Abstract—This paper presents a physics-based model for the threshold voltage in bulk MOSFETs valid from room down to cryogenic temperature (4.2 K). The proposed model is derived from Poisson’s equation including bandgap widening, intrinsic carrier-density scaling, and incomplete ionization. We demonstrate that accounting for incomplete ionization in the expression of the threshold voltage is critical for an accurate estimation of the current. The model is validated with our experimental results from nMOSFETs of a 28-nm CMOS process. The developed model is a key element for a cryo-CMOS compact model and can serve as a guide to optimize processes for high-performance cryo-computing and ultra-low-power quantum computing.

I. INTRODUCTION

Remarkable progress towards a silicon quantum computer in recent years has attracted attention to the operation of VLSI technologies at deep-cryogenic temperatures (<10 K) [1]–[6]. Dedicated cryo-CMOS circuits can reduce the overhead associated with the initialization, manipulation, coupling, and read-out of qubits. The power consumption of the CMOS circuits envisioned for qubit control is severely limited (∼1 W at 4.2 K). At the same time, the circuits need to meet sharp timing constraints on the control pulses applied to the qubits. This delicate trade-off between power consumption and performance at cryogenic temperatures ultimately asks for an accurate model of the MOSFET threshold voltage ($V_T$).

Besides quantum computing, temperature scaling can also provide a performance booster for conventional computing applications, if harnessed carefully. As shown in Fig. 1, cryogenic operation does not improve the on-state current ($I_{ON}$) of small commercial devices at the maximum supply voltage ($V_{DD}$). Furthermore, the available overdrive voltage ($V_{DD} - V_T$) is reduced by the substantial increase of $V_T$ down to cryogenic temperatures ($\Delta V_T \approx 0.1$ – 0.2 V). However, there is room for improvement since the leakage current is extremely low at cryogenic temperature. The $I – V$ curve can thus be shifted by selecting a custom work function difference and/or doping, hence supporting a larger $V_{DD} - V_T$ for high performance applications or a scaled $V_{DD}$ for ultra-low power applications.

Importantly, both technology optimization and the use of commercial devices require that $V_T$ can be accurately predicted. Since the MOSFET turns on over a few millivolts at cryogenic temperatures, a small mis-prediction of $V_T$ leads to orders of magnitude difference in the drain current. To arrive at the most accurate model for $V_T$, incomplete ionization of the dopants and other low-temperature phenomena need to be taken into account by starting from the physics. While many models have been proposed so far for the $V_T$ at low temperature, they are either semi-empirical, validated only down to moderate cryogenic temperature, or not accounting for the various physical phenomena contributing to the temperature dependence of $V_T$ [7]–[10]. For instance, the physics-based model from Jaeger and Fox [8] is valid down to 77 K and does not account for the dopant freezeout. Recently, Dao et al. [10] derived a $V_T$ model valid down to 6 K by introducing two empirical parameters to account for the field-assisted ionization of the dopants. In this work, we demonstrate that both freezeout and field-assisted ionization are accounted for by a single incomplete-ionization term in the Poisson-Boltzmann equation. This results in an entirely physics-based, long-channel $V_T$ model without introducing fitting parameters.

II. CRYOGENIC MEASUREMENTS IN 28-NM CMOS

Low-temperature measurements were performed in the four $n$-type width-over-length device-corners of a 28-nm bulk CMOS process. After cooling a Lakeshore CPX cryogenic probe station to its 4.2-K base temperature, the station was heated to 298 K with intermediate temperature steps at 20, 36, 50, 77, 110, 160, and 210 K. The transfer characteristics in linear ($V_{DS} = 5$ mV) and saturation ($V_{DS} = 0.9$ V) were acquired using a Keysight B1500A device analyzer. Temperature-variable probes were used on the bulk and source contact pads to ensure the same bulk reference at the different temperature steps. The measurements in saturation are shown in Fig. 2(a). For all devices, a monotonic increase is obtained, leveling off in the deep cryogenic regime (below 50 K). The drain-induced barrier lowering (DIBL) [Fig. 2(b)] follows a similar trend and is highest for a narrow-long device.

