Analytical Modeling of Source-to-Drain Tunneling Current Down to Cryogenic Temperatures

Hung-Chi Han, *Student Member, IEEE*, Hung-Li Chiang, *Member, IEEE*, Iuliana P. Radu, and Christian Enz, *Life Fellow, IEEE*

***Abstract*— The subthreshold swing (SS) of MOSFETs decreases with temperature and then saturates below a critical temperature. Hopping conduction via the band tail has been proposed as the possible cause for the SS saturation. On the other hand, numerical simulations have shown the source-to-drain tunneling (SDT) current limits the SS at low temperatures. It has been argued which transport mechanism dominates the cryogenic subthreshold current. Hence, for the first time, this letter presents an analytical model of the SDT current and the corresponding SS, which is validated by cryogenic measurement on devices from an advanced 16 nm FinFET technology.**

***Index Terms*— Cryo-CMOS, cryogenic, FinFET, MOSFET, subthreshold swing, tunneling, quantum computing.**

I. Introduction

THE study on transistors operating at low temperatures has been getting more attention due to quantum computing applications [1], [2], [3]. Subthreshold swing (SS) is the figure of merit that characterizes the switching efficiency between on-off states for a transistor. In MOSFETs, the thermionic SS should follow the Boltzmann limit, $SS_{th} = k_B T \ln(10)/q$ (Boltzmann constant k_B, temperature T, and elementary charge q) and reach 0.83 mV/dec at 4.2 K. In reality, SS does not improve that much at cryogenic temperatures due to the SS saturation at the range from sub-10 to 30 mV/dec at 4 K [4], [5], [6], [7], [8]. The physical mechanism that limits SS at extremely low temperatures is being argued. Bohuslavskyi et al. [5] and Beckers et al. [6], [9] claimed that the band tail leads to the SS saturation at cryogenic temperatures. In particular, the hopping transport happens in the localized states just below the conduction band or right above the valance band. Consequently, the localized states in the band tail result in insufficient SS. Another transport mechanism degrading SS at cryogenic temperatures is the source-to-drain tunneling (SDT), which is an intraband tunneling from the source to the drain side. In 2002, J. Wang and M. Lundstrom, using Non-Equilibrium Green Function (NEGF) simulation, showed that SDT sets an ultimate scaling limit due to large SS, and SDT makes SS saturate below a critical temperature [10]. Recently, many groups using NEGF simulations have investigated the influence of SDT on Cryo-CMOS in terms of doping concentration, channel length, and temperature [11], [12]. However, a rigorous experimental study has not been performed yet to evidence the SDT manifesting in a down-scaled channel. This work experimentally and theoretically presents the SDT in a 16 nm FinFET technology.

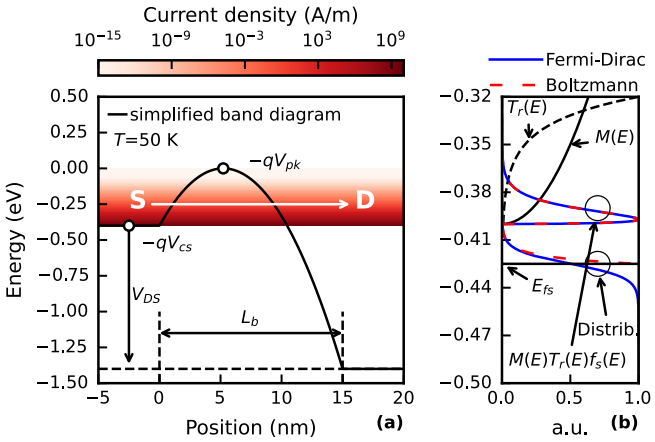

Fig. 1. Modeling of SDT current density in a short channel at 50K. (a) current density on the plot of energy versus position, the barrier energy is described by a quadratic expression. (b) a zoom-in plot from (a), justifying Boltzmann approximation at cryogenic temperatures with E_{fs} locating at 0.025eV below $-qV_{cs}$.

