Experimental characterization of BTI defects

B. Kaczer1, V. V. Afanas’ev2, K. Rott3,4, F. Cerbu2, J. Franco1, W. Goes4, T. Grassner4, O. Madia2, A. P. D. Nguyen2, A. Stesmans2, H. Reisinger3, M. Toledano-Luque1, P. Weckx1,2

1imec, Leuven, Belgium
2KU Leuven, Leuven, Belgium
3Infineon, Munich, Germany
4TU Wien, Vienna, Austria

Abstract—Selected physical and electrical characterization of gate-oxide defects is reviewed. Interfacial dangling bond defects at interfaces of Si and SiGe are characterized electrically and compared to the electron spin resonance data. It is found that additional silicon dangling bonds (Pb0 centers) acting as amphoteric interface traps are generated upon oxygen scavenging anneal. A close density of Pb0 centers is also found in Si/SiGe/Si/HfO2 stack. The Ge dangling bond defects (Ge Pb1 centers) found at interfaces of SiGe with thermal SiO2 are shown to behave as acceptor center traps. The properties of individual gate-oxide defects in FETs are studied with Time-Dependent Defect Spectroscopy (TDDS), an electrical characterization technique. The primary properties of a defect are its capture and emission time constants, both dependent on the local electric gate field and the temperature. Considering the wide distributions of these defect properties, complete characterization of many defects is required over the full gate and drain voltage range from 0 to VDD at different substrate doping, and with the decreasing FET area. The average impact per charged defect is found to be reduced in SiGe substrates, contributing to lower degradation of SiGe-based FETs. In low-doped substrates, other sources of variability, such as interface states, are shown to take over. The impact of a single charged gate-oxide defect on the full I-V characteristic is measured and corroborated by atomistic device simulations, thus enabling the path to compact modeling of degraded deeply-scaled FET devices. Finally, the model explaining the correlation between drain and gate current RTN is discussed.

I. INTRODUCTION

The so-called Bias-temperature instability (BTI) in CMOS technologies with both conventional and advanced gate stacks remains at the forefront of reliability concerns. The device instability arises from charging of defects (traps) in the gate dielectric and at its interface with the substrate. Here we review several recent insights into the properties of these traps obtained with both electrical and physical characterization techniques. In Section II, the properties of interface defects are studied in large devices. In Sections III and IV, we discuss the temporal and electrostatic properties of individual bulk oxide traps, as investigated in deeply-scaled devices.

II. SI/AND SiGe/SiO2 INTERFACE TRAPS

Compared to the workhorse (100) face, transport along the (110) crystallographic plane of Si offers enhancement in hole mobility as well as a more compact architecture of metal-oxide-semiconductor (MOS) field-effect transistor (MOSFET), e.g., vertical MOSFET and FinFET devices, while remaining within the framework of the highly developed Si MOS fabrication process. We evaluated the trap density at interfaces $D_{vth}$ of (110), (111) and (100)Si with thermal SiO2 systems by using conductance and capacitance methods. The results of these experiments are illustrated by the experimental $D_{vth}(E)$ profiles shown in Fig. 1. The inferred $D_{vth}$ is found highly sensitive to the Si crystal orientation, highest for the (111) face, and lowest for the (100) face, with (110)Si/SiO2 closely resembling the (111)Si/SiO2 case. Two peaks in the $D_{vth}(E)$ profile within Si band gap, at about 0.25 eV and 0.85 eV above the valence band (VB), are observed for all three orientations and correlate with the densities of Electron-Spin Resonance (ESR) active Pb interface defects (dangling bonds of the surface Si atoms), responsible for the majority of amphoteric interface traps.

As the next step, we addressed the effect of oxygen scavenging on the interface defect density in p-Si/HfO2 (1.8 nm)/TiN/poly-Si structures. The scavenging has been performed by applying laser annealing to the samples in O2-free ambient at 700, 900, or 1100 °C. Since these structures have aggressively downscaled insulator thickness (EOT=0.6 nm), they exhibit considerable leakage current which precluded us from the application of conventional interface trap characterization techniques. For the p-Si/HfO2 (1.8 nm) interface the total $N_{th}$ of about $5\times10^{12}$cm$^{-2}$ is estimated.

