BTI Reliability and Time-Dependent Variability of Stacked Gate-All-Around Si Nanowire Transistors

Adrian Chasin*, Jacopo Franco, Ben Kaczer, Vamsi Putcha[+], Pieter Weckx,
Romain Ritzenthaler, Hans Mertens, Naoto Horiguchi, Dimitri Linten
imec
[+]also at ESAT, KU Leuven
Leuven, Belgium
*phone: +3216288397, adrian.chasin@imec.be

Gerhard Rzepa
TU Wien
Vienna, Austria

***Abstract*—We report experimental results of the N/PBTI (Negative/Positive Bias Temperature Instability) reliability of vertically stacked Gate-All-Around (GAA) silicon nanowire (NW) MOSFETs. We benchmark the lifetime of these novel devices against FinFETs with different widths and similar gate-stack. We do not only compare the average degradation, but also the time-dependent variability. At last, we predict the impact of the nanowire diameter on the reliability using TCAD simulations. Both the experimental results and the simulations indicate that BTI reliability is not negatively impacted down to a nanowire diameter of 6nm.**

***Index Terms*—PBTI/NBTI, GAA, FinFETs, time-dependent variability, scaling**

I. Introduction

Cylindrical GAA transistors enable ultimate MOSFET scaling [1]. Short channel effects are substantially decreased as the natural length (λ) of such devices is reduced [1]. Horizontal nanowires exhibit strong similarities with the FinFET architecture. Therefore, horizontal GAA can be seen as the natural evolution of bulk FinFETs. Given their limited channel area, however, the normalized drive current of such devices is reduced. In order to tackle this limitation, several wires should be vertically stacked. Recently, unichannel [2] and CMOS [3] stacked nanowires with diameters of 8nm were fabricated on bulk Si substrates (cross-sections shown in **Fig.1**). These devices revealed excellent electrostatic control and effectively matched n- and p- V_{th}'s via a nanowire compatible dual-work function metal (DWFM) integration scheme [3]. In this work, we investigate the BTI reliability and time-dependent variability of these novel devices and compare them to FinFETs with similar gate-stack process.

II. Devices

We studied the reliability of five different GAA processes ('NW1-5'), shown in **Table 1**. All devices are based on the same Replacement Metal Gate (RMG) high-k process. Unichannel nMOS and pMOS devices have TiN/TiAl/TiN and TiN/TaN metal gate stacks, respectively, whereas CMOS processes (NW3, NW4 and NW5) follow the DWFM scheme schematically shown in **Fig.2**. The contribution of the underlying parasitic bulk device to the total current is reduced by adjusting its V_{th} to a higher value as compared to the V_{th} of the nanowire by means of a Ground Plane (GP) doping [2] at the beginning of the device fabrication. The nanowires reliability is compared to FinFETs based on the same RMG and extension/halo processes. The dimensions of both fins and nanowires were determined by TEM cross-sections (**Fig.1**).

TABLE I. Gate-All-Around processes used in this work. Five different wafers were used, in which two are unichannel devices (NW1 and NW2) and three are CMOS processes (NW3, NW4, NW5). These five devices are then compared with FinFETs with similar gate-stack process.

Nanowire number	Dimensions	Halo	Extension	p-metal etch duration	Ground Plane Doping
NW 1 (nMOS)	d=8nm	No	Yes	x	Strong
NW 2 (pMOS)	d=8nm	No	Extesionless	x	Strong
NW 3 (CMOS)	top wire: d=10nm bot. wire: d=8nm	Yes	Extensionless	180s	No
NW 4 (CMOS)	top wire: d=10nm bot. wire: d=8nm	Yes	Extensionless	360s	Weak
NW 5 (CMOS)	top wire: d=10nm bot. wire: d=8nm	No	Yes	360s	Strong

III. NBTI and PBTI

N/PBTI stress experiments have been performed using the eMSM technique [4] at 125C. ΔV_{th} is evaluated at t_{relax}=1ms. Maximum operating voltages ($V_{ov,max}$) are estimated by power-law extrapolations to 10 years (failure: ΔV_{th}=30mV). In order to determine the mean V_{th} shift, large test structures with 22/44 parallel fins/NWs and gate length (L_G) of 100nm are used to avoid stochastic effects [5]. At least 3 devices were measured for each stress voltage, so that the data shown is the average degradation of these devices.

