FP7-NMP-2010-SMALL-4

COLLABORATIVE PROJECT

Final report

Modelling of the reliability and degradation of next generation nanoelectronic devices

MORDRED

Final report

20/05/2015

Work programme: **NMP.2010.2.5-1 Modelling of degradation and reliability of crystalline materials**

Name of the coordinating person: **Prof. Adam S. Foster, Aalto University, Finland**

The objective of the MORDRED project is to develop multi-scale modelling technology supported by comprehensive experimental characterization techniques to study the degradation and reliability of next generation Complementary-Metal-Oxide-Semiconductor (CMOS) transistors and to construct meso-scale models to account for defect generation and impact on CMOS transistor and circuit performance and yield.

All deliverables and milestones were achieved, and the software infrastructure was successfully used to go from quantum mechanical calculations of defect states through trapping and transport models, up to compact modelling of devices. This infrastructure (being developed here http://physics.aalto.fi/~tjhynnin/mordred/software/) is available for future use, and is already being applied to materials and devices beyond the scope of the MORDRED project.

The following summarizes the highlights and key objectives of each of the workpackages within the project.

1. WP1 Interface structure and defect properties

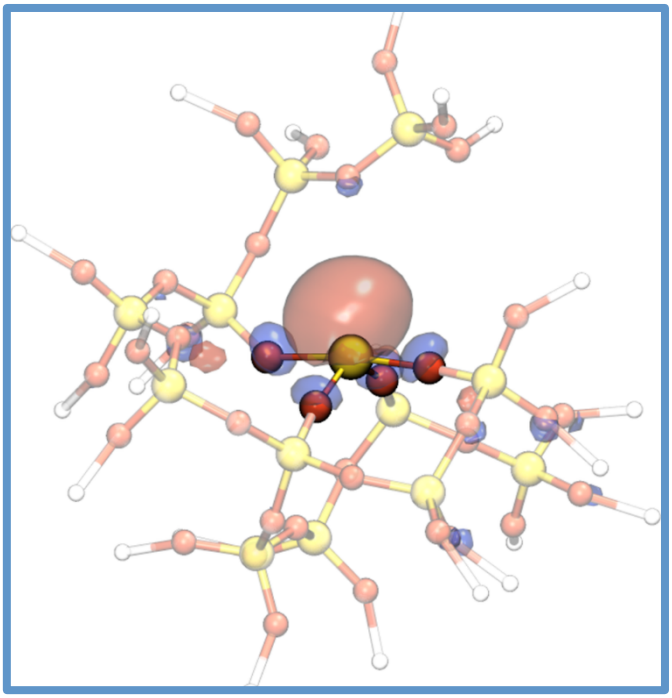


Figure 1: An intrinsic electron trapping site in amorphous silicon dioxide.

ELECTRON TRAPPING IN SILICA

The mechanisms of electron and hole trapping in SiO_2 and the nature of trapping sites are important for our understanding of a wide range of physical phenomena. Of particular importance to the MORDRED project, electron trapping is known to have a dramatic effect on the performance and reliability of electronic devices employing SiO_2 as gate insulator and charge trap flash memory devices. Using classical and ab initio calculations we demonstrate that extra electrons can trap spontaneously on pre-existing structural precursors in amorphous SiO_2. The precursors for electron trapping in amorphous SiO_2 comprise wide (132°) O-Si-O angles and elongated Si-O bonds at the tails of corresponding distributions. Optical absorption spectra for the intrinsic electron trap have been calculated and they show a strong absorption peak at 3.9 eV, which agrees well with a previously unassigned experimental absorption peak at 3.9 eV.

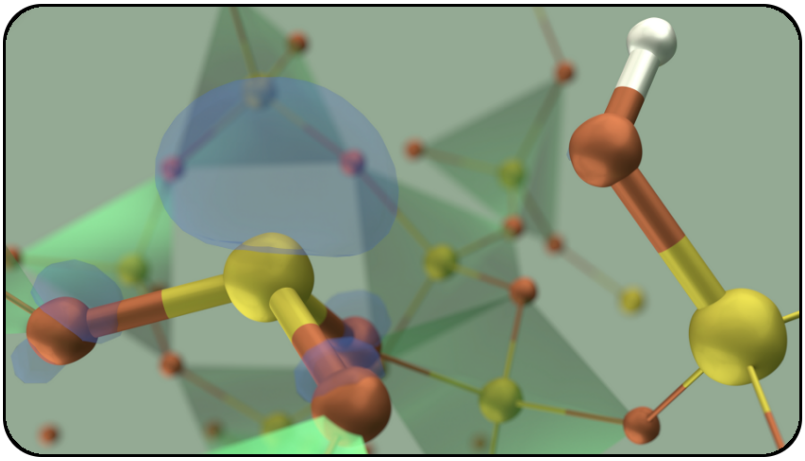


Figure 2: The hydrogen induced defect in silicon dioxide known as the hydroxyl E' centre. In the background, the SiO4 tetrahedra are intact. However, in the foreground, an H atom has broken a SiO4 tetrahedron, breaking the continuity of the network and making a defective site.

For decades, amorphous silicon dioxide has been at the core of the development of microelectronic technologies. To produce reliable devices, hydrogen is widely used in their fabrication processing, as hydrogen is widely perceived to heal defective sites in silicon dioxide. As part of the MORDRED project, we used state-of-the-art atomistic modelling techniques to show that atomic hydrogen can in fact interact rather strongly with silicon dioxide networks to generate additional defective sites, in stark contrast to its well known healing properties. These sites can trap charges and can take part in electronic device reliability issues. Hence, atomic hydrogen is not always a benign agent when processing SiO_2, as commonly perceived in the literature.

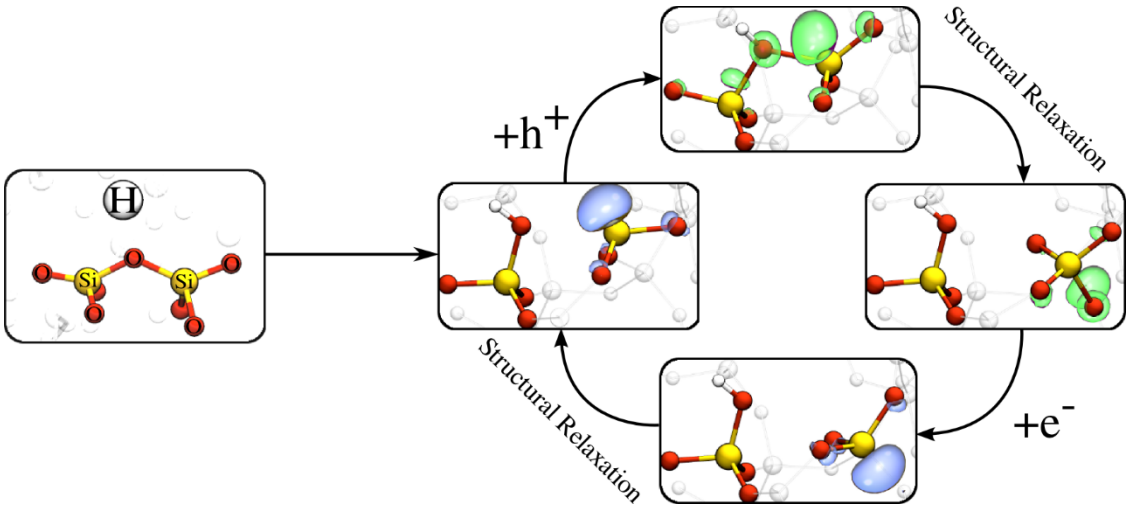


Figure 3: In collaboration with researchers in the MORDRED project, we have shown that atomic hydrogen can interact with amorphous silicon dioxide to produce a defective site known as the hydroxyl E' centre. Further studies of the hydroxyl E' centre demonstrate that it can participate in charge trapping reactions and can be cycled back into its original configuration. The energetics of this cycle is consistent with the electronic device reliability issue known as the negative bias temperature instability (NBTI).

SILICA/HAFNIA INTERFACES

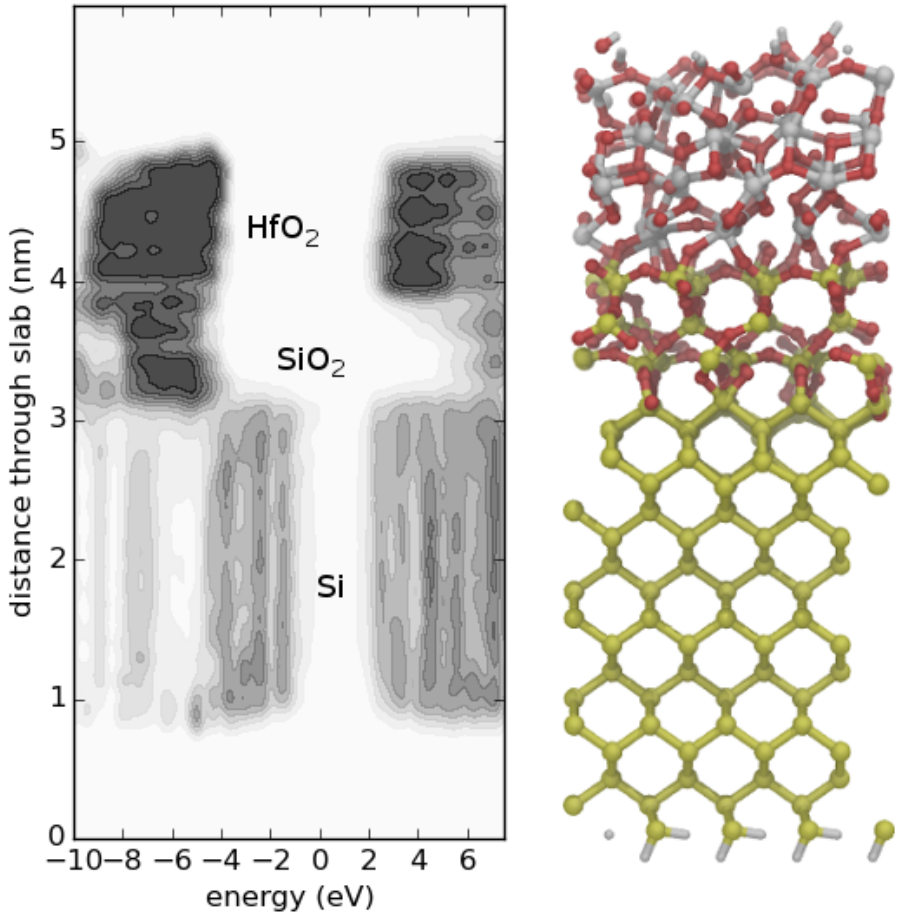


Figure 4: Model of a high-k stack – the layers of silicon (yellow), thin layer of SiO_2 (yellow and red), and HfO_2 (grey and red) can be seen in the atomistic model on the left.

Working closely with lead developers of the CP2k in Zurich, the range of systems that can be tackled using so-called hybrid density functional methods has extended to include some that come close to atom-atom models of transistor stacks – for instance the Si-SiO_2-HfO_2 interface shown with ~650 atoms. The lateral dimensions are 1.5x1.5 nm^2 and in the vertical direction through the slab the size of the SiO2 interfacial layer is the real size of the latest devices. The use of hybrid density functional methods is important because they give realistic band-gaps and band-alignments across the different materials – this allows us to model electron trapping at new types of defects located at the interfaces of the materials.

This efficiency of the calculations allows us to calculate defect energies at different location throughout the stack-model. For instance the formation energies of oxygen vacancy defects were calculated for oxygen atoms located between the Si-SiO_2 interface and the SiO_2-HfO_2 interface. By calculating large numbers of defect energies we can estimate statistical properties of the defects, including the distribution formation energies and the location of electron trapping levels. This is vital because many of the materials used in microelectronics applications (SiO_2 is the prime example) are amorphous – this means that each site in the material is unique. The importance of these distributions is emphasized in the collaboration with the Vienna group on identifying defects responsible for NBTI.

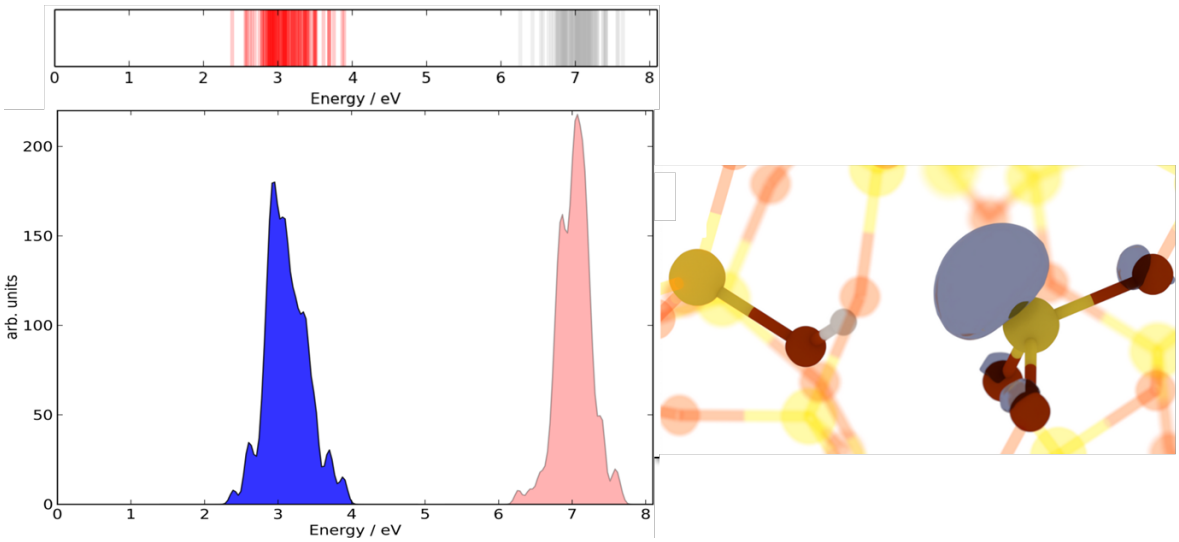


Figure 5: Distribution of donor and acceptor single electron levels in the hydroxyl E' centre of amorphous SiO_2. The energy is measured relative to the top of the SiO_2 valence band. This class defect has levels 2-4 eV above the SiO2 valence band that overlap with the energy position of the Si valence band, making it possible for some of these defects to trap holes.

In order to provide a faster route to optimizing of silica/hafnia stacks, we also compared our CP2k results to those produced using a much faster semi-empirical potential model. Two interfaces between SiO_2 (alpha-quartz) and HfO_2 (cubic hafnia) have been created from scratch, respecting the stoichiometry. After an initial relaxation within the VASP DFT code to obtain charges, we used the code LAMMPS to perform the MD simulations. We have used the official LAMMPS distribution of COMB and results can be seen by comparison of Figs. 6 and 7. In both cases the silica and hafnia parts are more spaced, and there is the presence of oxygen atoms that have diffused to the interface. Moreover there are no silicon atoms 5-fold coordinated in silica, meaning that Si maintains the right coordination. This behavior highlights characteristics of the COMB potential, that includes spurious energy minima, i.e., non-physical metastable states close to the actual minimum energy configuration. This is a common problem of phenomenological potentials constructed from pair and many-body interactions and it is not surprising to find this behavior also in COMB. Nonetheless, it means that simple optimization of structures carried out with the COMB potential will not necessarily find a physically valid structure.