II. Analytical Source-to-Drain Tunneling Model

In a short-channel device, the gate barrier gets thinned and lowered by a strong drain voltage electrostatically due to the short-channel effects, i.e., drain-induced barrier thinning (DIBT) and lowering (DIBL). A carrier with energy lower than the barrier peak has a higher chance of tunneling through the barrier. It leads to the SDT current in the subthreshold regime, which flows in parallel with the thermionic current. To model the SDT current, we adopt Landauer formalism [13]

$$I_{sdt} = \frac{2q}{h} \int M(E) T_r(E) \left(f_s(E) - f_d(E) \right) dE, \qquad (1)$$

Manuscript received 28 February 2023; accepted 6 March 2023. Date of publication 9 March 2023; date of current version 26 April 2023. This work was supported by the European Union's Horizon 2020 Research and Innovation Program under Agreement 871764 (SEQUENCE). The review of this letter was arranged by Editor V. Moroz. *(Corresponding author: Hung-Chi Han.)*

Hung-Chi Han and Christian Enz are with the Integrated Circuits Laboratory (ICLAB), Ecole Polytechnique Fédérale de Lausanne (EPFL), 2000 Neuchâtel, Switzerland (e-mail: hung.han@epfl.ch).

Hung-Li Chiang and Iuliana P. Radu are with the Corporate Research, Taiwan Semiconductor Manufacturing Company, Hsinchu 300-77, Taiwan.

Color versions of one or more figures in this letter are available at https://doi.org/10.1109/LED.2023.3254592.

Digital Object Identifier 10.1109/LED.2023.3254592

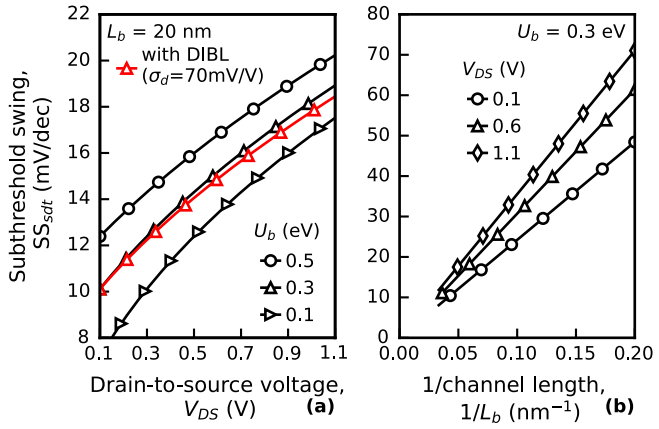

Fig. 2. Simulated subthreshold swing due to SDT. (a) SS_{sdt} versus V_{DS} at different gate barrier potential, U_b, the red line presents the DIBL effect on SS_{sdt}. (b) SS_{sdt} versus $1/L_b$ at different V_{DS}.

with h the Plank constant, $M(E)$ the number of conduction modes, $T_r(E)$ the transmission probability, and $f_{s,d}(E)$ the Fermi-Dirac distribution at source/drain. Since 16 nm FinFET technology is characterized in this work,

$$M(E) = \frac{g_v W \sqrt{2m^*(E - E_{cs})}}{\pi \hbar} \tag{2}$$

is used for the 2-dimensional electron gas with m^* the effective mass, g_v the valley degeneracy, W the channel width, E_{cs} the conduction band edge at the source, and $\hbar$ the reduced Planck constant [14]. The gate barrier of a short-channel device can be described by a quadratic expression [15], which yields a closed-form expression of the transmission probability with the Wentzel–Kramers–Brillouin (WKB) approximation. Hence, as shown in Fig. 1, the gate barrier energy along the y-axis is given by

