SiGe Channel Technology: Superior Reliability toward Ultra-Thin EOT devices. Part I: NBTI

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

***Abstract*—We report extensive experimental results of the 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 a significantly improved NBTI robustness compared to Si-channel devices, which can solve the reliability issue for sub 1-nm EOT devices. A physical model is proposed to explain the intrinsically superior NBTI robustness.**

***Index Terms*—Ge, NBTI, pMOSFET, Reliability, SiGe.**

I. Introduction

NEGATIVE Bias Temperature Instability (NBTI) is considered as the most severe reliability issue for scaled CMOS technologies [1]. The quasi-constant supply voltage scaling proposed by the International Technology Road Map [2] for the recent technology nodes enhances NBTI due to the ever increasing interfacial oxide electric field (E_{ox}). As a consequence, although several groups have already demonstrated well behaving CMOS devices with aggressively scaled EOT down to 0.5nm [3,4], a 10 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, 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 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. Already in 2009 we have observed that the incorporation of Ge into the channel significantly improves the 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 ultra-thin EOT devices. In particular, *a reliability-oriented gate stack optimization with a high Ge fraction, a thick QW and thin Si passivation layer* is shown to boost the allowed gate voltage overdrive for 10 year lifetime above the expected operating V_{DD} for devices with ultra-thin EOT (down to ~0.6nm 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 ultra-thin EOT devices.

Manuscript received June 8, 2012. This work was supported in part by the European Commission under the 7th Framework Programme (Collaborative project MORDRED, contract No. 261868).

imec, Kapeldreef 75, 3001 Leuven, BELGIUM (phone: +32-16-28-1085; fax: +32-16-28-1214; e-mail: Jacopo.Franco@imec.be).

J. Franco and G. Groeseneken are also with the ESAT Dept., Katholieke Universiteit Leuven, BELGIUM.

T. Grasser is with the Technische Universität Wien, AUSTRIA.

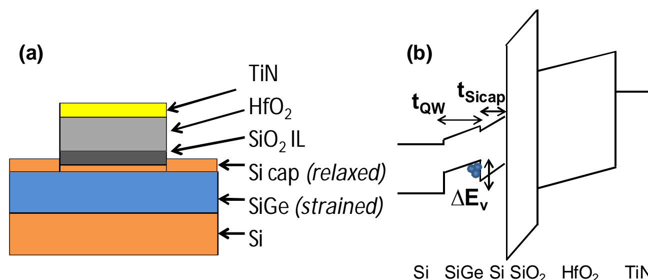

Fig. 1 (a) Sketch of the gate-stack of SiGe devices used in this work. (b) Band diagram sketch in inversion. Channel holes are confined into the SiGe QW due to the valence band offset (ΔE_v) between the SiGe channel and the Si cap. The Si cap thickness (t_{Sicap}) therefore contributes to the T_{inv}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

II. Experimental

The buried SiGe channel p-FETs used in this work were fabricated at 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 3nm and 7nm. 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 0.65nm and 2nm (as estimated from C-V curves and TEM pictures of the final device). A detailed description of the epi-process can be found elsewhere [14]. Gate stack fabrication started with a wet chemical oxidation (*imec clean* [15]) of the Si cap. On top of this IL, a ~1.8nm thick HfO_2 layer was deposited by ALD. Finally, a PVD TiN metal gate was deposited. 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 1.5x and 2.4x, depending on the gate-stack parameters [8].

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

NBTI stress experiments were performed using the extended measure-stress-measure technique [16]. The devices were stressed at T = 125°C with several gate overdrives, while the sensing bias was V_G=V_{th0}. To minimize NBTI relaxation effects for the device lifetime predictions, ΔV_{th} was evaluated at t_{relax}=1ms, 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, assumed at 30mV threshold voltage shift, was extracted. The 10 year lifetime operating overdrive (V_{op}) was then extrapolated by fitting a power law to the lifetime vs. gate overdrive dataset (see e.g. Figs. 2-4). For a dedicated experiment presented in section VI-D, an ultra-fast NBTI measurement setup (Keithley 4200 PMUs UF-BTI) was used to reduce the measurement delay to ~2us (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 varied separately in order to assess their individual impact on NBTI. In this preliminary set of experiments the EOT was not aggressively scaled (EOT~1.2nm). For comparison, a second set of standard Si channel devices with an identical gate stack was also used. Then the 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 the NBTI reliability. The extrapolated operating overdrive voltage for a 10 year lifetime (V_{op}) increased from 0.46V for the Si ref. up to 0.8V for 45% Ge fraction device with a SiGe layer thickness of 7nm and a Si cap thickness of 1.3nm. Increasing the Ge fraction to 55% while fixing the other parameters boosted the operating overdrive voltage even more, reaching 0.9V.

