This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

SiGe Channel Technology: Superior Reliability Toward Ultrathin EOT Devices—Part I: NBTI

Jacopo Franco, *Student Member, IEEE*, Ben Kaczer, Philippe J. Roussel, Jérôme Mitard, Moonju Cho, Liesbeth Witters, Tibor Grasser, *Senior Member, IEEE*, and Guido Groeseneken, *Fellow, IEEE*

***Abstract*—We report extensive experimental results of the negative bias temperature instability (NBTI) reliability of SiGe channel pMOSFETs as a function of the main gate-stack parameters. The results clearly show that this high-mobility channel technology offers significantly improved NBTI robustness compared with Si-channel devices, which can solve the reliability issue for sub-1-nm equivalent-oxide-thickness devices. A physical model is proposed to explain the intrinsically superior NBTI robustness.**

***Index Terms*—Ge, negative bias temperature instability (NBTI), pMOSFET, reliability, SiGe.**

I. Introduction

NEGATIVE bias temperature instability (NBTI) is considered the most severe reliability issue for scaled CMOS technologies [1]. The quasi-constant supply voltage scaling proposed by the International Technology Roadmap for Semiconductors (ITRS) [2] for the recent technology nodes enhances NBTI due to the ever increasing interfacial oxide electric field $E_{\rm ox}$. As a consequence, although several groups have already demonstrated well-behaving CMOS devices with aggressively scaled equivalent oxide thickness (EOT) down to 0.5 nm [3], [4], a ten-year lifetime cannot be guaranteed for the expected operating voltages [5], [6]. Hence, the reliability issue is coming up as a showstopper.

Meanwhile, the use of high-mobility channels (e.g., SiGe and Ge) is being considered for further enhancement of the CMOS performance [7]–[10]. The main benefit promised by the Ge-based technology can be briefly summarized as follows: 1) enhanced mobility, which can alleviate the mobility reduction caused by the defective high-k layer coming closer to the channel due to the scaling of the SiO_2 interfacial layer (IL), and 2) pMOS threshold voltage tuning toward the roadmap target.

In this paper, we report a complete study of the NBTI reliability of Ge-based quantum-well (QW) pMOSFETs. In 2009, we have already observed that the incorporation of Ge into the channel significantly improves NBTI robustness [11], [12]. Extensive experimental datasets are collected here, showing the reliability improvement to be process and architecture independent, while being intrinsically related to the incorporation of Ge. We thoroughly discuss extensive experimental results, including new insights and supporting data, and we propose a physical model that can explain all the experimental observations. It is made clear how incorporation of Ge into the pMOSFET channel opens a new degree of freedom for optimizing the NBTI reliability of ultrathin EOT devices. In particular, a reliability-oriented gate-stack optimization with a high Ge fraction, a thick QW, and a thin Si passivation layer is shown to boost the allowed gate voltage overdrive for a ten-year lifetime above the expected operating V_{DD} for devices with ultrathin EOT (down to ~0.6-nm EOT) [13]. The extensive experimental results collected on a variety of processed wafers and reported here strongly support SiGe channel technology as a promising candidate for future CMOS technology nodes, offering a solution to the reliability issue for ultrathin EOT devices.

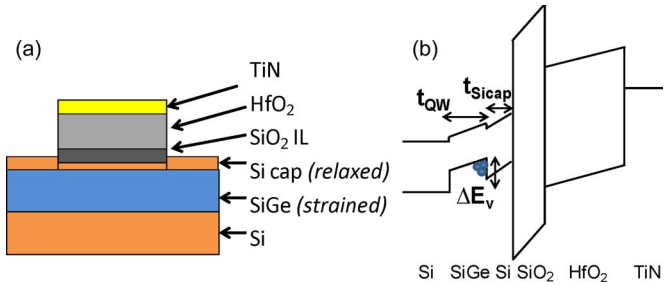

Fig. 1. (a) Sketch of the gate stack of SiGe devices used in this paper. (b) Band diagram sketch in inversion. Channel holes are confined into the SiGe QW due to the valence-band offset ΔE_v between the SiGe channel and the Si cap. The Si cap thickness $t_{\rm Sicap}$ therefore contributes to the $T_{\rm inv}$ of the gate stack.

II. Experimental

The buried SiGe channel pFETs used in this paper were fabricated at the Interuniversity Microelectronics Centre (IMEC) on 300-mm Si wafers. A sketch of the device gate stack and its band diagram in inversion are depicted in Fig. 1. The channel layer consists of an epitaxially grown compressively strained thin $Si_{1-x}Ge_x$ layer, with thickness varying between 3 and 7 nm. Ge fractions up to $x = 0.55$ were used. On top of the $Si_{1-x}Ge_x$ layer, a thin undoped Si cap was grown epitaxially. The physical thicknesses of this thin Si cap varied between 0.65 and 2 nm (as estimated from C–V curves and transmission

Manuscript received June 8, 2012; revised September 27, 2012; accepted October 4, 2012. This work was supported in part by the European Commission through the Seventh Framework Programme (collaborative project MORDRED) under Contract 261868. The review of this paper was arranged by Editor J. S. Suehle.

J. Franco and G. Groeseneken are with the imec, 3001 Leuven, Belgium, and also with the Department of Electrical Engineering (ESAT), Katholieke Universiteit Leuven, 3000 Leuven, Belgium (e-mail: Jacopo.Franco@imec.be).

B. Kaczer, P. J. Roussel, J. Mitard, M. Cho, and L. Witters are with the imec, 3001 Leuven, Belgium.

T. Grasser is with the Technische Universität Wien, 1040 Vienna, Austria.

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.2012.2225625

0018-9383/$31.00 © 2012 IEEE

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

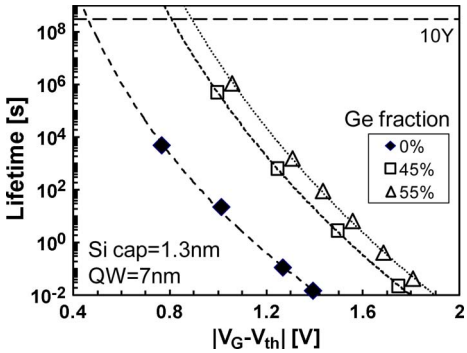

Fig. 2. Extrapolated lifetimes as a function of gate voltage overdrive for varying Ge content. A higher Ge fraction boosts NBTI robustness.

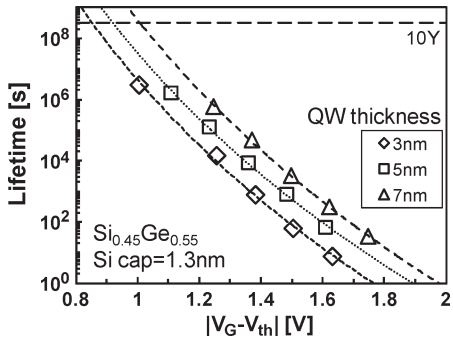

Fig. 3. Extrapolated lifetimes as a function of gate voltage overdrive for varying QW thickness (55% Ge fraction, 1.3-nm-thick Si cap). A thicker QW boosts NBTI robustness.

