Article

Bandstructure and Size-Scaling Effects in the Performance of Monolayer Black Phosphorus Nanodevices

Mirko Poljak * and Mislav Matić

Computational Nanoelectronics Group, Faculty of Electrical Engineering and Computing, University of Zagreb, HR 10000 Zagreb, Croatia; mislav.matic@fer.hr
* Correspondence: mirko.poljak@fer.hr

Abstract: Nanodevices based on monolayer black phosphorus or phosphorene are promising for future electron devices in high density integrated circuits. We investigate bandstructure and size-scaling effects in the electronic and transport properties of phosphorene nanoribbons (PNRs) and the performance of ultra-scaled PNR field-effect transistors (FETs) using advanced theoretical and computational approaches. Material and device properties are obtained by non-equilibrium Green's function (NEGF) formalism combined with a novel tight-binding (TB) model fitted on ab initio density-functional theory (DFT) calculations. We report significant changes in the dispersion, number, and configuration of electronic subbands, density of states, and transmission of PNRs with nanoribbon width (W) downscaling. In addition, the performance of PNR FETs with 15 nm-long channels are self-consistently assessed by exploring the behavior of charge density, quantum capacitance, and average charge velocity in the channel. The dominant consequence of W downscaling is the decrease of charge velocity, which in turn deteriorates the ON-state current in PNR FETs with narrower nanoribbon channels. Nevertheless, we find optimum nanodevices with $W > 1.4$ nm that meet the requirements set by the semiconductor industry for the "3 nm" technology generation, which illustrates the importance of properly accounting bandstructure effects that occur in sub-5 nm-wide PNRs.

Keywords: black phosphorus; phosphorene; nanoribbon; bandstructure; quantum transport; NEGF; nanodevice; field-effect transistor; scaling; average charge velocity

1. Introduction

Monolayer black phosphorus (BP) or phosphorene, illustrated in Figure 1a,b, is a promising two-dimensional (2D) material for the realization of future electron devices in integrated circuits due to its favorable electronic and transport properties, which include acceptable bandgap and carrier mobility [1,2]. Large-area BP field-effect transistors (FETs) have been experimentally demonstrated [2–4], and short-channel and wide-gate phosphorene FETs have been theoretically studied [5,6]. On the other hand, patterning monolayer BP into phosphorene nanoribbons (PNRs) provides a technologically relevant way of adjusting the electronic and transport properties by quantum confinement effects [7–10]. Ultra-narrow PNRs have been fabricated and characterized recently with widths down to ~0.5 nm [11,12], which further kindles interest in theoretical research of PNR-based nanoelectronic devices.

For a proper assessment of these nanostructured phosphorene FETs, methodology must be based on advanced theoretical formalisms such as quantum transport, e.g., non-equilibrium Green's function (NEGF) formalism, because the transport physics in nanodevices is governed by quantum effects [13–15]. Moreover, these nanostructures must be described by proper atomically-resolved Hamiltonians that consider the complex bandstructure of such materials at the nanoscale [16,17]. The best bandstructure description is naturally provided by ab initio methods such as density-functional theory (DFT) calculations. However, the Hamiltonians obtained by DFT are very large and dense matrices

which makes DFT Hamiltonians difficult to implement for quantum transport simulations for realistically sized FETs [5,17,18]. Tight-binding (TB) models are more readily applicable and result in more computationally efficient NEGF simulations. A TB model from the relevant literature (TBL model) for phosphorene, as described in [19], is widely used to explore the properties of BP and PNRs [9,20–22]. Nevertheless, in comparison to DFT results [23], the TBL model does not reproduce the intricate multi-valley bandstructure of PNRs with the widths under ~5 nm.

In [24,25], we introduced a new DFT- based TB Hamiltonian model (DFT-TB model) that describes the bandstructure of ultra-narrow PNRs more accurately. The DFT-TB model is more complicated than the TBL model, with larger and denser unit-cell and coupling matrices but is still numerically efficient in comparison to coupled DFT-NEGF simulations. In this paper, we employ NEGF and the DFT-TB model to study size-scaling and bandstructure effects in the electronic and transport properties of PNRs, and the performance of ballistic PNR FETs with nanoribbon widths under 5 nm. We show that PNR width downscaling significantly modifies the dispersion, including the bandgap and effective mass, together with a considerable impact on density of states and transmission through the nanoribbon. For PNR FETs, we demonstrate the deterioration of the ON-state current with decreasing PNR width; however, a surprising maximum of the width-normalized current is reported for ~2.5 nm-wide PNR FET. These findings are further investigated by examining the width- and bias-dependence of channel charge density and average charge velocity. We find that width-scaling-induced bandstructure effects in carrier velocity behavior are the dominant factor that determines the properties of PNR FET current driving capabilities in the ON-state.

2. Methods

2.1. DFT-Based Tight-Binding Hamiltonian

Due to the limitations of the TBL Hamiltonian model, especially in describing ultra-narrow PNRs, this work uses the recently developed DFT-TB model, which reproduces all important size-scaling effects on the bandstructure of PNRs [24,25]. In the development of the new DFT-TB model, the DFT simulations were performed using the OpenMX package [26,27], employing generalized gradient approximation (GGA) with Perdew–Burke–Ernzerhof (PBE) exchange-correlation (XC) functional. The DFT results were used as inputs into TBStudio [28], a new software package that implements the Slater–Koster (SK) method, and allows the choice of orbital number and type for fitting the DFT data [29]. In the new DFT-TB model, 4 orbitals are included for each phosphorus atom (s, p_x, p_y and p_z), and all relevant SK overlap integrals are accounted for ($ss\sigma$, $sp\sigma$, $pp\sigma$, $pp\pi$). We showed in [24,25] that our new DFT-TB model achieves excellent agreement with DFT bandstructure in the energy range of interest, i.e., within ~1 eV from the conduction band minimum (CBM) and valence band maximum (VBM). The model data needed for the construction of PNR Hamiltonians can be found in Supplementary Materials. Electronic properties, transport properties, and PNR FET performance obtained using the new DFT-TB model are compared to those obtained with a simpler widely-used TBL model [19], in order to demonstrate the strong impact of bandstructure effects. These two TB Hamiltonians are used to study width-dependent dispersion of ultra-scaled PNRs and as inputs in the NEGF equations, which enables the investigation of the transport properties of PNRs and the performance of PNR FETs.