B. SiGe QW thicknesses

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

C. Si cap thicknesses

The most significant impact on the NBTI reliability was observed when varying the Si cap thickness (Fig. 4). Interestingly, a reduced thickness of this layer clearly improved the NBTI robustness. Naively one would expect the thinner Si cap to act as a reduced tunneling barrier for holes but conversely V_{op} increased from 0.82V to 1.14V when the Si cap thickness was decreased from 2nm to 0.65nm. This counter-intuitive 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 the NBTI reliability, also reduces the device T_{inv} (thanks to reduced hole displacement, see Fig. 1b) and therefore enhances the current drive performance.

Finally, we note that a distinct relation between the initial V_{th0} and the NBTI-caused ΔV_{th} is consistently observed in our

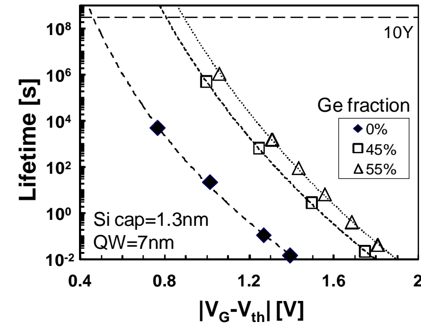

Fig. 2: Extrapolated lifetimes as a function of gate voltage overdrive for varying Ge content. A higher Ge fraction boosts the NBTI robustness.

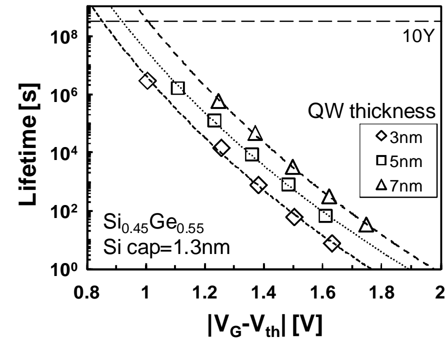

Fig. 3: Extrapolated lifetimes as a function of gate voltage overdrive for varying QW thickness, (55% Ge fraction, 1.3nm thick Si cap). A thicker QW boosts the NBTI robustness.

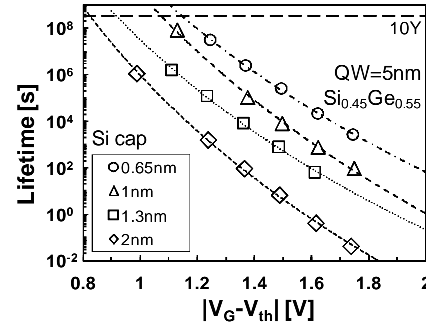

Fig. 4: Extrapolated lifetimes as a function of gate voltage overdrive for varying Si cap thickness (55% Ge fraction, 5nm thick QW) . A reduced Si cap thickness boosts the NBTI robustness while enabling T_{inv} reduction. The extrapolated V_{op} are plotted in Fig. 6 for fair benchmarking vs. T_{inv}.

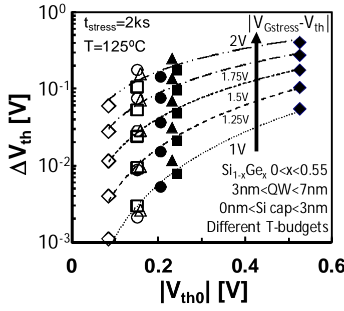

Fig. 5: A clear correlation between the initial V_{th0} and NBTI-caused ΔV_{th} is consistently observed on our SiGe devices with different gate-stacks: devices with lower initial V_{th0} always showed reduced V_{th} instability, at any given stress condition ($|V_{Gstress} - V_{th0}|$). This was not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

SiGe devices with different gate stacks (Fig. 5). Devices with lower initial V_{th0} always showed reduced V_{th} instability, at any given stress condition ($|V_{Gstress} – V_{th0}|$). 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_{op} extracted for different Si cap thicknesses are shown in a benchmark plot vs. the T_{inv} values (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_{op} boost together with a T_{inv} reduction. Such a V_{op} boost for reduced Si cap thickness was observed consistently for several IL thicknesses. This trend is clearly different w.r.t. the data collected on Si channel devices where a T_{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 ultra-thin EOT (10 year continuous operation at $|V_G-V_{th}|\approx0.6$V at $T_{inv}\approx1$nm, EOT≈0.6nm).

