Solid-State Electronics 199 (2023) 108524

Contents lists available at ScienceDirect

Solid State Electronics

journal homepage: www.elsevier.com/locate/sse

Forked-contact and dynamically-doped nanosheets to enhance Si and 2D-material devices at the limit of scaling

Aryan Afzalian *, Zubair Ahmed, Julien Ryckaert

imec, Leuven, Belgium

ARTICLE INFO

Handling Editor Francisco Gamiz

Keywords:
CMOS
CGP scaling
Si
2D materials
DFT-NEGF
Ab-initio
Nanosheet
Dynamic-doping

ABSTRACT

We propose a novel Forked-Contact, Dynamically-Doped Multigate transistor as ultimate scaling booster for both Si and 2D materials in aggressively-scaled nanosheet devices. Using accurate dissipative DFT-NEGF atomistic-simulation fundamentals and cell layout extrinsics, we demonstrate superior and optimal device characteristics and invertor energy – delays down to sub-30-nm pitches, i.e., a 10 nm scaling boost compared to the nanosheet MOSFET references, regardless of the material system used. This gain is linked to a more compact architecture but does not change the material-specific fundamental gate-length limit that we also assess here. By switching from Si to 2D materials, however, an additional 5 nm reduction in gate length scaling could be enabled.

1. Introduction

The Dynamically-Doped (D2) Field-Effect Transistor is a novel device architecture that scales better than its MOSFET nanosheet (*NS*) counterpart, owing to the suppression of ungated extensions (spacers) from the device Contacted Gate-pitch (CGP) equation (Fig. 1a) [1,2]. What used to be the *NS* chemically doped extensions are now electrically and dynamically-doped by the gate, i.e., a part of the channel. Hence, for a given CGP, the channel length L in the D2FET is twice the spacer length (L_{SPACER}) longer than L of a standard MOSFET, as it benefits from the full distance between the source (S) and drain (D) contact pads. The gate length L_G value could even be larger than L, if the gate would overlap over the contact region of length L_C (Fig. 1a).

For a single-gate (SG) single-sheet device, this can simply be enabled by having the gate contact on the side opposite to the contacts, i.e., using for instance a top-contact and an individually back-gated transistor [1]. To enable a D2 tri-gate with stacked sheets, however, we propose here a doubled forked structure (E2), where the sheets are connected to a forked gate on one side and to forked S & D contacts on the other side (Fig. 1 and Fig. 2). Our simulation results show, as expected, that such a multigate E2D2 architecture enables a better electrostatic control and improved drive current at scaled CGP, especially for Si where the film thickness can be relaxed, compared to the SG-D2 transistor (Fig. 1b).

We report here on the impact of the multigate E2D2 architecture innovation on intrinsic-device and loaded-invertor performance, when pitch is scaled well below 30 nm using accurate dissipative DFT-NEGF atomistic-simulation fundamentals and cell-layout extrinsics. The E2D2 architecture is benchmarked to *NS* MOSFETs using both Si and 2 emerging 2D transition metal dichalcogenide (TMD) monolayer (1ML) materials – one, WS_2, with predicted fundamental drive similar to that of Si, the other, HfS_2, featuring an enhanced fundamental drive current [1] – as test vehicles.

2. Methods

Current – Voltage ($I_D(V_G)$), and intrinsic device capacitances (C_{Gi}) for Si and 2D TMD E2D2 and standard *NS* references are simulated using our atomistic NEGF solver ATOMOS, including electron–phonon (e-ph) scattering [1,2]. For the 2D materials, density functional theory (DFT) simulations were carried out to compute the electronic properties of the materials, including Hamiltonian matrix elements. These matrix elements were then imported in ATOMOS. We used QUANTUM ESPRESSO [3] with the generalized gradient approximation (GGA) implementation of DFT with the optB86b exchange–correlation functionals [4], followed by a Wannierization step to obtain an orthogonal localized-orbital model. The simulations were directly performed with a real-space (RS)

* Corresponding author.
E-mail address: aryan.afzalian@imec.be (A. Afzalian).

https://doi.org/10.1016/j.sse.2022.108524

Available online 11 November 2022

NEGF model. The procedure is fully detailed in [1,2].

