Promising Properties of a Sub-5-nm Monolayer $MoSi_2N_4$ Transistor

Junsheng Huang,[1] Ping Li,[1] Xiaoxiong Ren,[1] and Zhi-Xin Guo[1,2,3,*]

[1] *State Key Laboratory for Mechanical Behavior of Materials, Center for Spintronics and Quantum System, School of Materials Science and Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China*

[2] *Key Laboratory of Polar Materials and Devices, Ministry of Education, East China Normal University, Shanghai 200241, China*

[3] *Department of Physics and Institute for Nanophysics, Xiangtan University, Xiangtan 411105, China*

(Received 4 June 2021; revised 27 August 2021; accepted 24 September 2021; published 13 October 2021)

Two-dimensional (2D) semiconductors have attracted tremendous interest as natural passivation and atomically thin channels could facilitate continued transistor scaling. However, air-stable 2D semiconductors with high performance are quite elusive. Recently, an extremely-air-stable $MoSi_2N_4$ monolayer was successfully fabricated [Hong *et al*., Science 369, 670 (2020)]. To further reveal its potential application in sub-5-nm metal-oxide-semiconductor field-effect transistors (MOSFETs), there is an urgent need to develop integrated circuits. Here, we report first-principles quantum-transport simulations on the performance limits of *n*- and *p*-type sub-5-nm monolayer (ML) $MoSi_2N_4$ MOSFETs. We find that the on-state current in the $MoSi_2N_4$ MOSFETs can be effectively manipulated by the length of gate and underlap, as well as the doping concentration. Very strikingly, we also find that for the *n*-type devices the optimized on-state currents can reach up to 1390 and 1025 $\mu A/\mu m$ for high-performance and low-power (LP) applications, respectively, both of which satisfy the International Technology Roadmap for Semiconductors (ITRS) requirements. The optimized on-state current can meet the LP application (348 $\mu A/\mu m$) for *p*-type devices. Finally, we find that the $MoSi_2N_4$ MOSFETs have an ultralow subthreshold swing and power-delay product, which have the potential to realize high-speed and low-power consumption devices. Our results show that $MoSi_2N_4$ is an ideal 2D channel material for future competitive ultrascaled devices.

DOI: 10.1103/PhysRevApplied.16.044022

I. INTRODUCTION

The downscaling of field-effect transistors (FETs) to the sub-7-nm channel length is in great demand for integrated circuits in the next decade [1,2]. However, traditional silicon FETs are suffering from the challenges of the short-channel effect, increased leakage current, and unnecessary power consumption [3–5]. To better solve these problems, numerous researchers have focused their attention on two-dimensional (2D) materials. Compared with that of three-dimensional materials, the uniform thickness and smooth surfaces of 2D materials are beneficial for high-speed application due to their suppressed carrier scattering and possible trap generation. Moreover, the atomically thin thickness of 2D materials make them ideal gate electrostatics with diminished short-channel effects and leakage currents [6–14].

To date, layered transition-metal dichalcogenides (TMDCs), such as MoS_2, are the most-studied 2D semiconductor material in FETs [15–19]. For example, the bilayer MoS_2 transistor with a 1-nm carbon-nanotube (CNT) gate has been experimentally fabricated and found to exhibit an excellent gate-control ability with a subthreshold swing (SS) of about 65 mV/dec and a remarkable on:off current ratio of 10^6 [15]. However, the on-state current ($<250\,\mu A\,\mu m^{-1}$) of 2D MoS_2 FETs is too low to meet the International Technology Roadmap for Semiconductors (ITRS) [20] standard for both high-performance (HP) and low-power (LP) applications [15–19], owing to the small carrier mobility of MoS_2. On the other hand, some 2D materials, such as 2D InSe and black phosphorene (BP), have much higher carrier mobility and excellent performance with high on currents, which can satisfy the ITRS requirements for both HP and LP standards with a device scaled down to about 5 nm [21–23]. However, their instability in air is an issue, which greatly limits device fabrication and applications [21,23]. Although numerous efforts have been devoted to exploring 2D materials, such as silicene, tellurene, Bi_2O_2Se, WSe_2, Ge-Se, ReS_2, BiN, AsP [24–33], for sub-5-nm FETs, a 2D material with both excellent device performance and air stability is still scarce.

Recently, the 2D material $MoSi_2N_4$ was successfully synthesized by chemical vapor deposition (CVD) [34]. This material has excellent ambient stability, e.g., it can be stable even upon immersion in an aqueous solution of HCl

*zxguo08@xjtu.edu.cn

for 24 h [34]. $MoSi_2N_4$ is also a semiconductor, the band gap of which can be efficiently tuned by either strain or an electric field [35–37]. In combination with the characteristics of highly controllable growth and good ohmic contact with electrodes, such as Ti, Sc, Ni, and NbS_2, $MoSi_2N_4$ has great potential applications in next-generation FET devices [34,38,39].

Here, we theoretically investigate the performance of *n*- and *p*-type sub-5-nm double-gated (DG) monolayer (ML) $MoSi_2N_4$ MOSFETs by using *ab initio* quantum-transport calculations. We find that the transmission channel of ML $MoSi_2N_4$ locates in the MoN_2 layer sandwiched between two Si-N layers (as indicated in Fig. 1). The main critical device properties, including the on-state current (I_{on}), SS, delay time (τ), and power-delay product (PDP), are taken into consideration. More importantly, we also find that the on-state current of the optimized *n*-type ML $MoSi_2N_4$ MOSFET for HP is as high as 1390 μA/μm. The *n*-type ML $MoSi_2N_4$ MOSFET can meet the ITRS standards for HP and LP devices when scaled down to 3 and 1 nm, respectively. The *p*-type ML $MoSi_2N_4$ MOSFET can only satisfy the ITRS LP requirement when it scaled down to 3 nm. Moreover, we find that the $MoSi_2N_4$ MOSFET has excellent gate controllability with quite a small SS, and its delay time and PDP are small enough to meet the ITRS HP and LP requirements.

II. METHOD

Geometric optimization and electronic structures of monolayer $MoSi_2N_4$ are calculated by density-functional theory (DFT) with the projector-augmented-wave (PAW) method, which is implemented in the Vienna *ab initio* simulation package (VASP) [40–42]. The exchange-correlation function is described based on the generalized gradient approximation (GGA) in the form of Perdew-Burke-Ernzerhof (PBE) parameterization [43]. The convergence standards of the atomic energy and positions are less than 1×10^{-6} eV per atom and 1×10^{-2} eV Å^{-1}, respectively. The cutoff energy of the wave function is set to 500 eV. The Brillouin zone is sampled by a $15 \times 15 \times 1$ Monkhorst-Pack *k*-point mesh [44] for geometrical optimization and $21 \times 21 \times 1$ mesh for electronic states.

