Study of Cryogenic MOSFET Sub-threshold Swing using *Ab Initio* Calculation

Tom Jiao, Edwin Antunez, and Hiu Yung Wong, *Senior Member, IEEE*

***Abstract*—The abnormal subthreshold swing (SS) in Silicon Metal-Oxide Semiconductor Field Effect Transistor (MOSFET) at cryogenic temperature is commonly attributed to band tail (BT) conduction. The cryogenic SS does not scale with the temperature, T, for T < 50K and it is observed to saturate at 10mV/dec ~ 20mV/dec at low T in most experiments. Hitherto, only analytical studies have been conducted for BT and its properties. It is not clear how much of its effect can be eliminated should there be an ideal manufacturing technology. In this paper, by using robust *ab initio* calculation with quantum transport, we have successfully calculated the BT in a Si nanowire (NW) and studied its characteristic length. By analyzing the transport properties of the NW with various gate lengths, L_G, at various temperatures, it is observed that for L_G < 20nm, the tunneling current dominates, and for L_G > 20nm, the BT current dominates at 3K. It is found that, in a *perfect* nanowire (as a gedanken experimental device), an SS as low as 1.4mV/dec can be achieved at 3K for 15 orders of magnitudes of current change with a minimum of 0.42mV/dec (L_G = 50nm). This also justifies the results in a recent experiment in which a very low SS (3.4mV/dec at 5.5K) was obtained. Moreover, it is also shown that for the 2nm node (L_G ~ 15nm), direct S/D tunneling will set the ultimate limit of SS at 3K.**

***Index Terms*—*Ab initio* Calculation, Cryogenic, Quantum Transport, Si Nanowire, Technology Computer-Aided Design (TCAD)**

I. Introduction

Cryogenic silicon MOSFETs below 77K are becoming more critical due to their role in quantum computer controlling electronics [1] and space exploration [2]. It is also envisaged that quantum computing at 4.2K may be possible on semiconductor chips [3]. Therefore, it is crucial to understand and model the physics of cryogenic Si MOSFETs down to 4.2K and below [4]-[10]. Among all cryogenic device physics, the abnormal subthreshold swing (SS) in MOSFETs may be the most interesting one. The SS of a MOSFET is expected to scale with temperature ($SS \sim nkT/q$, where n, k, T, and q are the ideality factor, Boltzmann constant, temperature, and elementary charge, respectively) when the current is dominated by thermionic emission. However, various experiments have shown that SS saturates at 10mV/dec to 20mV/dec at $T < 50K$ [7]-[9]. As a result, for $n = 1$, the SS cannot reach 0.36mV/dec and 0.26mV/dec for 4.2K and 3K, respectively.

There are two possible explanations for the abnormal SS. One attributes this to the existence of oxide/channel interface traps [4][10] and this was supported by its correlation with the increase in 1/f noise at cryogenic temperatures [11]. Another theory attributes the phenomenon to the existence of the Si band tail (BT) [7]-[9]. When the Si is an infinite perfect crystal, the band edge should be sharp and there are no states in the forbidden gap. Band-tail states thus exist when the Si is finite such as in a nanowire (in both the radial and channel directions) and due to the termination of Si at the gate insulator/Si interface. BT states also exist due to defects or dopant fluctuations in Silicon [12]. As the technology improves and also undoped channel is used in advanced technology, BT due to defects (in channel bulk and at the insulator/channel interface) and dopants are expected to reduce. It is thus important to understand if the abnormal SS will be solved in an undoped nanowire with ideal terminations (although the existence of gate dielectric cannot be eliminated in reality). It should also be noted that quantum mechanically, BT states are just a natural extension of the ideal conduction band (CB), and thus, BT state conduction is an extension of thermionic emission. *In this paper, band-tail current refers to the conduction due to the extra conduction states in the otherwise ideal bandgap.*

Hitherto, there are only analytical studies of Si MOSFET BT [7]-[9]. Abnormal SS due to interface traps can be modeled using Technology Computer-Aided-Design (TCAD) simulations [4]. But to understand the effect of the BT, it requires computational-intensive *ab initio* calculations coupled with non-equilibrium green's function approach (NEGF) at cryogenic temperature with the BT calculated automatically.