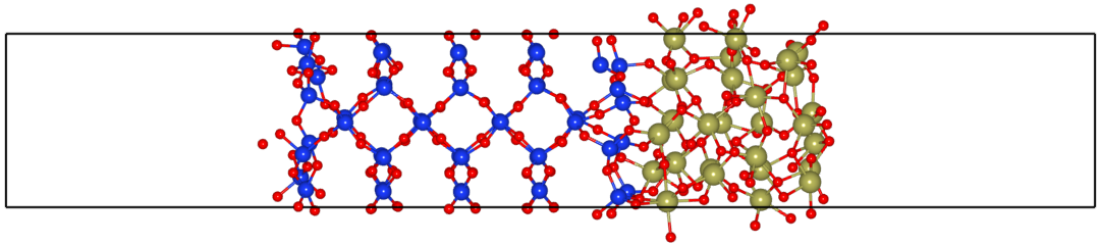


Figure 6: Silica/hafnia interface, relaxed with DFT.

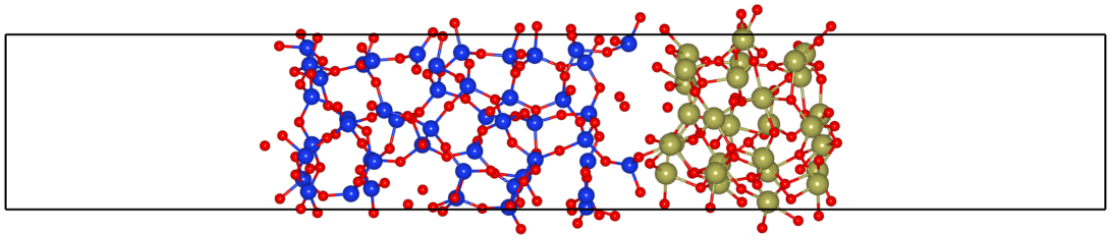


Figure 7: Silica/hafnia interface, annealed and cooled back with COMB.

HOLE-INJECTION LAYERS FOR TMDC-BASED ELECTRONIC DEVICES

We have used *ab-initio* simulations to investigate the role of differently functionalized graphene compounds in getting hole-injection transition metal dichalcogenides (TMDCs) – based devices. We show that graphene oxide (GO) and graphene fluoride (C_xF) inserted between MoS_2 and the contacts allow the field-effect transistor (FET) to be p-type. Given that the actual fabrication of p-FETs presents intrinsic challenges, our approach allows to have n- and p-type FETs on the same device, a fundamental requirement to produce logic circuits. One proposed device architecture to obtain a p-FET is shown in Fig. 8f, where MoS_2 is used as the active channel material, and GO or C_xF act as hole-injection layers between the contacts and the channel.

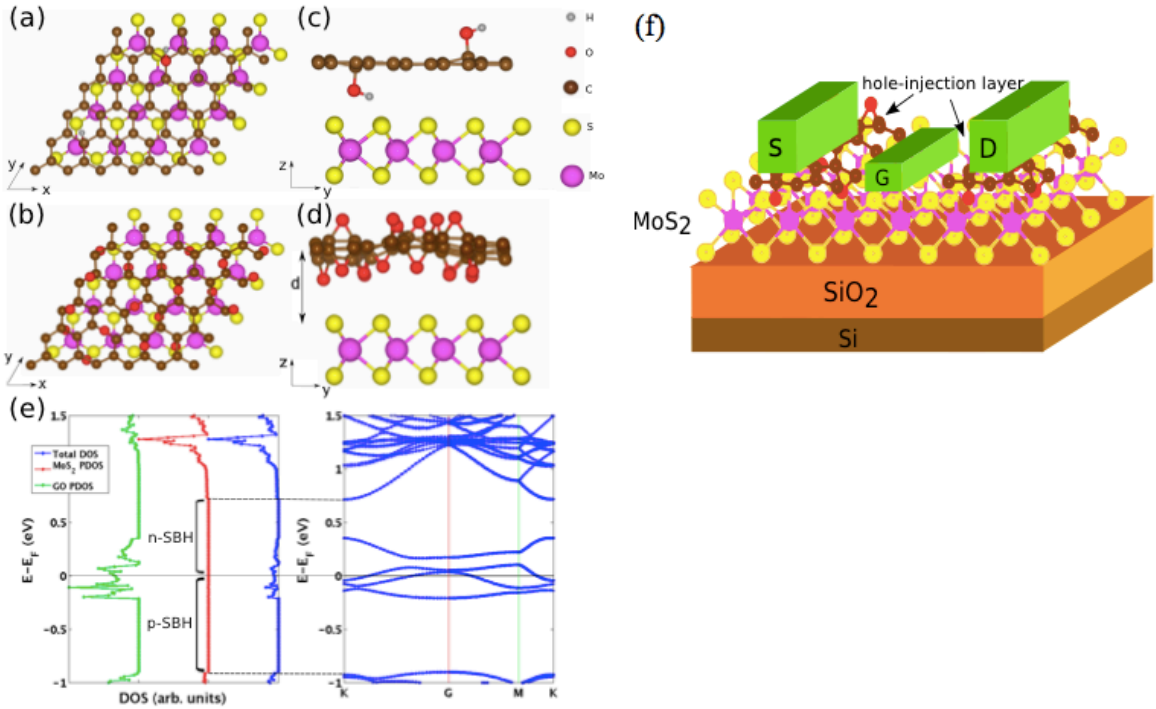


Figure 8: Graphene Oxide (GO) in contact with MoS_2 - example of relaxed supercells: (a,b) top view and (c,d) lateral view. In (a) and (c), the GO layer has a low oxygen concentration (4%) and is functionalized with hydroxyl groups alone. In (b) and (d), the GO layer is epoxy functionalized, with a high oxygen concentration (24%). Oxygen concentrations are given in atomic percent. (e) Example DOS plot showing the definition of n-type and p-type SBH, with the corresponding band structure. (f) Possible device architecture using GO or C_xF as hole injection layers between the MoS_2 monolayer and the contacts, to obtain a p-FET.

For each hybrid structure, the DOS and the band structure have been computed to obtain a value for the MoS_2 p-type SBH. We define the energy difference between the Fermi level and the minimum of the conduction band as the n-type SBH, and similarly the p-type SBH is given by the energy difference between the valence band maximum and the Fermi energy (see Figure 9e), consistent with previous DFT simulations of MoS_2/graphene contacts and MoS_2/metal contacts.

SOFTWARE INFRASTRUCTURE

Our strategy for achieving the software goals of the MORDRED project is to construct a programming interface for the simulation tool, and Python has become a popular interface language as it has both powerful scripting capability as well as advanced features of object oriented programming. As an interpreted language it is slow to execute, but it is possible to implement the computationally intensive parts in more efficient compiled languages such as C or Fortran in order to gain back computational speed.

In this work, we present a Python library for evaluations of classical atomistic force fields. The library is designed to work with the Atomistic Simulation Environment (ASE) [1], an established Python framework for atomistic and electron structure calculations. ASE follows the paradigm of object oriented programming to wrap simulation entities such as atomic structures or dynamics algorithms in Python objects, which are easy to manipulate by both human users and script.

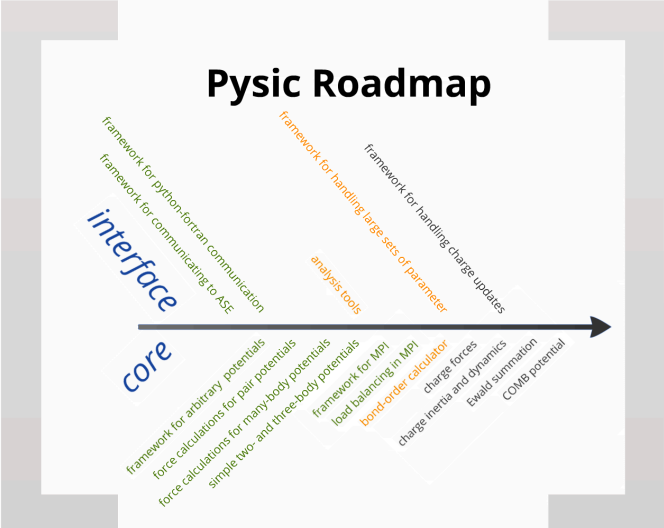


Figure 9: Implemented features in the Pysic infrastructure.

But even though ASE provides tools for building and evolving atomic structures, it relies on external programs to determine the interactions between the atoms. Interfaces exist to several such codes, called *calculators* in ASE, at classical (e.g., LAMMPS [2]) and density functional theory level (e.g., GPAW [3]). Our library, *Pysic*, is also a calculator for ASE at the classical level, but instead of providing just a Python interface to an external calculator, Pysic reduces atomistic potentials to Python objects allowing the user to build the interaction model from components. Pysic is not concerned with constructing an atomic structure or its dynamic evolution, as these are already handled in ASE. Instead Pysic calculates the energies and forces of a given structure, i.e., it defines the potential energy surface of the system.

The programmable interface allows one to control and tune the interactions through Python scripting. In particular, one can build an algorithm on top of the energy calculator in order to automate the optimization of parameters against any desired observables. This is not limited to parameters either, since the modularity of the program also enables changing the functional form describing the potential at script level. Also the analysis of force field models benefits from the possibility of breaking down the potential to components. It is possible to pick out the contribution of the different components one at a time, or in groups, and even automatically monitor for situations where a particular part of a potential is especially weak or strong.

2. WP2 Defect Dynamics

Following the International Technology Roadmap for Semiconductors (ITRS), modern Metal- Oxide-Semiconductor (MOS) transistors have been shrunk to scales where even a single charge trapped in the dielectric noticeably alters the transistor characteristics. These charge trapping events can severely impact the device functionality or may in the worst case even lead to device failure. They are often reversible but can also lead to a permanent degradation. The fluctuations due to these events are linked to serious reliability issues, such as Bias Temperature Instability (BTI), Random Telegraph Noise (RTN) in the drain current, and Stress-Induced Leakage Currents (SILC) amongst others, and hence have aroused a great deal of industrial interest. These reliability issues have also posed major scientific challenges. So far, no clear picture of the underlying microscopic processes could be established. However, it is commonly accepted that these phenomena originate from charge trapping into and out of defects located in the dielectric. Therefore, this work package of the MORDRED project has focused on the trapping dynamics as well as the chemical kinetics of those defects.

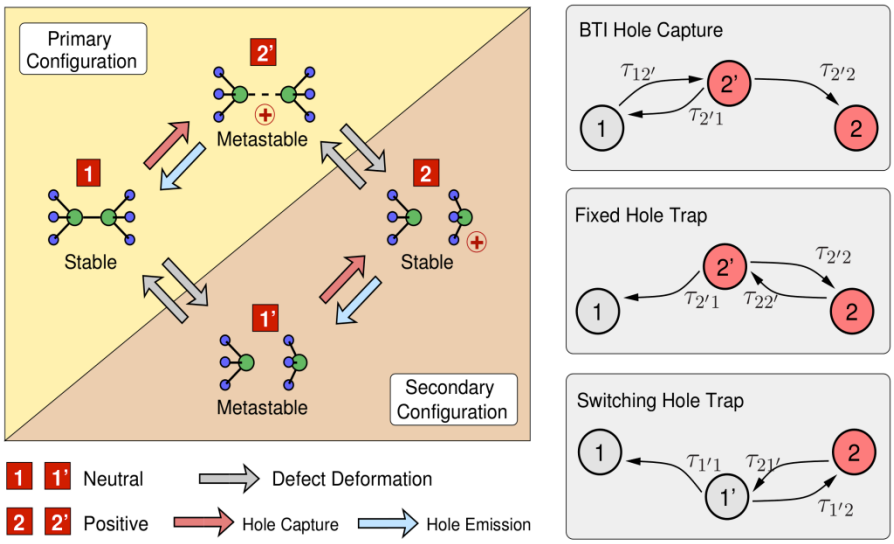
First advances have been made regarding the BTI phenomenon, which is the response of oxide defects in the dielectric if large negative/positive voltages (NBTI/PBTI) are applied to the gate terminal. Its effects were found to be more pronounced at higher temperatures. This phenomenon has traditionally been described within the framework of the Reaction-Diffusion (RD) model, in which the kinetics are governed by hydrogen diffusion from and towards the interface. However, studies based on Time-Dependent Defect Spectroscopy (TDDS) have demonstrated that the recoverable component of NBTI must be ascribed to a first-order reaction, which is inconsistent with any variant of the RD model. This measurement technique has also revealed a series of experimental features:

- The RTN/NBTI hole capture and emission times of individual defects are only weakly correlated.
- The capture times exhibit a strong gate bias dependence that levels off towards higher gate biases.
- Furthermore, they also feature a pronounced frequency dependence.
- The individual defects show two distinct kinds of detrapping behavior: Depending on whether their emission times are insensitive to the gate bias or not, they are classified as fixed positive or switching oxide traps, respectively.
- Some of the defects producing noise and BTI sporadically disappear and re-appear again, a phenomenon which is known as temporary or anomalous RTN and can probably be linked to the “volatility” of defect investigated in this WP.

The above findings were essential for the development of the multi-state non-radiative multiphonon (NMP) model as they served as hard criteria for model verification. The first prototype of this model focussed on NBTI and was capable of reproducing all its experimental features mentioned above. The defect was assumed to have a stable neutral (state 1) and positive charge state (state 2) as illustrated in Fig. 1. Most importantly, it also features a metastable configuration (1',2') for each charge state. The defect dynamics between those four states show a quite complex behavior: The hole capture process is dominated by

the transition from the neutral state 1 to the positive state 2 over the metastable state 2'. By contrast, the alternative pathway over the metastable state 1' is usually suppressed for typical BTI conditions. The corresponding emission process proceeds in the opposite direction but can take the pathway over state 1' as well as 2'. These complex dynamics must be correctly described using homogeneous continuous-time Markov chain theory. Then the effective capture and emission time constants follow from the time constants of the elementary reactions using first-passage times (see Fig. 1).

Fig. 1: Left: The state diagram of the multi-state NMP model. The actual charge capture or emission does only occur between the neutral (1/1') and the positive (2/2') charge state while transitions between a stable (1/2) and a metastable (1'/2') configuration involve a defect deformation. Right: Transition pathways for hole capture and emission.



In the multi-state NMP model the actual charge capture and emission rely on a nonradiative multiphonon (NMP) process, which is elucidated in the configuration coordinate diagram of Fig. 2. In the framework of the MORDRED project, these NMP transitions have been studied at various levels of sophistication. For instance, a quantum mechanical description has been developed, which is regarded as the most general formulation and is based on the so-called lineshape functions (LSF). In this formulation, the adiabatic potential energy surfaces are modelled as harmonic quantum oscillators, which are usually assumed in solid state theory. The oscillator wavefunctions of both charge states overlap and therefore allow for nuclear tunneling between them. This transition corresponds to the actual charge capture or emission event (1<->2' or 1'<->2) in the multi-state NMP model. The probability of nuclear tunneling is given by the LSF, which reflects the gate bias and temperature behavior of this process. It has been demonstrated (see the MORDRED deliverables D2.1.1 and D2.2.1) that the LSF can be efficiently calculated from DFT simulations using the single-mode coupling approximation. This general formulation yields accurate results and can become important for weak-electron phonon coupling in trap-assisted tunneling (TAT) through the dielectric for instance. As demonstrated in the course of this project, in most cases the classical high-temperature limit of the NMP theory provides a good approximation at usual device operation temperatures. This is due to the fact that the NMP processes are dominated by thermal activation over an NMP barrier rather than nuclear tunneling through this barrier.

Therefore, in the classical formulation the transition probability is governed by an Arrhenius-like barrier factor, which is more efficient for the application in TCAD tools and has been employed primarily.