The transport properties are simulated based on the DFT method combined with the nonequilibrium Green's function (NEGF) formalism, using the Atomistix ToolKit (ATK) 2019 package [45,46]. The drain current at a given bias voltage, V_b, and gate voltage, V_g, is calculated through the Landauer-Büttiker formula [47]:

$$I(V_g, V_b) = \frac{2e}{h}\int_{-\infty}^{+\infty}\{T(E, V_b, V_g)[f_S(E-\mu_S) - f_D(E-\mu_D)]\}dE, \quad (1)$$

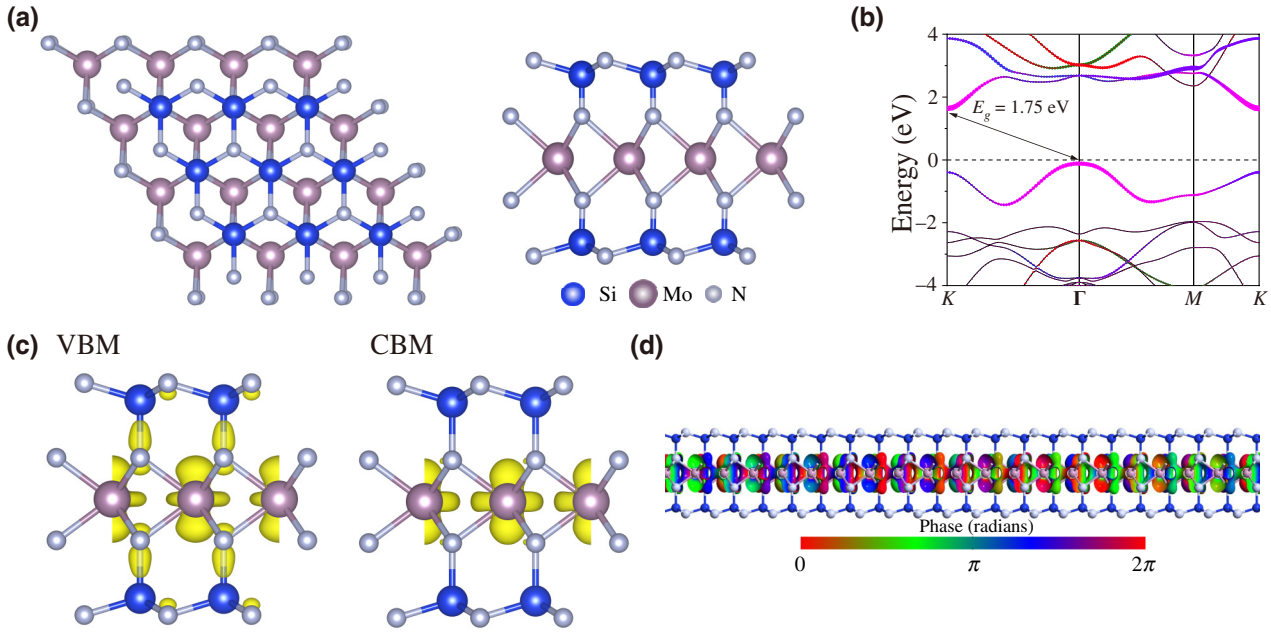

FIG. 1. (a) Top view and side view of ML $MoSi_2N_4$; dashed lines represent the primitive unit cell of ML $MoSi_2N_4$. Blue, purple, and white balls represent silicon, molybdenum, and nitrogen atoms, respectively. (b) Band-decomposed charge-density distributions corresponding to valence-band maximum (VBM) and conduction-band minimum (CBM) of ML $MoSi_2N_4$. Isovalue is 0.02 e/bohr^3. (c) Electronic band structure of monolayer $MoSi_2N_4$. Different orbitals of Mo are mapped with different colors: Mo d_{xy} orbital, blue; Mo d_{yz} orbital, green; Mo d_{xz} orbital, red; Mo d_{z^2} orbital, magenta; Mo $d_{x^2-y^2}$ orbital, purple; Mo *s* orbital, orange. (d) Transmission eigenstates at *K* point (1/3,1/3) and $E = 0.32$ eV under on-state conditions ($V_g = 0$ V). Isovalue is 0.2 arb. units.

where $T(E,V_b,V_g)$ is the transmission coefficient; f_S and f_D stand for the Fermi-Dirac distribution functions for the source and drain, respectively. μ_S and μ_D stand for the electrochemical potentials of the source and drain, respectively. In the calculations, the tier 3 basis set is adopted with Hartwigsen-Goedecker-Hutter pseudopotentials and GGA in the form of the PBE function is utilized to represent the exchange and correlation interactions. Because of the heavily screened electron-electron interaction by doping carriers, the DFT GGA-based single-electron approximation is good in the description of the device's electronic structure [22,48,49]. The real-space mesh cutoff is chosen to be 75 hartree, and the k-point meshes [44] are set as Monkhorst-Pack $11 \times 1 \times 129$ and $11 \times 1 \times 1$ meshes for the electrode region and the central region in the Brillouin zone. Moreover, the boundary condition along the transverse, vertical, and transport directions are set to be of periodic, Neumann, and Dirichlet type, respectively [50].

III. RESULT AND DISCUSSION

A. Channel materials and device configuration

The atomic structure of ML $MoSi_2N_4$ is shown in Fig. 1(a), which can be regarded as a MoN_2 layer sandwiched between two Si-N layers [44]. The optimized lattice parameter of ML $MoSi_2N_4$ is 2.91 Å, in good agreement with previous results [34]. The band-decomposed charge-density distributions [Fig. 1(b)] on the VBM and CBM show that the CBM and VBM of monolayer $MoSi_2N_4$ are mainly contributed to by the middle MoN_2 layer. Figure 1(c) further shows that the valence and conduction bands of ML $MoSi_2N_4$ are mainly contributed to by the d_{z^2} and $d_{x^2-y^2}$ orbitals of Mo atoms, and Si and N atoms contribute little to them (Fig. S1 within the Supplemental Material [51]). This feature means that the middle MoN_2 layer would be responsible for electron transmission, which is verified by transmission-eigenstate calculations, as shown in Fig. 1(d). Since the transmission pathway in MoN_2 is protected by the two outer Si-N bilayers, $MoSi_2N_4$ can be recognized as a natural microwire with a conducting wire surrounded by an insulating wire.

The performance of $MoSi_2N_4$ MOSFETs is further investigated by using the ATK package. In the simulation, double gates are adopted, and degenerately doped ML $MoSi_2N_4$ is selected as the two-probe electrodes [Fig. 2(a)]. Compared with the single-gate case, dual gates can obviously increase the competence of gate modulation [52]. The electron-transport direction is along the zigzag direction of $MoSi_2N_4$, and the gate is assumed to be an ideal rectangle in the device model. In calculations, electrodes with a length of about 1 nm are adopted, and the dielectric constant of silicon dioxide is set to be 3.9. As indicated in Fig. 2(a), in the channel, the gate region is usually shorter than the dielectric region, and the uncovered dielectric region is called the underlap region. Previous studies found that gate underlap, which is the spacer area between the gate and electrode, played an essential role in the scalability of the gate length for the modeled FETs [22]. An appropriate length of the gate underlap, L_{UL}, could improve the device performance. Thus, one can define the channel length (L_{ch}) as the sum of the gate length (L_g) and twice the length of the underlap (L_{UL}), namely, $L_{\mathrm{ch}} = L_g + 2L_{\mathrm{UL}}$. In calculations, the atomic-compensation-charge method is used for doping in the electrodes and a doping concentration of 1.0×10^{13} cm^{-2} is adopted, unless otherwise specified. According to the ITRS 2013 edition requirements for HP and LP standards of a sub-5-nm device in 2028, we use 0.64 V as the supply voltage (V_{dd}) and 0.41 nm as the equivalent oxide thickness (EOT) of a dielectric material (silicon dioxide). Various lengths of gate underlap, ranging from 0 to 4 nm, are considered for a comprehensive investigation of the performance of $MoSi_2N_4$ MOSFETs.

B. On-state current

On-state current (I_{on}) is a key parameter for evaluating the transition speed of a logic device. A high I_{on} is advantageous for efficient applications, such as high-performance servers with a high switching velocity. One can calculate I_{on} (HP) and I_{on} (LP) by applying $V_g(\text{on-HP}) = V_g(\text{off-HP}) \pm V_{dd}$ and $V_g(\text{on-LP}) = V_g(\text{off-LP}) \pm V_{dd}$, respectively. To obtain the value of the on-state current, the first step is to make these DG ML $MoSi_2N_4$ MOSFETs reach the off-state current. Following the requirements of ITRS for the off-state current, I_{off} of HP and LP devices are set to 0.1 and 5×10^{-5} μA/μm, respectively.

