Tunneling Attenuation and Leakage Current in MoS_2 Nanoribbon MOSFETs

I. Prevarić, M. Matić, and M. Poljak*
Computational Nanoelectronics Group
Micro and Nano Electronics Laboratory, Faculty of Electrical Engineering and Computing
University of Zagreb, HR-10000 Zagreb, Croatia
*E-mail: mirko.poljak@fer.hr

***Abstract*—We study the OFF-state leakage current in quasi-one-dimensional MoS_2 nanoribbon (MoS_2NR) FETs using *ab initio* Hamiltonians and quantum transport simulations based on Green's functions. Complex band structure is computed for these devices and the energy-dependent tunneling attenuation inside the bandgap is obtained. We investigate the tunneling component of the OFF-state leakage for sub-20 nm long and sub-3 nm wide MoS_2NR FETs, using the under-the-barrier (UTB) and top-of-the-barrier (ToB) ballistic models. We report that using the parabolically-approximated attenuation overestimates the OFF-state leakage significantly. Furthermore, we demonstrate that all MoS_2NR FETs show good tunneling suppression due to high attenuation even for the shortest devices where the OFF-state leakage is under 16.5 nA/μm for nFETs and lower than 22 nA/μm for pFETs.**

***Keywords*—MoS_2, molybdenum disulfide, leakage current, complex bandstructure, tunneling, quasi-one-dimensional, nanoribbon, quantum transport, NEGF, ab initio, DFT**

I. Introduction

Monolayer molybdenum disulfide (MoS_2), alongside graphene, is one of the most promising two-dimensional (2D) materials for future field-effect transistors (FETs) due to high stability and very high ON-OFF current ratio [1]. The 2D MoS_2 is a transition metal dichalcogenide (TMD) that exhibits a sandwiched S-Mo-S structure as illustrated in Fig. 1a and b. The electronic, transport, and device properties of devices with 2D MoS_2 monolayer were comprehensively studied previously [2]–[4], however, patterning MoS_2 into quasi-one-dimensional (quasi-1D) structures such as nanoribbons provides an additional degree of freedom. This nanostructure enables the tuning of electron and transport properties via quantum confinement effects and allows a strong influence of edge atoms on the characteristics of MoS_2 nanoribbons (MoS_2NRs). Consequently, a high number of possible MoS_2NR configurations exists and many of them have been studied previously in both armchair and zigzag directions [5]–[7]. Nevertheless, there are some interesting configurations, such as MoS_2NRs with hybrid OH-passivated edges shown in Fig. 1a and b, whose transport and device properties have not yet been explored. Electronic properties of hybrid OH-passivated MoS_2NRs were studied in [8] where this configurations was proven to be one of the most stable edge passivation configurations for future MoS_2-based nanodevices.

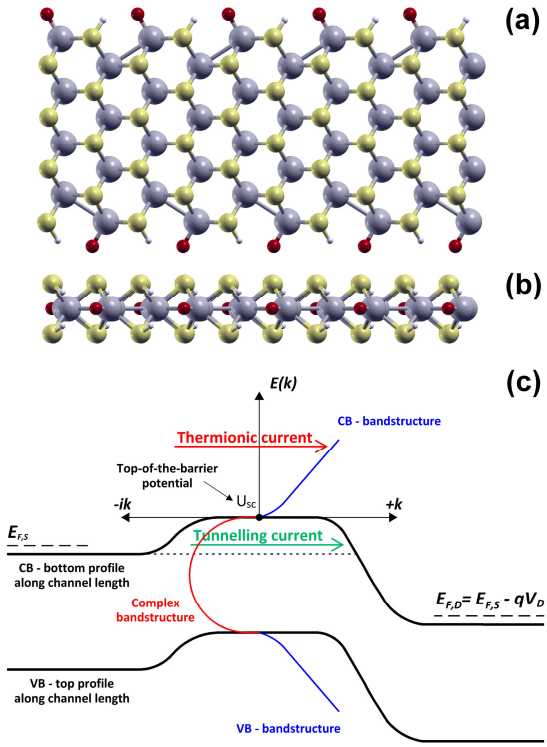

Fig. 1. (a) Top and (b) side view of MoS_2NR with OH passivation (explained in the text). (c) Conduction and valence band profiles along the channel length with an illustration of real and complex bandstructure used in UTB and TOB calculations.

