Reliability and Variability of Advanced CMOS Devices at Cryogenic Temperatures

A. Grill†, E. Bury†, J. Michl*, S. Tyaginov†, D. Linten†, T. Grasseral*, B. Parvais†, B. Kaczer†, M. Waltl*, I. Radu† and I. I. NTRODUCTION

Abstract—In this work, we present time-zero variability and degradation data obtained from a large set of on-chip devices in specifically designed arrays, from room temperature to 4 K. We show that the investigated nMOS transistors still suffer from significant PBTI and HC degradation down to the lowest temperatures. We further investigate the contribution of multiple-carrier mechanism versus single-carrier mechanism of Si-H bond dissociation across different temperatures. Finally, we extrapolate the time-to-failure for a large gate and drain bias space and show that HCD after on-state stress and off-state stress show opposite temperature trends with the off-state stress being worse at cryogenic temperatures.

Index Terms—Bias Temperature Instability (BTI), Hot Carrier Degradation (HCD), Degradation Maps, Variability, Smart Arrays, Cryoelectronics, Cryogenic, 4 K, 28 nm bulk CMOS

I. INTRODUCTION

Apart from its well-known applications in space industry, astronomy, and high-precision metrology [1], advanced CMOS technology operated at cryogenic temperatures is one of the next key steps to enable large-scale quantum computing [2]–[4] and to improve the computational performance of data centers. While the latter application is likely mostly limited to a temperature range of 77 K (LN₂), the bulk of integrated qubit control systems will be operated at liquid helium temperatures (4 K) (LN₂, RF oscillators, etc.) or even down to the mK regime depending on power and noise constraints of the specific qubit technology. The close integration of classical CMOS logic and qubits thereby not only helps to mitigate wiring constraints but also reduces signal distortions during read and write operations.

Recent publications on advanced CMOS technologies mostly focus on the improved device characteristics (sub-threshold swing, on-state current, leakage, etc.) at cryogenic temperatures [5]–[7]. Because of measurement constraints like the number of probes available in cryostats (usually up to 9), these studies are often carried out on single devices. This neglects the effects of temperature-dependent variability, which is especially pronounced in nanoscaled devices [8], [9].

In this work we analyze variability and reliability of nMOS transistors in a wide range of bias conditions from room temperature down to 4 K. The data is based on a large statistical set — 2 560 devices per temperature — obtained from dedicated array structures, which were custom designed for this task.

II. EXPERIMENTAL

The arrays used in this work consist of 12 blocks, each of them containing 2 560 individually selectable, minimally-spaced nanoscale nMOS devices (WxL = 100 nm × 28 nm) - adding up to 30 720 devices per chip. One block is divided into 256 programmable gate lines, each connected to ten devices. Each block has 256 digitally selectable gate lines with shift registers and latches, where each line addresses 10 devices in parallel with separate drain lines. The array was bonded to a PCB and measured using a custom designed defect probing instrument on a Lakeshore CPX-LVT setup equipped with a 48-line break-out box.

This work is funded in part by imec’s Industrial Affiliation Program on Quantum Computing and Cryoelectronics and the European Union’s Horizon 2020 program through the Marie Skłodowska-Curie Grant 794950. Furthermore, research leading to this work has been funded the Take-off program of the Austrian Research Promotion Agency FFG (projects no. 861022 and 867414) and the Austrian Science Fund (FWF) under Grant P31204-N30.
The chip is bonded to a custom designed PCB and connected to a 48-line break-out box on a modified Lakeshore CPX-LVT cryostat. The measurements are performed on specifically designed hardware [12]. One of the main benefits of using this system is that it not only allows to program and measure the arrays but also to change the sample temperature, which enables semi-automated measurement sequences.

### III. RESULTS AND DISCUSSION

For each temperature, a set of 2,560 devices is selected and an initial characterization is performed by recording $I_dV_g$ characteristics in the linear and saturation regimes in forward and reverse direction. The results of these measurements for the linear regime as well as the median of all devices can be seen in Fig. 2. Note that for the purpose of visualization, only a subset of all traces is plotted here.

Especially at low temperatures, additional kinks located in the subthreshold region of the device start to appear. Previous studies attributed this behavior to Coulomb blockades due to disorder at the channel interface or (partial) freeze-out of parasitic edge-transistors resulting in a double-threshold voltage ($V_{th}$) behavior [7], [13]. The outliers at 300 K can likely be attributed to extrinsic leakage stemming from damaged devices on one of the drain lines of this specific block of the array.