$$U(y) = -q\left(\frac{V_a(y - y_{pk})^2}{L_b^2} + V_{pk}\right), \tag{3}$$

with coefficient V_a, location of barrier peak y_{pk}, and potential of barrier peak V_{pk}. The term V_{pk} is a function of the gate voltage V_G and the drain-to-source voltage V_{DS} [16]. The boundary condition of $U(y)$ is given by $U(y = 0) = -qV_{cs}$, $U(y = y_{pk}) = -qV_{pk}$, and $U(y = L_b) = -q(V_{cs} + V_{DS})$ with conduction potential at source V_{cs} and channel length L_b. Therefore, V_a is given by

$$V_a = V_{DS} + 2(V_{cs} - V_{pk}) + 2\sqrt{(V_{cs} - V_{pk})(V_{DS} + V_{cs} - V_{pk})}. \tag{4}$$

The transmission probability $T_r(E)$ is therefore defined as [16]

$$T_r(E) = \exp\left(\frac{E + qV_{pk}}{W_{sdt}}\right), \tag{5}$$

where

$$W_{sdt} = \frac{\hbar\sqrt{V_a q}}{\pi L_b \sqrt{2m^*}} \tag{6}$$

is the characteristic decay of $T_r(E)$. Because of (5) and (6), the situations such as a small m^*, energy close to $-qV_{pk}$, or a short L_b, lead to a high $T_r(E)$. When high V_{DS} is applied, V_a gets larger due to (4). Also, the tunneling distance for carriers at $-qV_{cs}$ is much shorter than L_b, as shown in Fig. 1. On top of that, the term $f_d(E)$ in (1) can be neglected due to V_{DS} larger than a few k_BT. Hence, we can get the current density $\frac{2q}{h}M(E)T_r(E)f_s(E)$ at each energy as shown in Fig. 1(a). Although $T_r(E)$ exponentially increases and reaches $T_r(E = -qV_{pk}) = 1$, most of the tunneling carriers have energy close to $-qV_{cs}$ because $f_s(E)$ at low temperatures decays faster than the increase of $T_r(E)$.

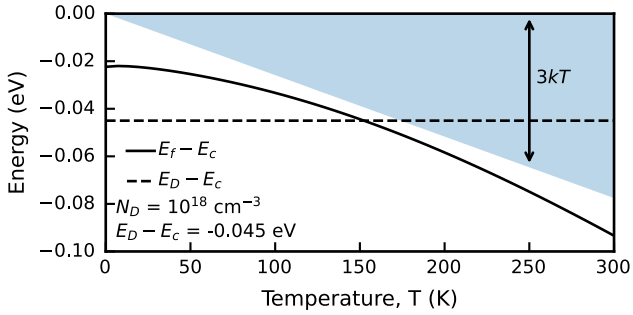

Fig. 3. Temperature dependency of Fermi level for a non-degenerate semiconductor. The term E_f is equal to E_{fs} for the LDD region.

In modern MOSFETs, the implementation of the lightly doped drain-source (LDD) structure is often used between the heavily doped contact and the channel [17]. Since LDD is close to the channel, it defines the position of the Fermi level close to the source E_{fs}. For the n-type LDD region, it has a doping level (N_D) around 10^{18} cm^{-3}. Accounting for incomplete dopant ionization and assuming the non-degenerate semiconductor give the Fermi level (E_f) referring to the conduction band (E_c) as [18]

$$E_f - E_c = k_BT \ln\left(\frac{-1 + \sqrt{1 + \frac{8N_D}{N_c}\exp\left(\frac{E_c - E_D}{k_BT}\right)}}{4\exp\left(\frac{E_c - E_D}{k_BT}\right)}\right), \tag{7}$$