In addition to stability, low OFF-state leakage current (I_{OFF}) is one of the critical requirements for future low power/high density devices. However, I_{OFF} has not yet been studied in ultra-scaled FETs with channels made of MoS_2NRs with hybrid OH passivation. In this paper, we present a study on I_{OFF} of MoS_2NR FETs using our in-house non-equilibrium Green's function (NEGF) formalism based quantum transport solver [9], [10]. The NEGF calculations are employed in conjunction with top-of-the-barrier (ToB) model [11] for the ballistic thermionic current and with the under-the-barrier (UTB) model [12],[13] that accounts for intra-band tunneling, as indicated in Fig. 1c. Highly accurate space-localized

Hamiltonians are obtained from *ab initio* plane-wave density functional theory (DFT) calculations using maximally-localized Wannier functions (MLWFs). We report strong suppression of tunneling for all the studied MoS_2NRs even in transistors with 5 nm-long channels, which indicates a robust performance and low I_{OFF} in ultra-scaled MoS_2NR FETs.

II. Methodology

Armchair MoS_2 nanoribbons passivated with OH of various widths (W) ranging from ~0.8 nm to ~3.0 nm are constructed for *ab initio* calculations. For example, the 1.11 nm-wide MoS_2NR is shown in the top (Fig. 1a) and side view (Fig. 1b) with the transport direction along the horizontal orientation. The hybrid OH passivation is achieved by terminating the edge Mo atoms with O atoms and S atoms with H atoms. Plane-wave DFT is employed in combination with transformation to tight-binding-like sparse matrices using MLWFs [14] that are suitable for NEGF simulations. Quantum Espresso [15] is used for DFT calculations, where a vacuum region of 20 Å is added in the confined directions due to the nature of DFT to assume periodicity in all three directions. The Brillouin zone (BZ) is sampled using an 1 × 15 × 1 Monkhorst-Pack grid [16], where 15 points are used in the transport direction. Perdew-Burke-Ernzerhof generalized gradient approximation (PBE-GGA) [17] exchange-correlation (XC) functional is used along projector augmented wave (PAW) pseudopotentials. The plane-wave cutoff energy is set to 1360 eV, whereas the convergence threshold is set to 10^{-3} eV/Å for the ionic force and to 10^{-4} eV for energy. Wannier90 program [18] is employed to transform the DFT Hamiltonians into the space-localized basis while preserving the accuracy of the DFT. The accuracy of transformation using MLWFs greatly depends on the trial projections on the Bloch manifold. In this work, for OH-passivated MoS_2NRs, d orbitals are chosen as trial orbitals for Mo atoms and p orbitals for S and O atoms. During wannierization, we obtain a Wannier spread lower than 2.5 Å^2 per atom for all MoS_2NR widths, which indicates successful transformation. After MLWF Hamiltonians of a MoS_2NR unit-cell are obtained, where the unit-cell designates a super-cell along the entire nanoribbon width, the cell is scaled in the transport direction to construct the total MLWF Hamiltonian of the given MoS_2NR with lengths (L) ranging from ~5 nm to ~20 nm.