2.2. Quantum Transport with NEGF

The NEGF formalism is utilized in this work to solve the Schrödinger's equation with open boundary conditions (OBCs). As will be explained later in Section 2.3, we need the transmission function and density of states of the PNR, so that only the equilibrium part of our in-house NEGF code for 2D material nanostructures [30,31] is needed. The

central quantity of the NEGF formalism is the retarded Green's function of the device and is obtained by

$$G^R(E) = \left[(E + i0^+)I - H - \Sigma_S^R(E) - \Sigma_D^R(E) \right]^{-1} \tag{1}$$

where H is the device Hamiltonian constructed using either the new DFT-TB or the existing TBL model. The size and sparsity of the Hamiltonian matrix depends on the model used, as well as on nanoribbon width (W) and length (L). The Σ^R matrices are the retarded contact self-energies that consider the OBCs to the two contacts (source, S, and drain, D). These self-energy matrices are found by the iterative and numerically efficient Sancho–Rubio method [32]. The NEGF calculations in this work assume ideal contacts, meaning that the S/D extensions or reservoirs are semi-infinite semiconducting PNRs. This choice is common in the literature as it eliminates noncoherent effects at the channel-contact interfaces, and introduces no additional contact resistance into the nanostructure [33]. The retarded (G^R) and advanced Green's function (G^A) of the device, where $G^A = G^{R\dagger}$, are then used to find the transmission function, $T(E)$, and density of states, $DOS(E)$, according to expressions given in e.g., [9,34,35].
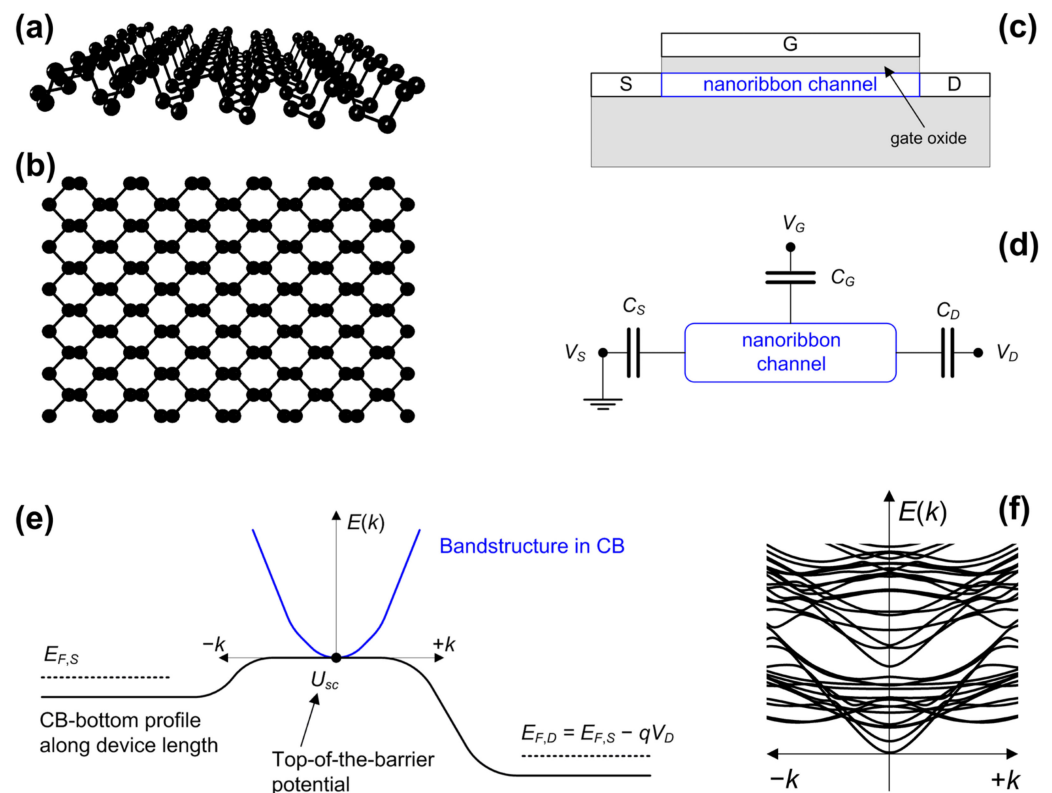

Figure 1. Illustration of a phosphorene nanoribbon with (**a**) side and (**b**) top view, as well as of (**c**) PNR FET cross-sectional view. ToB model description with (**d**) capacitive model, (**e**) bandstructure along the channel length, and (**f**) example of a PNR bandstructure calculated by the DFT-TB model.

2.3. Top-of-the-Barrier Device Model

The characteristics of ultra-scaled PNR FETs, illustrated in Figure 1c, and relevant device performance metrics are obtained using the top-of-the-barrier (ToB) device model, which self-consistently solves the Poisson equation that exploits the NEGF results and provides ballistic device characteristics [22,36,37]. The ToB model does not include tunneling, so its relevance is limited to the above-threshold operation region and can be utilized to adequately assess the ON-state performance in devices with channel lengths larger than ~10 nm [36,37].

In the ToB model, the central parameter is the top-of-the-barrier potential (U_{sc}), located at the maximum of the source-drain barrier. Assuming a grounded source, the capacitive model defined in Figure 1d results in the following potential:

$$U_{sc} = U_{sc0} - q(\alpha_G V_{GS} + \alpha_D V_{DS}) + \frac{q}{C_{ox}}(Q_s - Q_s(V_{GS} = 0)) \tag{2}$$

where the α parameters describe the capacitive coupling between the electrodes and the channel, e.g., gate coupling is defined as $\alpha_G = C_G/(C_S + C_D + C_G)$. We set $\alpha_G = 1$ and $\alpha_D = 0$, assuming that the gate electrode exhibits ideal control over the atomically-thin nanoribbon channel. In (2), C_{ox} is the gate oxide capacitance, and Q_s is the inversion charge density in the channel at top-of-the-barrier. The Q_s depends on the positions of U_{sc} and Fermi levels in source and drain regions ($E_{F,S}$ and $E_{F,D}$, respectively), which are shown in Figure 1e, and is found as follows:

$$Q_s = Q_{s,S} + Q_{s,D} \tag{3}$$

$$Q_{s,S(D)} = \frac{1}{2}\int_{-\infty}^{+\infty} DOS(E - U_{sc})f(E - E_{F,S(D)})dE \tag{4}$$

with $E_{F,D} = E_{F,S} - qV_{DS}$. After convergence is achieved for U_{sc} and Q_s, the drain current is calculated using the Landauer formula [34]. In all calculations, we assume that devices operate at room temperature, i.e., $T = 300$ K; studying temperature-related effects, including self-heating, is beyond the scope of this work.

The charge density defined in (4) and the Landauer's current formula take as inputs the $DOS(E)$ and $T(E)$, respectively, which are found by the NEGF simulations. Therefore, despite its relative simplicity, the ToB device model inherently includes all size-scaling and bandstructure-related effects with atomistic and orbital resolution. A simplified single-band effective-mass bandstructure in the conduction band is shown in Figure 1e, whereas our results are based on TB Hamiltonians, with an example of PNR dispersion obtained with the DFT-TB model shown in Figure 1f. Therefore, the described approach will allow us to explore accurately the impact of bandstructure and size scaling on the performance of ballistic PNR FETs.

3. Results and Discussions

3.1. Electronic and Transport Properties of Ultra-Narrow PNRs

Figure 2 reports the dispersion curves for various PNRs, with W ranging from 0.49 nm to 4.41 nm, calculated using the TBL and DFT-TB Hamiltonians. We observe that the DFT-TB Hamiltonian results in multi-valley bandstructure in both the conduction and valence band, which agrees with DFT studies of ultra-narrow PNRs reported in [23,38]. In contrast, the TBL Hamiltonian produces a symmetric single valley for both electrons and holes in the conduction and valence bands, respectively, irrespective of PNR width. Therefore, we expect a more accurate analysis of electronic, transport and device properties of PNRs and PNR FETs by using the DFT-TB model.