Before comparing the reliability of FinFETs and NWs, we investigated the impact of the DWFM on the CMOS NW lifetime. In such process, the p-metal work function metal (WFM) is selectively etched on top of the n-type transistor. Given the new device architecture, the SF6 etch time was

tuned such that the PWFM could be effectively removed in between and below the nMOS nanowires as schematically shown on top of **Fig. 2.** However, extensive etching times could have, in principle, a negative impact on the device reliability as it could damage the gate-stack dielectric. **Fig. 2** shows the (a) ΔV_{th} at 1s of stress and (b) lifetime as a function of overdrive voltage ($V_{ov}=V_{G,stress}-V_{th0}$) for unichannel and nMOS devices with two different p-metal etching times during the DWFM process. $V_{ov,max}$ is independent of the etching time,

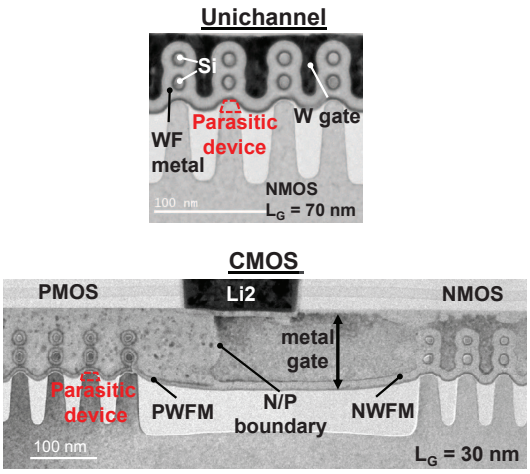

Figure 1. Cross-sections of the two NW processes used in this work. Unichannel [2] (top) and CMOS [3] (bottom) processes. Both devices are based on two-stacked horizontal nanowires. A parasitic fin, shown in red, is below each NW stack.

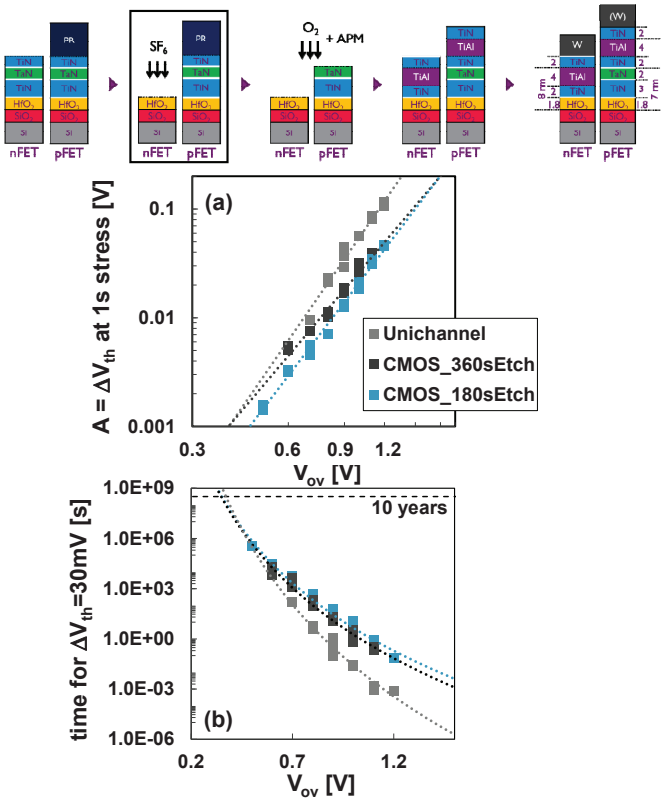

Figure 2. No impact of the p-metal etching time process (top schematics) on the nMOS device lifetime. PBTI reveal that both 180s and 360s SF6 etching results in the same lifetime as the unichannel nMOS device (b).

which validates the DWFM integration scheme.

Besides checking the integrity of the gate stack, it is also important to verify the impact of the ground plane doping on device reliability. This process step is simply based on an ion implantation on bare silicon wafer at the beginning of the process [2]. The GP doping objective is to suppress parallel conduction below the Si NWs in two ways: (1) by acting as a punch-through stopper, and (2) by increasing the V_{th} of the electrostatically less favorable planar-like transistor below the Si NWs (shown as red in **Fig.1**). The effectiveness of the GP doping can be indirectly verified by the device lifetime: if the device degradation depends on the GP doping strategy it means that the main conductions path also changes as a function of the GP process. Three different ground plane doping processes are then compared. NW4 and NW5 have medium and high doping concentrations, respectively, whereas in the NW3 process no GP doping was used. The impact of the ground plane process on device lifetime is verified and shown in **Fig.3**. BTI reliability is independent from GP doping, which indicates that all the three possibilities can be used to effectively turn-off the bulk parasitic device. It's important to mention that the parasitic device seems to be turned off even in NW3 where GP doping was not used, because of the presence of Halo in this device which acts like a substrate doping, replacing the initial GP doping.