The main topic of this paper is the study of I_{OFF} of n- and p- type MoS_2NR MOSFET devices with ideal contacts. The I_{OFF} consists of the thermionic (above barrier) and tunneling (under barrier) current components, as shown in Fig. 1c. The thermionic current is calculated using the ToB model that depends on the density of states (DOS) and transmission spectra obtained using the NEGF simulations [19]–[21], as implemented in our in-house code [9], [10], [22], [23]. Furthermore, UTB model [13] is employed to calculate the tunneling current through the bandgap of the S/D potential barrier. For the UTB model, tunneling attenuation is needed, which can be obtained from the complex bandstructure. While using real wave-vector k leads to dispersion of propagating electron waves, setting imaginary values of the wave-vector k and calculating the dispersion results in tunneling attenuation $\kappa(E)$ that describes evanescent electron waves. Transmission probability is calculated according to the Wentzel-Kramers-Brillouin (WKB) approximation

$$T_{WKB} = \sum_n \exp\left[-2\, l(E)\, \kappa(E)\right] \tag{1}$$

where n represents the number of complex subbands included in the calculation, while $l(E)$ is an energy-dependent tunneling barrier length. For simplicity, the potential barrier between the source and drain (S/D) is approximated with a parabolic function. Tunneling current is calculated using the Landauer formula that depends on WKB transmission, barrier height and Fermi-window, i.e. difference between S/D Fermi functions.

In order to provide meaningful comparison between various MoS_2NR devices we set a common ToB I_{OFF} of 10 nA/μm as defined in the International Roadmap for Devices and Systems (IRDS) for future high-performance (HP) CMOS nodes [24]. The I_{OFF} is set by a proper gate work function, which modifies the flat-band voltage, and accordingly lowers the OFF-state ToB potential that is used in the UTB model. Gate equivalent oxide thickness (EOT) is set to 1 nm and S/D doping is set at 0.01 molar fraction of the MoS_2NR areal atom density. Device OFF-state performance is studied when the tunneling attenuation $\kappa(E)$ is extracted directly from the complex bandstructure, and when $\kappa(E)$ is approximated with a parabolic function as is originally done in the literature [13]. The parabolic $\kappa(E)$ is defined by the maximum attenuation, occurring usually near the mid-gap, and the bandgap value. Furthermore, nanoribbon width and length scaling effects on the tunneling component of I_{OFF} are studied because the thermionic current is ballistic and, hence, independent of nanoribbon length.

III. Results and Discussion

Figure 2 reports the bandstructure consisting of the complex part for imaginary k as negative values, and the real part for real k as positive values, for $W = 0.81$ nm (Fig. 2a) and $W = 3.04$ nm (Fig. 2b). Scaling down the nanoribbon width decreases the number of subbands and, consequently, the number of complex subbands inside the bandgap also decreases. While scaling down the width generally decreases the bandgap, i.e. from 1.5 eV for $W = 3.04$ nm to 1.25 eV for the 0.81 nm-wide nanoribbon, the attenuation curves of the dominant subband do not change as much. We observe that the two attenuation characteristics exhibit qualitatively similar behavior with maximum attenuation of ~2 nm^{-1} being equal for both MoS_2NRs in Fig. 2.

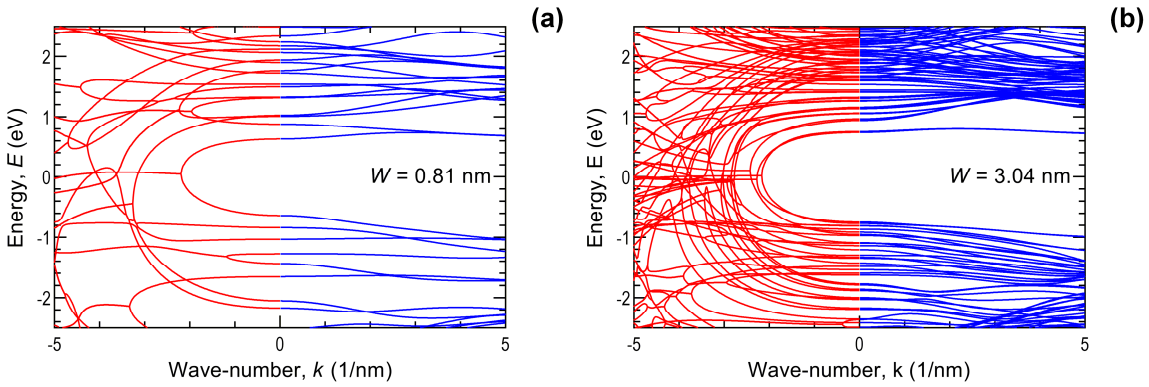

Fig. 2. Complex and real bandstructure of MoS_2NRs with the widths of (a) $W = 0.81$ nm, and (b) $W = 3.04$ nm. Real part is on the right-hand-side for positive k, and the complex part (attenuation) is on the left-hand-side for negative k.