For Si, we used a spds* tight-binding (TB) model [5]. A mode-space (MS) approach was used to fasten the simulations for the largest structures [6]. For MS, the scattering self-energies were not computed in mode space using a form-factor method [7,8], but directly computed in real space using the up-converted lesser and greater Green's functions. The self-energies were then down-converted back to mode space using a dedicated pool of workers to efficiently perform the task in parallel. It was verified that the MS results were in good agreement with those of the RS model (Fig. 3).

From these simulations the intrinsic single-sheet device fundamental performance vs CGP can be assessed (Fig. 4). For each CGP, a full device optimization is made including film thickness (t_S) scaling for Si and extension doping for the *NS*. The detrimental impact of quantum confinement, including a ballistic ratio decrease due to an increase of the electron–phonon wave-function overlap [9,10], dark space [11], and source-to-drain direct tunneling are included in our quantum transport solver.

For computing stacked-invertor energy-delay products (Fig. 5), the extracted extrinsic capacitance of the cell layout C_{cell} and the backend-of-line load, C_{BK} are used (Fig. 2). C_{cell} values are reported in Fig. 6a. The number of stacked sheets (n_S) used is computed to allow a total stack height of 60 nm for all devices. n_S is the same for E2D2 and *NS* of a same material ($n_S = 4$ for Si and 5 for the TMDs owing to their 1ML thickness of about 0.6 nm [1]). The available width for a single sheet, *W*, in our 5-track E2D2 layout cell is 12 nm. The standard *NS* layout is described in [12] and *W* is 12 nm as well.

3. Results

Owing to its 10 nm extended gate length at same CGP, the E2D2 SS and, hence, I_{ON} at fixed I_{OFF} are superior compared to those of the *NS*, as CGP is scaled below 30 nm for all materials. The E2D2 C_{Gi} are however larger at fixed CGP, the net effect being that the E2D2 optimal intrinsic delay is comparable to that of its *NS* counterpart, but shifted towards smaller CGPs by about 10 nm, i.e., $2 \times L_{SPACER}$ (Fig. 4). Hence the E2D2 architecture enables a significant scaling boost.

For Si, the optimal *NS* and E2D2 delays are obtained at CGP = 36, and 26 nm, respectively, i.e., $L_G = 10$ nm and $t_S = 3$ nm in both cases. For the 2D materials, a further 5 nm scaling boost is observed, and optimal

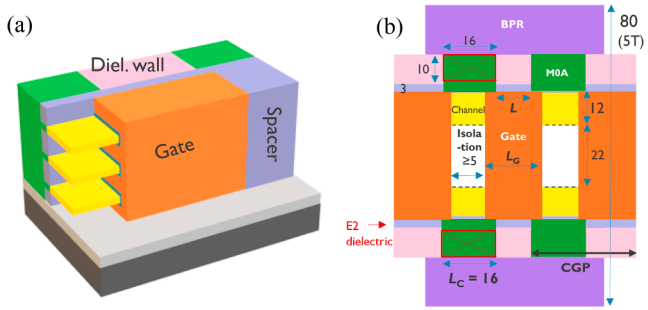

Fig. 2. E2D2 device structure. a) 3D view showing the doubled forked (E2) structure, b) cell layout of the 5-track (cell height = 80 nm) E2D2 invertor cell with buried power rail (BPR). The technological dimensions, we assumed for this study are indicated in the figure. We assumed $L_G = L$. $L_C = 16$ nm, $L_{SPACER} = 5$ nm. The width of an individual sheet is $W = 12$ nm. The P/N separation is 22 nm. For the gate oxide, we assumed a 2 nm hafnium oxide with $\varepsilon_R = 15.6$. The gate-stack metal thickness is 6 nm. The channel (white region in Fig. 1a) is intrinsic. The contact regions (in blue in Fig. 1a) are degenerately doped. $n_S = 4$ for Si and 5 for 2D. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

delays are achieved at CGP = 31 and 21 nm for the *NS* and E2D2 respectively, corresponding to $L_G = 5$ nm in both cases. For the TMDs, the 21-nm E2D2 CGP corresponds to the case where CGP is only limited by the contacts (L_C and the minimum isolation spacing, IS, required to separate subsequent pads, assuming IS = L_{SPACER} (Fig. 2)), hence ultimate gate scaling has been achieved. Further CGP reductions may still be achieved by scaling the contacts.

