

"Interface Engineering for Steep Slope Cryogenic MOSFETs"

B. Richstein, Y. Han, Q. Zhao, L. Hellmich, J. Klos, S. Scholz, L. Schreiber and J. Knoch

Abstract—Until today, MOSFETs show inverse subthreshold slope SS values that stay well above the theoretical Boltzmann limit of 0.8 mV/dec at 4.2 K. Recently, the saturation of SS has been attributed to band tailing which is described by a density of states (DOS) at the band edges which decays exponentially into the band gap. While band tailing is introduced by mobile states, localized states degrade SS as a function of the gate voltage close to the transition region around the threshold voltage which is called inflection. Both phenomena are observed independent of fabrication technology and imply the necessity of dedicated cryogenic CMOS devices. While the potential of mobile and localized states can be screened by a sufficient oxide scaling to a certain degree band tailing still contribute to diffusion current which finally saturates SS . In this work, we show gate stack and interface engineering as an approach to counteract band tailing and inflection which facilitate the realization of steep slope cryogenic MOSFETs. Fully-depleted Silicon-on-Insulator (FD-SOI) devices with modified gate stacks were fabricated and characterized at cryogenic temperatures.

Index Terms—Band tailing, interface states, interface engineering, inflection, cryogenic MOSFET, silicon nitride.

I. INTRODUCTION

The co-integration of classical cryogenic control electronics (cCMOS) with silicon-based spin qubits holds promise for paving the way towards up-scaling of quantum information processors (QIP) [1]–[4]. However, the limited cooling power of cryostats requires the cCMOS to be operated in the few millivolt regime. This in turn requires very steep inverse subthreshold slopes SS to properly switch the devices. Although a very steep SS is expected from classical MOSFET theory, when cooling down MOSFETs designed for room-temperature (RT) operation, SS saturates well above the theoretical Boltzmann limit. For devices fabricated on bulk-Si

SS ranges between 10 mV/dec and 30 mV/dec [5]–[9]. Fully depleted (FD) Silicon-on-Insulator (SOI) devices which benefit from a smaller depletion capacitance C_{depl} show smaller SS values. However, reported SS values of Fin-FETs [10] and FD-MOSFETs [11] stay above 7 mV/dec at 4.2 K, which is approximately ten times higher than the theoretical limit. The reason for the limited SS is band tailing due to electron-electron interactions, electron-phonon scattering and static disorder (i.e. surface roughness, defects) at the MOS interface leading to a density of mobile states (DOS_m) at the band edges that decays exponentially into the band gap (as illustrated in Fig. 1a), red line):

$$\text{DOS}_m \propto \exp\left(\frac{E - E_C}{\Delta E_C}\right) \quad (1)$$

where ΔE_C is the band tail extension into the band gap around the conduction band E_C . Note that the same applies to the valence band ($\Delta E_C = \Delta E_V$, $E_C = E_V$).

While at RT the convolution of the exponential decay of the Boltzmann tail of the Fermi function $f(E)$ with a step-function-like DOS limits SS , at cryogenic temperatures $f(E)$ becomes a step function (Fig. 1a), purple line) but band tailing limits SS . This in turn leads to a saturation of SS when T drops below an effective temperature $T^* = \Delta E_C/k_B$ [8], [11], [12]. In addition, Gaussian distributed, localized interface states around the band edges result in a V_{gs} -dependent capacitance $C_{\text{it}}^{\text{localized}}$ that significantly increases SS in the transition region around to the threshold voltage V_{th} which is called inflection [9], [12], [13]. As a result of band-tailing and inflection, SS can be written at cryogenic temperatures as:

$$SS = \frac{k_B T^*}{e} \ln(10) \frac{\partial \Phi_g}{\partial \Phi_0} = \frac{k_B T^*}{e} \ln(10) \left(1 + \frac{C_{\text{it}}}{C_{\text{ox}}}\right) \quad (2)$$

where Φ_0 is the potential maximum in the channel, $\Phi_g = -eV_{\text{gs}}$, $k_B T^* = \Delta E$ is the effective temperature due to the exponential decay of band tails, $C_{\text{it}} = C_{\text{it}}^{\text{mobile}} + C_{\text{it}}^{\text{localized}}$ and C_{ox} is the geometrical oxide capacitance. Note that the depletion capacitance is neglected since in this work FD-MOSFETs are characterized.