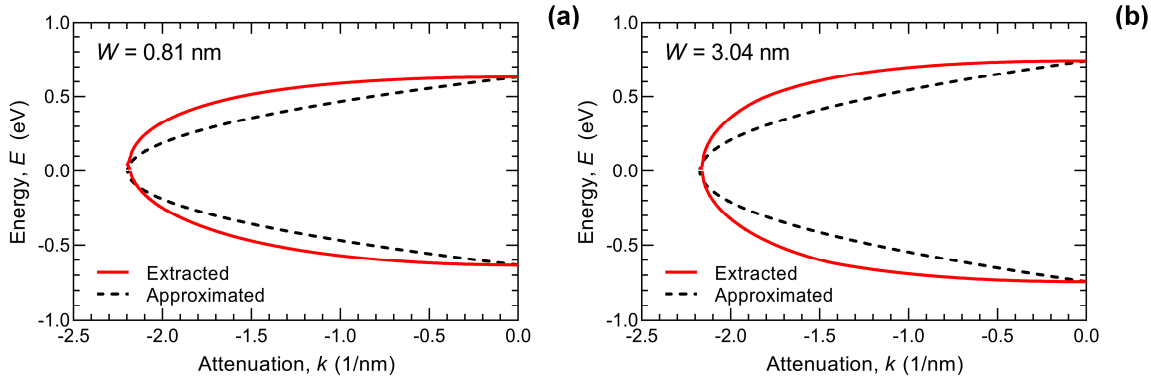

Fig. 3. Extracted and approximated attenuation characteristics of the dominant (with the lowest attenuation) complex subband for (a) 0.81 nm- and (b) 3.04 nm-wide MoS_2NRs.

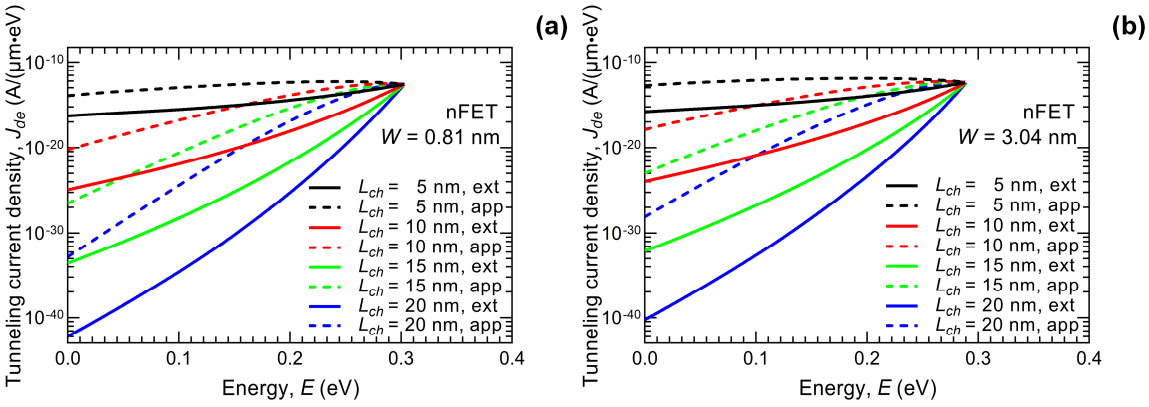

Fig. 4. Tunneling current energy density for MoS_2NR nFETs calculated with the exact extracted (ext) and approximated (app) attenuation for (a) $W = 0.81$ nm, and (b) $W = 3.04$ nm, and for four channel lengths equal to 5 nm, 10 nm, 15 nm and 20 nm.

