Monolayer WSi_2N_4: A promising channel material for sub-5-nm-gate homogeneous CMOS devices

Ying Li,[1,*] Chunyu Qi,[2] Xun Zhou,[2] Linqiang Xu,[1] Qiuhui Li,[1] Shiming Liu,[1] Chen Yang,[1] Shiqi Liu,[1] Lin Xu,[3] Jichao Dong,[1] Shibo Fang,[1] Zongmong Yang,[1] Yifan Chen,[1] Xiaotian Sun,[2,4,†] and Jing Lu[1,5,6,7,‡]

[1] *State Key Laboratory for Mesoscopic Physics and Department of Physics, Peking University, Beijing 100871, China*

[2] *College of Chemistry and Chemical Engineering, and Henan Key Laboratory of Function-Oriented Porous Materials, Luoyang Normal University, Luoyang 471934, China*

[3] *Department of Chemistry, The University of Hong Kong, Hong Kong 999077, China*

[4] *John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*

[5] *Collaborative Innovation Center of Quantum Matter, Beijing 100871, China*

[6] *Beijing Key Laboratory for Magnetoelectric Materials and Devices, Beijing 100871, China*

[7] *Peking University Yangtze Delta Institute of Optoelectronics, Nantong 226000, China*

(Received 29 March 2023; revised 8 October 2023; accepted 29 November 2023; published 26 December 2023)

Complementary metal oxide semiconductor (CMOS) devices require both *n*-type and *p*-type metal-oxide-semiconductor field-effect transistors (MOSFETs), but achieving both types that meet the requirements of the International Technology Roadmap for Semiconductors (ITRS) at ultrashort gate lengths is a challenge. Recently synthesized two-dimensional crystal WSi_2N_4 exhibits high theoretical hole and electron carrier mobilities. In this study, we investigate the performance limits of double-gated (DG) monolayer (ML) WSi_2N_4 MOSFETs with sub-5-nm gate lengths using first-principles density-functional theory and the nonequilibrium Green's function method. Our results show that both the *n*-type and *p*-type DG ML WSi_2N_4 MOSFETs can meet the requirements of ITRS 2013 for high-performance (HP) applications at the 2028 technology node, even when the gate length is scaled to 3 nm. For low-power applications, the scaling limits for the gate lengths of the *n*-type and *p*-type DG ML WSi_2N_4 MOSFETs are 4 and 5 nm, respectively. Importantly, the ON-state currents of the *n*-type and *p*-type HP DG ML WSi_2N_4 MOSFETs are highly symmetric, highlighting the potential of ML WSi_2N_4 as a promising material for building next-generation CMOS devices, especially for HP applications.

DOI: 10.1103/PhysRevApplied.20.064044

I. INTRODUCTION

Complementary metal oxide semiconductor (CMOS) technology is widely used in very large-scale integration due to its low static power and high noise immunity [1]. For the continuation of Moore's law [2], the thickness of the CMOS channel must be aggressively reduced to sustain gate electrostatics in the sub-5-nm technology node [3]. Bulk semiconductors such as Si are vulnerable to thickness scaling due to the detrimental mobility degradation caused by the increased scattering at the interfaces between the channel and dielectric [4]. Owing to their two-dimensional (2D) crystal nature and dangling-bond-free surface, 2D materials [5,6] can maintain high carrier mobility in atomic layer thicknesses [4,7–9]. Achieving high performance in CMOS devices requires symmetrical *n*-type and *p*-type metal-oxide-semiconductor field-effect transistors (MOSFETs) at the same time. However, while there have been a substantial number of 2D heterogeneous CMOS devices [10–12], their construction involves complex and costly design and manufacturing. It is favorable to fabricate homogeneous CMOS devices, which require symmetrical *n*-type and *p*-type MOSFETs from the same channel material [13].

Recently, Hong *et al.* proposed a new group of 2D materials *MA2Z4* (where *M* is a transition metal, *A* is Si or Ge, and *Z* is N, P, or As), and have successfully synthesized two of them, $MoSi_2N_4$ and WSi_2N_4 [14–17]. Calculations based on the deformation potential theory have shown that the hole and electron mobilities of monolayer (ML) $MoSi_2N_4$ and WSi_2N_4 are 4 to 6 times higher than that of ML MoS_2 [18], and their lattice thermal conductivity

*1601110140@pku.edu.cn
†sxt@lynu.edu.cn
‡jinglu@pku.edu.cn

can reach up to 500 W/mK [19]. First-principles quantum transport simulations [20,21] have demonstrated that both *n*-type and *p*-type sub-5-nm-gate double-gated (DG) ML $MoSi_2N_4$ MOSFET can meet the ON-state current requirement of the International Technology Roadmap for Semiconductors (ITRS) [22] for high-performance (HP) applications. However, the *n*-type DG ML $MoSi_2N_4$ MOSFET did not meet the ON-state current requirement of ITRS for low-power (LP) applications in the sub-5-nm-gate region. Compared with ML $MoSi_2N_4$, ML WSi_2N_4 has higher hole mobility (approximately 320 $cm^2\ V^{-1}\ s^{-1}$) and similar electron mobility (approximately 1200 $cm^2\ V^{-1}\ s^{-1}$) [19], making it a more promising candidate than ML $MoSi_2N_4$. However, currently, there is only a simple transport simulation of the *p*-type DG ML WSi_2N_4 MOSFET [23] based on a top-of-the-barrier model [24,25]. This model oversimplifies the device channel to a single point, ignoring the significant source-to-drain tunneling current in sub-5-nm-gate transistors. Therefore, it is necessary to further explore the transport characteristics of both the *n*-type and *p*-type ML WSi_2N_4 MOSFETs in the sub-5-nm-gate region using device-level simulations that consider the geometry of the device.

In this study, we employ first-principles density-functional theory (DFT) combined with nonequilibrium Green's function (NEGF) formalism to investigate the performance limits of the *n*-type and *p*-type DG ML WSi_2N_4 MOSFETs with sub-5-nm gate length. Our simulations demonstrate that both the *n*-type and *p*-type DG ML WSi_2N_4 MOSFETs can meet the ITRS standards for HP applications as the gate length scales down to 3 nm. Remarkably, the ON-state current of the *n*-type and *p*-type DG ML WSi_2N_4 MOSFETs for HP applications exhibits a high degree of symmetry in the gate length range of 3–5 nm, with an ON-state current ratio of approximately 1.2. Additionally, the *n*-type and *p*-type DG ML WSi_2N_4 MOSFETs can also meet the ITRS requirements for LP applications as the gate length is scaled down to 4 and 5 nm, respectively. Notably, the 5-nm-gate-length *n*-type LP DG ML WSi_2N_4 MOSFET exhibits an ultrahigh ON-state current of 700 μA/μm, which is 2.4 times that of the ITRS requirement and 6.1 times that of the LP DG ML $MoSi_2N_4$ MOSFET. Therefore, our findings indicate that ML WSi_2N_4 is a promising material for building homogeneous CMOS devices for both LP and HP applications at sub-5-nm gate lengths, especially for HP applications. In addition, we employed Pareto frontier analysis [26,27] to identify the optimal solution in the multiobjective optimization of power consumption and delay time for MOSFETs. Our analysis shows that the *n*-type DG ML WSi_2N_4 MOSFETs lie on the Pareto frontier of the delay time versus power-delay product plots for both HP and LP applications, indicating that they offer one of the best solutions among MOSFETs made from different ML materials.

II. METHOD

To model the DG ML WSi_2N_4 MOSFETs, a two-terminal device is employed, as illustrated in Fig. 1(a). The device consists of a left electrode (source), a central area, and a right electrode (drain). The electrodes extend slightly into the central region to shield it from the scattering of the central area, enabling us to treat the electrodes as semi-infinite bulks. We use DFT combined with NEGF to simulate the device from the first principles [28,29]. The method is integrated into the computational software QUANTUM ATOMISTIX TOOLKIT version 2020 [28,30–32].

The NEGF framework utilizes the Green's function method to transform the process of solving the Schrödinger equation into solving Green's function and integrating terms involving it [29,33]. In the DFT + NEGF framework [28,34], the device Hamiltonian H is formulated using DFT as a functional of the electron density $n(\mathbf{r})$,

$$H(n) = -\frac{1}{2}\nabla^2 + V_{\text{ext}} + V_{\text{H}} + V_{\text{XC}}(n). \tag{1}$$

The terms on the right-hand side represent the kinetic energy term; the external potential, which includes the potential originating from the nuclei; Hartree potential; and exchange-correlation potential, respectively. V_{XC} is a function of the electron density and V_{H} is obtained by solving the Poisson equation for the electron density with boundary conditions. The effect of the left (right) electrode is introduced into the device Green's function G via the self-energy $\Sigma_{\text{L(R)}}(E)$ of the left (right) electrode,

$$G(E) = [(E + i\eta)S - H - \Sigma_{\text{L}}(E) - \Sigma_{\text{R}}(E)]^{-1}, \tag{2}$$

where E, S, i, and η represent the energy, overlap matrix, imaginary unit, and positive infinitesimal, respectively.

The self-energy term $\Sigma_{\text{L(R)}}(E)$ describes the couplings to the electrodes. They are given by

$$\Sigma_{\text{L(R)}} = \tau_{\text{L(R)}} G_{\text{L(R)}} \tau^{+}_{\text{L(R)}}, \tag{3}$$

where $\tau_{\text{L(R)}}$ represents the coupling between the left (right) electrode and the channel, and $G_{\text{L(R)}}$ is the Green's function of the left (right) electrode,

$$G_{\text{L(R)}} = [EI - H_{\text{L(R)}} + i\eta]^{-1}. \tag{4}$$

Here, $H_{\text{L(R)}}$ represents the Hamiltonian of the isolated electrode, and I denotes the identity matrix. The self-energy is a general physical concept that can represent any interaction. If we decompose the self-energy into its real and

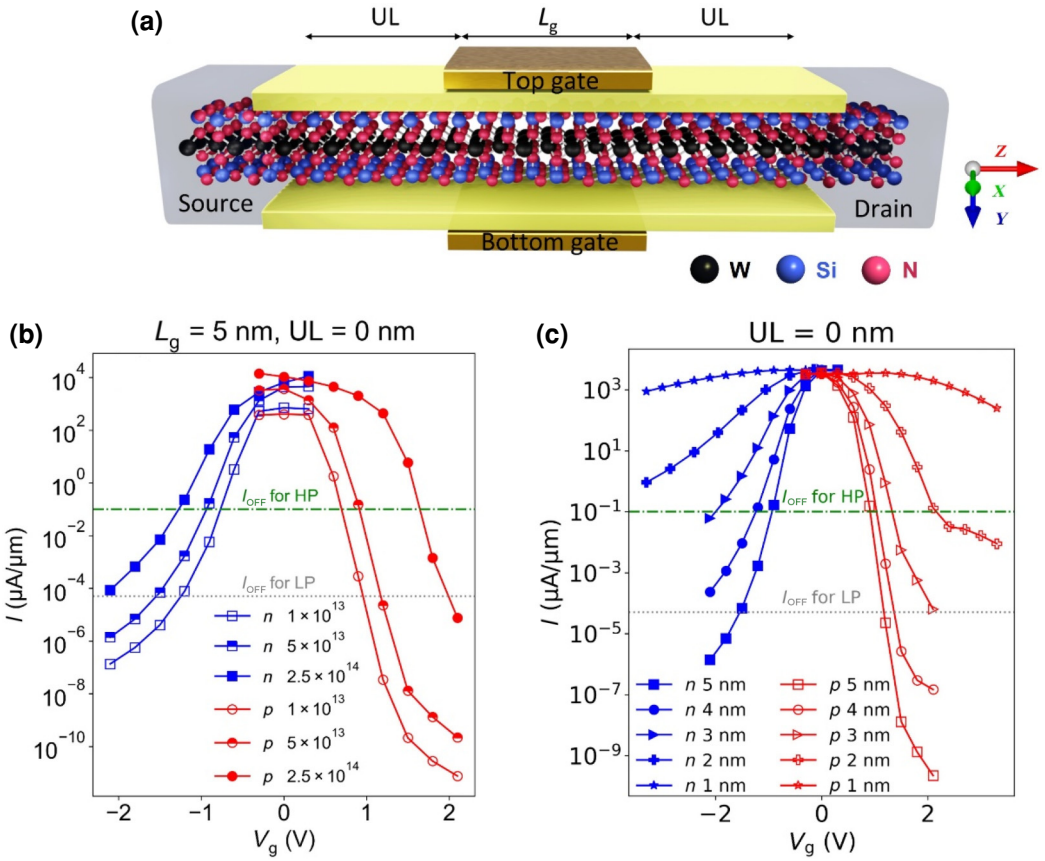

FIG. 1. (a) Schematic view of a DG ML WSi_2N_4 MOSFET. (b) Transfer characteristic curves of DG ML WSi_2N_4 MOSFETs with different doping levels (1×10^{13}, 5×10^{13}, and 2.5×10^{14} cm^{-2}) at $L_g = 5$ nm and UL = 0 nm. (c) Transfer characteristic curves of DG ML WSi_2N_4 MOSFETs with varying gate lengths ($L_g = 1$–5 nm) and UL = 0 nm. The green dash-dotted lines and gray dotted lines represent the OFF-state current requirement of ITRS for HP and LP applications, respectively. *n*-type and *p*-type MOSFETs are distinguished by blue and red colors, respectively.

imaginary parts,

$$\Gamma_{L(R)}(E) = i[\Sigma_{L(R)}(E) - \Sigma^{+}_{L(R)}(E)], \tag{5}$$

$$\Sigma_{H_{L(R)}}(E) = \frac{1}{2}[\Sigma_{L(R)}(E) + \Sigma^{+}_{L(R)}(E)]. \tag{6}$$

From Eq. (2) we can see that the real part $\Gamma_{L(R)}(E)$ modifies H to $H + \Gamma_L(E) + \Gamma_R(E)$, while the imaginary part $\Sigma_{H_{L(R)}}(E)$ represents and electron lifetime.