Having confirmed that the DWFM CMOS integration scheme and the GP doping are efficient and do not interfere on device reliability, the GAA degradation is compared to FinFETs with different fins width. **Fig.4(a)** compares the threshold voltage shift at 1 second of NBTI stress of the three CMOS NWs process with FinFETs having different widths. In this case, the NBTI degradation of NWs seems to be enhanced. The voltage acceleration factor (γ) is smaller for the NWs, which indicates a broader distribution of defects in the gate dielectric or a higher oxide field (E_{ox}) for the same V_{ov}. A small but consistent trend is also observed for the FinFETs, showing a higher degradation for wider fins at the same V_{ov}. We ascribe this observation to the reduced oxide electric field (E_{ox}) obtained at constant V_{ov} in fully depleted fins, similarly to the effect of Forward Body Biasing (FBB) in planar devices [6, 7]. In fully depleted devices, the depletion charge is limited

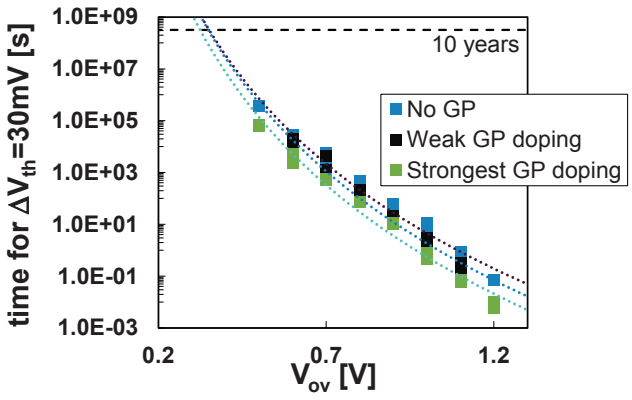

Figure 3: Comparison between NW3, NW4 and NW5 shows that device lifetime is independent from GP doping. It indicates that the parasitic device is off during device operation; it can be effectively turned off using GP doping, or by using Halo as a source of doping to the substrate (NW 3 – No GP).

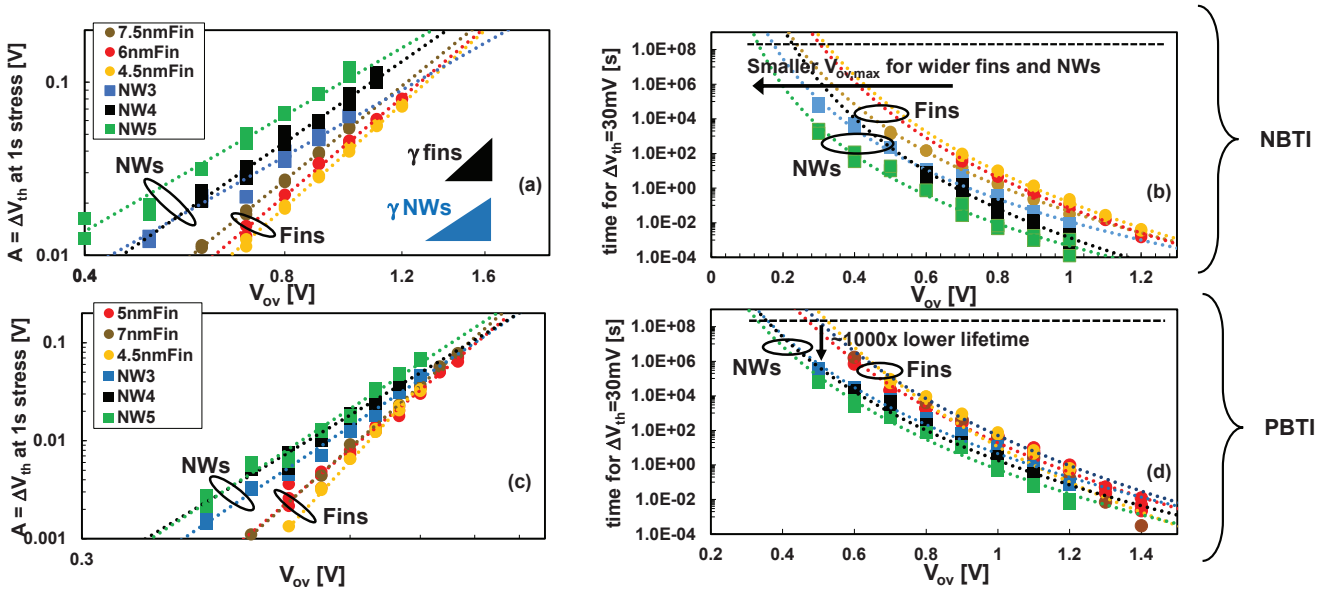

Figure 4. ΔV_{th} as a function of the overdrive voltage at 1s of stress for (a) NBTI and (c) PBTI. For both cases, NWs have higher degradation and smaller acceleration factor than FinFETs. (b) NBTI and (d) PBTI lifetime is smaller for NWs and wider Fins.

by the physical width of the device itself. For a fixed inversion charge Q_{inv} (as a result of the same V_{ov}), narrower fins will have a smaller depletion region, which will result in smaller E_{ox} ($Q_G=Q_{inv}+Q_{depl}$; $E_{ox}=Q_G/\varepsilon_0\varepsilon_r$). If the device lifetime is calculated (**Fig.4b**), the same trend is observed, as expected. GAAs and wider fins have smaller $V_{ov,max}$. Similar results are observed for PBTI. NWs have twice as large V_{th} shift at 1s of stress at the same V_{ov} as FinFETs and smaller γ (**Fig.4c**), resulting in a ~1000 times smaller lifetime at FinFETs' $V_{ov,max}$ (**Fig.4d**).