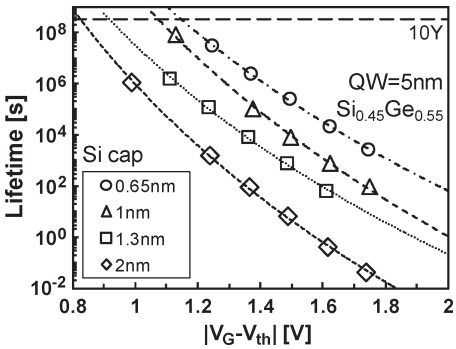

Fig. 4. Extrapolated lifetimes as a function of gate voltage overdrive for varying Si cap thickness (55% Ge fraction, 5-nm-thick QW). A reduced Si cap thickness boosts NBTI robustness while enabling $T_{\rm inv}$ reduction. The extrapolated $V_{\rm op}$ are plotted in Fig. 6 for fair benchmarking versus $T_{\rm inv}$.

electron microscope pictures of the final device). A detailed description of the epitaxial process can be found elsewhere [14]. Gate-stack fabrication started with wet chemical oxidation (*imec clean* [15]) of the Si cap. On top of this IL, an $\sim$1.8-nm-thick HfO_2 layer was deposited by atomic layer deposition. Finally, a TiN metal gate was deposited by physical vapor deposition (PVD). The metal gate thickness controlled the final IL thickness by means of the oxygen-scavenging technique, as discussed in [3]. The mobility enhancement factor of SiGe devices w.r.t. Si ranged between 1.5$\times$ and 2.4$\times$, depending on the gate-stack parameters [8].

Due to the valence-band offset between SiGe and the Si cap [see Fig. 1(b)], inversion holes are confined in the SiGe channel, which therefore acts as a QW. This causes the Si cap thickness to lower the inversion capacitance, as compared with the accumulation capacitance [11]. For fair benchmarking of these devices, it is therefore necessary to consider the capacitance-equivalent thickness in inversion $T_{\rm inv}$ (evaluated at $V_G = V_{\rm th} - 0.6$ V), which includes the contribution of the Si caps of varying thicknesses.

NBTI stress experiments were performed using the extended measure–stress–measure (eMSM) technique [16]. The devices were stressed at $T = 125\,^{\circ}$C with several gate overdrives, while the sensing bias was $V_G = V_{\rm th0}$. To minimize NBTI relaxation effects for the device lifetime predictions, $\Delta V_{\rm th}$ was evaluated at $t_{\rm relax} = 1$ ms, i.e., the minimum delay of the used setup (Keithley 2602 Fast Source Meter Units). This delay was fixed in the experiments to allow cross-comparison. For each gate voltage, the stress time needed to reach a failure criterion, which was assumed at a 30-mV threshold voltage shift, was extracted. The ten-year lifetime operating overdrive $V_{\rm op}$ was then extrapolated by fitting a power law to the lifetime-versus-gate-overdrive data set (see, e.g., Figs. 2–4). For a dedicated experiment presented in Section VI-D, an ultrafast NBTI measurement setup (Keithley 4200 PMUs UF-BTI) was used to reduce the measurement delay to $\sim$2 μs (i.e., comparable with the fastest reported NBTI measurements [17]).

First, the three major process parameters of the SiGe pMOSFETs, i.e., the Ge fraction, the SiGe layer thickness, and the Si cap thickness, were separately varied in order to assess their individual impact on NBTI. In this preliminary set of experiments, the EOT was not aggressively scaled (EOT $\sim$ 1.2 nm). For comparison, a second set of standard Si channel devices with an identical gate stack was also used. Then, experimental learning was used for a reliability-oriented optimization of the SiGe gate-stack at sub-1-nm EOT. Finally, a model explaining the superior reliability is proposed.

III. Experimental Results

A. Ge *Fraction*

As shown in Fig. 2, the introduction of Ge in the channel significantly improved NBTI reliability. The extrapolated operating overdrive voltage for a ten-year lifetime $V_{\rm op}$ increased from 0.46 V for the Si reference up to 0.8 V for a 45% Ge fraction device with a SiGe layer thickness of 7 nm and a Si cap thickness of 1.3 nm. Increasing the Ge fraction to 55% while fixing the other parameters boosted the operating overdrive voltage even more, reaching 0.9 V.

B. SiGe *QW Thicknesses*

Increasing the thickness of the SiGe QW gave an additional improvement of NBTI reliability (see Fig. 3): $V_{\rm op}$ increased from 0.85 up to 1.01 V when moving from a 3-nm-thick SiGe layer to a 7-nm one. This observation was made with fixed Si cap thickness (1.3 nm) and Ge fraction (55%).

C. Si *Cap Thicknesses*

The most significant impact on NBTI reliability was observed when varying the Si cap thickness (see Fig. 4). Interestingly, a reduced thickness of this layer clearly improved NBTI

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

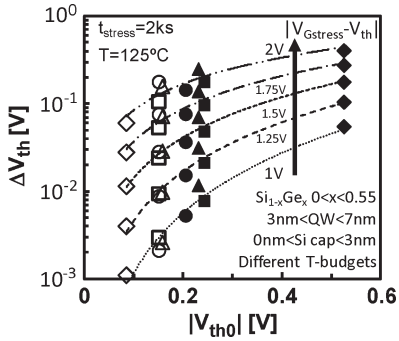

Fig. 5. A clear correlation between the initial $V_{\rm th0}$ and NBTI-caused $\Delta V_{\rm th}$ is consistently observed on our SiGe devices with different gate stacks: devices with lower initial $V_{\rm th0}$ always showed reduced $V_{\rm th}$ instability, at any given stress condition ($|V_{\rm Gstress} - V_{\rm th0}|$). This was not observed for Si devices.

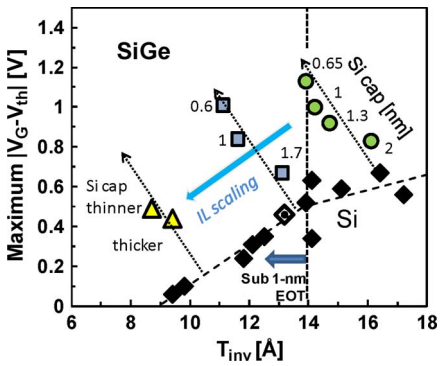

Fig. 6. Maximum operating overdrive for a ten-year lifetime ($T = 125$ °C, failure criterion $\Delta V_{\rm th} = 30$ mV) versus $T_{\rm inv}$ (evaluated at $V_G = V_{\rm th} - 0.6$ V). SiGe devices with a thin Si cap offer improved NBTI reliability, i.e., higher $V_{\rm op}$. Note: Open diamond and circles represent the Si reference gate stack of Fig. 2 and the SiGe gate stacks of Fig. 4, respectively.

robustness. Naively, one would expect the thinner Si cap to act as a reduced tunneling barrier for holes but, conversely, $V_{\rm op}$ increased from 0.82 to 1.14 V when the Si cap thickness was decreased from 2 to 0.65 nm. This counterintuitive observation is crucial for understanding the superior SiGe reliability, as we will discuss in Section VII. Moreover, the observation is particularly relevant since a reduced Si cap thickness, while improving NBTI reliability, also reduces the device $T_{\rm inv}$ [due to reduced hole displacement, see Fig. 1(b)] and therefore enhances the current drive performance.

Finally, we note that a distinct relation between the initial $V_{\rm th0}$ and the NBTI-caused $\Delta V_{\rm th}$ is consistently observed in our SiGe devices with different gate stacks (see Fig. 5). Devices with lower initial $V_{\rm th0}$ always showed reduced $V_{\rm th}$ instability, at any given stress condition ($|V_{\rm Gstress} - V_{\rm th0}|$). This correlation has not been observed for Si channel devices [18], and it will be discussed later in this paper in Section VII.