V. Process- and Architecture- Independent Results

The optimization presented above 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 (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. Si planar ref. when removing the Si cap (Fig. 8). Furthermore, the experiment with varying Si cap thickness

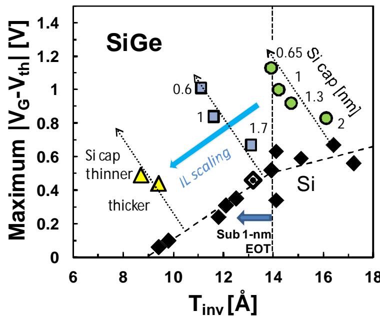

Fig. 6: Maximum operating overdrive for 10 year lifetime (T=125°C, failure criterion ΔV_{th}=30mV) vs. T_{inv} (evaluated at $V_G= V_{th}$-0.6V). SiGe devices with a thin Si cap offer improved NBTI reliability, i.e. higher V_{op}. Note: the open-diamond and the circles represent the Si ref. gate stack of Fig. 2 and the SiGe gate stacks of Fig. 4 respectively.

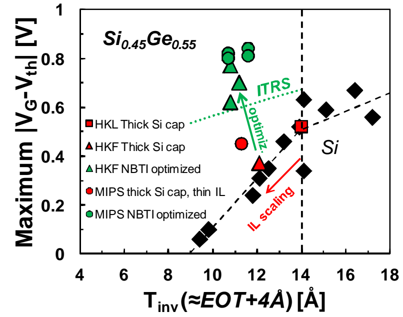

Fig. 7: A high Ge fraction (55%) in a 6.5nm thick QW, combined with a thin Si cap (0.8nm) boost V_{op} to meet the target V_{DD} at ultra-thin EOT in a MIPS flow (green circles, as compared to red circle). The optimization was also implemented in a RMG flow: high-k last SiGe sample with thick Si cap (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_{op} above the ITRS target. The results were reproduced for several process T-budgets.

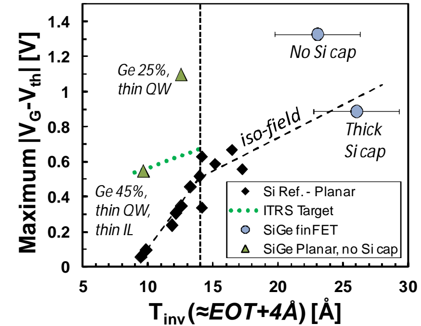

Fig. 8: SiGe channel bulk pFinFETs without a Si cap show improved NBTI reliability w.r.t. the same devices with a thick Si cap and w.r.t Si planar pFETs. The dashed trendline for T_{inv}>1.4nm demarcates planar Si pFET constant field scaling (*"iso-field"*). Uncertainty in T_{inv} is related to the finFET 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: $Si_{0.75}Ge_{0.25}$ 3nm-thick QW, and $Si_{0.55}Ge_{0.45}$ 3nm-thick QW with reduced IL thickness).

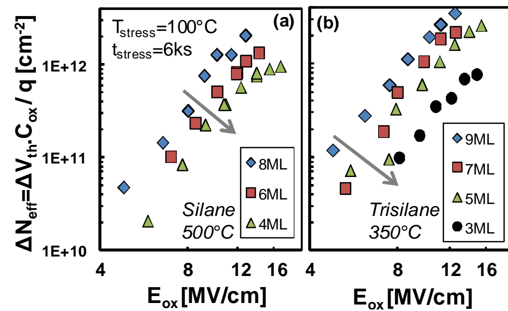

Fig. 9: A decreased thickness of the Si passivation layer improves the NBTI robustness of Ge pMOSFETs, independently of the Si cap epi-process used, (a) Silane 500ºC, or (b) Trisilane 350ºC.

was repeated on pure Ge channel pMOSFETs [21] with 4, 6, and 8 Si mono-layers (ML) epi-grown from Silane precursor at 500ºC. A reduced thickness of the Si-layer again resulted in a reduced NBTI at fixed stress conditions (electric field, stress time, stress temperature, sensing delay): 4 MLs devices degrade ~4× less than 8 MLs devices (Fig. 9a). The same trend was also observed for Si-caps grown using a 350°C epi-growth from Trisilane precursor: a ~8× NBTI reduction is observed when reducing the Si from 9 to 3MLs (Fig. 9b).

