Quantum Mechanical Charge Trap Modeling to Explain BTI at Cryogenic Temperatures

J. Michl†, A. Grill∗, D. Claes†, G. Rzepa†, B. Kaczer†, D. Linten†, I. Radu†, T. Grasser†, and M. Waltl†

†Christian Doppler Laboratory for Single-Defect Spectroscopy in Semiconductor Devices at the
‡Institute for Microelectronics, TU Wien, Gusshausstraße 27-29, 1040 Vienna, Austria
∗imec, Kapeldreef 75, 3001 Leuven, Belgium
⋄KU Leuven, Dep. of Electrical Engineering (ESAT), MICAS, Kasteelpark Arenberg 10, 3000 Leuven, Belgium
•Global TCAD Solutions GmbH, Bösendorferstraße 1/12, 1010 Vienna, Austria

Abstract—Electronics operating at cryogenic temperatures is crucial for scaling up single qubits to complex quantum computing systems. There are various studies concentrating on the characterization of advanced CMOS technologies operating at low temperatures, but so far little attention has been paid to reliability issues. Even though classical models predict BTI to freeze out, our measurements clearly reveal a significant threshold voltage degradation down to 4 K. This effect can be consistently explained by considering a quantum mechanical extension for the description of charge transitions in the transistor, which leads to an effective barrier lowering towards cryogenic temperatures. We implement this model in our reliability simulator Comphy and are finally able to fully explain BTI behaviour at temperatures down to 4 K.

Index Terms—28 nm bulk CMOS, Bias temperature instability (BTI), cryoelectronics, cryogenic, 4 K, MOS transistor modeling, Quantum mechanical extension

I. INTRODUCTION

Operating CMOS technologies at cryogenic temperatures is important in various fields such as astronomy [1], high performance computing, space exploration and quantum computing [2]. For the realization of the necessary classical interface for operating qubits of state-of-the-art quantum computing systems, several new devices as near-quantum-limited amplifiers [3], ultralow-loss resonators [4] or circulators [5] have been developed. This hardware is typically separated from the quantum mechanical layer and runs at room temperature [6]. One of the next key steps for up-scaling quantum computers is to integrate the classical interface in the refrigerator together with qubits to reduce signal distortion and thermal-noise [7]. For solid-state qubits it could even be possible to integrate the control system on the same chip as the qubits [8] which would reduce wiring challenges dramatically. These up-scaled systems could then provide an exponential speed-up over conventional computing systems for certain complexity classes [9]. Since for advanced CMOS technologies operating at cryogenic temperatures outstanding characteristics have been reported in terms of on-state current, leakage-current, subthreshold swing, or transconductance [10], reliability studies are highly relevant for understanding the degradation mechanisms down to 4 K.

II. DEVICES AND MEASUREMENTS

In our study we use commercial 28 nm high-k nMOSFETs with dimensions \( W \times L = 1 \mu m \times 0.5 \mu m \) (Fig. 1 (top)). Using liquid helium for cooling, we measure basic DC characteristics between 4 K and 300 K. Figure 2(a) shows the \( I_D(V_G) \) curves as a function of temperature between 0 V and 1 V applying \( V_{DS} = 25 \) mV. The subthreshold slope (Fig. 2(b)) shows the characteristic saturation below approx-
immediately 100 K, instead of following the Boltzmann thermal limit $SS = n k_B T / q \ln (10)$ to 0 meV/dec when $T \to 0$. K. This saturation can be seen across various technologies and is caused by band-tail tunneling of interface defects [12]. With lower temperatures the transconductance (Fig. 2(c)) improves, whereas the threshold voltage $V_{th}$ (Fig. 2(d)), which was determined using the maximum transconductance method, shifts about 150 mV towards higher values due to the shifting Fermi-level in the bulk and the temperature dependent Fermi-Dirac distribution [10].

For measuring the temperature dependence of the threshold voltage degradation $\Delta V_{th}$, we applied the extended Measure-Stress-Measure (eMSM) scheme illustrated in Figure 1 (bottom) [11]. After recording an initial $I_D(V_G)$, subsequent stress ($V_{GS} = V_{Ga}$) and recovery ($V_{GS} = V_{Gr}$) phases with increasing stress times $t_s$ and recovery times $t_r$ are measured. During the recovery phase, the source current is recorded and can be mapped to $V_{th}$ using the initial $I_D(V_G)$. Figure 3 shows the degradation after 1 ms for different stress times and temperatures, with $V_{Ga} = 2$ V and $V_{DS} = 25$ mV. Although $\Delta V_{th}$ decreases for lower temperatures, a significant threshold voltage shift can still be measured even at $T = 4$ K.