For simulations, a theoretical framework has been developed to integrate the multi-state NMP model into a Schrödinger-Poisson solver or a Non-equilibrium Green's function (NEGF) code. Both simulators were used to calculate the electrostatics of the semiconductor device. This information is needed to evaluate the NMP rate expressions of the NMP transitions, which enter the master equations given by the state diagram of the multi-state NMP model. From these equations the effective capture and emission times can be computed based on the concept of first-passage times. In a probabilistic treatment these equations can be solved using the Gillespie algorithm so that also the stochastic nature of the reliability issues can be investigated. For instance, this becomes important in the case of small-area devices where the impact of single charge capture or emission events is of interest. Furthermore, the master equations can also be solved for the average of a large ensemble of defect as it is necessary for large-area devices. With this general formulation of the defect dynamics, the foundation has been laid to investigate BTI in particular but also other reliability issues in greater detail.

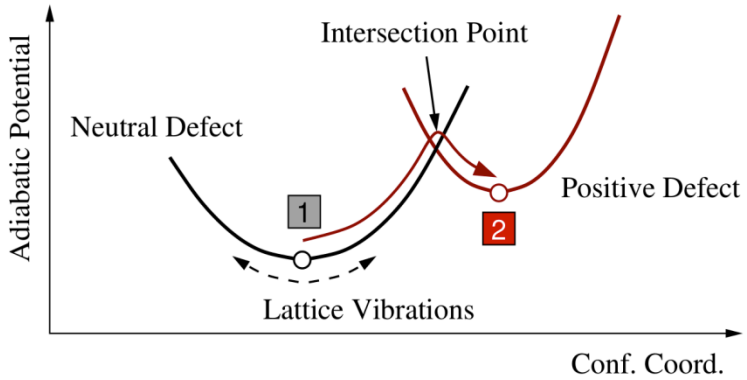


Fig. 2: Configuration coordinate diagram for an NMP process. In this kind of diagram, the defect energy is plotted as a function of the configuration coordinate, which describes the complex lattice vibrations using one quantity only. The resulting curves are termed adiabatic potential energy surfaces and exist for the neutral and the positive charge state of the defect. The arrow indicates an NMP transition for a hole which is emitted from the substrate into the defect. During this process, the defect must be thermally excited from the minimum of the left curve (neutral defect) up to the intersection point (IP) with the right curve (positive defect). At this point, the actual hole capture takes place. Subsequently, the defect can relax to the minimum energy of the right curve.

ELECTRON TRAPS

NBTI in pMOSFETs has been intensively studied over the last years while PBTI in nMOSFETs has received much less attention. As such, intensive efforts were made to investigate PBTI in this project. Due to smaller defect concentrations or better energetical alignment of the defect band with the substrate, the degradation in nMOSFETs is found to be about one order of magnitude lower than NBTI. Therefore, devices with a relatively large gate area were chosen in order to increase the probability for finding detectable defects in TDDS. As the threshold voltage shift due to a single electron capture event scales inversely

with the gate area, the experimental setup is required to accurately resolve the drain current steps in TDDS analysis. In this TDDS study, two defects were detected, whose electron capture and emission time are plotted in Fig. 3. Their capture and emission times were found to be temperature dependent with activation energies in the range of 0.4 – 0.9 eV, showing a strong exponential gate bias dependence. Also, the distinct behavior of fixed (defect #1) and the switching (defect #4) oxide traps were rediscovered in the nMOSFET. Due to the similarities to hole traps in NBTI, both defects were simulated using the four-state NMP model (see Fig. 3). It was found that both electron traps can be described by the multi-state NMP model.

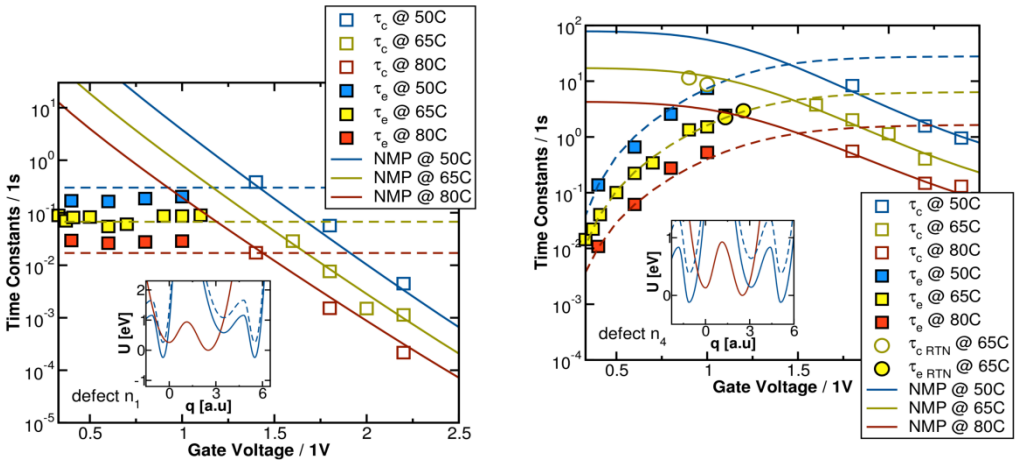


Fig. 3: Left: The gate voltage dependence of the capture and the emission time for defect #1 detected in an nMOSFET. The emission time appears to be independent of the applied gate bias (fixed charge trap) whereas the capture time shows a significant bias dependence. Good agreement between data (symbols) and model (lines) is obtained. The inset shows the configuration coordinate diagram of the NMP model. Right: Contrary to defect #1, the emission time of defect #4 (also found on an nMOSFET) shows a pronounced gate bias dependence (switching trap). Just like the fixed charge trap behavior, the switching trap behavior can be well explained by the four-state NMP model (lines).

BTI AND RTN

Charge trapping is believed to be the cause of drain current RTN as well as BTI and as such, it seems likely that both effects can be ascribed to the same defects. In order to verify this hypothesis, defects detected in TDDS and RTN measurements were thoroughly analyzed and compared. For this purpose, the defects of a small-area device (W × L = 150 nm × 100 nm) were characterized using the TDDS technique. Their capture and emission times were extracted for a wide voltage range starting from accumulation to deep inversion at 125°C and 175°C (see Fig. 4). From these measurements, the gate voltage was estimated at which this defect is expected to produce an RTN signal. This occurs in the region where the capture and emission times are within approximately one order of magnitude. Based on the predicted RTN time constants, defects studied with TDDS could also be found in the traces of the subsequent RTN experiments. This observation suggests that BTI and RTN originate from the same defects.

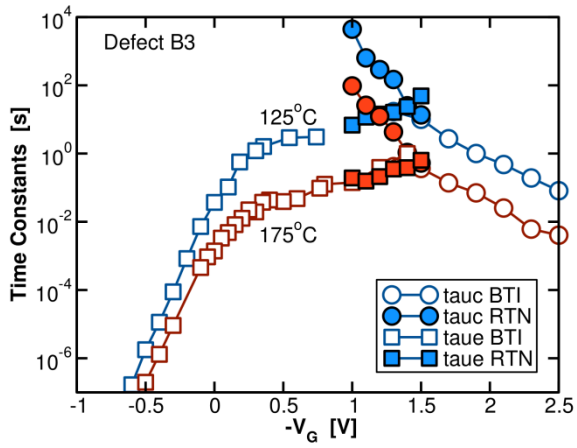


Fig. 4: The bias and temperature dependent capture and emission times determined by TDDS (open symbols) and RTN analysis (filled symbols) agree very well.

In order to verify the statistical relevance of these defects, the step heights and emission times were determined for the defects in 100 nano-scale devices in coupled BTI and RTN measurements. Their corresponding distributions compare surprisingly well, corroborating the common physical origin of BTI and drain current RTN. In view of these findings, the impact of RTN on reliability has been reassessed. In particular, it was found that the superposition of several RTN events can result in large threshold voltages shifts and therefore can cause a serious reliability problem. Supposing that a certain combination of charged defects exceeds the threshold voltage of a typical lifetime criterion, their charging time corresponds to the time of the first device failure. It was shown that this effect can even dominate over conventional BTI in the low-voltage range. However compared to the BTI conditions, the device leaves the state of "device failure" quickly.

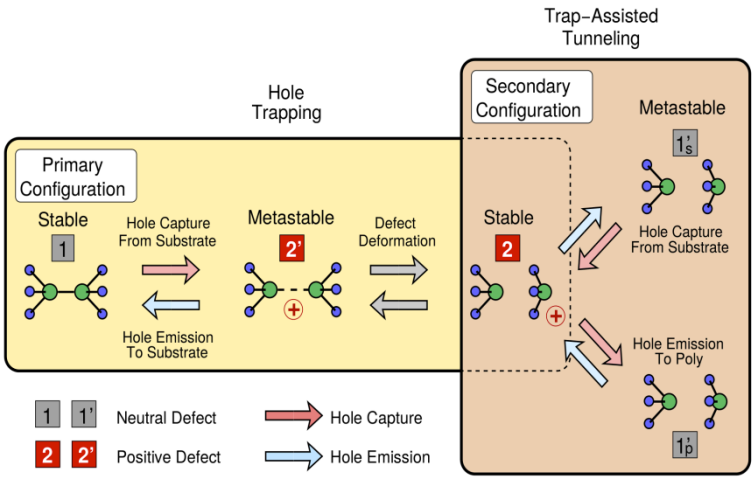
SILC

During the last decade drain current RTN has emerged as a serious reliability issue for MOS devices. Recently it was found that the fluctuations in the drain and the gate current can be correlated. In WP4 of the MORDRED project, drain current RTN has been traced back to the capture and emission of substrate charge carriers in the gate oxide. Therefore, it has been argued in literature that a captured charge locally repels the inversion layer, thereby decreasing the direct tunneling current through the gate oxide. This effect has been investigated using NEGF simulations for a series of random dopant configurations. In the worst case, they predict gate current fluctuations of less than 1% while measurements yield values around 8%. As a consequence, this electrostatic effect must be ruled out as an explanation for the gate current fluctuations.

Another explanation is based on the multi-state NMP model and relies on trap-assisted tunneling (TAT) through the defect producing the drain current noise. As illustrated in Fig. 5, the TAT current is described by two consecutive NMP transitions $1'_s \leftrightarrow 2 \leftrightarrow 1'_p$. It is emphasized that this current only flows upon hole capture, that is, if the defect is positive (state 2). However, it stops upon hole emission when the defect returns back to its neutral state 1. This switching behavior can explain the correlation between the drain and gate current noise seen in experiments. It has been demonstrated that the multi-state NMP model yields the correct field and temperature dependence of the gate current fluctuations while it still fits the RTN/BTI capture and emission time constants. Quite surprisingly, the gate current fluctuations are actually temperature insensitive. At a first glance, this fact may

contradict the above explanation since NMP transitions involve barriers, which must be thermally overcome.

Fig. 5: Refined state diagram of the multi-state NMP model explained using the oxygen vacancy as an example. Hole capture and emission still occur via transitions between the states 1 and 2 over the metastable state 2′. However, the state 1′ has to be split in order to differentiate whether the hole is located in the substrate (s) or the poly-gate (p). With this modification in the state diagram, one can define a TAT process in which the hole is transferred from the substrate into the poly-gate and can therefore contribute to the gate current.



However, this does not hold for the special case of NMP transitions in the "weak electron-phonon coupling". Such transitions were found to be the dominating contribution to the TAT current, thereby explaining the temperature independent gate currents.

The above investigations were focused on single defects, whose impact on the gate leakage current must be studied in terms of reliability. It was found that defects may start carrying a TAT current upon application of a gate bias - a behavior which is commonly known as stress-induced leakage current (SILC).

TWO DIMENSIONAL MATERIALS

Recently many attempts have been undertaken to replace the substrate material by a novel class of 2D materials, with graphene having received a lot of attention. To assess the true potential of these materials, they must also be evaluated with respect to their reliability behavior. Therefore it was tested whether they show the same experimental features as conventional silicon technologies and can consequently be described by the same model, i.e. the multi-state NMP model.

The experimental studies were based on large-area Graphene FETs (GFETs) with W×L = 4–80×1–4 µm, where the graphene channel was sandwiched between an aluminium oxide top gate insulator and a silicon dioxide back gate insulator. It was realized that the threshold voltage is an ill-defined quantity in GFETs and therefore the degradation behavior due to BTI was investigated in terms of Dirac point shifts. This is because the BTI stress does not only result in a horizontal shift of the Dirac point but also leads to a modulation of the drain current. In our BTI measurements, GFETs behave similarly to conventional silicon technologies (not shown here): The GFET devices degrade with a shift of the Dirac point, an effect that become more pronounced with an increasing stress time. During relaxation, the

Dirac point also returns back towards its initial value, with the shift being accelerated at higher temperatures. Interestingly, the degradation was found to have a fast trap component, which appears as a hysteresis in repeatedly recorded transfer characteristics. This component only contributes to the short-term degradation but already recovers within about 100s. By contrast, the long-term degradation follows the universal relaxation relation, with the fitting parameters being in excellent agreement with those from silicon technologies. This fact suggests that there is a similarity between the degradation processes occurring in both technologies. Interestingly, a permanent component seems to be absent in the graphene technology. These measurements were further used to calculate capture/emission time plots, which show the correlations between the capture and emission times. Like for silicon technologies, the obtained Capture/Emission Time (CET) maps can be described by bivariate Gaussian distributions. However, the contribution due to the permanent component was found to be missing again. Further investigations were performed regarding Hot-Carrier Degradatin (HCD) in GFETs. Interestingly, the recovery curves after HCD stress show a pronounced overrelaxation. This observation can only be explained by the involvement of electron as well as hole traps. Both kinds of trap are created during stress but the hole traps relax more slowly than their electron counterparts, resulting in a "permanent" component of a different sign.

In conclusion, BTI on GFETs seems to be dominated by the same kind of degradation process already found for conventional silicon technologies. As such, the theory of the multi-state NMP model should be transferable to graphene technologies. By contrast, an HCD model requires an extension that accounts for charge trapping of both electrons and holes during stress.

VOLATILITY

In this work package, experimental efforts have been undertaken to study also the long-time behavior of oxide defects on small-area devices. The most intriguing observation was that some defects which contribute to BTI, occasionally disappear and reappear. This phenomenon, which we termed "defect volatility", can affect device reliability and has therefore been studied in detail. As this volatility occurs at large time scales, the number of stress/recovery cycles and consequently the duration of experiments has been significantly increased. Sometimes an optional bake step has been introduced so that the initial state of the device prior degradation could be restored. During these long-time experiments, the following observations have been made:

• The volatility has been found not to be a rare event. Rather, it occurs quite frequently, i.e. for the majority of all defects in the device under test.

• The defects can disappear into a neutral and a positive charge state, thereby becoming temporarily or permanently inactive.

• The volatility occurs on widely distributed time scales, ranging from hours up to weeks.

• The extracted de-activation and re-activation times have been found to follow an exponential distribution, consistent with a reaction-limited process.

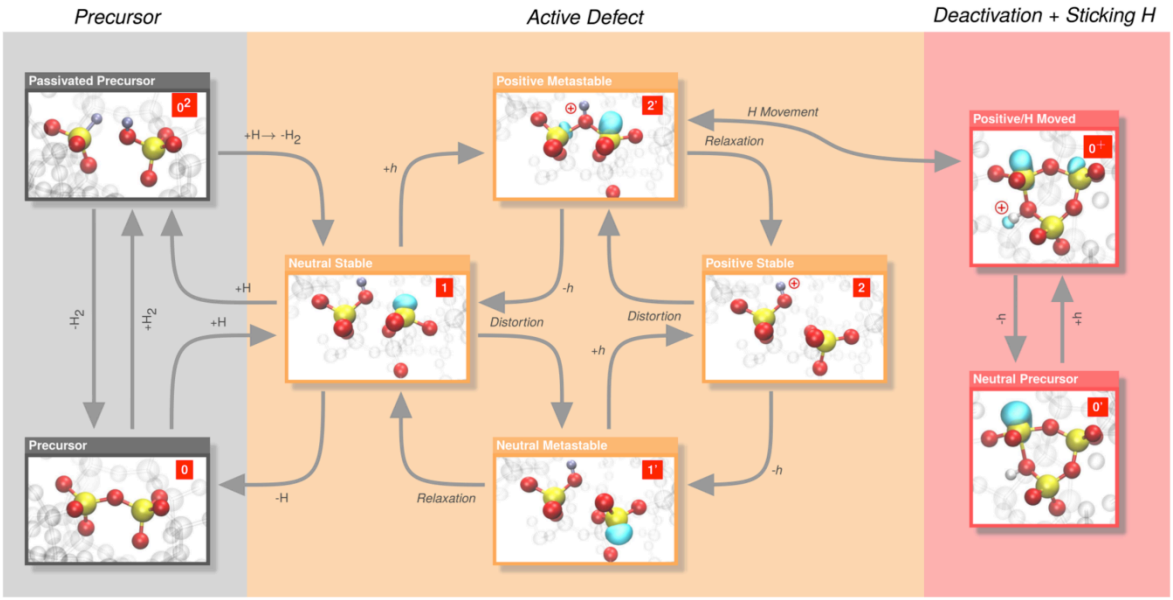
• Some defects may be re-actived by short bias pulses towards accumulation or high-temperature bake steps.

