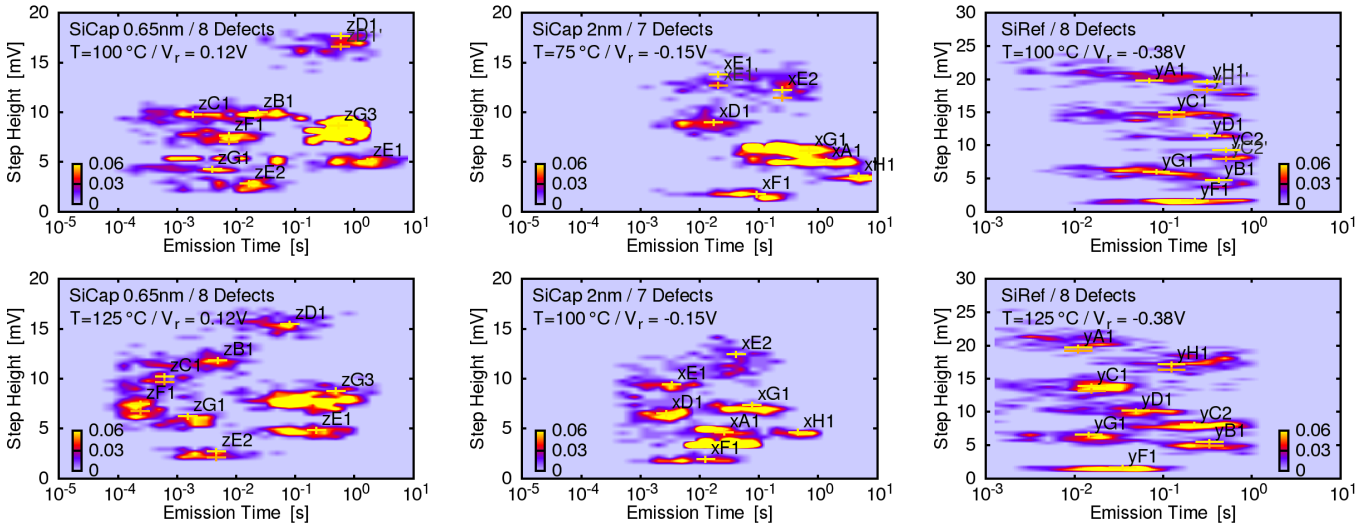

Fig. 7. Defects identified in the three different pMOSFET types are collected in spectral maps at two different temperatures (left: $d_{\mathrm{SiCap}} = 0.65$ nm, middle: $d_{\mathrm{SiCap}} = 2$ nm, and right: reference Si devices). As can be seen, the single clusters move toward shorter emission times when the device temperature is increased (compare **top** and **bottom** row).

as the reference devices degrade much more. This trend is consistent with the observations from our nanoscale devices [19].

IV. Results and Discussion

To study the trapping characteristics of single defects in our nanoscale DUTs, discrete charge transition events from typically 100 repetitions are analyzed and plotted as a distribution of the emission time versus their corresponding step heights (τ_e, d).

Such graphs are called spectral maps and are shown in Fig. 6 for each device variant. Each cluster visible in the spectral maps belongs to one individual defect and acts as the fingerprint of each defect. The intensity of a cluster increases with the number of emission events contributing to it. Moreover, the probability of a defect to capture a charge during stress increases with higher stress time. This allows for the extraction of the characteristic capture time τ_c for a defect using the relation

$$n_e/N = A \cdot (1 - e^{-t_s/\tau_c}) \tag{3}$$

with the stress time t_s, the total number of measured traces N, and the number of emission events n_e [6], [7]. The parameter A gives the probability of a single trap to be charged after indefinitely long device stress. For defect zD1, the occupancy nicely follows (3) and shows a capture time of $\tau_c = 89.3$ ms and $A = 0.99$ (see Fig. 8). Note that, defects with $A \approx 0.5$ produce random telegraph noise provided the time constants fall within the measurement window.

At the beginning of each stress/recovery cycle, the defect which is analyzed has to dwell in its neutral state. During the stress cycle, the defect can become charged, which is a necessary precondition for an emission event to occur during the subsequent recovery cycle. Finally, the vertical position of the clusters in the spectral maps reflect the average contribution of the single defects to the threshold voltage shift ΔVth at certain bias conditions and temperatures.