Figure 2. DC characteristics of nMOSFET transistors in the temperature range of 4 K to 300 K: (a) Source current at different temperatures. (b) The subthreshold slope is found to saturate instead of following the Boltzmann thermal limit to 0 meV/dec at $T = 0$. K using $n = 7/6$. This saturation is caused by band-tail tunneling [12]. (c) Transconductance increases towards low temperatures, due to a higher mobility of the charge carriers, i.e. fewer scattering events. (d) Temperature lowering down to $T = 4$ K increases the threshold voltage by around $150$ mV, because the temperature dependent Fermi-level in the bulk is shifted [10]. Note that the shape of $V_{th}(T)$ depends on the selected extraction method (with ‘gm’ being extraction of $V_{th}$ from transconductance and ‘cc’ using a constant current criteria).

Figure 3. $\Delta V_{th}$ measured with a delay of 1 ms at different temperatures and stress times $t_s$ for commercial 28 nm nMOSFETs. BTI degradation is lower towards 4 K (stars), but does not freeze out completely as might be expected from simple classical charge transfer theory.

III. RESULTS AND DISCUSSION

The compact physics simulator Comphy has been demonstrated to explain BTI behaviour of 28 nm nMOSFETs with different gate stacks operating above room temperature using an effective two-state non-radiative multiphonon (NMP) model to describe charge trapping and a double well model for interface state generation [13]. In the two-state NMP model, a defect can be either neutral or charged (Fig. 4) and a (classical) transition between those states is possible by overcoming the energy barriers $\varepsilon_{12}$ and $\varepsilon_{21}$. The corresponding transition rates are proportional to $\exp (-\varepsilon_{12} / k_B T)$ or $\exp (-\varepsilon_{21} / k_B T)$ [14], [15]. Applying a harmonic and adiabatic approximation for the potential energy surfaces of the defect, in the classical limit the energy barrier can be computed using the Huang-Rhys parameters $R$, $S$ and $E_T$ from the intersection of the parabolas [16]. As a consequence, however, for $T \to 0$ K charge transitions $\exp (-\varepsilon / k_B T) \to 0$ would be predicted to freeze out.

The consequence of this freeze-out can be seen in the comparison of Figures 5 and 6. The circles in Figure 5 represent eMSM measurements at 300 K for a gate stress of $V_{Ga,s} = 2$ V in the same 28 nm nMOSFET technology used in Figure 2. After optimizing the input parameters, Comphy can simulate these measurements by sampling defects as indicated in the energy diagram of Figure 5. Figure 6, on the other hand, shows eMSM measurements obtained under the same bias and timing conditions but at $T = 77$ K. For the longest stress time of $t_s = 1$ ks a degradation of nearly $\Delta V_{th} = 15$ mV is observed. However, using the same optimized set of parameters...
obtained at \( T = 300 \) K, the classical approximation severely underestimates this degradation.

To circumvent this weakness of the classical NMP model at low temperatures, a quantum mechanical extension as delineated in the configuration diagram in Figure 7 is incorporated in this work. By approximating the potential energy surfaces by parabolas, it is possible to solve the eigenproblem of the resulting quantum harmonic oscillator for the corresponding vibrational wavefunctions [17]. The vibrational wavefunctions overlap at energies below the classical transition point and allow charges to tunnel from one state to the other. The overlap \( I_{i\alpha j\beta} \) of two vibrational wavefunctions \( \Omega_{i\alpha}(q) \) and \( \Omega_{j\beta}(q) \) can be computed by the inner product

\[
|I_{i\alpha j\beta}|^2 = \int_{-\infty}^{\infty} \Omega_{i\alpha}(q) \Omega_{j\beta}(q) \mathrm{d}q,
\]

where \( i, j \in \{1, 2\} \), \( \alpha \) and \( \beta \) represent the levels of the vibrational wavefunction. The overlap integral has to be weighted

---

**Figure 4.** Classical charge transitions rely on the intersection point in the configuration coordinate diagram showing the potential energy surfaces of two states. Within the NMP model charge transitions are proportional to \( \exp(\frac{-\varepsilon_{12}}{k_B T}) \) and \( \exp(\frac{-\varepsilon_{21}}{k_B T}) \), which would lead to a freeze-out for \( T \to 0 \) K.

**Figure 5.** Measured degradation of \( \Delta V_{th} \) for different stress times at 300 K recorded during recovery (markers). Using our compact physics simulator Comphy we can explain the measurements using a two-state NMP model to describe charge trapping of oxide defects (lines), which are considered to be distributed as indicated in the inset. At these relatively low temperatures, the creation of interface states described by the double-well model is negligible and thus ignored.