Then the density matrix ρ of the device can be obtained from Green's function,

$$\rho = \int \left(\frac{dE}{2\pi}\right) G[f(E-\mu_L)\Gamma_L + f(E-\mu_R)\Gamma_R]G^{+}, \tag{7}$$

where f, $\mu_{L(R)}$, and $\Gamma_{L(R)}$ denote the Fermi distribution function, the chemical potential of the left (right) electrode, and the broadening function of the left (right) electrode, respectively. The electron density distribution in real space, $n(\mathbf{r})$, is the diagonal element of the density matrix ρ in the real space representation. To obtain G, we need to know $n(\mathbf{r})$. However, $n(\mathbf{r})$ is obtained from G itself. Therefore, the Green's function and the electron density must be solved through a self-consistent method and the self-consistent process is shown in Fig. S1 in the Supplemental Material [35].

After completing the self-consistent process, the transmission coefficient $T(E)$ can be obtained from the Green's function,

$$T(E) = \mathrm{Tr}[\Gamma_L(E)\, G(E)\, \Gamma_R(E)\, G(E)^{+}]. \tag{8}$$

The Landauer-Büttiker formula [36] is then used to determine the current $I_{current}$,

$$I_{current} = \frac{2e}{h}\int_{-\infty}^{+\infty}\{T(E)[f(E-\mu_L) - f(E-\mu_R)]\}dE \quad , \tag{9}$$

where e and h are the elementary charge and Planck's constant, respectively. The quantity $T(E)[f(E-\mu_L) - f(E-\mu_R)]$ is the current spectrum. $I_{current}$ is obtained by integrating the current spectrum with the energy.

We use the generalized gradient approximation (GGA) Perdew-Burke-Ernzerhof functional [37], PSEUDODOJO pseudopotential [38], the linear combination of atomic orbitals basis, a real-space grid cutoff energy of 75 hartree, and an electron temperature of 300 K in our calculations. The process of electronic transport involves excited states, which cannot be adequately described by DFT. This is because excited states involve nonlocal many-body interactions between electrons, which are not fully captured by DFT using local or semilocal exchange-correlation potentials. The *GW* method [39,40] incorporates self-energy corrections that can accurately describe nonlocal exchange-correlation interactions and is suitable for computing excited states. Therefore, the *GW* + NEGF method [41] allows for more precise calculations of electronic transport processes. In the *GW* method, DFT energy levels and wave functions are often used as the zeroth-order approximation. In low-dimensional device environments, the presence of gate dielectric regions significantly screens the nonlocal electron-electron interactions in the channel [42]. Additionally, the channels of FETs are heavily doped in the ON state, leading to further screening of the nonlocal electron-electron interactions. Under these conditions, DFT can approximate the excited states. For instance, in the case of ML MoS_2, the DFT GGA-calculated band gap is 1.7 eV, which is closer to the experimentally measured value of 1.9 eV in the device environment than the quasiparticle band gap of 2.8 eV from *GW* calculations [42]. For heavily doped ML $MoSe_2$, the band gap calculated using DFT GGA is 1.52 eV, which agrees well with the *GW* result of 1.59 eV and the angle-resolved photoelectron spectroscopy experimental value of 1.58 eV [43–45]. For these two reasons, DFT GGA functional can provide reasonable results in describing low-dimensional electronic transport.

We employ periodic boundary conditions, Newman boundary conditions, and Dirichlet boundary conditions in the *X*, *Y*, and *Z* directions, respectively [46], as shown in Fig. 1(a). The actual 2D device is finite in the *X* direction. In experiments, the width of 2D devices in the *X* direction is usually on the order of micrometers [47–49]. At this scale, quantum size effects can be ignored. Therefore, using periodic boundary conditions is reasonable. The electrodes use a Monkhorst-Pack grid of $32 \times 1 \times 100$ while the central area uses a Monkhorst-Pack grid of $32 \times 1 \times 100$.

Our idealized model simulates the device in the Ohmic contacted and ballistic transport limit. If the actual device can be pushed into this situation, then the experimental results can be compared with our results. Thanks to advancements in micro- and nanofabrication techniques, it is now possible to fabricate 2D material Ohmic contacted FETs with sub-5-nm gate lengths and total channel lengths less than 10 nm, as well as grow high-quality dielectric layers [50,51]. These devices have Ohmic contacts and work with a ballistic ratio above 80% [47,48,52]. The reliability of the DFT + NEGF method is validated by the good agreement between the experimental and simulated data [53]. For example, the subthreshold swing (SS) of a 1-nm-gate-length ML MoS_2 transistor, simulated using DFT + NEGF, is found to be 66 mV/dec, which closely matches the experimental measurement of 65 mV/dec [52,54,55]. Similarly, the transfer characteristics, ON-state current, delay time, and power-delay product of a 5-nm-gate-length carbon nanotube transistor, simulated using DFT + NEGF, are highly consistent with the experimental measurements [56–58]. The predicted high ON-state current (1497 μA/μm) for the ML 7-nm InSe FET at a supply voltage of 0.64 V [59] has been confirmed by experiment (1430 μA/μm) for a 20-nm three-layer InSe FET at a supply voltage of 0.7 V [48].

III. RESULTS AND DISCUSSION

A. Device structure and optimization

The ML WSi_2N_4 structure can be visualized as a single W-Si sublayer sandwiched between two Si-N sublayers [see Fig. S2(a) in the Supplemental Material [35]]. Our optimization process yields lattice parameters of $a = b = 2.909$ Å, in good agreement with the previous theoretical data (2.915 Å), although slightly lower than the experimental data (2.97 Å) obtained from the TEM measurements [14]. We calculate the band structure and projected density of states of the ML WSi_2N_4 [see Figs. S2(b) and S2(c) in the Supplemental Material [35]]. The indirect band gap is found to be 2.09 eV and the main components of the band edges come from the *d* orbitals of the W atom. The effective mass of the electron and the hole along the transport direction is $0.34m_0$ and $1.36m_0$ (m_0 is the rest mass of a free electron, with a value of approximately 9.109×10^{-31} kilograms), respectively.

We utilize heavily doped ML WSi_2N_4 as the electrode to directly inject carriers and DG structures to enhance gate controllability [53]. In the DG configuration, the ML WSi_2N_4 layer is capped by two symmetrical metal gates, with SiO_2 (dielectric constant, 3.9; thickness, 4.1 Å) serving as the dielectric layer, as illustrated in Fig. 1(a). Additionally, we incorporate underlap (UL) structures to mitigate short-channel effects [60]. The underlaps are made of the same material SiO_2 (dielectric constant, 3.9; thickness, 4.1 Å) as the dielectric layer. In our simulation, carriers are injected directly from the doped source electrode to the channel to simulate perfect electrode contact. Therefore, the current primarily flows through the ML WSi_2N_4, and there is no current in the gate and the dielectric regions. Hence, we treat both the metal and dielectric regions as continuous medium spatial regions [30]. This corresponds to an average approximation of the Hartree potential from the gate and dielectric regions. This kind of systematic error is believed to be insignificant, and

such approximation is widely used in atomistic device simulations [28,30,31].

At room temperature, nonelastic scattering caused by phonons can lead to decoherence. Since our model is based on ballistic transportation, it might overestimate the ON-state current. For instance, in a monolayer blue phosphorene FET with a channel length of 10.2 nm simulated using DFT + NEGF calculations, phonon scattering can cause a decrease in the ON-state current by about 20% [61]. It is important to note that the impact of phonon scattering decreases as channel and gate lengths decrease [62]. To demonstrate the applicability of our calculations at room temperature, we calculate the variation of the phonon-limited mean free paths of electrons and holes with temperature (refer to the Supplemental Material [35] for more details). At room temperature, the mean free paths of electrons and holes are 64 and 13 nm, respectively, which are longer than all the channel lengths involved in our calculations. Therefore, the DG ML WSi_2N_4 MOSFETs used in our study primarily operate in the ballistic zone. So, our computational results can be compared with experimental devices exhibiting ballistic transport.

To evaluate the performance of the device, we extract various parameters from the transfer curves, including SS, ON-state current (I_{ON}), ON:OFF current ratio (I_{ON}/I_{OFF}), total capacitance (C_t), delay time (τ), power-delay product (PDP), and energy-delay product (EDP). Owing to the slowdown of the physical scaling of devices in recent years, the shortest gate length that we can refer to in the latest International Roadmap for Devices and Systems (IRDS) 2022 is 12 nm [63]. So, we used the more stringent criteria from ITRS 2013 [22] for devices with a gate length of 5 nm to benchmark our sub-5-nm-gate DG ML WSi_2N_4 MOSFETs. It is worth noting that if the device meets the requirements of ITRS 2013, it will automatically meet the requirements of IRDS 2022. In the following sections, we refer to the requirements for sub-5-nm-gate devices from ITRS 2013 as "ITRS requirements" for convenience.

Firstly, we optimize the doping level while keeping the gate length ($L_g = 5$ nm) and UL (UL = 0 nm) fixed. Figure 1(b) shows that a low doping level (1×10^{13} cm^{-2}) results in a low saturation current and consequently a low ON-state current, while a high doping level (2.5×10^{14} cm^{-2}) leads to a high SS and consequently a low ON-state current. By considering both the SS and saturation current, we select a doping level of 5×10^{13} cm^{-2} for further simulations (the SS and I_{ON} for the HP applications are presented in Table S1 in the Supplemental Material [35]).

Subsequently, we analyze the transfer characteristic curves of the DG ML WSi_2N_4 MOSFETs without UL (UL = 0 nm) and with gate lengths ranging from 5 to 1 nm, as depicted in Fig. 1(c). We observe that for different gate lengths, the saturation currents remain constant. However, the gating capability weakens with a decrease in the gate length, due to the combined effect of drain-induced barrier lowering (DIBL) and a reduction in the width of the barrier under the gate, making it difficult to turn off the device. We find that, in the absence of UL structures, *n*-type HP DG ML WSi_2N_4 MOSFET can meet the ON-state current requirement of ITRS at $L_g = 5$ nm with $I_{ON} = 1339$ μA/μm, while the *p*-type HP DG ML WSi_2N_4 MOSFET can meet the ON-state current requirement of ITRS when the gate length is scaled to 4 nm with $I_{ON} = 1024$ μA/μm. However, neither *n*-type nor *p*-type DG ML WSi_2N_4 MOSFET can fulfill the ON-state current requirement of ITRS for LP applications in the sub-5-nm-gate-length region.

To improve gate control, we introduce UL structures that can act as a buffer between the source or drain and the gate, thus reducing DIBL. However, if the ULs are too long, the gate's controllability weakens due to the reduction in the proportion of the channel covered by the gate. Thus, we optimize the MOSFET's performance at a specific gate length by selecting an optimal UL length. We calculate the I-V_g curves of the *n*-type and *p*-type DG ML WSi_2N_4 MOSFETs for gate lengths ranging from 1 to 5 nm with varying UL lengths, as shown in Fig. S3 in the Supplemental Material [35]. The corresponding SS, I_{ON}, I_{ON}/I_{OFF}, C_t, τ, and PDP values for both HP and LP applications are summarized in Tables I and II, respectively.

B. Effects of UL

To illustrate the effects of UL structures, we choose the 3-nm-gate *n*-type DG ML WSi_2N_4 MOSFETs for HP applications as examples. As shown in Fig. 2(a), without ULs, a high gate voltage of over 2 V is needed to turn off the drain-source current, while with the aid of ULs, a gate voltage of less than 1 V is sufficient to turn off the drain-source current. Figure 2(b) shows that the ON-state current initially increases and then decreases with the increase of the UL length. This nonmonotonic relationship between the ON-state current and the UL length is common for the DG ML WSi_2N_4 MOSFETs, as demonstrated in Fig. S4 in the Supplemental Material [35].

To elucidate the physical mechanism underlying the effect of UL structures, we calculate the projected local density of states (PLDOS) and current spectra for the ON state and OFF state under different ULs, and extract the effective barrier change $\Delta\phi_B$ using the equation

$$\Delta\phi_B = \phi_{B_{OFF}} - \phi_{B_{ON}}, \tag{10}$$

where $\phi_{B_{ON}}$ and $\phi_{B_{OFF}}$ represent the effective barriers (ϕ_B) between the channel and the source of the ON state and OFF state, respectively [64,65]. The PLDOS figures in Fig. 3 allow us to determine $\phi_{B_{ON}}$ and $\phi_{B_{OFF}}$.