With downscaling the PNR width from 4.41 nm to 0.49 nm, bandgap (E_G) increases considerably for both TB models, i.e., from 1.57 eV to 2.61 eV (TBL) and from 0.71 eV to 1.61 eV (DFT-TB). The TBL model provides wider bandgaps due to different XC functionals used for the development of the two Hamiltonian models. Namely, PBE was used for our DFT-TB model [24,25], which results in a lower E_G in comparison to the HSE functional used for the development of the TBL model [19]. The PBE functionals are known to underestimate the bandgap, so the realistic E_G value is expected to be between those obtained by PBE and HSE DFT simulations. Nevertheless, our DFT-TB model provides more accurate dispersion properties for both electrons and holes. Since the ToB device model relies on the bandstructure properties not far away from CBM and VBM, we avoid artificial E_G adjustment in the DFT-TB model.

Due to the larger number of orbitals considered in the DFT-TB model, the related bandstructure plots in Figure 2 contain a larger number of subbands and, hence, conducting

modes than those obtained by the TBL model. Consequently, NEGF simulations using the DFT-TB model should result in higher transmission probabilities at the same energy away from the CBM or VBM in comparison to the TBL results. When PNR width is scaled down, the number of subbands decreases for both models. However, improved bandstructure description by the DFT-TB model shows that even the narrowest PNRs, with *W* of 0.49 nm (Figure 2a) and 1.47 nm (Figure 2b), exhibit a much richer dispersion so that higher transmission probabilities and more enhanced DOS are expected in these devices than predicted by the TBL model.
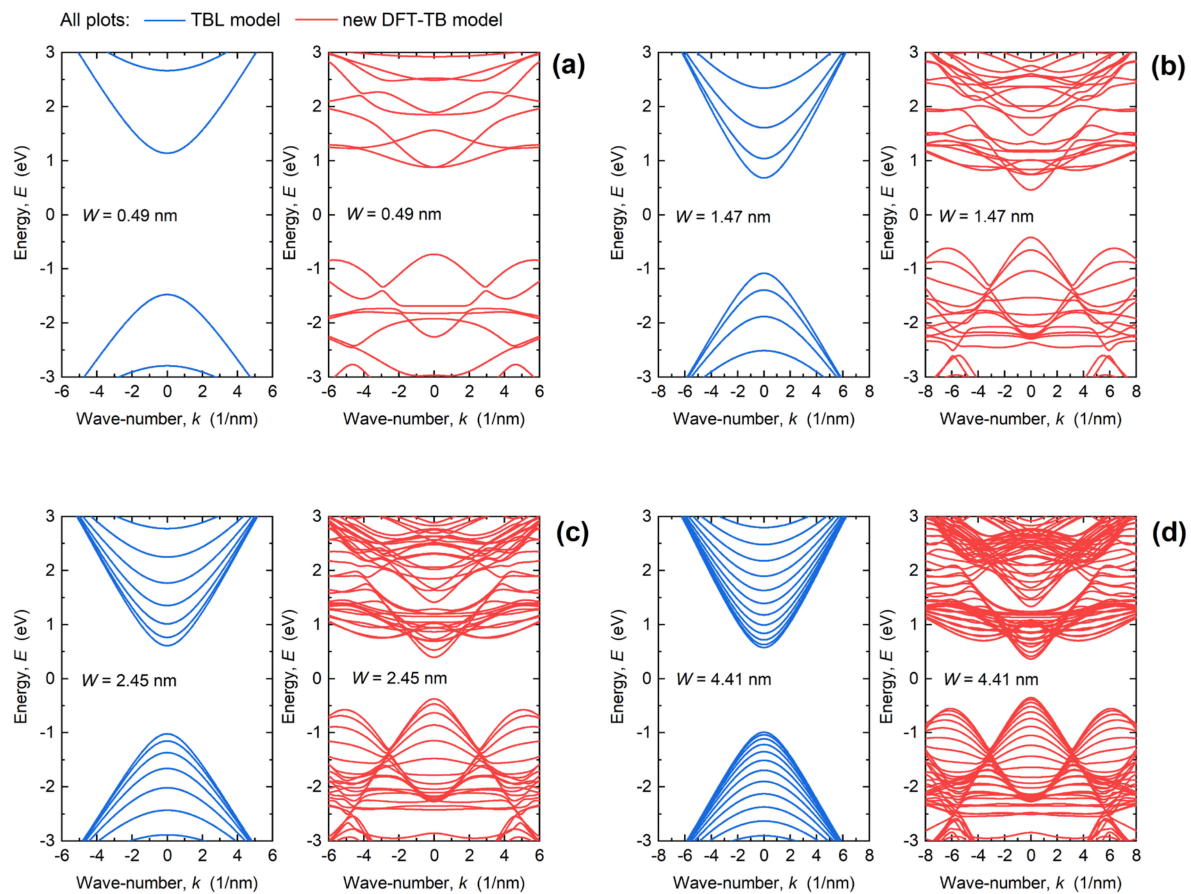
Figure 2. Comparison of dispersions obtained by the DFT-TB (panels on the right) and TBL (panels on the left) models for PNR widths of (**a**) 0.49 nm, (**b**) 1.47 nm, (**c**) 2.45 nm, and (**d**) 4.41 nm.

Regarding carrier effective masses (m^*), in wider PNRs, both TB models provide similar m^* values near the CBM and VBM, while the difference increases considerably when the width is scaled down. For W = 4.41 nm, m^* in the first subband equals $\sim 0.28m_0$ (TBL) and $\sim 0.21m_0$ (DFT-TB). For W = 0.49 nm, the DFT-TB model results in much heavier electrons in comparison to the TBL model. Namely, the DFT-TB model gives $m^* \sim 1.9m_0$ in the first subband and $m^* \sim 0.5m_0$ in the second subband, while the TBL model results in $m^* \sim 0.5m_0$ in the first subband. The lower dispersion curvature, i.e., heavier carriers, in either the conduction or valence band generally leads to increased DOS that benefits inversion charge density in the channel. At the same time, heavier carriers exhibit diminished carrier velocities, which negatively impacts the current drivability of PNR FETs. Due to interplay of different phenomena, from subband number to effective mass change, it is difficult to qualitatively estimate how the PNR width downscaling will impact device performance purely from analyzing the dispersions.

Size-scaling effects observed in dispersion characteristics in Figure 2 consequently have a strong impact on the DOS and transmission of ultra-narrow PNRs. Figure 3 shows