Next, we investigate the switching energy vs delay (EDP) of high-performance stacked E2D2 and *NS* loaded inverters for different CGPs at various V_{DD} (Fig. 5). For HfS_2 E2D2 and *NS* invertors, the optimal EDP is achieved at CGP = 21, $L_G = 5$ nm and CGP = 31, $L_G = 5$ nm respectively. We obtain a similar result for the WS_2 case (not shown here). For Si E2D2 and *NS* invertors, the optimal EDP is achieved at CGP = 26, $L_G = 10$ nm and CGP = 36, $L_G = 10$ nm respectively. Any further attempt to scale CGP by scaling L_G beyond this optimal value results in significant performance reduction (for the TMD E2D2 it is simply not possible to further scale CGP with L_G scaling). These results further confirm the 10-nm improved scalability, we obtained from the intrinsic device

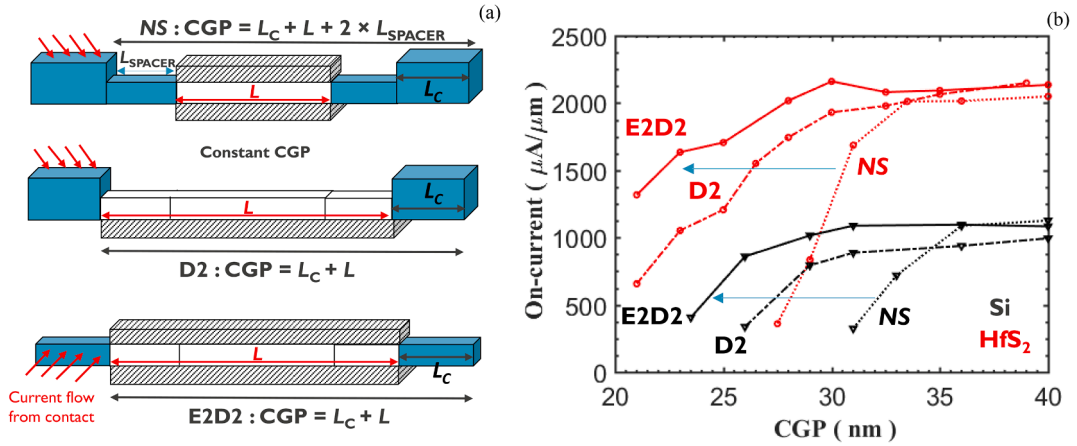

Fig. 1. A) side view schematic of a single-sheet multigate conventional *NS* (Top), D2 (middle), and E2D2 transistor (Bottom) with same CGP. b) On-current (I_{ON}), at fixed I_{OFF}, vs CGP for Si and for HfS_2 2D monolayer *NS, D2* and E2D2 architectures from ab-initio – NEGF transport simulations using ATOMOS [1]. For the *NS*, the gate length L_G = CGP – L_C – 2 × L_{SPACER}, while, for E2D2, L_G = CGP – L_C. $L_C = 16$ nm, $L_{SPACER} = 5$ nm. I_{ON} is normalized by the gate perimeter. $I_{OFF} = 5$nA/μm. V_{DD} = 0.6 V. Gate equivalent oxide thickness, EOT, = 0.5 nm. The channel (white regions in Fig. 1a) is intrinsic. The contact regions (in blue in Fig. 1a) are degenerately doped with $N_{SD} = 1$ and 4×10^{20} cm^{-3} for the Si and 2D sheets, respectively. The spacer regions (for *NS* devices only, also in blue in Fig. 1a) are doped as the contact regions, excepted for the $L_G = 5$ nm 2D *NS*, where an optimal value of 2×10^{20} cm^{-3} was used as a trade-off between short-channel effects and source starvation [1]. The junctions were simulated using abrupt doping profiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

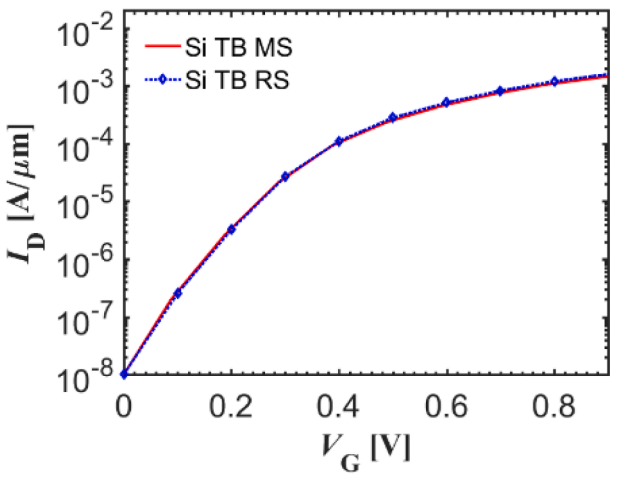

Fig. 3. $I_D(V_G)$ characteristics (including e-ph scattering) of a Si *NS* nMOS transistor computed from the RS TB model and from the optimized (cleaned) MS TB Model. Typical MS to RS speedups are > 100 ×. t_S = 3 nm. V_D = 0.6 V.

performance and delays (Fig. 4). The E2D2-invertor improved EDP performance compared to that of the *NS* is mostly linked to the reduced C_{BK} owing to CGP scaling.