Because SS depends on V_{gs} (due to the dependence of C_{it} on the gate voltage) plotting SS as a function of I_D facilitates a straightforward comparison of the impact of, e.g., different gate stacks on the switching behavior of different cCMOS devices. Figure 1b) shows schematically the impact of band tails and inflection where i) a step-function like DOS (black curve), ii) the exponentially decaying DOS of Eqn. (1) (red

This paragraph of the first footnote will contain the date on which you submitted your paper for review. It will also contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the Deutsche Forschungsgemeinschaft under Grants KN545/29 and KN545/24."

The next few paragraphs should contain the authors' current affiliations, including current address and e-mail. For example, F. A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: author@boulder.nist.gov).

S. B. Author, Jr., was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu).

T. C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp).

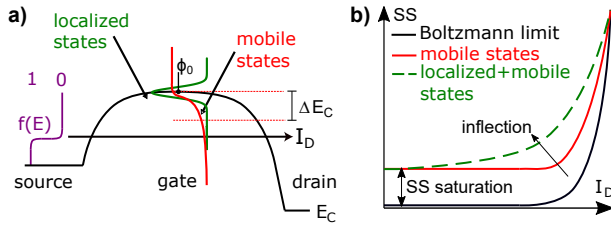


Fig. 1. (a) Schematic depiction of a MOSFET band structure at cryogenic temperatures considering a step-function like $f(E)$ (purple), an exponential decay of a DOS introduced by mobile states (red) and localized states (green). (b) SS behavior of a MOSFET at cryogenic temperatures affected by band tailing and inflection. (cf. [13]).

curve) and iii) the combined effect of band tails and localized states (green curve) are displayed.

While the impact of mobile and localized states (i.e. C_{it}) can be screened with sufficiently large C_{ox} resulting in $\partial\Phi_g/\partial\Phi_0 = 1$ (cf. Eqn. (2)), ΔE_C at the MOS interface is only moderately reduced [14]. The latter requires simultaneous oxide scaling and engineering of the MOS interface. In this work, we show experimentally that gate stack and interface engineering facilitates the realization of steep slope cryogenic MOSFETs. This is achieved by including thin interlayers between the gate oxide and the silicon surface. The growth of SiN layers with a temperature-dependent, self-limiting process has the advantage of a very high tunability and reproducibility of the silicon nitride layer thickness between 0.8 nm and 2 nm. A very important feature of the used SiN is that it leads to a larger density of interface states D_{it} compared to thermally grown SiO₂ [15]. This D_{it} lies within the Si band gap and would deteriorate the switching of room temperature transistors. However, these states are in an energy range that is irrelevant for cCMOS. On the contrary, the ability of SiN to prevent oxygen diffusion during the deposition of high-k dielectrics and hence avoid the oxidation of Si [16], [17] yields smooth interfaces with reduced disorder. This, in turn, appears effective in decreasing band tailing as capacitance voltage measurements of MOS-caps suggest [14].

II. DEVICE FABRICATION

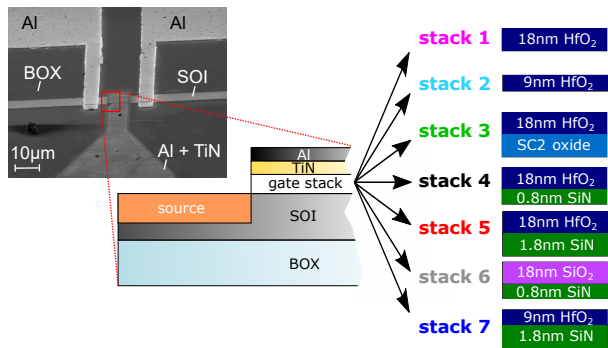


Fig. 2. SEM picture of the fabricated final device and the schematic assembly of the seven fabricated gate stacks.

