# Fundamental Limit to Scaling Si Field-Effect Transistors Due to Source-to-Drain Direct Tunneling

Stanislav Markov, Yanho Kwok, Jun Li, Weijun Zhou, Yi Zhou, and Guanhua Chen

***Abstract*— How far can the miniaturization of metal–oxide–semiconductor field-effect transistors (MOSFETs) continue is a recurring question, essential to all aspects of digital technology. Recent claims of well-performing MOSFETs with gate lengths below 4 nm apparently defy the fundamental limit of source-to-drain direct tunneling (SDDT). Here, we investigate that limit by simulating gate-all-around Si nanowire FETs with gate lengths between 8 and 3 nm using the state-of-the-art atomistic quantum transport modeling. We find that at 3-nm gate length, the current is dominated by SDDT, resulting in large source–drain leakage and poor switching performance even if the gate modulates the potential barrier between the source and drain sufficiently well. However, at 6-nm gate-length SDDT barely starts to degrade the subthreshold characteristics at large drain bias, and the ballistic ON-/OFF-current ratio is $\sim 10^6$ with a subthreshold swing of 70 mV/decade, on par with contemporary Si technology. This means that in the best case, the technology roadmap could potentially be extended for several generations beyond the currently projected nodes. In addition, the results substantiate that the experimental devices with the claimed gate lengths below 6 nm in fact operate with a longer effective gate lengths.**

***Index Terms*— Atomistic simulation, density functional tight binding, metal–oxide–semiconductor field-effect transistor (MOSFET), quantum transport.**

## I. Introduction

TRANSISTOR density of integrated circuits has approximately doubled every two years for several decades now and has been traditionally associated with the scaling of the gate length by a factor of 0.7–0.8 across technology generations [1], [2]. Although leading manufacturers have already announced their readiness for the 10-nm technology node [3], the corresponding gate length is much larger, ∼19 nm, and this number does not scale below 12 nm in the latest editions of the technology roadmap [2]. This is not far from the fundamental physical limit imposed by source–drain direct tunneling (SDDT) in metal–oxide–semiconductor field-effect transistors (MOSFETs), predictions for which vary broadly between 11- and 4-nm gate length [4]–[6]. Therefore, it is important to accurately determining where that limit is, in order to stimulate further technological innovation and miniaturization beyond the current roadmap. The experimental work toward this goal has focused on the observation of temperature independent subthreshold swing (SS)—the change in gate voltage needed for a ten-fold change in the drain current [6]. This happens when SDDT starts to dominate electron transport over the intended thermionic emission (TE) above the gate-induced potential barrier [see Fig. 1(a)]. It has first been observed below 77 K in a shallow-junction bulk-Si MOSFET with 8-nm gate length [6], and later up to 300 K in a thin-body silicon-on-insulator MOSFET with 10-nm gate length and an estimated channel length of 5 nm [7]. Consistent with that a few theoretical studies found that SS deteriorates fast below 7-nm gate length and that the fraction of tunnel current through the source–drain barrier comprises between 40% and 90% of the OFF-current as gate-length scales from 10 to 4 nm [5], [8], [9]. In contrast, recent experimental reports of MOSFETs with gate lengths of 4, 3, and even 1 nm show surprisingly small SS between 65–85 mV/dec [10]–[13]. Some theoretical support for such remarkable results exists [14], [15], but an alternative explanation ascribes them to a longer effective channel length [12], [13]. Unfortunately, the lack of temperature or gate-length dependence, or any details of the underlying transport mechanisms, leaves the contradiction unresolved.

Here, we settle the above question by applying the state-of-the-art atomistic quantum transport modeling to gate-all-around (GAA) silicon nanowire (SiNW) MOSFETs, which remain the optimal device for ultimate scaling [1], [2], [9], despite the intense pursuit of alternative channel materials, architectures, and operational principles [16]–[21]. Our focus is specifically on the impact of SDDT on the subthresholds swing of a device design that exhibits excellent immunity from short channel effects. Aspects of technology that currently impede ultimate miniaturization—e.g., contact resistance, variability, and mobility degradation—are purposefully excluded,

Manuscript received December 30, 2018; revised January 19, 2019; accepted January 21, 2019. Date of publication February 5, 2019; date of current version February 22, 2019. This work was supported in part by the University Grants Council, Hong Kong, within the Area of Excellence on Theory, Modeling and Simulation of Emerging Electronics under Contract AoE/P-04/08 and in part by SZSTI Project under Grant JCYJ20150629151046882. The review of this paper was arranged by Editor J. Mateos. *(Corresponding author: Stanislav Markov.)*

S. Markov, Y. Kwok, W. Zhou, Y. Zhou, and G. Chen are with the Department of Chemistry, University of Hong Kong, Hong Kong (e-mail: figaro@hku.hk; ghc@everest.hku.hk).

J. Li is with the Department of Chemistry, University of Hong Kong, Hong Kong, and also with the HKU-Shenzhen Institute of Research and Innovation, Shenzhen, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TED.2019.2894967