Despite the absence of reliable electrical data, density of interface defects can still be evaluated by ESR. The ESR measurements allowed us to estimate density of Pb0 centers which, as it has been discussed above, provide the major contribution to interface trap density at the interfaces of thermally oxidized silicon. Densities of paramagnetic Pb0 centers found from these measurements are summarized in Fig. 2 for 1.8 nm thick HfO2 and ZrO2 insulators. The results indicate an increase in the density of Pb0 centers after high-temperature.
annealing of Si/HfO₂ interface suggesting its structural and electrical degradation. Interestingly, if performing the O scavenging annealing by using spike-processing in He at 1035 °C (cf. Fig. 2), no measurable interface damage is found. This observation points towards interface degradation mechanism triggered by formation of Si-O molecules by reaction of a SiO₂-like IL with Si substrate:

\[
\text{SiO}_2 + \text{Si} \rightarrow 2\text{SiO(gas)},
\]

which has earlier been identified as the primary mechanism of degradation of Si/SiO₂ interfaces upon annealing in the O-deficient ambient. Application of He gas during annealing allows one to significantly suppress this kind of degradation [1]. No traces of E' centers which are frequently invoked to explain charge trapping in the oxides on silicon and other degradation phenomena are found in the studied samples.

The interaction of molecular hydrogen with dangling bond defects represents the key technological processing allowing one to reduce the density of harmful interface traps to the device-acceptable level. On the other hand the release (re-activation) of the passivated dangling bonds is considered nowadays as the basic process of electrical degradation of semiconductor interfaces including BTI. In an attempt to verify the applicability of this concept beyond the well-studied Si dangling bond defects, we addressed dangling bonds of germanium atoms (P₅-type defects) in Si₁₋ₓGeₓ alloys (x = 0.75) at the interface with well-studied thermal SiO₂ insulator as affected by annealing in molecular hydrogen (passivation) and by the subsequent annealing in high vacuum (de-passivation).

Analysis of the density of the Ge dangling bond defects remaining electrically and ESR active was conducted upon isochronal anneals in the range of temperatures from 100 to 425 °C shown in Fig. 3, as well as kinetics of isothermal passivation and de-passivation (cf. Fig. 4) using the standard first-order kinetics equations. These observations were conducted using both the electrical detection of the Ge P₅ centers as acceptors in SiGe [2] and by ESR. In general, both passivation and de-passivation processes are observed to occur at lower temperatures than at the Si/SiO₂ interfaces. The activation energies of germanium dangling bond passivation by H₂ molecule and the activation energy of Ge-H bond thermal dissociation were found equal to \( E_f = 1.44 \pm 0.04 \) eV and \( E_d = 2.23 \pm 0.04 \) eV, respectively.

Contrary to the well-studied case of dangling bonds of silicon atoms at the interfaces with the same insulator (SiO₂),
the sum of these two energies appears to be significantly lower than the thermodynamic binding energy of the H₂ molecule (4.4 eV). This discrepancy might be explained either by different charge state of the Ge Pb-type defects (negatively charged) during passivation or, else, by the presence of additional H₂ cracking centers. Furthermore, the statistical (site-to-site) spread of the activation energies for Ge Pb-center passivation and de-passivation appears to be significantly larger (by a factor of about 3) than for Si dangling bonds. This large variability makes it impossible to achieve the complete passivation of the Ge dangling bonds, about 40 % of which remain electrically active even after highest thermal budget of the H₂ passivation anneal (425 °C).

III. INDIVIDUAL TRAP PROPERTIES

The understanding of oxide trap behavior is crucial for a number of reliability issues, like the bias temperature instability, hot carrier degradation, time-dependent dielectric breakdown, random telegraph and 1/f noise. Recent results have demonstrated that hole capture and emission into oxide traps in pMOS transistors are more complicated than the usually assumed Shockley-Read-Hall-like process [3, 4]. In particular, both charging and discharging proceed via a non-radiative multiphonon (NMP) mechanism involving metastable defect states.

The Time-Dependent Defect Spectroscopy (TDDS) is an electrical characterization technique allowing the extraction of properties of individual gate-oxide defects in FETs. The primary properties of a defect are its capture and emission time constants, both dependent on the local electric gate field and the temperature. Considering the wide distributions of these defect properties, complete characterization of hundreds of defects is required over the full gate and drain voltage range from 0 to VDD and also at different temperatures. This is desirable in order to model degradation for large, “analog” MOSFETs and to model the ΔVth-variability of deeply-scaled FETs.

Employing the “conventional” TDDS technique illustrated in Fig.5, and by varying temperature and gate bias we could determine the field dependence and thermal activation of individual defects [5].