In order to investigate the impact of gate stack engineering on the SS saturation and inflection (i.e. reduce T^* and C_{it} , Eqn. (1) and (2)), n -type FD-SOI MOSFETs with 50 nm Si

(p -type 5–10 Ω cm) and 100 nm BOX were fabricated (cf. Fig. 2). The SOI layer was patterned into mesa structures using optical lithography and a SF₆/O₂ dry etching process at RT. Subsequently, another optical lithography and consecutive ion implantation of antimony with a dose of 2.5×10^{14} cm⁻² at 10 kV acceleration voltage and 1×10^{14} cm⁻² at 2 kV acceleration voltage were carried out in order to realize degenerately doped source/drain regions. The resist used as an implantation mask was removed in an O₂ plasma ashing process and a Piranha clean (H₂SO₄:H₂O₂ = 3:1). Next, dopant activation at 900 °C for 60 s in 2000 sccm Ar at 1000 mbar in a rapid thermal annealing tool was conducted. ECV measurements of bulk-Si test samples implanted and activated with the same parameters confirm a doping concentration of $0.9 - 1 \times 10^{20}$ cm⁻³ at the doped Si surface. Prior to the gate stack fabrication, a RCA-clean at 80 °C [18] was performed.

To study the impact of oxide scaling on SS stacks with different gate oxides and oxide thicknesses were fabricated. HfO₂ as a high-k dielectric and SiO₂ as a more common gate oxide with and without an interlayer were formed. Three different interlayer configurations were realized: i) no interlayer, ii) a SC2 (standard clean 2)-grown SiO₂ interlayer and iii) a rapid thermal processing (RTP)-grown SiN interlayer. In total, the following seven gate stacks were fabricated as shown in Fig. 2: Two reference stacks without an interlayer were covered with 18 nm (stack 1) and 9 nm (stack 2) HfO₂ after SC2-oxide removal. In addition, one stack leaving the SC2-oxide as an interlayer on the sample surface was fabricated (stack3). The 0.8 nm and 1.8 nm SiN layers of stack 4–7 were grown with rapid thermal nitridation (2000 sccm NH₃, 1 bar, 60 s) at 450 °C and 900 °C, respectively [17]. Prior, a short HF-dip (1%) was performed in order to remove the SC2-oxide grown during the RCA clean. The grown SiN was immediately covered with 9 nm, 18 nm of HfO₂ or 18 nm of SiO₂ deposited with atomic layer deposition (ALD) or chemical vapor deposition (CVD), respectively. The electrical permittivity of HfO₂ and SiO₂ were determined by C-V measurements of MOS capacitors (not shown here) and resulted in $\epsilon_r = 3.9$ and $\epsilon_r = 20$, respectively.

A gate electrode made of 20 nm TiN and 150 nm Al was deposited with sputtering and electron beam physical vapor deposition (EBPVD). This gate stack was successively patterned using optical lithography and wet-chemical etching. To this end, commercially available Al etch (TechniEtch Al80, MicroChemicals) (1 nm/min) and H₂O₂ at 60 °C (2.22 nm/min) were employed, respectively. In the source/drain regions the interlayer and the gate oxide were removed with buffered oxide etch (BOE) with an etch rate of 1 nm/s and 2.33 nm/min for SiO₂ and HfO₂, respectively. Note, that the etch rate for SiN does not scale linearly with the layer thickness. This is due to the complete dissociation of ammonia at elevated temperatures which increases the amount of Si-N bonds at the interface and thereby leads to a denser SiN layer [17], [19]. Consequentially, the etch rates for 0.8 nm and 1.8 nm SiN are 8 nm/min and 0.98 nm/min.