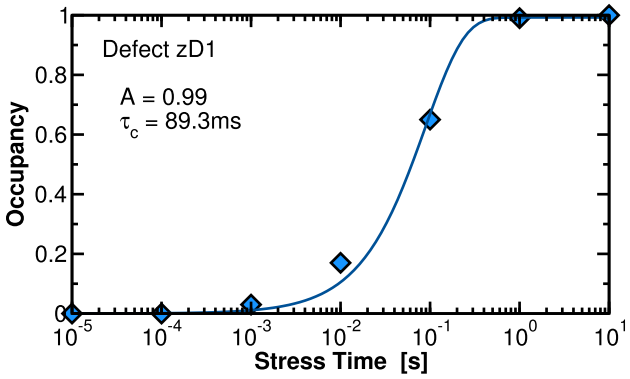

Fig. 8. To calculate the capture time for defect zD1, the TDDS experiments are repeated for seven increasing stress times. For each set of stress/recovery conditions, the spectral map is created and the occupancy, i.e., the ratio between the number of emissions and the number recorded traces, is calculated. Using (3), the capture time at certain bias conditions is obtained.

In the studied nanoscale DUTs, a total of 23 defects have been identified (see Fig. 6). To show the thermally activated nature of charge emission, the spectral maps are plotted at two different temperatures. We observed thermal activation energies E_a starting from approximately 0.5 eV up to more than 1.2 eV for both hole capture and emission (see Fig. 9) very similar to what has been observed for conventional SiON pMOSFETs [9]. We remark that the lower boundary of this energy distribution is determined by the temporal resolution of our equipment as well as the temperatures used, while the upper bound is mainly a consequence of our decision to conduct experiments within a certain time frame. In contrast to the single-trap investigations, the activation energy distribution of defects present in large-area devices cannot be extracted directly from the experimental data. Therefore, detailed TCAD simulations are performed, presented in the second part of our investigations [26], which show that the activation

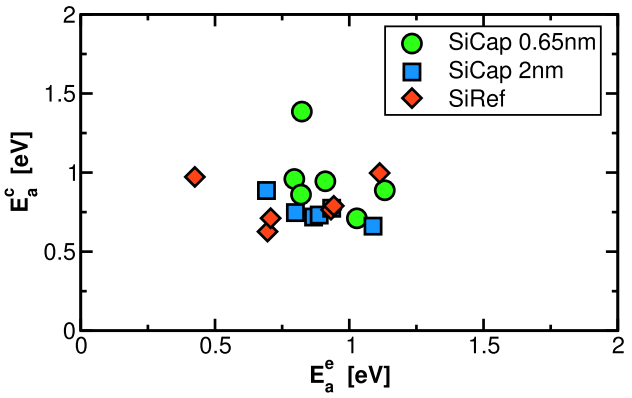

Fig. 9. 18 single traps in our nanoscale devices show a broad distribution of their activation energies for charge capture an emission.

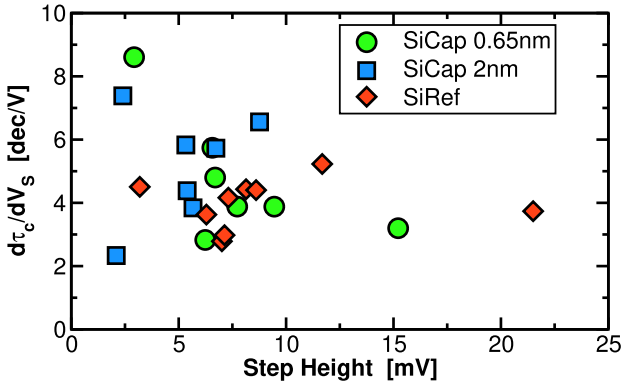

Fig. 10. To check for preferred locations of the traps within the gate stack, the stress bias dependence of the capture times is analyzed versus the step height of the 23 analyzed defects. In general, the bias dependence of the capture time is correlated with the trap position as is explained in more detailed in [26]. As can be seen, for all three technologies, a similar bias dependence of the capture times is observed. Following from that, a similar trap depth distribution is expected, which is confirmed by our calculated trap depth distributions [26].

energy distributions for nanoscale, and large devices are fully consistent.