• Measurements on devices with different hydrogen content suggest that the volatility is microscopically somehow linked to hydrogen.

• No indication could be found that the volatile defects behave differently to the others. Therefore, their volatility must be an extension to their current description based on the multi-state NMP model.

• Some defects may undergo a transformation, particularly when subjected to a baking step.

The findings from the above measurements and theoretical investigations require an extension of the multi-state NMP model, whose state diagram is depicted in Fig. 6. Still, the heart of the extended model is the bistable defect with its four states (1, 1', 2, and 2'), which describes charge capture and emission of the recoverable component of BTI when the defect is active. However, the long-time experiments revealed that these defects can become inactive. For this to happen, the defect may undergo a reaction with a hydrogen atom and remains neutral afterwards. In the state diagram, this reaction is represented by a transition from state 1 into one of the states 0 and 0^2. According to the DFT studies performed in WP 1, the resulting defect structure may interact with molecular hydrogen and thus can switch between the states 0 and 0^2. Some defects were also found to become inactive in a positive charge state. This occurs most likely from state 2' via a defect transformation. In the case of the hydroxyl E' center, the positively charged hydrogen atom moves from its current site (2') to a neighboring site (0^+) where it is attached to an oxygen again. The extended model described above accounts for the volatility of the defect in both the neutral and the positive charge state. The latter presumably corresponds to the permanent component of BTI so that



the extended variant of this model becomes a comprehensive description of BTI.

Fig. 6: State diagram of the extended model. The current state of research suggests the hydroxyl E' center as the most probable defect candidate, which is investigated in WP2 in more detail.

HOT-CARRIER DEGRADATION

Similar to BTI, HCD causes a substantial damage of the dielectric layer and/or its interface to the substrate and thereby severely affects the device functionality. However, it is triggered by

the application of high source-drain voltages, which accelerates charge carriers within the channel. These high energetic charge carriers can collide with interfacial SiH bonds and excite them to their antibonding state (AB), followed by bond breakage. The remaining silicon atom has an electrically active dangling bond, which can capture charge carriers and thus deteriorates the conductivity of the channel. The above scenario is referred to as the single-particle (SP) process. However, this process requires highly energetic charge carriers and therefore competes with a process referred to as multiple vibrational excitations (MPE). Here, the SiH bond is hit by series of colder carriers, finally resulting in the bond breakage. In the advanced variant of our HCD model, these two mechanisms are not only allowed to occur separately but also as a combination in which MVE processes promotes/facilitates/assists the SP process.

For different conditions, both mechanisms have a different contribution to HCD due to their different dependences of the energetical distribution function (DF) of the charge carriers. As such, the HCD model requires an accurate DF, which must be obtained from a carrier transport model accounting for the most important scattering mechanisms. Among them are the electron-electron scattering (EES), ionized-impurity scattering, surface scattering, and electron-phonon interactions, all of which are incorporated in the deterministic Boltzmann transport solver ViennaSHE.

For the validation of the model, we simulated SiON nMOSFETs of an identical architecture but with different gate lengths, i.e. L_G = 65, 100, and 150 nm, which corresponds to channel lengths of 45, 80, and 120 nm. The devices were subjected to hot-carrier stress at the worst-case HCD conditions, i.e. drain voltages of V_{ds} = 1.8 and 2.2 V. As shown in Fig. 7, good agreement between experiment and our improved HCD model could be achieved. It is emphasized here that the neglect of the AB-, the MVE-mechanism, or EES for example can lead to a severe underestimation of the predicted degradation. Therefore, each of these mechanisms must be properly treated by any HCD model of a TCAD device simulator.

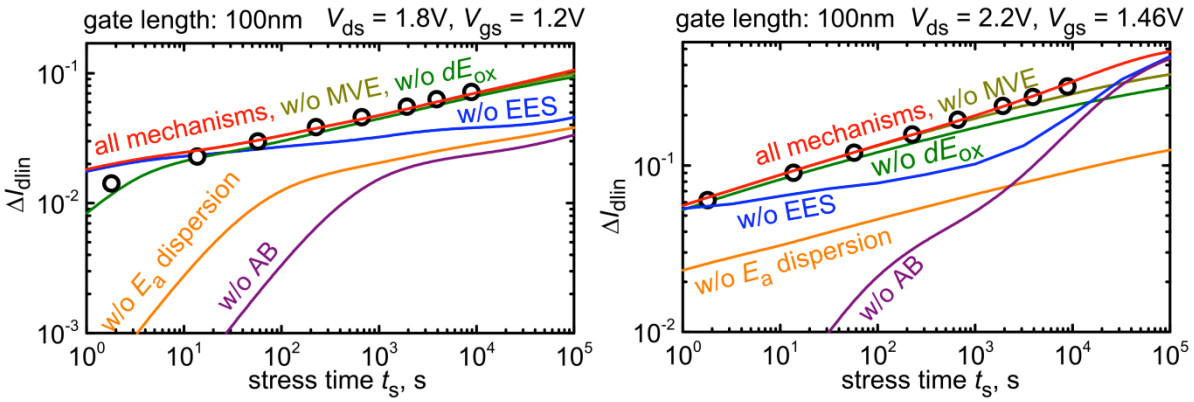


Fig. 7: Evaluation of the HCD model (solid line) against the experimental data (circles) of an nMOSFET with a 65nm gate length. Neglecting the AB or the MVE processes results in significant underestimation of the simulated degradation. It is furthermore noted that the agreement has been achieved using a unique set of parameters for all stress conditions.

CONCLUSIONS

In this work package we laid the theoretical foundation for the multi-state NMP model, which can now describe the dynamics of single defects as well as of ensembles of defects. Furthermore, it has been successfully demonstrated that the multi-state NMP model does not

only give an explanation for charge capture and emission in NBTI but also for PBTI, drain current RTN, and SILC. As such, the multi-state NMP model provides a comprehensive description of oxide defects, thereby further corroborating the validity of this model.

3. WP3 Electron Transport

In this report, we briefly discuss the deliverables and present the main main scientific results achieved in Work-Package 3 (WP3). WP3 has three specific objectives which have all being covered in the lifetime of the project. The first one is to develop various theoretical models, tools and methodology for physical simulation of the oxide wear-out associated charge trapping, which is the key phenomenon intervening in several degradation mechanisms and which dynamically affects not only the threshold voltage but also the ON current of dvices. The second aims to transfer the WP3.1 developed models into a Monte-Carlo Simulator in order to properly evaluate the ON current degradation and finally implemented reliability framework shall be included into a specifically developed quantum solver. WP3 interacts with other MORDRED WPs as well; atomistic level models developed in WP1 are underlying the WP3 capture and emission models, which are later calibrated against traps characteristics distribution measured in WP2. The final goal is to validate the models for a carefully calibrated device using the data obtained from the specifically produced FETs in Work-Package 4 (WP4). Finally statistical compact models required for the WP5 are extracted from the WP3 TCAD obtained statistical results. The work in this WP3 has been reported in three deliverables (D3.1, D3.2 and D3.3), which have been delivered on 16 months, 30 months and 48 months of the project lifetime correspondingly. All objectives of WP3 have been achieved and the obtained results, developed by the University of Glasgow computational framework, are reported in the deliverables as well.

INTRODUCTION

One of the biggest challenges that Reliability-Aware Design methodology has to face today, in order to keep pace with the aggressive transistor scaling, is the statistical variability associated with the discreteness of charge and granularity of matter in nanometre scale CMOS transistors. On the top of this so called Statistical Variability (SV) oxide wear out had to be re-evaluated as a time dependent statistical variability because of the discreteness of charged traps impact on the devices parameters, which arises from transistors downscaling, ending the reliability impact averaging. In the last few years several works have assessed the intrinsic connection between random fluctuations and reliability performance, showing that reliability-related parameters, e.g. the device lifetime, have to be re-interpreted as stochastic variables. Corroborated by a surge of new experimental evidences, an important paradigm shift has recently identified the oxide traps as the uniquely responsible phenomenon leading to random telegraph noise (RTN) and bias temperature instabilities (BTI). (It should be noted that BTI had previously been attributed to reaction-diffusion phenomena in the oxide.)

VARIABILITY

The intrinsic interplay between reliability and variability can be exemplified by considering the percolative nature of the source-to-drain conduction in nanoscale MOSFETs arising from the potential fluctuations associated with the discrete dopants in the channel and other statistical variability sources. The device reliability is related to the trap formation and the subsequent charge-trapping phenomena in the gate oxide. It is clear that a charge trapped over a percolative conduction path will have a large impact on the device threshold voltage shift, while the impact will be much smaller if the charge is trapped over a region which

already has low current density. This is demonstrated in Figure - 1, where the 3D simulated carrier density shows the percolative nature of conduction between the source and drain and the impact of a cluster of traps located over the preferential conduction path.

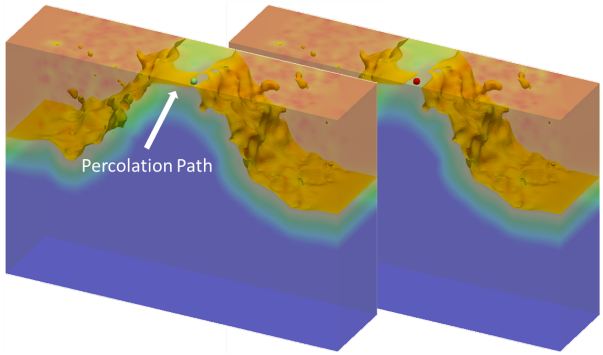


Figure - 1: (a) Example of a percolative conduction path between the source and drain in an atomistically doped 25MOSFET device. (b) The effect of a single trapped charge closing off the main conduction path during the threshold voltage reading operation.

Hence, the importance of developing three-dimensional (3D) physics-based simulation tools to elucidate on the interactions between reliability and statistical variability becomes evident. This is the main objective of Work-Package three (WP3) in the MORDRED project, where we developed three-dimensional (3D) physics-based simulation tools to illuminate on the phenomenology of reliability in presence of statistical variability.

The implementation strategies of these original tools are detailed in MORDRED deliverable 3.1; basically traps properties and traps dynamics models have been developed based on the refined understanding of the distribution of their properties brought through WP1 and WP2. Based on these physical models, a Kinetic Monte Carlo engine is charged with selecting the most probable charge trapping and detrapping to occur, emulating the stochastic character of this phenomenon. After the charge occupancy of the selected traps has been updated, the device electrostatic is solved again to obtain the specific impact of this trap on the performances. Some of the typical results reported in the deliverable 3.1 are presented in Figure - 2 and Figure - 3: A BTI charge experiment, progressively increasing the threshold voltages of the 100 atomistically different (due to SV) devices of our sample The lifetime distribution of these devices can be extracted based on a fixed criterion (here 30mV). RTN and BTI traces for two devices with random traps are illustrated in Fig.2, showing the different expressions of reliability with bias conditions and traps properties

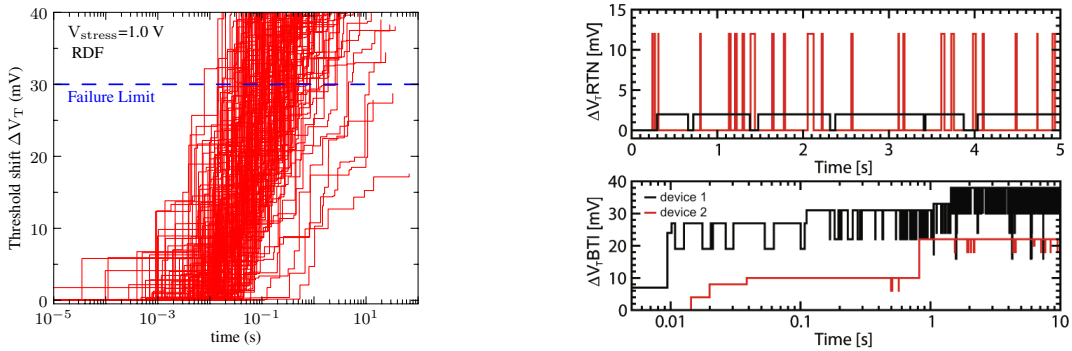


Figure - 2: Left: Threshold voltage stochastic degradation under BTI stress conditions obtained with the Drift-Diffusion simulation tools developed and submitted with (MORDRED DELIVERABLE 3.2). Random traps in atomisticcally diferent devices lead to a broad

dispersion of the time dependent devices parameters. Right: RTN and BTI signals for two different devices with random traps: interactions with SV as well as traps position and parameters dispersions causes different behaviours at device level.

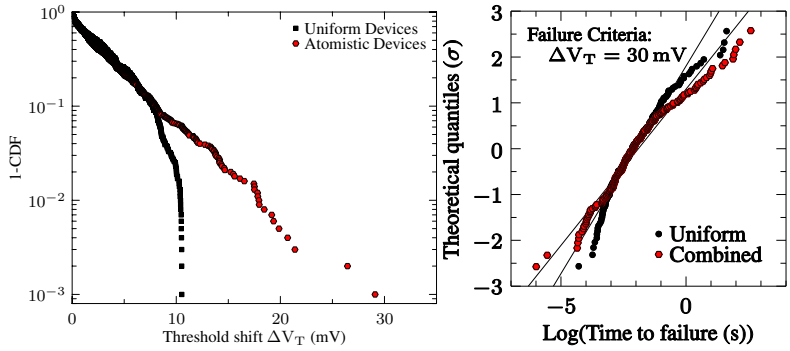


Figure - 3: Left: Single trapped charge induced threshold voltage shift distribution obtained with the Drift-Diffusion simulation tools developed and submitted with MORDRED DELIVERABLE 3.1. Right: time to failure distribution for uniform and Atomistic devices suffering from combined variability sources.

These DD simulation methodologies prove accurate and efficient for the study of the MOSFET behaviour in the sub-threshold region, where the carrier transport is mainly diffusive and the effect of scattering phenomena on drain current is negligible and from which the threshold voltage is extracted. However the ON current behaviour degradation and variability studies require more advanced Monte Carlo (MC) simulations frameworks accounting for scattering effects which are fundamental in determining the drain current. The DD simulation only mimics the scattering effects by means of calibrated bulk mobility models; for this reason WP3.2 focuses in transferring the methodologies and tools developed in WP3.1 into our MC simulator. In return the MC obtained enhanced ON current dispersion will be used as a reference to calibrate DD mobility models. In Figure 4, we reveal the degree of correlation between the drain current degradation at low and high gate biases, as predicted from the MC and DD simulations. In the case of low drain bias we can see that the MC results are less correlated than the DD results. This is a direct consequence of the additional variability of local scattering associated with each trapped charge in the MC simulation. On the contrary, at high drain bias the correlation increases and the difference between the MC and DD results is reduced because of the lowering of the impact of local scattering in this regime. Additionally, in Figure -5 we report the drain current degradation value as a function of the trap position along the channel length. The atomistic results obtained for several gate biases and at low and high drain biases confirm the trends observed for the case of a continuously doped device. However discrepancies are observed between DD and MC due to the fundamentally different charge injection regimes, dominated for MC approach by the velocity.