IV. Gate-Stack Optimization

The $V_{\rm op}$ extracted for different Si cap thicknesses are shown in a benchmark plot versus the $T_{\rm inv}$ values (see Fig. 6) and compared with benchmark data measured on Si channel pMOSFETs. As one can see, it is clear that reducing the Si cap thickness yields a significant $V_{\rm op}$ boost together with a $T_{\rm inv}$ reduction. Such a $V_{\rm op}$ boost for a reduced Si cap thickness was

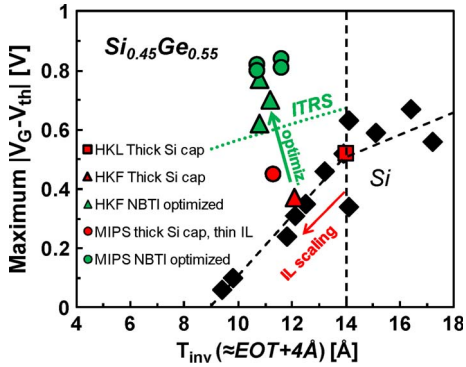

Fig. 7. A high Ge fraction (55%) in a 6.5-nm-thick QW, combined with a thin Si cap (0.8 nm), boosts $V_{\rm op}$ to meet the target V_{DD} at ultrathin EOT in a MIPS flow (green circles, as compared with the red circle). The optimization was also implemented in an RMG flow: high-k last SiGe sample with thick Si cap (red square) shows poor NBTI robustness; an IL reduction by means of O-scavenging in a high-k first process flow (red triangle), further reduces NBTI robustness; however, the SiGe gate-stack optimization (green triangles) boosts the $V_{\rm op}$ above the ITRS target. The results were reproduced for several process T-budgets.

consistently observed for several IL thicknesses. This trend is clearly different w.r.t. the data collected on Si channel devices, where a $T_{\rm inv}$ reduction (normally achieved by IL scaling) is always associated with a reliability reduction, as also shown in Fig. 6.

This remarkable property can be used to optimize the SiGe gate stack and salvage the NBTI reliability of devices with aggressively scaled IL. Fig. 7 reports that combining the beneficial effects of a high Ge fraction, a thicker QW, and a thinner Si cap, the NBTI lifetime was boosted above the ITRS target V_{DD} condition at ultrathin EOT (ten-year continuous operation at $|V_G - V_{\rm th}| \approx 0.6$ V at $T_{\rm inv} \approx 1$ nm, EOT ≈ 0.6 nm).

V. Process- and Architecture-Independent Results

The optimization presented earlier was demonstrated both in a metal-inserted poly-Si (MIPS) and a replacement metal gate (RMG) process flow, with reproducible results for different thermal budgets (see Fig. 7) [19]. These process-independent results already suggest the reliability improvement to be an intrinsic property of Ge-based devices.

Moreover, the improved reliability is observed to be also architecture independent: preliminary results on novel SiGe wrapped-channel bulk pFinFETs [20] show improved NBTI lifetime w.r.t. the Si planar reference when removing the Si cap (see Fig. 8). Furthermore, the experiment with varying Si cap thickness was repeated on pure Ge channel pMOSFETs [21] with 4, 6, and 8 Si monolayers (ML) epitaxially grown from a silane precursor at 500 °C. A reduced thickness of the Si layer again resulted in a reduced NBTI at fixed stress conditions (electric field, stress time, stress temperature, and sensing delay): 4-ML devices degrade $\sim 4\times$ less than 8-ML devices [see Fig. 9(a)]. The same trend was also observed for Si caps grown using a 350 °C epitaxial growth from a trisilane precursor: an $\sim 8\times$ NBTI reduction is observed when reducing the Si from 9 to 3 MLs [see Fig. 9(b)].

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

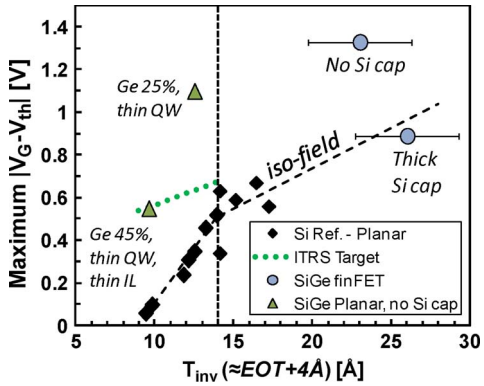

Fig. 8. SiGe channel bulk pFinFETs without a Si cap show improved NBTI reliability w.r.t. the same devices with a thick Si cap and w.r.t. Si planar pFETs. The dashed trend line for $T_{\rm inv} > 1.4$ nm demarcates planar Si pFET constant field scaling (*iso-field*). Uncertainty in $T_{\rm inv}$ is related to the finFET dimensions. Note: Planar SiGe devices without a Si cap also show improved NBTI reliability (e.g., two different planar gate stacks are shown, green triangles: $Si_{0.75}Ge_{0.25}$ 3 nm-thick QW and $Si_{0.55}Ge_{0.45}$ 3 nm-thick QW with a reduced IL thickness).

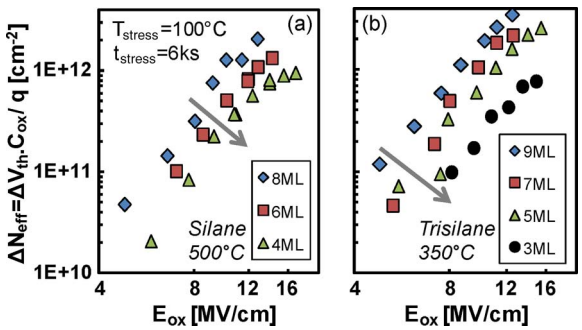

Fig. 9. A decreased thickness of the Si passivation layer improves the NBTI robustness of Ge pMOSFETs, independently of the Si cap epitaxial process used. (a) Silane, 500 °C. (b) Trisilane, 350 °C.

All these process- and architecture-independent results suggest that the reduced NBTI is an intrinsic property of the Ge-based channel structure, further emphasizing the use of a SiGe channel as a promising candidate for future CMOS technology nodes. In the next sections, the physical mechanism behind this experimentally observed property is discussed, and a model for the improved NBTI reliability is proposed.

VI. Discussion

A reduction of the Si cap thickness was shown to yield the most significant reliability boost on SiGe. It is then worth to discuss this remarkable experimental result in more detail.

A. Power-Law Time Exponent and $E_{\rm ox}$-Acceleration

Fig. 10 shows the typical NBTI $\Delta V_{\rm th}$ evolution versus the stress time for the Si reference and the SiGe devices with different Si caps. The SiGe devices show a significantly reduced $\Delta V_{\rm th}$, particularly for the samples with a reduced Si cap thickness. The NBTI $\Delta V_{\rm th}$ evolution is often described as a power law of the stress time ($\Delta V_{\rm th} = A t_{\rm stress}^n$), with a prefactor A dependent on the stress E_{ox}, and an apparent exponent n typically reported in the range of 0.15–0.25 [1] depending on the relaxation allowed by the measurement

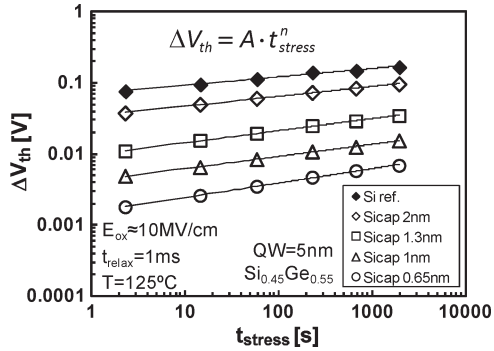

Fig. 10. Measured $\Delta V_{\rm th}$ during NBTI stress at fixed stress conditions on a Si reference and on SiGe devices with different Si cap thicknesses. The SiGe device with the thinnest Si cap shows most reduced $V_{\rm th}$ instability.