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_{ox}-acceleration

Fig. 10 shows the typical NBTI ΔV_{th} evolution vs. the stress time for the Si ref. and the SiGe devices with different Si caps. The SiGe devices show a significantly reduced ΔV_{th}, especially for the samples with reduced Si cap thickness. The NBTI ΔV_{th} evolution is often described as a power-law of the stress time ($\Delta V_{th}=A\ t_{stress}{}^{n}$), with a pre-factor A dependent on the stress E_{ox}, and an apparent exponent n typically reported in the range of 0.15 to 0.25 [1] depending on the relaxation allowed by the measurement delay [16]. Fig. 11 documents the extracted power-law pre-factors and exponents for all the devices here considered. The pre-factors clearly show a dramatic reduction for the SiGe devices, with a further reduction for a reduced Si cap thickness. Moreover, the pre-factors show a significantly stronger E_{ox}- acceleration for SiGe w.r.t. the Si ref., 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 ΔN_{it} exponent of ~0.25) and to a faster relaxation [16] in SiGe devices, as discussed later.

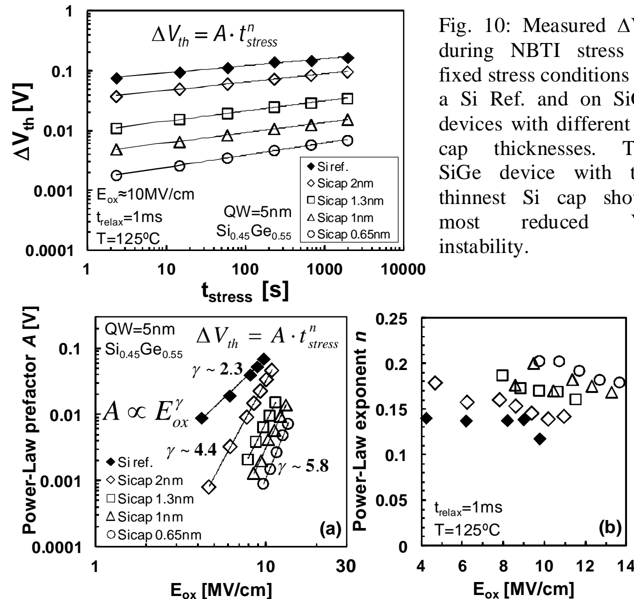

Fig. 10: Measured ΔV_{th} during NBTI stress at fixed stress conditions on a Si Ref. and on SiGe devices with different Si cap thicknesses. The SiGe device with the thinnest Si cap shows most reduced V_{th} instability.

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

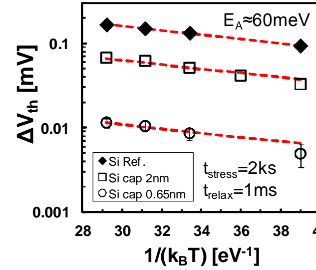

Fig. 12: NBTI-induced ΔV_{th} measured at different temperatures on the Si ref. devices, and on SiGe devices with two different Si cap thicknesses (2nm and 0.65nm). No clear difference in the apparent ΔV_{th}-activation energy is observed (extracted E_A≈60meV).

B. Temperature activation

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

C. Interface state creation (ΔN_{it}) and hole trapping (ΔN_{ot})

It was reported recently that NBTI is possibly ascribed to two components [22]: a recoverable (*R*) one related to hole trapping in pre-existing bulk oxide defects (ΔN_{ot}), and a so-called permanent one (*P*) typically associated with creation of new interfaces states (ΔN_{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 ΔN_{it} was monitored by CP, ΔN_{ot} was calculated by subtracting the ΔN_{it} contribution from the total ΔV_{th} measured. Fig. 13 reports ΔN_{it} and ΔN_{ot} evolutions measured on SiGe devices with two different Si cap thicknesses and on the Si ref. device for fixed stress conditions (E_{ox} = 10MV/cm, T=125ºC). Two main observations can be made: ΔN_{it} follows a power law of stress time with the same exponent of ~0.25 for the Si ref. 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 ΔV_{th}.