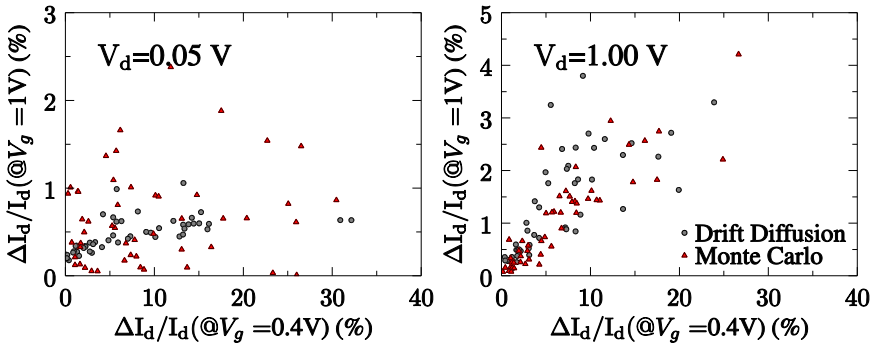


Figure - 4: The MC and DD trap impact on current degradation at high gate vs. trap impact at low gate. (MORDRED DELIVERABLE 3.2).

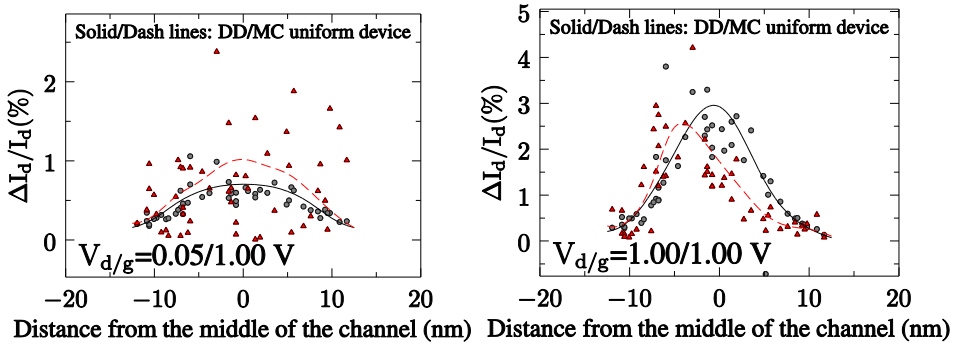


Figure - 5: Trap impact at several biases as a function of their position along the channel. (MORDRED DELIVERABLE 3.2).

On the contrary DD mobility model is uniform, therefor averaging the local mobility fluctuations arising from charge scattering simulated in MC. In particular in case of a charged trap, only the electrostatic impact is taken into account in the DD framework. To improve this point we successfully derived a local mobility model simulating the MC obtained impact. This model has been derived from the local current densities ratios MC/DD. Figure 6 presents both this ratio and the obtained traps impact characteristics with and without the local mobility correction.

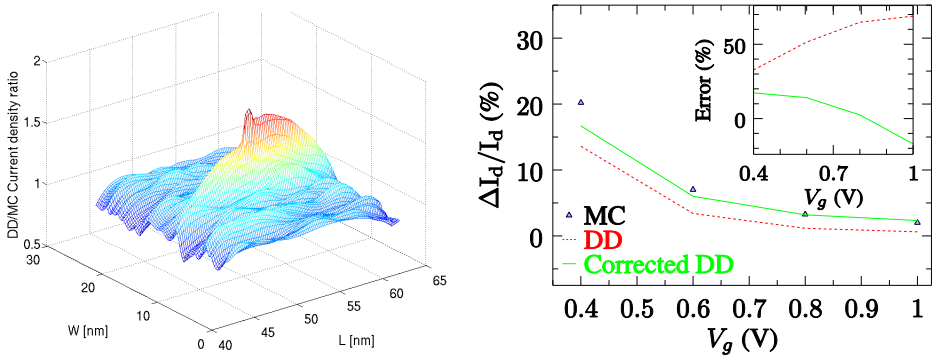


Figure - 6: Left: Current density ratio between the DD and MC simulation at V_G=0.8V on a 2D plane at 1nm below the channel/oxide interface. Right: Corrected and uncorrected DD, using the local mobility correction derived from MC simulations, to be compared with MC obtained trap impact (MORDRED DELIVERABLE 3.2).

In last year deliverable (D3.3), we discussed results based on the Non Equilibrium Green Functions (NEGF) simulation tool developed by us. This tool encapsulates accurate

physical models based on an important quantum mechanical effect in charge distribution and current transport in nanoscale devices. Introducing of quantum effects in current transistors is essential in order to fully unlock the variability and reliability issues and to draw accurate pictures about the device performance. Indeed, this is the main objective in task 3.3 and the work described in deliverable D3.3. In this deliverable we investigated the impact of a single discrete charge trapped at the oxide interface on the performance of scaled nMOS FinFET transistors, taking into account the quantum nature of the charge distribution.

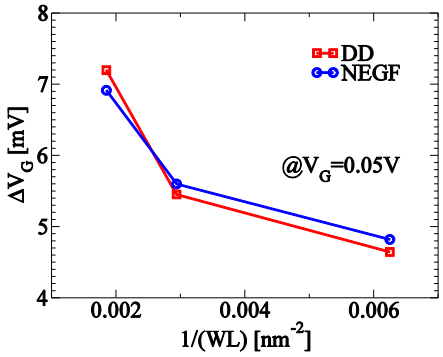


Figure - 4 Gate voltage shift due to a single discrete trapped charge on the top center of the fin as a function of the FinFET active area size obtained by the DD and NEGF simulations with a reading current criterion in the sub-threshold region. Note, $W=W_{FIN}+2\times H_{FIN}$. (MORDRED DELIVERABLE 3.2).

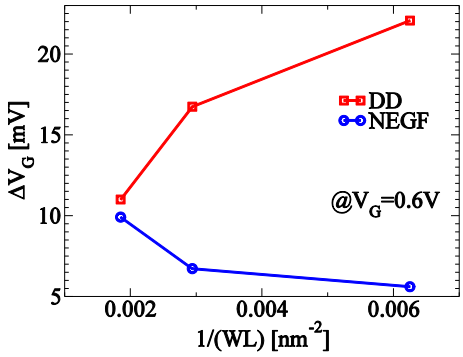


Figure - 5 Gate voltage shift due to a single discrete trapped charge on the top center of the fin as a function of the FinFET active area size obtained by the DD and NEGF simulations with a reading current criterion in nearly ON-state region. Note, $W=W_{FIN}+2\times H_{FIN}$. (MORDRED DELIVERABLE 3.2).

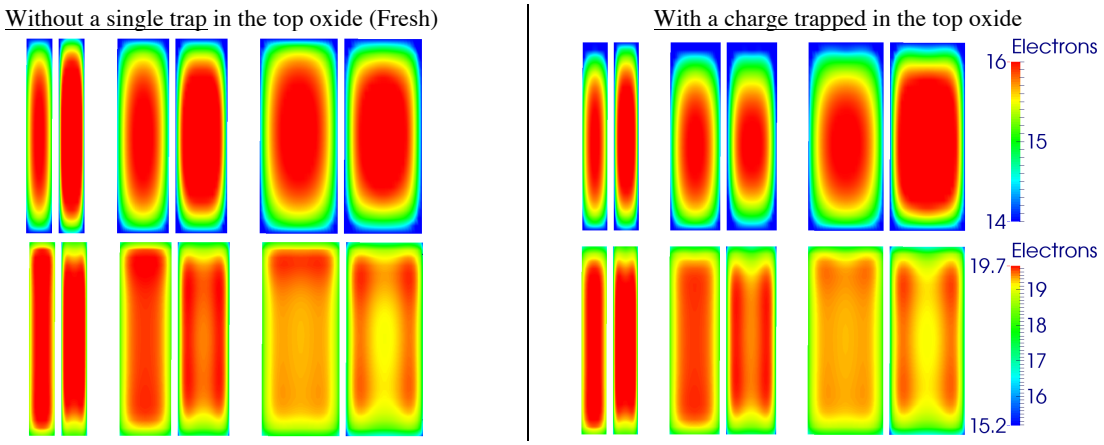


Figure - 6 2D cross-sections (at the center of the channel) of the simulated electron density at the threshold voltage conditions for a current criterion in the sub-threshold region (TOP) and in the ON-state region (BOTTOM). The left (right) of each couple of FinFET cross-sections represents the DD(NEGF) results (MORDRED DELIVERABLE 3.3).

Some results from deliverable D3.3 are presented in Figure - 4 and Figure - 5. Figure - 4 shows the gate voltage shift obtained in the subthreshold regime, as a function of scaling. Both the DD and NEGF simulations show similar results with a surprising "inverse-scaling" behaviour. Indeed, the impact of a trap located on the top-fin oxide increases when the fin size increases. This behaviour is counterintuitive if compared to the trends expected for the conventional planar MOSFETs.

Figure - 5 reveals the gate voltage shift obtained in the ON-state regime. In this case, the DD simulations yield results in accordance with the expected scaling trends. However, the NEGF simulations result in a lower gate voltage shift in line withe the trend in the subthreshold regime. It is important to note that both the DD and NEGF results confirm a positive trend between the gate voltage shift value and the gate drive voltage. In fact, the gate voltage shift increases, moving from the subthreshold to the ON-state region, in contrast to what is showed by the conventional planar MOSFETs.

In order to understand these trends observed in Figure - 4 and Figure - 5, in Figure - 6 we report the simulated charge density at a low and high gate bias. Firstly, it is evident that the channel inversion occurs deeply inside the bulk in the subthreshold regime, while it moves closer to the oxide interface in the ON-state regime. This explains why the trap impact increases with increasing of shift in the gate voltage (a common feature for the DD and NEGF simulations). It should be noted that this behaviour cannot be observed in a conventional planar MOSFET, where the inversion charge centroid is always close to the oxide interface and barely modulated by the gate voltage conditions. Secondly, Figure - 6 shows small differences between the DD and NEGF results in the subthreshold regime. This explains the similar gate voltage shift values reported in Figure - 4 for the two simulation approaches. On the other hand, the DD and NEGF charge distributions distinctly differ in the ON-state regime. While the DD charge moves uniformly towards the fin top interface, the NEGF charge follows a similar trend but maintaining two distinct peaks close to the lateral fin sidewalls.

CONCLUSIONS

In summary, this document reports briefly on the achievements and some of the results of WP3 of the MORDRED project. During the lifetime of the project all objectives and tasks of WP3 have been addressed and fulfilled. We also provided three deliverables, where we reported our progress and scientific results.

Overall, in WP3, we developed and implemented models and tools that allow us to address a broad range of reliability issues by means of statistical numerical calculations, using successively DD, MC and finally NEGF approaches. A reliability module has been implemented allowing the simulation of time dependent variability and the extraction of devices lifetime including interactions between traps and statistical variability. Additional scattering effects induced by charged traps have been studied using MC approaches and a local mobility correction model has been developed and validated to emulate the local mobility perturbations due to the trap into the higher level DD framework. Finally a quantum simulator has been implemented and NEGF results have been compared to classical simulation results. In particular scaling trends have been studied within the quantum and classical approaches. The results based on our work carried out in the MORDERED project have been published in numerous scientific papers and presented at international conferences.

4. WP4 Sample fabrication and characterization

The so-called Bias-temperature instability (BTI) and Random Telegraph Noise (RTN) 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. MORDRED Work Package 4 (WP4) was aimed at the *experimental* characterization of the properties of these traps, both *electrical* and *physical*, as well as sample fabrication for the entire project. The experimental results were typically first analyzed within the WP using standard techniques, and then provided to other WPs for advanced analysis, calibration, and modeling.

Within WP4, the properties of trap *populations* have been studied, such as trap densities, and energetical and spatial distributions in the oxide. This is showcased in Section II. The hallmark of this project has been, arguably, understanding the properties of *individual* traps, such as their capture and emission times and impact of device characteristics. In Sections III and IV, the temporal and electrostatic properties of individual bulk oxide trap are presented, as investigated in nm-sized devices.

Si/SiO$_2$ and SiGe/SiO$_2$ 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_{it} of (110), (111) and (100)Si with thermal SiO_2 systems by using conductance and capacitance methods. The results of these experiments are illustrated by the experimental $D_{it}(E)$ profiles shown in Fig. 1. The inferred D_{it} is found highly sensitive to the Si crystal orientation, highest for the (111) face, and lowest for the (100) face, with (110)Si/SiO_2 closely resembling the (111)Si/SiO_2 case. Two peaks in the $D_{it}(E)$ 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 P_b 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/HfO_2 (1.8 nm)/TiN_x/poly-Si structures. The scavenging has been performed by applying laser annealing to the samples in O-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/HfO_2 (1.8 nm) interface the total N_{it} of about $5\times10^{12}cm^{-2}$ is estimated.

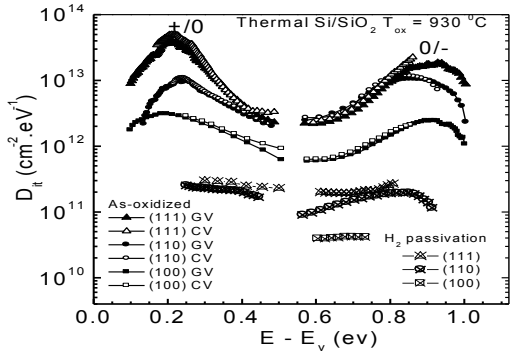


Fig. 1. $D_{it}(E)$ profiles of Si/SiO_2 interfaces derived from CV (solid symbols) and GV (open symbols) methods in Si/SiO_2 samples fabricated on (100), (110), and (111) faces of Si.

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 P_{b0} 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 P_{b0} centers found from these measurements are summarized in Fig. 2 for 1.8 nm thick HfO_2 and ZrO_2 insulators. The results indicate an increase in the density of P_{b0} centers after high-temperature annealing of Si/HfO_2 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 SiO_2-like IL with Si substrate:

$$SiO_2 + Si \rightarrow 2SiO(gas), \qquad (1)$$

which has earlier been identified as the primary mechanism of degradation of Si/SiO_2 interfaces upon annealing in the O-deficient ambient. Application of He gas during annealing allows one to significantly suppress this kind of degradation. 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 (P_b-type defects) in $Si_{1-x}Ge_x$ alloys ($x = 0.75$) at the interface with well-studied thermal SiO_2 insulator as affected by annealing in molecular hydrogen (passivation) and by the subsequent annealing in high vacuum (de-passivation).

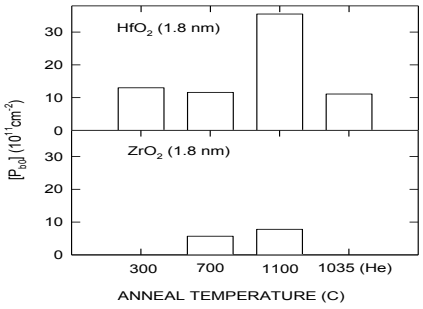


Fig. 2. Density of interfacial P_{b0} defects found at interfaces of (100)Si with 1.8-nm thick layers of HfO_2 (top panel) and ZrO_2 (bottom panel) after different annealing treatments.