There are three factors that can influence the ON-state current. Firstly, there are two types of conduction mechanisms, and UL structures can alter the dominant

TABLE I. Comparison of the device performance metrics for the sub-5-nm-gate *n*- and *p*-type DG ML WSi_2N_4 MOSFETs against the ITRS requirements for HP applications in 2028 horizon (2013 edition). The OFF-state current is fixed at 0.1 μA/μm according to the ITRS HP standards. The table presents subthreshold swing (SS), ON-state current (I_{ON}), ON:OFF ratio (I_{ON}/I_{OFF}), total capacitance (C_t), delay time (τ), and power-delay product (PDP).

L_g (nm)	UL (nm)	SS (mV/dec)		I_{ON} (μA/μm)		I_{ON}/I_{OFF}		C_t (fF/μm)		τ (ps)		PDP (fJ/μm)	
		n type	*p* type	*n* type	*p* type	*n* type	*p* type	*n* type	*p* type	*n* type	*p* type	*n* type	*p* type
5	0	120	78	1339	1506	1.34×10^4	1.49×10^4	0.346	0.36	0.165	0.153	0.142	0.147
	1	82	56	2130	1672	3.53×10^4	1.67×10^4	0.282	0.294	0.085	0.113	0.116	0.121
	2	69	59	2013	1344	2.01×10^4	1.34×10^4	0.202	0.206	0.064	0.098	0.083	0.084
4	0	180	97	255	1024	2.55×10^3	1.02×10^4	0.268	0.302	0.672	0.189	0.11	0.124
	1	106	75	1545	1451	1.55×10^4	1.45×10^4	0.25	0.258	0.103	0.114	0.102	0.106
	2	80	60	1675	1158	1.68×10^4	1.16×10^4	0.174	0.171	0.067	0.094	0.071	0.07
	3	70	63	1296	797	1.30×10^4	7.97×10^3	0.13	0.127	0.064	0.102	0.053	0.052
3	0	287	136	5	393	5.00×10^1	3.93×10^3	0.205	0.238	26.211	0.389	0.084	0.098
	1	154	92	585	913	5.85×10^3	9.13×10^3	0.17	0.194	0.186	0.136	0.07	0.079
	2	103	70	1288	1085	1.29×10^4	1.09×10^4	0.155	0.157	0.077	0.093	0.064	0.064
	3	84	55	1170	839	1.17×10^4	8.39×10^3	0.116	0.117	0.064	0.089	0.048	0.048
2	0	616	225	···	36	···	3.60×10^2	···	0.17	···	2.998	···	0.07
	1	250	121	34	541	3.40×10^2	1.21×10^3	0.116	0.126	2.19	0.149	0.048	0.051
	2	146	89	483	733	4.83×10^3	7.33×10^3	0.114	0.106	0.151	0.092	0.047	0.043
	3	109	78	406	777	4.06×10^3	7.77×10^3	0.071	0.092	0.112	0.076	0.029	0.038
	4	92	67	279	582	2.79×10^3	5.82×10^3	0.063	0.08	0.145	0.087	0.026	0.033
1	0	2343	1087	···	···	···	···	···	···	···	···	···	···
	1	429	189	1	88	1.00×10^1	8.80×10^2	0.068	0.077	43.403	0.556	0.028	0.031
	2	223	121	54	334	5.40×10^2	3.34×10^3	0.057	0.06	0.681	0.114	0.024	0.024
	3	156	95	330	412	3.30×10^3	4.12×10^3	0.05	0.053	0.097	0.082	0.02	0.022
	4	119	67	383	407	3.83×10^3	4.07×10^3	0.045	0.043	0.074	0.068	0.018	0.018
12 (IRDS)	···	75		924		9.24×10^4		0.37		0.78		0.47	
5.1 (ITRS)	···	···		900		9.00×10^3		0.6		0.423		0.24	

conduction mechanism [66]. As depicted in Fig. 3, the thermal current corresponds to the current spectrum above the effective barrier ϕ_B, while the tunneling current corresponds to the current spectrum below ϕ_B. Secondly, UL structures can change the effective barrier change $\Delta\phi_B$. The effective barrier ϕ_B changes when the MOSFET switches between the ON state ($\phi_{B\,ON}$) and OFF state ($\phi_{B\,OFF}$). However, due to the short-channel effect, the effective barrier change $\Delta\phi_B$ is typically smaller than the change in the gate voltage (0.64 eV in ITRS). A high value of $\Delta\phi_B$ will lead to a low SS, as illustrated in Fig. 2(c). Finally, the UL structures alter the ratio of L_g to the whole channel length. If the ULs are too long, the ULs not covered by the gate electrode will not be doped by electrostatics.

When UL = 0 nm, the DG ML WSi_2N_4 MOSFET operates entirely in the tunneling region, and thus a large negative gate voltage is required to raise the effective barrier $\phi_{B\,OFF}$ to 0.53 eV [Fig. 3(a)] to turn off the device. Meanwhile, due to DIBL, the effective barrier change $\Delta\phi_B$ is seriously reduced to only 0.3 eV [Fig. 2(c)], resulting in a high SS (287 mV/dec). Therefore, the ON-state current is low. As the UL length increases, the tunneling probability decreases exponentially, reducing the effective barrier needed to turn off the device (ϕ_{BOFF}). Meanwhile, the alleviation of DIBL brings a higher effective barrier change $\Delta\phi_B$, so the effective barrier of the ON state (ϕ_{BON}) gets lower. Hence the ON-state current gets higher. If we only consider these two factors, a longer UL will always lead to a higher ON-state current. However, when the UL is too long, the undoped UL region will decrease the transmission coefficient $T(E)$ for E in the transmission window $[\mu_s, \mu_d]$ (where μ_s, μ_d are the chemical potentials of the source and drain electrode, respectively), leading to a decline in the ON-state current. This decrease in the transmission coefficient $T(E)$ is reflected in the decline of the peak value of the current spectrum from 3.86 μA/eV to 3.42 μA/eV as the UL changes from 2 to 3 nm, as shown

TABLE II. Comparison of the device performance metrics for the sub-5-nm-gate *n*- or *p*-type DG ML WSi_2N_4 MOSFETs against the ITRS requirements for LP applications in 2028 horizon (2013 edition). The OFF-state current is fixed at 5×10^{-5} μA/μm in terms of the ITRS LP standards. The table presents subthreshold swing (SS), ON-state current (I_{ON}), ON:OFF ratio (I_{ON}/I_{OFF}), total capacitance (C_t), delay time (τ), and power-delay product (PDP).

L_g (nm)	UL (nm)	SS (mV/dec)		I_{ON} (μA/μm)		I_{ON}/I_{OFF}		C_t (fF/μm)		τ (ps)		PDP (fJ/μm)	
		n type	*p* type	*n* type	*p* type	*n* type	*p* type	*n* type	*p* type	*n* type	*p* type	*n* type	*p* type
5	0	120	78	···	219	···	4.38×10^6	···	0.322	···	0.942	···	0.132
	1	82	56	132	337	2.64×10^6	6.74×10^6	0.194	0.215	0.94	0.409	0.079	0.088
	2	69	59	700	350	1.40×10^7	7.56×10^6	0.138	0.153	0.126	0.28	0.057	0.063
4	0	180	97	···	39	···	7.80×10^5	···	0.254	···	4.223	···	0.104
	1	106	75	6	234	1.20×10^5	4.68×10^6	0.143	0.199	16.602	0.546	0.059	0.082
	2	80	60	191	273	3.82×10^6	5.46×10^6	0.126	0.127	0.423	0.297	0.052	0.052
	3	70	63	417	228	8.34×10^6	4.56×10^6	0.096	0.094	0.147	0.264	0.039	0.039
3	0	287	136	···	···	···	···	···	···	···	···	···	···
	1	154	92	···	69	···	1.38×10^6	···	0.153	···	1.423	···	0.063
	2	103	70	21	201	4.20×10^5	4.02×10^6	0.094	0.127	2.856	0.404	0.038	0.052
	3	84	55	129	210	2.58×10^6	4.20×10^6	0.089	0.096	0.441	0.293	0.037	0.039
2	0	616	225	···	···	···	···	···	···	···	···	···	···
	1	250	121	···	3	···	6.00×10^4	···	0.113	···	21.777	···	0.046
	2	146	89	···	66	···	1.32×10^6	···	0.088	···	0.852	···	0.036
	3	109	78	12	132	2.40×10^5	2.64×10^6	0.063	0.078	3.272	0.379	0.026	0.032
	4	92	67	17	138	3.40×10^5	2.76×10^6	0.063	0.067	2.416	0.312	0.026	0.028
1	0	2343	1087	···	···	···	···	···	···	···	···	···	···
	1	429	189	···	···	···	···	···	···	···	···	···	···
	2	223	121	···	4	···	8.00×10^4	···	0.054	···	7.659	···	0.022
	3	156	95	···	25	···	5.00×10^5	···	0.046	···	1.196	···	0.019
	4	119	67	12	62	2.40×10^5	1.24×10^6	0.036	0.038	1.94	0.392	0.015	0.016
12 (IRDS)	···	68		521		5.2×10^6		0.37		0.78		0.47	
5.9 (ITRS)	···	···		295		5.9×10^6		0.69		1.493		0.28	

in Fig. S5 in the Supplemental Material [35]. Hence, an optimal UL length is needed to optimize the ON-state current.

In the simulations, we use homogeneous doping in the electrode regions, which are hard to realize while the electrode regions are very short, so we need to discuss the effect of inhomogeneous doping on the relationship of the ON-state current and the UL. Within the initial increase relationship of the ON-state current with the UL, the enhancement of performance is attributed to the decrease of the SS. SS is defined as

$$\mathrm{SS} = \frac{\partial V_g}{\partial(\log_{10} I)} = \frac{\partial q V_g}{\partial \phi_B} \frac{\partial \phi_B}{\partial(\log_{10} I) q} = \frac{1}{C} \frac{\partial \phi_B}{\partial(\log_{10} I) q}, \tag{11}$$

where V_g and q are the gate voltage and the charge unit, respectively. C is the body factor, a geometric coefficient related to the specific device shape, which is less than 1, and $\partial\phi_B/\partial(\log_{10} I)q$ is the transport factor. The effective barrier change $\Delta\phi_B$ reflects the averages of the body factor C over the OFF state and ON state, since

$$\mathrm{average}(C) = \mathrm{average}\left(\frac{\partial \phi_B}{\partial q V_g}\right) = \frac{\Delta \phi_B}{\Delta q V_g}. \tag{12}$$

From Fig. 2(c), it can be observed that the reduction in SS is caused by an increase in $\Delta\phi_B$ when the UL increases. Since we employ the intrinsic semiconductor as the channel material and only introduce doping in the electrode regions to simulate perfect Ohmic contacts, the body factor C remains unaffected by changes in the doping if the doping does not penetrate to the UL region.

Considering the possibility of dopant diffusion into the UL region in actual processes, the literature [67] has used a Gaussian tail to simulate electrode doping that infiltrates into the UL region. In the simulation of sub-5-nm ML MoS_2 devices, UL structures still exhibit performance enhancement. Therefore, we believe that inhomogeneous

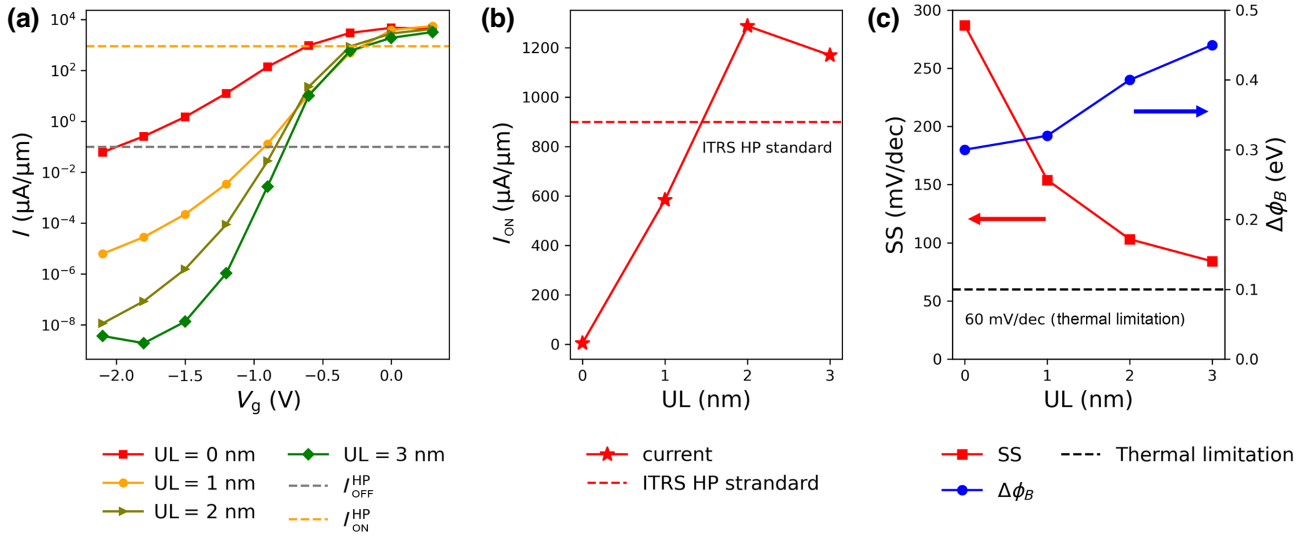

FIG. 2. Effects of the UL structures. (a) Transfer characteristic curves for the 3-nm-gate *n*-type DG ML WSi_2N_4 MOSFETs with UL lengths ranging from 0 to 3 nm. (b) ON-state currents extracted from (a) for HP applications, with the ITRS standard indicated by a red dashed line. (c) SS and effective barrier changes ($\Delta\phi_B$) for the 3-nm-gate *n*-type HP DG ML WSi_2N_4 MOSFETs with UL lengths of 0–3 nm, with SS on the left axis and $\Delta\phi_B$ on the right axis. The black dashed line denotes the thermal limit of the SS.

doping will not affect the conclusion that the optimized UL structure enhances the device's performance.