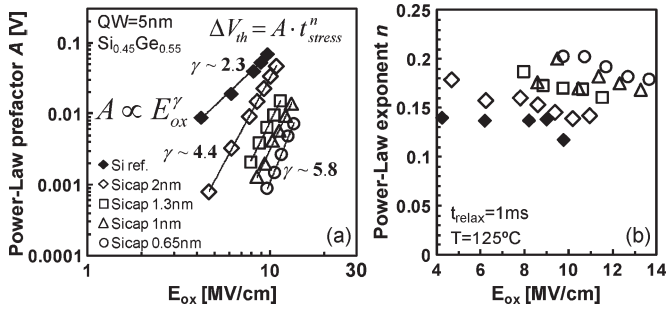

Fig. 11. (a) Extracted power-law prefactors: a significant reduction for the SiGe devices is observed, particularly with a reduced Si cap thickness. A stronger $E_{\rm ox}$-acceleration for SiGe w.r.t. the Si reference device is also noted. (b) Extracted power-law time exponents: SiGe devices with a reduced Si cap thickness show slightly higher apparent exponents.

delay [16]. Fig. 11 documents the extracted power-law prefactors and exponents for all the devices considered here. The prefactors clearly show a dramatic reduction for the SiGe devices, with a further reduction for a reduced Si cap thickness. Moreover, the prefactors show a significantly stronger $E_{\rm ox}$-acceleration for SiGe w.r.t. the Si reference, yielding further benefit at the lower operating fields. On the other hand, we note that SiGe devices with thin Si caps show a slightly higher apparent time exponent. This observation might be linked to a reduced hole trapping component (i.e., causing an apparent exponent closer to the typically observed $\Delta N_{\rm it}$ exponent of $\sim$0.25) and to a faster relaxation [16] in SiGe devices, as discussed later.

B. Temperature Activation

Fig. 12 reports NBTI data at different stress temperatures for Si and SiGe devices (Si cap thicknesses: 2 and 0.65 nm). No clear difference in the activation energies is observed for the different devices. The apparent $\Delta V_{\rm th}$-activation energy is $\sim$60 meV, in the typically reported range [1].

C. Interface State Creation ($\Delta N_{\rm it}$) and Hole Trapping ($\Delta N_{\rm ot}$)

It has been recently reported that NBTI is possibly ascribed to two components [22]: a recoverable (R) one related to

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

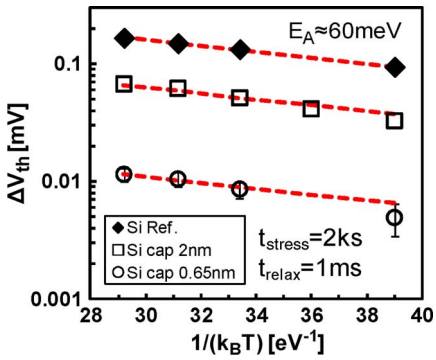

Fig. 12. NBTI-induced $\Delta V_{\rm th}$ measured at different temperatures on the Si reference devices and on SiGe devices with two different Si cap thicknesses (2 and 0.65 nm). No clear difference in the apparent $\Delta V_{\rm th}$-activation energy is observed (extracted $E_A \approx 60$ meV).

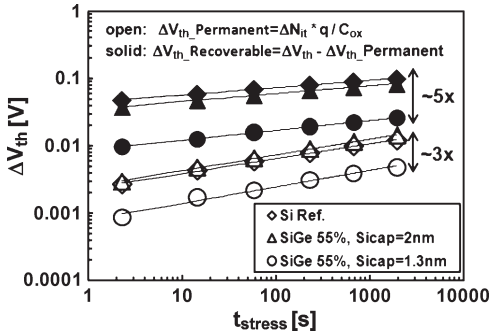

Fig. 13. Total $\Delta V_{\rm th}$ split into the so-called permanent (P) $\Delta V_{\rm th}$, assumed to be caused by $\Delta N_{\rm it}$, and the recoverable (R) $\Delta V_{\rm th}$, assumed to be caused by filling of preexisting oxide traps ($N_{\rm ot}$). $\Delta N_{\rm it}$ measured with CP during NBTI stress were converted to $\Delta V_{\rm th\_Permanent}$ ($= \Delta N_{\rm it}.q/C_{\rm ox}$) in order to decouple their contribution from the total measured $\Delta V_{\rm th}$. $\Delta N_{\rm it}$ follows a power law on the stress time with the same exponent (~0.25) on all three samples. However, SiGe devices with a thinner Si cap show both reduced P and R, with the reduction of R having a higher impact on the total $\Delta V_{\rm th}$.

hole trapping in preexisting bulk oxide defects ($\Delta N_{\rm ot}$) and a so-called permanent one (P) typically associated with creation of new interfaces states ($\Delta N_{\rm it}$). To get insights into the measured NBTI trends, the charge-pumping (CP) technique [23] was used to monitor the interface state creation during the NBTI stress. While $\Delta N_{\rm it}$ was monitored by CP, $\Delta N_{\rm ot}$ was calculated by subtracting the $\Delta N_{\rm it}$ contribution from the total $\Delta V_{\rm th}$ measured. Fig. 13 reports $\Delta N_{\rm it}$ and $\Delta N_{\rm ot}$ evolutions measured on SiGe devices with two different Si cap thicknesses and on the Si reference device for fixed stress conditions ($E_{\rm ox} = 10$ MV/cm, $T = 125$ °C). Two main observations can be made: $\Delta N_{\rm it}$ follows a power law of stress time with the same exponent of ~0.25 for the Si reference and for SiGe with different Si caps, suggesting the same interface bond-breaking process. However, the SiGe device with a thin Si cap shows both reduced R and P, with the R reduction being of higher relevance on the total $\Delta V_{\rm th}$.

D. Faster NBTI Relaxation

To investigate the supposedly faster NBTI relaxation suggested by the higher apparent $\Delta V_{\rm th}$ time exponent for SiGe devices with thin Si caps [see Fig. 11(b)], a dedicated ultrafast

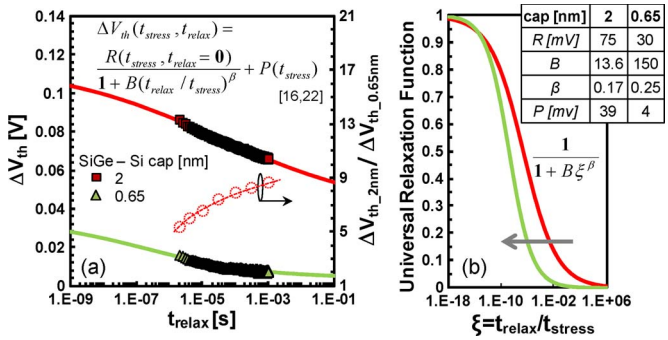

Fig. 14. (a) Ultrafast NBTI relaxation transients recorded on SiGe devices with 2- and 0.65-nm Si caps after the same stress (stress conditions $V_{\rm ov} = 1.95$ V, $T = 25$ C, $t_{\rm stress} = 100$ s). (b) The transients fitted with the universal relaxation model [16], [22] reveal faster relaxation for the thin cap device (inset table shows fitted parameter values; note larger B and β). This is also evident when looking at the increasing ratio of the two $\Delta V_{\rm th}$ curves as a function of the relaxation time (a).