For a fair comparison, it is important to determine the EOT of these devices, as $\Delta V_{th} \sim E_{ox}^{\gamma}$ and $E_{ox} \sim V_{ov}/EOT$. However, EOT estimation is challenging in aggressively scaled non-planar devices due to uncertainty in the actual device dimensions. TEM cross sections are accurate, but cannot predict the physical variations intrinsically present in 3D devices. Small variation of EOT can result in drastic differences of E_{ox}, and consequently, ΔV_{th}. $J_{leakage}$ is a suitable indicator of EOT, as it is exponentially sensitive to gate dielectric thickness, while only linearly proportional to device dimensions, which is obtained by TEM cross-sections. **Fig.5** shows that $J_{leakage}$ for the NWs is, in average, ten times larger than the FinFETs, which indicates a thinner EOT. If we assume that the increase of NW $J_{leakage}$ is solely due to SiO_2 interfacial layer (IL) scaling, **Fig.6** is obtained. It shows that for both PBTI and NBTI, $V_{ov,max}$ follows the 1V/nm isoelectric field scaling, as previously observed in EOT scaling BTI studies [8]. Therefore, the lower $V_{ov,max}$ of the NWs is liked due to process differences and caused by a thinner SiO_x interfacial layer. The ~0.2nm thinner EOT could have been caused by a smaller growth rate of SiO_x on the different GAAs crystalline planes.

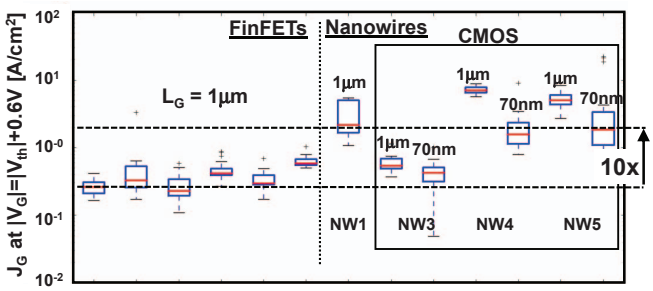

Figure 5. nMOS gate leakage current density. In average, NWs have a 10x higher current densities than the FinFETs which can indicate a thinner EOT. Each boxplot represents 20 devices measured in 5 different wafer positions.

IV. Time Dependent Variability

Having confirmed experimentally that ~8nm diameter NWs follow the same reliability trend as the FinFET devices, the next step is to compare their variability and how it changes with stress time.

In addition to the time-zero variability, an extra source of variability in deeply scaled devices arises during operation due to charging of pre-existing defects and generation of new defects [9]. It was shown that the statistics of this time-dependent variability of threshold voltage shifts ΔV_{th} can be expressed analytically in the so-called Defect-Centric (DC) model, which allows to derive a crucial relation between its first two moments $\langle\Delta V_{th}\rangle$ and $\sigma_{\Delta Vth}$, namely [9]:

$$\sigma^2_{\Delta Vth} = 2\eta\langle\Delta V_{th}\rangle \quad (1)$$

In this case, η represents the average contribution of a single charged trap to ΔV_{th}. As the variance increases as the population ages, the time dependent variability is best described in terms of η, which is independent of degradation.

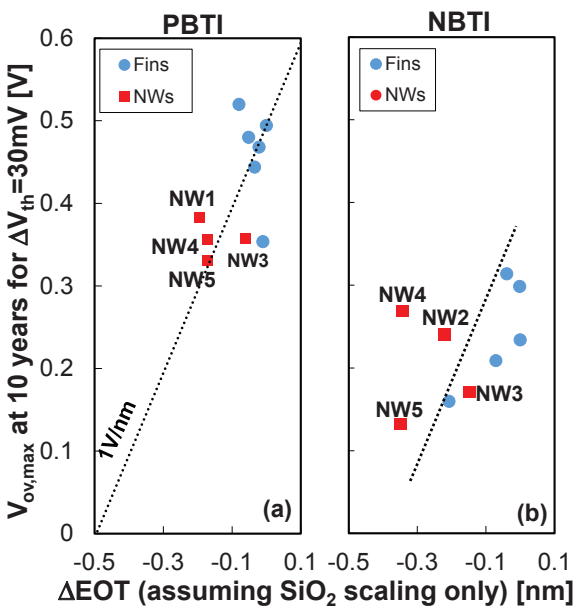

Figure 6. P/NBTI $V_{ov,max}$ for all measured devices as a function of the ΔEOT. The ΔEOT was calculated having the device with the lowest $J_{leakage}$ as reference and assuming that 0.2nm thinner SiO2 corresponds to a 10 times higher leakage current density. The dashed lines indicates the isoelectric field of 1V/nm which dictates the reliability for sub-nm EOT devices [6].