the DOS obtained with DFT-TB and TBL models for PNR widths of 0.49 nm, 1.47 nm, 2.45 nm, and 4.41 nm. When W is scaled down, DFT-TB DOS generally decreases away from the CBM and VBM due to reduced number of subbands in narrower nanoribbons. The two TB models provide approximately equal DOS in the vicinity of the CBM and VBM, within ~0.1 eV to ~0.2 eV. Further away from this energy range, the DFT-TB model gives much higher DOS values, which means that PNRs described by a more detailed Hamiltonian model will provide a higher amount of inversion charge that can be induced by appropriate adjustment of the quasi-Fermi level in the channel by the gate electrode, for both *n*- and *p*-type devices. For the narrowest PNR, the two DOS characteristics diverge considerably even near the CBM and VBM, which is caused by the large difference in m^* between the two models. Therefore, we can expect the largest discrepancy in PNR FET performance between the two TB models for 0.49-nm-wide PNR FETs.
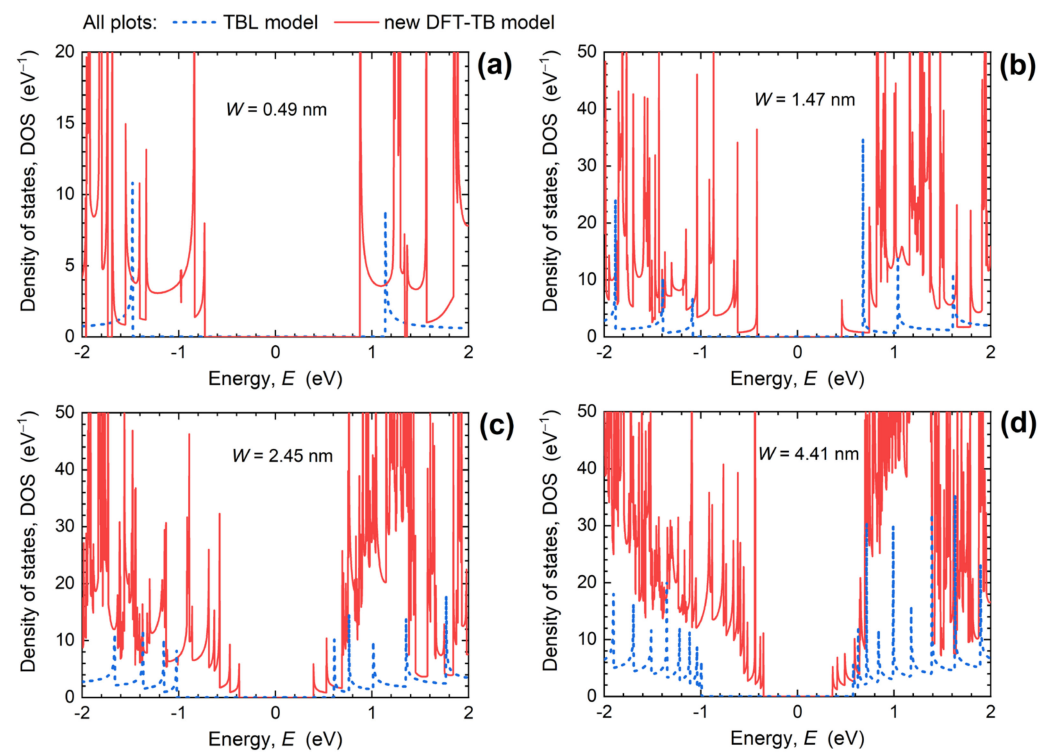

Figure 3. Density of states in PNRs obtained using the DFT-TB and TBL models for PNR widths of (**a**) 0.49 nm, (**b**) 1.47 nm, (**c**) 2.45 nm, and (**d**) 4.41 nm.

Connection between the bandstructure and transmission through the nanoribbon is illustrated in Figure 4a–c for the 1.47 nm-wide PNR. Namely, transmission function counts the number of conducting modes with a positive velocity at a certain energy and, hence, the DFT-TB model gives a generally higher transmission due to richer bandstructure than the TBL model. For $W = 1.47$ nm in Figure 4c, and other PNR widths from 0.49 nm to 4.41 nm reported in Figure 4d–f, we again observe that the TBL model reproduces the transmission well only near the CBM and VBM, while at higher energies $E >$ CBM + 0.2 eV, the DFT-TB model provides a more complex characteristic and higher transmission values. Due to the higher number of conducting modes, the improved bandstructure description by the DFT-TB model results in up to ~4× higher transmission in comparison to the TBL model.

As shown in Figure 4, downscaling of W leads to an increased transmission gap and a reduced transmission maximum. In the examined energy range, from −2 eV to 2 eV, the transmission maxima decrease from 27, over 16 and 10, to 4 when PNR width decreases from 4.41 nm, over 2.45 nm and 1.47 nm, down to 0.49 nm. At the same time, the transport gap extracted as the energy range where $T(E) < 0.01$ increases from 0.71 eV for $W = 4.41$ nm up to 1.61 eV for the 0.49 nm-wide PNR. This increase makes narrower PNRs more immune

to tunneling effects, which improves the performance of narrow PNR FETs in the OFF-state. It is expected that the reduced transmission in narrower PNRs will lead to poorer ON-state performance when W decreases, but device simulations are needed for a proper assessment due to several competing factors that play a role in device operation. For example, in the case of $W = 0.49$ nm, electrons are considerably heavier than in wider PNRs (Figure 2a), which results in high DOS (Figure 3a), and a high inversion charge density can be expected in 0.49 nm-wide PNR FETs. In addition, the transmission is boosted near the CBM in comparison to the TBL results (Figure 4d), which suggest a high drain current. However, the high effective mass ($m^* \sim 1.9 m_0$) points toward a lower carrier velocity and overall poorer current-driving capabilities of 0.49 nm-wide PNR FETs.
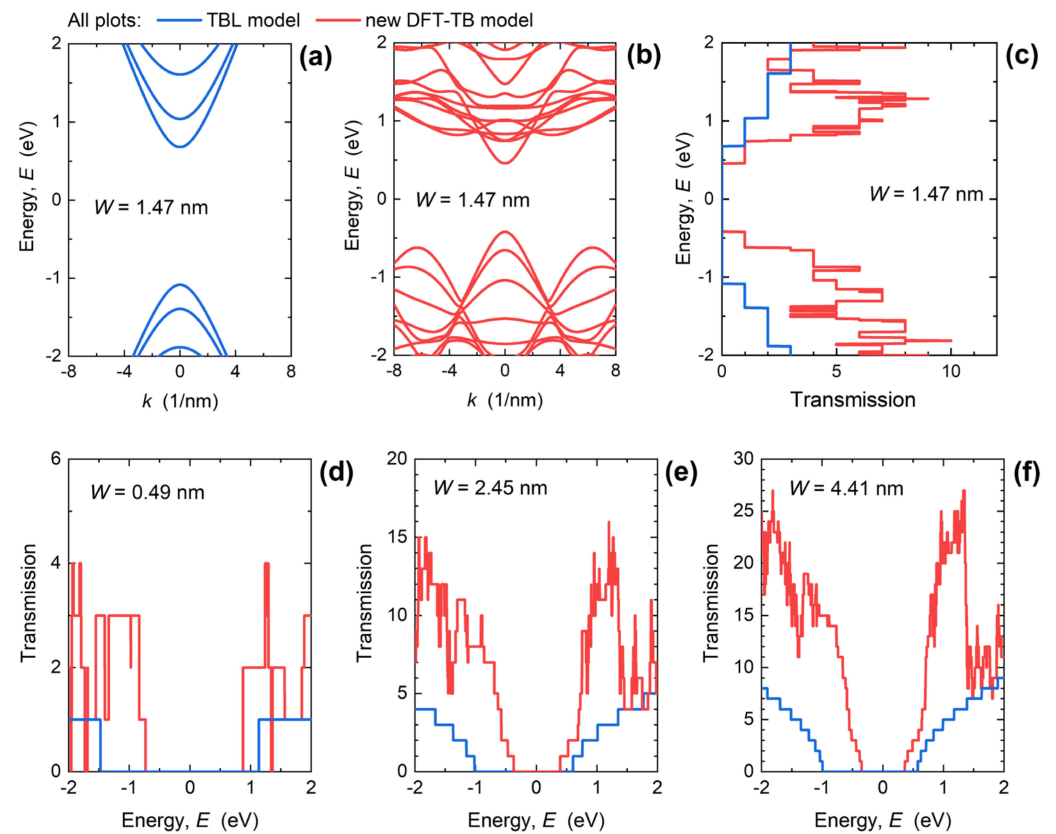

Figure 4. Dispersion obtained by (**a**) TBL and (**b**) DFT-TB model, and (**c**) transmission for the 1.47 nm-wide PNR. Comparison of transmission functions obtained with the two TB Hamiltonian models for W of (**d**) 0.49 nm, (**e**) 2.45 nm, and (**f**) 4.41 nm.