The calculated transfer characteristics of sub-5-nm gate-length $MoSi_2N_4$ MOSFETs are shown in Figs. 2(b)–2(g). As shown in Figs. 2(b)–2(g), all of the MOSFETs have a small enough source-drain leakage current to reach the off-state requirements for high-performance standards. This is due to the large band gap and simple energy-band state in the conductive region. However, for the low-power application with $L_{\mathrm{UL}} < 2$ nm, the 1-nm gate-length device cannot reach the requirement of the off-state current due to the short-channel effect. The introduction of the underlap region enlarges the effective channel length and improves the MOSFETs' ability to reach the off-state current, especially for the device with a gate length of $L_g = 1$ nm. When the gate length is long enough, increasing UL has little effect on the device performance. This is because the introduction of UL will inhibit the tunneling of carriers in the off state, especially for devices with a short gate, which has a relatively large tunneling effect. For devices with a long gate, the tunneling current is small and so is the impact of UL. These phenomena can be observed by comparing Figs. 2(b) and 2(d), where the I-V curves for the 5-nm gate-length device with different L_{UL} are almost

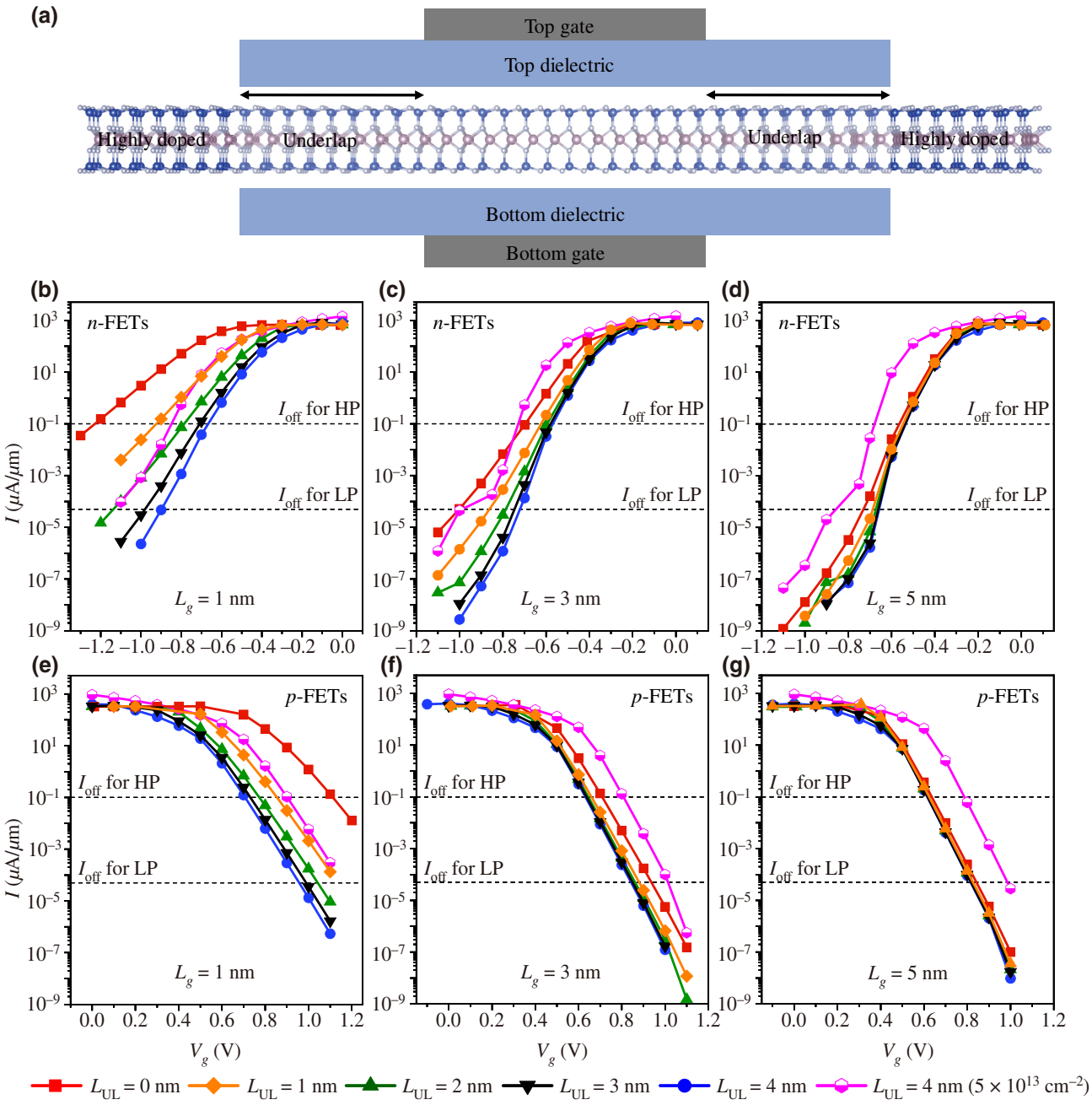

FIG. 2. (a) Schematic diagram of the DG ML $MoSi_2N_4$ MOSFETs. (b)–(g) I-V_g characteristics of n- and p-type FETs with different gate lengths and L_{UL} for $V_b = 0.64$ V.

overlapped in Fig. 2(d). We additionally explore the effect of increasing doping concentration on the MOSFET current, i.e., with 5.0×10^{13} cm^{-2} electron doping of the electrodes for the device of $L_{UL} = 4$ nm. Compared with the low-doping-concentration case, the increase in doping concentration obviously increases the current under the same V_g.

We further summarize the values of I_{on} for the sub-5-nm MOSFETs in Fig. 3. It is seen that I_{on} generally monotonically increases as the gate length increases. For the HP ML $MoSi_2N_4$ devices [Figs. 3(a) and 3(b)], the on-state current for the high-doping-concentration case (5.0×10^{13} cm^{-2}) is much higher than that of the low doping concentration (1.0×10^{13} cm^{-2}). It is seen that only the n-type MOSFET with a high concentration can meet the ITRS HP requirement (900 μA/μm), reaching up to 1206 and 1390 μA/μm for 3- and 5-nm gate lengths, respectively. Nevertheless, the highest I_{on} of the low-doping-concentration case reaches 817 μA/μm, fulfilling 91% of the ITRS standard [Fig. 3(a)]. As for the p-type MOSFETs with low