D. Faster NBTI relaxation

To investigate the supposedly faster NBTI relaxation suggested by the higher apparent ΔV_{th} time exponent for SiGe devices with thin Si caps (see Fig. 11b), a dedicated ultra-fast(minimum t_{relax}≈2μs) NBTI measurement was performed. Fig. 14 shows typical relaxation transients recorded on SiGe

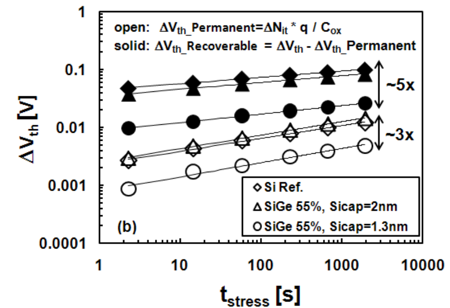

Fig. 13: Total ΔV_{th} split into the so-called permanent (*P*) ΔV_{th}, assumed to be caused by ΔN_{it}, and the recoverable (*R*) ΔV_{th}, assumed to be caused by filling of pre-existing oxide traps (N_{ot}). ΔN_{it} measured with charge pumping during NBTI stress were converted to $\Delta V_{th_Permanent}$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

devices with 2nm and 0.65nm Si cap after the same stress. While valence band electron re-population of interface states could be partially responsible for different ΔV_{th} recovery at short relaxation times (<1µs), differences in the relaxation shape are still found for longer relaxation times (>1µs): the relaxation transients fitted with the empirical Universal Relaxation model [16,22] show faster recovery for the 0.65nm Si cap devices. This can be observed also by looking at the ratio between the ΔV_{th} of the thick Si cap device vs. the ΔV_{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 non-zero measurement delay. This hypothesis is readily ruled out by observing the device current degradation during the stress (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.

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

(1) Reduced NBTI, i.e. lower power-law pre-factor;

(2) Stronger E_{ox}-acceleration;

(3) Similar temperature activation (E_A≈60meV) as Si ref.;

(4) Similar ΔN_{it} time exponent (~0.25) as Si ref.;

(5) Significantly reduced ΔN_{it} and ΔN_{ot}, with the latter reduction being of greater relevance.

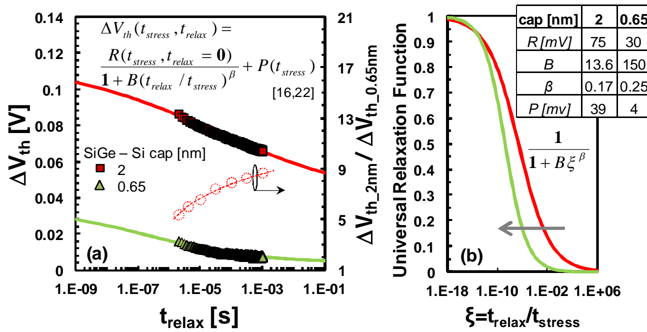

Fig. 14: (a) Ultra-fast NBTI relaxation transients recorded on SiGe devices with 2nm and 0.65nm Si cap after the same stress (stress conditions V_{ov}=1.95V, T=25C, t_{stress}=100s). (b) The transients fitted with the Universal Relaxation model [16,22] reveal faster relaxation for the thin cap device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

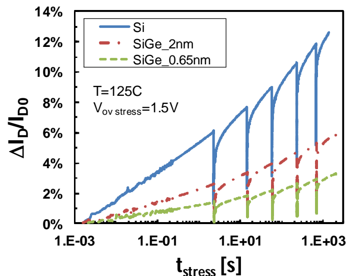

Fig. 15: Drain current degradation recorded on-the-fly during NBTI stress phases of the eMSM. SiGe devices with reduced Si cap thickness show lowest degradation also with zero-measurement delay, showing the improved NBTI is not an apparent effect due to faster relaxation behaviors. (note: the drops in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

(6) Slightly higher apparent ΔV_{th} time exponent due to strongly reduced ΔN_{ot} and faster relaxation.

VII. Model

Previous work attributed an improved NBTI robustness to induced strain at the interface [24]. This explanation does not apply to our devices: the SiGe layer thicknesses here considered 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% → reduced bandgap → higher valence band offset ΔE_v) and QW thickness (for a very thin QW, i.e. 3nm, 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 above, SiGe devices with reduced Si cap thickness show both reduced *P* and *R* components. In the next sub-sections, models for reduced *P* and *R* are proposed.