C. ON-state current

According to ITRS standards, HP applications require an I_{OFF} of 0.1 μA/μm at a gate length of 5.1 nm, while LP applications require an I_{OFF} of 5×10^{-5} μA/μm at a gate length of 5.9 nm. In addition, with a driving voltage of 0.64 V, ITRS specifies ON-state current requirements of 900 μA/μm for HP and 295 μA/μm for LP applications. Figure 4 shows that as gate length decreases, the optimized

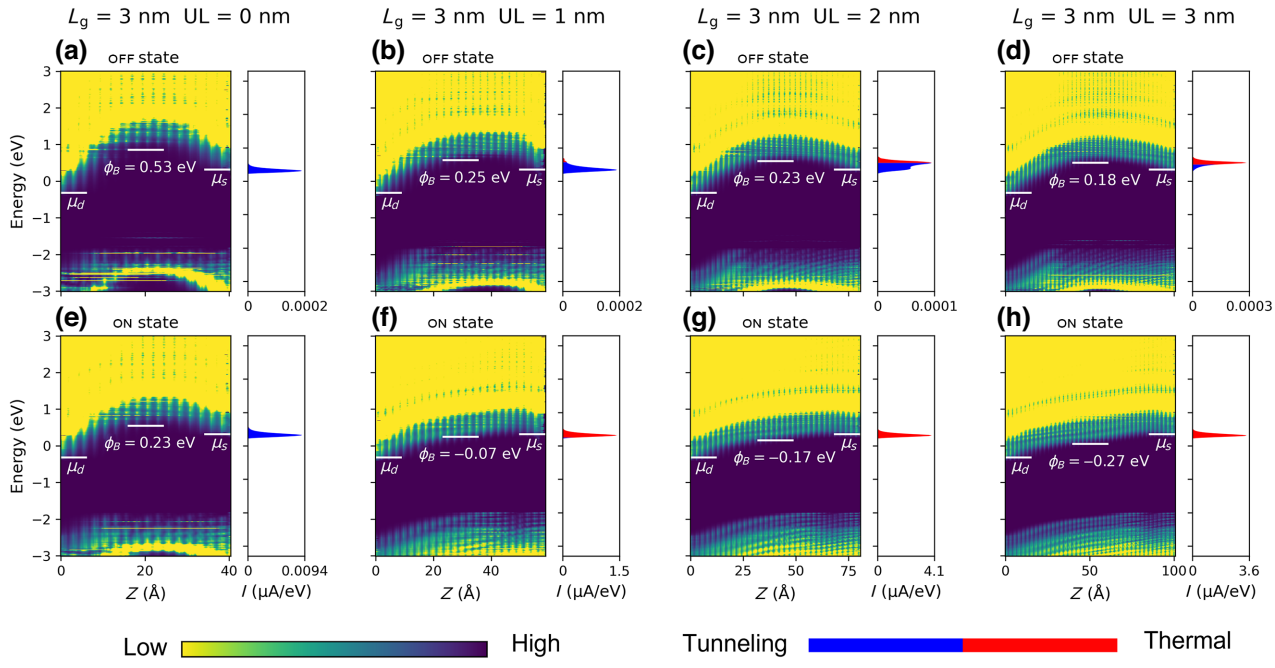

FIG. 3. Projected local density of states (PLDOS) (left in each panel) and current spectra (right in each panel) of the ON-state (a)–(d) and (e)–(h) OFF-state 3-nm-gate *n*-type HP DG ML WSi_2N_4 MOSFET with UL lengths of 0–3 nm. In each panel, μ_s and μ_d denote the chemical potentials of the source and drain, respectively, while ϕ_B represents the height of the effective barrier. The red and blue parts in the current spectra represent the thermal and tunneling currents, respectively.

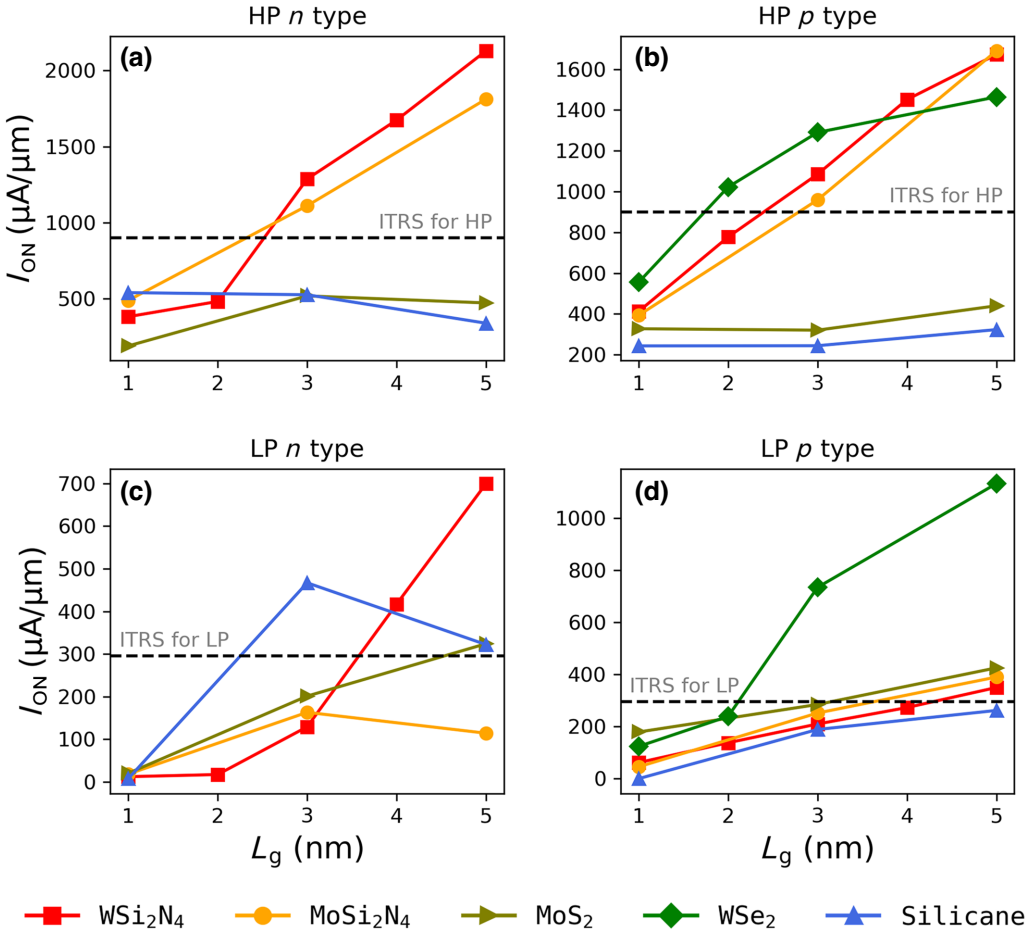

FIG. 4. Optimal ON-state current (I_{ON}) for the DG ML WSi_2N_4 MOSFETs with different gate lengths (L_g). Panels (a)–(d) show the I_{ON} for HP *n*-type devices, HP *p*-type devices, LP *n*-type devices, and LP *p*-type devices, respectively. The optimal I_{ON} for the DG MOSFETs using ML $MoSi_2N_4$, ML MoS_2, ML WSe_2, and ML silicane as the channel material are included for comparison [20,55,69,70]. The black dashed line represents the corresponding ITRS standard.

ON-state currents of *n*-type HP, *n*-type LP, *p*-type HP, and *p*-type LP DG ML WSi_2N_4 MOSFETs all decrease. It is worth noting that the current-optimized UL length of HP devices is always shorter than or equal to that of LP devices with the same doping type and gate length, as shown in Table S2 in the Supplemental Material [35]. This is because when the UL is too long, the undoped UL region accounts for a large proportion of the channel, which reduces the saturation current of the device. Since the working region of the LP device is farther away from the saturation current region than that of the HP device, the LP device is less limited by the saturation current.

For HP applications, both the *n*-type and *p*-type DG ML WSi_2N_4 MOSFETs with gate lengths of 3–5 nm can meet the ON-state requirements of ITRS for HP applications. In this gate region, the *n*-type HP DG ML WSi_2N_4 MOSFETs have higher ON-state currents of 1288–2130 μA/μm compared with the ML $MoSi_2N_4$, ML MoS_2, ML WSe_2, and ML silicane MOSFETs. The ON-state currents of the *p*-type HP DG ML WSi_2N_4 MOSFETs with gate lengths of 3–5 nm are in the range of 1085–1672 μA/μm, surpassing those of ML MoS_2 and ML silicane. The ratios of the ON-state currents for the *n*-type and *p*-type HP DG ML WSi_2N_4 MOSFETs with gate lengths of 3–5 nm are 1.19–1.27, indicating a high degree of symmetry.

For LP applications, the *n*-type LP DG ML WSi_2N_4 MOSFETs with gate lengths of 4–5 nm can meet the ON-state current requirement of ITRS, exhibiting ON-state currents as high as 417–700 μA/μm, which are higher than those of ML $MoSi_2N_4$, ML MoS_2, ML WSe_2, and ML silicane MOSFETs. The *p*-type DG ML WSi_2N_4 MOSFET with $L_g = 5$ nm meets the ON-state current requirement of ITRS with $I_{ON} = 350$ μA/μm, while the 4-nm-gate *p*-type DG ML WSi_2N_4 MOSFET with $I_{ON} = 273$ μA/μm can reach 93% of the corresponding ITRS standard. Therefore, the DG ML WSi_2N_4 MOSFETs enable the development of sub-5-nm-gate CMOS for both HP and LP applications, particularly for HP applications.

D. Subthreshold swing

The SS is a measure of the gate voltage required to change the source-drain current by an order of magnitude in the subthreshold region and is defined as Eq. (11). SS is often used to describe gate control, where a low SS represents strong control of the gate over the channel. Theoretical analysis has shown that SS has a lower bound of 60 mV/dec at room temperature (300 K) for transistors based on hot-electron injection [68].

Figure 5 shows that, in general, the optimal SS increases as the gate length decreases. However, the introduction of ULs can lead to nonmonotonic behaviors, as shown in the optimal SS of the *p*-type DG ML WSi_2N_4 MOSFET [Fig. 5(b)], due to the comprehensive effects discussed in Sec. III.B. The optimal SSs for the *n*-type and *p*-type DG ML WSi_2N_4 MOSFET fall within the ranges of 69–119 and 46–59 mV/dec, respectively. The optimal SSs for the 3-nm- and 5-nm-gate *p*-type DG ML WSi_2N_4 MOSFETs are below the thermal limit due to the tunneling conduction mechanism in such ultrashort channels. In comparison to the corresponding ML $MoSi_2N_4$ MOSFETs, the *n*-type DG ML WSi_2N_4 MOSFETs have nearly the same optimal SSs, while the *p*-type DG ML WSi_2N_4 MOSFETs have slightly higher optimal SS. The optimal SSs of the *p*-type DG ML WSi_2N_4 MOSFETs are significantly lower than those of their silicane counterparts. Figure S6 in the Supplemental Material [35] provides a more detailed relationship between the optimal SS, gate length, and ULs. It is easy to see from Fig. S6 in the Supplemental Material [35] that the SS of the DG ML WSi_2N_4 MOSFET decreases with its UL length, indicating relief of DIBL.

E. Total capacitance, delay time, and power-delay product

Transistor switching is the process of charging and discharging the channel under gate control. In this process, the total capacitance C_t consists of the intrinsic gate capacitance (C_{int}) and the fringe capacitance C_f, as described by Eqs. (13) and (14),

$$C_t = C_{\text{int}} + C_f, \tag{13}$$

$$C_{\text{int}} = \frac{\partial Q_{\text{ch}}}{\partial V_{\text{g}}}, \tag{14}$$

where Q_{ch} represents the charge in the channel under the gate. The fringing capacitance arises from the fringing fields between the fringes of the gate metal and the electrodes. According to the ITRS 2013 standard for the 5-nm-gate-length MOSFET, $C_f = 2C_{\text{int}}$. Smaller C_t results in faster transistor switching. For HP applications, the optimized C_t of both the *n*-type and *p*-type DG ML WSi_2N_4 MOSFET with gate lengths of 1–5 nm is in the range of 0.043–0.206 fF/μm, meeting the ITRS total capacitance requirement of <0.6 fF/μm. For LP applications, the corresponding optimal C_t is in the range of 0.036–0.153 fF/μm, also meeting the ITRS total capacitance requirement of <0.69 fF/μm. The detailed data can be found in Tables I and II.