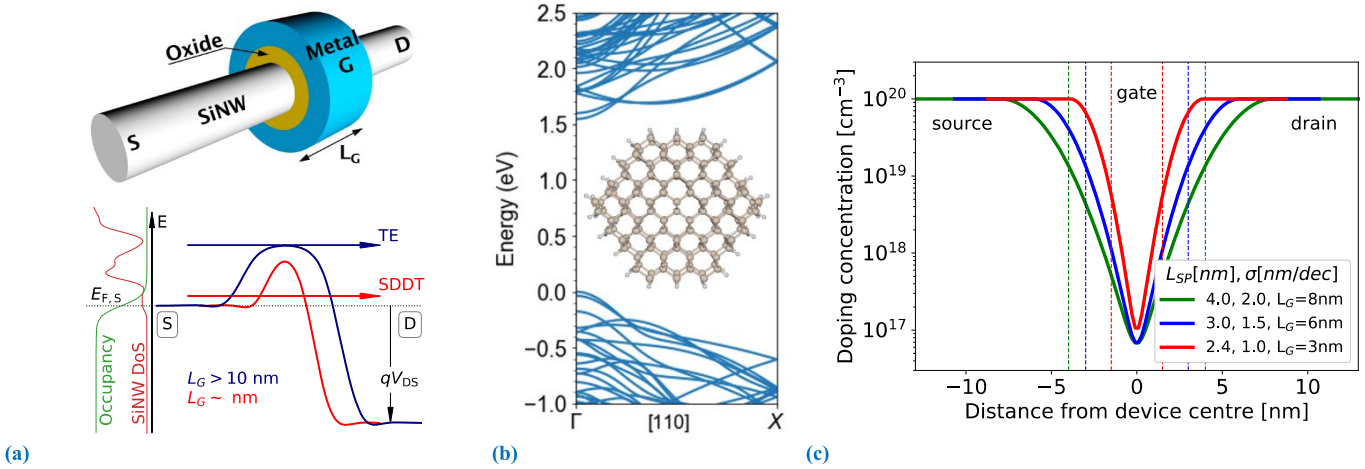


Fig. 1. Diagram of GAA SiNW MOSFET, transport mechanisms, and features of the simulation model. (a) Schematic of the transistor model, indicating source (S), drain (D), and gate (G) contacts. Hereafter, the [110]-SiNW axis is referred to as *x*-direction. Below, the band diagram to the right of the energy axis E depicts the potential barrier of two MOSFETs with different gate length $L_G$ and the associated dominant transport mechanism—TE, at larger $L_G$, and source–drain direct tunneling (SDDT) at $L_G$ approaching few nanometers. SiNW density of states and occupation probability at the source are shown to the left of the energy axis. (b) Band-structure and cross-sectional view of the atomic model of the SiNW. (c) Longitudinal profile of the donor concentration used in this paper, based on a Gaussian rolloff defined by the spacer width $L_{SP}$ and steepness $\sigma$ (as reported in the legend).

so that the obtained figures of merit represent aspirational goals for technology optimization and serve as a benchmark for emerging technologies and materials.

## II. Methodology

We model the transistors of three different gate lengths of 3, 6, and 8 nm, based on the same [110]-SiNW of approximately 2.2 nm diameter and an effective oxide thickness of 0.7 nm. This results in excellent control of the gate over the height of the potential barrier in the channel and enables us to determine the exact potential barrier width, at which SDDT starts dominating the subthreshold behavior of the transistor. Such insight cannot be obtained from experiment and has not been elucidated by modeling despite its importance. Our simulation approach couples the density functional tight-binding (DFTB) Hamiltonian to the nonequilibrium Green's function (NEGF) formalism self-consistently via the Poisson equation, which allows us to obtain the density matrix and transmission function and then calculate the electron current [22]–[24]. The choice of DFTB Hamiltonian makes it computationally affordable to treat atomic models of relevant size without involving material-related phenomenological parameters and ensures the highest accuracy at present. It is a significant advancement beyond the effective mass- and mode-space approximations [8], and beyond empirical tight binding [9]. However, its computational complexity makes it impossible to simulate the impact of device non-uniformities that are known to affect sub-threshold slope, like surface roughness and remote coulomb scattering centers [38], [39]. The coupling of DFTB with NEGF has been attempted in the past to model Si NW MOSFETs [15], [40], but it is only through a recent work on DFTB parameterization at the level of chemical elements that quantitative accuracy has been achieved both in terms of measured electronic structure and dielectric response in $Si/SiO_2$ system [25], [26]. This is reflected in Fig. 1(b), showing a cross section of the basic repeating unit used to construct the SiNW together with its band structure. Notable is the calculated band-gap of 1.545 eV which agrees well with the experimental value of 1.5 eV for a hydrogenated [110]-SiNW with a cross section of 2.5 nm [27]; in contrast, the density functional theory (DFT) in the generalized gradient approximation underestimates the gap for the same structure, yielding a gap of 1.2 eV despite the strong confinement [28]. The transport simulations are performed in the ballistic limit, without phonon scattering as our focus is on the subthreshold current, which is practically independent of phonon scattering even for longer channel lengths [29]. At the same time, we can establish the best possible on/off ratio for the selected nanowire cross section.

We construct the atomic models of the MOSFETs by repeating the unit shown as inset in Fig. 1(b). The resulting SiNWs consist of 3572, 6048, and 7344 atoms and their length is 18, 22, and 26 nm. Si atoms are described with spd basis [25], resulting in 25 432, 41 888, and 50 867 corresponding orbitals. In this paper, we use continuous donor-doping profile based on the virtual crystal approximation [32], rather than the explicit chemical doping used in [30]. The present choice significantly helps convergence, and at the same time avoids the intrinsic aspects of variability associated with the position of impurity atoms. The doping profile of each device—shown in Fig. 1(c)—assumes highly doped source and drain regions, with donor concentration of $10^{20}$ cm$^{-3}$, and a Gaussian rolloff from the edge of the spacer. The width of the spacer $L_{SP}$ and the steepness of the rolloff $\sigma$ are chosen to ensure donor concentration near the gate edge of approximately $10^{19}$ cm$^{-3}$ without having an excessively large concentration under the gate. The effective oxide thickness is realized via vacuum separation between the atomic model shown in Fig. 1(b) and a cylindrical Dirichlet boundary condition representing a metal gate. Further details of the simulation methodology are given in the Appendix.