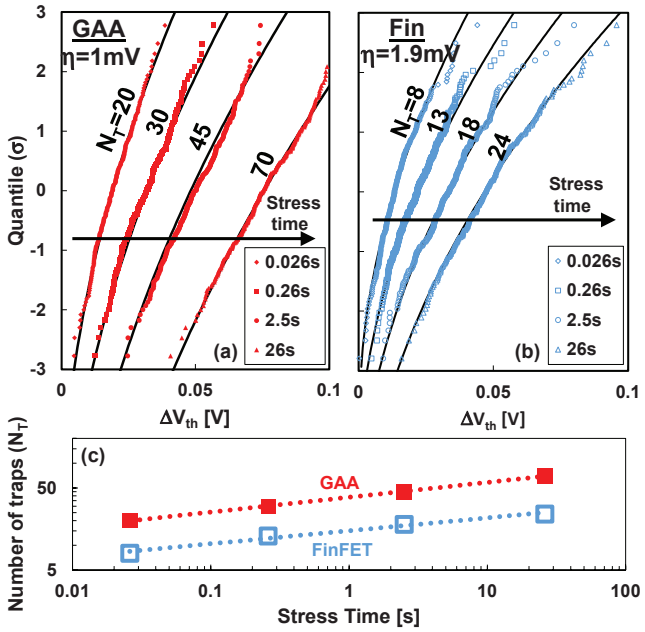

Figure 7. Data of 250 (a) GAA and (b) Fin nFETs stressed at PBTI condition of Vov=+1.2V for different stress times. The data can be fitted (black lines) with the defect-centric distribution with a single η. (c) The number of charged defects as a function of stress time can be fitted with a power law (dashed line) for both architectures. GAA have a higher number of N_T for the same stress time, in agreement with the higher mean degradation found previously in large devices.

Therefore, η can be used as metric for comparing different technologies. Besides η, the DC model allows to extract the number of charged traps per device by means of:

$$<\Delta V_{th}> = \eta N_T(t), \quad (2)$$

in which N_T is the expectation value of Poisson-distributed *n* charged defects in each device.

The measured devices have channel length (L_G) and number of fins/NWs equal to 28nm and 2/4, respectively. Similarly to the previous average BTI study, the experiments were performed using the eMSM technique at 125C. A continuous stress voltage of V_{ov}=+1.2V was used for 120 seconds and the ΔV_{th} is measured at relaxation time of 1ms.

A. Single devices

Initially, we compare the PBTI degradation of 10nm width FinFETs and NWs (NW3) by means of single device measurements.

Fig.7 shows the ΔV_{th} distribution of these two kind of devices as a function of stress time. As expected, the ΔV_{th} distributions can be described excellently for both technologies with the defect-centric model, assuming η of 1mV and 1.9mV, for the GAA and FinFET, respectively, and increasing N_T with stress time. **Fig. 7(c)** shows N_T as a function of stress time for both GAA and FinFET devices. GAA devices have more filled traps than FinFETs at the same stress time, which agrees with the higher mean degradation found in large devices in the previous section. The smaller η value of the GAA indicate that the defects in GAAs have smaller impact on the ΔV_{th} than in FinFETs, resulting in smaller time-dependent variability.

B. Matched Pairs

Similarly to its time-zero counterpart, time-dependent variability has systematic and random components. It was shown that the random component can be extracted using matched pairs [11], analogously to time-zero analysis [12]. In order to confirm the previous results obtained from ΔV_{th} of single devices, we also studied the PBTI of nMOS devices organized into matched pairs with varying gate length L_G and number of fins/wires. Instead of analysing the pure ΔV_{th} per device, we use the $\delta\Delta V_{th}$ which is simply the difference of the ΔV_{th} caused by stress between the left ($\Delta V_{th,L}$) and right ($\Delta V_{th,R}$) matched devices in the pair ($\delta\Delta V_{th} = \Delta V_{th,L} - \Delta V_{th,R}$). In this way, the random component of the time-dependent variability can be directly extracted.

Fig.8 shows the random component of both time zero and time-dependent variability of both 5nm width FiNFETs and NWs (NW3) extracted from the matched pairs. **Fig.8a** shows that time zero variability of both architectures are very similar and follow the expected $A_{-0.5}$ dependency (dashed line) [12] with a Pelgrom coefficient A_{VT} ~ 2 mV.μm. **Fig.8b** reveals the random component of the time-dependent variability. It follows the expected A_{-1} trend (dashed lines) [11]. Values of η for the smallest area devices are in close agreement with the one previously extracted based on single device distributions (Section A). It indicates that the measured devices do not deeply suffer from systematic variability (most probably because the single devices were located in close neighbours dies). Most important, it confirms that the GAA devices have lower η than FinFETs, and therefore, are less