A. Reduced P (ΔN_{it})

N_{it} creation during NBTI stress is commonly attributed to de-passivation of H-passivated Si dangling bonds (P_{b0}) at the Si/SiO_2 interface. We have previously reported [13] Electron Spin Resonance Spectroscopy (ESR) [28] measured on a Ge substrate with a thick Si cap which revealed a high P_{b0} density (~1x10^{12} cm^{-2}), while 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_{it} precursor defect density and therefore reduce ΔN_{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 which is mainly related to a significant reduction of the *R* component, as noted above. The *R* reduction is discussed next.

B. R (ΔN_{ot}) - A 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. to the pre-existing bulk oxide defect energy levels (see Fig. 16). Larger misalignment can cause carriers to interact with a reduced density of oxide traps N_{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

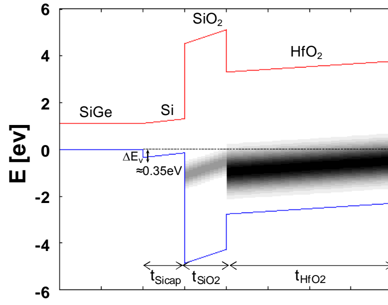

Fig. 16: A model including defect bands centered at 0.95eV and 1.4eV below the Si valence band in the IL and in the HfO_2 respectively (compatible with the expected O-vacancy levels). The channel Fermi level determines which part of the defect bands is accessible to channel holes. The defect band is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

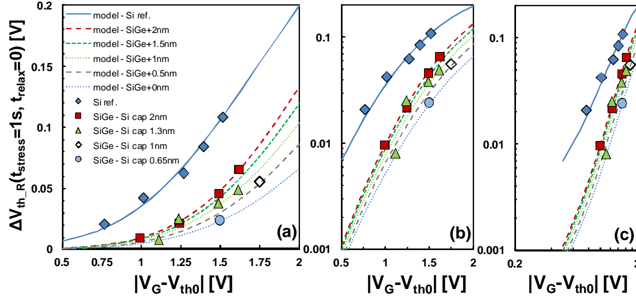

Fig. 17: Calculated ΔV_{th} vs. experimental data of the recoverable component. The model was first calibrated on the Si ref. data, then the same defect band parameters were used to calculate the expected ΔV_{th} for SiGe devices (including the valence band offset between the SiGe and the Si cap, and the voltage drop on different Si cap thickness). The simple model matches the experimental data remarkably well. (a) Lin-lin, (b) log-lin, (c) log-log scales.

value of the distributions were pinned at 0.95eV below the Si valence band for the IL (corresponding to the E'γ[$E_{0/+}$] center in SiO_2 [30]) and at 1.4eV below the Si valence band for the high-k (corresponding to the neutral oxygen vacancy [O^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, while all the defects below are neutral (note: no trapping/de-trapping kinetics is included in this calculation, i.e. thermodynamic equilibrium).

The model was first calibrated using the NBTI data on the Si ref. device: the standard deviations of the Gaussian distributions were used as a fitting parameter (in the range of 0.3~0.5eV) in order to capture the correct electric field dependence, while the defect densities were fitted in order to match the observed ΔV_{th} magnitude. The varying ΔV_{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 ΔV_{th} was calculated for SiGe channel devices, including the valence band offset of +0.35eV in the channel and including the varying voltage drop on Si cap with varying thicknesses.

As one can see in Fig. 17 the simple model matches excellently the experimental data relative to the recoverable component. The model readily captures: i) *the reduced NBTI;* ii) *the stronger field dependence* observed for SiGe devices with reduced Si cap thicknesses; iii). *the faster relaxation* observed for SiGe, since the higher energy difference between

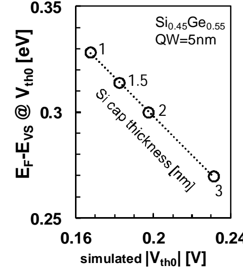

Fig. 18: MEDICI simulations show that a lower device V_{th0} corresponds to a higher channel Fermi level energy (E_F) w.r.t. the Si valence band (E_{VS}), which is beneficial for reducing the carrier-trap interaction, according to the model proposed in Fig. 16.

the defect levels and the SiGe channel is expected to enhance the trapped charge emission [26].

The model explains also 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_{th0} and the NBTI observed in SiGe devices (Fig. 5): as calculated with MEDICI for, e.g., a Si cap thickness split, gate-stacks with lower $|V_{th0}|$ have higher channel Fermi level energy (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 observed experimentally for planar devices and finFETs (Fig. 8).