![Fig. 5. Illustration of the 3 different measuring schemes used by TDDS: Shown left is the conventional DC TDDS measurement procedure, during which defects are charged at a stress voltage \( V_s \) and discharged during the subsequent recovery at \( V_r \). This sequence is typically repeated 100 times to allow for a statistical analysis of the discrete emission events. In the middle, the dynamic AC TDDS measurement is shown, where defects are subjected to an AC signal switching between \( V_s \) and \( V_r \), followed by a discharge period at \( V_r \). Finally, on the right, for the dynamic pulse TDDS, a pulse \( V_p \) is applied for the duration \( t_p \) between the charging and discharging biases [6].

Fig. 6 shows our proposed defect model, including two stable and two metastable states. The existence of these metastable states could be clearly demonstrated by extending the previously introduced time-dependent defect spectroscopy to a more general dynamic case, by employing AC stress [8]. Varying the AC stress frequency for 1kHz to 5MHz clearly reveals a frequency-dependence of the capture time of individual defects [7]. Such a frequency-dependence would not exist for a first order capture process and thus clearly confirms the existence of the intermediate metastable state 2’. Fig. 7 shows that charge emission can be drastically accelerated for some defects by applying a gate bias \( V_p \) into the depletion regime. Since the drain current is very small in depletion, emission events cannot be directly measured. Therefore the emission time constant is indirectly measured in this regime: After charging a given defect to an occupancy level of 100% by applying a sufficiently long stress pulse, a series of “recovery-pulse” with a given bias \( V_p \) and lengths \( t_p \) are applied, followed by a standard recovery trace testing the occupancy. \( \tau_e \) then is calculated from the decay rate \( \exp(-t_p/\tau_e) \), which describes the occupancy during the recovery pulse.

![Fig. 6. The four states of oxide defects extracted from TDDS experiments [7-10]. Each defect has two stable states, 1 and 2, and possible two metastable states 1’ and 2’. The metastable state 2’ seems to be always present, while the existence of metastable state 1’ decides on whether the trap behaves like a fixed positive or a switching trap [11].]
Fig. 7. An example for the wide range of extracted emission time constants for two defects using the “dynamic” TDDS technique. The range is extended to gate voltages far below threshold where no drain current can be measured. The emission time constant of defect B1 is bias-independent revealing B1 as a fixed positive trap. Defect B3 is a switching trap because of its bias-dependent emission time constant. The symbols are the data while the lines are from the NMP model [7].

In summary, while all investigated traps show a frequency-dependent capture time constant, suggesting them to be of the same microscopic origin, we find two different kinds of emission behavior, namely fixed positive and switching traps, as shown in Fig.3. The multi-state NMP model perfectly captures both cases.

IV. TRAP IMPACT ON FET PROPERTIES

The reduction of FET device dimensions to nanometer scales implies that literally only a handful of defects will be present in each device, while each defect will have a substantial impact on the device operation [12]. The impact of individual charged defects on transistor properties is now discussed. Clearly defined steps due to single-carrier discharge events are visible in Fig. 8a in the \( \Delta V_{th} \) relaxation curves in TDDS-like measurements (cf. Section III) performed on multiple deeply-scaled devices. Each device behaves differently, resulting in large time-dependent variability of the total \( \Delta V_{th} \). The \( \Delta V_{th} \) step heights appear exponentially distributed, with the cumulative distribution function (CDF)

\[ F_{\eta}(\Delta V_{th}) = 1 - \exp\left(-\frac{\Delta V_{th}}{\eta}\right), \]  

with an average step height \( \eta \) (Fig. 8b). Note that while the majority of charged traps results in small \( \Delta V_{th} \)’s, a small fraction of traps will significantly change \( V_{th} \) by up to tens of mV [12].

The average impact of a single defect \( \eta \) on the threshold voltage is a fundamental parameter determining the variability of deeply scaled technologies [13]. The time-dependent variance of the threshold voltage shifts of a population of devices (such as e.g. in Fig. 8a) can be shown to behave as

\[ \sigma_{\Delta V_{th}}^2(t) = 2\eta \left(\Delta V_{th}(t)\right), \]

where \( \Delta V_{th} \) is the average threshold voltage shift due to BTI [13, 14]. Because of the similarities between BTI and Random Telegraph Noise (RTN) [12], this parameter also reflects the average expected amplitude of the RTN signal in deeply-scaled devices [14].