with N_c the effective density of states and E_D the dopant ionization energy. Fig. 3 shows the T dependency of (7), where $E_c - E_f > 3k_BT$ validates the use of the Boltzmann approximation. As $T \to 0$ K, (7) can be simplified to $E_f - E_c = (E_D - E_c)/2 + k_BT\ln(\sqrt{2N_D}/(2\sqrt{N_c}))$. Consequently, E_f locates at $(E_D + E_c)/2$ for few Kelvins [19]. Hence, Fig. 1(b) shows the product of $M(E)T_r(E)f_s(E)$ using Boltzmann approximation, which is close to expression calculated by Fermi-Dirac distribution. Finally, integrating (1) from $-qV_{cs}$ to $-qV_{pk}$ results in a closed-form expression for the SDT current as

$$I_{sdt} = \frac{4\sqrt{2m^*}qg_vWW_c^{3/2}}{h^2}\exp\left(-\frac{U_b}{W_{sdt}}\right) \times \exp\left(\frac{E_{fs} + qV_{cs}}{k_BT}\right)\left(\frac{\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\frac{U_b}{W_c}}\right)}{2} - \sqrt{\frac{U_b}{W_c}}e^{-\frac{U_b}{W_c}}\right), \tag{8}$$

where $U_b = -q(V_{pk} - V_{cs})$, $\frac{1}{W_c} = \left|\frac{1}{W_{sdt}} - \frac{1}{k_BT}\right|$, and erf is the error function. The term U_b is the energy barrier height controlled by the gate and drain voltages; the latter is due to

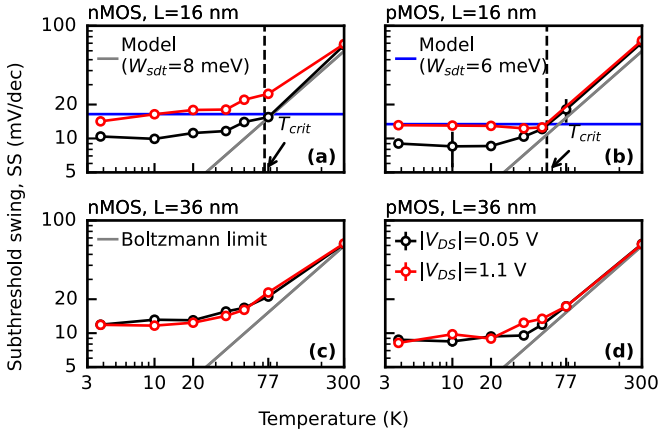

Fig. 4. Average subthreshold swing, SS, versus temperature, measured from nMOS/pMOS FinFETs with 16nm and 36nm of channel length. The blue lines are the model using (9).

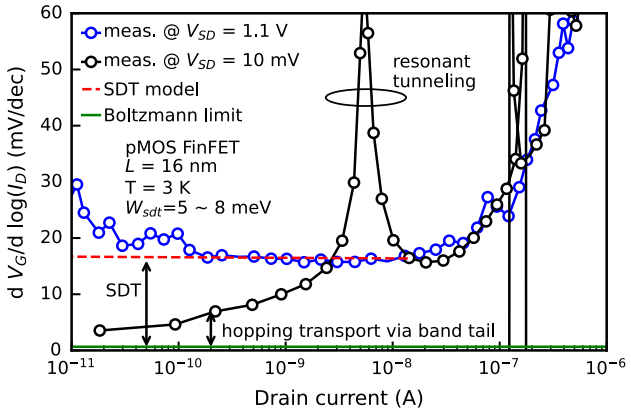

Fig. 5. Characterization of SDT versus drain current at 3K, where the device in linear and saturation modes is compared. The red dashed line is the model using (8) and (9).

the DIBL effect. Using (8), $SS_{sdt} = \partial V_G / \partial \log I_{sdt}$ can then be expressed as

$$SS_{sdt} \simeq n \left(\frac{W_{sdt}}{q}\right) \ln 10 \left(1 + \frac{V_{pk} - V_{cs}}{V_a} \frac{\partial V_a}{\partial V_{pk}}\right)^{-1}. \quad (9)$$