**Figure 6.** cMSM measurements at 77 K (markers) and classical simulation (lines) using the same set of parameters as for the simulation at \( T = 300 \) K (cf. Fig. 5). The simulation significantly underestimates the observed degradation due to the freeze-out of the transition rates.

---

**Table I**

<table>
<thead>
<tr>
<th>Extracted defect band parameters for shallow SiO(_2) and deep SiO(_2) (top), shallow HfO(_2) and deep HfO(_2) (bottom)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defect band parameters</td>
</tr>
<tr>
<td>Shallow SiO(_2)</td>
</tr>
<tr>
<td>Deep SiO(_2)</td>
</tr>
<tr>
<td>Unit</td>
</tr>
<tr>
<td>( E_T \pm \sigma_{E_T} )</td>
</tr>
<tr>
<td>( S \pm \sigma_S )</td>
</tr>
<tr>
<td>( R )</td>
</tr>
<tr>
<td>( N_T )</td>
</tr>
<tr>
<td>Deep HfO(_2)</td>
</tr>
<tr>
<td>Shallow HfO(_2)</td>
</tr>
<tr>
<td>Unit</td>
</tr>
<tr>
<td>( E_T \pm \sigma_{E_T} )</td>
</tr>
<tr>
<td>( S \pm \sigma_S )</td>
</tr>
<tr>
<td>( R )</td>
</tr>
<tr>
<td>( N_T )</td>
</tr>
</tbody>
</table>
with the occupancy $p_{i\alpha}$ of an excited state

$$p_{i\alpha} = \frac{\exp \left( -\frac{E_{i\alpha}}{k_B T} \right)}{\sum_{\gamma} \exp \left( -\frac{E_{i\gamma}}{k_B T} \right)}, \quad (2)$$

where $E_{i\alpha}$ is the corresponding eigenenergy of state $\Omega_{i\alpha}$, to get to an expression for the lineshape function

$$\xi_{ij} = \sum_{\alpha,\beta} |I_{ij\alpha\beta}|^2 p_{i\alpha} \delta(E_{i\alpha} - E_{j\beta}). \quad (3)$$

To account for lifetime broadening, the Dirac-delta function $\delta(E_{i\alpha} - E_{j\beta})$, which restricts the wavefunction overlaps to overlaps at the same energy level, has to be broadened to a Gaussian function

$$g(\Delta U) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left( -\frac{\Delta U^2}{2\sigma^2} \right), \quad (4)$$

where $\sigma$ depends on the lifetime of the defect. Since $\Delta U$ depends on the environment and is very complex to determine, it was chosen to be equal to the separation of the eigenenergies of the potential energy surface [17]. Using the relation

$$\theta_{ij} = \frac{2\pi}{\hbar} |\theta_{ij}|^2 \xi_{ij}, \quad (5)$$

the transition rates can be computed using the lineshape function. $\theta_{ij}$ can be approximated with the WKB method [18]:

$$\theta_{ij} \approx \tilde{k} \exp \left( -\frac{\hbar}{4m} \frac{1}{x_2 - x_1} \int_{x_1}^{x_2} (U(x) - E_0) \, dx \right), \quad (6)$$

where $m_t$ is the effective tunneling mass, $U(x)$ the barrier, $E_0$ the energy of the carrier and $x_2 - x_1$ the width of the barrier. The parameter $\tilde{k}$ needs to be calibrated to experimental data.

In comparison to the classical approximation of the two-state model, where charge transfer freezes out at cryogenic temperatures, the full quantum mechanical model still allows nuclear tunneling from one state to the other, even if the classical energy barrier could not be overcome. Since the calculation of $I_{ij\alpha\beta}$ is computationally expensive, a more efficient approximation was used in this work, which is called Saddlepoint method and was developed by Holstein [19]. In this approach, the transitions between the states are treated using a WKB method. In particular, temperature dependent transition rates can be computed by finding $E^*$

$$\frac{d\varphi(E)}{dE} \bigg|_{E=E^*} = \beta, \quad (7)$$

where

$$\varphi(E) = \frac{\sqrt{8m}}{\hbar} \int_{q_c(E)}^{q_e} \sqrt{U_2(q) - E} \, dq - \int_{q_a(E)}^{q_c} \sqrt{U_1(q) - E} \, dq. \quad (8)$$