Finally, source/drain contact pads (150 nm Al) were fabricated with EBPVD and a subsequent lift-off process followed by an annealing in forming gas (N₂/H₂) at 400 °C for 30 min.

III. RESULTS AND Discussions

The fabricated devices were characterized at RT and cryogenic temperatures between 40 K to 4.2 K using a Keithley 4200 SCS parameter analyzer with two pre amplifiers. For the measurements, samples were mounted on a dipstick and immersed in liquid helium ($T = 4.2$ K). Measurement temperatures > 4.2 K were adjusted by the dipping height of the sample above the liquid helium. The temperature during the measurements was observed using a Lakeshore diode DT 670-SD with an accuracy of ± 12 mK at 4.2 K which was mounted right next to the sample.

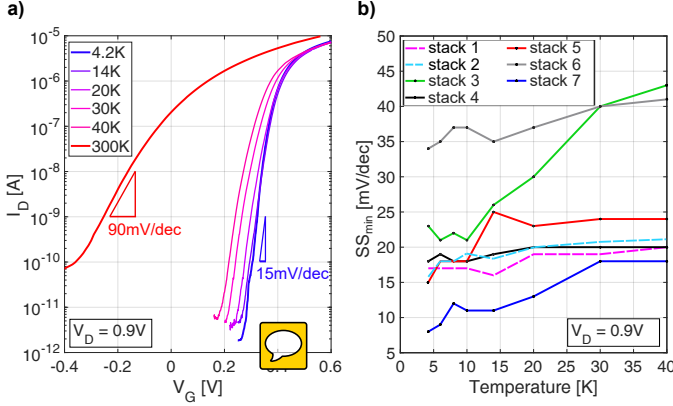


Fig. 3. (a) Transfer curves of a device with 18 nm HfO₂ and 1.8 nm SiN (stack 5) measured between 300 K and 4.2 K. (b) Extracted minimal $SS_{\min}$ as a function of temperature for all seven gate stacks. ($V_D = 0.9$ V)

Figure 3a) shows exemplary transfer curves of the device with stack 5 between 40 K and 4.2 K at an applied drain voltage of $V_D = 0.9$ V. Figure 3b) shows the corresponding temperature dependency of the minimal $SS_{\min}$ for all seven gate stacks. The saturation of $SS_{\min}$ is observable between 40 K and 4.2 K but interestingly, depends remarkably on the gate stack. In the following, the effect of gate stack modifications on band and tailing (i.e. T^*) will be discussed.

First, compare gate dielectric stack 6, 4, 1 and 2: Replacing SiO₂ (stack 6) with HfO₂ (stack 4) reduces $SS_{\min}$ from 34 mV/dec to 18 mV/dec at 4.2 K. Omitting the 0.8 nm SiN interlayer (stack 1), $SS_{\min}$ is slightly reduced to 17 mV/dec at 4.2 K. Finally, reducing the HfO₂ to 9 nm yields $SS_{\min} = 16$ mV/dec. These observations may lead to the conclusions that gate dielectric scaling reduces SS and that interface engineering with SiN is counterproductive. The first conclusion is only partly true: while the gate stack with the smallest EOT yields the lowest $SS_{\min}$, a significant reduction of $SS_{\min}$ is only obtained when replacing SiO₂ with HfO₂, i.e. when exchanging the material. Hence, the strong reduction appears to be rather related to material properties such as defect densities than the dielectric permittivity. This is consistent with the rather small decrease of $SS_{\min}$ when the same material is considered (stacks 1, 2 and 4). Next, we elaborate on the second conclusion above namely that interface engineering may be counterproductive. To elucidate the impact of interlayers we compare gate stacks 1, 3, 4 and 5 that all feature the same HfO₂ with 18 nm thickness. While the