A low value of \( \eta \) is generally desired. \( \eta \) has been argued to scale as

\[ \eta \approx \frac{t_{inv} N_{d}^{\alpha}}{A}, \]

where \( t_{inv} \) is the oxide thickness corresponding to capacitance in inversion, \( A \) the area of the device channel, and \( N_{d} \) the channel doping, with the exponent \( \alpha \) has been observed to be around 0.5 in simulations [13]. In order to test Eq. 4 experimentally, we collected an extensive dataset of NBTI relaxation transients devices with different areas. Figure 9a demonstrates \( \eta \) scales reciprocally with device gate area, in agreement with Eq. 4 [15]. Figure 9b then demonstrates that \( \eta \) decreases with oxide thickness and can be changed by back bias, which effectively modulates the number of charged dopants in the channel. Both of those observations are in line with Eq. 4 [16].
Over the whole measured range in Fig. 9a, \( \eta \) is observed ~2\( x \) higher than the “naively” expected impact \( \eta_0 \) given by the charge sheet approximation for a single charge at substrate/dielectric interface [14,15]

\[
\eta = \frac{q}{C_{ox}}.
\]  

(5)

Here, \( q \) is the elementary charge and \( C_{ox} \) the gate oxide capacitance (in Farads) in inversion. In pFET devices with SiGe substrates we have, however, observed that the relative value of \( \eta \) will be reduced for oxide defects closer to the gate (Fig. 10a) [15]. In addition to the low impact per charged defect \( \eta \), the superior NBTI robustness of SiGe-based pFET devices is also reflected in the significantly decreased number of active defects, documented in Fig. 10b [17,18].

In low doped channel devices, we have experimentally demonstrated that charged interface traps represent a new source of variability, as shown in Fig. 11a [14]. The value of \( \eta \) is then increased by further interface trap generation during operation, see Fig. 11b.

Precise circuit simulations will require models incorporating the impact of the small number of trapped charges on the entire FET current characteristics [12]. We have demonstrated the measurements of the impact of a single charged gate oxide defect on the entire \( I_D-V_G \) characteristic—in fact the ultimate reliability measurement (Fig. 12) [15]. The learning from measurements and full scale 3D ‘atomicistic’ simulations (Fig. 12c) can be captured in reliability-aware FET compact models [19].

Finally, in order to understand and to model the impact of gate oxide defects on the FET gate current \( I_G \), correlated drain and gate current RTN measurements were performed [20]. We have observed that the same defect controls both processes and constructed a corresponding defect state diagram involving a metastable state, akin to Fig. 6 [21]. Based on this picture we have proposed a qualitative model describing the correlated \( \Delta I_G \) and \( \Delta I_D \) distributions (Fig. 13) [21].

V. CONCLUSIONS

Reduction and mitigation of BTI in future CMOS technologies requires thorough understanding of its underlying causes—gate oxide interface and bulk defects. We have reviewed several recent findings of the properties of these defects, obtained with both electrical and physical characterization techniques.
Fig. 11. (a) A clear correlation is found between single-carrier discharge $\Delta V_{th}$ and $D_{it}$ in FinFET devices. Interface traps enhance channel conduction percolation between source and drain in the same way dopants do (inset). (b) Generation of $D_{it}$ after electrical stress is reflected in increase of single-carrier discharge $\Delta V_{th}$.

Fig. 12. The “ultimate” reliability measurement: (a) $I_D$-$V_D$ curves measured at $V_D = -0.1$ V on a $L\times W = 35 \times 90$ mm$^2$ pFET with a single oxide defect uncharged and charged. The corresponding $\Delta V_{th}(V_G)$ is defined as a simple “geometrical” horizontal distance between the two curves at a given $V_G$. (b) Multiple measurements of $\Delta V_{th}(V_G)$ are perfectly reproducible. (c) 3D atomistic simulations of a nanoscaled device showing an oxide defect located exactly above the critical confinement spot of the current percolation path (demarcated by the vertical arrow) can effectively suppress the channel current when charged. The simulations show the defect impact on the device characteristics depends on the radial distance $r$ from the critical spot. (d) Different $\Delta V_{th}(V_G)$ characteristics are expected as a function of $r$.

ACKNOWLEDGMENTS

The hereby presented work represents a concerted effort performed within the Work Package 4 (WP4, Sample fabrication and characterization) of the collaborative project MORDRED, contract No. 261868, supported by the European Commission under the 7th Framework Programme. The authors’ names have been ordered as follows: WP4 leader, WP4 task leaders, and WP4 contributing researchers in alphabetical order. Discussions with A. Asenov, S. M. Amoroso, L. Gerrer, and F. Adamu-Lema (WP3) are gratefully acknowledged.

REFERENCES