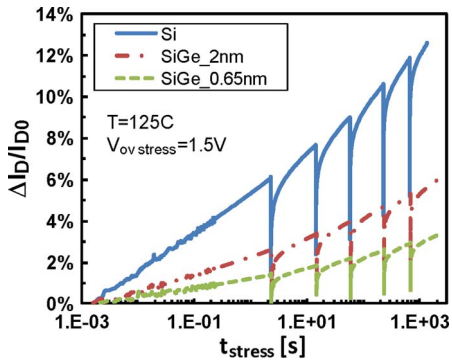

Fig. 15. Drain current degradation recorded on-the-fly during NBTI stress phases of the eMSM. SiGe devices with a reduced Si cap thickness show the lowest degradation also with zero-measurement delay, showing the improved NBTI is not an apparent effect due to faster relaxation behaviors. (Note: The drops in the measurement are due to the MSM technique, where relaxation transients are recorded in regular intervals.)

(minimum $t_{\rm relax} \approx 2$ μs) NBTI measurement was performed. Fig. 14 shows typical relaxation transients recorded on SiGe devices with 2- and 0.65-nm Si caps after the same stress. While valence-band electron repopulation of interface states could be partially responsible for different $\Delta V_{\rm th}$ recovery at short relaxation times (< 1 μs), differences in the relaxation shape are still found for longer relaxation times (> 1 μs): the relaxation transients fitted with the empirical universal relaxation model [16], [22] show faster recovery for the 0.65-nm Si cap devices. This can be also observed by looking at the ratio between the $\Delta V_{\rm th}$ of the thick Si cap device versus the $\Delta V_{\rm th}$ of the thin Si cap device as a function of the relaxation time: an increasing ratio suggests a faster recovery for the latter.

Due to the observed faster relaxation, one may argue that the improved NBTI reliability for SiGe devices with thin Si caps might be only an apparent effect related to the nonzero measurement delay. This hypothesis is readily ruled out by observing the device current degradation during the stress (see Fig. 15). Similarly to the on-the-fly BTI measurement techniques [1], this measurement is performed while the device is biased at the stress condition, i.e., with zero delay. As one can see, the reduced degradation for SiGe is still evident.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

The previously discussed experimental observations on SiGe devices with a reduced Si cap thickness can be summarized as follows:

1) reduced NBTI, i.e., lower power-law prefactor;
2) stronger $E_{\rm ox}$-acceleration;
3) similar temperature activation ($E_A \approx 60$ meV) as the Si reference;
4) similar $\Delta N_{\rm it}$ time exponent ($\sim$0.25) as the Si reference;
5) significantly reduced $\Delta N_{\rm it}$ and $\Delta N_{\rm ot}$, with the latter reduction being of greater relevance;
6) slightly higher apparent $\Delta V_{\rm th}$ time exponent due to strongly reduced $\Delta N_{\rm ot}$ and faster relaxation.

VII. Model

Previous work attributed improved NBTI robustness to induced strain at the interface [24]. This explanation does not apply to our devices: the SiGe layer thicknesses considered here were well below the critical relaxation thickness for the used epitaxial processes, causing the channel layer to be compressively strained [14]. Therefore, the Si cap was lattice matched to the underlying Si substrate, and thus, strain effects at the Si/SiO_2 interface were not involved.

Another hypothesis is related to the Si cap acting as a tunneling barrier for holes toward the dielectric. On the contrary, recent works have clearly disqualified the direct tunneling mechanism to be able to explain the trapping component of NBTI [25], [26]. Moreover, although this explanation might fit the observed trends for the Ge fraction (higher Ge% $\rightarrow$ reduced bandgap $\rightarrow$ higher valence-band offset ΔE_v) and QW thickness (for a very thin QW, i.e., 3 nm, quantum–mechanical effects increase the hole energy and therefore artificially reduce ΔE_v [27]), it clearly fails to explain why a thinner Si cap yields reduced NBTI.

As discussed earlier, SiGe devices with a reduced Si cap thickness show both reduced P and R components. In the next sections, models for reduced P and R are proposed.

A. Reduced P ($\Delta N_{\rm it}$)

$N_{\rm it}$ creation during NBTI stress is commonly attributed to depassivation of H-passivated Si dangling bonds (P_{b0}) at the Si/SiO_2 interface. We have previously reported [13] electron spin resonance spectroscopy [28] measured on a Ge substrate with a thick Si cap that revealed a high P_{b0} density ($\sim 1 \times 10^{12}$ cm^{-2}), whereas it could not detect these defects ($< 10^{11}$ cm^{-2}) for a very thin Si cap. This suggested that the higher Ge segregation at the Si/SiO_2 interface reported for thin Si caps [29] can reduce the $N_{\rm it}$ precursor defect density and therefore reduce $\Delta N_{\rm it}$ during NBTI stress. This reduced creation of interface states might play a role in the improved NBTI reliability observed for SiGe channel devices (see Fig. 13). However, it cannot completely explain the strongly reduced overall NBTI degradation that is mainly related to a significant reduction of the R component, as noted above. The R reduction is discussed next.

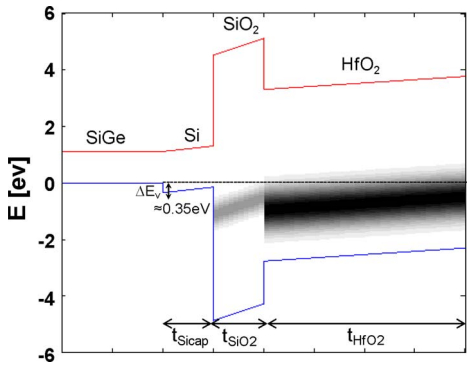

Fig. 16. Model including defect bands centered at 0.95 and 1.4 eV below the Si valence band in the IL and HfO_2, respectively (compatible with the expected O-vacancy levels). The channel Fermi level determines which part of the defect bands is accessible to channel holes. The defect band is modeled as a Gaussian distribution over energy. Charged defects at different spatial positions contribute differently to the total $\Delta V_{\rm th}$ due to electrostatic.

B. R ($\Delta N_{\rm ot}$)—Model for Improved NBTI

We propose that the R reduction is related to a favorable alignment shift of the Fermi level E_F in the SiGe QW w.r.t. the preexisting bulk oxide defect energy levels (see Fig. 16). Larger misalignment can cause carriers to interact with a reduced density of oxide traps $N_{\rm ot}$. To model this effect, we assumed the existence of a defect band both in the SiO_2 IL and in the high-k layer. We note that interacting defects have to be located in both the dielectric layers since the same NBTI trends on SiGe with different Si caps were consistently observed for scaling IL (see Fig. 6). As depicted in Fig. 16, the Fermi level in the channel determines which part of the defect band is accessible to channel holes. The defect bands are modeled as Gaussian distributions over energy. The mean value of the distributions were pinned at 0.95 eV below the Si valence band for the IL (corresponding to the $E'\gamma[E_{0/+}]$ center in SiO_2 [30]) and at 1.4 eV below the Si valence band for the high-k (corresponding to the neutral oxygen vacancy [O°] level in HfO_2 [31]). As a function of the applied gate voltage, all the defects located above the channel Fermi level are considered occupied by trapped holes, whereas all the defects below are neutral (note: no trapping/detrapping kinetics is included in this calculation, i.e., thermodynamic equilibrium).