To account for any correlation between the trap position inside the dielectrics and the device variants, the stress bias dependence of the capture time of our single traps is shown Fig. 10. According to NMP theory, $d\tau_c/dV_G^s$ of a single trap is strongly correlated with the trap depth. As the range of $d\tau_c/dV_G^s$ values for all traps among the three device variants is nearly the same, we expect a similar trap depth distribution for the three technologies. This is confirmed in part two of our investigations where we use our calibrated four-state NMP model to estimate the trap depth of our single traps [26]. Note that, as the trap depth is primarily determined from the bias dependence of the capture time and not by the step height of the single defect, the slope of the CCDF is on average not affected by the trap distribution. However, a change of the stress bias would affect the CCDF as the number of active traps N_T increases at higher stress bias and decreases at lower stress bias.

Based on the measurement data, we finally evaluate the device lifetime of the large-area devices (see in Fig. 11). The threshold voltage shifts are extracted from the eMSM

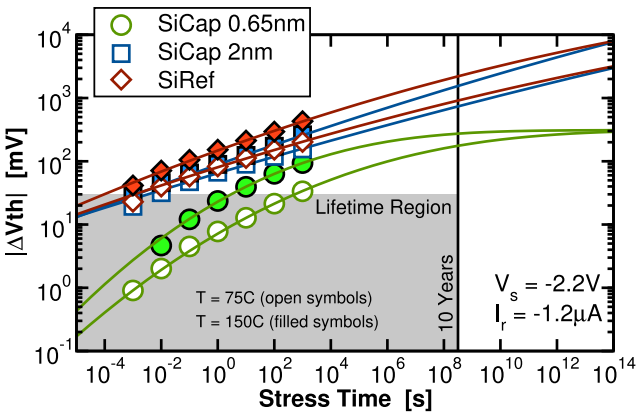

Fig. 11. To estimate the lifetime of the large-area DUTs, the dependence of the threshold voltage shift $\Delta V\text{th}$ at $t_r = 10\ \mu\text{s}$ after stress release on the stress time t_s is analyzed. Even at high stress voltages, the device lifetime is considerably increased by the introduction of a SiGe channel and a thin Si cap layer. The lifetime region is defined by the maximum allowed voltage shift of $\Delta V\text{th}^{\max} = 30$ mV and the maximum expected operation time of ten years and the nominal operating voltage of our DUTs is typically $V_{\text{DD}} = -1.2$ V. It has to be noted that the lifetime estimation is based on the measuement data at hand recorded at bias conditions tremendously larger than the nominal operating conditions. Using technology relevant biases will cause an additional boost of the device lifetimes.

measurement data from the large-area device at $t_r = 10\ \mu\text{s}$ after switching from stress to recovery bias conditions. As can be seen, the SiGe devices with the thinnest Si cap provide a superior lifetime, easily outperforming the Si reference device. Moreover, with increasing stress time, the absolute threshold voltage shift of the device with $d_{\text{SiCap}} = 0.65$ nm saturates and thus does not follow a simple power law. A rather more realistic lifetime prediction is achieved by assuming normally distributed activation energies of the capture time constants. Following from that, the threshold voltage shift dependence on the stress time can be described by [28]

$$\Delta V\text{th}(t_s) = \frac{\Delta V\text{th}^{\max}}{2}\text{erfc}\left(\frac{\bar{E}_a - k_B T \log(t_s/\tau_0)}{\sqrt{2}\sigma_c}\right) \quad (4)$$

with the maximum value for the threshold voltage shift $\Delta V\text{th}^{\max}$, the mean and standard deviation for normally distributed activation energies $\bar{E}_a$ and σ_c, the normalization constant τ_0, the Boltzmann constant k_B, and the absolute temperature T. Using the above-mentioned relation, a realistic lifetime prediction can be made (see Fig. 11). In contrast, using a simple power law quite obviously leads to an overly pessimistic extrapolation.

The lifetime region from Fig. 11 is limited by a maximum allowed threshold voltage shift of $\Delta V\text{th}^{\max} = 30$ mV and the maximum expected operation time of ten years. Even at a stress bias of $V_G^s = 2.2$ V, which is a much higher voltage than the nominal operating voltage of typically $V_{\text{DD}} = -1.2$ V for these devices, the lifetime of the devices with the thinnest Si cap layer is more than five decades larger compared with the other two technologies.