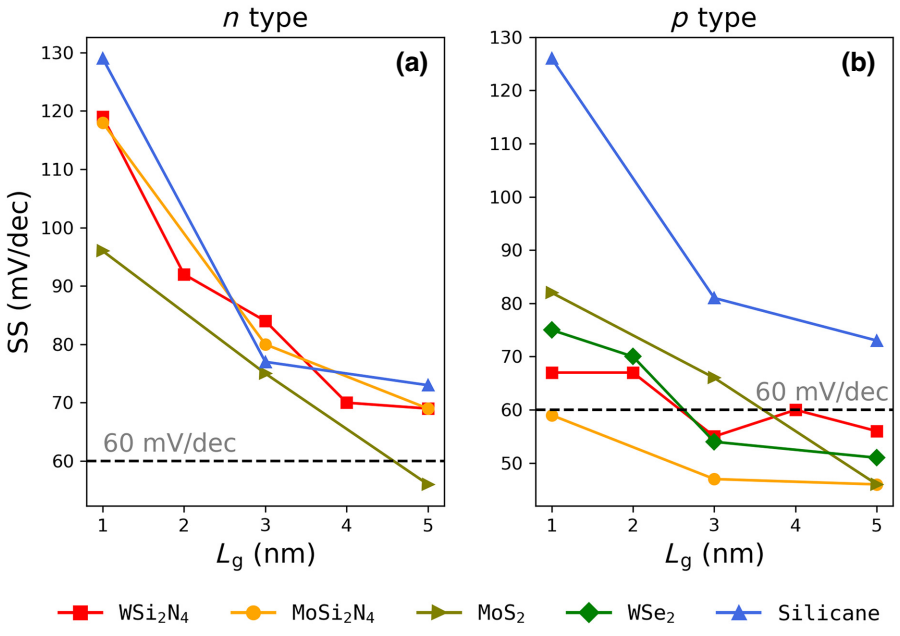

FIG. 5. Optimal SSs of the DG ML WSi_2N_4 MOSFETs with different gate lengths (L_g). (a),(b) Represent *n*-type and *p*-type devices, respectively. For comparison, the optimal SSs of the DG MOSFETs made of other typical 2D materials, including ML $MoSi_2N_4$, ML MoS_2, ML WSe_2, and ML silicane, are also shown [20,55,69,70]. The dashed black line represents the thermal limit.

The delay time (τ) is a measure of the time required for a transistor to switch between its ON state and OFF state. It can be calculated using Eq. (15),

$$\tau = C_t V_{dd}/I_{ON}, \tag{15}$$

where V_{dd}= 0.64 V is the driving voltage and I_{ON} is the ON-state current. A small τ indicates a faster switch. The delay time requirements of ITRS for HP and LP applications are 0.423 and 1.493 ps, respectively. In the case of HP applications [as shown in Figs. 6(a) and 6(b)], the optimal delay times for both the *n*-type and *p*-type sub-5-nm-gate DG ML WSi_2N_4 MOSFETs fall within the range of 0.064–0.112 ps, satisfying the ITRS standard. For LP applications [Figs. 6(c) and 6(d)], except for the *n*-type DG ML WSi_2N_4 MOSFETs with a gate length of 1–2 nm, whose ON-state currents are so low so that their delay times are too high, the optimal delay times of both *n*-type and *p*-type devices with sub-5-nm gate length range from 0.126 to 0.441 ps, meeting the ITRS standard. Overall, the delay times of the ML WSi_2N_4 devices are comparable with those of the ML $MoSi_2N_4$ devices and smaller than those of the MoS_2 devices in the whole sub-5-nm-gate region [Figs. 6(c) and 6(d)].

PDP is a factor reflecting the power consumption of the transistor's ON-OFF switch,

$$\mathrm{PDP} = V_{dd} I_{ON} \tau = C_t V_{dd}^2. \tag{16}$$

Since the driving voltage V_{dd} is fixed at 0.64 eV, the PDP is mainly influenced by C_t. As depicted in Figs. 6(e)–6(h), due to the low C_t values of the devices, the optimal PDPs of the sub-5-nm-gate DG ML WSi_2N_4 MOSFETs are low and all meet the ITRS standards for both HP (PDP < 0.24 fJ/μm) and LP (PDP < 0.28 fJ/μm) applications. For HP (LP) application, PDPs of the *n*-type and *p*-type sub-5-nm-gate DG ML WSi_2N_4 MOSFETs are in the range of 0.018–0.084 fJ/μm (0.016–0.063 fJ/μm). In the sub-5-nm-gate region, the ML WSi_2N_4 devices exhibit lower optimal PDPs compared with their ML WSe_2 and ML MoS_2 counterparts with the same gate length. We summarize the delay times and PDPs for the DG ML WSi_2N_4 MOSFETs with different ULs and gate lengths in Figs. S7 and S8 in the Supplemental Material [35]. In general, the introduction of the UL can decrease the τ and PDP at the same time.

F. Energy-delay product and multiobjective optimization

Ideal transistors should have low power consumption (low PDP) and quick switch speed (low τ) at the same time. The EDP is a useful metric that evaluates the comprehensive effect of PDP and τ,

$$\mathrm{EDP} = \mathrm{PDP}\tau. \tag{17}$$

A low EDP indicates that the transistor can switch quickly while consuming a small amount of power. Figure 7 shows

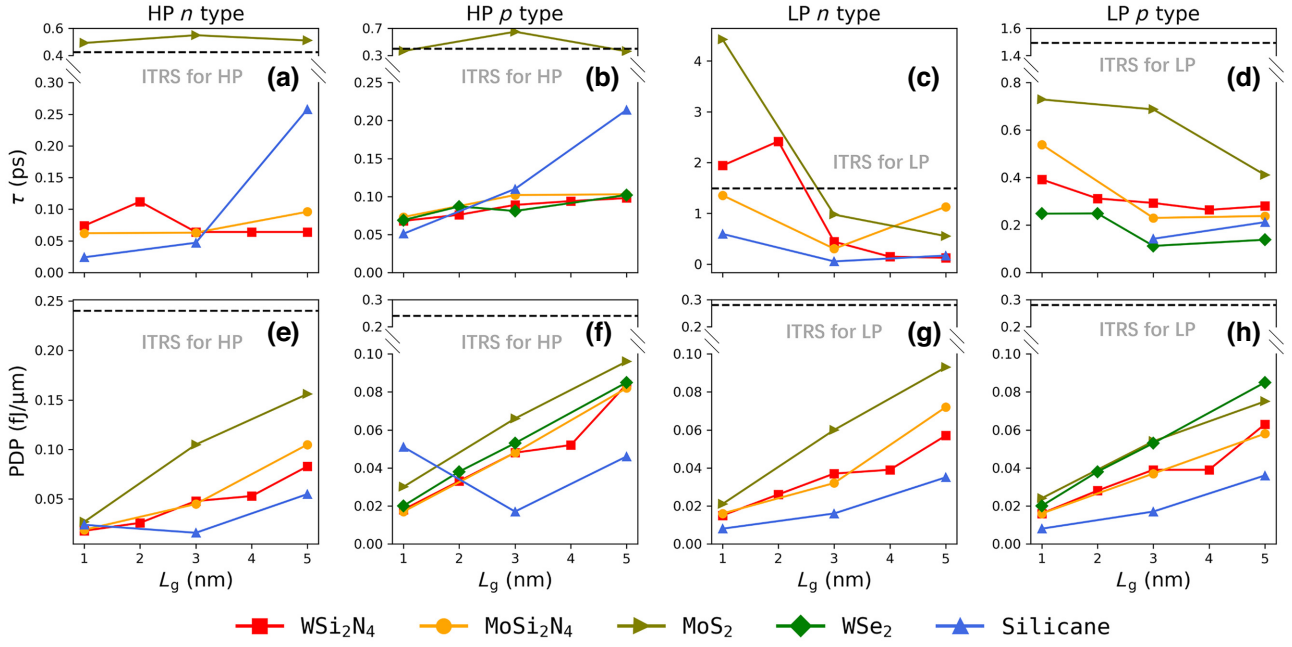

FIG. 6. Optimal delay times (τ) (a)–(d) and power-delay product (PDP) (e)–(h) of the DG ML WSi_2N_4 MOSFETs with different gate lengths. The optimal delay times and PDPs of the DG MOSFETs made of other typical 2D materials (ML $MoSi_2N_4$, ML MoS_2, ML WSe_2, ML silicane) are plotted for comparison [20,55,69,70]. The corresponding ITRS standards are denoted by black dashed lines.

the optimal τ vs PDP plots for the 5-nm-gate DG ML MOSFETs made of ML WSi_2N_4 and other typical 2D ML materials [20,55,66,69–73]. Since both the x axis and the y axis use logarithmic coordinates, the contour lines of EDP on the graph appear as a series of parallel straight lines. For the multiobjective optimization [74,75] of PDP and τ, the Pareto frontier [26,27] of the τ vs PDP plot represents a set of optimal solutions when the weights of τ and PDP changes, and a lower EDP, which is close to the lower left corner, is preferable. We calculated the Pareto frontiers of the τ vs PDP plots for HP and LP applications and show them as red dashed lines in Figs. 7(a) and 7(b), respectively.

As depicted in Fig. 7(a), for HP applications, we can use a thick purple dashed line to divide 2D materials into two tiers. The n-type and p-type ML WSi_2N_4 are in the first tier and located close to each other. Among them, the n-type ML WSi_2N_4 is closest to the lower left corner and lies on the Pareto frontier for HP applications. As shown in Fig. 7(b), for LP applications, these 2D materials can be divided into three tiers. The n-type and p-type ML WSi_2N_4 are in the first tier, with the n-type ML WSi_2N_4 on the Pareto frontier for LP applications. Both the n-type and p-type DG ML WSi_2N_4 MOSFETs with a gate length of 5 nm achieve optimal EDPs that meet the ITRS and IRDS standards for both HP and LP applications. We present the τ vs PDP plots for more gate lengths in Fig. S9 in the Supplemental Material [35]. The p-type 1-nm-gate DG ML WSi_2N_4 MOSFETs are on the Pareto frontiers for HP and LP applications, and the n-type 1-nm-gate DG ML WSi_2N_4 MOSFETs are very close to the Pareto frontiers for HP applications.

G. Discussion

To ensure the good performance of CMOSs, it is necessary to have symmetric characteristics between the n-type and p-type devices. To show the symmetry of the HP DG ML WSi_2N_4 MOSFETs, we compare the SS, I_{ON}, C_t, τ, and PDP of the current-optimized n-type and p-type HP DG ML WSi_2N_4 MOSFETs in Figs. 8(a)–8(e). The ratios of these parameters of the n-type devices to those of the p-type devices are also quantified in Fig. 8(f). As the devices with L_g in the range of 3–5 nm meet the ON-state current requirement of the ITRS for HP applications, we focus

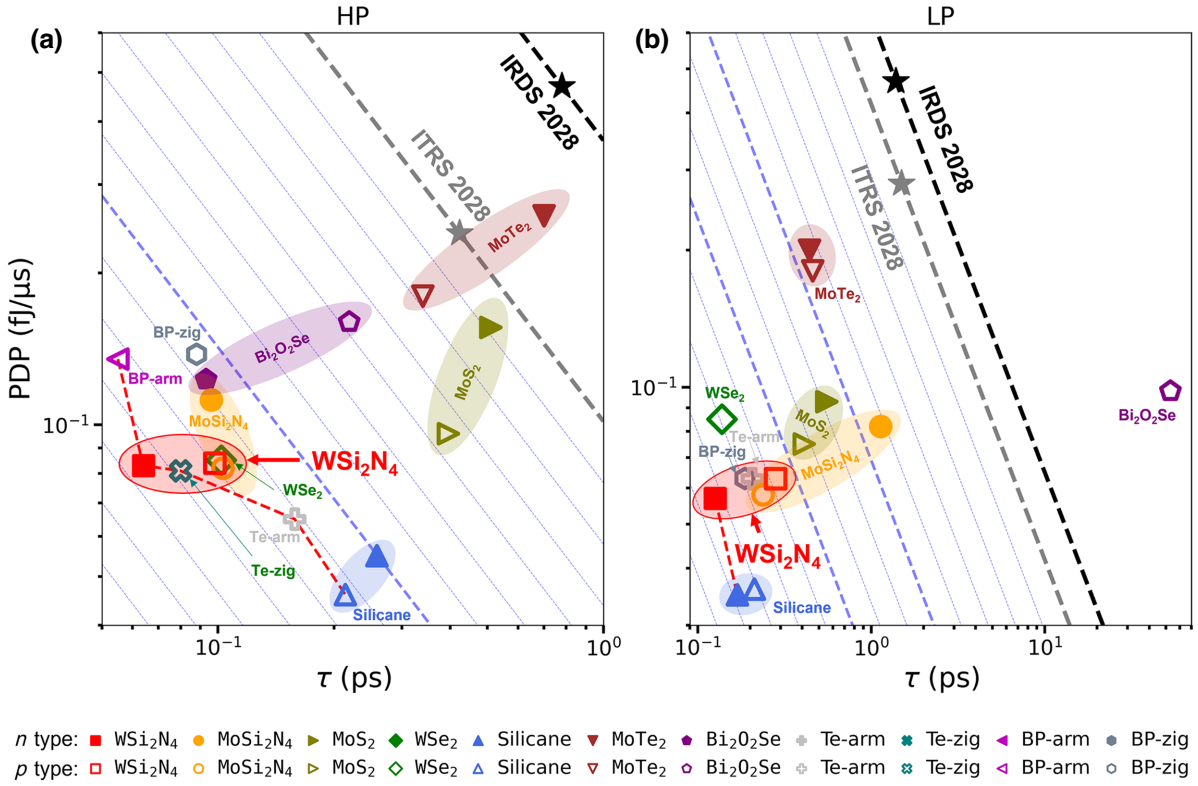

FIG. 7. Optimal delay time versus power dissipation (τ vs PDP) plots for the 5-nm-gate DG MOSFETs made of ML WSi_2N_4 and other typical 2D materials [20,55,66,69–73]. The criteria of ITRS 2013 and IRDS 2022 for the 2028 technical node are denoted by the black and gray stars, respectively. The purple dashed lines represent the contours of EDP. The 2D materials are divided into two and three tiers by the thick purple dashed lines for HP and LP applications, respectively. The red dashed lines denote the Pareto frontiers [26,27] for multiobjective optimization of τ and PDP. Abbreviations: tellurene (Te), black phosphorene (BP), armchair direction (arm), and zigzag direction (zig). The EDP of the n-type 5-nm-gate Bi_2O_2Se device is much higher than the IRDS and ITRS standards so we did not show this data in the figure.