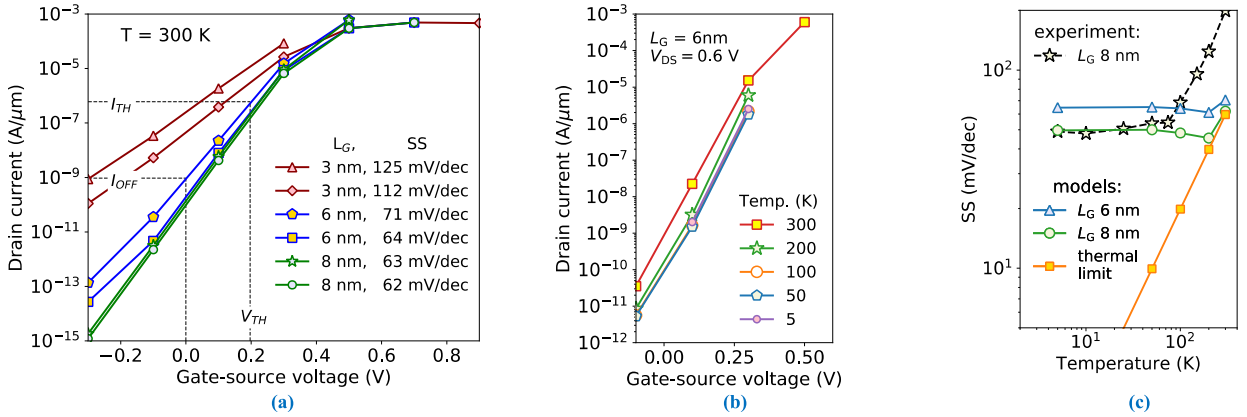


Fig. 2. Current–voltage characteristics and SS at different bias and temperature. (a) Transfer characteristics ($I_{DS}$ versus $V_{GS}$) of the MOSFETs at high drain bias $-V_{DS} = 0.6$ V (upper curve at given $L_G$: triangles, pentagons, and stars) and low drain bias $V_{DS} = 50$ mV (lower curve at given $L_G$: diamonds, squares, and circles). $V_{TH}$ indicates the threshold voltage, based on 600 nA/$\mu$m drain current of the 6-nm MOSFET at $V_{DS} = 0.6$ V. Notable are the SS of 63 mV/dec at $L_G = 8$ nm—close to the ideal 60 mV/dec at 300 K, and the twice bigger SS of 125 mV/dec at $L_G$ of 3 nm. The MOSFET with 6-nm $L_G$ exhibits an unexpectedly large change in SS when drain bias changes from low to high; also unexpected is the change in SS at low $V_{DS}$ and negative $V_{GS}$, both effects hinting at a change of dominant transport mechanisms. (b) Transfer characteristics of the 6-nm MOSFET at different temperatures and high drain bias, practically overlapping below 200 K, indicative of SDDT dominated transport. (c) Subthreshold swing versus temperature at $L_G = 6$ nm and $L_G = 8$ nm, showing a good agreement against reported experiment at 8 nm gate length [6]; the ideal temperature-dependent SS is labeled thermal limit.

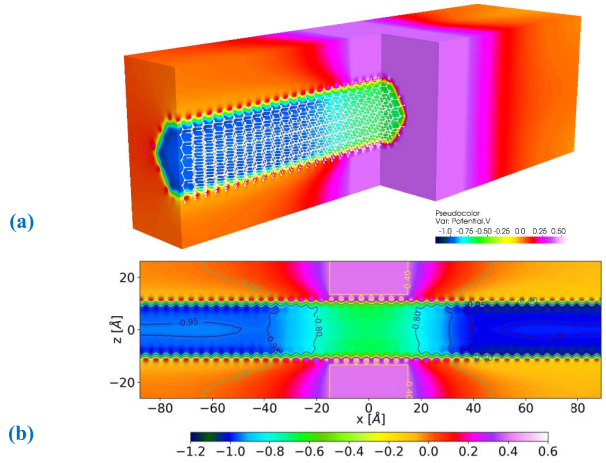


Fig. 3. Electrostatic potential in the simulation domain for 3-nm gate length; $V_{DS} = 50$ mV, $V_{GS} = 0.1$ V. (a) 3-D distribution of the potential. The cutout of the domain reveals the atomic structure of the SiNW model and the potential through the channel. (b) 2-D distribution of the potential in the ($xz$) plane through the SiNW axis.

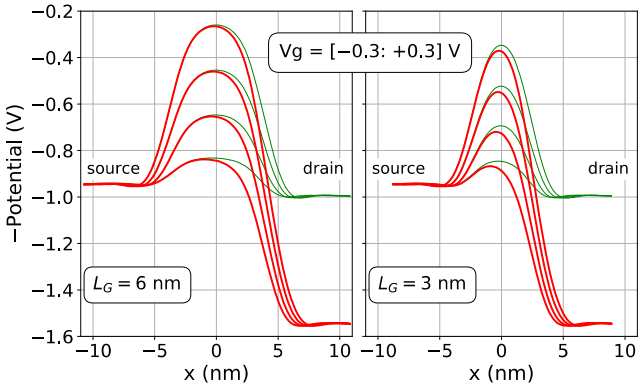


Fig. 4. Electrostatic potential through the axis of the SiNW for $L_G$ of 6 and 3 nm, for $V_{DS}$ of 0.6 V (red, thick lines) and 50 mV (green, thin lines). $V_{GS}$ varies in steps of 0.2 V between −0.3 and +0.3 V. There is a significant lowering of the peak potential barrier at high drain bias only at 3 nm $L_G$.

## III. Results and Discussion

Fig. 2(a) shows the current–voltage transfer characteristics ($I_{DS}$ versus $V_{GS}$) of the transistors at two drain biases, $V_{DS}$, of 50 mV and 0.6 V, and electronic temperature $T = 300$ K. The legend reports the SS for each curve and suggests an excellent subthreshold performance of the MOSFET with 8-nm gate length ($L_G$), having an SS of 63 mV/dec—close to the ideal SS of 60 mV/dec at 300 K. For $L_G = 6$ nm, the SS at high drain bias deteriorates significantly to 71 mV/dec, which is somewhat surprising, given that at low drain bias the SS remains rather small, 64 mV/dec. These figures should be compared to the 70 mV/dec reported for contemporary state-of-the art technology with a much longer gate length of ∼19 nm [3]. In contrast, the SS for 3 nm $L_G$ nearly doubles to 112 and 125 mV/dec at low and high drain bias, respectively. Regarding the switching performance, we obtain an ON-/OFF-current ratio close to $10^6$. The current at high gate bias is practically independent of the gate length, which is expected from the ballistic quantum transport formalism, and is around 600 $\mu$A/$\mu$m, or 4.1 $\mu$A, in line with previous NEGF calculations of SiNWs of similar cross section in the ballistic limit [20], [29]. Our focus however remains on the region of $V_{GS}$ below 0.3 V, where SDDT may be the reason of degraded subthreshold performance.