3.2. Performance of Ultra-Scaled PNR FETs

In this work, we investigate the performance of 15 nm-long PNR FETs, for which we assume SiO_2 as gate dielectric with a thickness of 1 nm and S/D doping of $m = 0.001$, where m is the molar fraction of the areal density of P atoms in PNRs, resulting in a doping density of $\sim 4 \times 10^{12}$ cm^{-2}. A common threshold voltage (V_{TH}) of 0.24 V, as projected in the International Roadmap for Devices and Systems (IRDS) at the "3 nm" CMOS node [39], is set for all devices by automatically adjusting the gate work function. Setting the same V_{TH}, and consequently the same OFF-state current (I_{OFF}), allows a meaningful and fair comparison of PNR FETs with different nanoribbon widths. The resulting I_{OFF} is ~ 1 nA/μm, defined for $V_{DS} = 0.7$ V and $V_{GS} = 0$ V. The supply voltage is 0.7 V, so the ON-state current (I_{ON}) is extracted from I-V characteristics for $V_{DS} = V_{GS} = 0.7$ V. Average charge velocity (v_{avg}) at ToB is calculated from the drain current and the obtained ToB charge density. The ON-state ToB charge density ($Q_{s,ON}$) and ON-state velocity (v_{ON}) are also extracted at the same bias point as I_{ON}, i.e., with gate and drain biased at the supply voltage.

Figure 5 reports the dependence of I_{ON} on PNR width for both the absolute magnitude of the current (Figure 5a) and for the width-normalized current (Figure 5b). The absolute I_{ON} monotonically decreases with the downscaling of PNR width, from 8.7 μA in the 4.41 nm-wide PNR FET down to 0.25 μA for $W = 0.49$ nm. In comparison to the TBL model, DFT-TB Hamiltonians with a more accurate PNR bandstructure provide higher driving currents, except for the narrowest device. These I_{ON} values in single PNRs are too low to be practically relevant, so several PNRs must be connected in parallel to provide sufficiently high I_{ON}. The plausibility of utilizing PNRs in extremely scaled FETs is quantified by the width-normalized I_{ON} reported in Figure 5b, which allows an assessment of PNR FET performance against IRDS requirements. As shown previously [22], I_{ON} obtained by the TBL model exhibits a generally decreasing trend with a weak modulation by W downscaling. Moreover, none of the examined PNR FETs using the TBL model fulfills the IRDS requirement for I_{ON} at the "3 nm" node, i.e., $I_{ON} > 1.9$ mA/μm [39]. In contrast, device simulations using DFT-TB Hamiltonians provide significant qualitative and quantitative changes. First, the I_{ON} vs. W characteristic is non-monotonic and the width-normalized I_{ON} exhibits a local maximum of 2.17 mA/μm for $W = 2.45$ nm, which means that 2.45 nm-wide PNR FETs are the most area-efficient devices in terms of current drivability. Regarding IRDS requirements, DFT-TB model reveals that PNR FETs with $W > 1.4$ nm can surpass the I_{ON} target set by the IRDS. The DFT-TB model reveals the severity of bandstructure effects in the narrowest devices because I_{ON} for $W = 0.49$ nm is only 0.51 mA/μm, which is ~2.7× lower than obtained by the TBL model. Going towards the widest examined nanoribbons, both Hamiltonian models converge to the same I_{ON}, which is expected as both approaches describe large-area phosphorene equally well.
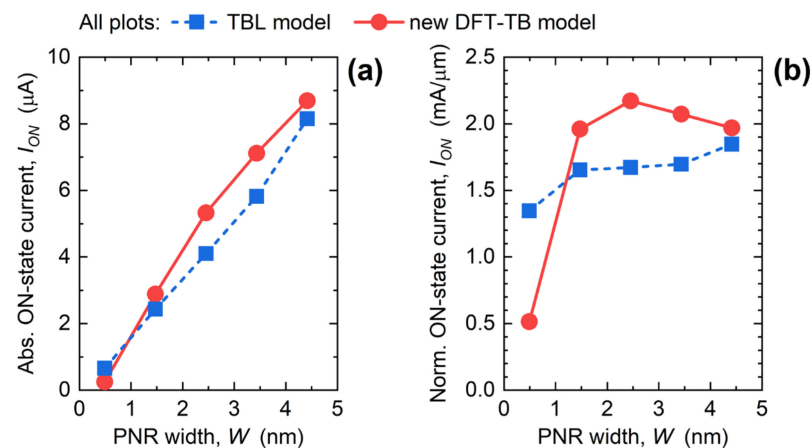

Figure 5. Impact of PNR width downscaling on (**a**) absolute I_{ON} and (**b**) width-normalized I_{ON} in PNR FETs. The plots compare the results obtained by DFT-TB and TBL models.

The size-scaling bandstructure effects on the device performance can be more easily understood by exploring the size- and bias-dependent properties of charge density (Figure 6) and average charge velocity (Figure 7) in the channel of PNR FETs. As shown in Figure 6a, Q_s increases with increasing V_{GS} up to ~7×10^{12} cm^{-2}, and generally increases with the downscaling of PNR width, although Q_s-V_{GS} curves are closely spaced for $W > 1.5$ nm. This Q_s behavior is attributed to the increasing DOS near the CBM when W decreases (see Figure 3). Figure 6b compares the bias-dependence of Q_s for the two TB models for 0.49 nm and 2.45 nm-wide phosphorene nanodevices. Clearly, the TBL model overestimates the channel charge density for $W = 2.45$ nm and underestimates it for $W = 0.49$ nm. The impact of width downscaling on Q_s in the ON-state ($Q_{s,ON}$) is reported in Figure 6c, and we observe that both models result in monotonic increase of $Q_{s,ON}$ when W decreases. The DFT-TB model provides a somewhat stronger modification of charge density with W, and the $Q_{s,ON}$ from DFT-TB simulations surpasses the TBL model results only for $W < 1.26$ nm. This result for $W = 0.49$ nm is expected because, from Figure 2a, it is

clear that DFT-TB bandstructure exhibits heavier electrons in the lowest subbands and a generally higher number of subbands near the CBM when compared to the simpler TBL model. On the other hand, wider PNRs defined by DFT-TB Hamiltonians have slightly lighter carriers, which results in $Q_{s,ON}$ being somewhat lower in the 1.5 nm to 3.5 nm width range. Since the Q_s behavior is in disagreement with I_{ON} trends, the reasons for improvement introduced by the DFT-TB model reported in Figure 5 must come from the gate capacitance (C_G) or the average charge velocity.
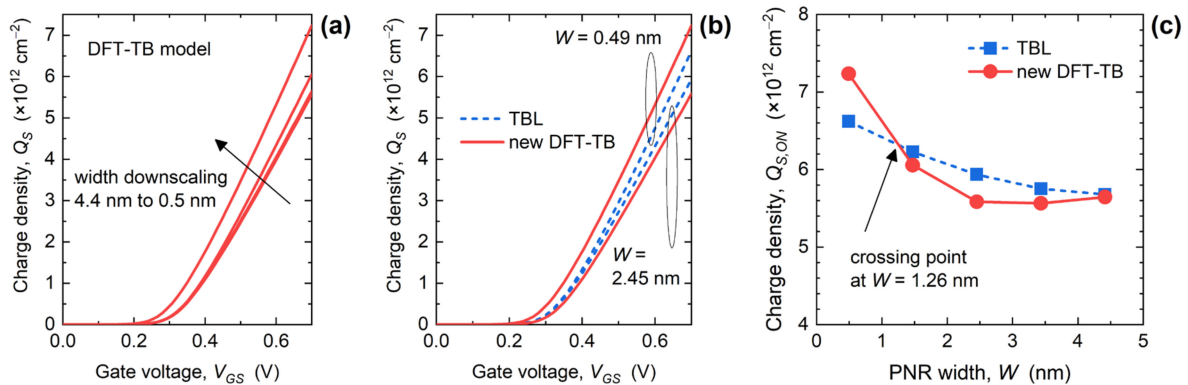