It should also be noted that the abnormal SS reduces the I_{ON}/I_{OFF} ratio which prevents the achievement of ultra-low standby power at cryogenic temperatures. However, due to its low leakage, direct S/D tunneling becomes important at cryogenic temperatures [14]. In Ref. [14], NEGF with analytical band structure without considering BT state was used to show that direct S/D tunneling can saturate the SS. It is

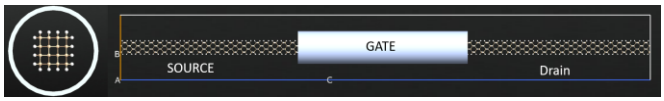

Fig. 1. The structure of the Silicon nano-wire simulated (L_G = 10nm is shown). Left: Traverse cross-section. Right: Longitudinal cross-section.

Manuscript received August 13, 2023. This material is based upon work supported by the NSF under Grant No. 2046220.T. Jiao, E. Antunez, and H. Y. Wong are with the M-PAC Lab, Department of Electrical Engineering, San Jose State University, San Jose, CA 95192 USA. (corresponding author: Hiu Yung Wong; e-mail: hiuyung.wong@sjsu.edu).

desirable to understand if it is still necessary to combat abnormal SS at cryogenic temperatures and if direct S/D tunneling will dominate in advanced technologies.

Therefore in this paper, we use *ab initio* calculation with NEGF to study the cryogenic SS behavior of Si nanowires. Various gate lengths (L_G) from 5 nm to 50 nm are used to deconvolve the effect of direct S/D tunneling and BT current. Hydrogen-passivated Si is constructed at the interface to form an ideal insulator/Si interface. A small cross-section nanowire is used to represent the worst-case BT due to the finite Si crystal.

II. *Ab Initio* Calculation Setup

QuantumATK [15] is used to create an n-type Silicon nanowire (NW) in the <100> direction and perform *ab initio* calculation (***Fig. 1***). 4×4 Si atoms are created in the transverse direction, which represents almost the most BT states one might have due to finite Si lattice. The approach in Ref. [13] is used to ensure good convergence in cryogenic *ab initio* calculation. Unlike Ref. [13], the channel is undoped (instead of p-type doped) with source and drain n-type doped to 5×10^{20} cm^{-3}. A vacuum is formed between the gate and the silicon nanowire which results in an equivalent oxide thickness (EOT) of about 1.6nm. The Si atoms are passivated with hydrogen using sp^3 hybridization. Therefore, this represents the most perfect insulator/channel interface one can have. However, such termination is still not ideal and cannot replace the infinite Si crystal.

The structure and LDOS are calculated using LCAO with GGA for exchange-correlation and PseudoDojo for pseudopotential. The calculated bandgap is about 3.2 eV due to the small diameter but is probably underestimated as it is known that other computationally intensive functionals such as HSE06 are required to obtain a correct bandgap [16]. However, the computation time is too long. Provided that this study does not cover interband mechanisms, this is expected to be sufficient, as demonstrated in [17]. We also note that the bandgap accuracy may be improved if DFT-1/2 is used [18]. To ensure good convergence [13], an initial calculation with 1000K energy-broadening is used, followed by a smaller

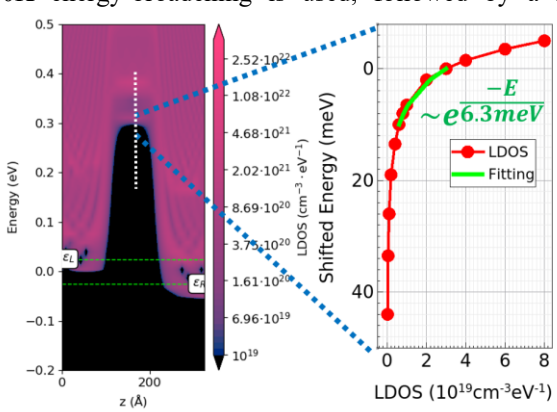

Fig. 2. Left: The local density of state (LDOS) plot at the CB region of L_G = 10nm when V_G = 0V and V_{DS} = 50mV. Right: 1-D plot of LDOS (x-axis) as a function of electron energy (y-axis) by taking the values along the white-dotted line at the middle of the device. Note that the energy is shifted so that the energy with LDOS = 3×10^{19}cm^{-3}eV^{-1} is 0meV.