Fig. 2(b) shows the subthreshold part of the transfer characteristics of the 6-nm MOSFET for temperature between 5 and 300 K. The curves for 5, 50, and 100 K practically overlap, meaning that the TE is not the primary transport mechanism at these temperatures. This is confirmed in Fig. 2(c), showing the dependence of SS versus temperature. For 6-nm $L_G$ the SS is practically the same, 65 mV/dec, between 5 and 100 K. For 8-nm $L_G$, the simulated SS of 49 mV/dec between 5 and 50 K is in excellent agreement with the reported experiment for 8-nm $L_G$ for the corresponding temperatures [6]. It should be noted that the experimental

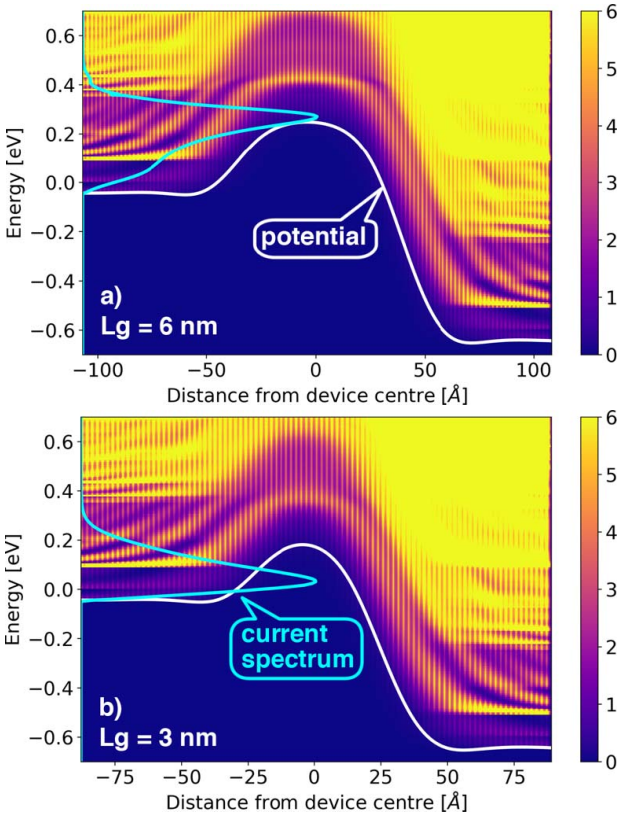


Fig. 5. Potential barrier profile through the nanowire axis and current spectrum superposed on the 2-D color map of the local density of states, for the MOSFETs with gate length of (a) 6 nm and (b) 3 nm, at $V_{DS}$ = 0.6 V, $V_{GS}$ = 0.1 V, and $T$ = 300 K. Energy reference is the Fermi level at the source contact. The potential is aligned with the bottom of the conduction band at the source. The current spectrum is normalized so that its peak reaches the 0 of the horizontal axis, allowing to visually comparing its alignment with the peak of the potential barrier. At 6-nm $L_G$ the current features substantially TE—the peak of the current spectrum is just above the peak of the potential barrier, while direct tunneling clearly dominates at 3-nm $L_G$—the peak of the current spectrum is close to the Fermi level at the source.

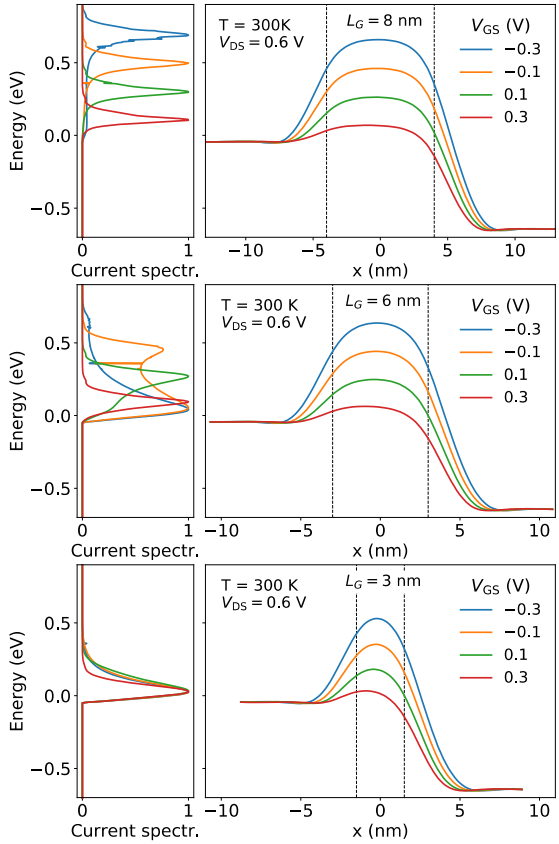


Fig. 6. Current spectrum and energy barrier profile along the Si NW axis at different gate lengths. Current spectrum is normalized by its peak to identify the dominant transport mechanism regardless of current magnitude. Vertical dashed lines indicate gate boundaries. Transport mechanism changes from TE at 8-nm $L_G$ (top) to SDDT at 3-nm $L_G$ (bottom), with transition happening at 6-nm $L_G$ (middle). At 6-nm $L_G$ both components are significant and SDDT becomes dominant at negative gate bias.

result pertains to a specialized bulk-like MOSFET structure with electrically variable ultrashallow junction, employing two different gates—a long upper gate that induces an inversion region serving as ultrashallow source/drain junctions, and an 8 nm lower gate that defines $L_G$ and modulates the conduction between these junctions. The device nevertheless exhibits poorer electrostatics in comparison to the present GAA nanowire model, which is reflected in the larger SS for temperatures above 50 K, where thermionic current is the principle leakage mechanism between source and drain. The specific features in the SS dependence on temperature are explained with the interplay between TE and SDDT components, and how the current spectrum is affected by the temperature.

To decouple the effect of gate-length reduction on TE current from the effect on SDDT, we first analyse the electrostatics within the channel and the corresponding potential barrier profile between the source and drain. Fig. 3(a) gives an overall 3-D perspective of the potential distribution in the simulation domain for the 3-nm MOSFET at 50-mV drain–source bias and 0.1-V gate bias, while Fig. 3(b) shows a more detailed 2-D view of the potential through the middle of the simulation domain. Fig. 4 compares the potential profiles along the axis of the nanowire between the MOSFETs with 6 and 3 nm gate lengths, for a set of bias conditions. When comparing these profiles, we recall that any increase in the peak of the barrier translates to a reduction in the TE current at a rate of 60 mV/dec (at room temperature)—the rate at which electron state occupancy decays for energies far from the Fermi level. However, we note in Fig. 4 that the difference in barrier height for different gate length at fixed bias amounts to 110 mV at most. This would imply less than 100-fold increase in the current for 3-nm $L_G$, but as seen in Fig. 2(a), at $V_{GS} = -0.3$ V, the increase is four orders of magnitude over the current for corresponding bias at 6-nm $L_G$. Such a discrepancy is due to enhanced SDDT, due to the narrowing of the potential barrier clearly seen in Fig. 4.