Figure 6. (**a**) Charge density vs. gate bias characteristics in PNR FETs. Comparison of (**b**) Q_s-V_{GS} curves for 0.49 nm and 2.45 nm-wide devices, and (**c**) Q_s-W characteristics in the ON-state obtained by the two TB models.

Regarding C_G, it can be assessed as a series of C_{ox} and quantum capacitance of the channel (C_q) defined as $C_q = d(qQ_s)/dV_{GS}$. Since gate oxide (SiO$_2$) thickness is 1 nm for all devices, $C_{ox} = 34.5$ fF/μm^2 and is independent of PNR width. On the other hand, C_q increases from ~102 fF/μm^2 for $W = 4.41$ nm up to ~248 fF/μm^2 (TBL) and ~298 fF/μm^2 (DFT-TB) for the 0.49 nm-wide device. Therefore, the dependence of C_q is qualitatively the same as Q_s in Figure 6c. In addition, C_q is much larger than C_{ox}, so the total C_G only slightly increases with the downscaling of PNR width, and the absolute values of C_G are 26–31 fF/μm^2. If the oxide thickness were to decrease, C_{ox} would rise, which in turn would enhance the relative impact of C_q in the total C_G of PNR FETs. Consequently, bandstructure effects reported using the DFT-TB model would be more pronounced, but a more detailed investigation is beyond the scope of the current paper.

Figure 7a plots V_{GS}-dependence of v_{avg} for $V_{DS} = 0.7$ V obtained using the DFT-TB model for various PNR widths. Generally, v_{avg} increases with increasing V_{GS}, while for $W = 0.49$ nm the velocity is independent of gate bias. The minimum v_{avg} of 0.44×10^7 cm/s is reported for the 0.49 nm-wide PNR FET, while the highest velocity is reached for $W = 2.45$ nm irrespective of V_{GS}. For $W = 2.45$ nm, v_{avg} equals 1.79×10^7 cm/s at threshold and grows to 2.43×10^7 cm/s in the ON-state. The changes introduced by the improved Hamiltonian model are illustrated in Figure 7b that compares v_{avg} in 0.49 nm and 2.45 nm-wide nanoribbons calculated using the two TB models. While both models give identical qualitative v_{avg} behavior with respect to V_{GS}, the TBL model underestimates the velocity for $W = 2.45$ nm and overestimates it for the 0.49 nm-wide PNR FET. In turn, these characteristics lead to weak W-dependence of I_{ON} for the TBL model reported in Figure 5b. Finally, the influence of PNR width downscaling on v_{ON} is illustrated in Figure 7c, and the curve exhibits a monotonic v_{ON} decrease in case of the TBL model. In contrast, a non-monotonic v_{ON} behavior is observed in case of the DFT-TB model with a maximum v_{ON} of 2.43×10^7 cm/s recorded in the 2.45 nm-wide PNR FET. This local v_{ON} maximum for $W = 2.45$ nm is a consequence of two competing mechanisms. As shown in Figure 2 and discussed in the related text, the first-subband effective mass monotonically increases in narrower PNRs, so a monotonic velocity decrease should occur with W downscaling. On the other hand, the average charge velocity considers all populated subbands, including higher subbands that exhibit larger effective masses. Therefore, the characteristics of the

first subband are important, but not the only one responsible for the overall v_{ON} behavior. When the PNR width decreases, the separation between subbands increases, which in turn decreases their population and their relative impact on v_{ON}. Hence, 2.45 nm presents an optimum width where these two mechanisms jointly result in maximum average charge velocity in the ON-state. Finally, by comparing the results in Figures 5b and 7c, we conclude that the I_{ON} characteristics reported in Figure 5b are dominantly caused by the bias- and width-dependence of carrier velocity in the channel.
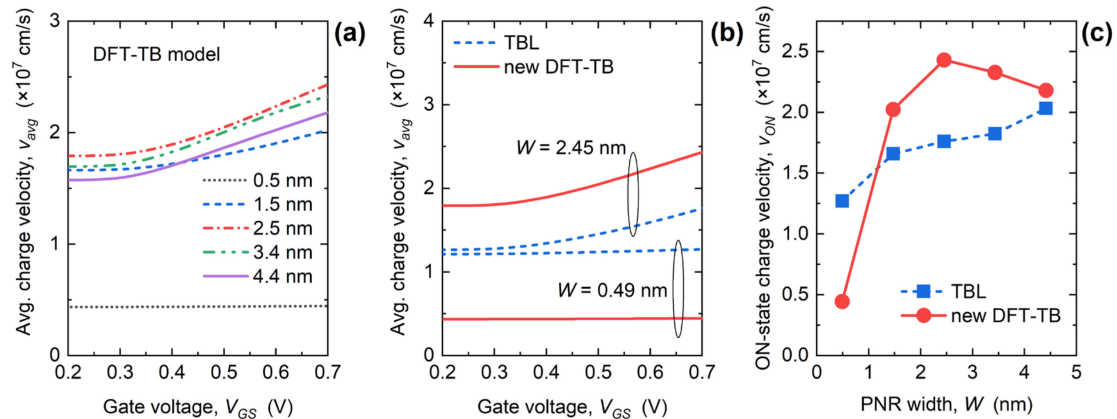

Figure 7. (**a**) Average charge velocity vs. V_{GS} in PNR FETs for the DFT-TB model. Comparison of (**b**) v_{avg}-V_{GS} curves for 0.49 nm and 2.45 nm-wide PNR FETs, and (**c**) ON-state v_{avg}-W characteristics obtained by the DFT-TB and TBL models.

4. Conclusions

Our results are obtained for ballistic transport across the channel, but in reality, carriers would likely experience scattering by intrinsic or extrinsic scattering centers, such as acoustic and optical phonons, Coulomb centers, defects, etc. [40,41]. Moreover, we neglect contact resistance in this study although it is a severe performance limiter in all 2D material-based FETs [42,43], including monolayer BP [4,44,45] and nanoribbon-based FET structures [22,33,46,47]. Nevertheless, our work illustrated the importance of using a proper Hamiltonian in the simulation of PNR nanodevices, quantified the magnitude of size-scaling and bandstructure effects, and demonstrated a considerable improvement of PNR FET figures-of-merit in comparison to the simpler TBL model. We found that for W = 2.45 nm (best device), the PNR FET is able to tolerate 13% current loss due to scattering and still meet the IRDS I_{ON} target at the "3 nm" technology node. The ballisticity level of 87% seems attainable because in a large-area phosphorene FET with a 15 nm-long channel, the ballisticity of ~90% was reported after including phonon scattering [48]. We note that the impact of electron-phonon scattering on the transport properties and performance of phosphorene nanodevices is under investigation, and the ballisticity level reported in [48] might be lower [18].