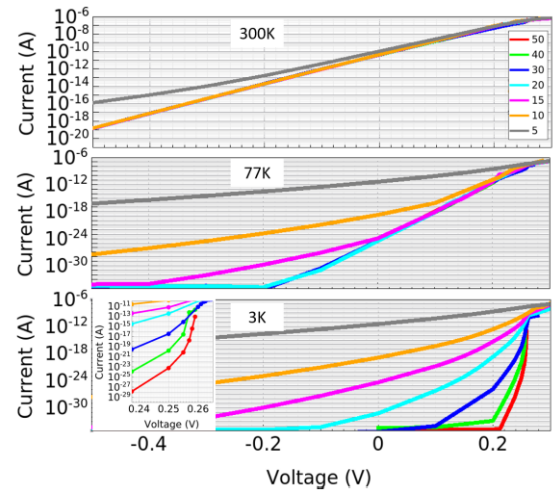

Fig. 3. I_DV_G of Si NW with various L_G and V_D = 50mV at 300K, 77K, and 3K, respectively. The inset at bottom-left shows the low S.S. region at 3K.

broadening until 5K broadening is reached. Energy broadening is a numeric technique to improve convergence and is shown in [13] that 5K or lower is necessary to obtain a correct result in T = 3K simulation. The reader may refer to [13] for details. It should be noted that the energy-broadening temperature is just a numerical technique and is not the same as the temperature of the device. The number of sampling points in the transport direction is set to 1000 to help convergence and the parallel conjugate gradient is chosen for the Poisson solver to reduce the calculation time. The drain voltage (V_{DS}) is set to 50mV in I_DV_G calculations. High V_{DS} (saturation mode) is not studied to avoid possible SS degradation due to high V_{DS}. Moreover, low V_{DS} has much better convergence than high V_{DS}. This approach is also believed to be valid for studying tunneling current because kT/q is much smaller than 50 mV at 3K.

7 gate lengths (L_G = 5 nm, 10 nm, 15 nm, 20 nm, 30 nm, 40 nm, 50 nm) have been simulated. The gate work function is adjusted such that when L_G = 15nm, the leakage is ~11nA/µm at V_{DS} = 50mV. This is similar to the requirement for the 2nm node in IRDS 2022 [19]. Although the V_{DS} for I_{off} calculation (50mV) is different from that in IRDS (0.65V), I_{off} is expected to change only a few times based on Ref. [20] which is used to derive the numbers in IRDS.

III. Calculation Results

Fig. 2 shows the Local Density of States (LDOS) of L_G = 10nm when V_G = 0V. By setting the CB edge at LDOS = 3×10^{19}cm^{-3}eV^{-1}, the BT has a characteristic length of 6.3meV. ***Fig. 3*** shows the I_DV_G curves of the NW with various L_G at T = 300K, 77K, and 3K, respectively, where T is the electronic temperature. At 300K, all devices have the same leakage current, except for L_G = 5nm, probably due to direct S/D tunneling. Thermionic emission over the barrier dominates except for L_G = 5nm at 300K. Even at 77K, thermionic emission still dominates for $L_G \geq 15$nm ($I > 10^{-25}$A). However, at 3K, thermionic emission becomes insignificant as the leakage current depends on L_G. It is not clear if this is due to direct S/D tunneling or BT current.

To identify the leakage mechanism, we first assume direct S/D tunneling dominates at 3K. Then the transmission (T_{max}) at the

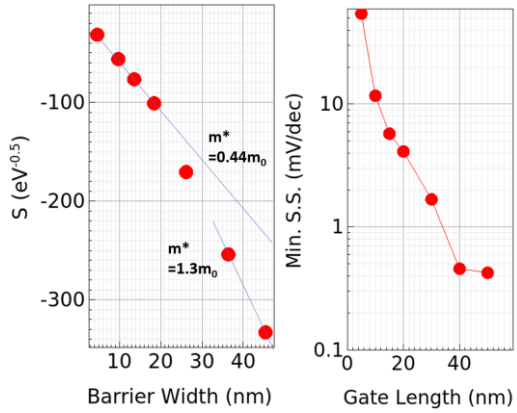

Fig. 4. Left: Extract S as a function of the tunneling width at the maximum transmission. Right: Minimum SS for various gate length at 3K.