Fig. 5 shows the local density of states, the potential barrier from drain to source, and the normalized current spectrum in the same two devices at 0.6 V drain–source bias and 0.1-V gate bias. While the local density of states has similar features between the two devices, there is a crucial difference in their current spectrum. Specifically, at 6-nm gate length,

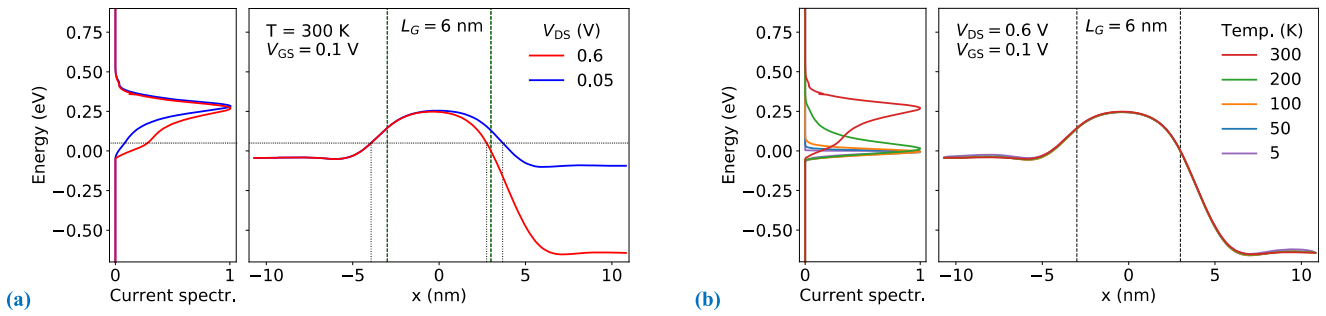


Fig. 7. (a) Substantial increase in the tunneling component due to the narrowing of the potential barrier that accompanies the increase of drain bias from 50 mV to 0.6 V at 6-nm $L_G$. The thin dotted lines identify the width of the tunnel barrier at the energy level where current spectrum features an emerging SDDT peak. The barrier width is 0.9 nm smaller—changing from 7.6 to 6.7 nm. (b) Current spectrum of the 6-nm MOSFET at different temperatures and fixed bias, showing that reducing the temperature to 200 K or lower suffice to make SDDT the dominant current component.

the peak of the current spectrum is above the peak of the potential barrier, consistent with the intended TE transport. In complete contrast, at 3-nm gate length, the current spectrum peaks around the Fermi level of the source, corresponding to SDDT transport. Therefore, the large subthreshold swing and OFF-current at 3-nm $L_G$ are not due to the inability of the gate to modulate the height of the potential barrier, but to its limited ability to control the width of the barrier, to which the direct tunneling is exponentially sensitive.

A closer observation of the current spectrum in Fig. 5(a) reveals a significant tunneling component even at 6-nm gate length, judged by the area of the current spectrum below the peak of the potential barrier. This suggests that 6 nm is the gate length at which SDDT start to significantly affect the subthreshold characteristics. The issue becomes clearer in Fig. 6, showing the current spectrum and potential barriers for each of the three transistors for a set of $V_{GS}$. At the two extremes are the cases of 3- and 8-nm gate length. In the former case, $L_G = 3$ nm, all peaks of the current spectrum cluster near the Fermi level, reflecting the dominant SDDT current. In the latter case, $L_G = 8$ nm, the peaks of the current spectrum are above the corresponding peaks of the potential, reflecting TE transport, with some tunneling enhancing the current. It is at 6-nm gate length that the photograph is in-between. The current spectrum shows a growing fraction of direct tunneling as gate voltage is reduced and an eventual dominance of the SDDT as $V_{GS}$ becomes negative. That the 6-nm gate length is critical is evident from its sensitivity to the drain bias too. The increase of $V_{DS}$ from 50 mV to 0.6 V shortens the barrier by almost 1 nm, from ~8 to ~7 nm. This leads to the substantial increase in SDDT current, as shown in Fig. 7(a), and explains the relatively large difference in SS between low and high drain bias at this $L_G$, i.e., 64 mV/dec versus 71 mV/dec, despite a negligible variation in the peak of the potential barrier. Finally, in Fig. 7(b), we show the potential barrier and current spectrum for different temperatures for 6-nm $L_G$. The subthreshold current at 200 K and below is entirely dominated by SDDT, as anticipated from the temperature dependence of its SS.

## IV. Conclusion

In summary, we demonstrated that a GAA SiNW MOSFET can be potentially scaled down to 6-nm gate length without significant degradation of the subthreshold behavior despite a clear manifestation of SDDT. The SS of 71 mV/dec at high drain bias and DIBL of 43 mV/V match or better the state-of-the-art technology [3], and therefore may result in tolerable leakage power density even at much higher transistor density. Therefore, 6-nm gate length should be on the technology roadmap. However, further shortening, the gate enters a regime entirely dominated by SDDT, which will make it hardly possible to control the leakage power by the usual transistor design parameters. In relation to the experimental reports of MOSFETs with 4-nm gate length and below [10]–[13], it is worth noting that the potential barrier is effectively elongated due to doping profile, and fringing fields specific to the gate shape or the high permittivity of gate spacers, as discussed in [12] and [13]. Unfortunately, such strategy in general does not warrant overall transistor scaling corresponding to the small gate length. Therefore, we conclude that the scaling of Si MOSFETs below 6-nm physical gate length and a correspondingly scaled gate pitch is not viable. Our results serve as an important reference for the development of phenomenological models that will facilitate the technology and circuit advancement toward this limit [31] and a benchmark in the exploration of alternative channel materials and designs [21].

## Appendix
## Computational Approach

We treat the electrical current in the MOSFETs as quantum transport in an open system, by using a well-established approach, consisting of coupling an atomistic Hamiltonian to the NEGF formalism and to the Poisson equation in a self-consistent loop, in order to obtain the density matrix and transmission function, and then calculate current [23], [33], [34]. We use self-consistent density-functional-based tight-binding (DFTB) Hamiltonian to represent the atomic model [22], which is only recently gaining popularity in the context of open system modeling, due to the emergence of accurate parameterization for relevant chemical elements. Below, we elaborate on a few details to the extent necessary to understand its applicability to device modeling, as well as to distinguish it from other atomistic approaches.