III. OVERVIEW OF $V_T$ TEMPERATURE DEPENDENCES

We assume an $n$-type MOSFET with uniform acceptor doping density ($N_A$) in the silicon body [Fig. 3(a)]. The source and drain junctions can be assumed completely ionized at all temperatures due to impurity-band formation as a result of degenerate doping [9]. Based on the MOSFET band diagrams in Fig. 3 we discuss the physical phenomena that impact the $V_T$ of a large MOSFET. The vacuum energy level ($E_{vac}$) is taken as the reference and a potential drop ($\psi$) is defined with respect to the intrinsic energy level $E_i$ in the bulk ($E_i$), i.e., $\psi = -(E_i - E_{i,b})/q$, where $q$ is the elementary charge.

1) Bandgap Widening and Scaling of Fermi-Dirac Occupation Function: The Si bandgap increases from $\approx 1.12$ to 1.17 eV from 298 to 4.2 K, plotted in Fig. 4. The Fermi-Dirac occupation function $f(E)$ scales exponentially, approaching a step function at 4.2 K [compare $f(E)$ in Fig. 4(b) and (d)]. More band bending is required to generate sufficient overlap of $f(E)$ with the density-of-states (DOS) in the conduction band. Both phenomena will thus act to increase...
3) Increase in Bulk Fermi-potential: The decrease in $n_i$ is sufficiently strong to make the Fermi potential (assuming complete ionization) increase, given by $\Phi_F = U_T \ln (N_A / n_i) [\text{Fig 3(b) vs. 3(c)}]$, despite its direct dependence on $U_T = kT/q$, the thermal voltage. Indeed, in order to satisfy charge neutrality in the case of a completely ionized $N_A$ at 4.2 K $[p_p = n_i \exp (\Phi_F / U_T) = N_A]$, $E_F$ needs to lie very close to the valence band edge ($E_v$) to compensate for $n_i$ [Fig 3(c)]. The increase in $\Phi_F$ ranges from about 0.1 V to 0.5 V depending on $N_A$, as shown by the dashed lines in Fig 3. It can be checked that $\lim_{T\to 0 K} \Phi_F = E_g/2$. This increase in $\Phi_F$ constitutes a large portion of the $V_T$ increase. However, the $\Phi_F$ increase is too strong and not sufficiently accurate. At 4.2 K, the silicon body would be degenerate ($E_g/2 - \Phi_F < 3kT$), which does not stroke with the measured electrical behavior in MOSFETs at 4.2 K. Incomplete ionization of $N_A$ needs to be taken into account in the Fermi potential to avoid this.

4) Incomplete Ionization (thermal): Returning back to the charge neutrality, this gives $p_p = N_A = N_A f(E_A)$, where $f(E_A)$ is the ionization probability of the dopants, both due to temperature (thermal ionization) and the applied voltages (field-assisted ionization). Under charge neutrality, the field is zero and thus $f(E_A)$ equals the thermal ionization probability, given by $f_\theta = 1/\{1 + \exp [\Phi_F / V_\Sigma B]\}$, plotted in Fig 6(b).

Below $f_\theta \approx 20\%$, the silicon body will enter freezeout. At 4.2 K, incomplete ionization results in the position of $E_F^*$ instead of $E_F$ in Fig 3(c), or mathematically:

$$\Phi_F^* = U_T \ln \left( \frac{N_A}{n_i} \right) - U_T \ln \left( 1 + \sqrt{1 + \left( \frac{4\alpha N_A}{n_i} \right)} \right) - \frac{\Delta \Phi_F}{2},$$