The time-zero distributions in Fig. 3 show that a significant temperature dependence of the variability can only be seen for subthreshold slope mobility ($SS$) and $g_m$ distributions are affected by the additional features present only at cryo-temperatures as well (see $g_m$ plot in Fig. 2).

The statistical moments of the distributions can be seen in Fig. 4. Both, mean mobility and $V_{th}$ increase with lower temperatures, but also show clear signs of saturation. For $V_{th}$, this can be explained with the Fermi-level scanning additional states close to the conduction band edge (e.g. from doping or band-tail states) and the temperature dependence of the Fermi-Dirac distribution and the silicon band gap. For the mobility, on the other hand, other scattering mechanisms like Coulomb and surface scattering become more and more dominant compared to phonon scattering.
After the initial characterization of the devices, each gate line (i.e., 10 parallel devices) is stressed with a certain stress condition \( \{V_{gs}, V_{ds}\} \) for stress times of 3 s, 10 s, 30 s, and 100 s. After stress, the same set of \( I_d V_T \) traces is recorded again to be able to track the degradation of the relevant transistor parameters. The results at each bias condition are then collected and plotted in degradation maps. Such maps are shown in Fig. 5 for \( \Delta V_{th} \) and \( \Delta I_{d, \text{sat}} \) at three different temperatures.

For pure BTI conditions (high \( V_{gs}, V_{ds} \approx 0 \) V), the degradation slightly decreases at lower temperatures, but much less than expected from a typical temperature activation described by a simple Arrhenius law. This effect can be explained by a lower effective transition barrier caused by nuclear tunneling in the framework of the NMP model [17], [18].

The opposite trend holds true for HCD. While the off-state stress seems to be largely unaffected by temperature, on-state stress clearly increases at lower temperatures. This is consistent with our previous observation that the temperature behavior of hot-carrier degradation changes when device dimensions shrink [19].

On the one hand, HCD has been reported to be weaker at elevated temperatures in long-channel devices [16], [20], [21]. The reason behind this behavior is that increased scattering at elevated temperatures helps to depopulate the high-energy fraction of the carrier ensemble and therefore scattering-out of hot carriers becomes more efficient if a transistor is heated, thereby suppressing HCD.

For ultra-scaled transistors, on the other hand, HCD has been shown to be accelerated at higher temperatures [22]–[24]. It was suggested that the mechanism responsible for this trend is increased electron-electron scattering which populates the high-energy part of the carrier spectrum [15]. The same paper discusses the transition between these two opposite trends occurring at channel lengths of 70-100 nm.

A closer look at the degradation data in Fig. 5 reveals an even more complicated behavior. One can see that the threshold voltage shift \( \Delta V_{th} \) acquired at \( V_{gs} = V_{ds} = 1.2 \) V (magenta) decreases (i.e., more blue) with temperature, with the same trend holding true also for the linear drain current change \( \Delta I_{d, \text{lin}} \). As for the \( V_{th} \) and \( I_{d, \text{lin}} \) values obtained at higher drain voltages, e.g., at \( V_{ds} = 1.8 \) V and \( V_{th} = 1.2 \) V (white), the trend is the opposite, i.e., the degradation becomes more pronounced (i.e., more red) at higher temperatures.

It should be emphasized that according to Rauch et al. [15], HCD should become more severe for all hot-carrier stress conditions for a channel length of 28 nm. Our experimental data is however consistent with our modeling results presented in [25] for an nMOS with a channel length of \( \sim 44 \) nm. Among other things, we showed that the rate of the multiple-carrier mechanism of Si-H bond dissociation becomes lower at higher temperatures. As a result, in regimes when hot-carrier degradation is dominated by this mechanism, the degradation becomes weaker at higher temperature.

Vice versa, for the single-carrier mechanism, the bond rupture rates increase with temperature. As a consequence, the relative contribution of the single-carrier mechanism becomes more prominent at higher \( V_{ds} \). For a certain drain bias \( V_{ds} \), this mechanism starts to dominate, leading to an increased HCD with temperature. Both of the previously mentioned trends can be seen in Fig. 5. This confirms the soundness of our HCD model proposed in [25], [26].

The voltage and current degradations after 143 s accumulated stress time for different bias conditions are depicted in Fig. 6. For pure BTI conditions, the voltage acceleration approximately follows a power-law, with the slope getting slightly steeper for low temperatures (top row).

Hot-carrier degradation is caused by the dissociation of Si-H bonds at the Si/SiO\(_2\) interface [19]. This can be either by high-energy carriers which trigger the single-carrier bond breakage mechanism or be a series of “cold” carriers which contribute to the multiple-carrier mechanism. The multiple- and single-carrier mechanism driven parts of HCD can be distinguished by plotting TTF \( \times I_{d, \text{stres}} \) over \( I_{d, \text{stres}} \) [27].