device with a SC2-grown SiO₂ interlayer (stack 3) performs worst the steepest slope with 15 mV/dec is obtained with a 1.8 nm SiN interlayer (stack 5). It is important to note that $SS_{\min}$ in the case of stack 5 is even smaller than the device with 9 nm HfO₂ and no interlayer (stack 2). These observations clearly show how important proper interface engineering of the silicon surface for the saturation of $SS_{\min}$ is and that a SiN interlayer of appropriate thickness yields the best switching behavior. The fact that the gate stack with 0.8 nm SiN (stack 4) performs worse than without interlayer is due to the incomplete dissociation of NH₃ and the formation of Si-H bonds at the silicon-SiN interface during the rapid thermal nitridation process at 450 °C (consistent with our earlier observation that 0.8 nm of SiN performs worse as Fermi level depinning layer than 1.8 nm in metal semiconductor contacts [17]).

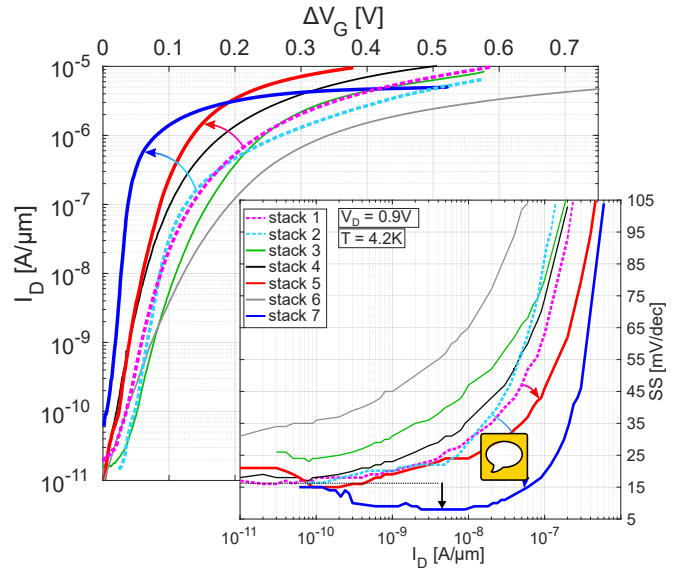


Fig. 4. Comparison of transfer curves for stack 1–7 at 4.2 K measured at $V_D = 0.9$ V. The increased curvature close to the transition region indicates reduced inflection (indicated by flexed arrows) compared to the corresponding reference sample without SiN (stack 1 and stack 2). The smallest $SS_{\min}$ of 8 mV/dec was reported for stack 7. The inset shows SS as a function of I_D at 4.2 K and $V_D = 0.9$ V for stack 1–7. The reduced inflection for stack 5 and stack 7 is indicated by flexed arrows. The suppressed saturation of SS for stack 7 is indicated by a black arrow.

Finally, by using a 1.8 nm SiN interlayer and oxide scaling simultaneously (stack 7) $SS_{\min}$ is further reduced to 8 mV/dec. Hence, stack 7 nearly halves $SS_{\min}$ in comparison to stack 2 without the interlayer but the same HfO₂ thickness. This corroborates that only a simultaneous interface modification and oxide scaling reduces band tailing. Next, the effect of gate stack engineering on the inflection at 4.2 K will be discussed.

The main panel of Fig. 4) depicts the transfer curves of all devices at 4.2 K for $V_D = 0.9$ V. Since the threshold voltage of each device differs, the relative gate voltage ΔV_G is used for the comparison. Furthermore, the inset of Fig. 4 shows SS as a function of $\log(I_D)$. Comparing devices exhibiting stack 1 with stack 5 as well as stack 2 with stack 7 (indicated by a blue and red flexed arrows), inflection is substantially reduced