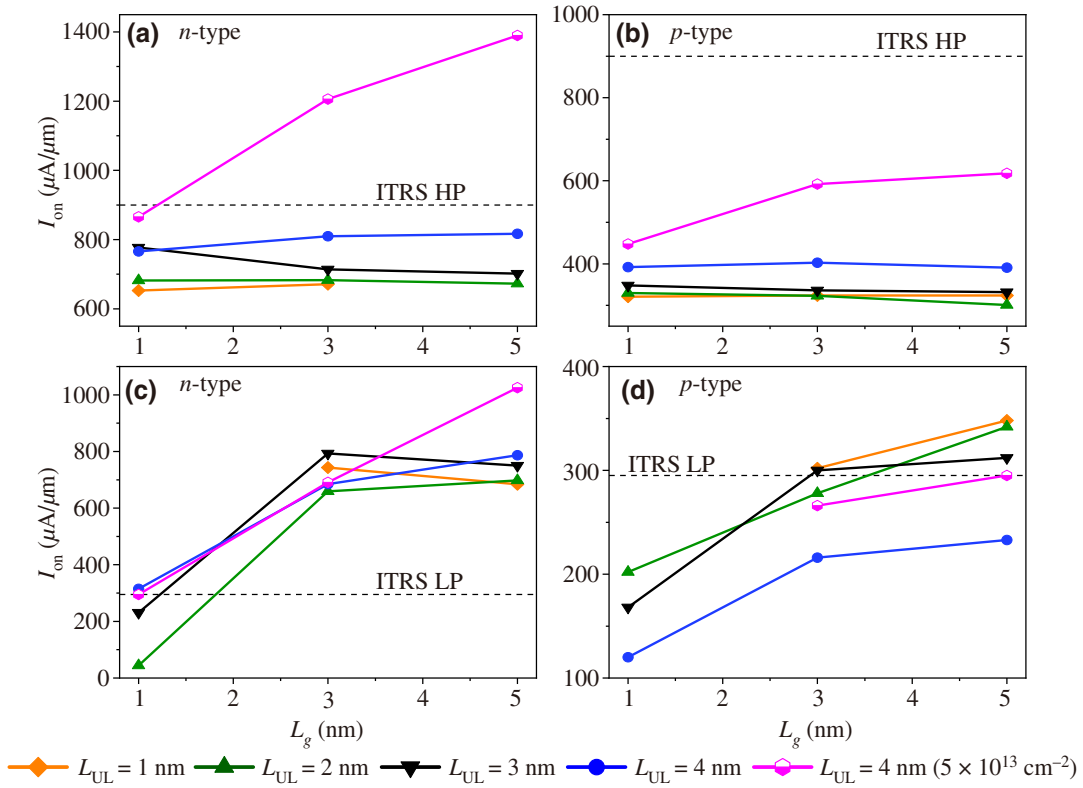

FIG. 3. (a)–(d) On-state current as a function of gate length for *n*- (a, c) and *p*-type (b, d) FETs with different L_{UL} for the HP (a, b) and LP (c, d) applications, respectively. Black dashed lines represent the ITRS HP and LP requirements.

concentration, the highest I_{on} appears for a 3-nm gate-length MOSFET with $L_{UL} = 4$ nm [Fig. 3(b)]. Similar to the *n*-type MOSFETs, I_{on} increases as the doping concentration increasing, where the *p*-type 5-nm gate-length MOSFET could reach an $I_{on} \approx 618$ μA/μm with a doping concentration of 5.0×10^{13} cm^{-2}. This value fulfills 69% of the ITRS standard. Notably, the introduction of the underlap region can help to increase I_{on} for the HP ML $MoSi_2N_4$ devices, in most cases.

As for the LP application, I_{on} of the *n*-type and *p*-type ML $MoSi_2N_4$ devices increases as the gate length increases from 1 to 3 nm, both of which can meet the ITRS HP standard (295 μA/μm) under certain L_{UL} [Figs. 3(c) and 3(d)]. The largest I_{on} of the *n*-type MOSFETs are 793 and 1025 μA/μm for a low doping concentration and a high doping concentration, respectively. The I_{on} value of *n*-type MOSFETs can satisfy the ITRS LP standard requirement, even as the gate length decreases to 1 nm with $L_{UL} = 4$ nm (315 μA/μm) [Fig. 3(c)]. As for the *p*-type device, the minimum gate length to fulfill the LP requirement is 3 nm [Fig. 3(d)], where the largest I_{on} (348 μA/μm) appears with the 5-nm gate length and $L_{UL} = 1$ nm.

It is observed that the increase of underlap does not always improve the on-state current of the device [as indicated in Figs. 3(c) and 3(d)]. This feature can be due to two competing mechanisms: on one hand, an increase of the underlap region makes the channel barrier longer, which decreases the transmission possibility and suppresses the short-channel effect (positive effect). On the other hand, the gate-controlling capability of the underlap region becomes weaker as its length increasing, which would degrade the performance of the device (negative effect). These two conflicting effects imply that the length of the underlap should be optimized to obtain the highest on-state current.

To illustrate the function of underlap and the modulation mechanism of the gate more clearly, we calculate the local density of states (LDOS) and transmission spectra of the 1-nm-gate ML $MoSi_2N_4$ MOSFETs with different L_{UL} (Fig. 4) for the HP case. Here, we define the maximum-electron-barrier height, Φ_m, as the energy barrier for transport from the source to the drain. As shown in Figs. 4(a)–4(c), with the same off-state current of 0.1 μA/μm, Φ_m is reduced from 0.26 eV at $L_{UL} = 0$ nm to 0.19 and 0.17 eV at $L_{UL} = 2$ and 4 nm, respectively. The calculated transmission spectra confirm the variation of the transmission barrier [Fig. 4(d)]. When a voltage of 0.64 V is applied, the CBM of the ML $MoSi_2N_4$ in the channel region would move downward, leading to the on state of the MOSFETs [Figs. 4(e)–4(g)]. From Fig. 3(a), I_{on} with $L_{UL} = 4$ nm is the highest, followed by the case with $L_{UL} = 2$ nm, and I_{on} in the one without UL is the lowest. This is because the CBM becomes lower as the L_{UL} increases, which

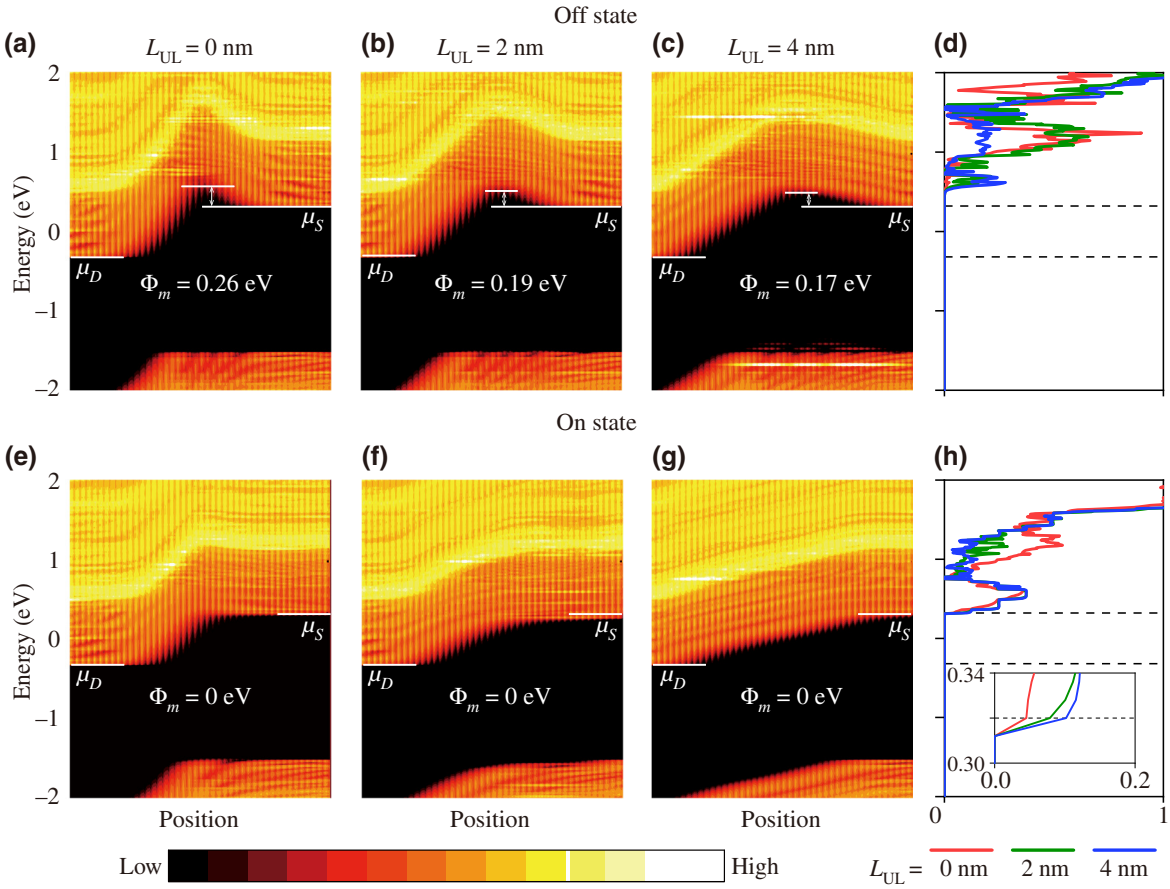

FIG. 4. Spatially resolved LDOS and transmission spectra for *n*-type FETs in the off (a)–(d) and on states (e)–(h) with 1-nm gate length. (a), (b), (c) represent the LDOS of off state with underlap length of 0 nm, 2 nm, 4 nm, respectively. (e), (f), (g) are the on-state LDOS with underlap length of 0 nm, 2 nm, 4 nm, respectively. (d) and (h) correspond to the transmission spectra of off and on states with different underlap length, respectively. Transmission spectra near the CBM of $MoSi_2N_4$ are shown in the inset in (h). μ_S and μ_D are the electrochemical potentials of the source and drain, respectively.

leads to a higher on-state current. Correspondingly, the on-state transmission-spectra edge at $L_{UL} = 4$ nm is uppermost within the bias window, which is consistent with the result of the highest on-state current for $L_{UL} = 4$ nm [Fig. 4(h)]. The variation of the UL leads to different carrier barrier heights, Φ_m and barrier lengths, resulting in differences of gate control.