Monolayer BP or phosphorene is an attractive candidate for future electronic devices, while nanostructured BP in the form of PNRs offers an additional avenue for adjusting the electronic, transport, and device properties by quantum confinement effects. The nature of nanoscale BP devices demands advanced theoretical approaches based on quantum transport and appropriate atomically-resolved device Hamiltonians. In this work, we explored ultra-narrow PNRs and PNR FETs with the widths under ~5 nm using NEGF simulations based on a recently developed DFT-TB Hamiltonian model that accurately reproduces multi-valley dispersion and valence-conduction band asymmetry observed in PNRs by ab initio calculations. We explored the dispersion, DOS, and transmission through PNRs of various widths and reported a strong impact of width-scaling on the number and shape of conducting subbands, bandgap, and carrier effective mass. Focusing on 15 nm-long ballistic PNR FETs, we found that the ON-state inversion charge density increases up to ~7×10^{12} cm^{-2}, and that the average charge velocity decreases considerably

(from ~2.2 × 10^7 cm/s to ~0.4 × 10^7 cm/s) with the downscaling of PNR width. Velocity decrease was found to be the dominant factor in current-driving properties, so the ON-state current in PNR FETs also declines in narrower nanoribbons, from ~2 mA/μm (W = 4.41 nm) to ~0.5 mA/μm (W = 0.49 nm). Nevertheless, using an improved bandstructure description with the DFT-TB model revealed that ballistic PNR FETs with $W > 1.4$ nm can meet the IRDS requirement for I_{ON} at the "3 nm" CMOS technology node. Moreover, an optimum PNR FET with W = 2.45 nm was found, which exhibits I_{ON} ~ 2.2 mA/μm and which can operate at 87% of the ballistic limit and still meet the IRDS target for the ON-state current.

References

1. Liu, H.; Neal, A.T.; Zhu, Z.; Luo, Z.; Xu, X.; Tománek, D.; Ye, P.D. Phosphorene: An Unexplored 2D Semiconductor with a High Hole Mobility. *ACS Nano* **2014**, *8*, 4033–4041. [CrossRef] [PubMed]
2. Das, S.; Demarteau, M.; Roelofs, A. Ambipolar Phosphorene Field Effect Transistor. *ACS Nano* **2014**, *8*, 11730–11738. [CrossRef] [PubMed]
3. Li, L.; Yu, Y.; Ye, G.J.; Ge, Q.; Ou, X.; Wu, H.; Feng, D.; Chen, X.H.; Zhang, Y. Black Phosphorus Field-Effect Transistors. *Nat. Nanotechnol.* **2014**, *9*, 372–377. [CrossRef] [PubMed]
4. Haratipour, N.; Robbins, M.C.; Koester, S.J. Black Phosphorus P-MOSFETs with 7-Nm HfO2 Gate Dielectric and Low Contact Resistance. *IEEE Electron Device Lett.* **2015**, *36*, 411–413. [CrossRef]
5. Klinkert, C.; Szabó, Á.; Stieger, C.; Campi, D.; Marzari, N.; Luisier, M. 2-D Materials for Ultrascaled Field-Effect Transistors: One Hundred Candidates under the Ab Initio Microscope. *ACS Nano* **2020**, *14*, 8605–8615. [CrossRef]
6. Cao, X.; Guo, J. Simulation of Phosphorene Field-Effect Transistor at the Scaling Limit. *IEEE Trans. Electron Devices* **2015**, *62*, 659–665. [CrossRef]
7. Guo, H.; Lu, N.; Dai, J.; Wu, X.; Zeng, X.C. Phosphorene Nanoribbons, Phosphorus Nanotubes, and van Der Waals Multilayers. *J. Phys. Chem. C* **2014**, *118*, 14051–14059. [CrossRef]
8. Taghizadeh Sisakht, E.; Zare, M.H.; Fazileh, F. Scaling Laws of Band Gaps of Phosphorene Nanoribbons: A Tight-Binding Calculation. *Phys. Rev. B* **2015**, *91*, 085409. [CrossRef]
9. Poljak, M.; Suligoj, T. Immunity of Electronic and Transport Properties of Phosphorene Nanoribbons to Edge Defects. *Nano Res.* **2016**, *9*, 1723–1734. [CrossRef]
10. Poljak, M.; Suligoj, T. The Potential of Phosphorene Nanoribbons as Channel Material for Ultrascaled Transistors. *IEEE Trans. Electron Devices* **2018**, *65*, 290–294. [CrossRef]
11. Watts, M.C.; Picco, L.; Russell-Pavier, F.S.; Cullen, P.L.; Miller, T.S.; Bartuś, S.P.; Payton, O.D.; Skipper, N.T.; Tileli, V.; Howard, C.A. Production of Phosphorene Nanoribbons. *Nature* **2019**, *568*, 216–220. [CrossRef]
12. Zhang, W.; Enriquez, H.; Tong, Y.; Mayne, A.J.; Bendounan, A.; Smogunov, A.; Dappe, Y.J.; Kara, A.; Dujardin, G.; Oughaddou, H. Flat Epitaxial Quasi-1D Phosphorene Chains. *Nat. Commun.* **2021**, *12*, 5160. [CrossRef]
13. Luisier, M.; Klimeck, G. Atomistic Full-Band Simulations of Silicon Nanowire Transistors: Effects of Electron-Phonon Scattering. *Phys. Rev. B* **2009**, *80*, 155430. [CrossRef]
14. Pizzi, G.; Gibertini, M.; Dib, E.; Marzari, N.; Iannaccone, G.; Fiori, G. Performance of Arsenene and Antimonene Double-Gate MOSFETs from First Principles. *Nat. Commun.* **2016**, *7*, 12585. [CrossRef]

15. Afzalian, A.; Pourtois, G. ATOMOS: An ATomistic MOdelling Solver for Dissipative DFT Transport in Ultra-Scaled HfS$_2$ and Black Phosphorus MOSFETs. In Proceedings of the 2019 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Udine, Italy, 4–6 September 2019; pp. 1–4.