VIII. Performance vs. 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 passivation: with a thinner Si cap more Ge from the channel segregates to the interface [29], causing a higher density of pre-existing interface states. This can be seen in the inset of Fig. 19a, where N_{it0} values extracted from CP measurements are reported for three different Si cap thickness. However, it is worth emphasizing that the higher ΔN_{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 the one of the devices with reduced Si cap thickness (Fig. 19a).

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

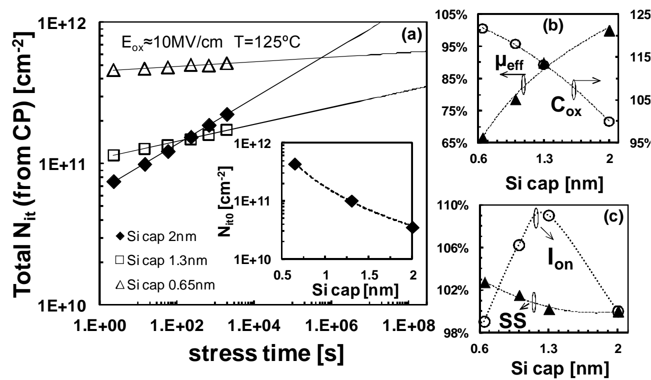

Fig. 19: (a) A reduced Si cap thickness yields a higher interface state density on the fresh device (N_{it0}, inset). However, the larger NBTI-induced ΔN_{it} on a 2nm thick Si cap sample (see Fig. 13), causes the total N_{it} to soon overtake the values measured on a medium-thick Si cap sample. (b) Thinner Si cap samples show reduced mobility due to poorer interface passivation, but also increased C_{ox} thanks to reduced hole displacement (see Fig. 1b). (c) This trade-off (mobility vs. C_{ox}) yields an optimum I_{ON} for a medium Si cap. The sub-threshold swing is almost independent of the Si cap, thanks to the higher C_{ox} reducing the effect of a poorer interface passivation for thin Si caps.

Fig. 19c), thanks to the higher C_{ox} reducing the detrimental effect of a poorer interface passivation (see Fig. 19a inset).

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

IX. Conclusions

The NBTI reliability of Ge-based channel pMOSFETs was investigated. The results clearly showed significantly 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 ultra-thin EOT SiGe devices with 10 year NBTI reliability at operating V_{DD}. The NBTI reduction was ascribed mainly to a favorable alignment shift of the Fermi level in the SiGe channel w.r.t. pre-existing 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 here reported strongly support SiGe technology as a promising candidate for future CMOS technology nodes, offering a solution to the reliability issue for ultra-thin EOT devices.

References

[1] V. Huard *et al.*, "NBTI degradation: from physical mechanism to modeling", in Micr. Rel., Vol. 46, No. 1, pp. 1-23, 2006;

[2] International Technology Roadmap for Semiconductors available at http://public.itrs.net;

[3] L.-Å. Ragnarsson *et al.*, "Ultra low-EOT (5Å) gate-first and gate-last high performance CMOS achieved by gate-electrode optimization", in *Proc.* IEEE IEDM, pp. 663-666, 2009;

[4] T. Ando *et al.*, "Understanding mobility mechanisms in extremely scaled HfO_2 (EOT 0.42nm) using remote interfacial layer scavenging technique and V_t-tuning dipoles with gate-first process", in *Proc.* IEDM, pp. 423-426, 2009;

[5] E. Cartier *et al.*, "Fundamental Aspects of HfO_2-based High-k Metal Gate Stack Reliability and Implication on tinv-Scaling", in *Proc.* IEDM, pp. 441-444, 2011;

[6] M. Cho *et al.*, "Insight into Negative and Positive Bias Temperature Instability (N/PBTI) mechanism in sub-nanometer EOT devices", in IEEE TED, Vol. 59, no. 8, pp. 2042-2048, 2012;

[7] L. Witters *et al.*, "8Å Tinv gate-first dual channel technology achieving low-V_t high performance CMOS", in *Proc.* Symp. on VLSI Technology, pp. 181-182, 2010;

[8] J. Mitard *et al.*, "Sub-nm EOT SiGe-55% pFETs for high-speed low-V_{DD} technology: a study from capacitor to circuit level", in *Proc.* IEDM, pp. 249-252, 2010;

[9] 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 *Proc.* IEDM, pp. 634-637, 2011;

[10] K.J. Kuhn, "Considerations for Ultimate CMOS Scaling", in IEEE TED, vol.59, no. 7, pp. 1813-1828, 2012;