The first subband belonging to evanescent waves, which connects the highest valence band with the lowest conduction band, can be extracted and then approximated with a parabola. Although using the approximation is much simpler to implement in the simulation environment, a large difference in attenuation is observed when comparing the extracted and approximated attenuation. The difference is reported in Fig. 3a for the 0.81 nm-wide nanoribbon, and an even larger mismatch is observed for $W = 3.04$ nm in Fig. 3b. Attenuation is the lowest near the conduction band minimum (CBM) and valence band maximum (VBM) which means that the tunneling OFF-state current is dominantly determined in the energy regions near the CBM for nFETs, and near the VBM for pFETs. Going towards mid-gap, the attenuation increases in both MoS_2NRs and reaches ~2.2 nm^{-1}, which turns off significant tunneling in this energy range.

In order to assess the differences in using the extracted and approximated attenuation curves, the tunneling current energy density (J_{de}), hereafter current density, is shown in Fig. 4 for various MoS_2NRs nFETs, with four different

channel lengths in the range from ~5 nm to ~20 nm, and the widths of $W = 0.81$ nm (Fig. 4a) and $W = 3.04$ nm (Fig. 4b). The tunneling transport occurs between $E = 0$ eV (mid-gap) and $E \sim 0.3$ eV (CBM). The current density calculated with the approximated attenuation is up to 12 orders of magnitude higher than the current density calculated using the exact extracted attenuation, which demonstrates the inadequacy of the parabolic approximation for quantitative prediction of the tunneling component of I_{OFF}. In both attenuation models, J_{de} increases while scaling down MoS_2NR channel length from 20 nm to 5 nm due to decreasing tunneling length. When the extracted $\kappa(E)$ is used, the maximum J_{de} occurs at the top of the barrier for all channel lengths. On the other hand, J_{de} calculated using the parabolic $\kappa(E)$ exhibits a maximum at the top of the barrier only for channels longer than ~15 nm. As shown in Fig. 4, for shorter channels the J_{de} maximum shifts towards the mid-gap, which is clearly unphysical. Significant difference in J_{de} magnitude and the shift of the maximum show that the approximated parabolic $\kappa(E)$ model is not appropriate to assess the OFF-state performance and, therefore, in further calculations only the extracted $\kappa(E)$ is used.

The extracted attenuation exhibits differences between electron and hole attenuation characteristics, i.e. $\kappa(E)$ curves from the mid-gap towards either CBM or VBM, as shown in Fig. 5a. In turn, nFET and pFET devices with MoS_2NR channels are expected to have different tunneling characteristics. Scaling down MoS_2NR width increases the mismatch between the electron and hole attenuations, with hole $\kappa(E)$ being lower than the electron $\kappa(E)$. This difference points to a higher OFF-state current for pFET devices generally, and the difference is expected to increase as W decreases. Maximum attenuation equals ~2.2 nm^{-1} for all MoS_2NRs with the exception for $W = 1.11$ nm where the maximum $\kappa(E)$ is slightly higher at ~2.5 nm^{-1}, as reported in Fig. 5b.

The total OFF-state current per nanoribbon width is calculated for various MoS_2NR widths and channel lengths in the range from 20 nm to 5 nm and is shown in Fig. 6. The total I_{OFF} is a sum of the tunneling current, calculated by integrating the current density J_{de} over the Fermi window, and of the L-independent thermionic current, set to 10 nA/μm for all devices in ToB simulations. Scaling down channel length results in a significant increase of I_{OFF} in both nFETs (Fig. 6a) and pFETs (Fig. 6b), especially in 5 nm-long devices with the shortest S/D tunnel barrier. The total I_{OFF} is of the same magnitude for all n- and p- type MoS_2NR FETs with $L \geq 10$ nm, while a considerable difference is observed only for $L = 5$ nm. For nFETs with the 5 nm-long channel, maximum I_{OFF} of 16.5 nA/μm is observed for $W = 0.81$ nm and minimum