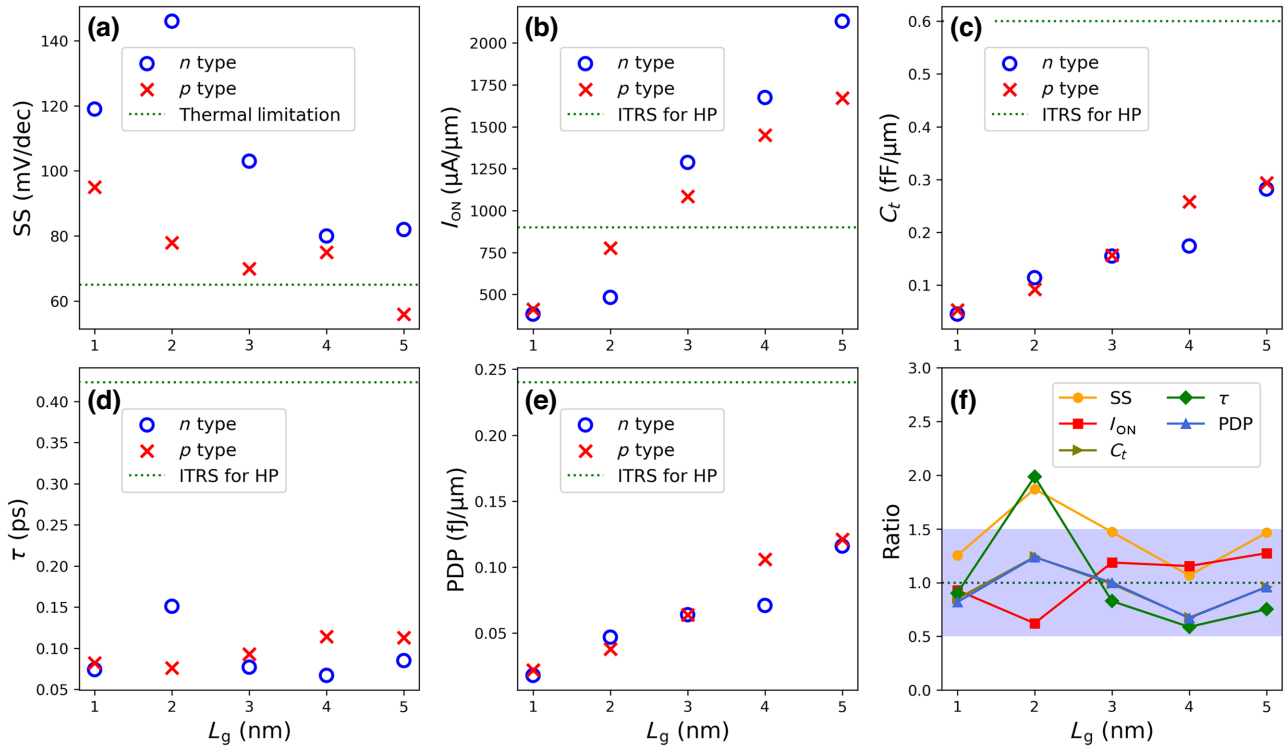

FIG. 8. Comparison of the *n*-type and *p*-type current-optimized DG ML WSi_2N_4 MOSFETs for HP applications at different gate lengths: (a) subthreshold swing (SS), (b) ON-state current (I_{on}), (c) total capacitance (C_t), (d) delay time (τ), (e) power-delay product (PDP), (f) ratios of the parameters of the *n*-type devices to those of the *p*-type devices.

on the ratios in this gate-length range. In this gate-length range, the ratios of the SS, I_{ON}, C_t, τ, and PDP are all in the range of 0.5–1.5, indicating a high degree of symmetry. In particular, the ratio of the ON-state current falls between 1.19 and 1.27, demonstrating excellent symmetry.

Our calculations are based on ballistic transport, so the parameters we calculate apply to temperatures where ballistic transport is maintained. At low temperatures, phonon scattering is weak, and the mean free path of carriers is larger than the device with ultrashort channels, in which case the device can be considered to be operating under ballistic transport. As the temperature increases, phonon scattering becomes stronger, and the device gradually transitions from ballistic transport to diffusive transport. We calculate the mobilities for electrons and holes of the monolayer WSi_2N_4 as a function of temperature based on the deformation potential theory [76,77] calibrated by the more accurate nonadiabatic molecular dynamics results [78]. The corresponding mean free paths can be obtained by the 2D free electron gas model [79]. (For more detailed information about the calculation of mobility and mean free paths, please refer to the Appendix.) It can be observed that both the electron and hole mean free paths decrease with increasing temperature, with the electron mean free path generally higher than that of holes, as shown in Fig. 9. Therefore, the properties of the *p*-type device deteriorate faster with increasing temperature, and the symmetry of the device decreases. Fortunately, both the electron and hole mean free paths can remain longer than 10 nm below 400 K, which is longer than all the channel lengths involved in our calculations, allowing our devices to primarily operate in the ballistic transport regime. It is undeniable that the symmetry of the *n*-type and *p*-type device parameters decreases with increasing temperature, but this can be partially compensated by separately optimizing the UL lengths for the *n*-type and *p*-type devices.

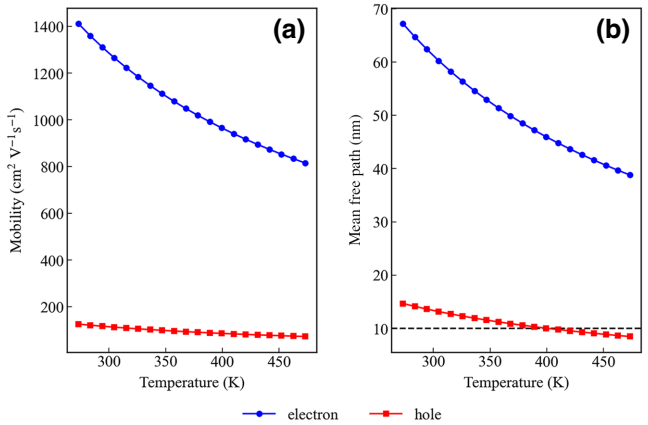

FIG. 9. Variation of the mobility (a) and the mean free path (b) of the ML WSi_2N_4 with temperature.

So, we estimate that the temperature range for symmetric behavior can be maintained below 400 K. If the temperature exceeds this threshold, the figure of merit for the *p*-type devices will significantly deteriorate compared to that of the *n*-type devices.

Many previous studies on ultrashort-channel FETs have discussed the relationship between ballistic rate and channel length. In the case of sub-10-nm channel lengths, 2D transistors primarily operate in the ballistic regime. The NEGF calculations based on the Born approximation, considering electron-phonon interactions at room temperature, have shown that in the ML MoS_2 devices, when the channel length is 10, 5, and 3 nm, the ballistic rate is 76%, 89%, and 96%, respectively [62]. In recent experiments with a 10-nm channel length, the three-layer InSe FET exhibited a measured room-temperature ballistic rate of 83% [48], and the three-layer MoS_2 with a 10-nm channel length showed a measured room-temperature ballistic rate of 80% [52].

Since substitutional doping will break the crystal structure of the 2D material, doping is usually realized by direct contact with metals [80], where the impurity of atoms is absent. If employing van der Waals contacts, the impact on the crystal structures of the 2D materials is minimized [47]. In our calculations, the doping caused by contacts is simulated by introducing the atomic compensation charges [81] for each atom in the source and drain regions,

$$\tilde{\rho}_i^{\text{atom}}(r) = c_i \times \rho_i^{\text{atom}}, \tag{18}$$

where ρ_i^{atom} is the intrinsic electron density of the ith atom in the neutral system and $\tilde{\rho}_i^{\text{atom}}(r)$ is the electron density of the ith atom after introducing the compensation charges. c_i is the rescaling factor. (Refer to the Supplemental Material [35] for more details.) In the calculations described previously, we use uniform doping (a constant c_i for all the atoms in the doped regions) without considering the disorder in doping concentration. Actually, the electron density distributions may not be uniform. To quantify this disorder, we sample from a Gaussian distribution and assign the atomic compensation charges proportional to this sampled factor,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)((x-\mu)/\sigma)^2}. \tag{19}$$

Here, μ corresponds to the rescaling factor in the case of uniform doping. We set σ to be 0, 1% of μ, and 5% of μ, respectively, and perform simulations on the *n*-type devices with $L_g = 5$ nm and UL = 1 nm. The resulting transport curves and ON-state currents are shown in Fig. 10. It can be observed that the disorder in charge density has minimal impact on the transport curves, and both the HP and LP ON-state currents slightly decrease with increasing disorder percentage.

The effective mass in the parabolic approximation characterizes the band-edge shape and has a significant impact on the ON-state current of a MOSFET. To qualitatively describe the relationship between ON-state current and effective mass, we use the classic equation $I = n\,e\,v$, where n, e, v represent carrier density, electron charge, and carrier velocity, respectively. Effective mass affects both carrier density and carrier velocity. Carrier density is proportional to the density of states (DOS) near the band edge, which can be calculated using DOS $= \frac{g_s g_v}{2\pi\hbar^2}\sqrt{m_x^* m_y^*}$, where g_s, g_v, $\hbar$, m_x^*, and m_y^* are the spin degeneracy, valley degeneracy, reduced Planck constant, transverse effective mass, and transport effective mass, respectively. Meanwhile, carrier velocity can be expressed as $v = \frac{eEt}{m^*}$, where E and t are the electric field and relaxation time, respectively.

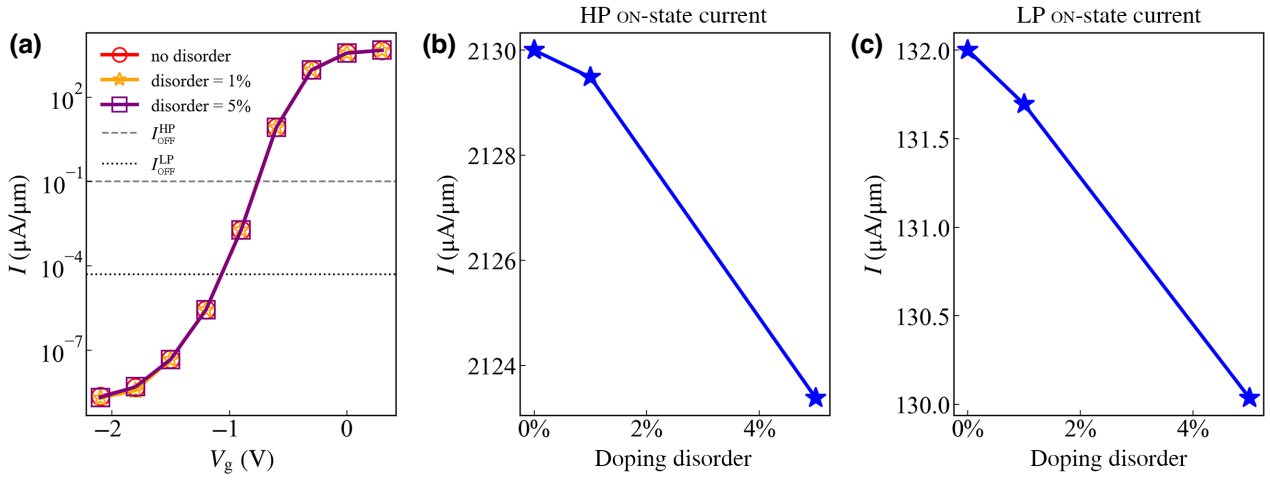

FIG. 10. Impact of disorder in doping concentration. (a) Transport curves of the *n*-type DG ML WSi_2N_4 transistors with $L_g = 5$ nm and UL = 1 nm, for disorder percentages of 0% of μ, 1% of μ, and 5% of μ. (b) and (c) show the ON-state currents for HP and LP applications, respectively, extracted from (a).