The terms W_c and U_b, in the second line of (8), are functions of V_G. However, their contribution to SS_{sdt} is negligible, compared to $\exp(-U_b/W_{sdt})$ in the RHS of (8). Hence, SS_{sdt} is simplified to (9), where $\partial V_a/\partial V_{pk}$ is analytical thanks to (4). The term n is the slope factor defined by $\partial V_G/\partial V_{pk}$, which is affected by depletion, interface states, and short-channel effects [20], [21]. Eq. (9) further shows that SS_{sdt} is temperature-independent, which leads to the SS saturation at cryogenic temperatures for extremely short-channel devices. Additionally, SS_{sdt} is a function of U_b, V_{DS}, and L_b, because of V_a and $V_{pk} - V_{cs}$. As shown in Fig. 2(a), the increase in V_{DS} leads to a higher SS_{sdt}. Also, a device biased in deep subthreshold regime, i.e., high U_b, has the worse SS_{sdt}. The red line shown in Fig. 2(a) accounts for the first-order DIBL effect, where U_b is lowered by $\sigma_d V_{DS}$ (DIBL parameter, $\sigma_d \sim 70\,\text{mV/V}$ [7]). Particularly, the DIBL effect compensates for the degradation from SDT, where the SS_{sdt} accounting for the DIBL effect (red triangle line) shows a lower value than SS_{sdt} without DIBL effect (black triangle line). In addition, Fig. 2(b) shows that SS_{sdt} exponentially degrades as L_b gets shorter.

A critical temperature can be defined as SS_{th} becomes equal to SS_{sdt}, it yields

$$T_{crit} = \frac{W_{sdt}}{k_B} \left(1 + \frac{V_{pk} - V_{cs}}{V_a} \frac{\partial V_a}{\partial V_{pk}}\right)^{-1}. \quad (10)$$

As $T < T_{crit}$, the SS saturates at the value given by (9) for a short-channel device in the presence of SDT.

III. Experimental Validation and Discussion

Fig. 4 evidences the SDT manifesting in the extremely short-channel devices at low temperatures by presenting $SS(T)$ for 16 nm and 36 nm FinFET. Because of (9), SS slightly changes over the whole subthreshold region. Hence, SS is extracted by taking the average from where the normalized drain current ranges from 10^{-3} to 10^{-2} μA/μm. As reported in the literature [5], [6], SS tends to deviate from the Boltzmann limit below a critical temperature and then saturates at $10 \sim 30\,\text{mV/dec}$. In most cases, the SS saturation at low temperatures is attributed to the hopping conduction in the band tail. However, in Fig. 4(a, b), the 16 nm devices show the worse SS when $|V_{DS}| = 1.1\,\text{V}$, such phenomenon does not happen in the 36 nm devices. This degradation in SS is due to the SDT current that takes over the thermionic and hopping currents. In other words, the W_{sdt} is larger than the characteristic decay of the exponential band tail ($W_t \simeq 4\,\text{meV}$) in [6]. Consequently, SS at $T < T_{crit}$ is dominated by SS_{sdt}, which has a more inefficient swing than that of the thermionic and hopping currents. Finally, the model proposed in (9) is in excellent agreement with the measurement, as shown in Fig. 4(a, b). The experimental SS starts to saturate to SS_{sdt} as $T < T_{crit}$ ($T_{crit} = 72\,\text{K}$ for nMOS and 54 K for pMOS). Fig. 5 highlights the subthreshold behavior of a short pMOS at 3 K in linear and saturation regimes. Because of the short gate length, the resonant tunneling through the ionized dopant appears at low V_{SD} [7]. The SS ascribed to the band tail can then be defined at the ultra-low current level, far from the event of resonant tunneling. Conversely, the resonant tunneling is eliminated as the device is biased at $V_{SD} = 1.1\,\text{V}$. Additionally, the experiment shows that the SS increases at the lower current level due to SDT; this phenomenon is well captured by the proposed model in (8) and the NEGF simulation in [11].