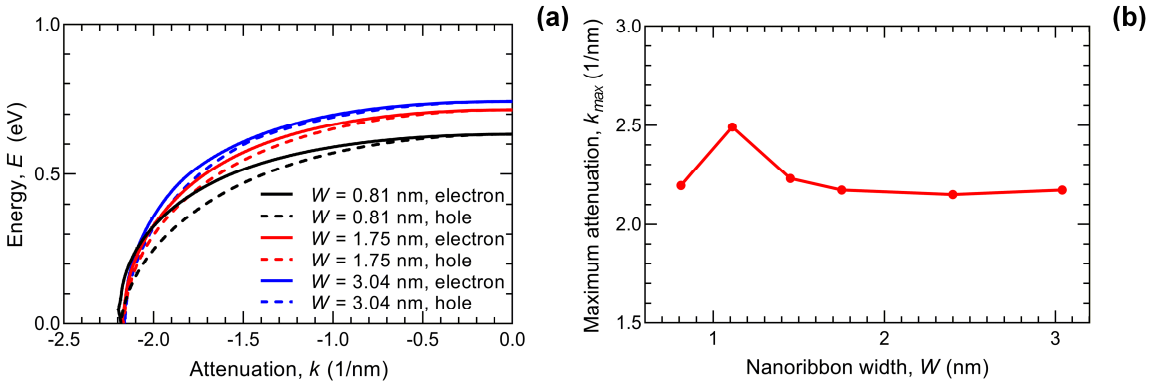

Fig. 5. (a) Extracted attenuation characteristics for electrons and holes for various widths of the MoS_2NR. The energy range spans from mid-gap at 0 eV to band minimum or maximum. (b) Dependence of the maximum attenuation of the dominant complex subband on the MoS_2NR width.

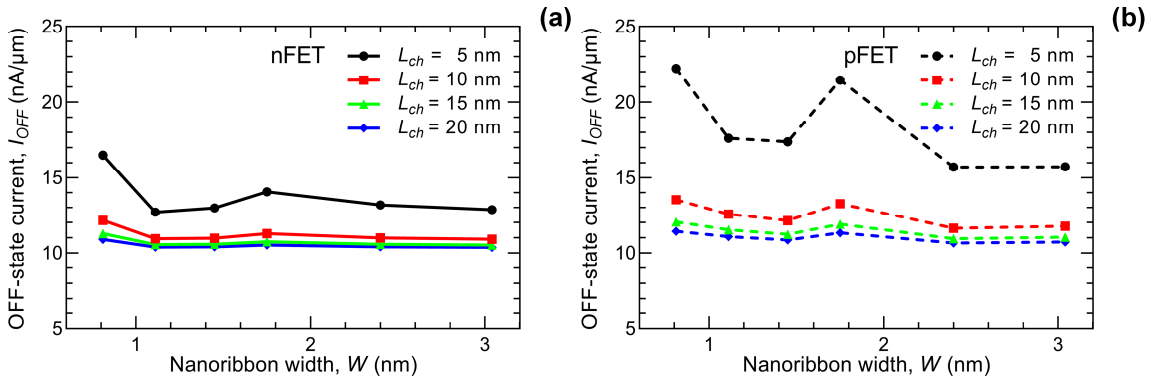

Fig. 6. Impact of MoS_2NR width scaling on total OFF-state current for (a) nFET and (b) pFET for various channel lengths ranging from 5 nm to 20 nm.