DFTB derives as an approximation of density-functional theory by the decomposition of the total energy functional in three terms: $E = E_{BS}[n_0] + E_{SCC}[n_0, \delta n] + E_{REP}(n_0)$.

These terms correspond to: (BS)—band-structure energy arising from a Hamiltonian built on a reference electron density $n_0(\boldsymbol{r})$, (SCC)—energy due to Coulomb and exchange-correlation (XC) interactions of the second order fluctuation in charge density $\delta n(\boldsymbol{r})$, (REP)—repulsive potential energy—a correction term accounting for ion-ion interaction, and double-counting and XC terms. The corresponding terms of the Hamiltonian are $\mathbf{H}^{(0)}$, $\mathbf{H}^{(2)}$, and $\mathbf{H}_{\mathrm{R}}$. The matrix elements of $\mathbf{H}^{(0)}$ and the overlap matrix $\mathbf{S}$ are precomputed with *ab initio* DFT using a linear combination of nonorthogonal optimized atomic orbitals as basis and subject to the two-center approximation, and are tabulated versus interatomic distance, as reported in [25]. $\mathbf{H}^{(2)}$ depends on the overall electrostatics of the device and drives the need for the self-consistent loop, while $\mathbf{H}_{\mathrm{R}}$ is ignored in here since it represents only a correction to the total energy of a closed system.

The matrix elements of $\mathbf{H}^{(2)}$ would generally have the form $h^{(2)}_{\mu\nu} = \langle \phi_\mu \,|\, \delta V_H(\boldsymbol{r}) \,|\, \phi_\nu \rangle$, where $\delta V_H(\boldsymbol{r})$ is the Hartree potential, solution of the Poisson equation for a given fluctuation density $\delta n(\boldsymbol{r})$. However, explicit use of the atomic basis $\{\phi_\mu\}$ is avoided in DFTB to reduce numerical complexity. Instead, the fluctuation density is decomposed into atomic contributions: $\delta n(\boldsymbol{r}) = \sum_\mu \Delta q_\mu F_\mu(\boldsymbol{r})$, where the induced Mulliken population $\Delta q_\mu$ is defined by subtracting the reference atomic charges $q^{(0)}_\mu$ from the Mulliken population $q_\mu = \sum_{i\in\mu}[\boldsymbol{\varrho}\boldsymbol{S}]_{ii}$, $\boldsymbol{\varrho}$ being the density matrix, and $F_\mu$ being a projection function for atom $\mu$. $\delta V_H(\boldsymbol{r})$ is also projected on the atomic sites to which $\mu$ and $\nu$ correspond, so that evaluation of $h^{(2)}_{\mu\nu}$ reduces to $h^{(2)}_{\mu\nu} = \langle \phi_\mu \,|\, ((v_\mu + v_\nu)/2) \,|\, \phi_\nu \rangle = (1/2)s_{\mu\nu}(v_\mu + v_\nu)$, $s_{\mu\nu}$ being the precomputed matrix element of the overlap matrix, and the projections $v_{\mu,\nu}$, known as onsite shifts, are independent of $\boldsymbol{r}$ [23]. Once that $\mathbf{H}^{(2)}$ and therefore $\mathbf{H}$ are known, the density matrix $\boldsymbol{\varrho}$ is found by applying the NEGF equations, as elaborated elsewhere [23], and the Mulliken populations are updated accordingly.

The discussion so far is summarized in Fig. 8, showing the self-consistent loop for finding $\mathbf{H}$, and the general equations that govern it. Last thing to clarify here is the projection function $F_\mu = (\tau^3/8\pi)e^{-\tau R}$, which corresponds to an exponentially decaying normalized spherical density typical of an $s$-type radial wave-function; $R$ is the distance from the atom to which orbital $\mu$ belongs, and $\tau = 3.2U_\mu$, where $U_\mu$ is obtained *ab initio* for each chemical element as the second derivative of the total atomic energy with respect to the orbital occupation number.

Two salient points must be understood from the above. First, the tabulation of $\mathbf{H}^{(0)}$ and $\mathbf{S}$ warrants a good computational efficiency, while the parameterization of the pseudoatomic orbitals allows to overcome the band-gap deficiency of standard DFT Hamiltonians, and to correct for the various approximations underlying DFTB [25], [35]. Notably, the parameterization happens at atomic level, within the pseudoatomic orbitals for computing $\mathrm{H}^{(0)}$. Second, the inclusion of $\mathbf{H}^{(2)}$ captures both charge transfer due to bond asymmetry and induced polarization arising from electrostatic boundary conditions. In this way, the method improves transferability of the tight-binding scheme, not limited to bulk or

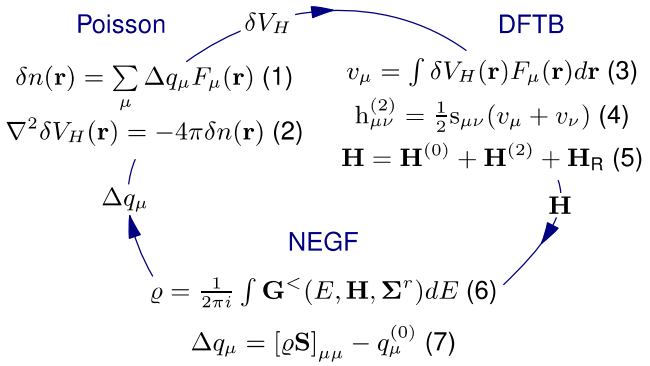


Fig. 8. Summary of the computational approach—self-consistent loop for determining the device Hamiltonian $\mathbf{H}$. All equations, except (2) and (6), are pertinent to the formulation with DFTB Hamiltonian. The iterative sequence starts with an initial guess for the atomic charge fluctuations $\Delta q_\mu$ ($\mu$ labels an atomic orbital), which enables us to evaluate the corresponding Hartree potential $\delta V_H$, and construct the second-order term $\mathbf{H}^{(2)}[\mathrm{h}^{(2)}_{\mu\nu}]$ of $\mathbf{H}$. With $\mathbf{H}$ and the self-energies $\mathbf{\Sigma}^r$ we find the lesser Green's function $\mathbf{G}^<$, which yields the density matrix $\varrho$ upon integration. Refined values for $\Delta q_\mu$ are obtained via Mulliken population analysis. The reference Hamiltonian and overlap matrices, $\mathbf{H}^{(0)}$ and $\mathbf{S}[\mathrm{s}_{\mu\nu}]$ are precomputed in DFTB. $F_\mu$ is a projection function between atomic bases and real space.

a particular device structure, and yields the electronic contribution to the dielectric response of the system as well—both electronic structure and permittivity emerge from the model with a good accuracy without phenomenological parameters [25], [26], and the Poisson equation is solved in vacuum. Details of the in-house developed implementation used in this paper, Lodestar, may be found elsewhere [36].