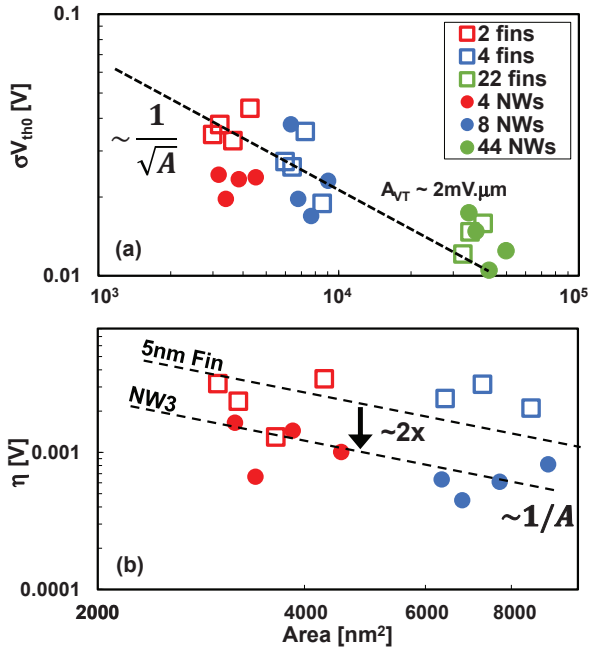

Figure 8. The random component of both (a) time-zero and (b) time-dependent variabilities, extracted using matched pairs for both GAA (solid symbols) and 5nm width FinFETs (open symbols), follow area scaling.

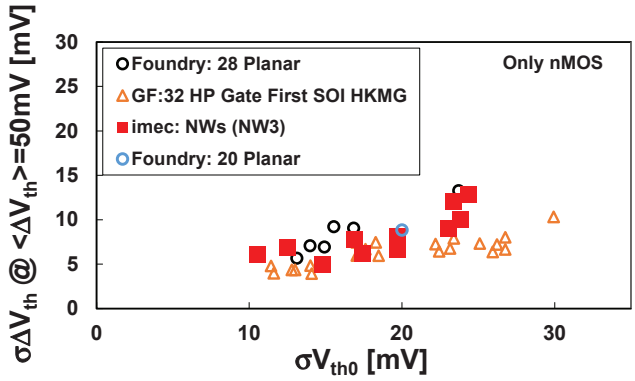

Figure 9. Benchmark σV_{th0} and $\sigma\Delta V_{th}$ for nMOS devices of different technologies. NWs discussed here (red squares) show relatively low $\sigma\Delta V_{th}$

impacted by time-dependent variability. This phenomena can be related to the better electrostatic control of the GAA [10] and/or to the fact that the current in GAA is not exactly at the interface with the gate stack dielectric due to volume inversion, being physically distant to the charged defect [1] and, therefore, less prone to its effects.

Finally, the relative time-dependent variability ($\sigma\Delta V_{th}/\sigma V_{th0}$) is benchmarked in **Fig.9** for nMOS devices of different technologies [13]. NW3 shows relatively low $\sigma\Delta V_{th}$, which is comparable to planar SOI devices in production.

V. Simulations

The impact of NW diameter on the device reliability is investigated with TCAD [14]. The defect model is based on the four-state non-radiative multiphonon model [15] calibrated for the recoverable component ($\Delta V_{th,R}$) of our RMG HK gate-stack. The simulated $\Delta V_{th,R}$ is based on a normally distributed defect level and uniformly distributed traps within SiO_x (hole traps) and HfO_x (electron traps) layers [16]. In this case, the distribution of defect levels E_t, (with regard to the respective valence band edge) are 3.8eV and 4.44eV, for the HfO_x and SiO_2, respectively. The defect