C. Gate control

The gate-control ability of the FET in the subthreshold region is usually described by the SS, which impacts on the device's performance and decides the operating voltage of the device. The definition of the SS is

$$S = \frac{\partial V_g}{\partial(\log_{10} I_{DS})} \tag{2}$$

where I_{DS} is the drain current. Note that the smaller SS corresponds to a better gate-control ability. Figure 5 shows the calculated SS with different gate lengths (L_g) and underlap (L_{UL}). It is found that SS generally increases with decreasing L_g and L_{UL}. In addition, for the *n*-type device without underlap ($L_{UL} = 0$), the SS value sharply increases from 57 to 166 mV/dec as L_g decreases from 5 to 1 nm. This feature shows that underlap is beneficial for reducing SS. This is because devices with a long L_{UL} can reach the off-state current with a smaller V_g, which leads to a larger slope in the I-V_g curve. Such an effect is more noticeable in the relatively-short-gate-length devices. For example, in the *n*-type case with $L_{UL} = 4$ nm, the SS of $L_g = 1$ nm MOSFETs can be reduced by 55% (from 166 to 75 mV/dec), while it can only be reduced by 23% (from 57 to 44 mV/dec) for $L_g = 5$ nm. A similar phenomenon is also observed in the *p*-type case. It is observed that the minimum SS of the *n*-type and *p*-type $MoSi_2N_4$ MOSFETs are 44 and 58 mV/dec, respectively. Both of these values are smaller than the Boltzmann tyranny (60 mV/dec), which is believed to be the fundamental limit of SS in MOSFETs at room temperature [22]. This limit is

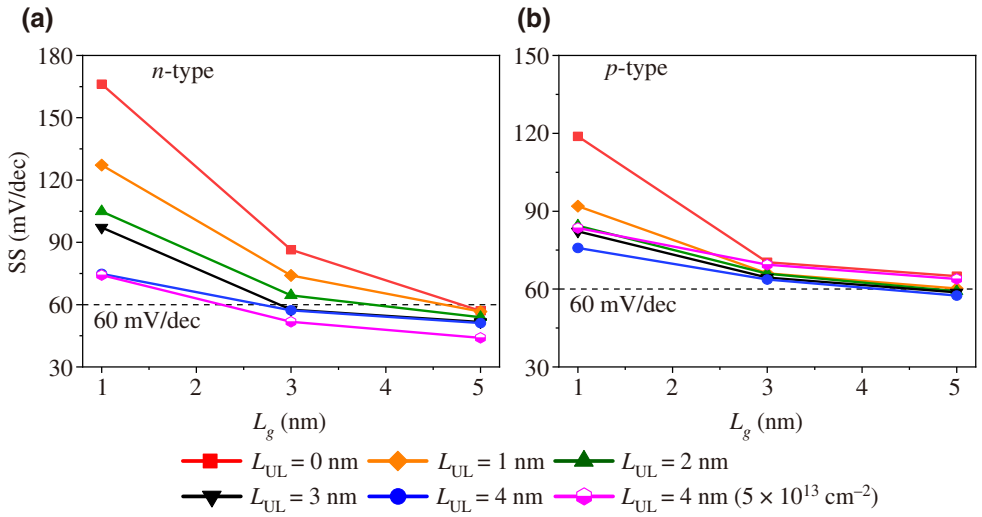

FIG. 5. SS as a function of gate length for (a) *n*- and (b) *p*-type FETs with different L_{UL}. Black dashed lines indicate the Boltzmann limit of 60 mV/dec for SS at room temperature.

suitable for classical transistors with long channels, the current of which is mainly composed of thermionic injection. In the MOSFETs with an ultrashort channel down to a few nanometers, the contribution of the tunneling current could help to make the value of SS fall below the Boltzmann tyranny. Taking both the tunneling and thermionic currents into consideration, namely, $I_{DS} = I_{\text{tunnel}} + I_{\text{therm}}$, SS can be expressed as follows [22]:

$$S = \frac{\partial V_g}{\partial (L_g I_{DS})} = \left[\frac{r_{\text{tunnel}}}{S_{\text{tunnel}}} + \frac{1 - r_{\text{tunnel}}}{S_{\text{therm}}} \right]^{-1}, \quad (3)$$

where

$$r_{\text{tunnel}} = \frac{I_{\text{tunnel}}}{I_{DS}}, \; S_{\text{tunnel}} = \frac{\partial V_g}{\partial (L_g I_{\text{tunnel}})},$$
$$S_{\text{therm}} = \frac{\partial V_g}{\partial (L_g I_{\text{therm}})}.$$

In the transistors with long channels, I_{tunnel} can be neglected, leading to $r_{\text{tunnel}} = 0$, $S = S_{\text{therm}}$, which has a fundamental limit of 60 mV/dec. In transistors with an ultrashort channel, tunneling should be considered, $r_{\text{tunnel}} \neq 0$; the SS can be below the Boltzmann tyranny when S_{tunnel} is small enough. The tunneling current is presented as $I_{\text{tunnel}} = e^{-w\sqrt{m^* \Phi_B}}$, where Φ_B is the average barrier height, and w is the width of the barrier. According to the LDOS maps, Φ_B decreases rapidly with changes to V_g, leading to a great change of I_{tunnel}. Thus, S_{tunnel} can be very small, which could help to make the value of SS below the limit of 60 mV/dec.

D. Delay time and power consumption

We additionally explore the property of switching speed in the $MoSi_2N_4$ MOSFETs, which is an essential figure of merit for a digital circuit. The switching speed can be characterized directly by the intrinsic delay time (τ), as

$$\tau = \frac{C_g V_{dd}}{I_{\text{on}}}, \quad (4)$$

where C_g the is total gate capacitance, which is defined as the sum of the channel capacitance (C_{ch}) and the gate fringing capacitance (C_f) per width. C_f is speculated to be 2 times the intrinsic channel capacitance, and C_{ch} can be calculated as

$$C_{\text{ch}} = \frac{\partial Q_{\text{ch}}}{W \partial V_g}, \quad (5)$$

with Q_{ch} being the total charge in the central region and W being the channel width. The calculated values of C_g for the *n*- and *p*-type sub-5-nm ML $MoSi_2N_4$ MOSFETs are given in Tables SI and SII within the Supplemental Material [51], respectively. It is found that C_g (0.099–0.14 for *n* type, 0.102–0.149 fF/μm for *p* type) of the $MoSi_2N_4$ MOSFETs is much smaller than either the HP (0.6 fF/μm) or LP (0.69 fF/μm) ITRS standard. Figures 6(a) and 6(b) further show the values of τ for *n*- and *p*-type sub-5-nm $MoSi_2N_4$ devices as a function of L_g. It is shown that the intrinsic delay time, τ, of both the *n*- and *p*-type devices with various gate lengths and L_{UL} can fulfill the ITRS requirement (0.423 ps) for HP devices. Also, all the intrinsic delay time, τ, values can meet the ITRS LP standard (1.493 ps) for the LP devices, except for the one in the *n*-type device with $L_g = 1$ nm and $L_{UL} = 2$ nm due to the very low I_{on} (44 μA/μm). Moreover, the value of τ under certain L_g and L_{UL} can be several (10) times smaller than the HP (LP) ITRS standard, indicating great potential applications in high-switching-speed $MoSi_2N_4$ MOSFETs.