Because of the much stronger bias dependence for the thin cap layer devices, this ratio will further improve at use conditions as will be shown in part two.

V. Conclusion

In this paper, we compare the experimental results of NBTI in nanoscale and large-area SiGe pMOS transistors. Devices with two different Si cap layer thicknesses and reference devices without a SiGe channel are studied in much greater detail than before to reveal the microscopic mechanism responsible for device degradation. First, we use the TDDS to study single trap in the nanoscale DUTs. We demonstrate that unimodal and bimodal CCDFs are found for the devices with $d_{\mathrm{SiCap}} = 0.65$ nm and $d_{\mathrm{SiCap}} = 2$ nm, respectively. This suggests that two conducting channels are present in SiGe devices with a thick Si cap layer. By evaluation of the CCDF together with the average recovery trace for the nanoscale devices, we demonstrate that the thin Si cap layer significantly decreases the device sensitivity to NBTI. It was also shown that a correlation between the slope of the CCDF and the variance of the average recovery trace calculated from 100 single traces is observed. The smaller the variance becomes, the smaller η value and the smaller the trap density N_T gets.

Using an eMSM scheme, the recovery of large-area devices is studied. We show that a significantly weaker dependence of the device degradation on the stress time is present in the SiGe devices with the thin Si cap compared with the two other device variants. This observation confirms the superior NBTI behavior as a consequence of the thin Si cap layer.

In all nanoscale DUTs, we identify 23 single defects. For these defects, we found activation energies of up to 1.2 eV for hole capture and emission. In addition to that, the three studied device technologies show a similar bias dependence of their capture times and thus similar trap depth distributions are expected.

Finally, a lifetime projection based on the experimental data is presented and confirms the massive boost for the device lifetime. It is demonstrated that by introducing a thin Si cap layer, the lifetime is increased by more than four decades in time.

Acknowledgment

This paper has been performed in part in the frame of the imec Core Partner Program on Ge-based devices.

References

[1] G. D. Wilk, R. M. Wallace, and J. M. Anthony, "High-*k* gate dielectrics: Current status and materials properties considerations," *J. Appl. Phys.*, vol. 89, no. 10, p. 5243, Jan. 2001.

[2] J. Robertson, "High dielectric constant oxides," *Eur. Phys. J. Appl. Phys.*, vol. 28, pp. 265–291, Dec. 2004.

[3] H. L. Byoung *et al.*, "Ultrathin hafnium oxide with low leakage and excellent reliability for alternative gate dielectric application," in *IEDM Tech. Dig.*, Dec. 1999, pp. 133–136.

[4] E. P. Gusev *et al.*, "Ultrathin high-*k* gate stacks for advanced CMOS devices," in *IEDM Tech. Dig.*, Dec. 2001, pp. 20.1.1–20.1.4.

[5] T. Grasser *et al.*, "The paradigm shift in understanding the bias temperature instability: From reaction–diffusion to switching oxide traps," *IEEE Trans. Electron Devices*, vol. 58, no. 11, pp. 3652–3666, Nov. 2011.

[6] T. Grasser, H. Reisinger, P.-J. Wagner, F. Schanovsky, W. Goes, and B. Kaczer, "The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability," in *Proc. Int. Rel. Phys. Symp. (IRPS)*, May 2010, pp. 16–25.

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

[8] H. Reisinger, T. Grasser, and C. Schlünder, "A study of NBTI by the statistical analysis of the properties of individual defects in pMOSFETS," in *Proc. Int. Integr. Rel. Workshop*, Oct. 2009, pp. 30–35.

[9] T. Grasser *et al.*, "On the microscopic structure of hole traps in pMOSFETs," in *IEDM Tech. Dig.*, Dec. 2014, pp. 21.1.1–21.1.4.

[10] M. Waltl, W. Goes, K. Rott, H. Reisinger, and T. Grasser, "A single-trap study of PBTI in SiON nMOS transistors: Similarities and differences to the NBTI/pMOS case," in *Proc. Int. Rel. Phys. Symp. (IRPS)*, Jun. 2014, pp. XT18.1–XT18.5.

[11] T. Wang *et al.*, "A novel transient characterization technique to investigate trap properties in HfSiON gate dielectric MOSFETs-from single electron emission to PBTI recovery transient," *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 1073–1079, May 2006.