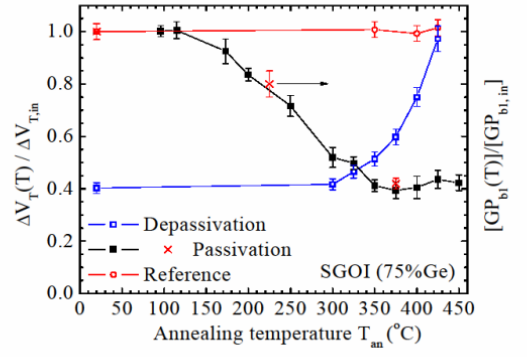


Fig. 3. Density of electrically active paramagnetic Ge P_{b1} defects observed at the interfaces of $Si_{0.25}Ge_{0.75}$ alloy with thermal SiO_2 as a function of isochronal (30 min) anneal temperature in 1.1 atm. H_2 (passivation) or in vacuum (dissociation).

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 P_{b1} centers as acceptors in SiGe and by ESR. In general, both passivation and de-passivation processes are observed to occur at lower temperatures than at the Si/SiO_2 interfaces. The activation energies of germanium dangling bond passivation by H_2 molecule and the activation energy of Ge-H bond thermal dissociation were found equal to E_f =1.44 ± 0.04 eV and E_d= 2.23 ±0.04 eV, respectively.

Contrary to the well-studied case of dangling bonds of silicon atoms at the interfaces with the same insulator (SiO_2), the sum of these two energies appears to be significantly lower than the thermodynamic binding energy of the H_2 molecule (4.4 eV). This discrepancy might be explained either by different charge state of the Ge P_b-type defects (negatively charged) during passivation or, else, by the presence of additional H_2 cracking centers. Furthermore, the statistical (site-to-site) spread of the activation energies for Ge P_b-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 H_2 passivation anneal (425 °C).

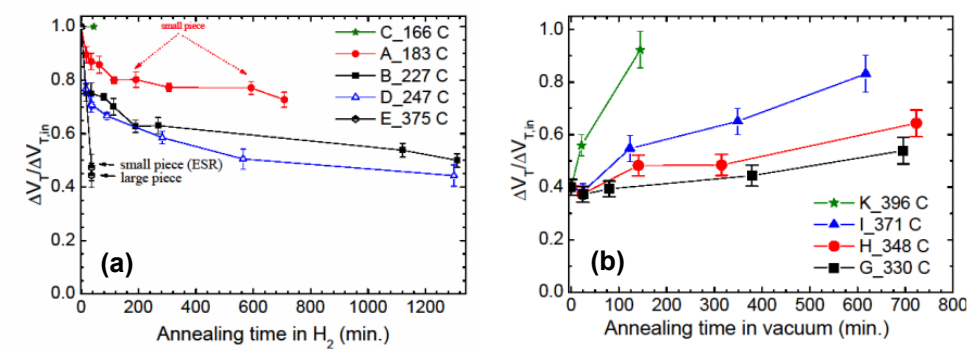


Fig. 4. Density of electrically active paramagnetic Ge P_{b1} defects observed at the interfaces of $Si_{0.25}Ge_{0.75}$ alloy with thermal SiO_2 as a function of isothermal anneal time in (a) 1.1 atm. H_2 (passivation) or (b) in vacuum (dissociation).

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. 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 ΔV_{th}-variability of deeply-scaled FETs.

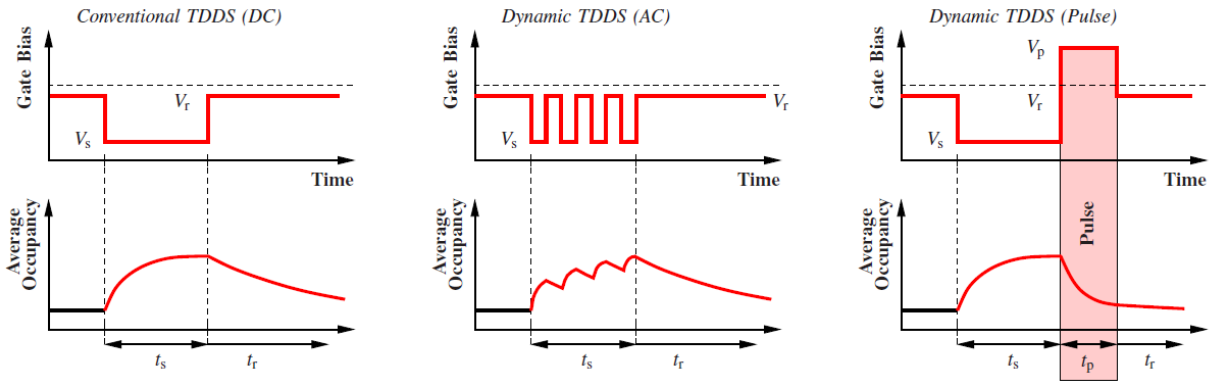


Fig. 5. Illustration of the 3 different measuring schemes used by TDDS: Shown left is the conventional DC TDDS measurement procedure, during which defects are charged at a stress voltage V_s and discharged during the subsequent recovery at V_r. This sequence is typically repeated 100 times to allow for 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.

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.

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. Varying the AC stress frequency for 1kHz to 5MHz clearly reveals a frequency-dependence of the capture time of individual defects. 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. τ_e then is calculated from the decay rate $\exp(-t_p/\tau_e)$, which describes the occupancy during the recovery pulse.

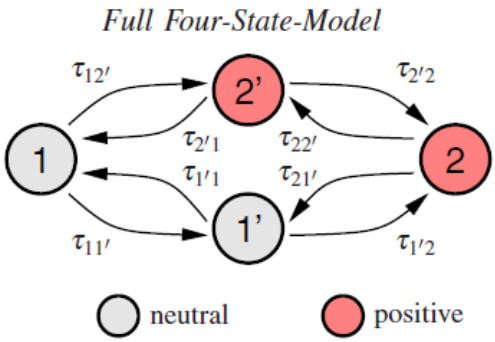


Fig. 6. The four states of oxide defects extracted from TDDS experiments. Each defect has two stable states, 1 and 2, and possible two metastable states 1' and 2'. The metastable state 2' seems to be always present, while the existence of metastable state 1' decides on whether the trap behaves like a fixed positive or a switching trap.

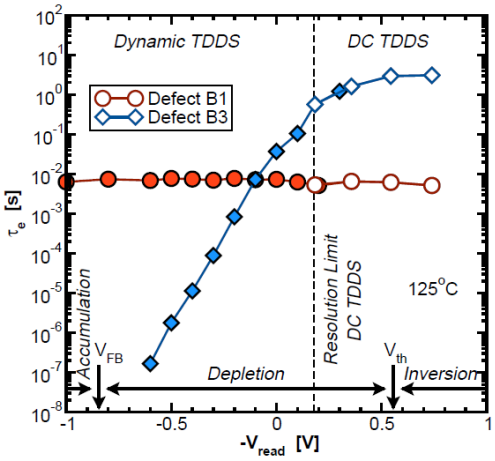


Fig. 7. An example for the wide range of extracted emission time constants for two defects using the "dynamic" TDDS technique. The range is extended to gate voltages far below threshold where no drain current can be measured. The emission time constant of defect B1 is bias-independent revealing B1 as 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.

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.

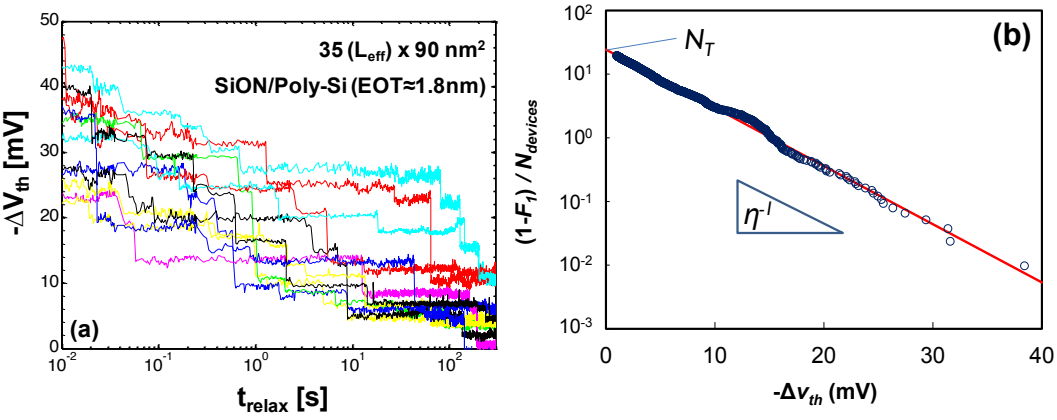


Fig. 8. (a) Typical NBTI relaxation transients, each recorded on a fresh nanoscaled pFET (W=90nm, L_{eff}=35nm, SiON/Poly-Si, EOT≈1.8nm) [14]. Steps of varying heights due to single defect-discharge events are clearly visible. (b) The ΔV_{th} step heights corresponding to the individual discharge events observed in the relaxation transients plotted on a complementary cumulative distribution function (1-*CDF*) plot. When normalized by the number of devices, the intercept with the y-axis gives the average number of defects per device N_T that emitted in the measured relaxation interval.

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. 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 ΔV_{th} relaxation curves in TDDS-like measurements performed on *multiple* deeply-scaled devices. Each device behaves differently, resulting in large time-dependent variability of the total ΔV_{th}. The ΔV_{th} step heights appear exponentially distributed, with the cumulative distribution function (CDF)

$$F_{\eta}(\Delta v_{th}) = 1 - \exp\left(-\frac{\Delta v_{th}}{\eta}\right), \qquad (2)$$

with an average step height η (Fig. 8b). Note that while the majority of charged traps results in small ΔV_{th}'s, a small fraction of traps will significantly change V_{th} by up to tens of mV.

The average impact of a single defect η on the threshold voltage is a fundamental parameter determining the variability of deeply scaled technologies. 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

$$\sigma^2_{\Delta V_{th}}(t) = 2\eta\langle\Delta V_{th}(t)\rangle, \quad (3)$$

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

A low value of η is generally desired. η has been argued to scale as

$$\eta \cong \frac{t_{inv} N_A^{\alpha}}{A}, \qquad (4)$$

where t_{inv} is the oxide thickness corresponding to capacitance in inversion, A the area of the device channel, and N_A the channel doping, with the exponent α has been observed to be around 0.5 in simulations. In order to test Eq. 4 experimentally, we collected an extensive dataset of NBTI relaxation transients devices with different areas. Figure 9a demonstrates η scales reciprocally with device gate area, in agreement with Eq. 4. Figure 9b then demonstrates that η 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.

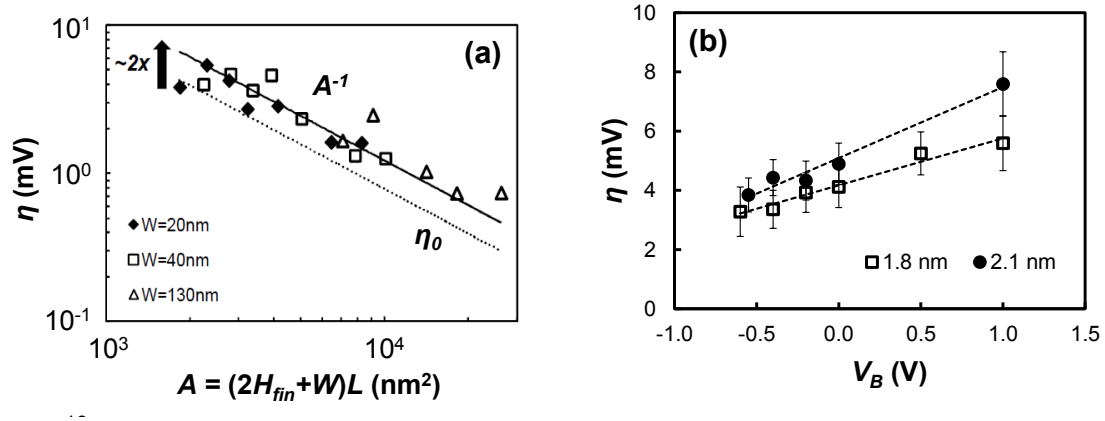


Fig. 9. (a) The average step height η scales as A^{-1} on Si pFinFETs (high-k/MG, t_{inv} ≈ 1.7nm) with varying fin width W and gate length L (fin height H is fixed). Each point is extracted from a set of multiple devices with identical dimensions as in Fig. 8. (b) The average step height η values extracted from distributions measured for varying back-bias V_B on two wafers with identical Si/SiON/Poly-Si planar pFETs and identical doping levels, but slightly different oxide thicknesses (t_{inv} = ~1.8 and ~2.1nm). Thicker oxide increases η while forward (reverse) back bias reduces (increases) the depletion width and thus reduces (increases) η, as per Eq. 4.

Over the whole measured range in Fig. 9a, η is observed ~2× higher than the "naively" expected impact η_0, given by the charge sheet approximation for a single charge at substrate/dielectric interface

$$\eta_0 = \frac{q}{C_{ox}} \quad . \qquad (5)$$

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 η will be reduced for oxide defects closer to the gate (Fig. 10a). In addition to the low impact per charged defect η, the superior NBTI robustness of SiGe-based pFET devices is also reflected in the significantly decreased number of active defects, documented in Fig. 10b.

In low-doped channel devices, we have experimentally demonstrated that charged *interface* traps represent a new source of variability, as shown in Fig. 11a. The value of η is then increased by further interface trap generation during operation, see Fig. 11b.

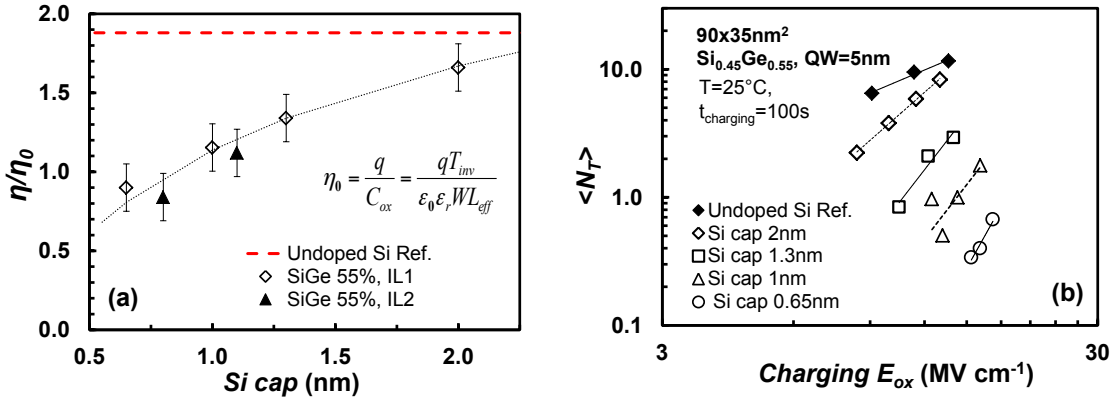


Fig. 10. (a) Extracted average ΔV_{th} step heights η for SiGe devices with different Si cap and for undoped Si channel devices, after a charging phase at E_{ox}≈12MV/cm. The extracted values of η are normalized by η_0. SiGe devices with the thinnest Si cap show a significantly lower η. The observation is confirmed on SiGe devices with two different SiO_2 interfacial layer thicknesses. The red dashed line demarcates the benchmark value experimentally estimated on undoped Si channel ref. devices. (b) Consistently with the significant NBTI reduction in large area devices, nanoscaled SiGe channel pMOSFETs with a reduced Si cap thickness show reduced average number of charging/discharging defects per device $\langle N_T \rangle$, and a stronger field acceleration.

Precise circuit simulations will require models incorporating the impact of the small number of trapped charges on the *entire* FET current characteristics . 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). The learning from measurements and fullscale 3D 'atomistic' simulations (Fig. 12c) can be captured in reliability-aware FET compact models.