in devices with the 1.8 nm SiN interlayer. Overall, stack 7 shows the steepest switching behavior with $SS < 10$ mV/dec over two orders of magnitude (as discussed before) and also the smallest inflection. Note that the smaller on-current of stack 7 is due to a parasitic drain resistance stemming from an accidental gate underlap; therefore, the reduction of inflection may be even underestimated. Even though the gate control is decreased when inserting the SiN interlayer but keeping the HfO_2 thickness constant, inflection is reduced with SiN indicating an effective lowering of localized states (i.e. C_{it}). Reducing the HfO_2 thickness increases C_{ox} such that the ratio C_{it}/C_{ox} (cf. Eqn. (2)) and hence inflection is strongly reduced in the device with stack 7.

IV. CONCLUSION

MOSFETs with interface engineered gate stacks were characterized at cryogenic temperatures. It was shown that a lowering of the SS saturation only by oxide scaling is not achievable. However, the combination of oxide scaling and silicon nitride interface engineering led to the steepest switching with $SS < 10$ mV/dec at 4.2 K over two orders of magnitude of I_D . Furthermore, inflection was strongly reduced due to an effective suppression of localized states around the band edges. From these observations, mobile and localized states were reduced by gate stack and interface engineering paving the way for steep slope cryogenic transistors.