IV. Conclusion

This letter presents an analytical expression of the SDT current and its impact on the subthreshold swing SS saturation of MOSFETs at cryogenic temperatures. It is shown that for extremely short-channel MOSFETs at high V_{DS} and for $T < T_{crit}$, the subthreshold SDT current is taking over the thermionic and hopping currents. Moreover, the model highlights that SS_{sdt} is temperature-independent but depends on the characteristic decay energy W_{sdt} of the tunneling probability. When W_{sdt} is larger than the characteristic energy of the band tail W_t, then SS_{sdt} becomes dominant. The proposed analytical model is successfully validated against the experimental data from a 16 nm FinFET technology.

REFERENCES

[1] E. Charbon, F. Sebastiano, A. Vladimirescu, H. Homulle, S. Visser, L. Song, and R. M. Incandela, "Cryo-CMOS for quantum computing," in *IEDM Tech. Dig.*, Dec. 2016, p. 13, doi: 10.1109/IEDM.2016.7838410.

[2] B. Patra, R. M. Incandela, J. P. G. Van Dijk, H. A. R. Homulle, L. Song, M. Shahmohammadi, R. B. Staszewski, A. Vladimirescu, M. Babaie, F. Sebastiano, and E. Charbon, "Cryo-CMOS circuits and systems for quantum computing applications," *IEEE J. Solid-State Circuits*, vol. 53, no. 1, pp. 309–321, Jan. 2018, doi: 10.1109/JSSC.2017.2737549.

[3] H. L. Chiang, T. C. Chen, J. F. Wang, S. Mukhopadhyay, W. K. Lee, C. L. Chen, W. S. Khwa, B. Pulicherla, P. J. Liao, K. W. Su, K. F. Yu, T. Wang, H. S. P. Wong, C. H. Diaz, and J. Cai, "Cold CMOS as a power-performance-reliability booster for advanced FinFETs," in *Proc. IEEE Symp. VLSI Technol.*, Jun. 2020, pp. 1–2, doi: 10.1109/VLSITechnology18217.2020.9265065.

[4] A. Beckers, F. Jazaeri, and C. Enz, "Characterization and modeling of 28-nm bulk CMOS technology down to 4.2 K," *IEEE J. Electron Devices Soc.*, vol. 6, pp. 1007–1018, Mar. 2018, doi: 10.1109/JEDS.2018.2817458.

[5] H. Bohuslavskyi, A. G. M. Jansen, S. Barraud, V. Barral, M. Casse, 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 Lett.*, vol. 40, no. 5, pp. 784–787, May 2019, doi: 10.1109/LED.2019.2903111.

[6] A. Beckers, F. Jazaeri, and C. Enz, "Theoretical limit of low temperature subthreshold swing in field-effect transistors," *IEEE Electron Device Lett.*, vol. 41, no. 2, pp. 276–279, Feb. 2020, doi: 10.1109/LED.2019.2963379.

[7] H.-C. Han, F. Jazaeri, A. D'Amico, A. Baschirotto, E. Charbon, and C. Enz, "Cryogenic characterization of 16 nm FinFET technology for quantum computing," in *Proc. IEEE 47th Eur. Solid State Circuits Conf. (ESSCIRC)*, Sep. 2021, pp. 71–74, doi: 10.1109/ESSCIRC53450.2021.9567747.

[8] H.-C. Han, F. Jazaeri, A. D'Amico, Z. Zhao, S. Lehmann, C. Kretzschmar, E. Charbon, and C. Enz, "Back-gate effects on DC performance and carrier transport in 22 nm FDSOI technology down to cryogenic temperatures," *Solid-State Electron.*, vol. 193, Jul. 2022, Art. no. 108296. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038110122000685, doi: 10.1016/j.sse.2022.108296.

[9] A. Beckers, D. Beckers, F. Jazaeri, B. Parvais, and C. Enz, "Generalized Boltzmann relations in semiconductors including band tails," *J. Appl. Phys.*, vol. 129, no. 4, Jan. 2021, Art. no. 045701, doi: 10.1063/5.0037432.