Note that the simulations at high gate bias—when the potential barrier is hardly perceptible—converge very difficult, and one may obtain unphysical solution, with a potential that is not flat near the leads. This is a problem even though the simulations employ double contour integration within the NEGF formalism [37]. Therefore, we have excluded such simulation points from the $I$–$V$ curves in Fig. 2(a) of the main text.

## Acknowledgment

Computations were performed using facilities offered by Information Technology Services, the University of Hong Kong, Hong Kong.

## References

[1] K. J. Kuhn, "Considerations for ultimate CMOS scaling," *IEEE Trans. Electron Devices*, vol. 59, no. 7, pp. 1813–1828, Jul. 2012, doi: 10.1109/TED.2012.2193129.

[2] IRDS. (2017). *The International Roadmap for Devices and Systems*. Accessed: Jun. 13, 2018. [Online]. Available: https://irds.ieee.org

[3] C. Auth *et al.*, "A 10 nm high performance and low-power CMOS technology featuring $3^{rd}$ generation FinFET transistors, Self-Aligned Quad Patterning, contact over active gate and cobalt local interconnects," in *IEDM Tech. Dig.*, Dec. 2017, pp. 1–29, doi: 10.1109/IEDM.2017.8268472.

[4] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, "Device scaling limits of Si MOSFETs and their application dependencies," *Proc. IEEE*, vol. 89, no. 3, pp. 259–288, Mar. 2001, doi: 10.1109/5.915374.

[5] 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.

[6] H. Kawaura, T. Sakamoto, and T. Baba, "Observation of source-to-drain direct tunneling current in 8 nm gate electrically variable shallow junction metal–oxide–semiconductor field-effect transistors," *App. Phys. Let.*, vol. 76, no. 25, pp. 3810–3812, 2000, doi: 10.1063/1.126789.

[7] J. Lolivier *et al.*, "Experimental characterization of source-to-drain tunneling in 10 nm SOI devices," in *Proc. IEEE Int. SOI Conf.*, Oct. 2005, pp. 34–35, doi: 10.1109/SOI.2005.1563525.

[8] N. Seoane and A. Martinez, "A detailed coupled-mode-space non-equilibrium Green's function simulation study of source-to-drain tunnelling in gate-all-around Si nanowire metal oxide semiconductor field effect transistors," *J. Appl. Phys.*, vol. 114, no. 10, p. 104307, 2013, doi: 10.1063/1.4820390.

[9] M. Luisier, M. Lundstrom, D. A. Antoniadis, and J. Bokor, "Ultimate device scaling: Intrinsic performance comparisons of carbon-based, InGaAs, and Si field-effect transistors for 5 nm gate length," in *IEDM Tech. Dig.*, Dec. 2011, pp. 11.2.1–11.2.4, doi: 10.1109/IEDM.2011.6131531.

[10] M. Li *et al.*, "Sub-10 nm gate-all-around CMOS nanowire transistors on bulk Si substrate," in *Proc. Symp. VLSI Technol.*, Jun. 2009, pp. 94–95.

[11] S. Migita, Y. Morita, M. Masahara, and H. Ota, "Electrical performances of junctionless-FETs at the scaling limit (LCH= 3 nm)," in *IEDM Tech. Dig.*, Dec. 2012, pp. 8.6.1–8.6.4, doi: 10.1109/IEDM.2012.6479006.

[12] S. Migita, Y. Morita, T. Matsukawa, M. Masahara, and H. Ota, "Experimental demonstration of ultrashort-channel (3 nm) junctionless FETs utilizing atomically sharp V-grooves on SOI," *IEEE Trans. Nanotechnol.*, vol. 13, no. 2, pp. 208–215, Mar. 2014, doi: 10.1109/TNANO.2013.2296893.

[13] S. B. Desai *et al.*, "$MoS_2$ transistors with 1-nanometer gate lengths," *Science*, vol. 354, no. 6308, pp. 99–102, 2016, doi: 10.1126/science.aah4698.

[14] Y.-R. Jhan, V. Thirunavukkarasu, C.-P. Wang, and Y.-C. Wu, "Performance evaluation of silicon and germanium ultrathin body (1 nm) junctionless field-effect transistor with ultrashort gate length (1 nm and 3 nm)," *IEEE Electron Device Lett.*, vol. 36, no. 7, pp. 654–656, 2015, Jul. doi: 10.1109/LED.2015.2437715.

[15] L. Ansari, B. Feldman, G. Fagas, J.-P. Colinge, and J. C. Greer, "Subthreshold behavior of junctionless silicon nanowire transistors from atomic scale simulations," *Solid-State Electron.*, vol. 71, pp. 58–62, May 2012, doi: 10.1016/j.sse.2011.10.021.

[16] F. Schwierz, "Graphene transistors," *Nature nanotechnol.*, vol. 5, no. 7, pp. 487–496, 2010, doi: 10.1038/nnano.2010.89.

[17] B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, and A. Kis, "Single-layer $MoS_2$ transistors," *Nature Nanotechnol.*, vol. 6, no. 3, pp. 147–150, 2011, doi: 10.1038/nnano.2010.279.

[18] A. D. Franklin *et al.*, "Carbon nanotube complementary wrap-gate transistors," *Nano Lett.*, vol. 13, no. 6, pp. 2490–2495, 2013, doi: 10.1021/nl400544q.

[19] J. P. Llinas *et al.*, "Short-channel field-effect transistors with 9-atom and 13-atom wide graphene nanoribbons," *Nature Commun.*, vol. 8, no. 1, 2017, Art. no. 633, doi: 10.1038/s41467-017-00734-x.