I_{OFF} of 12.7 nA/μm for the nFET with the 1.11 nm-wide MoS_2NR channel. On the other hand, pFET with the channel length of 5 nm shows the maximum I_{OFF} of 22 nA/μm for $W = 0.81$ nm and the minimum I_{OFF} of 15.7 nA/μm for $W = 3.04$ nm. These results show that a direct consequence of the lower hole attenuation compared to electron $\kappa(E)$ is the higher tunneling current in pFETs that ranges from 5.7 to 12 nA/μm in pFETs, in comparison to 2.7 to 6.5 nA/μm in nFETs. Finally, we note that the general increase and non-monotonic dependence of I_{OFF} on W reported in Fig. 6 is a consequence of non-monotonic features of $\kappa(E)$ (see Fig. 5b) and width-normalization of the drain current.

IV. Conclusions

Some features of NEGF calculations combined with the UTB and ToB models are used to calculate the tunneling and thermionic OFF-state currents in ultra-scaled MoS_2NR FETs with n- and p-type channels. Regarding the UTB model, we report a significant overestimation of tunneling leakage when a simple parabolic approximation of $\kappa(E)$ is used, rendering it useless for quantitative predictions about tunneling in ultra-scaled MoS_2 nanostructures. We report that scaling down channel length results in the expected growth of the tunneling current that is especially evident for sub-10 nm-long MoS_2NRs n- and pFETs. Due to weaker attenuation pFETs exhibit larger tunneling leakage, i.e. for $L = 5$ nm the total I_{OFF} is lower than 16.5 nA/μm for nFETs and under 22 nA/μm for pFETs. Generally, all MoS_2NRs show strong attenuation that ensures an appropriate suppression of tunneling in sub-3 nm-wide and sub-20 nm-long MoS_2NR FETs.

Acknowledgments

This work was supported by the Croatian Science Foundation under the project CONAN2D (Grant No. UIP-2019-04-3493).

References

[1] B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, and A. Kis, "Single-layer MoS2 transistors," *Nature Nanotechnology*, vol. 6, no. 3, pp. 147–150, 2011, doi: 10.1038/nnano.2010.279.

[2] W. Cao, J. Kang, D. Sarkar, W. Liu, and K. Banerjee, "Performance evaluation and design considerations of 2D semiconductor based FETs for sub-10 nm VLSI," in *2014 IEEE International Electron Devices Meeting*, Dec. 2014, p. 30.5.1-30.5.4. doi: 10.1109/IEDM.2014.7047143.

[3] A. Nourbakhsh *et al.*, "MoS2 Field-Effect Transistor with Sub-10 nm Channel Length," *Nano Lett.*, vol. 16, no. 12, pp. 7798–7806, Dec. 2016, doi: 10.1021/acs.nanolett.6b03999.

[4] Y. Lee, S. Fiore, and M. Luisier, "Ab initio mobility of single-layer MoS2 and WS2: comparison to experiments and impact on the device characteristics," in *2019 IEEE International Electron Devices Meeting (IEDM)*, Dec. 2019, p. 24.4.1-24.4.4. doi: 10.1109/IEDM19573.2019.8993477.

[5] C. Ataca, H. Şahin, E. Aktürk, and S. Ciraci, "Mechanical and Electronic Properties of MoS2 Nanoribbons and Their Defects," *J. Phys. Chem. C*, vol. 115, no. 10, pp. 3934–3941, Mar. 2011, doi: 10.1021/jp1115146.

[6] K. Dolui, C. D. Pemmaraju, and S. Sanvito, "Electric Field Effects on Armchair MoS2 Nanoribbons," *ACS Nano*, vol. 6, no. 6, pp. 4823–4834, Jun. 2012, doi: 10.1021/nn301505x.

[7] Q. Yue *et al.*, "Bandgap tuning in armchair MoS2 nanoribbon," *J. Phys.: Condens. Matter*, vol. 24, no. 33, p. 335501, Jul. 2012, doi: 10.1088/0953-8984/24/33/335501.

[8] L. Zhang *et al.*, "Modulation of Electronic Structure of Armchair MoS_2 Nanoribbon," *J. Phys. Chem. C*, vol. 119, no. 38, pp. 22164–22171, Sep. 2015, doi: 10.1021/acs.jpcc.5b04747.