Here $\beta = 1/k_B T$, $m$ is the effective mass of the oscillator, $q$ is the configuration coordinate and $U_1(q)$ and $U_2(q)$ are the potential energy curves of the states of the defect [20]. $q_c$ is the configuration coordinate of the classical intersection point of the energy surfaces and $q_a(E)$ and $q_b(E)$ are the configuration coordinates corresponding to energy $E$, which can be in the interval between $\max(E_{1\text{min}}, E_{2\text{min}})$ and the
energy of the intersection point of the two energy surfaces (Fig. 4). Subsequently the lineshape function can then be computed by

\[ \xi_{ij} = e^{\psi(\beta)}, \quad (9) \]

where

\[ \psi(\beta) = \varphi(E^*(\beta)) - \beta E^*(\beta). \quad (10) \]

Figure 8 shows the agreement of this approximation with the full quantum mechanical solution for different displacements of the potential energy surfaces \( \Delta q \). For the computation of the dashed lines, representing the lineshape function computed with the Saddlepoint method, an effective displacement \( \Delta q_{\text{eff}} = \Delta q - 0.8 \sqrt{\Delta} \) was used. Since \( \Delta q \) is a fitting parameter only occurring in the computation of the transition rates, the shift to an effective displacement has no impact on the electrostatics.

For the quantum mechanical treatment, the limits of the integrals in equation (8) depend on the parameters \( q_{\text{on}}(E) \) and \( q_{\text{off}}(E) \). These values depend on the parameter \( \Delta q \), which describes the displacement between the potential energy surfaces. Note that this parameter is not necessary in the classical picture, where this transition only depends on the height of the barrier. Using this model, an effective barrier lowering can be calculated (Fig. 9) by approximating the lineshape function with an Arrhenius law for different displacements \( \Delta q \). While in the classical limit only a slight temperature dependence of the effective energy barrier is observed, the quantum mechanical model shows a tremendous barrier lowering towards \( T = 0 \) K, which becomes more pronounced for small displacements \( \Delta q \). This barrier lowering \( E_B \rightarrow 0 \) eV when \( T \rightarrow 0 \) K explains why \( \exp(E_B/k_B T) \) does not freeze out and charge transitions still occur at low temperatures.

By implementing this model into Comphy, we are able to correctly simulate \( \Delta V_{\text{th}} \) degradation from 300 K down to 4 K by using the very same set of trap parameters as used for the classical simulations (see Figure 10). Using the Nelder-Mead method the parameter \( \Delta q = 5.40 \sqrt{\Delta} \) has been found, while the parameters of Table III have been fixed. In comparison to the classical model, \( \Delta V_{\text{th}} \) at low temperatures can be reproduced nicely and does not freeze out completely. It is noticeable that the quantum mechanical correction shows already an effective barrier lowering at 300 K (Fig. 9), therefore \( \Delta V_{\text{th}} \) during stress and recovery is slightly overestimated. This could be compensated by reoptimizing the Huang-Rhys parameters, which was not done here for the sake of a better comparability. For temperatures below 25 K, numerical problems in the electrostatics were observed, which is a well known issue in simulations at cryogenic temperatures [10], [21]. A commonly used way to avoid this is the introduction of additional constraints, which serve as lower boundaries in the temperature regime to avoid division by zero errors [22]. Using a temperature constraint of 25 K for the computation of the charge carrier density we have been able to correctly compute the threshold voltage shift down to 4 K.

### IV. Conclusions

After characterizing commercial 28 nm nMOSFETs between 4 K and 300 K, we measure their BTI behaviour using an eMSM scheme. The results show a degradation of the threshold voltage, which does not freeze out at cryogenic temperatures. This is in contradiction with classical charge transition models, where a charge transition takes place by overcoming the energy barrier between the charge states. In a quantum mechanical picture the vibrational wavefunctions of the two charge states, which can be interpreted as potential energy surfaces, overlap below the classical energy barrier. This leads to an effective barrier lowering and delivers considerably higher transition rates, especially for low temperatures. This quantum mechanical extension was implemented in the compact physics simulator Comphy, which employs an effective two-state NMP model to describe charge transitions. With this improved model we can consistently explain measurements between 4 K and room temperature and understand why BTI is not completely frozen out at cryogenic temperatures.

### References

Figure 10. Comphy simulations compared to the 300 K, 150 K, 77 K and 4 K measurements using a quantum mechanical extension (QME) for the computation of the charge transition rates of the defects while employing the same set of Comphy parameters as used in Figure 5. The introduction of tunneling of the nuclei at low temperatures clearly improves the result below 300 K (compare Fig. 6). At 300 K, both the stress and the relaxation rates, are slightly overestimated, which results from effective barrier lowering also occurring at room temperature (Fig. 9). However, the enhancement of our trapping model significantly extends its accuracy and enables the description of charge trapping over a broad temperature range from 4 K to 600 K. 