For medium values of \( V_{ds} \), hot carrier degradation is dominated by the multiple-carrier process of bond rupture. At higher values of \( V_{ds} \), especially for low values of \( V_{gs} \), the TTF clearly deviates from this trend and degradation seems to be governed by the single-carrier mechanism. This effect is also more pronounced at cryogenic temperatures. In this regime, phonon scattering vanishes and carriers are therefore able to gain more energy from the drain field on average.

The localization of the defects generated by HCD can be estimated by plotting the relative error between the degradations recorded in forward and backward direction [11]. If the drain side of the device is stressed (i.e. in forward direction), the defects are mainly generated in the pinch-off region close to the
Figure 5. Threshold voltage drift and drain current degradation as a function of gate and drain bias at different temperatures. Each dot in the maps represents the data extracted from 50 devices, which roughly is equivalent to a 2-sigma interval. Note that hot-carrier degradation increases at lower temperatures likely due to reduced phonon scattering [14]–[16]. The cumulative impact of different degradation mechanisms — BTI, current-, and energy-driven degradation - on the observed data and their de-convolution is shown in [11].

drain contact. During a reverse sweep performed immediately afterwards on the same device, many of those defects will be screened by channel carriers. Therefore, a clear asymmetry can be seen between these two measurements (see Fig. 7). The effect again becomes stronger at low temperatures, supporting the picture of more high-energetic carriers causing additional damage at the drain side of the device. In other words, HCD becomes more strongly localized at lower temperatures.

To extrapolate the time to failure for different temperatures, a simple power-law acceleration across time is used. The exponents and pre-factors are chosen to be independent for each stress condition. The benefit of this approach is that it only relies on measurement data. Therefore, no assumptions on the different physical degradation mechanisms and their voltage and temperature acceleration are required (which probably are not accurately covered by available models at cryo-temperatures anyways). This, however, comes at the expense of the range of extrapolated stress times limited mainly by the noise floor of the data. The results of a time-to-failure extraction based on this method can be seen in Fig. 8. Most interestingly, the temperature dependence between on-state stress and off-state stress is the opposite. While for high-current conditions, lifetime gets worse with lower temperatures, the off-state degradation slightly improves on average.

IV. CONCLUSIONS

We have extracted the time-zero distributions of a large statistical dataset of nMOS transistors at different temperatures, ranging from 300 K to 4 K employing our dedicated Smart Arrays. The device mobility and sub-threshold slope show a significantly larger variability at lower temperatures because of additional mechanisms like Coulomb blockades or parasitic edge transistors appearing at low temperatures.

We recorded degradation maps for 82 sets of bias conditions with a sample size of 50 different devices per stress case. The results indicate increased HCD as well as a significant BTI degradation still being present at lower temperatures. Hot carrier induced damage at the drain side of the devices is increased at cryogenic temperatures as carriers get more energetic due to reduced phonon scattering, which is inconsistent with previous reports on short-channel devices.

Finally, time to failure plots have been extracted from the stress time dependence. They indicate opposite trends between off-state and on-state stress, with the time to failure getting worse for on-state stress. Most likely this is due to reduced phonon scattering and therefore more energetic carriers on average at cryogenic temperatures.
Figure 6. Time to failure (TTF) times stress current plotted against the stress current (top row) after 143 s of cumulative stress. For mid-$V_{ds}$ and high-$V_{gs}$ conditions, the degradation is current driven, while for high-$V_{gs}$ conditions under off-state stress, the degradation is more energy driven. This trend is even more pronounced at lower temperatures. For BTI conditions, the degradation follows an approximate power-law voltage acceleration across all temperatures (bottom row).

Figure 7. The lateral location of defects generated by hot carrier stress can be estimated by measuring the forward and reverse currents of the device [27]. Defects generated in the pinch-off region during forward stress are screened by the channel during reverse stress. In general, a higher damage localization can be observed at lower temperatures.

Figure 8. Power-law time exponents extracted for $\Delta V_{th}$ and $\Delta I_{d,sat}$ at 4K (top) and the corresponding time-to-failure extracted from the measurements (bottom). The overall trend of an increased hot-carrier degradation at lower temperatures especially for on-state stress is confirmed for both, voltage and current criteria. Note that the lack of degradation at lower voltages prohibits safe-operating area projection for longer TTF.

REFERENCES