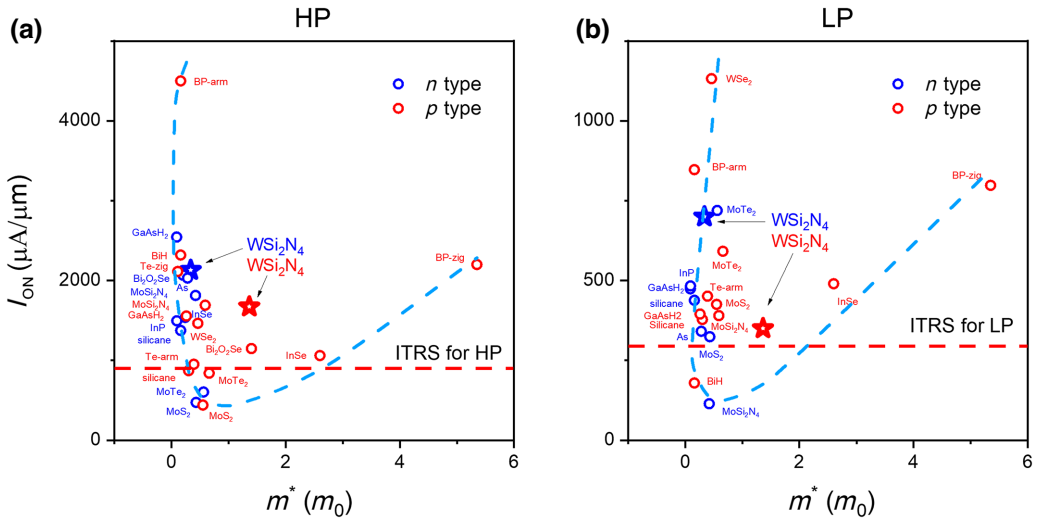

FIG. 11. Relationship of the ON-state current of the MOSFETs for HP (a) and LP (b) applications and the effective mass of the ML channel materials. We choose the optimal ON-state currents at the gate length of 5 nm for comparison. All the data are from the first-principles quantum transport calculations [20,55,59,66,69–73,82–85]. The dashed light blue arcs are guides to the eyes. The blue and red markers represent the *n*-type and *p*-type MOSFETs, respectively.

The optimal ON-state currents of the 5-nm-gate MOSFETs made from ML materials with different effective masses are presented in Fig. 11. All the data are obtained from *ab initio* quantum transport calculations [20,55,59,66,69–73,82–85]. Notably, the ON-state current exhibits a U-shaped dependence on the effective mass, with the bottoms of the U-shape occurring at around $0.7m_0$ and $0.5m_0$ for HP and LP applications, respectively. For ML WSi_2N_4, neither the electron ($0.34m_0$) nor hole ($1.36m_0$) effective mass along the transport direction falls within the low-current range of $0.5m_0$–$0.7m_0$, leading to high ON-state currents. In contrast, the electron ($0.43m_0$ and $0.43m_0$) and hole ($0.59m_0$ and $0.45m_0$) effective masses of ML $MoSi_2N_4$ and MoS_2 are close to the $0.5m_0$ edge, resulting in low ON-state currents.

IV. CONCLUSION

In conclusion, we demonstrate the potential of the DG ML WSi_2N_4 MOSFETs with a sub-5-nm gate length for both HP and LP applications using first-principles DFT combined with NEGF methods. Our study shows that both *n*-type and *p*-type DG ML WSi_2N_4 MOSFETs meet the requirements of ITRS for HP applications until L_g scales down to 3 nm. Additionally, the ON-state current ratios between the *n*-type and *p*-type HP devices show a high degree of symmetry. For LP applications, the scale limits of the *n*-type and *p*-type DG ML WSi_2N_4 MOSFETs are 4 and 5 nm, respectively, and the ON-state current of the *p*-type 5-nm-gate-length DG ML WSi_2N_4 MOSFET is significantly higher than those of other typical ML material MOSFETs. We find that the optimal 5-nm-gate-length *n*-type DG ML WSi_2N_4 MOSFETs for HP and LP applications both lie on the Pareto frontiers of the optimal τ vs PDP plots, demonstrating their potential for achieving optimal trade-offs between speed and energy efficiency. Our results suggest that employing DG ML WSi_2N_4 MOSFETs can enable the development of homogenous CMOS devices in the sub-5-nm-gate region for both HP and LP applications, particularly for HP applications.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Science and Technology of China (Grant No. 2022YFA1203904), the National Natural Science Foundation of China (Grants No. 12274002 and No. 91964101), the Foundation of He'nan Educational Committee (Grant No. 23A430015), the open research fund of the National Center for International Research on Intelligent Nano-Materials and Detection Technology in Environmental Protection (Grant No. SDGH2106), and the High-performance Computing Platform of Peking University and the MatCloud + high throughput materials simulation engine. Y.L. is grateful to Dr. Hengwei Luan, Dr. Sixuan Li, and Miss Yuqi Chen for fruitful discussions.

APPENDIX: CALCULATION OF MOBILITY AND MEAN FREE PATHS

Based on the deformation potential theory proposed by Bardeen and Shockley [76], the carrier mobility of 2D materials can be calculated using the following

equation [77]:

$$\mu_{2D} = \frac{e\hbar^3 C_{2D}}{k_B T m^* m_a (E_l^i)^2}, \tag{A1}$$

where e, $\hbar$, and k_B are the elementary charge, reduced Planck's constant, and Boltzmann constant, respectively. T, m^*, m_a, and C_{2D}, represent the temperature, effective mass in the transport direction, average effective mass, and elastic modulus, respectively. E_l^i is the deformation potential constant for either the hole located at the valence band maximum or the electron located at the conduction band minimum along the transport direction. Since the deformation potential theory often overestimates the carrier mobility, we use the carrier mobility at 300 K obtained from the more accurate nonadiabatic molecular dynamics calculations for calibration [78]. Further, we estimate the mean free path l based on the 2D free electron gas model [79],

$$l = \frac{\hbar\mu}{e}\sqrt{\frac{4\,\pi\,n}{g_s}}, \tag{A2}$$

where μ, n, and g_s represent the carrier mobility, carrier concentration, and degeneracy, respectively.

[1] S. Calebotta, CMOS, the ideal logic family, Natl. Semicond. CMOS Datab. Rev. **1**, 2 (1975).

[2] G. E. Moore, Cramming more components onto integrated circuits, Proc. IEEE **86**, 82 (1998).

[3] A. P. Jacob, R. Xie, M. G. Sung, L. Liebmann, R. T. Lee, and B. Taylor, Scaling challenges for advanced CMOS devices, Int. J. High Speed Electron. Syst. **26**, 1740001 (2017).

[4] Y. Liu, X. Duan, H.-J. Shin, S. Park, Y. Huang, and X. Duan, Promises and prospects of two-dimensional transistors, Nature **591**, 43 (2021).

[5] K. Khan, A. K. Tareen, M. Aslam, R. Wang, Y. Zhang, A. Mahmood, Z. Ouyang, H. Zhang, and Z. Guo, Recent developments in emerging two-dimensional materials and their applications, J. Mater. Chem. C **8**, 387 (2020).

[6] R. Bian, C. Li, Q. Liu, G. Cao, Q. Fu, P. Meng, J. Zhou, F. Liu, and Z. Liu, Recent progress in the synthesis of novel two-dimensional van der Waals materials, Natl. Sci. Rev. **9**, nwab164 (2022).

[7] S. Das, A. Sebastian, E. Pop, C. J. McClellan, A. D. Franklin, T. Grasser, T. Knobloch, Y. Illarionov, A. V. Penumatcha, and J. Appenzeller, Transistors based on two-dimensional materials for future integrated circuits, Nat. Electron. **4**, 786 (2021).

[8] M. Chhowalla, D. Jena, and H. Zhang, Two-dimensional semiconductors for transistors, Nat. Rev. Mater. **1**, 1 (2016).

[9] M.-Y. Li, S.-K. Su, H.-S. P. Wong, and L.-J. Li, How 2D semiconductors could extend Moore's law, Nature **567**, 169 (2019).

[10] A.-J. Cho, K. C. Park, and J.-Y. Kwon, A high-performance complementary inverter based on transition metal dichalcogenide field-effect transistors, Nanoscale Res. Lett. **10**, 1 (2015).

[11] P. J. Jeon, J. S. Kim, J. Y. Lim, Y. Cho, A. Pezeshki, H. S. Lee, S. Yu, S.-W. Min, and S. Im, Low power consumption complementary inverters with n-MoS_2 and p-WSe_2 dichalcogenide nanosheets on glass for logic and light-emitting diode circuits, ACS Appl. Mater. Interfaces **7**, 22333 (2015).

[12] M. H. Chiu, H. L. Tang, C. C. Tseng, Y. Han, A. Aljarb, J. K. Huang, Y. Wan, J. H. Fu, X. Zhang, and W. H. Chang, Metal-guided selective growth of 2D materials: Demonstration of a bottom-up CMOS inverter, Adv. Mater. **31**, 1900861 (2019).

[13] J. Y. Lim, A. Pezeshki, S. Oh, J. S. Kim, Y. T. Lee, S. Yu, D. K. Hwang, G. H. Lee, H. J. Choi, and S. Im, Homogeneous 2D $MoTe_2$ p–n junctions and CMOS inverters formed by atomic-layer-deposition-induced doping, Adv. Mater. **29**, 1701798 (2017).

[14] Y.-L. Hong, Z. Liu, L. Wang, T. Zhou, W. Ma, C. Xu, S. Feng, L. Chen, M.-L. Chen, and D.-M. Sun, Chemical vapor deposition of layered two-dimensional $MoSi_2N_4$ materials, Science **369**, 670 (2020).

[15] L. Wang, Y. Shi, M. Liu, A. Zhang, Y.-L. Hong, R. Li, Q. Gao, M. Chen, W. Ren, and H.-M. Cheng, Intercalated architecture of MA_2Z_4 family layered van der Waals materials with emerging topological, magnetic and superconducting properties, Nat. Commun. **12**, 1 (2021).

[16] Q. Wang, L. Cao, S.-J. Liang, W. Wu, G. Wang, C. H. Lee, W. L. Ong, H. Y. Yang, L. K. Ang, and S. A. Yang, Efficient Ohmic contacts and built-in atomic sublayer protection in $MoSi_2N_4$ and WSi_2N_4 monolayers, npj 2D Mater. Appl. **5**, 1 (2021).

[17] S. B. Touski and N. Ghobadi, Vertical strain-induced modification of the electrical and spin properties of monolayer $MoSi_2X_4$ (X = N, P, As and Sb), J. Phys. D: Appl. Phys. **54**, 485302 (2021).

[18] Y. Cai, G. Zhang, and Y.-W. Zhang, Polarity-reversed robust carrier mobility in monolayer MoS_2 nanoribbons, J. Am. Chem. Soc. **136**, 6269 (2014).

[19] B. Mortazavi, B. Javvaji, F. Shojaei, T. Rabczuk, A. V. Shapeev, and X. Zhuang, Exceptional piezoelectricity, high thermal conductivity and stiffness and promising photocatalysis in two-dimensional $MoSi_2N_4$ family confirmed by first-principles, Nano Energy **82**, 105716 (2021).

[20] X. Sun, Z. Song, N. Huo, S. Liu, C. Yang, J. Yang, W. Wang, and J. Lu, Performance limit of monolayer $MoSi_2N_4$ transistors, J. Mater. Chem. C **9**, 14683 (2021).

[21] J. Huang, P. Li, X. Ren, and Z.-X. Guo, Promising properties of a sub-5-nm monolayer $MoSi_2N_4$ transistor, Phys. Rev. Appl. **16**, 044022 (2021).

[22] S. International Technology Roadmap for Semiconductors (ITRS) 2013 Edition. https://www.semiconductors.org/resources/2013-international-technology-roadmap-for-semiconductors-itrs/.

[23] N. Ghobadi, M. Hosseini, and S. B. Touski, Field-effect transistor based on $MoSi_2N_4$ and WSi_2N_4 monolayers under biaxial strain: A computational study of the electronic properties, IEEE Trans. Electron Devices **69**, 863 (2022).

[24] M. W. Chuan, K. L. Wong, A. Hamzah, S. Rusli, N. E. Alias, C. S. Lim, and M. L. P. Tan, A review of the top of the barrier nanotransistor models for semiconductor nanomaterials, Superlattices Microstruct. **140**, 106429 (2020).

[25] A. Rahman, J. Guo, S. Datta, and M. S. Lundstrom, Theory of ballistic nanotransistors, IEEE Trans. Electron Devices **50**, 1853 (2003).

[26] A. V. Lotov and K. Miettinen, in *Multiobjective Optimization* (Springer, Berlin, 2008), p. 213.

[27] R. T. Marler and J. S. Arora, The weighted sum method for multi-objective optimization: New insights, Struct. Multidiscip. Optim. **41**, 853 (2010).

[28] M. Brandbyge, J.-L. Mozos, P. Ordejón, J. Taylor, and K. Stokbro, Density-functional method for nonequilibrium electron transport, Phys. Rev. B **65**, 165401 (2002).

[29] S. Datta, *Quantum Transport: Atom to Transistor* (Cambridge university Press, New York, 2005).

[30] S. Smidstrup, T. Markussen, P. Vancraeyveld, J. Wellendorff, J. Schneider, T. Gunst, B. Verstichel, D. Stradi, P. A. Khomyakov, and U. G. Vej-Hansen, QuantumATK: An integrated platform of electronic and atomic-scale modelling tools, J. Phys.: Condens. Matter **32**, 015901 (2019).