This is confirmed in Fig. 6, where the loaded-invertor EDPs are shown for the 3 material cases at their optimal CGP and L_G values with and without C_{BK} included in the load. Regardless of the material system used, without C_{BK} (Fig. 6a), the E2D2 performance are similar to that of their *NS* counterparts, while they are enhanced when C_{BK} is included (Fig. 6b) (the E2D2 and *NS* devices also share the same optimal L_G of 10 nm for Si and 5 nm for the TMDs).

4. Conclusions

We proposed a compact E2D2 multigate architecture that enables sub-30-nm CGP, i.e., an improved 2 × L_{SPACER} pitch scaling, compared to a *NS* reference, owing to the suppression of ungated extensions from the CGP equation. This E2D2 scaling benefits were measured in term of similar intrinsic performance and optimal delay but at a 10 nm reduced CGP. A similar conclusion was found comparing E2D2 and *NS* stacked-invertor cells. For backend-loaded invertors, the E2D2 EDP performance is further enhanced due to CGP and, hence, C_{BK} reduction. Similar relative benefits were observed regardless of the material system used. Compared to Si, a mature 2D-material technology could potentially enable an extra 5-nm CGP scaling boost. This boost arises from a smaller L_G, both for the E2D2 and *NS* architectures, with same or improved performance, if respectively WS_2, a material with a fundamental drive similar to Si, or HfS_2, a higher mobility material, were used.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Aryan Afzalian has patent licensed to imec. Julien Ryckaert has patent licensed to imec.