Another significant concern for FET applications is the switching-energy cost by PDP, which can be calculated by

$$\text{PDP} = V_{dd} I_{\text{on}} \tau = C_g V_{dd}^2. \quad (6)$$

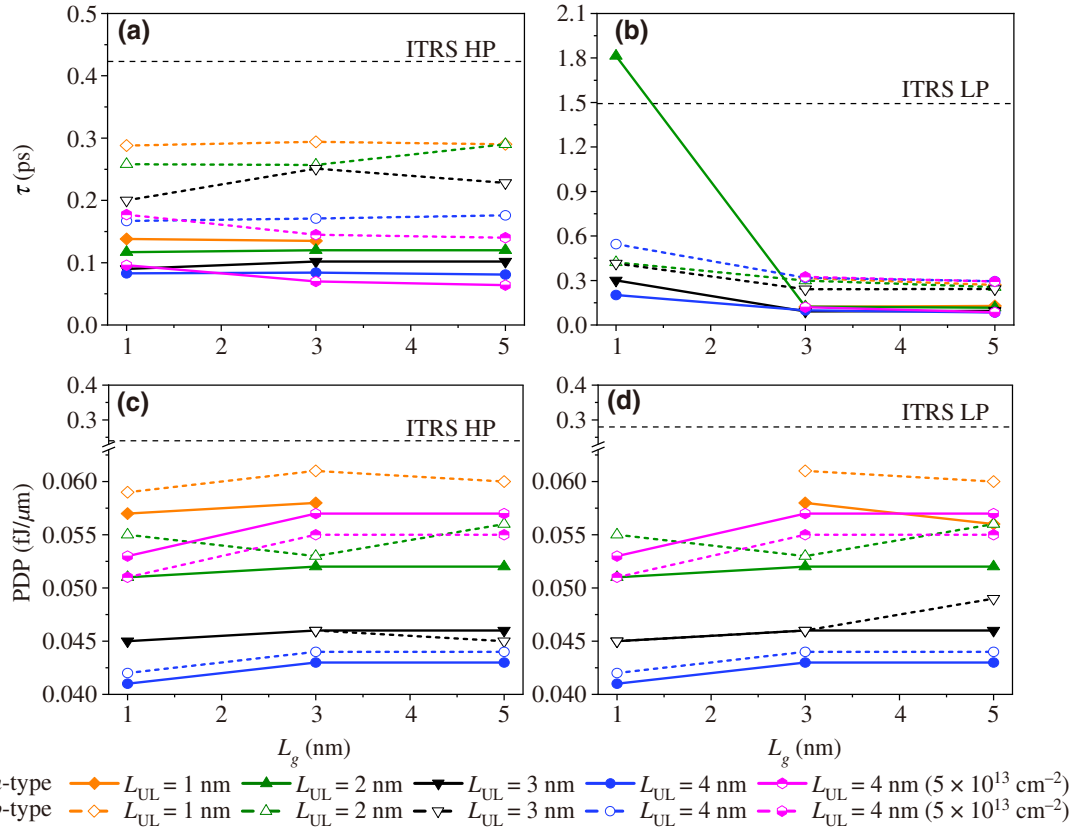

FIG. 6. Intrinsic delay time (τ) (a),(b) and PDP (c),(d) as a function of gate length for n- and p-type FETs with different L_{UL}. Black dashed lines are the ITRS HP and LP requirements for τ and PDP, respectively.

Figures 6(c) and 6(d) show the calculated PDPs of n-type and p-type sub-5-nm $MoSi_2N_4$ devices as functions of L_g, respectively. In both cases, PDP monotonously decreases with increasing L_{UL}. In addition, the PDPs of both n-type (0.023–0.101 fJ/μm) and p-type (0.026–0.121 fJ/μm) sub-5-nm MOSFETs are much lower than the ITRS requirements for HP (0.24 fJ/μm) and LP (0.28 fJ/μm) standards. This feature verifies that the $MoSi_2N_4$ MOSFET devices also have the advantage of low-power consumption.

E. Discussion

A comparison of the main parameters, namely, on-state current, subthreshold swing, delay time, and power-delay product, of ML $MoSi_2N_4$ MOSFETs and other 2D MOSFETs with $L_g \leq 5$ nm based on *ab initio* quantum-transport calculations for HP and LP devices is shown in Tables I and II, respectively. There are some 2D MOSFETs with very high on-state currents for HP applications, such as phosphorene (4500 μA/μm), BiH (2320 μA/μm), tellurene (2114 μA/μm), and arsenene (2030 μA/μm). However, they are all p-type MOSFETs and are not suitable for n-type doping, which hinders their application in complementary metal-oxide semiconductors. For the n-type MOSFETs, the performance of ML $MoSi_2N_4$ is better than that of other 2D materials, especially for the LP application.

TABLE I. Comparison of the upper-performance limit of the ML $MoSi_2N_4$ MOSFETs with other 2D MOSFETs with $L_g \leq 5$ nm for HP devices.

	Doping type	I_{on} (μA/μm)	SS (mV/dec)	τ (ps)	PDP (fJ/μm)
MoS_2 [53]	n	473	58	1.287	0.195
	p	440	46	0.396	0.096
WSe_2 [54]	p	1464	82	0.168	0.156
Bi_2O_2Se [33]	n	916	114	0.240	0.141
	p	585	96	0.375	0.140
BiH [55]	p	2320	77	0.020	0.029
Ge-Se [52]	n	518	130	0.124	0.041
	p	1703	60	0.054	0.059
Phosphorene [22]	p	4500	76	0.055	0.135
Tellurene [29]	p	2114	102	0.068	0.098
Silicane [56]	n	1374	65	0.042	0.037
	p	871	67	0.075	0.043
Arsenene [57]	p	2030	77	0.017	0.032
$MoSi_2N_4$	n	1390	44	0.064	0.057
	p	618	64	0.140	0.055

TABLE II. Comparison of the upper-performance limit of the ML $MoSi_2N_4$ MOSFETs with other 2D MOSFETs with $L_g \leq 5$ nm for LP devices.

	Doping type	I_{on} (μA/μm)	SS (mV/dec)	τ (ps)	PDP (fJ/μm)
MoS_2 [53]	*n*	324	56	0.552	0.093
	p	425	46	0.411	0.075
WSe_2 [54]	*p*	1132	63	0.149	0.108
ReS_2 [58]	*p*	329	72	0.700	0.150
BiH [55]	*p*	179	67	0.168	0.018
Ge-Se [52]	*n*	274	90	0.320	0.055
Phosphorene [22]	*p*	857	85	0.193	0.108
Tellurene [29]	*p*	451	57	0.206	0.063
Silicane [56]	*n*	467	77	0.054	0.016
	p	378	67	0.136	0.012
Arsenene [57]	*p*	341	77	0.101	0.023
$MoSi_2N_4$	*n*	1025	44	0.086	0.057
	p	355	70	0.265	0.060

Two-dimensional MoS_2 transistors have been extensively studied for many years, both experimentally and in simulations. The electronic structures of MoS_2 and $MoSi_2N_4$ are very similar. However, the performance of the ML $MoSi_2N_4$ transistor is much better than that of the ML MoS_2 transistor. The current along the transport direction is defined as $I = Nev$, where N is the number of carriers and v is the velocity of carriers and is defined as $v = \mu E$, where μ and E are the carrier mobility and electric field, respectively. The on-state current is proportional to the carrier mobility. For 2D materials, the intrinsic carrier mobility can be calculated by

$$\mu_{2D} = \frac{2e\hbar^3 C}{3k_B T|m^*|^2 E_1^2}, \tag{7}$$

where C is the elastic modulus, m^* is the effective mass, T is the temperature, and E_1 is the deformation potential (DP) constant. The elastic modulus of ML $MoSi_2N_4$ is about 4 times higher than that of MoS_2 (about 530 N/m for $MoSi_2N_4$ and 128 N/m for MoS_2, for both holes and electrons), which makes the carrier mobility of monolayer $MoSi_2N_4$ (about 1227 $cm^2\,V^{-1}\,s^{-1}$ for holes and 288 $cm^2\,V^{-1}\,s^{-1}$ for electrons) higher than that of MoS_2 (about 200 $cm^2\,V^{-1}\,s^{-1}$ for holes and 152 $cm^2\,V^{-1}\,s^{-1}$ for electrons) [34]. Therefore, the ML $MoSi_2N_4$ transistors have much higher on-state currents than that of ML MoS_2.