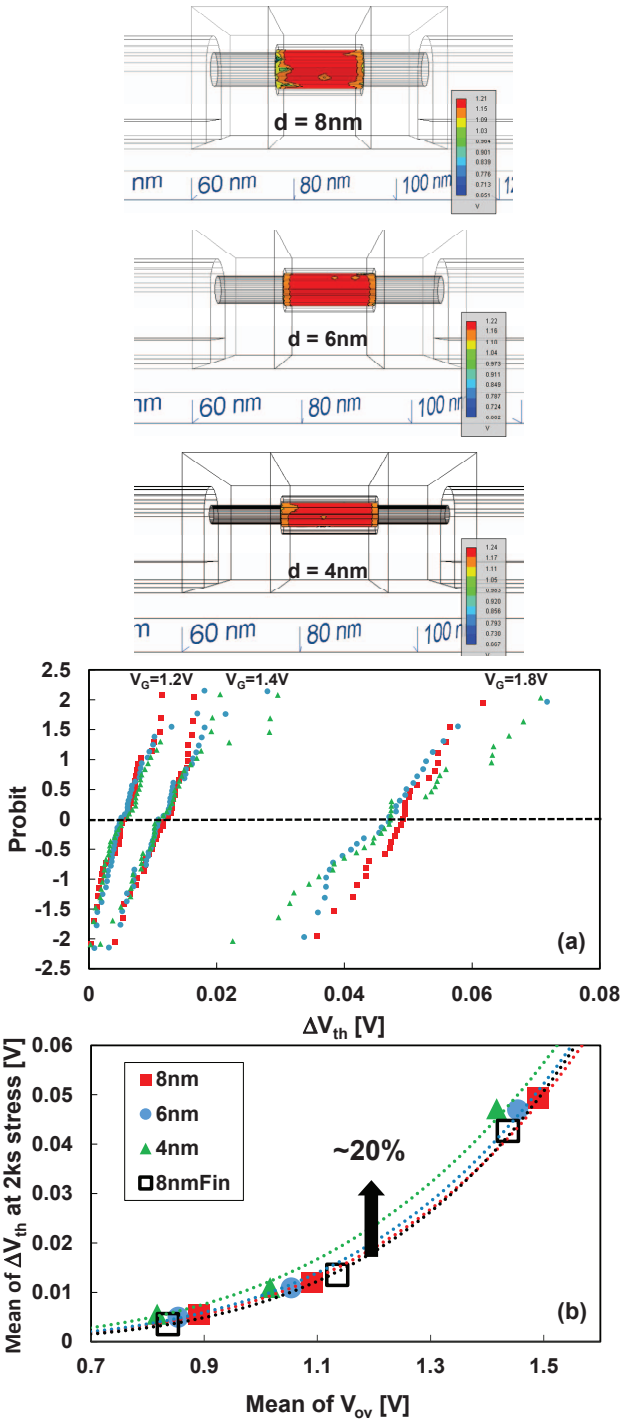

Figure 10. (a) Simulated $\Delta V_{th,R}$ of nMOS NWs with three different diameters for three different stress voltages. Each simulation is based on a randomly generated defect distribution. For each stress voltage, the mean degradation and V_{ov} is calculated which are then used to verify the impact of the NW dimensions on the PBTI reliability shown in (b).

concentration N_{OT}, on the other hand, is equal to $2.0x10^{20}$ and $0.9x10^{20}$ for the HfO_x and SiO_2, respectively [16].

The simulation methodology mimics the experimental one, in which ~30 devices are simulated to extract both the variability and the mean degradation, as shown in **Fig.10a** for NWs with three different diameters and stress voltages. Each simulation is based on a randomly generated defect distribution. The variability increases as a function of the mean degradation, as expected by the DC model. If the mean degradation is compared, 4nm wires have a 20% higher PBTI degradation than 8nm diameter/width NW/FinFETs (**Fig.10b**). This higher degradation for smaller wires is directly related to the increase of oxide electric field as shown in **Fig.11**. Electrostatics dictates that in cylindrical coordinates, the oxide field is not constant within the gate dielectric, increasing for smaller wires at the same stress voltage. Therefore, electric field crowding in scaled wires can be a potential threat for their reliability.

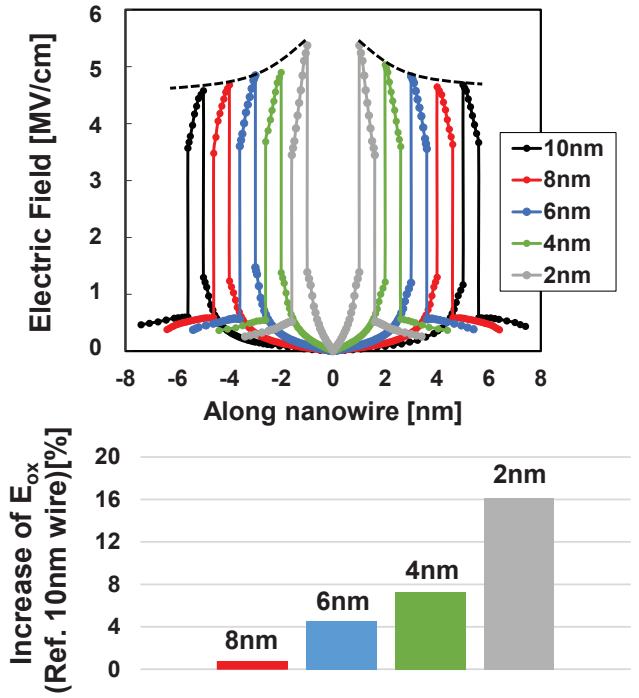

Figure 11. (a) Simulated oxide electric field within a NW for 5 different diameters. Due to electric field crowding in cylindrical coordinates, the oxide field is not constant within the gate dielectric and (b) it increases dramatically for scaled nanowires.

VI. Conclusions

In this work, we investigated the P/NBTI reliability of GAA devices. Two main experimental results were obtained: first, after interlayer SiO_2 thinning correction, NWs have comparable mean P/NBTI lifetime as FinFETs. Second, the time dependent variability of NWs is about 2 times smaller than 5nm and 10nm width FinFETs with similar gate-stack process. Such phenomena can be related to better electrostatic control and/or volume inversion in these novel devices. Finally, simulations based on a calibrated four-state non-radiative multiphonon model forecast no substantial PBTI enhancement in scaled NWs down to a diameter of 6nm. Further scaling is likely to be limited by BTI enhancement caused by the oxide electric field increase which is inherent to cylindrical devices.

Acknowledgment

GTS (Global TCAD Solutions) represented by F. Schanovsky and C. Kernstock are acknowledged for their support.

References

[1] J.-P. Colinge and J.C. Greer, *Nanowire Transistors*, Cambridge University Press, 2016.