The model was first calibrated using the NBTI data on the Si reference device: the standard deviations of the Gaussian distributions were used as a fitting parameter (in the range of 0.3–0.5 eV) in order to capture the correct electric field dependence, whereas the defect densities were fitted in order to match the observed $\Delta V_{\rm th}$ magnitude. The varying $\Delta V_{\rm th}$ contribution of defects located at varying depths due to their electrostatic effect was also included. Then, with the same defect band parameters, the expected $\Delta V_{\rm th}$ was calculated for SiGe channel devices, including the valence-band offset of $+0.35$ eV in the channel and including the varying voltage drop on Si cap with varying thicknesses.

As shown in Fig. 17, the simple model excellently matches the experimental data relative to the recoverable component. The model readily captures the following: 1) the reduced NBTI; 2) the stronger field dependence observed for SiGe devices with reduced Si cap thicknesses; and 3) the faster relaxation

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

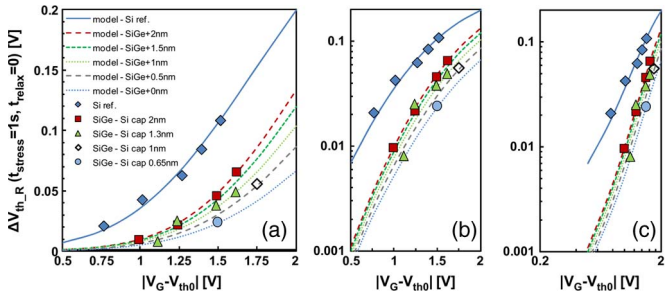

Fig. 17. Calculated $\Delta V_{\rm th}$ versus experimental data of the recoverable component. The model was first calibrated on the Si reference data, then the same defect band parameters were used to calculate the expected $\Delta V_{\rm th}$ for SiGe devices (including the valence-band offset between the SiGe and the Si cap, as well as the voltage drop on different Si cap thickness). The simple model matches the experimental data remarkably well. (a) Linear–linear scale. (b) Logarithmic–linear scale. (c) Logarithmic–logarithmic scale.

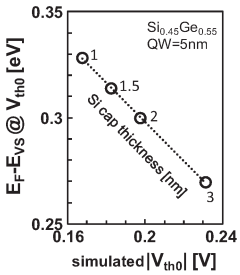

Fig. 18. MEDICI simulations show that a lower device $V_{\rm th0}$ corresponds to a higher channel Fermi level energy E_F w.r.t. the Si valence band $E_{\rm VS}$, which is beneficial for reducing the carrier-trap interaction, according to the model proposed in Fig. 16.

observed for SiGe since the higher energy difference between the defect levels and the SiGe channel is expected to enhance the trapped charge emission [26].

The model also explains the other experimental observations previously made concerning the Ge fraction and the Si cap thickness. In order to minimize the fraction of accessible defects, i.e., in order to *push up* the Fermi level in the channel w.r.t. the defect band, the valence-band offset between SiGe and Si has to be maximized: higher Ge fraction (reduced bandgap and higher ΔE_v) and thick QW (to reduce quantization) are therefore beneficial.

Moreover, this model can also explain the relation between the fresh device $V_{\rm th0}$ and the NBTI observed in SiGe devices (see Fig. 5): as calculated with MEDICI for, e.g., a Si cap thickness split, gate stacks with lower $|V_{\rm th0}|$ have higher channel Fermi level energy (see Fig. 18) and therefore benefit from reduced interaction between holes and oxide defects. Finally, the model also predicts improved NBTI reliability for SiGe channel devices even *without* a Si cap (no voltage drop on the cap, i.e., maximum Fermi energy shift) as experimentally observed for planar devices and finFETs (see Fig. 8).

VIII. Performance Versus Reliability

We have shown that a reduced Si cap thickness is the key for improved reliability. However, previous work reported reduced hole mobility for SiGe devices with a reduced Si cap thickness [21]. This mobility loss was ascribed to poorer interface

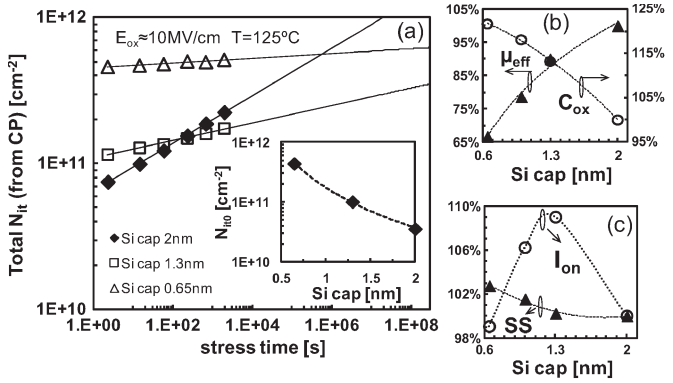

Fig. 19. (a) A reduced Si cap thickness yields a higher interface state density on the fresh device ($N_{\rm it0}$, inset). However, the larger NBTI-induced $\Delta N_{\rm it}$ on a 2-nm-thick Si cap sample (see Fig. 13) causes the total $N_{\rm it}$ to soon overtake the values measured on a medium-thick Si cap sample. (b) Thinner Si cap samples show not only reduced mobility due to poorer interface passivation but also increased $C_{\rm ox}$ due to reduced hole displacement [see Fig. 1(b)]. (c) This tradeoff (mobility versus $C_{\rm ox}$) yields an optimum $I_{\rm ON}$ for a medium Si cap. The subthreshold swing is almost independent of the Si cap due to the higher $C_{\rm ox}$, reducing the effect of a poorer interface passivation for thin Si caps.

passivation: with a thinner Si cap, more Ge from the channel segregates to the interface [29], causing a higher density of preexisting interface states. This is shown in the inset in Fig. 19(a), where $N_{\rm it0}$ values extracted from CP measurements are reported for three different Si cap thicknesses. However, it is worth emphasizing that the higher $\Delta N_{\rm it}$ observed during NBTI stress for devices with thicker Si caps (see Fig. 13) quickly causes the interface quality of these samples to become worse than that of the devices with a reduced Si cap thickness [see Fig. 19(a)].

Furthermore, it is worth noting that a thinner Si cap, while causing a mobility reduction, increases the gate-stack $C_{\rm ox}$ due to reduced hole displacement [reduced $T_{\rm inv}$, see Fig. 1(b)], as shown in Fig. 19(b). When looking at the $I_{\rm ON}$ performance of the SiGe devices for different Si cap thicknesses, the best performance is typically obtained for a medium-thickness Si cap of around 1.2 nm, where a tradeoff between higher $C_{\rm ox}$ and reduced mobility is obtained [see Fig. 19(c)]. However, the $I_{\rm ON}$ stays within a $\pm 5\%$ range for the whole Si cap thickness range considered here. Looking at the subthreshold swing of the devices (which ultimately determines the $I_{\rm OFF}$ figure of merit when combined with the device $V_{\rm th0}$), a very small increase is observed for the thinnest Si cap [$<3\%$, Fig. 19(c)] due to the higher $C_{\rm ox}$, reducing the detrimental effect of a poorer interface passivation [see Fig. 19(a), inset].