[10] J. Wang and M. Lundstrom, "Does source-to-drain tunneling limit the ultimate scaling of MOSFETs?" in *IEDM Tech. Dig.*, Dec. 2002, pp. 707–710, doi: 10.1109/IEDM.2002.1175936.

[11] K.-H. Kao, T. R. Wu, H.-L. Chen, W.-J. Lee, N.-Y. Chen, W. C.-Y. Ma, C.-J. Su, and Y.-J. Lee, "Subthreshold swing saturation of nanoscale MOSFETs due to source-to-drain tunneling at cryogenic temperatures," *IEEE Electron Device Lett.*, vol. 41, no. 9, pp. 1296–1299, Sep. 2020, doi: 10.1109/LED.2020.3012033.

[12] K. Yilmaz, B. Iniguez, F. Lime, and A. Kloes, "Cryogenic temperature and doping analysis of source-to-drain tunneling current in ultrashort-channel nanosheet MOSFETs," *IEEE Trans. Electron Devices*, vol. 69, no. 3, pp. 1588–1595, Mar. 2022, doi: 10.1109/TED.2022.3145339.

[13] S. Datta, *Quantum Transport: Atom to Transistor*. Cambridge, U.K.: Cambridge Univ. Press, 2005. 10.1017/CBO9781139164313.

[14] Y. A. Kruglyak, "A generalized Landauer-Datta-Lundstrom electron transport model," *Russ. J. Phys. Chem. A*, vol. 88, no. 11, pp. 1826–1836, Feb. 2014, [Online]. Available:https://link.springer.com/article/10.1134/S0036024414110119, doi: 10.1134/S0036024414110119.

[15] G. Hiblot, Q. Rafhay, F. Boeuf, and G. Ghibaudo, "Analytical relationship between subthreshold swing of thermionic and tunnelling currents," *Electron. Lett.*, vol. 50, no. 23, pp. 1745–1747, 2014, doi: 10.1049/el.2014.3206.

[16] Y.-K. Lin, J. P. Duarte, P. Kushwaha, H. Agarwal, H.-L. Chang, A. Sachid, S. Salahuddin, and C. Hu, "Compact modeling source-to-drain tunneling in sub-10-nm GAA FinFET with industry standard model," *IEEE Trans. Electron Devices*, vol. 64, no. 9, pp. 3576–3581, Sep. 2017, doi: 10.1109/TED.2017.2731162.

[17] Z.-H. Liu, C. Hu, J.-H. Huang, T.-Y. Chan, M.-C. Jeng, P. Ko, and Y. Cheng, "Threshold voltage model for deep-submicrometer MOSFETs," *IEEE Trans. Electron Devices*, vol. 40, no. 1, pp. 86–95, Jan. 1993, 10.1109/16.249429.

[18] S. Sze and K. K. Ng, *Physics of Semiconductor Devices*. Hoboken, NJ, USA: Wiley, 2006, doi: 10.1002/0470068329.

[19] A. Beckers, F. Jazaeri, and C. Enz, "Cryogenic MOS transistor model," *IEEE Trans. Electron Devices*, vol. 65, no. 9, pp. 3617–3625, Sep. 2018, doi: 10.1109/TED.2018.2854701.

[20] N. G. Tarr, D. J. Walkey, M. B. Rowlandson, S. B. Hewitt, and T. W. MacElwee, "Short-channel effects on MOSFET subthreshold swing," *Solid-State Electron.*, vol. 38, no. 3, pp. 697–701, Mar. 1995, doi: 10.1016/0038-1101(94)00147-8.

[21] B. Ray and S. Mahapatra, "Modeling of channel potential and subthreshold slope of symmetric double-gate transistor," *IEEE Trans. Electron Devices*, vol. 56, no. 2, pp. 260–266, Feb. 2009, doi: 10.1109/TED.2008.2010577.