where $\alpha = g_A \exp [(E_A - E_v) / (kT)] = g_A \exp (-\Phi_A / U_T)$. The acceptor potential $\Phi_A \equiv (E_i - E_A)/q$ as well as $\Phi_F^*$ are shown in Fig 4 with solid lines. The function $\Delta \Phi_F = \Phi_F - \Phi_F^*$ is plotted in Fig 6(a) for different $N_A$ and is in the order of millivolts. It can be checked that $\lim_{T\to 0 K} \Delta \Phi_F = \delta / 2$ with $\delta = E_A - E_v = 0.045 eV$, for typical, hydrogen-like Si:B doping and thus $\lim_{T\to 0 K} \Phi_F^* = \Phi_A + \delta / 2$ as shown in Fig 4. It might seem that $\Delta \Phi_F$ due to incomplete ionization is negligible for $V_T$ at cryogenic temperatures compared to the overall increase of $\Phi_F$ or $\Phi_F^*$ in Fig 4. However, Fig 5 shows a large discrepancy in inverted density ($n_p$) at 2$\Phi_F$ and 2$\Phi_F^*$.

5) Incomplete Ionization (field): A second effect associated with the incomplete dopant ionization is the process of the frozen-out dopants ionizing due to the field, as shown by the band bending in Fig 3(d). However, this field-assisted process of ionization (e.g., Poole-Frenkel [12]) happens in early depletion and is already completed near the surface before reaching the on-set of strong inversion where $V_T$ is evaluated. This process can thus be possibly neglected in $V_T$ and is not a justification for fitting parameters as done in [10].

IV. ANALYTICAL $V_T$ MODEL

From standard MOS device physics it follows that the gate-to-bulk voltage is given by $V_{GB} = \psi_s + \Phi_{ms} + \varepsilon_s \varepsilon_s (\psi_s) / C_{ox}$, where $\psi_s$ is the surface potential, $\varepsilon_s$ the silicon permittivity,
\[ E_s \] is the electric field normal to the semiconductor/oxide interface, \( C_{ox} \) is the gate oxide capacitance, and \( \Phi_{ms} \) is the difference between the metal (\( \Phi_m \)) and semiconductor (\( \Phi_s \)) work functions. Note that we do not take into account the oxide and interface-trapped charges. The threshold voltage is usually defined as the \( V_{GB} \) evaluated at a given value of \( \psi_s \), i.e., the inversion threshold \( \psi'_s \), for which \( n_p \) equals the totally ionized \( N_A \). Note that in the strict meaning of the inversion threshold when incomplete ionization is present, we have to define the threshold at \( n_p = N_A \) since then the density of carriers in the channel has been inverted from the flatband case \( (p_p = N_A) \). However, this density can be very low due to freezeout and it is thus more meaningful to keep defining the threshold at \( n_p = N_A \). The temperature dependence of the four terms of \( V_T \) are investigated separately below:

1) \( \psi'_s \): Near 298 K, the inversion threshold is taken at 2\( \Phi_F \) since then \( n_p = N_A \) when complete ionization can be assumed. However, this threshold needs to be reassessed for the low and cryogenic temperatures. From \( n_p = N_A \), we find that \( \psi'_s = U_T \ln(N_A/n_i) + \Phi_F + V_{ch}(x) + V_{SB} \), where \( V_{ch} \) is the channel voltage. We evaluate \( V_T \) at source (\( x = 0 \)) and assume \( V_{SB} = 0 \). It can be checked that this yields back 2\( \Phi_F \) for complete ionization (\( \alpha = 0 \)). However, using (1), the correct inversion threshold is \( \psi'_s = \Phi_F + \Phi_F^* = 2\Phi_F^* + \Delta \Phi_F = 2\Phi_F - \Delta \Phi_p \), and not 2\( \Phi_F \), nor 2\( \Phi_F^* \).

2) \( \Phi_{ms} \): From Fig.3, we can obtain that \( \Phi_{ms} = \Phi_m - (\chi + E_g/2 + \Phi_F) \). \( E_{F,m} \) is assumed temperature independent due to metallic gate (or degeneracy in poly-Si). Due to lack of available data on the temperature dependence of \( \chi \) in silicon down to 4.2 K, it is assumed that \( \chi \) decreases with the same amount as \( E_g \) increases. Under these assumptions, the temperature dependence of \( \Phi_{ms} \) is defined by \( \Phi_F^* \).