REFERENCES

- [1] P. Galy, J. Camirand Lemyre, P. Lemieux, F. Arnaud, D. Drouin, and M. Pioro-Ladriere, "Cryogenic Temperature Characterization of a 28-nm FD-SOI Dedicated Structure for Advanced CMOS and Quantum Technologies Co-Integration," *IEEE Journal of the Electron Devices Society*, vol. 6, pp. 594–600, 2018.
- [2] L. L. Guevel, G. Billiot, X. Jehl, S. De Franceschi, M. Zurita, Y. Thonnart, M. Vinet, M. Sanquer, R. Maurand, A. G. M. Jansen, and G. Pillonnet, "19.2 A 110mK 295μW 28nm FDSOI CMOS Quantum Integrated Circuit with a 2.8GHz Excitation and nA Current Sensing of an On-Chip Double Quantum Dot," in *2020 IEEE International Solid-State Circuits Conference - (ISSCC)*, pp. 306–308, Feb. 2020. ISSN: 2376-8606.
- [3] E. Charbon, F. Sebastiano, A. Vladimirescu, H. Homulle, S. Visser, L. Song, and R. M. Incandela, "Cryo-CMOS for quantum computing," in *2016 IEEE International Electron Devices Meeting (IEDM)*, pp. 13.5.1–13.5.4, Dec. 2016. ISSN: 2156-017X.
- [4] S. Schaal, S. Barraud, J. J. L. Morton, and M. F. Gonzalez-Zalba, "Conditional Dispersive Readout of a CMOS Single-Electron Memory Cell," *Physical Review Applied*, vol. 9, p. 054016, May 2018. Publisher: American Physical Society.
- [5] H. Homulle, L. Song, E. Charbon, and F. Sebastiano, "The Cryogenic Temperature Behavior of Bipolar, MOS, and DTMOs Transistors in Standard CMOS," *IEEE Journal of the Electron Devices Society*, vol. 6, pp. 263–270, 2018. Conference Name: IEEE Journal of the Electron Devices Society.
- [6] H. Hanamura, M. Aoki, T. Masuhara, O. Minato, Y. Sakai, and T. Hayashida, "Operation of bulk CMOS devices at very low temperatures," *IEEE Journal of Solid-State Circuits*, vol. 21, pp. 484–490, June 1986. Conference Name: IEEE Journal of Solid-State Circuits.
- [7] R. M. Incandela, L. Song, H. Homulle, F. Sebastiano, E. Charbon, and A. Vladimirescu, "Nanometer CMOS characterization and compact modeling at deep-cryogenic temperatures," in *2017 47th European Solid-State Device Research Conference (ESSDERC)*, pp. 58–61, Sept. 2017. ISSN: 2378-6558.
- [8] A. Beckers, F. Jazaeri, and C. Enz, "Theoretical Limit of Low Temperature Subthreshold Swing in Field-Effect Transistors," *IEEE Electron Device Letters*, vol. 41, pp. 276–279, Feb. 2020. Conference Name: IEEE Electron Device Letters.
- [9] A. Kamgar, "Subthreshold behavior of silicon MOSFETs at 4.2 K," *Solid-State Electronics*, vol. 25, pp. 537–539, July 1982.
- [10] H. Achour, R. Talmat, B. Cretu, J.-M. Routoure, A. Benfdila, R. Carin, N. Collaert, E. Simoen, A. Mercha, and C. Claeys, "DC and low frequency noise performances of SOI p-FinFETs at very low temperature," *Solid-State Electronics*, vol. 90, pp. 160–165, Dec. 2013.
- [11] H. Bohuslavskyi, A. G. M. Jansen, S. Barraud, V. Barral, M. Cassé, L. Le Guevel, X. Jehl, L. Hutin, B. Bertrand, G. Billiot, G. Pillonnet, F. Arnaud, P. Galy, S. De Franceschi, M. Vinet, and M. Sanquer, "Cryogenic Subthreshold Swing Saturation in FD-SOI MOSFETs Described With Band Broadening," *IEEE Electron Device Letters*, vol. 40, pp. 784–787, May 2019. Conference Name: IEEE Electron Device Letters.
- [12] G. Ghibaudo, M. Aouad, M. Casse, S. Martinie, T. Poiroux, and F. Balestra, "On the modelling of temperature dependence of subthreshold swing in MOSFETs down to cryogenic temperature," *Solid-State Electronics*, vol. 170, p. 107820, Aug. 2020.
- [13] A. Beckers, F. Jazaeri, and C. Enz, "Inflection Phenomenon in Cryogenic MOSFET Behavior," *IEEE Transactions on Electron Devices*, vol. 67, pp. 1357–1360, Mar. 2020.
- [14] J. Knoch, B. Richstein, Y. Han, D. König, M. Frentzen, L. Hellmich, J. Klos, S. Scholz, and Q. Zhao, "Cryogenic steep slope transistors," *Silicon Nanoelectronics Workshop*, pp. 1–2, 2022.
- [15] D. Molinnus, H. Iken, A. L. Johnen, B. Richstein, L. Hellmich, A. Poghosian, J. Knoch, and M. J. Schöning, "Miniaturized pH-Sensitive Field-Effect Capacitors with Ultrathin Ta2O5 Films Prepared by Atomic Layer Deposition," *physica status solidi (a)*, vol. 219, no. 8, p. 2100660, 2022.
- [16] H. Kobayashi, K. Imamura, K. ichi Fukayama, S.-S. Im, O. Maida, Y.-B. Kim, H.-C. Kim, and D.-K. Choi, "Complete prevention of reaction at hfo2/si interfaces by 1nm silicon nitride layer," *Surface Science*, vol. 602, no. 11, pp. 1948–1953, 2008.
- [17] B. Richstein, L. Hellmich, and J. Knoch, "Silicon Nitride Interface Engineering for Fermi Level Depinning and Realization of Dopant-Free MOSFETs," *Micro*, vol. 1, pp. 228–241, Dec. 2021. Multidisciplinary Digital Publishing Institute.
- [18] W. Kern, "The Evolution of Silicon Wafer Cleaning Technology," *Journal of The Electrochemical Society*, vol. 137, pp. 1887–1892, June 1990.
- [19] S. Fischer, H. I. Kremer, B. Berghoff, T. Maß, T. Taubner, and J. Knoch, "Dopant-free complementary metal oxide silicon field effect transistors: Dopant-free MOSFETs," *physica status solidi (a)*, vol. 213, pp. 1494–1499, June 2016.

The text should be about 2 and 2/3 pages for the first submission . The experimental part is too long and too much in detail.