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. 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. Based on this picture we have proposed a qualitative model describing the correlated ΔI_G and ΔI_D distributions (Fig. 13).

CONCLUSIONS

Substantial novel insights have been generated in WP4 over the four years of the project, enabling new approaches to understanding reliability and reliability statistics in present and especially upcoming CMOS technologies.

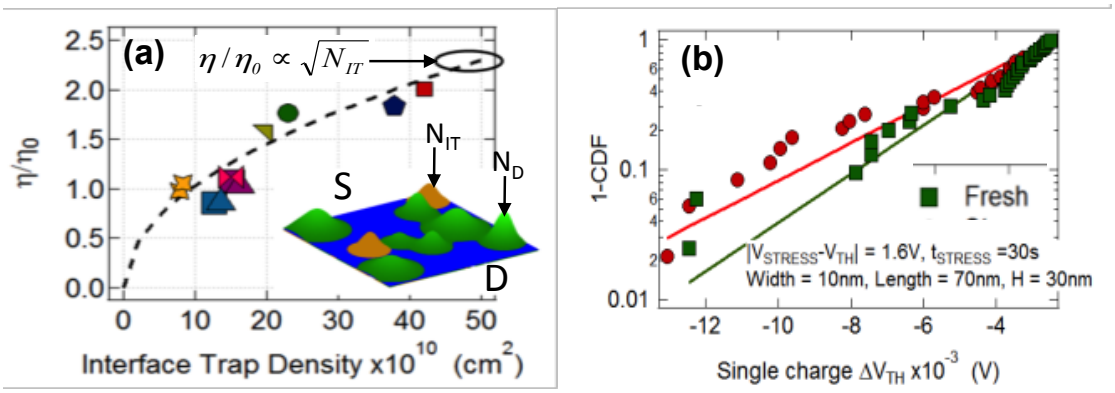


Fig. 11. (a) A clear correlation is found between single-carrier discharge ΔV_{th} and D_{it} in FinFET devices. Interface traps enhance channel conduction percolation between source and drain in the same way dopants do (inset). (b) Generation of D_{it} after electrical stress is reflected in increase of single-carrier discharge ΔV_{TH} .

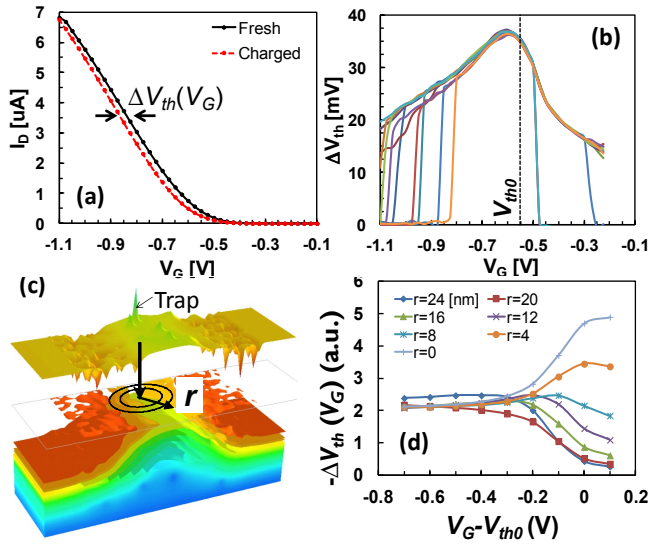


Fig. 12. The "ultimate" reliability measurement: (a) I_D-V_G curves measured at V_D = -0.1 V on a $L \times W$ = 35×90 nm^2 pFET with a single oxide defect uncharged and charged. The corresponding $\Delta V_{th}(V_G)$ is defined as a simple "geometrical" horizontal distance between the two curves at 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*.

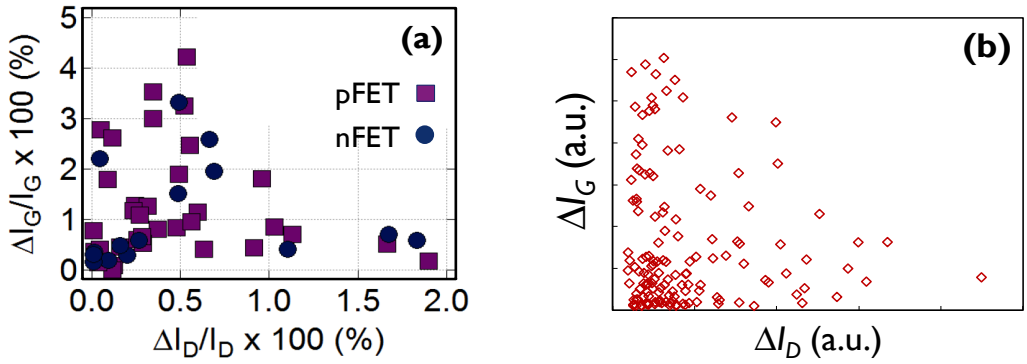


Fig. 13: (a) Correlation plot of relative changes in I_D and I_G (I_G = ~10 pA). Note that upper right corner (large I_D and large I_G fluctuations) is unpopulated. The proposed mechanism involves enhanced conduction through a gate oxide trap when it is unoccupied. Such a process would be most efficient for traps close to the center of the oxide, readily explaining why there are no defects causing both large ΔI_G and large ΔI_D. (b) The same correlation can be qualitatively constructed using this assumption, combined with the impact of a charged trap on I_D depending on the distance of the trap from the critical point of a source-drain percolation path (Fig. 12).

5. WP5 Compact models

Transistor degradation mechanisms such as negative bias temperature instability (NBTI) and hot carrier injection (HCI) were and are considered as important aging effects. Both NBTI and HCI generate traps in the transistor gate oxide which affect parameters like the threshold voltage of a transistor and remain a concern across different technologies. On the other hand, positive bias temperature instability (PBTI) is a more recent type of degradation that became more apparent with the use of HK material in the gate stack. All these ageing effects have an impact on transistors. Circuits designed using such transistors also suffer from aging effects.

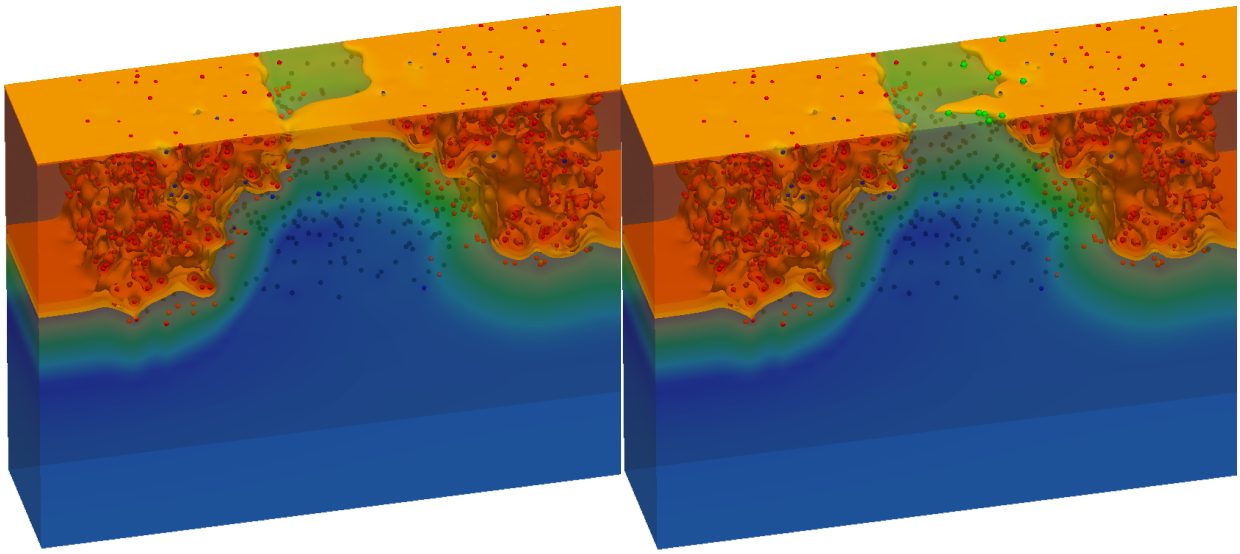
The backbone of this workpackage is the development of analytical ageing models, a flexible statistical compact model extraction methodology and tools and and statistical circuit simulator engine which will enable the design for reliability and design technology co-optimisation techniques required in advanced technologies.

Here, we present a summary of the full research activities for Workpackage 5. In particular, we are will summarize the main results achieved in WP5:

- Delivered an optimal set of statistical parameters for NBTI, PBTI and HCD
- A statistical SPICE compact model extraction methodology and reliability ready statistical compact model extractor.
- Delivered time dependent statistical compact model strategies and incorporated them into a design PDK.
- Delivered an evaluation of the impact of degradation on SRAM and simple standard cells and basic analog cells.
- Delivered an evaluation of compact models against measured circuits which compares analytical compact models with silicon measurement.
- Delivered a Reliability ready statistical circuit simulator via the GSS RandomSpice simulator and the GSS ModelGEN technology.

RELIABILITY READY STATISTICAL COMPACT MODEL EXTRACTOR AND SIMULATOR

SPICE Compact Models are the link between device technology and design. In D5.1 a selection of compact model parameters, which are capable of capturing the impact of NBTI, PBTI and HCI degradation on transistor performance have been identified. We have developed robust extraction methodologies for SPICE models of devices under the influence of statistical degradation, specifically, the impact of trapped charges at the oxide interface of the MOSFET as illustrated in Fig. 1.



(a) (b)

Figure 1: The electron density in a single atomistically simulated device. In figure (a) a clear current path exists between the source and drain. Figure (b) shows the effect of trapped charges (shown in green), which have completely closed off the current path resulting in a large change in device threshold voltage.

A template 25nm device developed by GSS was adopted for this demonstration, and for which the extraction strategy for the nominal characteristics and for statistical variability have been already published.

Over the course of the MORDRED project we have developed significant capabilities in the GSS GARAND simulation software, which allow the modelling of the impact of individual trapped charges on technology performance. Using GARAND, TCAD simulations have been performed for an ensemble of thousands of microscopically different 25nm N-Channel MOSFETs which are affected by statistical variability caused by the discreteness of charge and matter at the nanometer scale. These sources of variability cause differences in the behaviour of individual transistors and include the impacts of Random Discrete Dopants (RDD), Line edge roughness (LER) and Metal Gate Granularity (MGG). Each of these thousands of devices was simulated at 3 different levels of degradation where trapped charges with different concentrations are modelled as individual charges and the combined impact on the performance of the device ensemble was analysed.

In order to accurately model the impact of variability at the device level on circuits GSS has developed an advanced statistical compact model extraction tool called Mystic and a robust extraction methodology that has been specifically designed to capture the effect of statistical variability on modern transistor devices. Details of how this statistical compact model extraction methodology is used in order to very accurately re-produce both the behaviour of the nominal device and the ensemble of variable devices is provided in D 5.1 of the project.

Using ensembles of 1000 atomistic TCAD simulations three different levels of degradation have been modelled by introducing trapped charge at the Semiconductor-Oxide interface with densities of $1\times10^{11}cm^{-2}$, $5\times10^{11}cm^{-2}$ and $10^{12}cm^{-2}$ these represent low, medium and high levels of degradation seen in Silicon devices. The current voltage characteristics obtained from

these simulations are matched using the previously developed statistical compact model extraction strategy and an example is show in Fig. 2.

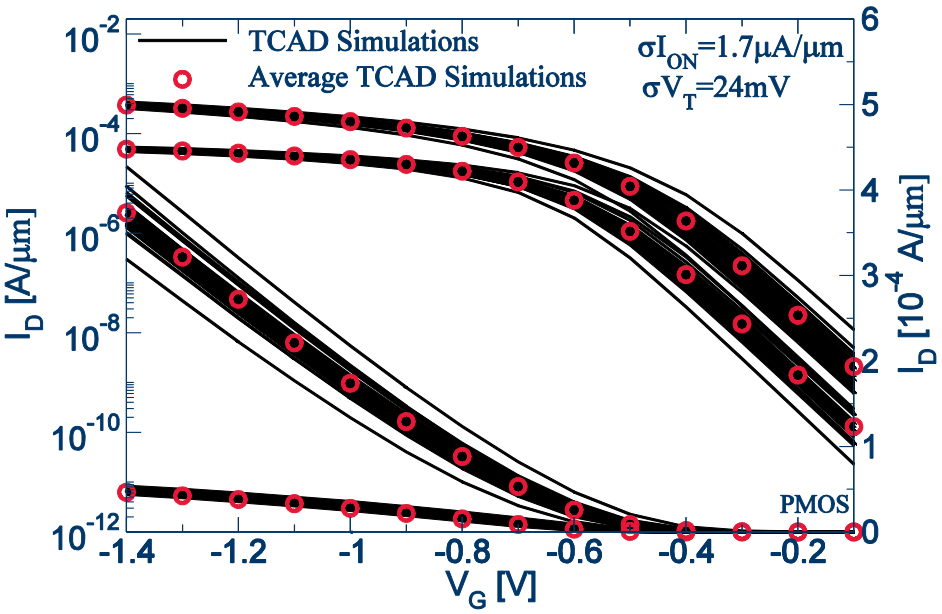


Figure 2: Current voltage characteristics of a P-type MOSFET in the presence of statistical variability and reliability effects.

The accuracy of the extracted models was assessed with respect to the level of degradation in order to determine whether the compact model parameter set utilised in this strategy is capable of capturing the effects of increased statistical variability after degradation or whether it requires modification.

D5.1.1 showed that a set of 7 BSIM 4 parameters was sufficient to fully capture both static statistical variability, as well as the additional variations introduced by including the effects of NBTI/PBTI in the TCAD simulations.

We then extended the work performed in D5.1 describing how the compact modelling results can be extended to circuit level simulation via several different compact modelling approaches. We demonstrated how these results can be incorporated into a strategy where interpolation can be performed over arbitrary trapped charge densities. Such interpolation over trapped charge density can be coupled to a model of time-dependent changes in such as those developed by KUL. It has been shown that these changes can be described by a power law function of time, and GSS implemented a similar model as a simple proof of concept approach in the GSS statistical circuit simulation engine RandomSpice (D5.3.3).

Three different compact modelling strategies have been compared, namely Gaussian V_T, full lookup table based models and more sophisticated fitting methods in RandomSpice. Simulations of a template SRAM cell were carried out using each of the three methods to be compared, showing how these methods can be employed in circuit and cell simulation. Cell metrics such as static noise margin (SNM) and projected yield at different levels of degradation have been compared across the three approaches. This resulted in the development of the GSS ModelGEN generation technology which has been implemented in the commercial version of the RandomSpice simulation engine which has been developed as part of the MORDRED project.

Furthermore, simulations carried out using the ModelGEN interpolators have been validated against both raw TCAD results, to ensure that the statistics of device performance are preserved, and against circuit simulations carried out using full models based on the same TCAD simulations, in order to demonstrate the validity and accuracy of the interpolators.

KUL-E has developed and demonstrated the strategies of time dependent compact models for NBTI, PBTI and HCD. The model calibration has been done under close collaboration with IMEC (see linked work with WP4). These models are targeted for circuit level reliability simulation. The effects of oxide degradation affect different transistor parameters (e.g. the threshold voltage V_{TH}, the carrier mobility μ and the output conductance g_0). The analysis was focused on V_{TH} degradation, since expressions for all degrading transistor parameters can be derived from the threshold voltage V_{TH}. It has been demonstrated that the time-dependent behavior of V_{TH} can, for both HC and NBTI, be described as a power-law function of time. The details of test chip design to validate and calibrate the proposed models and related simulation results details have been reported in D5.2.