[11] B. Kaczer *et al.*, "Improvements in NBTI Reliability of Si-passivated Ge/high-k/metal-gate pFETs", in Micr. Eng., vol. 86, no. 7-9, pp. 1582-1584, 2009;

[12] J. Franco *et al.*, "Improvements of NBTI reliability in SiGe p-FETs", in *Proc.* IEEE IRPS, pp. 1082-1085, 2010;

[13] J. Franco *et al.*, "6Å EOT $Si_{0.45}Ge_{0.55}$ pMOSFET with Optimized Reliability (V_{DD}=1V): Meeting the NBTI Lifetime Target at Ultra-Thin EOT , in *Proc.* IEDM, pp. 70-73, 2010;

[14] A. Hikavyy *et al.*, "SiGe SEG Growth For Buried Channel p-MOS Devices", ECS Transactions, Vol. 25, No. 7, pp. 201-210, 2009;

[15] M. Meuris *et al.,* "The IMEC clean: A new concept for particle and metal removal on Si surfaces", Solid S. Tech., 38(7), pp. 109-113, 1995;

[16] B. Kaczer *et al.*, "Ubiquitous Relaxation in BTI Stressing–New Evaluation and Insights", in *Proc.* IRPS, pp. 20-27, 2008;

[17] H. Reisinger *et al.*, "Analysis of NBTI Degradation- and Recovery-Behavior Based on Ultra Fast VT-Measurements", in *Proc.* IRPS, pp. 448-456, 2006**;**

[18] B. Kaczer *et al.*, "Origin of NBTI Variability in Deeply Scaled pFETs", in *Proc.* IRPS, pp. 26-32, 2011**;**

[19] J. Franco *et al.*, "Superior NBTI Reliability of SiGe Channel pMOSFETs: Replacement Gate, FinFETs, and impact of Body Bias", in *Proc.* IEDM, pp. 445-448, 2011;

[20] T. Chiarella *et al.*, "Benchmarking SOI and bulk FinFET alternatives for PLANAR CMOS scaling succession", in Solid-State El., Vol. 54, No. 9, pp. 855-860, 2010;

[21] J. Mitard, *et al.* "Impact of Epi-Si Growth Temperature on Ge-pFET Performance", in *Proc.* ESSDERC, pp. 411-414, 2009;

[22] T. Grasser and B. Kaczer, "Negative Bias Temperature Instability: recoverable versus permanent degradation", in *Proc.* ESSDERC, pp. 127-130, 2007;

[23] G. Groeseneken, H.E. Maes, N. Beltran, R.F. De Keersmaecker, "A reliable approach to Charge Pumping measurements in MOS transistors", in IEEE TED, Vol. 31, no. 1, pp. 42-53, 1984;

[24] A.E. Islam *et al.*, "Universality of Interface Trap Generation and Its Impact on I_D Degradation in Strained/Unstrained PMOS Transistors Under NBTI Stress", in *Proc.* IEEE IEDM, pp. 107-110, 2008;

[25] M. Toledano-Luque *et al.*, "Response of a Single Trap to AC Negative Bias Temperature Stress, in *Proc.* IRPS, pp. 364-371, 2011;

[26] T. Grasser *et al.*, "The Paradigm Shift in Understanding the Bias Temperature Instability: from Reaction-Diffusion to Switching Oxide Traps", in IEEE TED, Vol. 58, no. 11, pp. 3652-3666, 2011;

[27] M.V. Fischetti and S.E. Laux, "Band Structure, Deformation Potentials, and Carrier Mobility in Strained Si, Ge, and SiGe Alloys", in Journal of Applied Physics, Vol. 80, No. 4, pp. 2234-2252, 1996;

[28] A. Stesmans and V. Afanas'ev, "ESR of interfaces and nanolayers in semiconductor heterostructures", in Characterization of Semiconductor Heterostructures and Nanostructures, Elsevier, pp. 435-489, 2008;

[29] M. Caymax *et al.*, "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;

[30] W. Gös, "Hole Trapping and the Negative Bias Temperature Instability", Ph.D. dissertation, T.U. Wien, available at http://www.iue.tuwien.ac.at/phd/goes/dissse19.html;

[31] A. S. Foster, F. Lopez Gejo, A. L. Shluger, and R. M. Nieminen, "Vacancy and interstitial defects in hafnia", Phys. Rev. B (65), 2002.