[31] S. Smidstrup, D. Stradi, J. Wellendorff, P. A. Khomyakov, U. G. Vej-Hansen, M.-E. Lee, T. Ghosh, E. Jónsson, H. Jónsson, and K. Stokbro, First-principles Green's-function method for surface calculations: A pseudopotential localized basis set approach, Phys. Rev. B **96**, 195309 (2017).

[32] D. Stradi, U. Martinez, A. Blom, M. Brandbyge, and K. Stokbro, General atomistic approach for modeling metal-semiconductor interfaces using density functional theory and nonequilibrium Green's function, Phys. Rev. B **93**, 155302 (2016).

[33] K. Y. Camsari, S. Chowdhury, and S. Datta, in *Springer Handbook of Semiconductor Devices* (Springer, Berlin, German, 2022), p. 1583.

[34] W. Y. Kim and K. S. Kim, Carbon nanotube, graphene, nanowire, and molecule-based electron and spin transport phenomena using the nonequilibrium Green's function method at the level of first principles theory, J. Comput. Chem. **29**, 1073 (2008).

[35] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevApplied.20.064044 for the flowchart of the self-consistent DFT + NEGF; the structure, band structure, and projected local density of states (PLDOS) of ML WSi_2N_4; transfer characteristic curves for devices with different UL; relationships between I_{ON} and UL under different L_g; peak value of the ON-state current spectra changes with UL; SS, τ, and PDP as a function of L_g under various UL; τ vs PDP plots for MOSFETs made of ML WSi_2N_4 and other typical 2D materials with various L_g; realization of gate electrodes, dielectric regions, and doping in the simulation model; performance in the doping test; optimal UL length in terms of the ON-state currents with various L_g.

[36] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, New York, 1997).

[37] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. **77**, 3865 (1996).

[38] M. J. van Setten, M. Giantomassi, E. Bousquet, M. J. Verstraete, D. R. Hamann, X. Gonze, and G.-M. Rignanese, The PseudoDojo: Training and grading a 85 element optimized norm-conserving pseudopotential table, Comput. Phys. Commun. **226**, 39 (2018).

[39] M. Shishkin and G. Kresse, Self-consistent *GW* calculations for semiconductors and insulators, Phys. Rev. B **75**, 235102 (2007).

[40] M. S. Hybertsen and S. G. Louie, Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies, Phys. Rev. B **34**, 5390 (1986).

[41] K. S. Thygesen and A. Rubio, Conserving *GW* scheme for nonequilibrium quantum transport in molecular contacts, Phys. Rev. B **77**, 115333 (2008).

[42] J. Ryou, Y.-S. Kim, S. Kc, and K. Cho, Monolayer MoS_2 bandgap modulation by dielectric environments and tunable bandgap transistors, Sci. Rep. **6**, 1 (2016).

[43] Y. Zhang, T.-R. Chang, B. Zhou, Y.-T. Cui, H. Yan, Z. Liu, F. Schmitt, J. Lee, R. Moore, and Y. Chen, Direct observation of the transition from indirect to direct bandgap in atomically thin epitaxial $MoSe_2$, Nat. Nanotechnol. **9**, 111 (2014).

[44] Y. Liang and L. Yang, Carrier plasmon induced nonlinear band gap renormalization in two-dimensional semiconductors, Phys. Rev. Lett. **114**, 063001 (2015).

[45] S. Liu, J. Li, B. Shi, X. Zhang, Y. Pan, M. Ye, R. Quhe, Y. Wang, H. Zhang, and J. Yan, Gate-tunable interfacial properties of in-plane ML MX_2 1T′–2H heterojunctions, J. Mater. Chem. C **6**, 5651 (2018).

[46] W. A. Strauss, *Partial Differential Equations: An Introduction* (John Wiley & Sons, Chichester, England, 2007).

[47] W. Li, X. Gong, Z. Yu, L. Ma, W. Sun, S. Gao, Ç Köroğlu, W. Wang, L. Liu, and T. Li, Approaching the quantum limit in two-dimensional semiconductor contacts, Nature **613**, 274 (2023).

[48] J. Jiang, L. Xu, C. Qiu, and L.-M. Peng, Ballistic two-dimensional InSe transistors, Nature **616**, 470 (2023).

[49] S. B. Desai, S. R. Madhvapathy, A. B. Sachid, J. P. Llinas, Q. Wang, G. H. Ahn, G. Pitner, M. J. Kim, J. Bokor, and C. Hu, MoS_2 transistors with 1-nanometer gate lengths, Science **354**, 99 (2016).

[50] W. Li, J. Zhou, S. Cai, Z. Yu, J. Zhang, N. Fang, T. Li, Y. Wu, T. Chen, and X. Xie, Uniform and ultrathin high-κ gate dielectrics for two-dimensional electronic devices, Nat. Electron. **2**, 563 (2019).

[51] C. Zhang, T. Tu, J. Wang, Y. Zhu, C. Tan, L. Chen, M. Wu, R. Zhu, Y. Liu, and H. Fu, Single-crystalline van der Waals layered dielectric with high dielectric constant, Nat. Mater. **22**, 832 (2023).

[52] J. Jiang, L. Xu, L. Du, L. Li, G. Zhang, C. Qiu, and L. Peng, Yttrium-induced phase-transition technology for forming perfect ohmic contact in two-dimensional MoS_2 transistors, Research Square (2023).

[53] R. Quhe, L. Xu, S. Liu, C. Yang, Y. Wang, H. Li, J. Yang, Q. Li, B. Shi, and Y. Li, Sub-10 nm two-dimensional transistors: Theory and experiment, Phys. Rep. **938**, 1 (2021).

[54] Z. Ni, M. Ye, J. Ma, Y. Wang, R. Quhe, J. Zheng, L. Dai, D. Yu, J. Shi, and J. Yang, Performance upper limit of sub-10 nm monolayer MoS_2 transistors, Adv. Electron. Mater. **2**, 1600191 (2016).

[55] H. Zhang, B. Shi, L. Xu, J. Yan, W. Zhao, Z. Zhang, Z. Zhang, and J. Lu, Sub-5 nm monolayer MoS_2 transistors toward low-power devices, ACS Appl. Electron. Mater. **3**, 1560 (2021).

[56] L. Xu, J. Yang, C. Qiu, S. Liu, W. Zhou, Q. Li, B. Shi, J. Ma, C. Yang, and J. Lu, Can carbon nanotube transistors be scaled down to the sub-5 nm gate length?, ACS Appl. Mater. Interfaces **13**, 31957 (2021).

[57] L. Liu, J. Han, L. Xu, J. Zhou, C. Zhao, S. Ding, H. Shi, M. Xiao, L. Ding, and Z. Ma, Aligned, high-density semiconducting carbon nanotube arrays for high-performance electronics, Science **368**, 850 (2020).

[58] C. Qiu, F. Liu, L. Xu, B. Deng, M. Xiao, J. Si, L. Lin, Z. Zhang, J. Wang, and H. Guo, Dirac-source field-effect transistors as energy-efficient, high-performance electronic switches, Science **361**, 387 (2018).

[59] Y. Wang, R. Fei, R. Quhe, J. Li, H. Zhang, X. Zhang, B. Shi, L. Xiao, Z. Song, and J. Yang, Many-body effect and device performance limit of monolayer InSe, ACS Appl. Mater. Interfaces **10**, 23344 (2018).

[60] V. Trivedi, J. G. Fossum, and M. M. Chowdhury, Nanoscale FinFETs with gate-source/drain underlap, IEEE Trans. Electron Devices **52**, 56 (2004).

[61] J. Wang, Q. Cai, J. Lei, G. Yang, J. Xue, D. Chen, B. Liu, H. Lu, R. Zhang, and Y. Zheng, Performance of monolayer blue phosphorene double-gate MOSFETs from the first principles, ACS Appl. Mater. Interfaces **11**, 20956 (2019).

[62] L. Liu, Y. Lu, and J. Guo, On monolayer MoS_2 field-effect transistors at the scaling limit, IEEE Trans. Electron Devices **60**, 4133 (2013).

[63] International Roadmap for Devices and Systems (IRDS) 2022 Edition. https://irds.ieee.org/editions/2022.

[64] M. S. Lundstrom, *Fundamentals of Nanotransistors* (World Scientific Publishing Company, Singapore, 2017), Vol. 6.

[65] M. Lundstrom and Z. Ren, Essential physics of carrier transport in nanoscale MOSFETs, IEEE Trans. Electron Devices **49**, 133 (2002).

[66] R. Quhe, Q. Li, Q. Zhang, Y. Wang, H. Zhang, J. Li, X. Zhang, D. Chen, K. Liu, and Y. Ye, Simulations of quantum transport in sub-5-nm monolayer phosphorene transistors, Phys. Rev. Appl. **10**, 024022 (2018).

[67] Z. Dong and J. Guo, Assessment of 2-D transition metal dichalcogenide FETs at sub-5-nm gate length scale, IEEE Trans. Electron Devices **64**, 622 (2017).

[68] I. Ferain, C. A. Colinge, and J.-P. Colinge, Multigate transistors as the future of classical metal–oxide–semiconductor field-effect transistors, Nature **479**, 310 (2011).

[69] X. Sun, L. Xu, Y. Zhang, W. Wang, S. Liu, C. Yang, Z. Zhang, and J. Lu, Performance limit of monolayer WSe_2 transistors; significantly outperform their MoS_2 counterpart, ACS Appl. Mater. Interfaces **12**, 20633 (2020).

[70] Y. Pan, J. Dai, L. Xu, J. Yang, X. Zhang, J. Yan, J. Li, B. Shi, S. Liu, and H. Hu, Sub-5-nm monolayer silicane transistor: A first-principles quantum transport simulation, Phys. Rev. Appl. **14**, 024016 (2020).

[71] Q. Li, J. Yang, Q. Li, S. Liu, L. Xu, C. Yang, L. Xu, Y. Li, X. Sun, and J. Yang, Sub-5 nm gate length monolayer $MoTe_2$ transistors, J. Phys. Chem. C **125**, 19394 (2021).

[72] R. Quhe, J. Liu, J. Wu, J. Yang, Y. Wang, Q. Li, T. Li, Y. Guo, J. Yang, and H. Peng, High-performance sub-10 nm monolayer Bi_2O_2Se transistors, Nanoscale **11**, 532 (2019).

[73] J. Yan, H. Pang, L. Xu, J. Yang, R. Quhe, X. Zhang, Y. Pan, B. Shi, S. Liu, and L. Xu, Excellent device performance of sub-5-nm monolayer tellurene transistors, Adv. Electron. Mater. **5**, 1900226 (2019).

[74] L. Zadeh, Optimality and non-scalar-valued performance criteria, IEEE Trans. Autom. Control **8**, 59 (1963).

[75] K. Deb, in *Search Methodologies* (Springer, New York, 2014), p. 403.

[76] J. Bardeen and W. Shockley, Deformation potentials and mobilities in non-polar crystals, Phys. Rev. **80**, 72 (1950).

[77] J. Xie, Z. Zhang, D. Yang, D. Xue, and M. Si, Theoretical prediction of carrier mobility in few-layer BC_2N, J. Phys. Chem. C **5**, 4073 (2014).

[78] P. Zhao, Z.-Y. Jiang, J.-M. Zheng, Y.-M. Lin, and A. Du, Theoretical study of a novel $WSi_2N_4/MoSi_2N_4$ heterostructure with ultrafast carrier transport, J. Phys. Chem. C **126**, 11380 (2022).

[79] Solid State Physics Part I: Transport Properties of Solids, https://web.mit.edu/6.732/www/6.732-pt1.pdf.

[80] Y. Wang, S. Liu, Q. Li, R. Quhe, C. Yang, Y. Guo, X. Zhang, Y. Pan, J. Li, and H. Zhang, Schottky barrier heights in two-dimensional field-effect transistors: from theory to experiment, Rep. Prog. Phys. **84**, 056501 (2021).

[81] Doping methods available in QUANTUMATK, https://docs.quantumatk.com/manual/technicalnotes/doping_methods/doping_methods.html.

[82] M. Ye, S. Liu, H. Zhang, B. Shi, J. Li, X. Zhang, J. Yan, and J. Lu, Sub-5 nm monolayer BiH transistors, ACS Appl. Electron. Mater. **1**, 2103 (2019).

[83] Q. Li, S. Fang, S. Liu, L. Xu, L. Xu, C. Yang, J. Yang, B. Shi, J. Ma, and J. Yang, Performance limit of ultrathin GaAs transistors, ACS Appl. Mater. Interfaces **14**, 23597 (2022).

[84] L. Xu, R. Quhe, Q. Li, S. Liu, J. Yang, C. Yang, B. Shi, H. Tang, Y. Li, and X. Sun, Device performance and strain effect of sub-5 nm monolayer InP transistors, J. Mater. Chem. C **10**, 2223 (2022).

[85] Y. Wang, P. Huang, M. Ye, R. Quhe, Y. Pan, H. Zhang, H. Zhong, J. Shi, and J. Lu, Many-body effect, carrier mobility, and device performance of hexagonal arsenene and antimonene, Chem. Mater. **29**, 2191 (2017).