In conclusion, the Si cap shows an overall limited impact on the $I_{\rm ON}/I_{\rm OFF}$ device metrics, whereas it has a dramatic impact on the device reliability. Therefore, when implementing a SiGe channel process, we suggest performing a Si cap thickness optimization based on performance/leakage metrics first and then reducing this thickness as slightly as needed to meet the NBTI reliability specifications.

IX. Conclusion

The NBTI reliability of Ge-based channel pMOSFETs has been investigated. The results clearly showed significantly

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

improved NBTI reliability for this family of high-mobility channel devices. A reliability-aware gate-stack optimization, with high Ge fraction, thick QW, and reduced Si cap thickness, was developed to demonstrate ultrathin EOT SiGe devices with ten-year NBTI reliability at operating $V_{\rm DD}$. The NBTI reduction was mainly ascribed to a favorable alignment shift of the Fermi level in the SiGe channel w.r.t. preexisting defect energy levels in the dielectric layers. The proposed model readily explains all the experimental observations. Finally, it was shown that the reliability improvement is obtained not to the detriment of the device performance. The extensive experimental results reported here strongly support SiGe technology as a promising candidate for future CMOS technology nodes, offering a solution to the reliability issue for ultrathin EOT devices.

REFERENCES

[1] V. Huard, M. Denais, and C. Parthasarathy, "NBTI degradation: From physical mechanism to modeling," *Microelectron. Reliab.*, vol. 46, no. 1, pp. 1–23, Jan. 2006.

[2] International Technology Roadmap for Semiconductors. [Online]. Available: http://public.itrs.net

[3] L.-Å. Ragnarsson, Z. Li, J. Tseng, T. Schram, E. Rohr, M. J. Cho, T. Kauerauf, T. Conard, Y. Okuno, B. Parvais, P. Absil, S. Biesemans, and T. Y. Hoffmann, "Ultra low-EOT (5Å) gate-first and gate-last high performance CMOS achieved by gate-electrode optimization," in *Proc. IEEE IEDM*, 2009, pp. 1–4.

[4] T. Ando, M. M. Frank, K. Choi, C. Choi, J. Bruley, M. Hopstaken, M. Copel, E. Cartier, A. Kerber, A. Callegari, D. Lacey, S. Brown, Q. Yang, and V. Narayanan, "Understanding mobility mechanisms in extremely scaled HfO_2 (EOT 0.42 nm) using remote interfacial layer scavenging technique and V_t-tuning dipoles with gate-first process," in *Proc. IEDM*, 2009, pp. 1–4.

[5] E. Cartier, A. Kerber, T. Ando, M. M. Frank, K. Choi, S. Krishnan, B. Linder, K. Zhao, F. Monsieur, J. Stathis, and V. Narayanan, "Fundamental aspects of HfO_2-based high-k metal gate stack reliability and implication on Tinv-scaling," in *Proc. IEDM*, 2011, pp. 18.4.1–18.4.4.

[6] M. Cho, J.-D. Lee, M. Aoulaiche, B. Kaczer, P. Roussel, T. Kauerauf, R. Degraeve, J. Franco, L. Ragnarsson, and G. Groeseneken, "Insight into negative and positive bias temperature instability (N/PBTI) mechanism in sub-nanometer EOT devices," *IEEE Trans. Electron Devices*, vol. 59, no. 8, pp. 2042–2048, Aug. 2012.

[7] L. Witters, S. Takeoka, S. Yamaguchi, A. Hikavyy, D. Shamiryan, M. Cho, T. Chiarella, L.-A. Ragnarsson, R. Loo, C. Kerner, Y. Crabbe, J. Franco, J. Tseng, W. E. Wang, E. Rohr, T. Schram, O. Richard, H. Bender, S. Biesemans, P. Absil, and T. Hoffmann, "8Å Tinv gate-first dual channel technology achieving low-V_t high performance CMOS," in *Proc. Symp. VLSI Technol.*, 2010, pp. 181–182.

[8] J. Mitard, L. Witters, M. Garcia Bardon, P. Christie, J. Franco, A. Mercha, P. Magnone, M. Alioto, F. Crupi, L.-Å. Ragnarsson, A. Hikavyy, B. Vincent, T. Chiarella, R. Loo, J. Tseng, S. Yamaguchi, S. Takeoka, W.-E. Wang, P. Absil, and T. Hoffmann, "High-mobility 0.85nm-EOT $Si_{0.45}Ge_{0.55}$-pFETs: Delivering high performance at scaled $V_{\rm DD}$," in *Proc. IEDM*, 2010, pp. 249–252.

[9] S. Krishnan, U. Kwon, N. Moumen, M. W. Stoker, E. C. T. Harley, S. Bedell, D. Nair, B. Greene, W. Henson, M. Chowdhury, D. P. Prakash, E. Wu, D. Ioannou, E. Cartier, M.-H. Na, S. Inumiya, K. Mcstay, L. Edge, R. Iijima, J. Cai, M. Frank, M. Hargrove, D. Guo, A. Kerber, H. Jagannathan, T. Ando, J. Shepard, S. Siddiqui, M. Dai, H. Bu, J. Schaeffer, D. Jaeger, K. Barla, T. Wallner, S. Uchimura, Y. Lee, G. Karve, S. Zafar, D. Schepis, Y. Wang, R. Donaton, S. Saroop, P. Montanini, Y. Liang, J. Stathis, R. Carter, R. Pal, V. Paruchuri, H. Yamasaki, J.-H. Lee, M. Ostermayr, J.-P. Han, Y. Hu, M. Gribelyuk, D.-G. Park, X. Chen, S. Samavedam, S. Narasimha, P. Agnello, M. Khare, R. Divakaruni, V. Narayanan, and M. Chudzik, "A manufacturable dual channel (Si and SiGe) high-k metal gate CMOS technology with multiple oxides for high performance and low power applications," in *Proc. IEDM*, 2011, pp. 28.1.1–28.1.4.

[10] K. J. Kuhn, "Considerations for ultimate CMOS scaling," *IEEE Trans. Electron Devices*, vol. 59, no. 7, pp. 1813–1828, Jul. 2012.

[11] B. Kaczer, J. Franco, J. Mitard, P. J. Roussel, A. Veloso, and G. Groeseneken, "Improvements in NBTI reliability of Si-passivated Ge/high-k/metal-gate pFETs," *Micr. Eng.*, vol. 86, no. 7–9, pp. 1582–1584, Jul.–Sep. 2009.

[12] J. Franco, B. Kaczer, M. Cho, G. Eneman, G. Groeseneken, and T. Grasser, "Improvements of NBTI reliability in SiGe p-FETs," in *Proc. IEEE IRPS*, 2010, pp. 1082–1085.

[13] J. Franco, B. Kaczer, G. Eneman, J. Mitard, A. Stesmans, V. Afanas'ev, T. Kauerauf, P. J. Roussel, M. Toledano-Luque, M. Cho, R. Degraeve, T. Grasser, L.-A. Ragnarsson, L. Witters, J. Tseng, S. Takeoka, W.-E. Wang, T. Y. Hoffmann, and G. Groeseneken, "6Å EOT $Si_{0.45}Ge_{0.55}$ pMOSFET with optimized reliability ($V_{\rm DD} = 1$ V): Meeting the NBTI lifetime target at ultra-thin EOT," in *Proc. IEDM*, 2010, pp. 4.1.1–4.1.4.