[12] J. Zou *et al.*, "New insights into AC RTN in scaled high-κ/ metal-gate MOSFETs under digital circuit operations," in *IEEE Symp. VLSI Technol. Dig. Tech. Papers*, Jun. 2012, pp. 139–140.

[13] M. Toledano-Luque *et al.*, "Temperature and voltage dependences of the capture and emission times of individual traps in high-*k* dielectrics," *Microelectron. Eng.*, vol. 88, no. 7, pp. 1243–1246, Jul. 2011.

[14] S. Deora *et al.*, "Intrinsic reliability improvement in biaxially strained SiGe p-MOSFETs," *IEEE Electron Device Lett.*, vol. 32, no. 3, pp. 255–257, Mar. 2011.

[15] S. Krishnan *et al.*, "A manufacturable dual channel (Si and SiGe) high-*k* metal gate CMOS technology with multiple oxides for high performance and low power applications," in *IEDM Tech. Dig.*, Dec. 2011, pp. 28.1.1–28.1.4.

[16] P. Srinivasan *et al.*, "SiGe composition and thickness effects on NBTI in replacement metal gate / high-*k* technologies," in *Proc. Int. Rel. Phys. Symp. (IRPS)*, Jun. 2014, pp. 6A.3.1–6A.3.6.

[17] J. Franco *et al.*, "SiGe channel technology: Superior reliability toward ultrathin EOT devices—Part I: NBTI," *IEEE Trans. Electron Devices*, vol. 60, no. 1, pp. 396–404, Jan. 2013.

[18] J. Franco *et al.*, "SiGe channel technology: Superior reliability toward ultra-thin EOT devices—Part II: Time-dependent variability in nanoscaled devices and other reliability issues," *IEEE Trans. Electron Devices*, vol. 60, no. 1, pp. 405–412, Jan. 2013.

[19] M. Waltl *et al.*, "Nanoscale evidence for the superior reliability of SiGe high-*k* pMOSFETs," in *Proc. Int. Rel. Phys. Symp. (IRPS)*, Apr. 2016, pp. XT-02-1–XT-02-6.

[20] T. Grasser *et al.*, "Advanced characterization of oxide traps: The dynamic time-dependent defect spectroscopy," in *Proc. Int. Rel. Phys. Symp. (IRPS)*, Apr. 2013, pp. 2D.2.1–2D.2.7.

[21] O. Baumgartner, Z. Stanojevic, K. Schnass, M. Karner, and H. Kosina, "VSP—A quantum-electronic simulation framework," *J. Comput. Electron.*, vol. 12, no. 4, pp. 701–721, Dec. 2013.

[22] *Global TCAD Solutions, Minimos-NT User Manual—Release 2014.03*, 2015.

[23] M. Waltl, P.-J. Wagner, H. Reisinger, K. Rott, and T. Grasser, "Advanced data analysis algorithms for the time-dependent defect spectroscopy of NBTI," in *Proc. Int. Integr. Rel. Workshop*, Oct. 2012, pp. 74–79.

[24] B. Kaczer, P. J. Roussel, T. Grasser, and G. Groeseneken, "Statistics of multiple trapped charges in the gate oxide of deeply scaled MOSFET devices—Application to NBTI," *IEEE Electron Device Lett.*, vol. 31, no. 5, pp. 411–413, May 2010.

[25] B. Kaczer *et al.*, "Origin of NBTI variability in deeply scaled pFETs," in *Proc. Int. Rel. Phys. Symp. (IRPS)*, May 2010, pp. 26–32.

[26] M. Waltl *et al.*, "Superior NBTI in high-*k* SiGe transistors—Part II: Theory," *IEEE Trans. Electron Devices*, vol. 64, no. 5, pp. 2099–2105, May 2017, doi: 10.1109/TED.2017.2686454.

[27] B. Kaczer *et al.*, "Ubiquitous relaxation in BTI stressing—New evaluation and insights," in *Proc. Int. Rel. Phys. Symp. (IRPS)*, Apr. 2008, pp. 20–27.

[28] G. Pobegen and T. Grasser, "On the distribution of NBTI time constants on a long, temperature-accelerated time scale," *IEEE Trans. Electron Devices*, vol. 60, no. 7, pp. 2148–2155, Jul. 2013.

Authors' photographs and biographies not available at the time of publication.