3) \( \varepsilon_s \), \( \varepsilon_s(\psi'_s) \), \( C_{ox} \): The temperature dependences of \( \varepsilon_s \) and \( C_{ox} \) are assumed negligible. From the Poisson-Boltzmann equation, \( \partial^2 \psi(y)/\partial y^2 = -\rho(\psi)/\varepsilon_s \), including incomplete ionization, \( \varepsilon_s \) in depletion can be derived as [11]:

\[
\varepsilon_s = \left( U_T/L_D \right) \sqrt{\psi_s/U_T + F}, \text{ where } \psi_s \text{ is the surface potential, } L_D = \sqrt{\varepsilon_s U_T/(2qN_A)}, \text{ and } F \text{ an extra term due to the incomplete ionization of } N_A, \text{ given by } F = \ln f_0 + \\
\ln \{1 + \alpha \exp [(-\psi_s + \Phi_F^* + V_{ch} + V_{SB})/U_T]\}. \text{ The second term, dependent on } \psi_s, \text{ is due to the field-assisted ionization process. Evaluating } \varepsilon_s \text{ at } \psi'_s = 2\Phi_F - \Delta \Phi_F + V_{ch} + V_{SB} \text{ gives}
\]

\[
\varepsilon_s(\psi'_s) = \frac{U_T}{L_D} \sqrt{\frac{2\Phi_F - \Delta \Phi_F}{U_T}} + \ln f_0 + \ln \left( 1 + \alpha e^{-\Phi_F^*/T} \right).
\]

for \( V_{ch} = V_{SB} = 0 \). As evidenced by Figs.6(c)-(d), the two logarithmic terms can be neglected tolerating a \( \approx 1\% \) error on the electric field. This approximation can be viewed as neglecting the field-assisted ionization process which is completed at threshold \( (\psi'_s \geq 0) \). Note that the different inversion threshold due to incomplete ionization, \( 2\Phi_F - \Delta \Phi_F \), is still taken into account. Considering all the temperature dependences together from the previous subsections, \( V_T \) is derived at the source as:

\[
V_T = 2\Phi_F - \Delta \Phi_F + V_{SB} + \Phi_F^* - \Phi_F^* + \Gamma_b \sqrt{2\Phi_F - \Delta \Phi_F + V_{SB}}.
\]

Fig. 3. a) Cross-section of an nMOSFET with incomplete ionization of \( N_A \) (thermal ionization probability given by \( f_p \)) and thermal + field-assisted ionization given by \( f = \text{Fermi-Dirac} \). Band diagrams from interface to bulk are shown: b) flat-band at 298 K, c) flat-band at 4.2 K \( (p_p = N_A \text{ gives } \Phi_F, \text{ and } p_p = N_A \text{ gives } \Phi_F) \), and d) strong-inversion threshold in the strict meaning \( (n_p = N_A) \). To reach this point, a band bending of 2\( \Phi_F^* \) is needed when including incomplete ionization. An additional band bending over \( q\Delta \Phi_p = q\Phi_F - q\Phi_F^* \) is needed to increase the inverted electron density from \( n_p = N_A \) to \( n_p = N_A \).
where $\Phi_m = \Phi_m - (\chi + E_g/2)$ and $\Gamma = \sqrt{2qN_A\varepsilon_{si}/C_{ox}}$. It is now interesting to develop the error introduced by assuming complete ionization at all temperatures in (3) ($\Phi_F = \Phi_F$, $\Delta\Phi_F = 0$), reducing $V_T$ to $V_T^{co} = V_T(\psi_s^* = 2\Phi_F)$. The absolute error $V_T^{co} - V_T$ (note that $V_T^{co} > V_T$) is not large (millivolts) compared to a typical $V_T$ at 4.2 K in a commercial device ($\approx 0.4$ V). However, due to the steepness of the subthreshold swing at 4.2 K, the absolute and relative error on the current are very large at cryogenic temperatures, as shown in Figs. (g) and (h). The current is considered in the linear regime, $I_{DS} = -(W/L)\mu Q_mV_{DS}$, and at constant mobility, $\mu$, to illustrate this discrepancy. The used $Q_m$ including incomplete ionization is given by