[9] M. Matić;, T. Župančić, and M. Poljak, "Parallelized Ab Initio Quantum Transport Simulation of Nanoscale Bismuthene Devices," in *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, 2022, pp. 118–123. doi: 10.23919/MIPRO55190.2022.9803335.

[10] M. Matić and M. Poljak, "Ab initio quantum transport simulations of monolayer GeS nanoribbons," *Solid-State Electronics*, vol. 197, p. 108460, Nov. 2022, doi: 10.1016/j.sse.2022.108460.

[11] A. Rahman, J. Guo, S. Datta, and M. S. Lundstrom, "Theory of ballistic nanotransistors," *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1853–1864, Sep. 2003, doi: 10.1109/TED.2003.815366.

[12] M. Matic, M. Leljak, and M. Poljak, "Estimating OFF-state Leakage in Silicene Nanoribbon MOSFETs from Complex Bandstructure," in *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*, Opatija, Croatia: IEEE, Sep. 2021, pp. 85–89. doi: 10.23919/MIPRO52101.2021.9596876.

[13] A. Szabo and M. Luisier, "Under-the-Barrier Model: An Extension of the Top-of-the-Barrier Model to Efficiently and Accurately Simulate Ultrascaled Nanowire Transistors," *IEEE Trans. Electron Devices*, vol. 60, no. 7, pp. 2353–2360, Jul. 2013, doi: 10.1109/TED.2013.2263386.

[14] N. Marzari and D. Vanderbilt, "Maximally localized generalized Wannier functions for composite energy bands," *Phys. Rev. B*, vol. 56, no. 20, Art. no. 20, Nov. 1997, doi: 10.1103/PhysRevB.56.12847.

[15] P. Giannozzi *et al.*, "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials," *J. Phys.: Condens. Matter*, vol. 21, no. 39, Art. no. 39, Sep. 2009, doi: 10.1088/0953-8984/21/39/395502.

[16] H. J. Monkhorst and J. D. Pack, "Special points for Brillouin-zone integrations," *Phys. Rev. B*, vol. 13, no. 12, Art. no. 12, Jun. 1976, doi: 10.1103/PhysRevB.13.5188.

[17] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Phys. Rev. Lett.*, vol. 77, no. 18, Art. no. 18, Oct. 1996, doi: 10.1103/PhysRevLett.77.3865.

[18] G. Pizzi *et al.*, "Wannier90 as a community code: new features and applications," *J. Phys.: Condens. Matter*, vol. 32, no. 16, p. 165902, Apr. 2020, doi: 10.1088/1361-648X/ab51ff.

[19] S. Datta, *Electronic Transport in Mesoscopic Systems*. Cambridge University Press, 1997.

[20] M. Pourfath, *The Non-Equilibrium Green's Function Method for Nanoscale Device Simulation*. in Computational Microelectronics. Wien: Springer-Verlag, 2014. doi: 10.1007/978-3-7091-1800-9.

[21] Y. He, T. Kubis, M. Povolotskyi, J. Fonseca, and G. Klimeck, "Quantum transport in NEMO5: Algorithm improvements and high performance implementation," in *2014 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sep. 2014, pp. 361–364. doi: 10.1109/SISPAD.2014.6931638.

[22] M. Poljak, "Electron Mobility in Defective Nanoribbons of Monoelemental 2D Materials," *IEEE Electron Dev. Lett.*, vol. 41, no. 1, pp. 151–154, Jan. 2020, doi: 10.1109/LED.2019.2952661.

[23] M. Poljak, M. Matić, and A. Zeljko, "Minimum Contact Resistance in Monoelemental 2D Material Nanodevices With Edge-Contacts," *IEEE Electron Device Letters*, vol. 42, no. 8, pp. 1240–1243, Aug. 2021, doi: 10.1109/LED.2021.3087908.

[24] "'IEEE Intl. Roadmap for Devices and Systems (IRDS), 2019 Update.' https://irds.ieee.org/."