[20] S. Manipatruni, D. E. Nikonov, and I. A. Young, "Beyond CMOS computing with spin and polarization," *Nature Phys.*, vol. 14, no. 4, pp. 338–343, 2018, doi: 10.1038/s41567-018-0101-4.

[21] G. Iannaccone, F. Bonaccorso, L. Colombo, and G. Fiori, "Quantum engineering of transistors based on 2D materials heterostructures," *Nature Nanotechnol.*, vol. 13, no. 3, pp. 183–191, 2018, doi: 10.1038/s41565-018-0082-6.

[22] M. Elstner *et al.*, "Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties," *Phys. Rev. B, Condens. Matter*, vol. 58, no. 11, p. 7260, 1998, doi: 10.1103/PhysRevB.58.7260.

[23] A. Pecchia, G. Penazzi, L. Salvucci, and A. Di Carlo, "Non-equilibrium Green's functions in density functional tight binding: Method and applications," *New J. Phys.*, vol. 10, no. 6, p. 065022, 2008, doi: 10.1088/1367-2630/10/6/065022.

[24] C. Yam, L. Meng, Y. Zhang, and G. Chen, "A multiscale quantum mechanics/electromagnetics method for device simulations," *Chem. Soc. Rev.*, vol. 44, no.7, pp. 1763–1776, 2015, doi: 10.1039/C4CS00348A.

[25] S. Markov, B. Aradi, C. Y. Yam, H. Xie, T. Frauenheim, and G. Chen, "Atomic level modeling of extremely thin silicon-on-insulator MOSFETs including the silicon dioxide: Electronic structure," *IEEE Trans. Electron Devices*, vol. 62, no. 3, pp. 696–704, Mar. 2015. doi: 10.1109/TED.2014.2387288.

[26] S. Markov *et al.*, "Permittivity of oxidized ultra-thin silicon films from atomistic simulations," *IEEE Electron Device Lett.*, vol. 36, no. 10, pp. 1076–1078, Oct. 2015, doi: 10.1109/LED.2015.2465850.

[27] D. D. D. Ma, C. S. Lee, F. C. K. Au, S. Y. Tong, and S. T. Lee, "Small-diameter silicon nanowire surfaces," *Science*, vol. 299, no. 5614, pp. 1874–1877, 2003, doi: 10.1126/science.1080313.

[28] R. J. Bondi, S. Lee, and G. S. Hwang, "First-principles study of the structural, electronic, and optical properties of oxide-sheathed silicon nanowires," *ACS Nano*, vol. 5, no. 3, pp. 1713–1723, 2011, doi: 10.1021/nn102232u.

[29] M. Luisier and G. Klimeck, "Atomistic full-band simulations of silicon nanowire transistors: Effects of electron-phonon scattering," *Phys. Rev. B, Condens. Matter*, vol. 80, no. 15, p. 155430, 2009, doi: 10.1103/PhysRevB.80.155430.

[30] C. Y. Yam *et al.*, "A multi-scale modeling of junctionless field-effect transistors," *Appl. Phys. Lett.*, vol. 103, no. 6, p. 062109, 2015, doi: 10.1063/1.4817911.

[31] Y.-K. Lin *et al.*, "Compact modeling source-to-drain tunneling in sub-10-nm GAA FinFET with industry standard model," *IEEE Electron Device Lett.*, vol. 64, no. 9, pp. 3576–3581, Sep. 2017, doi: 10.1109/TED.2017.2731162.

[32] N. J. Ramer and A. M. Rappe, "Virtual-crystal approximation that works: Locating a compositional phase boundary in $Pb(Zr_1\text{-}xTi_x)O_3$," *Phys. Rev. B, Condens. Matter*, vol. 62, no. 2, p. R743, 2000, doi: 10.1103/PhysRevB.62.R743.

[33] J. Taylor, H. Guo, and J. Wang, "Ab initio modeling of quantum transport properties of molecular electronic devices," *Phys. Rev. B, Condens. Matter*, vol. 63, no. 24, p. 245407, 2001, doi: 10.1103/PhysRevB.63.245407.

[34] M. Luisier, A. Schenk, W. Fichtner, and G. Klimeck, "Atomistic simulation of nanowires in the $sp^3d^5s^*$ tight-binding formalism: From boundary conditions to strain calculations," *Phys. Rev. B, Condens. Matter*, vol. 74, no. 20, p. 205323, 2006, doi: 10.1103/PhysRevB.74.205323.

[35] F. Bi *et al.*, "Enhanced photovoltaic properties induced by ferroelectric domain structures in organometallic halide perovskites," *J. Phys. Chem. C*, vol. 121, no. 21, pp. 11151–11158, 2017, doi: 10.1021/acs.jpcc.7b03091.

[36] C. Yam, L. Meng, Y. Zhang, and G. Chen, "A multiscale quantum mechanics/electromagnetics method for device simulations," *Chem. Soc. Rev.*, vol. 44, no. 7, pp. 1763–1776, 2015, doi: 10.1039/C4CS00348A.

[37] R. Li *et al.*, "A corrected NEGF + DFT approach for calculating electronic transport through molecular devices: Filling bound states and patching the non-equilibrium integration," *Chem. Phys.*, vol. 336, nos. 2–3, pp. 127–135, 2007, doi: 10.1016/j.chemphys.2007.06.011.

[38] C. Buran, M. Pala, M. Mouis, and S. Poli, "Full-3D real-space simulation of surface-roughness effects in double-gate MOSFETs," in *Proc. 13th Int. Workshop Comput. Electron.*, May 2009, pp. 1–4, doi: 10.1109/IWCE.2009.5091115.

[39] S. Poli, M. G. Pala, and T. Poiroux, "Full quantum treatment of remote coulomb scattering in silicon nanowire FETs," *IEEE Trans. Electron Devices*, vol. 56, no. 6, pp. 1191–1198, Jun. 2009, doi: 10.1109/TED.2009.2019380.

[40] F. Sacconi *et al.*, "Electronic and transport properties of silicon nanowires," *J. Comput. Electron.*, vol. 6, pp. 329–333, Sep. 2007, doi: 10.1007/s10825-006-0138-y.


Authors' photographs and biographies not available at the time of publication.