IV. CONCLUSION

To summarize, we explore the performance limit of sub-5-nm *n*- and *p*-type ML $MoSi_2N_4$ MOSFETs by applying precise *ab initio* quantum-transport simulations. We find that the middle MoN_2 layer of $MoSi_2N_4$ is responsible for electron transmission, and the variation of on-state current in the $MoSi_2N_4$ MOSFETs can be effectively manipulated by the length of gate and underlap, as well as the doping concentration. More importantly, a competing mechanism for the influence of underlap length on the on-state current is also found, indicating there is an optimized on-state current for a certain assembly of gate length and underlap length. In addition, we also find that for *n*-type devices the optimized on-state current can reach 1390 and 1025 μA/μm for the HP and LP applications, respectively, both of which satisfy the ITRS requirements. The optimized on-state current can meet the LP application (348 μA/μm) for *p*-type devices. Finally, we find that sub-5-nm $MoSi_2N_4$ MOSFETs can have an unusually short intrinsic delay time and low power-delay product compared with the standards of ITRS, which lead to devices with high speed and low-power consumption. Considering that $MoSi_2N_4$ is remarkably stable in air, its sub-5-nm MOSFET devices with high performance are expected to be widely realized in the near future.

ACKNOWLEDGMENTS

We are grateful for useful discussions with Prof. Ying Guo. This work is supported by the National Key R&D Program of China (Grant No. 2018YFB0407600); National Natural Science Foundation of China (Grants No. 12074301 and No. 12004295); Fundamental Research Funds for Central Universities (Grant No. xzy012019062); and Open Research Fund of Key Laboratory of Polar Materials and Devices, Ministry of Education. P.L. also thanks the China Postdoctoral Science Foundation for funding the project (Grant No. 2020M673364).

J.-S. Huang and P. Li contributed equally to this work.

[1] S. Salahuddin, K. Ni, and S. Datta, The era of hyper-scaling in electronics, Nat. Electron. **1**, 442 (2018).

[2] N. G. Orji, M. Badaroglu, B. M. Barnes, C. Beitia, B. D. Bunday, U. Celano, R. J. Kline, M. Neisser, Y. Obeng, and A. E. Vladar, Metrology for the next generation of semiconductor devices, Nat. Electron. **1**, 532 (2018).

[3] R. Chau, B. Doyle, S. Datta, J. Kavalieros, and K. Zhang, Integrated nanoelectronics for the future, Nat. Mater. **6**, 810 (2007).

[4] T. N. Theis and P. M. Solomon, It's time to reinvent the transistor!, An Emergent Change Of, Science **327**, 1600 (2010).

[5] Y. Liu, X. Duan, Y. Huang, and X. Duan, Two-dimensional transistors beyond graphene and TMDCs, Chem. Soc. Rev. **47**, 6388 (2018).

[6] D. Akinwande, C. Huyghebaert, C. H. Wang, M. I. Serna, S. Goossens, L. J. Li, H. S. P. Wong, and F. H. L. Koppens, Graphene and two-dimensional materials for silicon technology, Nature **573**, 507 (2019).

[7] Z. Ni, M. Ye, J. Ma, Y. Wang, R. Quhe, J. Zheng, L. Dai, D. Yu, J. Shi, J. Yang, S. Watanabe, and J. Lu, Performance

upper limit of sub-10 Nm monolayer MoS_2 transistors, Adv. Electron. Mater. **2**, 1600191 (2016).

[8] M. Zeng, Y. Xiao, J. Liu, K. Yang, and L. Fu, Exploring two-dimensional materials toward the next-generation circuits: From monomer design to assembly control, Chem. Rev. **118**, 6236 (2018).

[9] K. Majumdar, C. Hobbs, and P. D. Kirsch, Benchmarking transition metal dichalcogenide MOSFET in the ultimate physical scaling limit, IEEE Electron Device Lett. **35**, 402 (2014).

[10] S. Zhang, S. Guo, Z. Chen, Y. Wang, H. Gao, J. Gómez-Herrero, P. Ares, F. Zamora, Z. Zhu, and H. Zeng, Recent progress in 2D group-VA semiconductors: From theory to experiment, Chem. Soc. Rev. **47**, 982 (2018).

[11] F. Schwierz, Graphene transistors, Nat. Nanotechnol. **5**, 487 (2010).

[12] G. Fiori, A. Betti, S. Bruzzone, and G. Iannaccone, Lateral graphene-HBCN heterostructures as a platform for fully two-dimensional transistors, ACS Nano **6**, 2642 (2012).

[13] W. Zhou, J. Chen, P. Bai, S. Guo, S. Zhang, X. Song, L. Tao, and H. Zeng, Two-dimensional pnictogen for field-effect transistors, Research **2019**, 1 (2019).

[14] A. Allain, J. Kang, K. Banerjee, and A. Kis, Electrical contacts to two-dimensional semiconductors, Nat. Mater. **14**, 1195 (2015).

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

[16] L. Xie, M. Liao, S. Wang, H. Yu, L. Du, J. Tang, J. Zhao, J. Zhang, P. Chen, X. Lu, G. Wang, G. Xie, R. Yang, D. Shi, and G. Zhang, Graphene-contacted ultrashort channel monolayer MoS_2 transistors, Adv. Mater. **29**, 1702522 (2017).

[17] K. Xu, D. Chen, F. Yang, Z. Wang, L. Yin, F. Wang, R. Cheng, K. Liu, J. Xiong, Q. Liu, and J. He, Sub-10 Nm nanopattern architecture for 2D material field-effect transistors, Nano Lett. **17**, 1065 (2017).

[18] A. Nourbakhsh, A. Zubair, R. N. Sajjad, T. K. G. Amir, W. Chen, S. Fang, X. Ling, J. Kong, M. S. Dresselhaus, E. Kaxiras, K. K. Berggren, D. Antoniadis, and T. Palacios, MoS_2 field-effect transistor with sub-10 Nm channel length, Nano Lett. **16**, 7798 (2016).

[19] M. Fang, F. Wang, Y. Han, Y. Feng, T. Ren, Y. Li, D. Tang, Z. Song, and K. Zhang, Controlled growth of bilayer-MoS_2 films and MoS_2-based field-effect transistor (FET) performance optimization, Adv. Electron. Mater. **4**, 1700524 (2018).

[20] International Technology Roadmap for Semiconductors (ITRS), 2013. http://www.itrs2.net/

[21] P. H. Ho, Y. R. Chang, Y. C. Chu, M. K. Li, C. A. Tsai, W. H. Wang, C. H. Ho, C. W. Chen, and P. W. Chiu, High-mobility InSe transistors: The role of surface oxides, ACS Nano **11**, 7362 (2017).