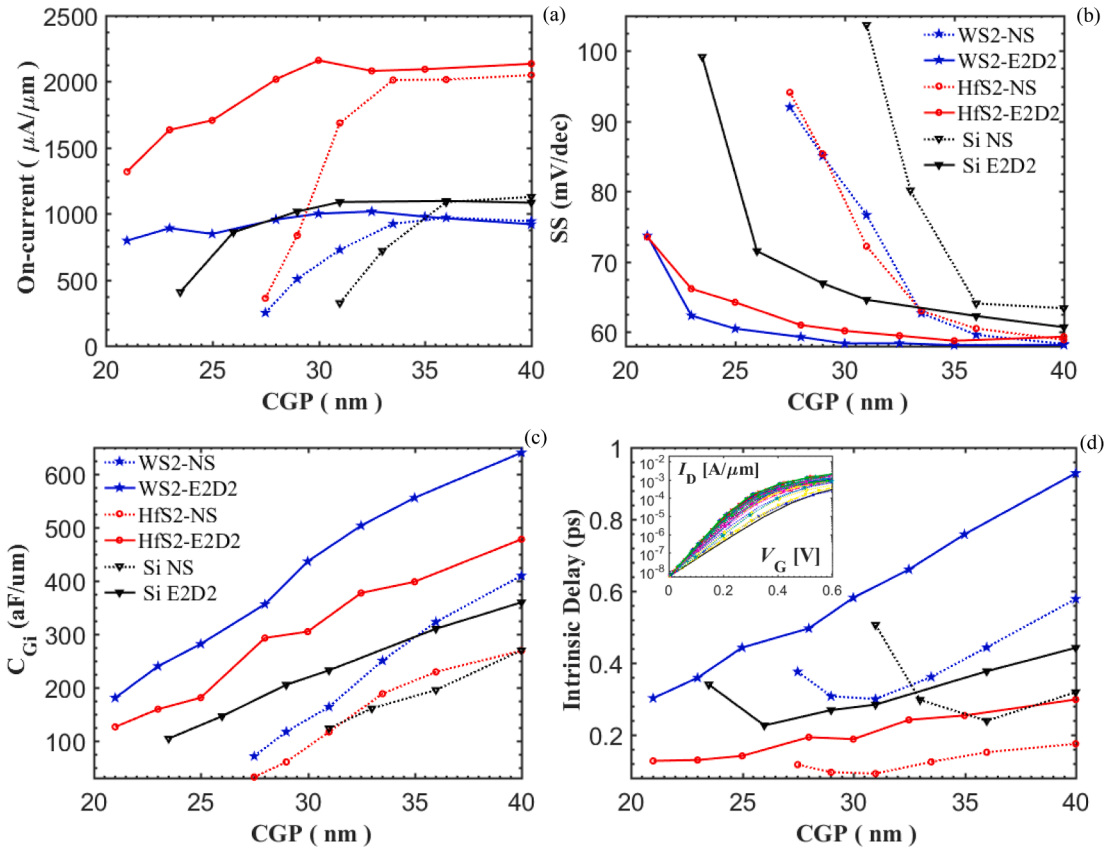

Fig. 4. A) on-current (I_{ON}) at fixed I_{OFF}, b) subthreshold slope (*SS*), c) intrinsic gate capacitance (C_{Gi}) and d) intrinsic delay – assuming an effective current i_{eff} = 0.45 × I_{ON} – vs CGP for Si and for WS_2 and HfS_2 2D monolayer *NS* and E2D2 architectures from atomistic – NEGF transport simulations (the trace of the $I_D(V_G)$ characteristics are in the inset) [1]. For the *NS*, L_G = CGP – L_C – 2 × L_{SPACER}, while L_G = CGP – L_C for E2D2. L_C = 16 nm, L_{SPACER} = 5 nm. I_{ON} is normalized by the gate perimeter. I_{OFF} = 5 nA/μm. V_{DD} = 0.6 V.

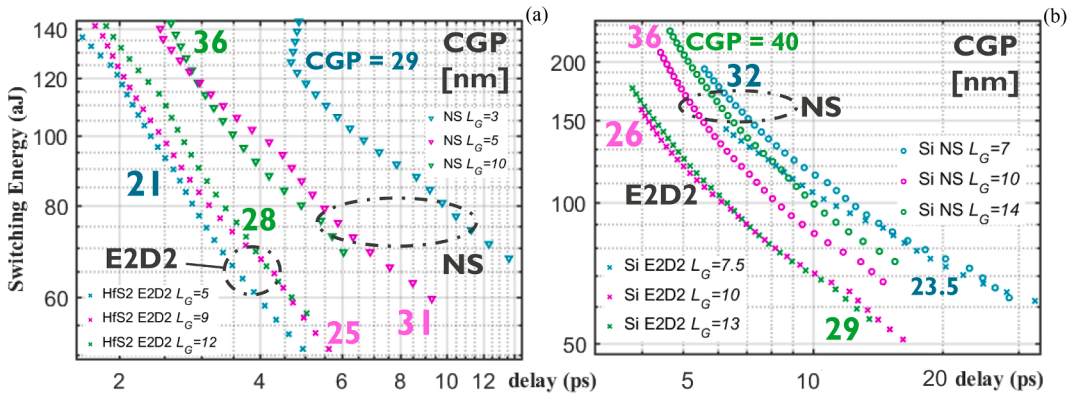

Fig. 5. Switching energy vs delay (EDP) of high-performance, stacked E2D2 and *NS* inverter cells for different CGP and L_G, as indicated in the figures, at various V_{DD} (0.4 V to 0.7 V). The devices are made of a) 1ML-HfS_2 with $n_S = 5$ sheets/device, b) Si with $n_S = 4$ sheets/device and optimized Si thickness t_S ranging from 3 to 5 nm. The inverters are loaded with the cell layout capacitances, C_{Cell}, and a 50 CGP-long metal line with capacitance C_{BK} = 198 aF/μm [13]. I_{OFF} = 5 nA/μm.