$$Q_m = -\gamma \sqrt{\frac{n_i}{N_A} \exp(\beta)} + \frac{\psi_s}{U_T} + F + \frac{\sqrt{\psi_s}}{U_T} + F,$$

where $\gamma = \varepsilon_{si}U_T/L_D$ and $\beta = (\psi_s - \Phi_F - V_{ch} - V_{SB})/U_T$. Fig. 7 validates the $V_T$ model from (3) with the experimental results in a large $n$MOS at $V_{DS} = 5$ mV with negligible DIBL.

V. CONCLUSION

The increase of $V_T$ down to 4.2 K is physically explained by $E_g$ widening and Fermi-Dirac scaling. This leads to a strong decrease in $n_i$ and thus increase in the inversion threshold $2\Phi_F$. Bulk incomplete ionization or freezeout makes the threshold $2\Phi_F - \Delta\Phi_F$, where $\Delta\Phi_F$ is temperature- and doping-dependent. Although in mV-range, $\Delta\Phi_F$ is significant for the cryogenic current. The process of field-assisted dopant ionization is negligible at inversion threshold. Good agreement is obtained between the $V_T$ model and the experimental results.

ACKNOWLEDGMENT

This project has received funding from the European Union’s Horizon 2020 Research & Innovation Programme under grant agreement No. 688539 MOS-Quito. The authors thank AQUA (EPFL) for access to a cryo-probe station, and Dr. Grabinski for guidance with measurement automation.

REFERENCES

[1] M. Veldhorst, C. Yang et al., “A two-qubit logic gate in silicon,” Nature, vol. 526, no. 7573, p. 410, 2015.
[2] L. Bourdet, L. Hutin et al., “All-Electrical Control of a Hybrid Electron Spin/Valley Quantum Bit in SOI CMOS Technology,” IEEE Transactions on Electron Devices, vol. 65, no. 11, pp. 5151–5156, Nov 2018.
[3] E. Charbon, F. Sebastiano et al., “Cryogenic CMOS for quantum computing,” in 2016 IEEE IEDM, Dec 2016, pp. 13.5.1–13.5.4.
[4] A. Beckers, F. Jazaeri, et al., “Cryogenic characterization of 28 nm bulk CMOS technology for quantum computing,” in 2017 47th European Solid-State Device Research Conference (ESSDERC), pp. 62–65.
[5] A. Beckers, F. Jazaeri, and C. Enz, “Characterization and modeling of 28-nm Bulk CMOS Technology Down to 4.2 K,” IEEE Journal of the Electron Devices Society, vol. 6, pp. 1007–1018, 2018.
[6] A. Beckers, F. Jazaeri et al., “Characterization and modeling of 28-nm FDSOI CMOS technology down to cryogenic temperatures,” Solid-State Electronics, 2019.
[7] E. A. Gutierrez-Díez, J. Deen, and C. Claeyss, “Low temperature electronics: physics, devices, circuits, and applications,” Elsevier, 2000.
[8] R. M. Fox and R. C. Jaeger, “MOSFET behavior and circuit considerations for analog applications at 77 K,” IEEE T-ED, vol. 34, no. 4, 1987.
[9] A. Akturk, J. Allnutt et al., “Device modeling at cryogenic temperatures: Effects of incomplete ionization,” IEEE T-ED, vol. 54, no. 11, 2007.
[10] N. C. Dan, A. E. Kass et al., “An enhanced MOSFET threshold voltage model for 6–300K temperature range,” Microelectron. Reliability, 2017.
[11] A. Beckers, F. Jazaeri, and C. Enz, “Cryogenic MOS Transistor Model,” IEEE Trans. on Electron Devices, vol. 65, no. 9, pp. 3617–3625, 2018.
[12] D. P. Foyt, “Impurity ionization in MOSFETs at very low temperatures,” Cryogenics, vol. 30, no. 12, pp. 1056 – 1063, 1990.