[22] R. Quhe, Q. Li, Q. Zhang, Y. Wang, H. Zhang, J. Li, X. Zhang, D. Chen, K. Liu, Y. Ye, L. Dai, F. Pan, M. Lei, and J. Lu, Simulations of Quantum Transport in Sub-5-Nm Monolayer Phosphorene Transistors, Phys. Rev. Appl. **10**, 024022 (2018).

[23] J. D. Wood, S. A. Wells, D. Jariwala, K. S. Chen, E. Cho, V. K. Sangwan, X. Liu, L. J. Lauhon, T. J. Marks, and M. C. Hersam, Effective passivation of exfoliated black phosphorus transistors against ambient degradation, Nano Lett. **14**, 6964 (2014).

[24] W. Zhou, S. Zhang, Y. Wang, S. Guo, H. Qu, P. Bai, Z. Li, and H. Zeng, Anisotropic in-plane ballistic transport in monolayer black arsenic-phosphorus FETs, Adv. Electron. Mater. **6**, 1901281 (2020).

[25] W. Zhou, S. Zhang, S. Guo, Y. Wang, J. Lu, X. Ming, Z. Li, H. Qu, and H. Zeng, Designing Sub-10-Nm Metal-Oxide-Semiconductor Field-Effect Transistors via Ballistic Transport and Disparate Effective Mass: The Case of Two-Dimensional Bi N, Phys. Rev. Appl. **13**, 044066 (2020).

[26] J. Zhao, H. Liu, Z. Yu, R. Quhe, S. Zhou, Y. Wang, C. C. Liu, H. Zhong, N. Han, J. Lu, Y. Yao, and K. Wu, Rise of silicene: A competitive 2D material, Prog. Mater. Sci. **83**, 24 (2016).

[27] S. Zhang, M. Xie, F. Li, Z. Yan, Y. Li, E. Kan, W. Liu, Z. Chen, and H. Zeng, Semiconducting group 15 monolayers: A broad range of band gaps and high carrier mobilities, Angew. Chemie **128**, 1698 (2016).

[28] E. Zhang, Y. Jin, X. Yuan, W. Wang, C. Zhang, L. Tang, S. Liu, P. Zhou, W. Hu, and F. Xiu, ReS_2-based field-effect transistors and photodetectors, Adv. Funct. Mater. **25**, 4076 (2015).

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

[30] L. Tao, E. Cinquanta, D. Chiappe, C. Grazianetti, M. Fanciulli, M. Dubey, A. Molle, and D. Akinwande, Silicene field-effect transistors operating at room temperature, Nat. Nanotechnol. **10**, 227 (2015).

[31] L. Li, Y. Yu, G. J. Ye, Q. Ge, X. Ou, H. Wu, D. Feng, X. H. Chen, and Y. Zhang, Black phosphorus field-effect transistors, Nat. Nanotechnol. **9**, 372 (2014).

[32] S. Guo, Y. Zhang, Y. Ge, S. Zhang, H. Zeng, and H. Zhang, 2D V-V binary materials: Status and challenges, Adv. Mater. **31**, 1902352 (2019).

[33] J. Yang, R. Quhe, Q. Li, S. Liu, L. Xu, Y. Pan, H. Zhang, X. Zhang, J. Li, J. Yan, B. Shi, H. Pang, L. Xu, Z. Zhang, J. Lu, and J. Yang, Sub 10 Nm bilayer Bi_2O_2Se transistors, Adv. Electron. Mater. **5**, 1800720 (2019).

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

[35] Q. Wu, L. Cao, Y. S. Ang, and L. K. Ang, Semiconductor-to-Metal transition in bilayer $MoSi_2N_4$ and WSi_2N_4 with strain and electric field, Appl. Phys. Lett. **118**, 113102 (2021).

[36] S. D. Guo, Y. T. Zhu, W. Q. Mu, and W. C. Ren, Intrinsic piezoelectricity in monolayer MSi_2N_4 (M = Mo, W, Cr, Ti, Zr and Hf), EPL **132**, 57002 (2020).

[37] H. Zhong, W. Xiong, P. Lv, J. Yu, and S. Yuan, Strain-induced semiconductor to metal transition in MA2Z4 bilayers (M = Ti, Cr, Mo; A = Si; Z = N, P), Phys. Rev. B **103**, 085124 (2021).

[38] Q. Wang, L. Cao, S.-J. Liang, W. Wu, G. Wang, C. H. Lee, W. L. Ong, H. Y. Yang, L. K. Ang, S. A. Yang, and Y. S. Ang, Efficient ohmic contacts and built-in atomic sublayer protection in $MoSi_2N_4$ and WSi_2N_4 monolayers, Npj 2D Mater. Appl. **5**, 71 (2021).

[39] L. Cao, G. Zhou, Q. Wang, L. K. Ang, and Y. S. Ang, Two-dimensional van Der Waals electrical contact to monolayer $MoSi_2N_4$, Appl. Phys. Lett. **118**, 013106 (2021).

[40] D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B - Condens. Matter Mater. Phys. **59**, 1758 (1999).

[41] G. Kresse and J. Hafner, Ab initio molecular dynamics for liquid metals, Phys. Rev. B **47**, 558 (1993).

[42] R. A. Vargas-Hernández, Bayesian optimization for calibrating and selecting hybrid-density functional models, J. Phys. Chem. A **124**, 4053 (2020).

[43] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. **77**, 3865 (1996).

[44] H. J. Monkhorst and J. D. Pack, Special points for brillouin-zone integrations, Phys. Rev. B **13**, 5188 (1976).

[45] M. Brandbyge, J. L. Mozos, P. Ordejón, J. Taylor, and K. Stokbro, Density-Functional method for nonequilibrium electron transport, Phys. Rev. B - Condens. Matter Mater. Phys. **65**, 1654011 (2002).

[46] QuantumATK Q-2019.12, Synopsys QuantumATK, https://www.synopsys.com/silicon/quantumatk.html

[47] S. Datta and H. van Houten, Electronic transport in mesoscopic systems, Phys. Today **49**, 70 (1996).

[48] H. Zhong, R. Quhe, Y. Wang, Z. Ni, M. Ye, Z. Song, Y. Pan, J. Yang, L. Yang, M. Lei, J. Shi, and J. Lu, Interfacial properties of monolayer and bilayer MoS_2 contacts with metals: Beyond the energy band calculations, Sci. Rep. **6**, 21786 (2016).

[49] Y. Wang, R. X. Yang, R. Quhe, H. Zhong, L. Cong, M. Ye, Z. Ni, Z. Song, J. Yang, J. Shi, J. Li, and J. Lu, Does P-type ohmic contact exist in WSe_2-metal interfaces?, Nanoscale **8**, 1179 (2016).

[50] A. H. D. Cheng and D. T. Cheng, Heritage and early history of the boundary element method, Eng. Anal. Bound. Elem. **29**, 268 (2005).

[51] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevApplied.16.044022 for specific data on the transport properties of FETs and the band structure of monolayer $MoSi_2N_4$.

[52] Y. Guo, F. Pan, G. Zhao, Y. Ren, B. Yao, H. Li, and J. Lu, Sub-5 nm monolayer germanium selenide (GeSe) MOSFETs: towards a high performance and stable device, Nanoscale **12**, 15443 (2020).

[53] 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).

[54] 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).

[55] 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).

[56] Y. Pan, J. Dai, L. Xu, J. Yang, X. Zhang, J. Yan, J. Li, B. Shi, S. Liu, H. Hu, M. Wu, and J. Lu, Sub-5-nm Monolayer Silicane Transistor: A First-Principles Quantum Transport Simulation, Phys. Rev. Appl. **14**, 024016 (2020).

[57] 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).

[58] R. Quhe, J. Chen, and J. Lu, A sub-10 nm monolayer ReS_2 transistor for low-power applications, J. Mater. Chem. C **7**, 1604 (2019).