[14] A. Hikavyy, R. Loo, L. Witters, S. Takeoka, J. Geypen, B. Brijs, C. Merckling, M. Caymax, and J. Dekoster, "SiGe SEG growth for buried channel p-MOS devices," *ECS Trans.*, vol. 25, no. 7, pp. 201–210, 2009.

[15] M. Meuris, P. Mertens, A. Opdebeeck, H. Schmidt, M. Depas, G. Vereecke, M. Heyns, and A. Philipossian, "The IMEC clean: A new concept for particle and metal removal on Si surfaces," *Solid State Technol.*, vol. 38, no. 7, pp. 109–114, Jul. 1995.

[16] B. Kaczer, T. Grasser, P. J. Roussel, J. Martin-Martinez, R. O'Connor, B. J. O'Sullivan, and G. Groeseneken, "Ubiquitous relaxation in BTI stressing—New evaluation and insights," in *Proc. IRPS*, 2008, pp. 20–27.

[17] H. Reisinger, O. Blank, W. Heinrigs, A. Muhlhoff, W. Gustin, and C. Schlunder, "Analysis of NBTI degradation- and recovery-behavior based on ultra fast VT-measurements," in *Proc. IRPS*, 2006, pp. 448–453.

[18] B. Kaczer, T. Grasser, P. J. Roussel, J. Franco, R. Degraeve, L. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger, "Origin of NBTI variability in deeply scaled pFETs," in *Proc. IRPS*, 2010, pp. 26–32.

[19] J. Franco, B. Kaczer, G. Eneman, P. J. Roussel, T. Grasser, J. Mitard, L.-A. Ragnarsson, M. Cho, L. Witters, T. Chiarella, M. Togo, W. Wang, A. Hikavyy, R. Loo, N. Horiguchi, and G. Groeseneken, "Superior NBTI Reliability of SiGe channel pMOSFETs: Replacement gate, FinFETs, and impact of body bias," in *Proc. IEDM*, 2011, pp. 18.5.1–18.5.4.

[20] T. Chiarella, L. Witters, A. Mercha, C. Kerner, M. Rakowski, C. Ortolland, L.-A. Ragnarsson, B. Parvais, A. De Keersgieter, S. Kubicek, A. Redolfi, C. Vrancken, S. Brus, A. Lauwers, P. Absil, S. Biesemans, and T. Hoffmann, "Benchmarking SOI and bulk FinFET alternatives for planar CMOS scaling succession," *Solid State Electron.*, vol. 54, no. 9, pp. 855–860, Sep. 2010.

[21] J. Mitard, K. Martens, B. DeJaeger, J. Franco, C. Shea, C. Plourde, F. E. Leys, R. Loo, G. Hellings, G. Eneman, J. C. Lin, B. Kaczer, K. DeMeyer, T. Hoffmann, S. DeGendt, M. Caymax, M. Meuris, and M. M. Heyns, "Impact of Epi-Si growth temperature on Ge-pFET performance," in *Proc. ESSDERC*, 2009, pp. 411–414.

[22] T. Grasser and B. Kaczer, "Negative bias temperature instability: Recoverable versus permanent degradation," in *Proc. ESSDERC*, 2007, pp. 127–130.

[23] G. Groeseneken, H. E. Maes, N. Beltran, and R. F. De Keersmaecker, "A reliable approach to charge pumping measurements in MOS transistors," *IEEE Trans. Electron Devices*, vol. ED-31, no. 1, pp. 42–53, Jan. 1984.

[24] A. E. Islam, J. H. Lee, W.-H. Wu, A. Oates, and M. A. Alam, "Universality of interface trap generation and its impact on I_D degradation in strained/unstrained PMOS transistors under NBTI stress," in *Proc. IEEE IEDM*, 2008, pp. 1–4.

[25] M. Toledano-Luque, B. Kaczer, P. J. Roussel, T. Grasser, G. I. Wirth, J. Franco, C. Vrancken, N. Horiguchi, and G. Groeseneken, "Response of a single trap to AC negative bias temperature stress," in *Proc. IRPS*, 2011, pp. 4A.2.1–4A.2.8.

[26] T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P. Wagner, F. Schanovsky, J. Franco, M. T. Luque, and M. Nelhiebel, "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.

[27] M. V. Fischetti and S. E. Laux, "Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys," *J. Appl. Phys.*, vol. 80, no. 4, pp. 2234–2252, Aug. 1996.

[28] A. Stesmans and V. Afanas'ev, "ESR of interfaces and nanolayers in semiconductor heterostructures," in *Characterization of Semiconductor Heterostructures and Nanostructures*. Amsterdam, The Netherlands: Elsevier, 2008, pp. 435–489.

[29] M. Caymax, F. Leys, J. Mitard, K. Martens, L. Yang, G. Pourtois, W. Vandervorstb, and M. Meuris, "The influence of the epitaxial growth process parameters on layer characteristics and device performance in Si-passivated Ge pMOSFETs," *J. Electrochem. Soc.*, vol. 156, no. 12, pp. H979–H985, 2009.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[30] W. Gös, "Hole trapping and the negative bias temperature instability," Ph.D. dissertation, T.U. Wien, Vienna, Austria, 2012.
[31] A. S. Foster, F. Lopez Gejo, A. L. Shluger, and R. M. Nieminen, "Vacancy and interstitial defects in hafnia," *Phys. Rev. B*, vol. 65, no. 17, p. 174 117, 2002.

Jacopo Franco (S'10) received the M.Sc. degree in electronic engineering from the Università della Calabria, Cosenza, Italy, in 2008. He is currently with the reliability group of imec, Leuven, Belgium, working toward the Ph.D. degree at KU Leuven, Leuven.

Ben Kaczer received the Ph.D. degree in physics from The Ohio State University, Columbus, in 1998.

In 1998, he joined the reliability group of imec, Leuven, Belgium, where he is currently a Principal Scientist.

Philippe J. Roussel received the Diploma in electrical engineering from Industri Hogeschool of Ghent, Ghent, Belgium, in 1983.

In 1987, he joined imec, Leuven, Belgium, where he is currently working in a multidisciplinary team.

Jérôme Mitard received the Ph.D. degree in microelectronic engineering from the Polytechnic University School of Marseille, Marseille, France, in 2003.

He is currently with imec, Leuven, Belgium, working on the integration of Ge and III–V channels

Moonju Cho received the Ph.D. degree in materials science and engineering from Seoul National University, Seoul, Korea, in 2007.

Since 2007, she has been with the reliability group of imec, Leuven, Belgium.

Liesbeth Witters received the M.Sc. degree in chemical engineering from the Institut Français du Pétrole, Paris, France, in 1993.

Since 2001, she has been with imec, Leuven, Belgium, working on CMOS process development.

Tibor Grasser (SM'05) is currently a Professor at TU Wien, Vienna, Austria.

Guido Groeseneken (F'05) received the Ph.D. degree in applied sciences from KU Leuven, Leuven, Belgium, in 1986.

Since 1987, he has been with imec, Leuven. Since 2001, he has been also a Professor at KU Leuven.