16. Marin, E.G.; Perucchini, M.; Marian, D.; Iannaccone, G.; Fiori, G. Modeling of Electron Devices Based on 2-D Materials. *IEEE Trans. Electron Devices* **2018**, *65*, 4167–4179. [CrossRef]

17. 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 Devices* **2021**, *68*, 5372–5379. [CrossRef]

18. Afzalian, A. Ab Initio Perspective of Ultra-Scaled CMOS from 2D-Material Fundamentals to Dynamically Doped Transistors. *npj 2D Mater. Appl.* **2021**, *5*, 5. [CrossRef]

19. Rudenko, A.N.; Katsnelson, M.I. Quasiparticle Band Structure and Tight-Binding Model for Single- and Bilayer Black Phosphorus. *Phys. Rev. B* **2014**, *89*, 201408. [CrossRef]

20. Poljak, M. Electron Mobility in Defective Nanoribbons of Monoelemental 2D Materials. *IEEE Electron Dev. Lett.* **2020**, *41*, 151–154. [CrossRef]

21. Yin, D.; Yoon, Y. Design Strategy of Two-Dimensional Material Field-Effect Transistors: Engineering the Number of Layers in Phosphorene FETs. *J. Appl. Phys.* **2016**, *119*, 214312. [CrossRef]

22. Poljak, M.; Matić, M.; Zeljko, A. Minimum Contact Resistance in Monoelemental 2D Material Nanodevices with Edge-Contacts. *IEEE Electron Device Lett.* **2021**, *42*, 1240–1243. [CrossRef]

23. Liang, L.; Wang, J.; Lin, W.; Sumpter, B.G.; Meunier, V.; Pan, M. Electronic Bandgap and Edge Reconstruction in Phosphorene Materials. *Nano Lett.* **2014**, *14*, 6400–6406. [CrossRef]

24. Poljak, M.; Matić, M. Bandstructure Effects in Phosphorene Nanoribbon MOSFETs from NEGF Simulations Using a New DFT-Based Tight-Binding Hamiltonian Model. In Proceedings of the 2021 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Dallas, TX, USA, 27–29 September 2021; pp. 180–183.

25. Poljak, M.; Matić, M. DFT-Based Tight-Binding Model for Atomistic Simulations of Phosphorene Nanoribbons. In Proceedings of the 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 27 September–1 October 2021; pp. 80–84.

26. Ozaki, T. Variationally Optimized Atomic Orbitals for Large-Scale Electronic Structures. *Phys. Rev. B* **2003**, *67*, 155108. [CrossRef]

27. OpenMX Website. Available online: http://www.openmx-square.org/ (accessed on 9 February 2021).

28. Nakhaee, M.; Ketabi, S.A.; Peeters, F.M. Tight-Binding Studio: A Technical Software Package to Find the Parameters of Tight-Binding Hamiltonian. *Comput. Phys. Commun.* **2020**, *254*, 107379. [CrossRef]

29. Slater, J.C.; Koster, G.F. Simplified LCAO Method for the Periodic Potential Problem. *Phys. Rev.* **1954**, *94*, 1498–1524. [CrossRef]

30. Poljak, M.; Song, E.B.; Wang, M.; Suligoj, T.; Wang, K.L. Influence of Edge Defects, Vacancies, and Potential Fluctuations on Transport Properties of Extremely Scaled Graphene Nanoribbons. *IEEE Trans. Electron Devices* **2012**, *59*, 3231–3238. [CrossRef]

31. Poljak, M.; Glavan, M.; Kuzmić, S. Accelerating Simulation of Nanodevices Based on 2D Materials by Hybrid CPU-GPU Parallel Computing. In Proceedings of the 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 20–24 May 2019; pp. 51–56.

32. Sancho, M.P.L.; Sancho, J.M.L.; Sancho, J.M.L.; Rubio, J. Highly Convergent Schemes for the Calculation of Bulk and Surface Green Functions. *J. Phys. F Met. Phys.* **1985**, *15*, 851. [CrossRef]

33. Poljak, M.; Matić, M. Metallization-Induced Quantum Limits of Contact Resistance in Graphene Nanoribbons with One-Dimensional Contacts. *Materials* **2021**, *14*, 3670. [CrossRef] [PubMed]

34. Datta, S. *Quantum Transport: Atom to Transistor*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2005; ISBN 0-521-63145-9.

35. Poljak, M. Intrinsic Capabilities for Digital Switching of Silicene Nanoribbons with Edge Defects. *IEEE Trans. Electron Devices* **2020**, *67*, 354–359. [CrossRef]

36. Rahman, A.; Guo, J.; Datta, S.; Lundstrom, M.S. Theory of Ballistic Nanotransistors. *IEEE Trans. Electron Devices* **2003**, *50*, 1853–1864. [CrossRef]

37. Kaneko, S.; Tsuchiya, H.; Kamakura, Y.; Mori, N.; Ogawa, M. Theoretical Performance Estimation of Silicene, Germanene, and Graphene Nanoribbon Field-Effect Transistors under Ballistic Transport. *Appl. Phys. Express* **2014**, *7*, 035102. [CrossRef]

38. Carvalho, A.; Rodin, A.S.; Neto, A.H.C. Phosphorene Nanoribbons. *Europhys. Lett.* **2014**, *108*, 47005. [CrossRef]

39. IEEE International Roadmap for Devices and Systems (IRDS), 2021 Edition. Available online: https://irds.ieee.org/ (accessed on 1 October 2021).

40. Gaddemane, G.; Vandenberghe, W.G.; Van de Put, M.L.; Chen, S.; Tiwari, S.; Chen, E.; Fischetti, M.V. Theoretical Studies of Electronic Transport in Monolayer and Bilayer Phosphorene: A Critical Overview. *Phys. Rev. B* **2018**, *98*, 115416. [CrossRef]

41. Gaddemane, G.; Van de Put, M.L.; Vandenberghe, W.G.; Chen, E.; Fischetti, M.V. Monte Carlo Analysis of Phosphorene Nanotransistors. *J. Comput. Electron.* **2021**, *20*, 60–69. [CrossRef]

42. Driussi, F.; Venica, S.; Gahoi, A.; Gambi, A.; Giannozzi, P.; Kataria, S.; Lemme, M.C.; Palestri, P.; Esseni, D. Improved Understanding of Metal–Graphene Contacts. *Microelectron. Eng.* **2019**, *216*, 111035. [CrossRef]

43. Gahoi, A.; Kataria, S.; Driussi, F.; Venica, S.; Pandey, H.; Esseni, D.; Selmi, L.; Lemme, M.C. Dependable Contact Related Parameter Extraction in Graphene–Metal Junctions. *Adv. Electron. Mater.* **2020**, *6*, 2000386. [CrossRef]

44. Du, Y.; Liu, H.; Deng, Y.; Ye, P.D. Device Perspective for Black Phosphorus Field-Effect Transistors: Contact Resistance, Ambipolar Behavior, and Scaling. *ACS Nano* **2014**, *8*, 10035–10042. [CrossRef]

45. Telesio, F.; le Gal, G.; Serrano-Ruiz, M.; Prescimone, F.; Toffanin, S.; Peruzzini, M.; Heun, S. Ohmic Contact Engineering in Few–Layer Black Phosphorus: Approaching the Quantum Limit. *Nanotechnology* **2020**, *31*, 334002. [CrossRef]

46. Poljak, M.; Matić, M. Quantum Transport Simulations of Phosphorene Nanoribbon MOSFETs: Effects of Metal Contacts, Ballisticity and Series Resistance. In Proceedings of the 2020 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Kobe, Japan, 23 September–6 October 2020; pp. 371–374.

47. Poljak, M. Impact of Width Scaling and Parasitic Series Resistance on the Performance of Silicene Nanoribbon MOSFETs. *IEEE Trans. Electron Devices* **2020**, *67*, 4705–4708. [CrossRef]

48. Szabo, A.; Rhyner, R.; Carrillo-Nunez, H.; Luisier, M. Phonon-Limited Performance of Single-Layer, Single-Gate Black Phosphorus n- and p-Type Field-Effect Transistors. In Proceedings of the 2015 IEEE International Electron Devices Meeting (IEDM), Washington, DC, USA, 7–9 December 2015; pp. 297–300.