[2] H. Mertens, R. Ritzenthaler, A. Hikavyy, M. S. Kim, Z. Tao, K. Wostyn, S. A. Chew, A. De Keersgieter, G. Mannaert, E. Rosseel, T. Schram, K. Devriendt, D. Tsvetanova, H. Dekkers, S. Demuynck, A. Chasin, E. Van Besien, A. Dangol, S. Godny, B. Douhard, N. Bosman, O. Richard, J. Geypen, H. Bender, K. Barla, D. Mocuta, N. Horiguchi, A.V-Y Thean, "Gate-All-Around MOSFETs based on Vertically Stacked Horizontal Si Nanowires in a Replacement Metal Gate Process on Bulk Si Substrates", in VLSI, p.158, 2016.

[3] H. Mertens, R. Ritzenthaler, A. Chasin, T. Schram, E. Kunnen, A. Hikavyy, L.-Å. Ragnarsson, H. Dekkers, T. Hopf, K. Wostyn, K. Devriendt, S. A. Chew, M. S. Kim, Y. Kikuchi, E. Rosseel, G. Mannaert, S. Kubicek, S. Demuynck, A. Dangol, N. Bosman, J. Geypen, P. Carolan, H. Bender, K. Barla, N. Horiguchi, D. Mocuta, "Vertically Stacked Gate-All-Around Si Nanowire CMOS Transistors with Dual Work Function Metal Gates", in IEDM, p.524, 2016.

[4] B. Kaczer, T. Grasser, Ph. J. Roussel, J. Martin-Martinez, R. O'Connor, B. J. O'Sullivan, G. Groeseneken, "Ubiquitous Relaxation in BTI stressing – new evaluation and insights", in IRPS, p.20, 2008.

[5] B. Kaczer, T. Grasser, Ph. J. Roussel, J. Franco, R. Degraeve, L.-A. Ragnarsson, E. Simoen, G. Groeseneken, H. Reisinger, "Origin of NBTI Variability in Deeply Scaled pFETs", in IRPS, p.915, 2010.

[6] J. Franco, B. Kaczer, A. Chasin, H. Mertens, L.-Å. Ragnarsson, R. Ritzenthaler, S. Mukhopadhyay, H. Arimura, Ph.J. Roussel, E. Bury, N. Horiguchi, D. Linten, G. Groeseneken, and A. Thean, "NBTI in Replacement Metal Gate SiGe core FinFETs: impact of Ge concentration, fin width, fin rotation and interface passivation by high pressure anneals", in IRPS, p.4B-2-1, 2016.

[7] J. Franco, B. Kaczer, G. Eneman, Ph.J. Roussel, T. Grasser, J. Mitard, L.-Å. Ragnarsson, M. Cho, L. Witters, T. Chiarella, M. Togo, W.-E 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 IEDM, p.445, 2011.

[8] Moonju Cho, Jae-Duk Lee, Marc Aoulaiche, Ben Kaczer, Philippe Roussel, Thomas Kauerauf, Robin Degraeve, Jacopo Franco, Lars-Åke Ragnarsson, and Guido Groeseneken, "Insight Into N/PBTI Mechanisms in Sub-1-nm-EOT Devices", T-ED, vol. 59, p.2042, 2012.

[9] B. Kaczer, Ph. J. Roussel, T. Grasser, and G. Groeseneken, "Statistics of Multiple Trapped Charges in the Gate Oxide of Deeply Scaled MOSFET Devices—Application to NBTI", EDL, vol. 36, p.300, 2010.

[10] Andrea Ghetti, Christian Monzio, Compagnoni, Alessandro S. Spinelli, Angelo Visconti, "Comprehensive Analysis of Random Telegraph Noise Instability and Its Scaling in Deca–Nanometer Flash Memories", TED, vol. 56, p.1746, 2009.

[11] Ben Kaczer, Jacopo Franco, Philippe J. Roussel, Guido Groeseneken, Thomas Chiarella, Naoto Horiguchi, and Tibor Grasser, "Extraction of the Random Component of Time-Dependent Variability Using Matched Pairs", EDL, vol. 35, p.300, 2015.

[12] M. J.M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," IEEE J. Solid-State Circuits, vol. 24, no. 5, pp. 1433–1439, 1989.

[13] P. Weckx, B. Kaczer, P. Raghavan, J. Franco, M. Simicic, Ph. J. Rousse, D. Linten, A. Thean, D. Verkest, F. Catthoor, and G. Groeseneken, "Characterization and simulation methodology for time-dependent variability in advanced technologies", in CICC, 2015.

[14] www.globaltcad.com

[15] T.Grasser, “Stochastic charge trapping in oxides: From random telegraph noise to bias temperature instabilities”, Microelectronics Reliability, vol. 52, p.39, 2012.
[16] G. Rzepa, M. Walt, W. Goes, B. Kaczer, J. Franco, T. Chiarella, N. Horiguchi, and T. Grasser, “Complete Extraction of Defect Bands Responsible for Instabilities in n and pFinFETs” in VLSI, p.208, 2016