maximum spectral current density, is extracted for a given L_G at every V_G in the subthreshold regime. It is worth noting that T_{max} defined here is not the maximum transmission as the spectral current depends on the source/drain electron distribution and also the transmission. The barrier width, a (dependent on L_G), and barrier height, V (dependent on V_G), for the energy corresponding to T_{max} are extracted and substituted into the Wentzel–Kramers–Brillouin (WKB) transmission equation for a rectangular barrier, which is $T_{max} = \mathrm{A}\exp(-2a\sqrt{2m^*V}/\hbar)$, where A, m^*, and $\hbar$ is a pre-exponential factor, electron effective mass, and reduced Planck constant, respectively. By transforming the equation into $\ln T_{max} = -2a\sqrt{2m^*}\sqrt{V}/\hbar + \ln A$, one may extract $S = -2a\sqrt{2m^*}/\hbar$ from the $\ln T_{max}$-$\sqrt{V}$ plot for each a. One can then plot S vs. a to extract the corresponding m^*. ***Fig. 4*** shows that when L_G < 20nm, $m^* = 0.44m_0$, which is reasonable. This is similar to the value calculated in a 1nm diameter <100> NW in Ref. [21]. Therefore, direct S/D tunneling is the major leakage mechanism and dominates at L_G < 20nm at 3K. On the other hand, $m^* = 1.3m_0$ is needed to fit the data points for $L_G \geq$ 40nm if we assume direct S/D tunneling is the dominant leakage mechanism. However, $m^* = 1.3m_0$ is likely unreasonable and m^* is believed not to be dependent on L_G. Therefore, it is believed that the leakage mechanism is dominated by BT current for $L_G \geq$ 40nm where direct S/D tunneling is negligible.

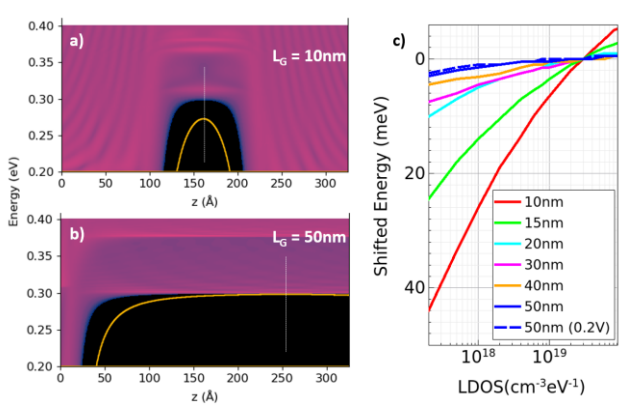

Fig. 5. Left: The local density of state (LDOS) plots at the CB region of L_G = 10nm (a) and L_G = 50nm (b) when V_G = 0V and V_{DS} = 50mV. The CB edge has LDOS = 3×10^{19}cm^{-3}eV^{-1}. The yellow contour lines show location when LDOS = 10^{18}cm^{-3}eV^{-1}. Right: 1-D plot of LDOS (x-axis) as a function of electron energy (y-axis) at the device center of various gate lengths at V_G = 0V. V_G = 0.2V for 50nm is also shown. Note that the energy is shifted so that the energy with LDOS = 3×10^{19}cm^{-3}eV^{-1} is 0meV.

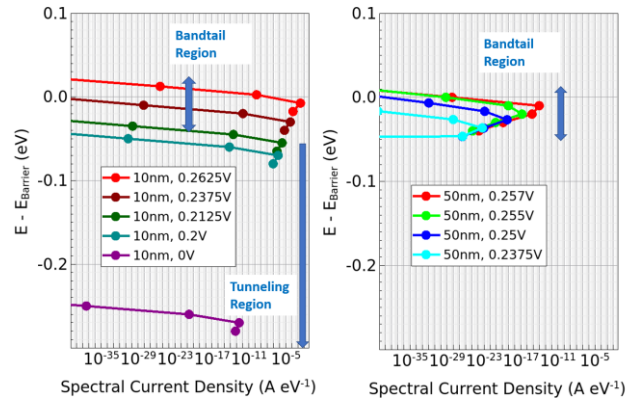

Fig. 6. Spectral current density for 10nm (left) and 50nm (right) devices at various biases at 3K. The energy is referenced to the average channel barrier at LDOS = 3×10^{19}cm^{-3}eV^{-1}. Tunneling and BT current regions are identified.