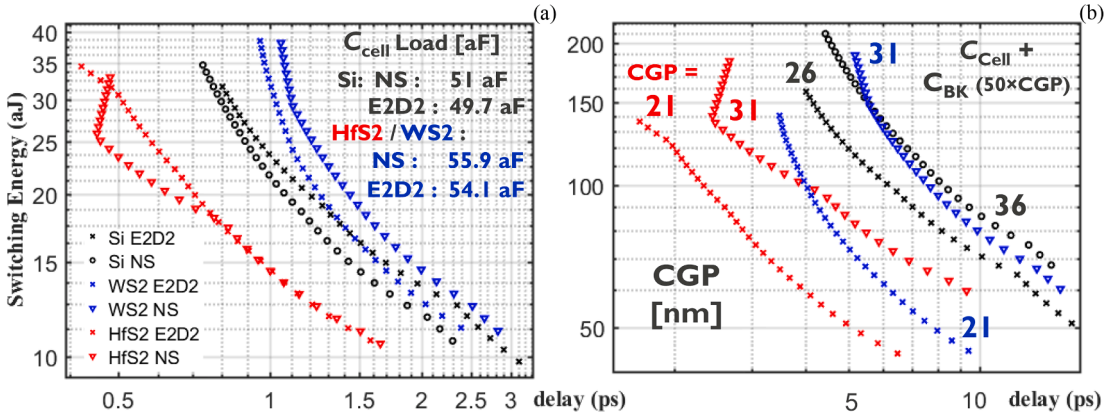

Fig. 6. EDP of the Si, WS_2 and HfS_2 stacked E2D2 and *NS* inverter cells at optimal CGP, as indicated in Fig. 6b, at various V_{DD} (0.4 V to 0.7 V). The inverters are loaded with a) C_{cell} only, its value for each invertor case is indicated in the figure, b) C_{cell} and a 50 CGP-long metal line C_{BK}. I_{OFF} = 5 nA/μm.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] Afzalian A. Ab initio perspective of ultra-scaled CMOS from 2d material fundamentals to dynamically doped transistors. npj 2D Mater Appl 2021;5(1).

[2] Afzalian A, Akhoundi E, Gaddemane G, Duflou R, Houssa M. Advanced DFT–NEGF transport techniques for novel 2-D material and device exploration including HfS2/WSe2 van der Waals heterojunction TFET and WTe2/WS2 metal/semiconductor contact. IEEE Trans Electron Dev 2021;68(11):5372–9. https://doi.org/10.1109/TED.2021.3078412.

[3] Giannozzi P, et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. J Phys Condens Matter 2009;21(39):395502. https://doi.org/10.1088/0953-8984/21/39/395502.

[4] Klimeš J, Bowler DR, Michaelides A. Chemical accuracy for the van der Waals density functional. J Phys: Cond Matter 2010;22:022201.

[5] Jancu J-M, Scholz R, Beltram F, Bassani F. Empirical spds* tight-binding calculation for cubic semiconductors: General method and material parameters. Phys Rev B 1998;57(6493). https://doi.org/10.1103/PhysRevB.57.6493.

[6] Afzalian A, Vasen T, Ramvall P, Shen T-M, Wu J, Passlack M. Physics and performance of III-V nanowire broken-gap heterojunction TFETs using an efficient tight-binding mode-space NEGF model enabling million-atom nanowire simulations. 254002 J Phys Condens Matter 2018;30(25):16. https://doi.org/10.1088/1361-648X/aac156.

[7] Afzalian A. Computationally efficient self-consistent born approximation treatments of phonon scattering for coupled-mode space non-equilibrium Green's function. J Appl Phys 2011;110(9):094517.

[8] Afzalian A, Doornbos G, Shen T-M, Passlack M, Wu J. A high-performance InAs/GaSb core-shell nanowire line-tunneling TFET: an atomistic mode-space NEGF study. In: IEEE J Electron Dev Society, Nov.; 2018. https://doi.org/10.1109/JEDS.2018.2881335.

[9] Afzalian A. Ultimate FDSOI multi-gate MOSFETs and multi-barrier boosted Gate Modulated Resonant tunneling-FETs for a new high-performance low-power paradigm. In: Iniewski K, editor. Nano-semiconductors: devices and technology. Boca Raton: CRC Press; 2012. https://doi.org/10.1201/9781315217468.

[10] Afzalian A. Computationally efficient self-consistent born approximation treatments of phonon scattering for coupled-mode space non-equilibrium Green's function. J Appl Phys 2011;110(9):094517.

[11] Afzalian A, Lee C, Dehdashti Akhavan N, Yan R, Ferain I, Colinge J. Quantum confinement effects in capacitance behavior of multigate silicon nanowire MOSFETs. IEEE Trans Nanotechnol 2011;10:300–9.

[12] Ahmed Z, et al. Introducing 2D-FETs in device scaling roadmap using DTCO. IEDM; 2020. pp. 22.5.1.

[13] https://irds.ieee.org/ editions/2018.