In order to study realistic SRAM cells, GSS performed an analysis by calibrating TCAD process and Device simulations against experimental data for a 65nm device technology provided by IMEC as part of WP4 of the project. These devices feature a 60nm long and 90nm wide poly-silicon gate and a 2.2 nm thick SiON gate oxide. The simulation calibration methodology followed an iterative process designed to provide an accurate description of the doping profile and electrical behaviour of the technology and is divided into three stages: the first one deduces the realistic average doping profile of the devices, the second one ensure a match with the measured statistical variability and the last stage targets reliability. Details of this 3-step calibration results are detailed in Deliverable 5.3.1.

After TCAD simulation calibration, Compact model parameter extraction was performed using the GSS Mystic tool and the methodology developed over the project. This process follows a similar three stage process to the TCAD model development: after a global nominal compact model extraction including channel length and back-bias dependences, statistical compact models are extracted from TCAD results and finally statistical reliability compact models are extracted from the TCAD simulations at different levels of degradation. The advantage of extracted statistical compact models over analytically generated parameters dispersions lies in the preservation of the correlation between the statistical parameters and the transistor figures of merit, as illustrated in Fig. 3. Once the compact model parameters distributios are obtained, the statistical sample can be extended to reach the high sigma levels via the ModelGEN technology allowing proper SRAM reliability assessments.

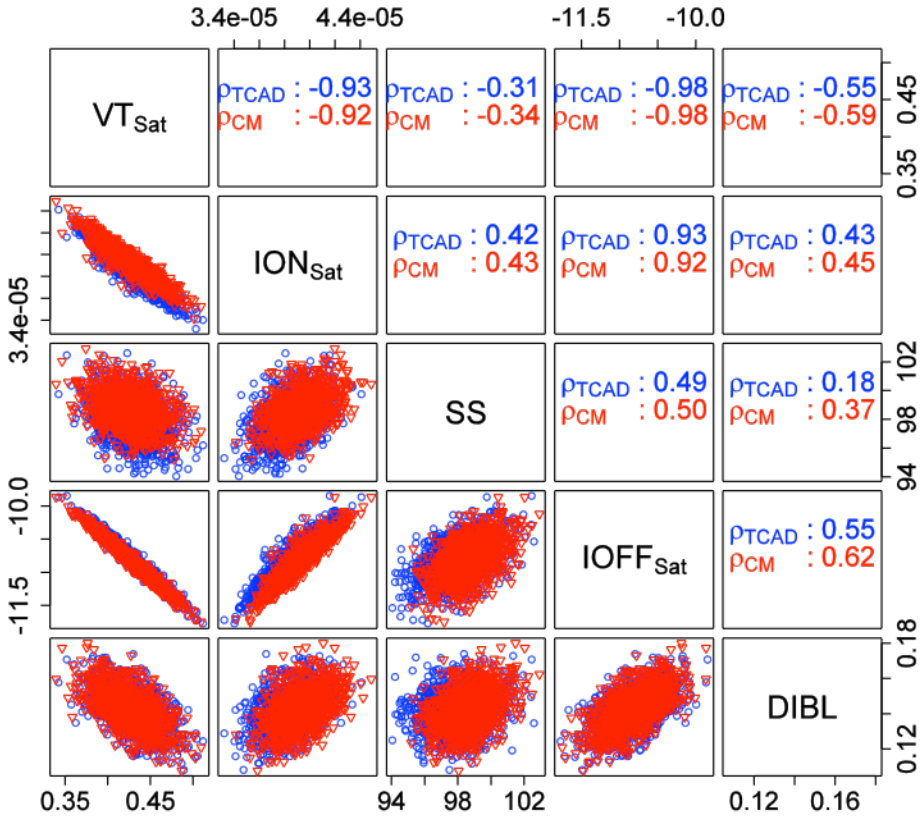


Figure 3: PMOS Compact Model and TCAD FoM distributions and correlations.

Once these statistical compact model libraries are obtained they can be applied to the investigation of the sensitivity of 6T-SRAM to ageing effects. SRAM is a memory circuit which is of significant importance to the global electronics industry as large portions of modern chips are given over to these circuits. Due to the very large numbers of SRAM cells used and the fact that SRAM is one of the most aggressively scaled portions of any technology it's correct operation in very rare cases if of paramount importance to the success of a product. Several scenarios of degradation have been considered which cover various stress and ageing conditions that may occur in SRAM operation. It has become clear that a holistic design technology co-optimisation process is required in order to get the best from a given semiconductor technology and the tools and methodologies developed in the MORDRED project are key enablers in this goal.

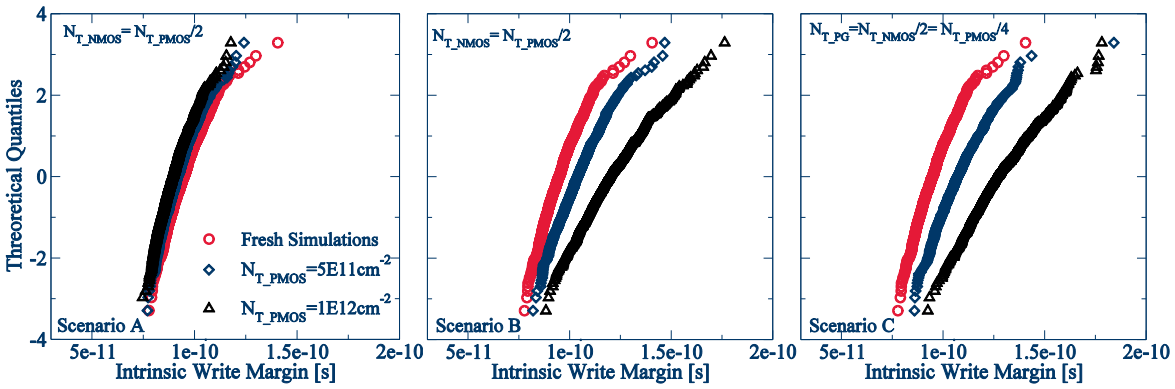


Figure 4: Example of SNM obtained from simulations of SRAM in different degradation scenarios.

Static Noise Margin (SNM) is a measure of the capability of an SRAM cell to retain the data stored in it. Statistical simulations of the SNM of cells have been studied for fresh devices with no degradation of their characteristics and different levels of charge trapping that correspond to low, medium and high levels of ageing related degradation (as illustrated in Fig. 4). As well as SNM, Write Time (WT) variation of the cell has also been investigated in the presence of combined statistical variability and reliability. Additionally, the impact of statistical variability and reliability on example digital standard cells and simple analogue circuits have been investigated providing a glimpse of how the software and techniques developed in this project could be used to optimise different aspects of the design process.

EVALUATION OF ANALYTICAL COMPACT MODELS AGAINST MEASURED CIRCUITS

In this work BTI and HCI compact models are described and compared to the ones available in literature. Principles of their usage in spice simulations to predict circuit ageing already at design time are explained. Because circuit designers usually require transient simulations and it is known that degradation in dynamic conditions is different from static conditions, our models were expanded to include the dynamic degradation (shown in Fig. 5). Single transistor measurements were used to calibrate the models. Voltage and temperature scaling together with the time evolution of degradation were measured and fitted by the models.

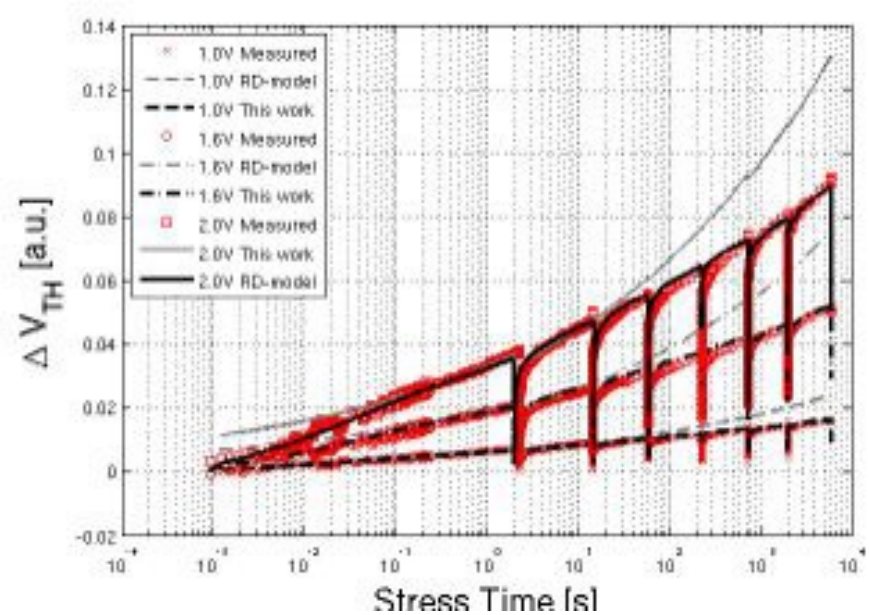


Figure 5 - NBTI compact model evaluated against pFET measurements at 125 ^{0}C for several stress voltages.

Using the knowledge gained through developing degradation compact models and preliminary simulations, a test chip design was created. Again, single transistors are used to calibrate the degradation compact models, but the evaluation needs to be done on circuits. Our test chip includes single devices along several common analogue and digital circuits including two operational trans-conductance amplifiers (OTA), an LC oscillator, two ring oscillators, NAND and NOR logic gates. A measurements scheme for the OTA is also proposed. Along to the test chip, which was produced in commercial 28 nm HKMG technology, also devices produced in the 65 nm SiON imec technology were measured an example of the response of the OTA to degradation is shown in Fig. 6.

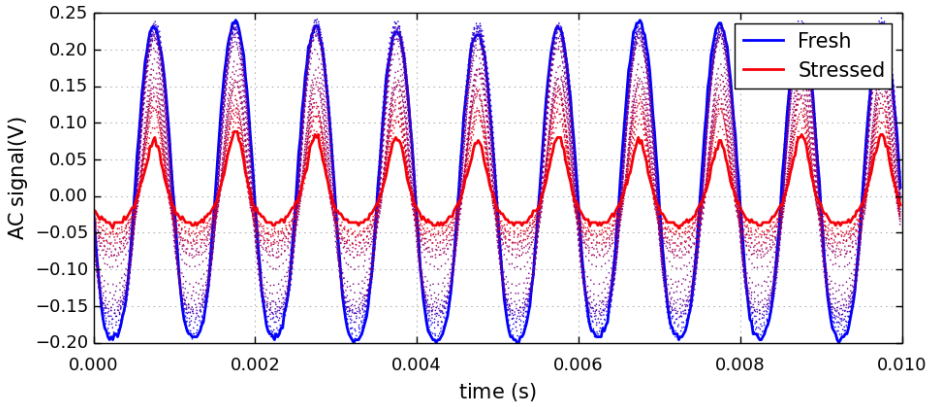


Figure 6 - Degradation of the OTA output AC signal. The large sinusoidal signal (blue) is measured before stress, while the small distorted sinusoidal (red) is measured after stress. The dotted signals are measured in intermediate steps between the first and last stress sequence.

As already mentioned, an inverter produced in IMEC technology within work package 4 is measured (Fig. 7). The results are then compared with simulations, using the degradation models which were calibrated to the same technology. As shown in Fig. 8, the fit of the measurements to the NBTI model simulations is not perfect due to the lack of a comprehensive electrical models of the characterised devices. However, even though the absolute values do not match well, the trend is clearly reproduced. The degradation compact models are thus validated against circuit measurements and the compact modelling loop is closed.

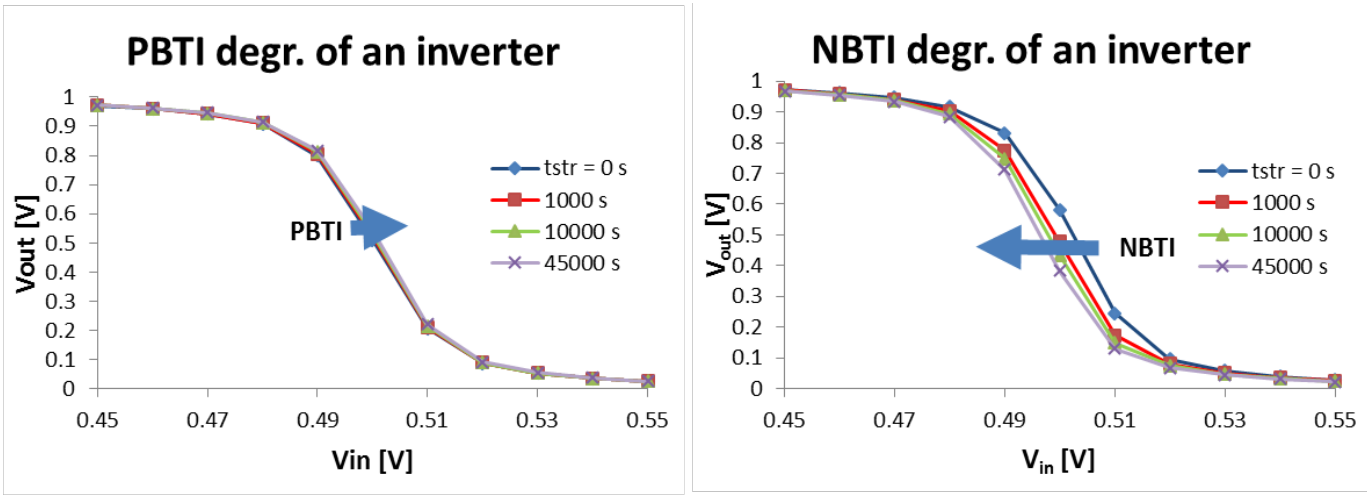


Figure 7 - PBTI degradation of the nMOS (left) and NBTI degradation of the pMOS (right) transistor of an inverter. The static characteristic of an virgin device is compared to the one after stress. Under equivalent conditions, the pMOS degrades more than the nMOS.

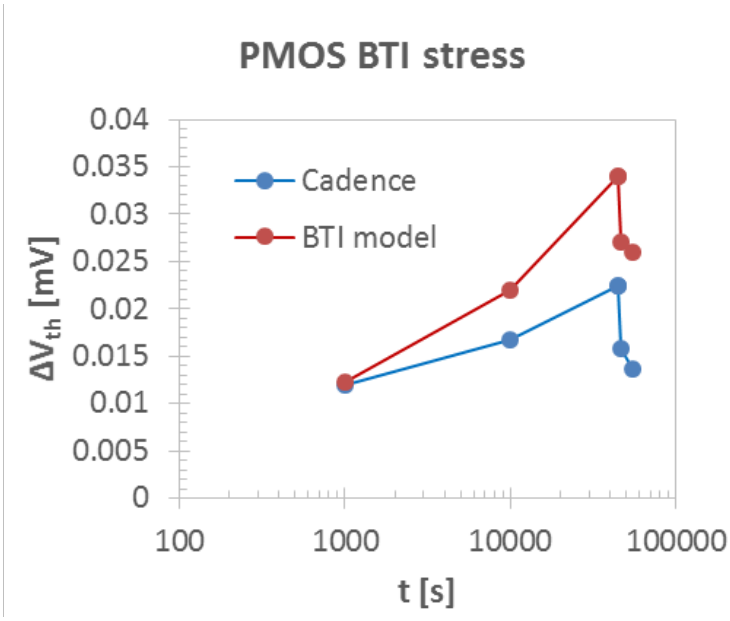


Figure 8 - Simulated ΔVth induced by NBTI degradation compared to measurements The degradation models (for NBTI, PBTI and HC) developed by KU Leuven in this project and described here, have been incorporated into RandomSpice, in collaboration

with GSS. This was needed since only the degradation for static stress can be computed at the time being. As the degradation after dynamic stress requires signal analysis, more effort is needed to embed it. Nevertheless, the developed degradation models have been proven to be compatible with RandomSpice, which makes the computation of time-zero and time-dependant variability possible for both single devices and circuits.