To understand the nature of the BT states, the LDOS of different L_G are plotted in ***Fig. 5***. It is believed that some of the BT states are induced "laterally" by the CB from the S/D (same lateral distance between the yellow contour line and the band edge in 10nm and 50nm devices). For small L_G, LDOS increases at the middle of the channel due to the "S/D-induced states". ***Fig. 6*** plots the spectral current density (SCD) at various biases for L_G = 10nm and 50nm cases. BT conduction region is identified as where the SCD peak is near the channel barrier. ***Fig. 4*** also plots the minimum SS for each L_G at 3K. Combining with the results in ***Fig. 2*** and ***Fig. 5***, BT conduction can be seen clearly: 1) For L_G = 10nm, the BT characteristic length (***Fig. 2***) is consistent with Ref. [8] and its BT region (V_G > 0.2125V in Fig. 6) SS is consistent with the experiments (10mV/dec to 20mV/dec) [7]-[9]. 2) For L_G = 50nm, SS is between 0.42mV/dec and 0.8mV/dec when V_G changes from 0.255V to 0.258V, which corresponds to the band tail region in ***Fig. 6*** where the SCD peak is near the channel CB edge and the energy corresponding to T_{max} sweeps from LDOS ~ 10^{16}cm^{-3}eV^{-1} to LDOS ~ 3×10^{19}cm^{-3}eV^{-1}.

Although the transport is believed to be dominated by BT current for L_G = 50nm, 0.42mV/dec can still be achieved, which is less than double the Boltzmann limit of 0.26mV/dec. For L_G = 50nm, the average SS is about 1.4mV/dec in 15 orders of magnitude of change of current (***Fig. 3***). Therefore, with an ideal interface and undoped channel, very small SS can be achieved (instead of 10mV/dec to 20mV/dec observed in most experiments). Indeed, a recent experiment shows that it is possible to achieve a SS of 3.4mV/dec at 5.5K [22]. It should also be noted that $n = 1.08$ is extracted for L_G = 50nm at 300K. Therefore, the lower limit of SS can be even smaller.

IV. Conclusion

By using *ab initio* calculation and NEGF, the BT is calculated in undoped Si NW with an ideal interface. It is found that BT current dominates at large L_G. However, a minimum SS of 0.42mV/dec and an average SS of 1.4mV/dec are realized for L_G = 50nm gedanken-experimental device. For the 2nm technology node, direct S/D tunneling can be the limiting factor on the cryogenic SS.

REFERENCES

[1] E. Charbon, F. Sebastiano, A. Vladimirescu, H. Homulle, S. Visser, L. Song, and R M. Incandela, "Cryo-CMOS for Quantum Computing," 2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, 2016, pp. 13.5.1-13.5.4, doi: 10.1109/IEDM.2016.7838410.

[2] R. L. Patterson, A. Hammoud, J. E. Dickman, S. Gerber, M. Elbuluk and E. Overton, "Electronics for Deep Space Cryogenic Applications," Proceedings of the 5th European Workshop on Low Temperature Electronics, Grenoble, France, 2002, pp. 207-210, doi: 10.1109/WOLTE.2002.1022482.

[3] K. Ono, T. Mori, and S. Moriyama, "High-temperature operation of a silicon qubit," Sci Rep 9, 469 (2019). https://doi.org/10.1038/s41598-018-36476-z

[4] Prabjot Dhillon, Nguyen Cong Dao, Philip H. W. Leong, and Hiu Yung Wong, "TCAD Modeling of Cryogenic nMOSFET ON-State Current and Subthreshold Slope," 2021 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), 2021, pp. 255-258, doi: 10.1109/SISPAD54002.2021.9592586.

[5] Hiu Yung Wong, "Calibrated Si Mobility and Incomplete Ionization Models with Field Dependent Ionization Energy for Cryogenic Simulations," IEEE 2020 International Conference on Simulation of Semiconductor Processes and Devices, pp193-196, doi: 10.23919/SISPAD49475.2020.9241599.

[6] Prabjot Dhillon and Hiu Yung Wong, "A Wide Temperature Range Unified Undoped Bulk Silicon Electron and Hole Mobility Model," in IEEE Transactions on Electron Devices, doi: 10.1109/TED.2022.3152471. (2022)

[7] A. Beckers, "Cryogenic MOSFET Modeling for Large-Scale Quantum Computing," Ph.D. dissertation, Faculté des sciences et techniques de l'ingénieur, Lausanne, EPFL, 2021.

[8] A. Beckers, F. Jazaeri and C. Enz, "Theoretical Limit of Low Temperature Subthreshold Swing in Field-Effect Transistors," in IEEE Electron Device Letters, vol. 41, no. 2, pp. 276-279, Feb. 2020, doi: 10.1109/LED.2019.2963379.

[9] A. Beckers, F. Jazaeri and C. Enz, "Revised Theoretical Limit of Subthreshold Swing in Field-effect Transistors," arXiv:1811.09146, 2019.

[10] 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.

[11] H. Oka, T. Matsukawa, K. Kato, S. Iizuka, W. Mizubayashi, K. Endo, T. Yasuda, and T. Mori, "Toward Long-coherence-time Si Spin Qubit: The Origin of Low-frequency Noise in Cryo-CMOS," VLSI Technology Symposium 2020.

[12] E. O. Kane, "Thomas-Fermi Approach to Impure Semiconductor Band Structure," Physical Review, vol. 131, no. 1, pp. 79–88, 1963.

[13] Tom Jiao and Hiu Yung Wong, "Robust Cryogenic Ab-initio Quantum Transport Simulation for LG=10nm Nanowire," Solid-State Electronics, Volume 197, 2022, 108440, doi.org/10.1016/j.sse.2022.108440

[14] K. -H. Kao et al., "Subthreshold Swing Saturation of Nanoscale MOSFETs Due to Source-to-Drain Tunneling at Cryogenic Temperatures," in IEEE Electron Device Letters, vol. 41, no. 9, pp. 1296-1299, Sept. 2020, doi: 10.1109/LED.2020.3012033.

[15] QuantumATK™ Reference Manual Version S-2021.06, June 2021.

[16] H. Y. Wong, "Ab Initio Study of HfO2/Ti Interface VO/Oi Frenkel Pair Formation Barrier and VO Interaction With Filament," in IEEE Transactions on Electron Devices, vol. 69, no. 9, pp. 5130-5137, Sept. 2022, doi: 10.1109/TED.2022.3188227.

[17] Shiqi Liu et al., "Performance Limit of Gate-All-Around Si Nanowire Field-Effect Transistors: An Ab Initio Quantum Transport Simulation," Phys. Rev. Applied 18, 054089, 2022.

[18] Luiz G. Ferreira, Marcelo Marques, and Lara K. Teles, "Slater half-occupation technique revisited: the LDA-1/2 and GGA-1/2 approaches for atomic ionization energies and band gaps in semiconductors," AIP Adv., 1(3):032119, 2011. doi:10.1063/1.3624562.

[19] (2022). International Roadmap for Devices and Systems—More Moore. [Online]. Available: https://irds.ieee.org/images/files/pdf/2022/2022IRDS_MM.pdf

[20] G. Yeap et al., "5nm CMOS Production Technology Platform featuring full-fledged EUV, and High Mobility Channel FinFETs with densest 0.021µm2 SRAM cells for Mobile SoC and High Performance Computing Applications," 2019 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2019, pp. 36.7.1-36.7.4, doi: 10.1109/IEDM19573.2019.8993577.

[21] N. Neophytou et al., "Bandstructure Effects in Silicon Nanowire Electron Transport," arXiv: 0801.0125, 2008

[22] Y. Han et al., "Steep Switching Si Nanowire p-FETs With Dopant Segregated Silicide Source/Drain at Cryogenic Temperature," in IEEE Electron Device Letters, vol. 43, no. 8, pp. 1187-1190, Aug. 2022, doi: 10.1109/LED.2022.3185781.